REVIEWS OF INFECTIOUS DISEASES • VOL. 12, NO. 1 • JANUARY-FEBRUARY 1990
 0612-0882/90/1201-0029$02.00

Global Genetic Structure and Molecular Epidemiology of Encapsulated *Haemophilus influenzae*

James M. Musser,* J. Simon Kroll, Dan M. Granoff, E. Richard Moxon, Bernard R. Brodeur, Jose Campos, Henri Dabernat, Wilhelm Frederiksen, Josee Hamel, Gregory Hammond, E. Arne Høiby, Kristin E. Jonsdottir, Mustafa Kabeer, Ingegerd Kallings, Waheed N. Khan, Mogens Kilian, Kathleen Knowles, H. J. Koornhof, Barbara Law, Karl I. Li, Janet Montgomery, Patricia E. Pattison, Jean-Claude Piffaretti, Aino K. Takala, Mee Len Thong, Robert A. Wall, Joel I. Ward, and Robert K. Selander

From the Department of Biology, Mueller Laboratory, Pennsylvania State University, University Park; Infectious Disease Unit, Department of Paediatrics, John Radcliffe Hospital, Headington, Oxford; Edward Mallinckrodt Department of Pediatrics, Washington University School of Medicine, St. Louis; Hybridoma Section, Bureau of Microbiology, Laboratory Centre for Disease Control, Health and Welfare Canada, Ottawa; Departments of Microbiology and Pediatrics, Hospital Infantil San Juan de Dios, Barcelona; Laboratoire Central de Microbiologie, C. H. U. Toulouse Purpan, Toulouse; Statens Seruminstitut, Copenhagen; Cadham Provincial Laboratory, Winnipeg; National Institute of Public Health, Oslo; Department of Bacteriology, University of Iceland, Reykjavik; The National Bacteriology Laboratory, Stockholm; Children's Hospital National Medical Center, Washington, D.C.; Department of Oral Biology, Royal Dental College, Aarhus; Montreal Children's Hospital, Montreal; South African Institute for Medical Research, Johannesburg; Department of Pediatrics and Child Health, University of Manitoba, Winnipeg; Children's Hospital of Pittsburgh, Pennsylvania; Papua New Guinea Institute of Medical Research, Goroka; Instituto Cantonale Batteriosierologico, Lugano; National Public Health Institute, Helsinki; Division of Microbiology, Department of Pathology, North Brisbane Hospitals Board, Brisbane; Medical Research Council Laboratories, Faraja; Department of Pediatrics, Harbor-UCLA Medical Center, University of California, Los Angeles

A collection of 2,209 isolates of six polysaccharide capsule types of *Haemophilus influenzae*, including 1,975 serotype b isolates recovered in 30 countries was characterized for electrophoretically demonstrable allele profiles at 17 metabolic enzyme loci. Two hundred eighty distinct multilocus genotypes were distinguished, and cluster analysis revealed two primary phylogenetic divisions. The population structure of encapsulated *H. influenzae* is clonal. Currently, most of the invasive disease worldwide is caused by serotype b strains of nine clones. Strains producing serotype c, e, and f capsules belong to single divisions and have no close genetic relationships to strains of other serotypes. Serotype a and b strains occur in both primary phylogenetic divisions, probably as a result of transfer and recombination of serotype-specific sequences of the *cap* region between clonal lineages. A close genetic relatedness between serotype d isolates and some strains of serotypes a and b was identified. There are strong patterns of geographic variation, on an intercontinental scale, in both the extent of genetic diversity and the clonal composition of populations of encapsulated strains. The analysis suggests that the present distribution of clones is, in part, related to patterns of racial or ethnic differentiation and historical demographic movements of the human host populations.

Received for publication 22 February 1989 and in revised form 13 June 1989.

Research leading to this review was supported by grant no. AI24332 from the National Institute of Allergy and Infectious Diseases (to R. K. S. and D. M. G.), by contract AID-2429-8-312 from the Board on Science and Technology for International Development (BOSTID), National Research Council (to D. M. G.), by a Program Grant from the Medical Research Council, U.K. (to E. R. M.), and by grant 85.3.16.8.E from the Caisse Nationale d'Assurances Maladie des Travailleurs Salaries (to H. D.). J. S. K. is a Lister Institute Research Fellow.

This paper is dedicated to Dr. Margaret Pittman.

C. M. Sommers and L. M. Tremblay provided technical assistance. The authors thank T. S. Whittam for help with data analysis. The senior author thanks K. A. Zeznick for extensive encouragement.

* Present affiliation: Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania.

Please address request for reprints to Dr. Robert K. Selander, Department of Biology, Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802.

Encapsulated strains of *Haemophilus influenzae* are a major cause of meningitis and other serious invasive diseases, including septicemia, obstructive epiglottitis, cellulitis, and septic arthritis, in young children in many parts of the world [1–3]. Because of their great public health importance, encapsulated strains, especially those expressing the serotype b polysaccharide, have been the subject of intensive research in the United States and Europe during the past 10 years. Early immunologic and metabolic studies demonstrated heterogeneity among strains [4–6], but attempts to understand the epidemiology of disease caused by encapsulated organisms initially were hindered by the lack of high-resolution, stable genetic marker systems for strain discrimination and classification.

Early in the present decade, techniques were developed for classifying serotype b isolates on the basis of variation in the electrophoretic mobility patterns of the major outer-membrane proteins (OMPs) [7–10] and lipopolysaccharide [11, 12], serologic diversity in lipopolysaccharide antigens [9, 13, 14], and other phenotypic characters [15–17]. An examination of 51 serotype b isolates recovered from children in St. Louis with invasive infections identified nine distinctive OMP pattern subtypes and demonstrated that five subtypes, designated as 1H, 1L, 2H, 2L, and 3L, accounted for 92% of the strains [18]. In a survey of 256 invasive isolates from a variety of clinical settings in 22 states in the United States, Granoff et al. [19] distinguished 21 OMP subtypes and discovered that about 70% of cases were caused by strains of three subtypes (1H, 2L, and 3L). Subsequently, van Alphen et al. [9] examined 80 isolates from invasive infections in the Netherlands where, in striking contrast to those in the United States, 84% of isolates had the same OMP pattern (type 1; subtype 3L in the Granoff laboratory system); moreover, no strains of 1H, 1L, or 2H were identified.

In an effort to measure genetic diversity and evolutionary genomic relationships among serotype b isolates classified by OMP type and biotype, Musser et al. [20] analyzed genetically determined electrophoretic variation in 16 metabolic enzymes in 177 isolates from the United States. They identified 32 distinctive multilocus enzyme genotypes (electrophoretic types, ETs) [20] and determined that 73% of invasive disease episodes were caused by strains of only three ETs. This work also demonstrated that serotype b capsule is expressed in strains belonging to four rather distantly related groups of chromosomal genotypes, each of which is associated with a characteristic set of OMP patterns [20]. To explain the occurrence of strong nonrandom associations of multilocus enzyme genotype, OMP subtype, and biotype, and the repeated recovery of isolates with identical properties in widely separated geographic regions and over a 40-year period, researchers hypothesized [20] and subsequently confirmed [21, 22] that the structure of natural populations of serotype b *H. influenzae* is basically clonal as a consequence of infrequent recombination of chromosomal genes.

Two important characteristics of the epidemiology of serotype b *H. influenzae* were identified in these studies. First, although there is extensive genetic diversity among strains expressing type b capsule, most disease is caused by a very small number of clonal chromosomal genotypes, which are marked by ETs. Second, comparison of results from studies in the United States and the Netherlands suggested the existence of strong patterns of geographic variation, on an intercontinental scale, in both the extent of genetic diversity and the clonal composition of populations.

Little is known about the genetics of populations of *H. influenzae* in areas of the world other than the United States and western Europe [23]. To study genetic structure and molecular epidemiology on a global scale, we assembled a large collection of encapsulated *H. influenzae* strains recovered in many countries over the past 40 years. One objective was to generate baseline data that would permit prospective assessment of the effects of vaccination on the genetic diversity and structure of populations of invasive and carrier strains. Additionally, we wanted to examine possible correlations between patterns of geographic variation in genetic diversity and clonal composition of the serotype b populations and patterns of human racial/ethnic differentiation and historical demographic movement. For immunoprophylactic research, development, and application, it may be important to understand the global clone distribution of type b strains.

We here present a comprehensive analysis of the genetic structure and molecular epidemiology of encapsulated *H. influenzae*, based on 2,209 isolates from six continents. The population genetic analysis is based primarily on electrophoresis of 17 chromosomally encoded metabolic enzymes, with additional information being supplied by OMP subtyp-

ing [8, 24, 25] and the restriction fragment length polymorphism (RFLP) pattern of the *cap* region [24, 26].

Materials and Methods

Bacterial Isolates

A collection of 2,209 isolates of encapsulated *H. influenzae* recovered from individuals in 30 countries on six continents was examined (table 1). The sample includes 52 isolates of serotype a, 1,975 of serotype b, 13 of serotype c, 27 of serotype d, 92 of serotype e, and 50 of serotype f. Of the serotype b strains, 1,814 were cultured from blood, cerebrospinal fluid, or other normally sterile body fluid, 91 from the nasopharynx of healthy individuals, 54 from severe lower respiratory tract infections (predominantly in Malaysia), and six from the eyes of patients with conjunctivitis. Most isolates of other serotypes were recovered from asymptomatic carriers or from patients with surface infections. However, 11 isolates of serotype a, six of serotype f, and two each of serotypes d and e were cultured from patients with invasive infections. The collection includes 24 isolates obtained between 1939 and 1954 [27], but most of the isolates were recovered between 1965 and 1987.

Of the carrier isolates, 46 were recovered from associated children in three day care centers in Spain [28], 10 were from Amerinds living on the Fort Apache Indian Reservation in Arizona, 20 were from Alaskan Natives in a variety of areas in the state, nine were from associated South Korean children taken to Norway for adoption, and six were from unknown sources in the United States and Canada.

Most isolates were obtained from a collection assembled by the senior author and from collections maintained by the coauthors. In addition, the following individuals each supplied a small number of isolates: J. O. Achola, University of Nairobi, Kenya; M. Arpi, Statens Seruminstitut, Copenhagen, Denmark; M. Catalano, National Research Council, Buenos Aires, Argentina; B. W. Catlin, Medical College of Wisconsin, Milwaukee, Wisconsin; P. D. Ellner, Columbia-Presbyterian Medical Center, New York, New York; R. J. Fallon, Ruchill Hospital, Glasgow, Scotland; P. Fleming, Hospital for Sick Children, Toronto, Ontario, Canada; E. J. Hansen, University of Texas Health Science Center, Dallas, Texas; H. Heffernan, National Health Institute, Wellington, New Zealand; T. Konda, National Institute of Health, Tokyo, Japan; J. Levy, Rochester General Hospital, Rochester, New York; M. Loeb, University of Rochester School of Medicine and Dentistry, Rochester, New York; M. Magdasy, Central Public Health Laboratory, Montevideo, Uruguay; K. McGowan, St. Christopher's Hospital, Philadelphia, Pennsylvania; T. Oguri, Juntendo University, Tokyo, Japan; S. J. Oppenheimer, Hospital Universiti Sains Malaysia, Kelantan, Malaysia; T. H. Pennington, University of Aberdeen, Scotland; C. A. Reichart, Johns Hopkins Medical Institutions, Baltimore, Maryland; S. H. Sell, Vanderbilt University Medical School, Nashville, Tennessee; A. L. Smith, Children's Orthopedic Hospital and Medical Center, Seattle, Washington; M. Soraekit, Mahidal University, Thailand; J. Spainhour, Children's Hospital of Philadelphia, Philadelphia, Pennyslvania; Y. Terawaki, Shinshu University School of Medicine, Matsumoto, Japan; T. Tupasi, Makati Medical Center, Philippines; M. K. Wagner, Kapiolani Women's and Children's Medical Center, Honolulu, Hawaii; L. B. Weiner, Upstate Medical Center, Syracuse, New York; and R. Yogev, Children's Memorial Hospital, Chicago, Illinois.

Electrophoresis of Enzymes

Isolates were grown overnight at 37°C in 150 mL of brain-heart infusion broth (Difco; Detroit, Mich.), supplemented with NAD and hemin, each at 2 mg/mL. Cells were harvested by centrifugation, suspended in 2 mL of 50 m*M* Tris-hydrochloride, pH 8.0, containing 5 m*M* EDTA, and sonicated (Branson Sonifier Cell Disruptor, model 200, with microtip; Danbury, Conn.) for 30 seconds at 50% pulse, with ice-bath cooling. After centrifugation at 20,000 $\times$ *g* for 20 minutes at 4°C, the clear supernatant (lysate) was stored at −70°C.

Techniques of horizontal starch-gel electrophoresis and the demonstration of specific enzyme activity were similar to those described elsewhere [20, 29, 30]. The 17 enzymes assayed were carbamylate kinase (CAK), nucleoside phosphorylase (NSP), phosphoglucose isomerase (PGI), malic enzyme (MAE), malate dehydrogenase (MDH), glucose-6-phosphate dehydrogenase (G6P), glutamic oxaloacetic transaminase (GOT), adenylate kinase (ADK), 6-phosphogluconate dehydrogenase (6PG), leucylalanine peptidase 1 (PE1), leucylalanine peptidase 2 (PE2),

Table 1. Composition of the sample of encapsulated *H. influenzae*, grouped by geographic source.

Geographic source	Collection period	No. of isolates: Serotype b	No. of isolates: Other	No. of ETs: Serotype b	No. of ETs: Other
North America		936		104	
Canada	1969–1986	376[a]		40	
United States	1939–1954,				
	1968–1987	560[b]	12	77	10
Europe		623		60	
Denmark[c]	1940–1942,				
	1980–1986	46		18	
England	1983–1985	21	149	8	52
Finland	1985	100[d]		12	
France	1980s	77[e]		10	
Iceland	1977–1986	40[f]		6	
The Netherlands	1975–1982	8		4	
Norway	1980–1985	39		11	
Scotland	1983–1986	20		7	
Spain	1980s	71[g]		3	
Sweden	1982–1986	78[h]		13	
Switzerland	1982–1986	123[i]		18	
Asia		120		17	
Japan	1981–1985	12[j]		3	
Malaysia	1971–1979	80[k]	39	11	19
Philippines	1986	9		4	
South Korea	1985	9[l]		3	
Thailand	1986	10		4	
Africa		93		27	
The Gambia	1983–1984	30[m]	5	6	2
Ghana	1983–1984	5		3	
Kenya	1980s	10	9	4	8
Rep. of South Africa	1984–1986	48[n]		18	
Papua New Guinea	1980–1985	60	4	9	4
Other		143		29	
Argentina	1986	4[o]		2	
Australia	1984–1986	40[p]		10	
Guatemala	1986	1		1	
Hawaii	1985–1986	22		11	
Mexico	1986	1[q]		1	
New Zealand	1985–1986	19[r]		5	
Dominican Republic	1980	55	2	14	2
Uruguay	1986	1[s]		1	
Unknown	. . .		14		8
Total		1,975	234	182	98

[a] Alberta (26), Ontario (81), Quebec (84), Manitoba (145), New Brunswick (6), Prince Edward Island (26), and British Columbia (8).

[b] Alabama (1), Alaska (88), Arizona (34), California (1), Colorado (2), Florida (1), Illinois (11), Indiana (1), Louisiana (1), Maryland (31), Massachusetts (1), Minnesota (45), Missouri (7), New York (78), North Carolina (2), Oklahoma (52), Pennsylvania (136), Rhode Island (1), Tennessee (13), Texas (27), Washington, D.C. (1), Washington state (10), Wisconsin (16), Hawaii excluded.

[c] Includes Greenland.

[d] Espoo (1), Helsinki (19), Joensuu (9), Jyvaskyla (6), Kajaani (1), Kemi (1), Kokkola (3), Kotka (2), Kuopio (9), Lappeenranta (3), Oulu (18), Pori (3), Porvoo (1), Rovaniemi (2), Savonlinna (2), Tammisaari (2), Turku (7), EPKS (6), PHKS (2), VAKS (1), and unknown (2).

[e] Aix en Provence (3), Bordeux (2), Clermont Ferrand (3), Colombes (1), Compiegne (2), Creteil (10), Le Mans (2), Longjumeau (2), Mulhouse (2), Nantes (2), Nice (2), Reims (5), Rennes (3), Strasbourg (11), Toulouse (21), Versailles (4), and Villeneuve Saint Georges (2).

[f] Akureyri (3), Breiddalsvik (1), Budardalur (1), Djupivogur (1), Egilsstadir (1), Eskifjordur (1), Gardabaer (1), Hafnarfjordur (1), Husavik (1), Keflavik (2), Keflavik (U.S.) naval base (1), Kopavogur (2), Reykjavik (23), and Vestmannaeyjar (1).

[g] Barcelona area (47), Catalonia (18), northern Spain (3), southern Spain (3).

[h] Boden (1), Eskilstuna (4), Galve (1), Jonkoping (28), Kalmar (2), Karlskrona (1), Linkoping (2), Norrkoping (2), Skovde (3), Stockholm (19), Umea (1), Vasteras (4), Vastervik (2), Vaxdo (2), Visby (3), and unknown (3).

[i] Bern (24), Geneva (55), Lausanne (15), Lugano (1), San Gall (16), Winterthur (2), Zurich (10).

[j] Matsumoto (7), Tokyo (5).

[k] Kuala Lumpur.

[l] Isolated from nose or throat of adopted South Korean children minutes after arrival at Fornebu Airport, Oslo, Norway.

[m] Many areas throughout the country.

[n] Cape Town (7), Durban (2), Johannesburg (32), unknown (7).

[o] Buenos Aires.

[p] Brisbane (38), Melbourne (2).

[q] Mexico City.

[r] Auckland (7), Christchurch (1), Hastings (4), Nelson (1), Palmerston North (2), Taumarunui (1), Wellington (2), Whakatane (1).

[s] Montevideo.

leucine aminopeptidase (LAP), phosphoglucomutase (PGM), catalase (CAT), glutamate dehydrogenase (GLD), glyceraldehyde-3-phosphate dehydrogenase (G3P), and fumarase (FUM).

Electromorphs (allozymes) of each enzyme were equated with alleles at the corresponding structural gene locus, and distinctive combinations of alleles over the 17 enzyme loci, representing multilocus chromosomal genotypes, were designated as electrophoretic types (ETs) [30].

Serotyping

Serotypes were determined by slide agglutination with serotype-specific sera or by the antiserum agar method. Strains of serotype b that were of ETs not previously identified [20, 29] were reserotyped in the laboratory of D. M. G. by countercurrent immunoelectrophoresis with specific type b rabbit antisera, prepared under contract for the Institute of Medicine. Isolates with multilocus enzyme genotypes that were very different from those of other isolates of the same serotype were reserotyped and tested for *cap* region pattern in the laboratory of E. R. M.

Electrophoresis of OMPs

Isolates of serotype a were typed previously for OMP pattern by Allan et al. [25]. Some serotype b isolates also were typed previously [24] and the OMP patterns of additional serotype b isolates were determined in the laboratories of J. I. W., E. R. M., and K. I. L. by the methods of Barenkamp et al. [8] or Loeb et al. [7], with standards supplied by D. M. G. In brief, electrophoresis was performed on detergent-soluble outer-membrane derivatives in an 8%–17.5% Laemmli linear gradient polyacrylamide gel system [31]. The serotype a isolates were assigned to categories on the basis of the gradient gel results and were further classified as H, L, or U, depending upon the electrophoretic mobility of a heat-modifiable protein (P1, with an apparent molecular mass of ~45,000 [25]) on 11% acrylamide gels [32]. For serotype b strains, the heat-modifiable protein P1 has three electrophoretic variants, designated H (50 kDa), L (49 kDa), and U (<49 kDa), based on mobility in the Laemmli gel system. Because a detailed comparison of the OMP patterns of serotype a and serotype b strains was not made, pattern designations are not necessarily cognate between serotypes.

RFLP Analysis of the *cap* Region

Isolates were typed for *cap* region RFLP pattern by the methods described elsewhere [24, 26]. Briefly, chromosomal DNA was digested with restriction endonuclease *Eco*RI, and fragments were resolved according to size on agarose gels, transferred to nitrocellulose filters, and probed with radiolabeled pUO38, a cloned fragment of serotype b *H. influenzae* chromosomal DNA carrying genes involved in capsule synthesis [33]. The hybridization pattern was visualized by autoradiography. Strains were selected to include ETs representing the breadth of genotypic diversity and geographic origin in each of the major lineages of encapsulated *H. influenzae*.

Statistical Analysis

Genetic diversity at an enzyme locus among ETs or isolates was calculated as $h = (1 - \Sigma x_i^2)[n/(n - 1)]$, where x_i is the frequency of the ith allele and n is the number of ETs or isolates in a sample [34]. Mean genetic diversity (H) is the arithmetic average of h values for all loci.

Genetic distance between pairs of ETs was expressed as the proportion of enzyme loci at which different alleles were represented (mismatches), and clustering of ETs was performed from a matrix of pairwise genetic distances by the average-linkage method [30, 35].

Results

Genetic and Genotypic Diversity

All 17 enzyme loci assayed were polymorphic for from three (G3P) to 11 (PGI) alleles encoding electrophoretically distinctive variants, with an average of 6.4 alleles per locus. The failure of extracts of some isolates to show activity for one or two enzymes was attributed to the presence of null alleles, the highest frequency of which occurred at the NSP locus (2.8% of the isolates). A total of 280 distinctive multilocus genotypes (ETs) was identified (table 2), among which mean genetic diversity per locus (H) was 0.467 (table 3). Most ETs (63%) were represented by single isolates, but 103 ETs (37%) had multiple isolates (range, 2–497 isolates). The number of ETs per serotype ranged from seven for serotype d isolates to 182 for serotype b isolates. There was no sharing of ETs between serotypes.

Table 2. Allele profiles and representative isolates of 280 ETs of encapsulated *H. influenzae*.

Group	ET	MHAEM no.*	Allele at indicated enzyme locus																	Sero-type	OMP type	RFLP pattern	Geographic source	Year of isolation	Other designation
			CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM						
A1a	1.9	712	9	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	1H	G	United States	1980	SLCH 668
A1a	1.9	513	9	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L	V	United States	1982	RM1130
A1a	1.9	1769	9	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L	S	Iceland	1984	B-7184
A1a	2	1847	9	5	1	5	5	3	3	2	2	4	5	2	4	2	4	2	2	b	1H		Canada	1980s	LAW 10
A1a	3	2910	9	5	1	5	5	3	1	2	2	4	5	2	4	2	4	2	2	b			United States		LOEB S138
A1a	4	2368	9	5	1	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b	1L		United States	1980s	HARBOR 117
A1a	5	1081	9	5	1	5	5	3	5	2	4	4	5	2	4	2	4	2	2	b	1H	G	United States	1970s	SLCH 1338
A1a	6	1065	9	5	1	5	5	3	6	2	4	4	5	2	4	2	4	2	2	b	1L	G	United States	1950s	EAGAN
A1a	7	2383	9	5	1	5	5	3	6	2	3	4	5	2	4	2	4	2	2	b	1L		United States	1980s	HARBOR 185
A1a	1.6	3078	6	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L		Alaska	1980s	JW 32
A1a	1.7	725	7	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b		V	Australia	1985	RM7118
A1a	1.8	3079	8	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	1L		Norway	1980s	D 143
A1a	8	1811	8	5	1	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b	3L	S	Kenya	1980s	RM7430
A1a	9	1992	9	5	1	5	5	3	5	3	2	4	5	2	4	2	4	2	2	b			Canada	1980s	LAW 155
A1a	10	3081	9	5	1	5	5	3	5	2	2	4	5	2	6	2	4	2	2	b			United States	1983	RGH 4649
A1a	11	1074	9	5	1	5	5	1	5	2	2	4	5	2	4	2	4	2	2	b	1L		United States		SLCH 3279
A1a	13	2949	9	5	1	5	5	6	5	2	2	4	5	2	4	2	4	2	2	b	1H		United States	1986	SLCH 10,026A
A1a	14	3080	9	5	1	5	5	3	5	2	2	4	5	2	4	3	4	2	2	b	1H		Alaska	1980s	JW 5
A1a	15	3077	9	5	1	5	2	3	5	2	2	4	5	2	4	2	4	2	2	b	2L		United States	1985	PITT 8524
A1a	16	1253	9	5	1	5	3	3	5	2	2	4	5	2	4	2	4	2	2	b			Canada	1980s	PERRIER
A1a	17	1982	9	5	1	5	5	3	5	2	2	4	2	2	4	2	4	2	2	b			Canada	1980s	LAW 145
A1a	18	3082	9	5	1	3	5	3	5	2	2	4	5	6	4	2	4	2	2	b	2L		Alaska	1980s	JW 11
A1a	19	3083	9	5	1	3	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L		Alaska	1980s	JW 36
A1a	20	3732	9	5	1	5	5	3	5	2	2	4	5	6	4	2	4	2	2	b	1H		United States	1984	PITT 8417
A1a	22	3733	9	5	1	5	5	3	5	2	2	4	5	3	4	2	4	2	2	b	2L		United States	1984	PITT 8430
A1a	23	2793	5	5	1	5	5	3	5	2	2	4	5	3	4	2	4	2	2	b			Scotland	1984	PEN† 7866
A1a	24	2958	9	5	5	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L		United States	1986	SLCH 4691b
A1a	26	1081	9	5	9	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	1H	G	United States	1983	SLCH 3765
A1a	27	2932	9	5	9	5	5	3	5	2	2	2	5	2	4	2	4	2	2	b	1L‡		United States	1985	SLCH 4996A
A1a	28	250	9	5	1	5	5	3	5	2	1	4	5	2	4	2	4	2	2	b			United States	1980s	
A1a	29	1079	9	5	1	5	5	3	5	2	1	0	5	2	4	2	4	2	2	b	2L	V	United States	1982	SLCH 3718
A1a	30	1052	9	5	1	5	5	3	5	2	1	2	5	2	4	2	4	2	2	b	9L	V	United States	1980	SLCH 628B
A1a	31	562	9	5	1	5	5	3	5	2	2	2	5	2	4	2	4	2	2	b		G	England	1985	RM6107
A1a	32	1054	9	0	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2H		United States	1980	SLCH 851b
A1a	33	1868	9	4	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b			United States	1980s	LAW 31
A1a	34	1080	9	0	1	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b	1L	G	United States	1983	SLCH 3752
A1a	35	1358	8	0	1	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b			South Africa	1980s	JOH§ 39746
A1b	36	1075	9	5	1	5	5	3	3	2	1	4	5	3	4	2	4	2	2	b	1L	G	United States	1980s	SLCH 3438
A1c	37	1200	9	5	4	5	5	3	5	2	2	4	5	2	4	2	4	2	2	b	2L		Canada	1980s	HAMEL QUE 167
A1c	38	1083	9	5	4	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b	1U		United States	1984	SLCH 3805
A1c	39	3093	5	5	4	5	5	3	6	2	2	4	5	2	4	2	4	2	2	b	3L	S	Norway	1970s	KILIAN 10390
A1c	40	1126	5	5	4	5	5	3	1	2	2	4	5	2	4	2	4	2	2	b	3L		Spain	1980s	R 1713
A1c	41	2069	5	5	4	5	5	3	5	2	2	4	5	6	4	2	4	2	2	b			France	1980s	DABERNAT 1762
A1c	42	977	5	5	4	5	5	3	5	2	2	4	5	3	4	2	4	2	2	b	3L		Switzerland	1985	N 76
A1c	43	945	5	5	4	5	5	7	5	2	2	4	5	3	4	2	4	2	2	b	11L		Switzerland	1985	N 108
A1c	44	3084	9	5	1	5	5	3	6	2	2	4	5	3	4	2	4	2	2	b	1L		United States	1984	PITT 8434
A1c	45	1477	8	5	1	5	5	3	6	2	2	4	5	3	4	2	4	2	2	b	1L		The Gambia	1982	WALL 9
A1c	46	2325	8	5	4	5	5	3	6	2	2	4	5	3	4	2	4	2	2	b			Australia	1980s	RBH 24
A1c	47	2976	7	5	4	5	5	3	3	2	2	4	5	3	4	2	4	2	2	b			Scotland	1986	PEN† 3943
A1c	48	1176	7	5	4	5	5	3	6	2	2	4	5	3	4	2	4	2	2	b	26L		Canada	1980s	HAMEL C 126/83
A1d	49	2957	9	5	1	5	5	7	3	2	2	4	5	2	4	2	4	2	2	b	1H		United States	1986	SLCH 4630C
A1d	50	701	8	5	1	5	5	7	3	2	2	4	5	2	4	2	4	2	2	b		V	Ghana	1983	RM7020
A1d	51	2422	8	5	4	5	5	6	3	2	2	4	5	2	4	2	4	2	2	b			Scotland	1980s	
A1e	52	812	8	5	1	5	1	6	3	2	2	4	5	3	4	2	4	2	2	b			Ghana	1980s	RM7017
A1f	53	2819	6	5	1	5	5	3	6	2	2	4	4	6	4	2	4	2	2	b		V	Denmark	1944	ENGBAEK S41
A2a	54	3085	8	5	1	5	5	3	6	2	2	4	5	6	4	2	4	2	2	b	1L		Canada	1984	HAM 7
A2a	55	1964	8	5	4	5	5	3	6	2	2	4	5	6	4	2	4	2	2	b			Canada	1980s	LAW 127
A2a	56	1491	8	5	4	5	5	3	5	2	2	4	5	6	4	2	4	2	2	b			The Gambia	1983	WALL 40
A2a	57	1072	8	5	4	5	5	3	1	2	2	4	5	6	4	2	4	2	2	b	3L		United States	1980s	SLCH 3254
A2a	58	1653	8	5	1	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	16L		Malaysia	1971	SN0053
A2a	12.8	1666	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		United States	1980	SLCH 3010
A2a	12.8	1064	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	1L	S	Thailand	1979	SLCH 1493A
A2a	59	2141	8	5	4	5	5	3	3	2	2	4	5	3	4	2	4	2	2	b	1L		Thailand	1980s	SLCH 8035
A2a	60	1800	8	5	4	5	5	3	3	2	2	4	5	5	4	2	4	2	2	b	3L	S	Kenya	1980s	RM7419
A2a	61	2821	8	5	4	5	5	8	3	2	2	4	5	6	4	2	4	2	2	b			Denmark	1983	ARPI 20
A2a	62	1071	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	1	2	b	3L		United States	1941	SLCH 3207
A2a	63	3086	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	3	2	b	3L		Norway	1980s	KILIAN 5687

(*continued*)

Table 2. (*continued*)

Group	ET	MHAEM no.*	Allele at indicated enzyme locus: CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM	Sero-type	OMP type	RFLP pattern	Geographic source	Year of isolation	Other designation
A2a	64	1078	8	5	4	5	5	3	3	2	2	4	1	6	4	2	4	2	2	b	3L	S	United States	1982	SLCH 3715A
A2a	65	1888	8	5	4	5	5	3	3	2	2	4	7	6	4	2	4	2	2	b			Canada	1980s	LAW 51
A2a	66	1062	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	3	b	11L	S	United States	1980	SLCH 1445
A2a	67	3734	8	5	4	5	5	3	3	2	2	4	5	6	4	2	11	2	2	b	3L		Norway	1970s	KILIAN 9703
A2a	68	1313	8	5	4	5	5	3	3	2	2	4	5	6	4	2	1	2	2	b	3L		Norway	1970s	KILIAN 3754
A2a	69	1595	8	5	4	5	6	3	3	2	2	4	5	6	4	2	4	2	2	b			Malaysia	1975	SN1352
A2a	70	319	8	5	4	0	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Papua New Guinea	1980	SLCH 4294
A2a	71	558	8	5	4	5	5	3	3	2	2	2	5	6	4	2	4	2	2	b		G	England	1983	RM6094
A2a	72	1751	8	5	4	5	5	3	3	3	2	2	5	6	4	2	4	2	2	b	3L		Iceland	1977	B-3786
A2a	73	1245	8	5	4	5	5	3	3	2	2	5	5	6	4	2	4	2	2	b			Canada	1980s	HAMEL MONTREAL
A2a	74	1301	8	5	2	5	5	3	3	2	2	5	5	6	4	2	4	2	2	b	2L		South Africa	1980s	CAPE TOWN 6559
A2a	12.6	1058	6	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	14L	S	United States	1980	SLCH 1256
A2a	12.10	476	0	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	11L		Hawaii	1980s	WAGNER 741
A2a	12.9	1373	9	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	16L	S	Japan	1982	KAWAKAMI 39
A2a	75	3089	9	5	4	5	5	3	3	2	2	4	5	6	1	2	4	2	2	b	3L		Denmark	1980s	KILIAN M1840
A2a	12.11	2358	1	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	29L		Philippines	1980s	SLCH 8661
A2a	76	1348	1	5	4	5	5	3	3	2	2	4	5	3	4	2	4	2	2	b			South Africa	1980s	JOH[§] 40422
A2a	12.7	3735	7	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L	S	United States	1984	PITT 8416
A2a	77	930	7	5	4	5	2	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		Switzerland	1983	N 201
A2a	78	1055	9	5	4	5	2	3	3	2	2	4	5	6	4	2	4	2	2	b	16L	S	United States	1970s	SLCH 1059
A2a	79	942	5	5	4	5	2	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		Switzerland	1985	N 111
A2a	12.5	1060	5	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L	S	United States	1950s	SLCH 1333
A2a	80	2833	5	5	4	5	5	3	3	2	2	4	5	3	4	2	4	2	2	b			Denmark	1983	ARPI 68
A2a	81	1343	5	5	4	5	5	3	3	2	2	4	5	5	4	2	4	2	2	b	3L		South Africa	1980s	JOH[§] LANCER
A2a	82	2852	5	5	4	5	5	3	3	2	2	4	5	2	4	2	4	2	2	b			Sweden	1982	NK 653/82
A2a	83	2315	5	5	4	5	5	3	3	2	2	4	5	6	2	2	4	2	2	b			Australia	1980s	RBH 14
A2a	84	3090	5	5	4	5	5	3	3	2	2	4	5	6	1	2	4	2	2	b			Canada	1984	HAM 13
A2a	85	3087	5	5	4	6	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		Canada	1984	HAM 16
A2a	21.8	2575	8	5	4	5	5	6	3	2	2	4	5	6	4	2	4	2	2	b	1L		Finland	1985	IHI23222
A2a	21.9	2820	9	5	4	5	5	6	3	2	2	4	5	6	4	2	4	2	2	b		S	Sweden	1985	RM7109
A2a	21.5	2310	5	5	4	5	5	6	3	2	2	4	5	6	4	2	4	2	2	b			Australia	1984	RBH 1
A2a	86	2955	5	5	4	5	5	8	3	2	2	4	5	6	4	2	4	2	2	b	3L		United States	1985	SLCH 4585B
A2a	87	2784	5	5	4	5	5	1	3	2	2	4	5	6	4	2	4	2	2	b			Finland	1985	IHI23368
A2a	88	2546	5	5	4	5	5	7	3	2	2	4	5	6	4	2	4	2	2	b			Finland	1985	IHI23255
A2a	89	2077	9	5	4	5	5	7	3	2	2	4	5	6	4	2	4	2	2	b			France	1980s	DABERNAT 1775
A2a	12.0	1694	0	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		New Zealand	1985	AS850619
A2a	90	2237	0	0	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Sweden	1982	NK 952
A2a	91	1840	7	0	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		Canada	1980s	LAW 3
A2a	92	1804	8	0	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b		S	United States	1983	RM1005
A2a	93	2979	5	0	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Scotland	1983	PEN[†] 62782
A2a	94	2714	9	4	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			United States	1980s	CHOP[#] 19
A2a	95	2708	8	4	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			United States	1986	CHOP[#] 13
A2a	96	2824	5	4	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Greenland	1983	ARPI 89-83
A2a	97	2609	7	4	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Canada	1980s	HAMEL PEI[‖] 26
A2a	98	3088	7	7	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			United States	1980s	YOGEV 111
A2a	99	1625	8	7	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b	3L		Malaysia	1975	SN1204
A2a	100	1673	7	5	4	5	5	3	3	4	2	4	5	6	4	2	4	2	2	b	3L		Canada	1980	SLCH 1687F
A2a	101	1067	7	5	4	5	5	3	3	2	3	4	5	6	4	2	4	2	2	b	14.1L	S	United States	1980	SLCH 1993
A2a	102	1077	7	5	4	5	5	3	3	2	4	4	5	6	4	2	4	2	2	b	22L	S	United States	1980	SLCH 3516
A2a	103	1057	7	5	4	5	5	3	3	2	2	4	5	6	6	2	4	2	2	b	5.1L		Alaska	1980	SLCH 1211
A2a	104	995	7	5	4	5	5	3	3	2	2	4	5	6	2	2	4	2	2	b	3L		Switzerland	1986	N 58
A2a	105	1036	8	5	4	5	5	3	3	2	2	4	5	6	6	2	4	2	2	b			Switzerland	1985	N 16
A2a	106	3091	8	5	4	5	1	3	3	2	4	4	5	6	4	2	4	2	2	b	13L		Alaska	1985	JW 19
A2a	107	3092	8	5	4	5	5	3	3	2	1	4	5	6	4	2	4	2	2	b			Canada	1969	TOR 31
A2a	108	933	5	5	4	5	5	3	3	2	1	4	5	6	4	2	4	2	2	b	3L		Switzerland	1983	N 204
A2a	109	1305	8	5	2	5	5	3	3	2	1	4	5	6	4	2	4	2	2	b	2L		South Africa	1980s	CAPE TOWN 9723
A2a	110	3736	5	5	4	5	5	3	3	2	2	4	5	6	4	2	12	2	2	b	1L		Denmark		KILIAN 114
A2a	111	1255	5	5	4	5	5	3	3	2	2	4	5	6	4	2	12	3	2	b			Canada	1980s	
A2b	112	1937	7	5	5	5	5	3	3	2	2	4	5	6	4	2	4	2	2	b			Canada	1980s	LAW 100
A2b	113	1244	8	5	5	5	5	3	3	2	2	4	5	6	4	3	4	2	2	b			Canada	1980s	CHILETTE
A2b	114	3737	6	5	1	5	5	3	3	2	2	4	5	6	4	3	4	2	2	b	5L		Alaska	1980s	JW 81
A2c	115	1314	8	5	4	5	5	3	3	2	2	4	4	5	4	2	4	2	2	b	2L		South Africa	1980s	DURBAN 63
A2d	116	188	8	5	4	5	3	6	3	2	2	4	5	6	4	2	4	2	2	b	16L		Dominican Republic	1980s	CEF 79
A2d	117	212	5	5	4	5	3	2	3	2	2	4	5	6	4	2	4	2	2	b			Dominican Republic	1980s	SULB 37
A2d	118	1059	7	5	4	5	3	3	3	2	2	4	5	6	4	2	4	2	1	b	5L	S	United States	1980s	SLCH 1287
A2e	21.6	1053	6	5	4	5	5	6	3	2	2	4	5	6	4	2	4	2	2	b	13L		United States	1979	SLCH 808A
A2e	119	1068	6	5	4	5	5	6	3	2	3	4	5	6	4	2	4	2	2	b	11L		United States	1980s	SLCH 3012
A2e	120	3095	6	5	4	5	5	6	3	2	2	4	5	6	6	2	4	2	2	b	13L		Alaska	1980s	JW 80

(*continued*)

Table 2. (*continued*)

Group	ET	MHAEM no.*	Allele at indicated enzyme locus																	Sero-type	OMP type	RFLP pattern	Geographic source	Year of isolation	Other designation
			CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM						
A2e	121	472	6	5	4	5	5	6	3	2	2	4	4	6	4	2	4	2	2	b	13L		Hawaii	1980	2862
A2e	122	1778	6	5	5	5	5	6	3	2	2	4	5	6	4	2	4	2	2	b		S	United States	1984	RM8012
A2e	123	1092	6	5	4	3	5	6	3	2	2	4	5	6	4	2	4	2	2	b			United States	1980s	
A2e	124	3096	6	0	4	5	5	6	3	2	1	4	5	6	4	2	4	2	2	b	13L		Alaska	1980s	JW 67
A2e	125	2716	6	5	4	5	5	6	3	2	2	4	7	3	4	2	4	2	2	b			United States	1980s	CHOP# 21
A2e	126	2809	6	5	4	5	5	7	3	2	2	5	5	6	4	2	4	2	2	b			Denmark	1944	ENGBAEK T28
A2f	127	2804	6	5	4	3	2	3	3	2	2	4	5	6	4	2	4	2	2	b			Denmark	1943	ENGBAEK K7
A3	128	484	6	5	1	3	5	3	3	2	2	4	5	3	4	12	4	2	2	b			Hawaii	1980s	WAGNER 6546
A4	129	1670	6	5	1	3	5	7	1	2	2	4	8	8	4	2	4	2	2	b			Canada	1980s	SLCH 1687C
A5	130	1300	4	7	1	5	6	6	3	2	1	4	5	2	4	2	4	2	2	b	1L		South Africa	1980s	CAPE TOWN 5274
B1a	131	3097	6	5	4	5	3	6	3	2	2	4	5	3	6	2	4	2	2	b			United States	1980s	RGH 5531
B1a	132	1629	6	5	4	5	3	6	3	1	2	4	5	3	6	2	4	2	2	b			Malaysia	1973	SN0573
B1b	25.6	1063	6	5	5	3	3	6	3	2	2	4	5	3	6	2	4	2	2	b	6U	S	United States	1981	SLCH 1481a
B1b	133	798	6	5	5	3	3	6	3	2	3	4	5	3	6	2	4	2	2	b		S	United States	1985	RM8069
B1b	134	1973	6	5	5	3	3	6	6	2	2	4	5	3	6	2	4	2	2	b			Canada	1980s	LAW 136
B1b	135	2163	6	5	5	3	3	6	3	2	2	3	5	3	6	2	4	2	2	b	6U		Papua New Guinea	1980	SLCH 8362a
B1b	136	2033	6	5	5	3	3	6	3	2	2	4	5	6	6	2	4	2	2	b			Hawaii	1986	WAGNER 2954
B1b	137	2135	6	5	5	3	3	6	3	2	2	4	5	6	4	2	4	2	2	b			France		DABERNAT 2280
B1b	138	881	6	5	5	3	3	6	3	2	2	4	5	3	4	2	4	2	2	b	6U	S	The Netherlands	1977	770177
B1b	139	2139	6	5	5	3	3	6	3	2	2	4	5	5	4	2	4	2	2	b	6U		Kenya	1980	SLCH 8009
B1b	140	1795	6	5	5	3	3	7	3	2	2	4	5	5	4	2	4	2	2	b		S	Kenya	1980s	RM7414
B1b	141	1481	6	5	5	3	3	6	3	2	2	4	5	5	6	2	4	2	2	b	6U		The Gambia	1982	WALL 13
B1b	142	1480	6	5	5	3	3	6	3	2	2	4	5	5	6	2	4	2	1	b			Hawaii	1980s	WAGNER 6618
B1b	143	1056	6	5	5	3	3	6	3	2	2	4	5	3	6	2	4	2	1	b	23U	S	United States	1985	SLCH 1209
B1b	144	1493	6	5	5	3	3	3	3	2	2	4	5	5	6	2	4	2	2	b			The Gambia		WALL 42
B1b	145	1066	6	5	5	3	3	6	3	2	2	4	5	2	6	2	4	2	2	b	24U	S	United States	1982	SLCH 1971
B1b	146	3098	6	5	5	3	3	7	3	2	2	4	5	2	6	2	4	2	2	b			United States	1984	PITT 8447
B1b	147	171	6	7	5	3	3	6	3	2	2	4	5	2	6	2	4	2	2	b			Dominican Republic	1980	CEF 8
B1b	148	215	6	7	5	3	3	6	3	2	2	4	5	3	6	2	4	2	2	b			Dominican Republic	1980s	SULB 83
B1b	149	225	6	7	5	3	3	3	3	2	2	4	5	2	6	2	4	2	2	b	6U		Dominican Republic	1980s	SULB 85
B1b	150	874	6	5	5	3	3	6	3	2	2	4	7	5	4	2	4	2	2	b	6U		South Africa	1984	840062
B1b	151	912	6	5	5	3	3	6	3	2	2	4	7	6	4	2	4	2	2	b		S	Switzerland	1985	N 207
B1b	152	1334	6	5	5	3	3	6	3	2	2	4	7	5	6	2	4	2	2	b	6U	S	South Africa	1980	12681
B1b	153	1347	6	0	5	3	3	6	3	2	2	4	7	5	6	2	4	2	2	b	6U		South Africa	1980s	JOH§ 128011
B1c	154	206	5	5	5	3	3	3	3	2	2	4	5	6	6	2	4	2	2	b			Dominican Republic	1980s	SULB 31
B1d	155	3100	6	5	5	3	3	1	3	2	2	4	5	12	6	2	4	1	2	b	1U		Alaska	1980s	JW 65
B1e	156	3099	6	5	5	3	3	3	3	2	2	3	5	3	4	2	4	2	2	b			South Korea	1980s	D14
B1f	157	571	6	5	5	3	5	6	3	2	2	4	5	3	6	2	4	2	2	d		D	England	1963	RM6137
B1f	158	518	6	5	5	3	5	6	3	2	2	4	5	3	6	2	4	3	2	d		D	United States		RM1168
B1f	159	584	6	5	5	3	5	6	1	2	2	4	5	3	6	2	4	2	2	d		D	England	1985	RM6150
B1f	160	787	6	5	5	5	5	6	1	2	2	4	5	3	6	2	4	2	2	d		D	United States	1984	RM8039
B1f	161	706	6	5	5	3	5	6	1	2	2	4	5	3	6	2	11	2	2	d		D	Papua New Guinea	1980s	RM7033
B1g	162	747	6	5	5	3	5	6	3	2	2	2	5	3	6	5	4	2	2	d		D	Malaysia	1972	RM7271
B1g	163	1810	6	5	5	3	5	6	3	2	2	2	5	5	4	2	4	2	2	d		D	Kenya	1986	RM7429
B2a	164	506	6	5	6	5	5	6	3	2	2	4	5	2	1	2	4	2	3	a	1U	T	United States	1981	RM1042
B2b	165	704	6	5	5	3	5	6	3	2	2	4	5	2	1	2	4	2	2	a	1U	T	Papua New Guinea	1980s	RM7031
B2b	166	705	6	5	5	3	5	6	3	2	2	4	5	2	1	2	4	3	2	a	1U	T	Papua New Guinea	1980s	RM7032
B2b	167	728	6	5	5	3	5	6	3	2	2	4	5	3	1	2	4	2	3	a	1U	T	Malaysia	1973	RM7191
B2b	168	735	6	5	5	3	5	6	3	2	2	4	6	3	1	2	4	2	3	a	1U	T	Malaysia	1974	RM7198
B2c	169	727	6	5	5	3	5	6	1	2	2	2	5	6	1	2	4	2	3	a	1U	T	Malaysia	1973	RM7190
B2d	170	1802	6	5	5	3	5	6	6	2	2	4	5	5	3	2	4	2	3	a	1U	N	Kenya	1986	RM7421
B3	171	1485	6	5	11	2	3	6	5	2	2	4	5	3	4	2	4	2	2	b			The Gambia	1983	WALL 18
B4	172	1370	6	5	2	3	2	6	1	2	2	4	5	2	4	2	4	2	2	b			Japan	1985	KAWAKAMI 10
B4	173	742	6	5	4	3	2	6	1	2	2	4	5	5	4	2	4	2	3	a	5L	N	The Gambia	1980s	RM7205
B4	174	1797	6	5	4	3	2	6	1	2	2	4	5	3	4	2	4	2	3	a	5L	N	Kenya	1986	RM7416
B5	175	1788	6	5	1	3	3	8	6	2	2	5	4	3	6	2	4	2	2	b			Iceland	1986	B-1014
C1	176	993	4	5	4	2	3	8	5	2	3	5	5	6	4	3	4	1	2	b			Switzerland	1985	N 60
D1a	177	517	6	5	10	5	1	6	5	2	2	4	1	3	6	2	4	3	2	c		C1	United States		RM1167
D1a	178	746	6	5	10	5	1	6	5	2	2	4	1	3	6	2	4	2	3	c		C1	Malaysia	1975	RM7270

(*continued*)

Table 2. (*continued*)

Group	ET	MHAEM no.*	Allele at indicated enzyme locus																	Sero- type	OMP type	RFLP pattern	Geographic source	Year of isolation	Other designation
			CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM						
D1b	179	520	6	0	10	5	1	6	5	2	2	4	1	2	6	2	4	2	2	c			Unknown	1968	RM1271
D1b	180	743	6	0	10	5	1	6	5	2	2	4	3	5	6	2	4	2	2	c		C1	Malaysia	1973	RM7267
D1b	181	569	6	0	10	5	5	6	5	2	2	4	5	2	6	2	4	2	2	c		C1	England	1981	RM6135
D2	182	565	6	5	10	5	5	6	5	2	2	4	3	6	6	2	4	2	1	c		C1	Unknown	1970	RM6129
D2	183	566	6	5	10	5	5	5	5	2	2	4	3	6	4	2	4	2	2	c		C1	England	1964	RM6132
D3	184	785	6	0	13	5	1	3	5	2	2	4	5	6	6	2	4	3	2	c		C1	United States	1983	RM8032
D4	185	568	6	5	7	5	1	7	3	2	2	4	4	3	6	2	4	3	2	c		C2	England	1975	RM6134
D5	186	1803	6	5	7	5	1	7	5	2	2	3	4	2	4	2	4	2	2	c		C2	Kenya	1986	RM7422
E1	187	1805	5	5	1	5	5	6	1	2	2	5	7	1	3	2	4	2	1	c			Kenya	1986	RM7424
F1	188	602	5	5	9	2	5	6	3	2	2	4	3	3	7	2	5	2	2	e		E	England	1964	RM6168
F1	189	603	5	5	9	2	5	6	3	2	2	4	3	3	7	2	1	2	2	e		E	England	1964	RM6169
F1	190	619	5	5	9	2	5	6	3	2	2	4	5	3	7	2	1	2	2	e		E	England	1965	RM6185
F2a	191	492	9	5	9	2	5	1	6	2	4	2	3	3	7	2	5	2	2	e		E	United States		RM1018
F2a	192	784	9	5	9	2	5	1	6	2	2	2	3	3	7	2	5	2	2	e			United States	1983	RM8031
F2a	193	634	9	5	9	2	5	1	6	2	4	2	3	5	7	2	5	2	2	e		E	England	1967	RM6200
F2a	194	755	9	5	9	2	5	3	6	2	2	2	3	6	7	2	5	2	2	e		E	Malaysia	1973	RM7279
F2a	195	585	9	7	9	2	5	1	3	2	4	2	3	3	7	2	5	2	2	e			Unknown		RM6151
F2a	196	644	9	0	9	2	5	1	3	2	4	2	3	3	7	2	5	2	2	e		E	England	1972	RM6210
F2a	197	592	9	5	9	2	5	1	3	2	4	2	3	3	7	2	5	2	2	e			England	1962	RM6158
F2a	198	596	9	5	9	2	5	1	3	2	4	2	3	3	8	2	5	2	2	e		E	England	1963	RM6162
F2a	199	589	9	5	9	2	5	1	3	2	4	2	3	3	7	2	5	3	2	e		E	Unknown		RM6155
F2a	200	590	9	5	9	2	5	1	3	2	4	2	2	3	6	2	5	2	2	e		E	Unknown		RM6156
F2a	201	624	9	5	9	2	5	1	3	2	4	2	2	3	7	2	5	2	2	e		E	England	1966	RM6190
F2b	202	615	5	5	9	2	5	1	5	2	4	2	3	3	7	2	1	2	2	e		Anom**	England	1965	RM6181
F2b	203	709	5	5	9	2	5	1	3	2	1	2	3	3	7	2	1	2	2	e		E	Papua New Guinea	1980s	RM7066
F2c	204	601	9	5	9	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1963	RM6167
F2c	205	1796	9	5	9	2	5	3	3	2	2	2	3	3	5	2	5	2	2	e			Kenya	1986	RM7415
F2c	206	756	9	5	9	2	4	3	3	2	2	2	3	3	7	2	5	2	2	e		E	Malaysia	1973	RM7280
F2c	207	600	9	5	9	2	5	3	3	2	2	2	3	4	7	2	5	2	2	e		E	England	1963	RM6166
F2c	208	612	9	5	9	2	5	3	3	2	2	2	3	3	7	2	5	2	1	e		E	England	1964	RM6178
F2c	209	611	9	0	9	2	5	3	3	2	2	2	3	3	7	2	5	2	1	e			England	1964	RM6177
F2c	210	645	9	0	9	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1972	RM6211
F2c	211	662	0	0	9	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1976	RM6228
F2c	212	649	9	8	9	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e			England	1973	RM6215
F2c	213	763	9	5	9	2	5	3	3	2	2	2	3	5	6	2	5	2	2	e			Malaysia	1970s	RM7287
F2c	214	642	9	5	9	2	3	4	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1971	RM6208
F2c	215	663	5	0	10	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1977	RM6229
F2c	216	653	5	5	10	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1974	RM6219
F2c	217	661	0	5	10	2	5	3	3	2	2	2	3	3	7	2	5	2	2	e		E	England	1976	RM6227
G1	218	1804	5	5	9	2	2	1	5	2	2	2	3	3	4	2	2	2	3	e		E	Kenya	1986	RM7423
H1a	219	515	7	5	4	1	3	1	3	2	5	2	3	2	6	3	1	2	2	a	2H	M	Unknown		RM1147
H1a	220	553	7	5	4	1	5	1	3	2	5	2	3	2	6	3	1	2	2	a	4H	M	England	1977	RM6083
H1a	221	543	7	5	4	1	3	1	3	2	5	2	2	2	6	3	1	2	2	a	2H	M	England	1966	RM6073
H1b	222	529	7	4	4	1	5	1	3	2	5	2	3	3	4	3	1	2	2	a	6H	M	England	1964	RM6059
H1b	223	534	7	5	4	1	5	1	3	2	5	2	3	3	4	3	1	2	2	a	6H	M	England	1966	RM6064
H1b	224	540	7	4	4	1	5	1	3	2	5	2	3	2	4	3	1	2	2	a	6H	M	England	1962	RM6070
H1b	225	549	0	4	4	1	5	1	3	2	5	2	3	2	4	3	1	2	2	a	6H	M	England	1967	RM6079
H1b	226	538	7	5	4	1	5	1	3	2	5	2	3	2	4	4	1	2	2	a	6H	M	England	1968	RM6068
I1a	227	532	1	5	5	1	5	1	2	1	5	2	3	1	4	2	4	2	2	a	4H	M	England	1965	RM6062
I1a	228	539	1	5	5	1	5	1	2	1	5	2	3	1	4	3	4	2	2	a	4H	M	England	1963	RM6069
I1a	229	550	1	5	5	1	5	1	2	1	5	1	3	1	4	2	4	2	2	a	4H	M	England	1967	RM6080
I1b	230	724	1	5	4	1	5	1	3	1	4	2	3	1	4	2	4	2	2	a	5H	M	Dominican Republic	1980s	RM7115
J1a	231	3106	7	5	4	2	5	1	4	1	2	3	5	2	2	5	4	2	2	b			United States	1968	CATLIN 6325
J1a	232	3107	7	5	4	2	5	1	4	1	5	3	5	2	2	5	4	2	2	b		O	United States	1968	CATLIN 5788
J1b	233	1076	7	5	4	2	5	1	3	1	4	2	5	2	1	3	4	2	2	b	8-H††		United States	1983	SLCH 3491
J1b	234	3108	7	6	4	2	5	1	3	1	4	3	5	2	1	3	4	2	2	b	8H		Alaska	1980s	JW 28
J1b	235	3109	7	7	4	2	5	1	3	1	4	3	5	2	1	3	4	2	2	b	8H	O	Alaska	1980	JW 43
J1b	236	1069	7	5	4	2	5	1	3	1	3	3	5	2	1	3	4	2	2	b	8H		United States	1954	SLCH 3191
J1b	237	1070	7	5	4	2	5	1	3	1	3	3	5	2	1	3	4	3	2	b	8H		United States	1940	SLCH 3205
J2	238	1303	7	5	4	5	5	1	3	1	4	4	5	2	2	2	4	2	2	b			South Africa	1980s	CAPE TOWN 8811
J3	239	1084	7	5	4	3	5	1	7	1	4	2	1	3	2	3	4	2	2	b	17H		United States	1947	RABINOWITZ
K1a	240	511	0	4	3	2	5	1	3	1	5	2	2	3	4	3	4	2	2	f		O	United States	1979	RM1121

(*continued*)

Table 2. (*continued*)

Group	ET	MHAEM no.*	Allele at indicated enzyme locus																	Sero-type	OMP type	RFLP pattern	Geographic source	Year of isolation	Other designation
			CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM						
K1b	241	686	7	5	3	2	3	1	3	1	5	2	2	3	7	3	4	2	2	f			England	1966	RM6252
K1c	242	697	7	5	2	2	5	1	3	1	5	2	2	2	7	3	4	2	2	f		F	England		RM6263
K1c	243	698	7	5	2	2	5	1	3	1	5	2	2	3	7	3	4	2	2	f			England	1984	RM6265
K1c	244	771	7	5	2	2	5	1	3	1	5	2	2	3	2	3	4	2	2	f		F	Malaysia	1970s	RM7299
K1c	245	757	7	5	2	2	5	1	3	1	5	2	2	3	4	3	4	2	2	f			Malaysia	1971	RM7281
K1c	246	758	7	5	2	2	5	1	3	1	5	2	2	2	4	3	4	2	2	f			Malaysia	1970s	RM7282
K1c	247	768	7	5	2	2	5	1	3	1	5	2	2	5	4	3	4	2	2	f			Malaysia	1970s	RM7295
K1c	248	761	7	5	2	2	5	1	3	1	5	2	2	6	4	3	4	2	2	f			Malaysia	1970s	RM7285
K1c	249	759	7	5	2	2	5	1	3	1	5	2	2	1	4	3	4	2	2	f		F	Malaysia	1970s	RM7283
K1c	250	770	7	5	2	2	5	1	3	1	5	2	2	1	2	3	4	2	2	f		F	Malaysia	1970s	RM7298
K1c	251	764	5	5	2	2	5	1	3	1	5	2	2	5	4	3	4	2	2	f		F	Malaysia	1970s	RM7290
K1c	252	760	7	0	2	2	5	1	3	1	5	2	2	1	4	3	4	2	2	f			Malaysia	1970s	RM7284
K1c	253	762	7	0	2	2	5	1	3	1	5	2	2	3	4	3	4	2	2	f		F	Malaysia	1970s	RM7286
K1c	254	765	5	0	2	2	5	1	3	1	5	2	2	6	4	3	4	2	2	f		F	Malaysia	1970s	RM7292
K1d	255	776	7	0	2	5	5	1	3	1	5	2	2	3	2	3	4	2	2	f			United States	1984	RM8009
K2a	256	514	7	5	3	2	5	3	4	1	5	2	2	3	6	3	1	2	2	f		O	United States		RM1137
K2a	257	689	7	5	3	2	5	3	4	1	5	2	2	3	7	3	1	2	2	f			England	1967	RM6255
K2a	258	699	7	5	3	2	5	3	4	1	5	2	2	2	7	3	1	2	2	f		UN‡‡	England	1984	RM6266
K2a	259	693	7	5	3	2	5	3	4	1	5	1	2	3	7	3	1	2	2	f		O	England	1967	RM6259
K2a	260	688	7	0	3	2	5	3	4	1	5	2	2	3	7	3	1	2	2	f		O	England	1966	RM6254
K2a	261	696	7	0	3	2	5	3	4	1	5	2	2	2	7	3	1	2	2	f		O	England	1968	RM6262
K2a	262	669	7	5	3	2	5	3	4	1	5	2	1	3	6	3	1	2	2	f		O	England	1963	RM6235
K2a	263	678	7	5	3	2	5	3	4	1	5	2	1	3	7	3	1	2	2	f		O	England	1964	RM6244
K2a	264	673	7	5	3	2	5	3	4	1	5	2	1	5	6	3	1	2	2	f		UN‡‡	England	1963	RM6239
K2a	265	671	7	0	3	2	5	3	4	1	5	2	1	3	6	3	1	2	2	f		O	England	1963	RM6237
K2a	266	679	7	0	3	2	5	3	4	1	5	2	1	3	7	3	1	2	2	f			England	1964	RM6245
L1	267	1798	7	5	3	2	5	1	4	1	1	2	2	5	5	2	1	3	2	f			Kenya	1986	RM7417

NOTE. See text or footnote to table 3 for enzyme abbreviations.

* Musser *Haemophilus*.

† Pennington.

‡ Similar to subtype 1L, but not identical.

§ Johannesburg.

Children's Hospital of Philadelphia.

‖ Prince Edward Island.

** Anomalous.

†† Similar to subtype 8, but apparently lacks P1 band.

‡‡ Undefined.

Genetic Relationships Among Multilocus Enzyme Genotypes

The dendrogram in figure 1 summarizes estimates of the genetic relationships of the 280 ETs, based on allelic variation at the 17 enzyme loci. At a genetic distance of 0.20, there were 56 branches, each represented by a single ET or a cluster of ETs. Twelve major lineages diverging at distances greater than 0.42 were designated by the capital letters A through L; clusters of ETs in these lineages were numbered (e.g., A1, A2, etc.), and groups of ETs in a cluster were designated with lower case letters. Lineages A through G were separated from lineages H through L at a genetic distance of 0.66, which means that ETs in these two primary divisions of the dendrogram (I and II) differed, on average, at 10 of the 17 loci assayed.

Major lineage A, in division I, was composed entirely of ETs of serotype b isolates. (Cluster A1 and clusters A2 through A5 included ETs previously assigned by Musser et al. [20, 29] to clone families [groups of closely allied clones] designated as A and B, respectively.) Lineage B, which diverged from lineage A at a genetic distance of 0.43, contained ETs of isolates producing capsule types a, b, and d. (Serotype b isolates in this lineage previously were assigned to clone family C [20, 29].) Clusters D1 through E1 contained the ETs of all the serotype c isolates studied, and lineages F and G, which were separated from lineages A through E at a genetic distance of 0.58, included all ETs represented by serotype e isolates. Lineages H and I, clusters J1 through J3, and clusters K1 through L1 included ETs of isolates of serotype a, b, and f, respectively. (Serotype b strains in clusters

Table 3. Mean genetic diversity (*H*) at 17 enzyme loci in 280 ETs of encapsulated *H. influenzae*.

Enzyme locus*	Mean genetic diversity (no. of alleles)						
	Serotype						
	a	b	c	d	e	f	Total
CAK	0.700 (4)	0.808 (9)	0.182 (2)	0.000 (1)	0.486 (3)	0.204 (3)	0.812 (9)
NSP	0.257 (2)	0.224 (5)	0.509 (2)	0.000 (1)	0.385 (4)	0.474 (3)	0.286 (6)
PGI	0.567 (3)	0.606 (6)	0.600 (4)	0.000 (1)	0.181 (2)	0.519 (2)	0.770 (11)
MAE	0.552 (3)	0.398 (5)	0.000 (1)	0.286 (2)	0.000 (1)	0.071 (2)	0.623 (6)
MDH	0.343 (3)	0.391 (5)	0.509 (2)	0.000 (1)	0.187 (4)	0.071 (2)	0.358 (6)
G6P	0.514 (2)	0.532 (6)	0.600 (4)	0.000 (1)	0.630 (4)	0.495 (2)	0.647 (8)
GOT	0.538 (4)	0.501 (6)	0.345 (3)	0.571 (2)	0.340 (3)	0.508 (2)	0.538 (7)
ADK	0.324 (2)	0.135 (4)	0.000 (1)	0.000 (1)	0.000 (1)	0.000 (1)	0.274 (4)
6PG	0.567 (3)	0.260 (5)	0.000 (1)	0.000 (1)	0.495 (3)	0.071 (2)	0.449 (5)
PE2	0.552 (3)	0.214 (5)	0.345 (3)	0.476 (2)	0.181 (2)	0.071 (2)	0.485 (6)
PE1	0.605 (4)	0.147 (6)	0.855 (5)	0.000 (1)	0.185 (3)	0.304 (2)	0.526 (8)
LAP	0.748 (5)	0.661 (6)	0.836 (5)	0.286 (2)	0.243 (4)	0.680 (5)	0.731 (8)
PGM	0.652 (4)	0.355 (4)	0.473 (3)	0.286 (2)	0.299 (5)	0.738 (5)	0.589 (8)
CAT	0.552 (3)	0.135 (4)	0.000 (1)	0.286 (2)	0.000 (1)	0.071 (2)	0.300 (5)
GLD	0.495 (2)	0.044 (4)	0.000 (1)	0.286 (2)	0.288 (3)	0.508 (2)	0.345 (6)
G3P	0.095 (2)	0.065 (3)	0.436 (2)	0.286 (2)	0.065 (2)	0.071 (2)	0.090 (3)
FUM	0.467 (2)	0.043 (3)	0.473 (3)	0.000 (1)	0.185 (3)	0.000 (1)	0.116 (3)
Average	0.502 (3.0)	0.325 (5.1)	0.363 (2.5)	0.162 (1.5)	0.242 (2.8)	0.286 (2.4)	0.467 (6.4)

* Abbreviations: CAK, carbamylate kinase; NSP, nucleoside phosphorylase; PGI, phosphoglucose isomerase; MAE, malic enzyme; MDH, malate dehydrogenase; G6P, glucose 6-phosphate dehydrogenase; GOT, glutamic oxaloacetic transaminase; ADK, adenylate kinase; 6PG, 6-phosphogluconate dehydrogenase; PE2, leucylalanine peptidase-2; PE1, leucylalanine peptidase-1; LAP, leucine aminopeptidase; PGM, phosphoglucomutase; CAT, catalase; GLD, glutamate dehydrogenase; G3P, glyceraldehyde-3-phosphate dehydrogenase; and FUM, fumarase.

J1 through J3 previously were assigned to clone family D [29].)

In summary, multilocus enzyme electrophoresis revealed two major genetic divisions among encapsulated *H. influenzae*. Division I contained all ETs of serotype c and d strains, one phylogenetic line of serotype a strains, and most ETs (95%) and isolates (99%) of serotype b. Division II included all ETs of serotype f, a second major line of ETs expressing serotype a capsule, and a genetically heterogeneous second group of ETs of serotype b isolates. Thus, ETs of isolates of serotypes a and b occurred in each of the two primary phylogenetic divisions that are separated at a genetic distance of 0.66.

The allele profiles and additional characteristics of representative isolates of the 280 ETs of encapsulated *H. influenzae* are presented in table 2; information on the geographic origin, OMP type, and *cap* region RFLP pattern of the isolates assigned to each ET is given in table 4. ETs were numbered in consecutive order according to their positions in the dendrogram (figure 1), with the exception of ETs that were identical at all loci except the CAK locus to the four common multilocus genotypes (ETs 1, 12, 21, and 25) earlier identified by Musser et al. [20], for which are reserved the numbers previously assigned. After the appearance of the paper describing these four ETs [20], two additional enzymes were identified, carbamylate kinase and nucleoside phosphorylase, that can be assayed in *H. influenzae* isolates. Carbamylate kinase is strongly polymorphic in serotype b isolates, and many electrophoretic variants of this enzyme are represented among isolates previously grouped in each of the four common ETs (ETs 1, 12, 21, and 25). Consequently, isolates of each of these ETs are now differentiated into a series of ETs on the basis of alleles of CAK (e.g., ET 12.5, 12.6, 12.7, 12.8, etc.).

The great majority (80%) of serotype b ETs are in groups A1a, A1c, A2a, A2e, and B1b, and most isolates are of ETs in groups A1a (28%), A2a (60%), and B1b (6%). The four ETs represented by the largest number of isolates are ET 1.9 (81% of isolates in group A1a), ET 12.5 and 12.8 (75% of isolates in group A2a), and ET 25.6 (40% of isolates in group B1b). Isolates of these four multilocus geno-

Table 4. Characteristics of 280 ETs (represented by 2,209 isolates) of encapsulated *H. influenzae*.

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
A1a 1.9	Alaska (26)	1H(13), 2L(6), 2H(4), 1L(2), 4H(1)	
A1a 1.9	Australia (3)		
A1a 1.9	Canada (143)	2L(22), 1L(2), 1H(2), 19H(1)	
A1a 1.9	Denmark (3)	1L(2)	
A1a 1.9	Dominican Republic (1)		S(1)
A1a 1.9	England (3)		G(2)
A1a 1.9	Finland (2)		
A1a 1.9	France (1)		
A1a 1.9	Hawaii (7)	1H(1), 11L(1)	
A1a 1.9	Iceland (33)	2L(31), 1H(1), G(1), S(1), V(1)	
A1a 1.9	Mexico (1)		
A1a 1.9	Norway (6)	1H(4), 2H(1)	
A1a 1.9	Papua New Guinea (1)		
A1a 1.9	South Africa (2)	2L(2)	S(1)
A1a 1.9	Sweden (6)		G(1)
A1a 1.9	Switzerland (2)	1H(1), 2L(1)	
A1a 1.9	United States (210)	1H(47), 2L(25), 2H(4), 7H(1), 10H(1), 15L(1), 18L(1), V(25), G(5)	
A1a 2	Canada (4)	1H(1)	
A1a 2	United States (1)	1H(1)	
A1a 3	United States (1)		
A1a 4	Alaska (6)	1L(6)	
A1a 4	Canada (5)		
A1a 4	Iceland (1)		
A1a 4	Sweden (1)		
A1a 4	United States (19)	1L(11), 1H(2)	G(1)
A1a 5	United States (1)	1H(1)	G(1)
A1a 6	United States (1)	1L(1)	G(1)
A1a 7	United States (1)	1L(1)	
A1a 1.6	Canada (2)		
A1a 1.6	Alaska (1)	2L(1)	
A1a 1.7	Australia (1)		V(1)
A1a 1.8	Norway (1)	1L(1)	
A1a 8	Australia (1)		
A1a 8	Hawaii (1)		
A1a 8	Kenya (4)	3L(3)	G(4)
A1a 8	South Africa (5)		
A1a 8	Switzerland (1)		
A1a 9	Canada (1)		
A1a 10	United States (1)		
A1a 11	United States (1)	1L(1)	
A1a 13	United States (2)	1H(1), 2L(1)	
A1a 14	Alaska (1)	1H(1)	
A1a 15	United States (1)	2L(1)	
A1a 16	Canada (2)		

(*continued*)

Table 4. (*continued*)

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
A1a 16	England (1)		
A1a 16	United States (1)		
A1a 17	Canada (2)		
A1a 18	Alaska (1)	2L(1)	
A1a 19	Alaska (3)	2L(3)	
A1a 20	Canada (1)		
A1a 20	United States (1)	1H(1)	
A1a 22	Canada (1)		
A1a 22	The Netherlands (1)		
A1a 22	Scotland (1)		
A1a 22	United States (2)	2L(2)	
A1a 23	Scotland (1)		
A1a 24	United States (1)	2L(1)	
A1a 26	United States (1)	1H(1)	G(1)
A1a 27	United States (1)	1L(1)	
A1a 28	United States (1)		
A1a 29	United States (1)	2L(1)	V(1)
A1a 30	United States (2)	1H(1), 9L(1)	V(1)
A1a 31	England (1)		G(1)
A1a 32	Canada (1)		
A1a 32	England (1)		G(1)
A1a 32	United States (3)	1L(1), 2H(1)	V(1)
A1a 33	Canada (1)		
A1a 34	United States (1)	1L(1)	G(1)
A1a 35	South Africa (1)		
A1b 36	United States (1)	1L(1)	G(1)
A1c 37	Canada (1)	2L(1)	
A1c 38	United States (1)	1U(1)	
A1c 39	Norway (1)	3L(1)	S(1)
A1c 40	France (1)		
A1c 40	Spain (1)	3L(1)	
A1c 41	France (1)		
A1c 42	Switzerland (1)	3L(1)	
A1c 43	France (1)		
A1c 43	Switzerland (1)	11L(1)	
A1c 44	United States (1)	1L(1)	
A1c 45	The Gambia (3)	1L(1)	
A1c 46	Australia (1)		
A1c 47	Scotland (1)		
A1c 48	Canada (1)	26L(1)	
A1d 49	United States (1)	1H(1)	
A1d 50	Ghana (2)		G(1), V(1)
A1d 51	Scotland (1)		
A1e 52	Ghana (1)		
A1f 53	Denmark (1)		
A2a 54	Canada (1)	1L(1)	
A2a 55	Canada (1)		
A2a 56	The Gambia (1)		
A2a 56	United States (1)		
A2a 57	Hawaii (1)		
A2a 57	United States (1)	3L(1)	

(*continued*)

Table 4. (*continued*)

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
A2a 58	Malaysia (5)	16L(1)	G(1)
A2a 12.8	Alaska (3)	3L(3)	
A2a 12.8	Argentina (1)	2L(1)	
A2a 12.8	Australia (2)		
A2a 12.8	Canada (72)	3L(6)	
A2a 12.8	Denmark (2)		
A2a 12.8	Dominican Republic (20)		
A2a 12.8	England (1)		
A2a 12.8	Finland (44)	3L(3)	
A2a 12.8	France (5)	2L(1)	
A2a 12.8	The Gambia (18)	2L(6)	
A2a 12.8	Ghana (2)		S(2)
A2a 12.8	Hawaii (4)		
A2a 12.8	Japan (10)	3L(3), 16L(1)	
A2a 12.8	Malaysia (52)	3L(10), 18L(3), 1L(1), 2L(1), 13.1L(1), 27L(1)	S(2)
A2a 12.8	New Zealand (4)	3L(4)	
A2a 12.8	Norway (5)		
A2a 12.8	Papua New Guinea (22)	1L(9), 3L(5), 13L(1)	S(3)
A2a 12.8	Philippines (6)	1L(3), 3L(2), 18L(1)	
A2a 12.8	South Africa (9)	2L(2)	S(1)
A2a 12.8	Scotland (1)		
A2a 12.8	South Korea (2)		
A2a 12.8	Sweden (14)		
A2a 12.8	Switzerland (9)		
A2a 12.8	Thailand (7)	2L(3), 3L(4), 1L(1)	S(1)
A2a 12.8	United States (76)	3L(20), 13L(6), 14L(1)	S(2)
A2a 12.8	Uruguay (1)	2L(1)	
A2a 59	Malaysia (1)	3L(1)	
A2a 59	Thailand (1)	1L(1)	
A2a 60	Denmark (1)		
A2a 60	Kenya (4)	3L(1)	S(4)
A2a 60	Malaysia (4)	3L(3)	
A2a 60	South Africa (1)		
A2a 60	Sweden (1)		
A2a 61	Denmark (1)		
A2a 62	United States (1)	3L(1)	
A2a 63	Norway (1)	3L(1)	
A2a 64	United States (1)	3L(1)	S(1)
A2a 65	Canada (1)		
A2a 66	United States (1)	11L(1)	S(1)
A2a 67	Denmark (1)		
A2a 67	Finland (1)		
A2a 67	Norway (1)	3L(1)	
A2a 68	Norway (1)		
A2a 68	South Africa (1)		
A2a 69	Malaysia (4)		
A2a 70	Papua New Guinea (2)		
A2a 71	England (1)		S(1)
A2a 72	Iceland (1)	3L(1)	
A2a 73	Canada (1)		
A2a 74	South Africa (1)	2L(1)	
A2a 12.6	Australia (1)		
A2a 12.6	Denmark (1)		
A2a 12.6	France (3)		
A2a 12.6	Philippines (1)	18L(1)	
A2a 12.6	Switzerland (1)		
A2a 12.6	United States (4)	14L(1), 18L(1)	S(1)
A2a 12.10	Hawaii (1)	11L(1)	
A2a 12.9	Canada (1)		
A2a 12.9	Dominican Republic (1)		
A2a 12.9	Finland (3)		
A2a 12.9	France (1)		
A2a 12.9	Japan (2)	16L(2)	
A2a 12.9	Netherlands (1)		
A2a 12.9	Norway (2)		
A2a 12.9	South Korea (5)		
A2a 12.9	Sweden (5)		S(1)
A2a 12.9	Switzerland (2)		
A2a 75	Denmark (1)	3L(1)	
A2a 12.11	Philippines (1)	29L(1)	
A2a 76	South Africa (1)		
A2a 12.7	Alaska (14)	3L(14)	
A2a 12.7	Australia (1)		
A2a 12.7	Canada (23)	31L(1)	S(1)
A2a 12.7	Denmark (1)	3L(1)	
A2a 12.7	England (1)		
A2a 12.7	Hawaii (2)		
A2a 12.7	Malaysia (1)		
A2a 12.7	Norway (2)		
A2a 12.7	Switzerland (4)		
A2a 12.7	United States (27)	3L(9), 13L(3)	
A2a 77	Switzerland (2)	3L(1)	
A2a 78	United States (2)	1L(1)†, 16L(1)	S(1)
A2a 79	Switzerland (1)	3L(1)	
A2a 12.5	Alaska (2)	3L(2)	
A2a 12.5	Argentina (3)	3L(1)	
A2a 12.5	Australia (24)		
A2a 12.5	Canada (63)	3L(7)	
A2a 12.5	Denmark (22)	3L(2)	
A2a 12.5	Dominican Republic (18)		
A2a 12.5	England (12)		
A2a 12.5	Finland (23)	3L(2)	
A2a 12.5	France (60)	3L(6)	
A2a 12.5	Greenland (3)		
A2a 12.5	Hawaii (2)		
A2a 12.5	Iceland (3)	3L(3)	S(1)
A2a 12.5	New Zealand (11)	3L(11)	
A2a 12.5	Netherlands (5)	3L(3)	
A2a 12.5	Norway (16)	3L(3)	
A2a 12.5	Papua New Guinea (2)	3L(2)	
A2a 12.5	South Africa (3)		

(*continued*)

Table 4. (*continued*)

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
A2a 12.5	Scotland (13)		
A2a 12.5	Spain (69)		
A2a 12.5	Sweden (34)		
A2a 12.5	Switzerland (88)	3L(10)	
A2a 12.5	United States (21)	3L(8)	S(1)
A2a 80	Denmark (1)		
A2a 80	United States (1)		
A2a 81	South Africa (1)	3L(1)	
A2a 82	France (2)		
A2a 82	Sweden (6)		
A2a 82	United States (1)	3L(1)	
A2a 83	Australia (1)		
A2a 84	Canada (1)		
A2a 85	Canada (1)	3L(1)	
A2a 21.8	Alaska (1)	1L(1)†	
A2a 21.8	Dominican Republic (1)		
A2a 21.8	Finland (19)	1L(3)	
A2a 21.8	Iceland (1)	1L(1)	
A2a 21.8	Malaysia (6)	1L(1)	
A2a 21.8	Sweden (4)		
A2a 21.8	United States (1)	1L(1)	
A2a 21.9	Sweden (1)		S(1)
A2a 21.5	Australia (3)		
A2a 86	United States (1)	3L(1)	
A2a 87	Finland (1)		
A2a 88	Finland (2)		
A2a 89	France (1)		
A2a 12.0	Alaska (1)	3L(1)	
A2a 12.0	Australia (2)		
A2a 12.0	Canada (4)	3L(1)	S(1)
A2a 12.0	Denmark (1)		
A2a 12.0	Finland (1)		
A2a 12.0	New Zealand (1)	3L(1)	
A2a 12.0	Switzerland (1)		
A2a 90	Canada (1)		
A2a 90	Sweden (1)		
A2a 91	Canada (1)	3L(1)	
A2a 92	United States (1)		S(1)
A2a 93	Canada (1)		
A2a 93	Dominican Republic (1)		
A2a 93	Finland (1)		
A2a 93	Scotland (2)		
A2a 93	Spain (1)		
A2a 93	Sweden (2)		
A2a 94	United States (1)		
A2a 95	United States (2)		
A2a 96	Denmark (1)		
A2a 96	Greenland (1)		
A2a 97	Canada (1)		
A2a 98	United States (1)		
A2a 99	Malaysia (2)	3L(1)	
A2a 100	Canada (17)	3L(3)	S(1)
A2a 100	Finland (2)		
A2a 101	United States (1)	14.1L(1)	S(1)
A2a 102	United States (1)	22L(1)	S(1)
A2a 103	Alaska (7)	5L(6), 3L(1), 5.1L(1)	S(1)
A2a 103	United States (1)		
A2a 104	Switzerland (2)	3L(1)	
A2a 105	Canada (1)		
A2a 105	Dominican Republic (1)		
A2a 105	South Africa (2)		
A2a 105	Switzerland (2)		
A2a 106	Alaska (1)	13L(1)	
A2a 107	Canada (1)		
A2a 108	Switzerland (1)	3L(1)	
A2a 108	United States (1)		
A2a 109	South Africa (1)	2L(1)	
A2a 110	Denmark (1)	1L(1)	
A2a 110	Finland (1)		
A2a 111	Canada (1)		
A2b 112	Canada (1)		
A2b 113	Canada (1)		
A2b 114	Alaska (1)	5L(1)	
A2c 115	South Africa (1)	2L(1)	
A2d 116	Dominican Republic (1)	16L(1)	
A2d 117	Dominican Republic (1)	16L(1)	
A2d 118	United States (1)	5L(1)	S(1)
A2e 21.6	Alaska (12)	13L(12)	
A2e 21.6	Canada (2)		
A2e 21.6	Denmark (1)		
A2e 21.6	Guatemala (1)		
A2e 21.6	New Zealand (1)	13L(1)	
A2e 21.6	Papua New Guinea (1)	13L(1)	
A2e 21.6	Philippines (1)		
A2e 21.6	Thailand (1)	13L(1)	
A2e 21.6	United States (37)	13L(21), 11L(1), 3L(1)†	S(8)
A2e 119	United States (1)	11L(1)	
A2e 120	Alaska (1)	13L(1)	
A2e 121	Canada (1)		
A2e 121	Hawaii (1)	13L(1)	
A2e 122	United States (2)		S(2)
A2e 123	United States (1)		
A2e 124	Alaska (1)	13L(1)	
A2e 125	United States (1)		
A2e 126	Denmark (1)		
A2f 127	Denmark (1)		
A3 128	Hawaii (1)		
A4 129	Canada (1)		
A5 130	South Africa (1)	1L(1)	
B1a 131	United States (1)		
B1a 132	Malaysia (1)	6U(1)	

(*continued*)

Table 4. (*continued*)

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
B1b 25.6	Canada (6)	6U(1)	S(2)
B1b 25.6	Malaysia (4)	6U(2)	
B1b 25.6	Papua New Guinea (26)	6U(4)	S(1)
B1b 25.6	Thailand (1)	6U(1)	
B1b 25.6	United States (10)	6U(9)	S(3)
B1b 133	United States (1)		S(1)
B1b 134	Canada (1)		
B1b 135	Papua New Guinea (2)	6U(2)	
B1b 136	Hawaii (1)		
B1b 137	Denmark (2)		
B1b 137	France (1)		
B1b 138	Netherlands (1)	6U(1)	S(1)
B1b 138	Papua New Guinea (1)	6U(1)	
B1b 139	Kenya (1)	6U(1)	
B1b 139	Malaysia (1)		S(1)
B1b 139	Sweden (1)		
B1b 140	Kenya (1)		S(1)
B1b 141	Canada (2)		
B1b 141	The Gambia (5)	6U(3)	
B1b 141	South Africa (4)		
B1b 141	Sweden (2)		
B1b 142	Hawaii (1)		
B1b 143	United States (1)	23U(1)	S(1)
B1b 144	The Gambia (1)		
B1b 145	Alaska (1)	6U(1)	
B1b 145	Canada (1)		
B1b 145	Dominican Republic (1)		
B1b 145	New Zealand (2)	6U(1), 12U(1)	S(1)
B1b 145	Papua New Guinea (3)		S(1)
B1b 145	United States (1)	24U(1)	S(1)
B1b 146	United States (1)		
B1b 147	Dominican Republic (6)		
B1b 148	Dominican Republic (1)		
B1b 149	Dominican Republic (1)	6U(1)	
B1b 150	South Africa (2)	6U(1)	
B1b 150	Switzerland (2)		
B1b 151	Switzerland (2)	6U(1)	S(1)
B1b 152	South Africa (12)	6U(2)	S(1)
B1b 153	South Africa (1)		
B1c 154	Dominican Republic (1)		
B1d 155	Alaska (1)	1U(1)	
B1e 156	South Korea (1)		
B1f 157	England (11)		d(9)
B1f 158	United States (1)		d(1)
B1f 159	England (4)		d(2)
B1f 160	United States (1)		
B1f 161	Papua New Guinea (1)		d(1)
B1g 162	Malaysia (8)		d(8)
B1g 163	Kenya (1)		d(1)
B2a 164	United States (1)	a1U(1)	aT(1)
B2b 165	Papua New Guinea (2)	a1U(2)	aT(2)
B2b 166	Papua New Guinea (1)	a1U(1)	aT(1)
B2b 167	The Gambia (2)	a1U(2)	aT(2)
B2b 167	Malaysia (7)	a1U(7)	aT(7)
B2b 168	Malaysia (1)	a1U(1)	aT(1)
B2c 169	Malaysia (2)	a1U(2)	aT(2)
B2d 170	Kenya (1)	a1U(1)	aN(1)
B3 171	The Gambia (2)		
B4 172	Japan (sero b) (1)		
B4 173	The Gambia (sero a) (3)	a5L(3)	aN(3)
B4 173	Malaysia (sero a) (1)	a1U(1)	aT(1)
B4 174	Kenya (sero a) (1)	a5L(1)	aN(1)
B5 175	Iceland (1)		
C1 176	Switzerland (1)		
D1a 177	United States (1)		c1(1)
D1a 178	Malaysia (1)		c1(1)
D1b 179	Unknown (1)		
D1b 180	Malaysia (3)		c1(3)
D1b 181	England (1)		c1(1)
D2 182	Unknown (1)		c1(1)
D2 183	England (1)		c1(1)
D3 184	United States (1)		c1(1)
D4 185	England (1)		c2(1)
D5 186	Kenya (1)		c2(1)
E1 187	Kenya (1)		
F1 188	England (1)		e(1)
F1 189	England (5)		e(2)
F1 190	England (2)		e(2)
F2a 191	England (2)		
F2a 191	United States (3)		e(1)
F2a 191	Unknown (1)		
F2a 192	United States (1)		
F2a 193	England (3)		e(1)
F2a 194	Malaysia (1)		e(1)
F2a 195	Unknown (3)		
F2a 196	England (2)		e(1)
F2a 197	England (4)		
F2a 198	England (1)		e(1)
F2a 199	Unknown (1)		e(1)
F2a 200	Unknown (1)		e(1)
F2a 201	England (1)		e(1)
F2b 202	England (1)		Anom‡
F2b 203	Papua New Guinea (1)		e(1)
F2c 204	England (31)		e(3)
F2c 205	Kenya (2)		
F2c 206	Malaysia (1)		e(1)
F2c 207	England (1)		e(1)
F2c 208	England (2)		e(1)

(*continued*)

Table 4. (*continued*)

Group ET	Geographic source (no. of isolates)	OMP pattern*	*cap* region RFLP pattern
F2c 209	England (1)		
F2c 210	England (2)		e(1)
F2c 211	England (1)		e(1)
F2c 212	England (1)		
F2c 213	Malaysia (1)		
F2c 214	England (2)		e(2)
F2c 215	England (5)		e(1)
F2c 216	England (6)		e(3)
F2c 217	England (1)		e(1)
G1 218	Kenya (1)		e(1)
H1a 219	England (13)	a2H(10), a7H(3)	aM(13)
H1a 219	Unknown (1)	a2H(1)	aM(1)
H1a 220	England (1)	a4H(1)	aM(1)
H1a 221	England (1)	a2H(1)	aM(1)
H1b 222	England (1)	a6H(1)	aM(1)
H1b 223	England (1)	a6H(1)	aM(1)
H1b 224	England (3)	a6H(3)	aM(3)
H1b 225	England (1)	a6H(1)	aM(1)
H1b 226	England (1)	a6H(1)	aM(1)
I1a 227	England (2)	a4H(2)	aM(2)
I1a 228	England (1)	a4H(1)	aM(1)
I1a 229	England (3)	a4H(2), a8H(1)	aM(3)
I1b 230	Dominican Republic (1)	a5H(1)	aM(1)
J1a 231	United States (1)		
J1a 232	United States (1)		O(1)
J1b 233	United States (1)	8-H§(1)	
J1b 234	Alaska (1)	8H(1)	
J1b 235	Alaska (2)	8H(2)	O(1)
J1b 235	Norway (1)	30H(1)	
J1b 236	Canada (1)	8H(1)	
J1b 236	United States (1)	8H(1)	
J1b 237	United States (1)	8H(1)	
J2 238	South Africa (1)	25L(1)	
J3 239	United States (1)	17H(1)	
K1a 240	United States (1)		
K1b 241	England (2)		O(1)
K1c 242	England (1)		F(1)
K1c 243	England (1)		
K1c 244	Malaysia (1)		F(1)
K1c 245	Malaysia (1)		
K1c 246	Malaysia (1)		
K1c 247	Malaysia (1)		
K1c 248	Malaysia (1)		
K1c 249	Malaysia (2)		F(1)
K1c 250	Malaysia (1)		F(1)
K1c 251	Malaysia (2)		F(1)
K1c 252	Malaysia (1)		
K1c 253	Malaysia (1)		F(1)
K1c 254	Malaysia (1)		F(1)
K1d 255	United States (1)		
K2 256	United States (1)		O(1)
K2 256	Unknown (1)		
K2 257	England (1)		
K2 258	England (1)		Un#(1)
K2 259	England (1)		O(1)
K2 260	England (7)		O(1),Un#(1)
K2 261	England (1)		O(1)
K2 262	England (2)		O(1)
K2 262	Unknown (2)		O(2)
K2 263	England (3)		O(1)
K2 264	England (1)		Un#(1)
K2 265	England (5)		O(3)
K2 266	England (4)		
L1 267	Kenya (1)		

* For most geographic source groups, strains were randomly selected for OMP subtype analysis. The OMP subtype designations for serotype b isolates are those reported by the Granoff laboratory [8, 19], and serotype a OMP patterns have been distinguished by the Moxon laboratory [25]. To avoid confusion, serotype a pattern designations are preceded by the letter "a." Note that serotype a and serotype b strains with the same OMP type designation may actually not be identical in OMP electrophoretic pattern. The *cap* region RFLP patterns have been reported previously [24, 26]. The number of isolates identified of each OMP pattern or *cap* region RFLP pattern is given in parentheses.

† Pattern similar to, but not identical with, designated OMP subtype.

‡ Anomalous.

§ Similar to type 8 but apparently lacking the P1 band.

Unclassified.

types account for 70% of all serotype b strains in the sample, and, moreover, 81% of all type b strains examined belong to only nine of the 280 multilocus enzyme genotypes identified (table 4).

cap Region RFLP Patterns and Genotypic Diversity in *cap* Region RFLP Pattern

The *cap* region is a 20- to 40-kilobase (kb) segment of the chromosome containing genes necessary for the production of capsule polysaccharide [33, 36, 37]. Previously 14 *cap* region RFLP patterns were identified among 222 isolates of six serotypes examined [21]. The approximate sizes (in kilobases) of the DNA fragments characterizing each pattern are indicated in figure 2; their distributions among the major branches of the dendrogram are shown in figure

1 and described in detail elsewhere [21]. All RFLP patterns were serotype specific.

For both serotypes a and b, the *cap* region patterns of isolates in the two primary phylogenetic divisions were very different. The a(T) and a(N) patterns of serotype a strains in division I have three of four fragments of common size, but fragments of similar size do not occur in the a(M) pattern of serotype a strains in division II. Serotype b strains in division I had patterns b(S), b(V), and b(G), which share most of their fragments, whereas serotype b strains in division II had the b(O) pattern, which has only two of its seven fragments (2.1 and 2.7 kb, which are type b specific) in common with the other serotype b patterns.

Each of the 14 *cap* region RFLP patterns occurred in isolates of several multilocus genotypes, with as many as 30 ETs being identified among the 53 isolates with pattern b(S). For patterns b(S), b(O), d, and e, mean genetic diversity (H) among ETs was roughly equivalent to that recorded for ETs of all isolates of the same serotype, whereas among ETs of the other probe patterns, H was, on average, equal to only 64% of that of ETs of the same serotype (table 5) [21]. These results reflect the circumstance that *cap* patterns are found in isolates that are nonrandom samples of clones of each serotype.

Table 5. Mean genetic diversity (H) among ETs within *cap* locus RFLP patterns of *H. influenzae*.

Pattern	No. of isolates	Mean no. of alleles	No. of ETs	H_{ETs}
a(T)	17	1.77	7	0.235
a(N)	5	1.29	3	0.196
a(M)	30	1.94	12	0.306
b(G)	14	1.88	10	0.261
b(V)	14	1.53	6	0.208
b(S)	53	2.65	30	0.314
b(0)	4	1.59	3	0.373
c(1)	9	1.82	7	0.283
c(2)	2	1.29	2	0.294
d	22	1.41	6	0.161
e*	30	2.65	23	0.257
f(F)	7	1.47	7	0.148
f(0)	12	1.53	8	0.183
f(un)†	3	1.29	3	0.176
Total	222	1.72	124	0.467

* One isolate with a slightly anomalous hybridization pattern was not included in the sample.
† un = unclassified.

Genotypic Diversity Within OMP Patterns of Serotype b Strains

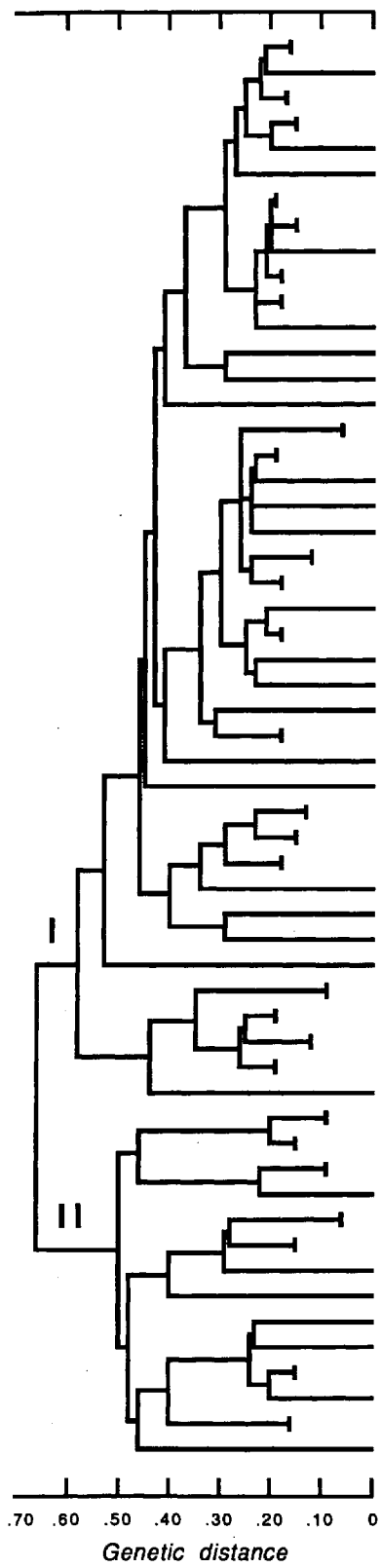
The OMP electrophoretic patterns of 587 serotype b isolates were determined and assigned to 33 categories. These include 25 patterns previously described [19, 20] and eight newly identified subtypes, as follows: 13.1L, which is similar to but distinguishable from 13L, and 25L, 26L, 27L, 29L, 31L, 30H, and 12U. Most OMP patterns were confined to somewhat closely related clones, with an average of 37% of the total genetic diversity and 54% of serotype b genetic diversity being apportioned among ETs represented by isolates of the nine more common OMP subtypes (table 6).

OMP types in cluster A1 were 1H (80 isolates), 2H (11 isolates), 1L (33 isolates), 2L (115 isolates), 3L (three isolates), 11L (two isolates), and 4H, 19H, 9L, 15L, 18L, 26L, and 1U, each of which was recorded in a single isolate. Cluster A2 contained OMP types 1L (26 isolates), 2L (16 isolates), 3L (172 isolates), 5L (eight isolates), 11L (four isolates), 13L (51 isolates), 14L (two isolates), 16L (six isolates), 18L (six isolates), and one isolate each of 5.1L, 13.1L, 14.1L, 22L, 27L, 29L, and 31L. All 6U isolates were in cluster B1, and 8H isolates were confined to cluster J1.

Mean genetic diversity per locus among ETs synthesizing the H, L, or U outer membrane protein P1 was, on average, 81%, 50%, and 47%, respectively, of that in the total sample of encapsulated strains and ~60% of that recorded for serotype b strains

Table 6. Genetic diversity in relation to OMP pattern and P1 type in serotype b *H. influenzae*.

OMP	No. of isolates	No. of ETs	H_{ETs}	Mean no. of alleles
1H	80	10	0.122	1.6
1L	59	19	0.223	2.2
2L	117	14	0.205	1.9
3L	180	30	0.162	2.8
11L	6	6	0.224	1.7
13L	50	7	0.171	1.5
16L	7	7	0.182	1.6
6U	33	11	0.165	1.6
8H	6	4	0.118	1.2
H (division I)	94	11	0.122	1.6
H (division II)	8	5	0.216	2.1
H (pooled)	102	16	0.377	2.6
L	447	76	0.233	3.8
U	38	14	0.218	2.1
Total	587	103	0.319	4.4



Cluster	N	*n*	Sero-type	RFLP cap pattern	Representative OMP Types	RFLP iga pattern
A1a	35	549	b	b(S), b(V), b(G)	1H,1L,2H,2L	II
A1b	1	1	b			II
A1c	12	16	b	b(S)	1L,2L,3L,11L,26L	II
A1d	3	4	b	b(V), b(G)	1H	
A1e	1	1	b			
A1f	1	1	b			
A2a	69	1189	b	b(S)	1L,2L,3L,13L,18L	I
A2b	3	3	b		5L	
A2c	1	1	b		2L	
A2d	3	3	b	b(S)	5L,16L	I
A2e	9	67	b	b(S)	11L,13L	I
A2f	1	1	b			
A3	1	1	b			
A4	1	1	b			
A5	1	1	b		1L	
B1a	2	2	b		6U	
B1b	22	114	b	b(S)	6U,12U,23U,24U	III
B1c	1	1	b			
B1d	1	1	b		1U	
B1e	1	1	b			
B1f	5	18	d	d		
B1g	2	9	d	d		
B2a	1	1	a	a(T)	1U	
B2b	4	13	a	a(T)	1U	
B2c	1	2	a	a(T)	1U	
B2d	1	1	a	a(N)	1U	
B3	1	2	b			
B4	3	6	a,b	a(N), a(T)	1U,5L	
B5	1	1	b			
C1	1	1	b			
D1a	2	2	c	c(1)		
D1b	3	5	c	c(1)		
D2	2	2	c	c(1)		
D3	1	1	c	c(1)		
D4	1	1	c	c(2)		
D5	1	1	c	c(2)		
E1	1	1	c			
F1	3	8	e	e(1), e(2)		
F2a	11	24	e	e(1)		
F2b	2	2	e	e(1)		
F2c	14	57	e	e(1), e(2), e(3)		
G1	1	1	e	e(1)		
H1a	3	16	a	a(M)	2H,4H	
H1b	5	7	a	a(M)	6H	
I1a	3	6	a	a(M)	4H	
I1b	1	1	a	a(M)	5H	
J1a	2	2	b	b(O)		
J1b	5	8	b	b(O)	8H,8-H,30H	
J2	1	1	b		25L	
J3	1	1	b		17H	IV
K1a	1	1	f			
K1b	1	2	f	f(O)		
K1c	13	15	f	f(F)		
K1d	1	1	f			
K2	11	30	f	f(O), f(Un.)		
L1	1	1	f			

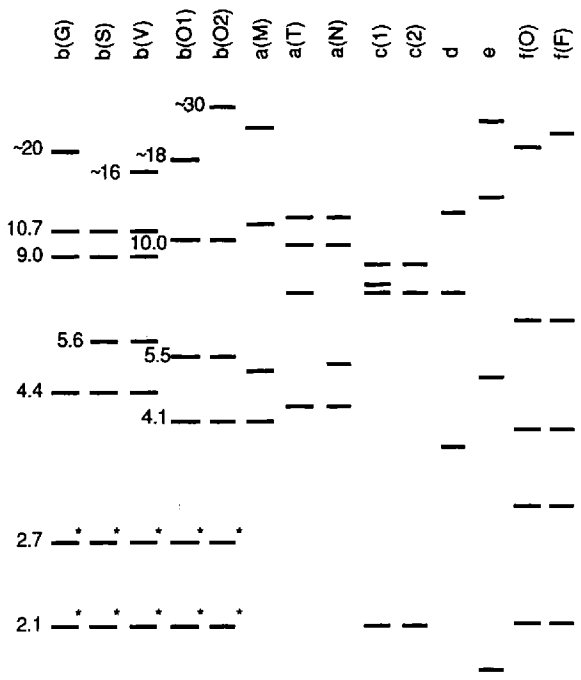


Figure 2. RFLP patterns of the *cap* region in encapsulated *H. influenzae*. The figure represents a Southern blot of *Eco*RI-digested chromosomal DNA from the known variants of the six serotypes hybridized to pUO38 [31], a DNA probe containing the entire *cap* locus from a serotype b strain. Bands marked with an asterisk denote DNA confined to serotype b strains. The sizes of fragments identified in serotype b strains are shown in kilobases. Similarity in size in strains of different serotype does not carry the implication of nucleotide sequence identity.

(table 6). This limited genetic diversity within P1 types (especially OMP subtype 1H) reflects the fact that isolates of the same OMP type are likely to be closely similar or identical in overall genetic character, a conclusion earlier deduced from observed associations between OMP type and biotype and between OMP type and ET in a sample from the United States [20].

Table 7. Proportion of ETs shared between pairs of serotype b *H. influenzae* isolates of various OMP patterns.

OMP pattern	OMP pattern								
	1H	1L	2L	3L	11L	13L	16L	6U	8H
1H	–	0.074	0.091	0.000	0.067	0.000	0.000	0.000	0.000
1L		–	0.065	0.043	0.042	0.040	0.083	0.000	0.000
2L			–	0.023	0.053	0.050	0.050	0.000	0.000
3L				–	0.029	0.088	0.028	0.000	0.000
11L					–	0.083	0.000	0.000	0.000
13L						–	0.077	0.000	0.000
16L							–	0.000	0.000
6U								–	0.000
8H									–

Nine OMP types were represented by isolates assigned to four or more ETs, and the largest number of ETs for these nine classes was recorded for OMP 3L, which occurred in isolates of 30 ETs. However, most OMP subtypes were confined to isolates of ETs in only one or two groups of serotype b strains; exceptions were subtypes 1L, 2L, 3L, and 11L, which occurred in isolates assigned to 5, 4, 3, and 4 groups, respectively, and there was extensive sharing of ETs among these and other OMP subtypes (table 7).

The common OMP subtypes were 1H (13% of isolates), 1L (10%), 2L (22%), 3L (29%), 13L (8%), and 6U (5%). Seventy-nine (99%) of the 1H isolates were ET 1.9 or other closely related ETs in group A1a; 98% of 3L isolates were of ETs in group A2a; and 97% of 6U isolates were assigned to group B1b. OMP subtypes 1L and 2L were unusual in that they occurred in significant proportions of isolates assigned to two different clusters (A1 and A2) of the dendrogram: for 1L isolates, 50% and 43%, respectively, were in groups A1a and A2a; and 2L isolates occurred in groups A1a (88%) and A2a (12%).

In summary, with a single exception, all isolates assigned to cluster A1 were of OMP types with the P1 protein designated H or L; in contrast, all iso-

Figure 1. Dendrogram showing genetic relationships of the 280 ETs of encapsulated *H. influenzae* and serotypes, *cap* region RFLP patterns, common OMP subtypes, and *iga* RFLP patterns for the various lineages. The dendrogram was generated by the average-linkage method of clustering from a matrix of coefficients of pairwise genetic distance, based on 17 enzyme loci. There are two primary divisions (I and II) and 12 major lineages (A through L). Groups represented by multiple ETs diverging from one another at a genetic distance of 0.20 or less were truncated; the point of truncation indicates the deepest level of divergence of families of ETs within the cluster. The number of ETs (N) and number of isolates (*n*) in each group are indicated. For serotype a and serotype b isolates, all OMP subtypes occurring in association with each RFLP pattern in each group are indicated, except for clusters A1 and A2, for which several additional OMP subtypes have been identified [20, 21]. The distribution of the *iga* RFLPs is based upon data in [110]. Note that serotype a and b strains with the same OMP subtype designation may not be identical in actual OMP electrophoretic pattern (see Materials and Methods). Additional OMP subtypes identified in each group are listed in table 4.

lates in cluster A2 had the L OMP variant, and all isolates in cluster B1 had the U variant. The H band also was present in isolates assigned to clusters J1 and J3, and the single strains comprising cluster J2 had the L variant. We identified 76 ETs among isolates expressing the L variant, 17 ETs among H-variant strains, and 14 ETs among strains typed as U.

Genotypic Diversity Within OMP Patterns of Serotype a Strains

The distribution of OMP patterns of serotype a isolates was reported previously [21]. Briefly, all isolates of the seven ETs in cluster B2 were OMP type 1U; of the five serotype a isolates in cluster B4, four were OMP subtype 5L and the fifth was OMP 1U.

All 30 serotype a isolates in lineages H and I in division II had the H mobility variant of the 45-kDa heat-modifiable OMP. The most common OMP subtype among the 23 type a isolates in cluster H1 was 2H, which occurred in 12 isolates; one, seven, and three isolates had OMP subtypes 4H, 6H, and 7H, respectively. Three OMP patterns were identified among the seven serotype a isolates in lineage I1: 4H (five isolates) and 5H and 8H (one isolate each).

In sum, all isolates of serotype a assigned to ETs in division I had the heat-modifiable U or L mobility variant of the 45-kDa OMP, and all type a isolates in division II had the H mobility variant of this protein.

Genetic Diversity in Relation to Disease Type

Because a variety of serious diseases are caused by strains of *H. influenzae* [3, 38], we were interested in determining the extent of association of the phylogenetically diverse types with particular diseases and anatomic sites of infection. Our analysis revealed no strong statistical association on a global or continental level between ETs and disease type or anatomic site of recovery of case isolates, but in Finland there was a tendency for serotype b strains of ET 21.8 to cause more meningitis and less epiglottitis compared with stains of the ET 12 clone family ($P = .078$, Fisher's exact test) [39, 40]. However, this trend was not observed among isolates of ET 21.8 from other parts of the world.

Ten of 22 serotype a isolates of ETs in clusters B2 and B4 (in division I), which are allied with abundant invasive clones of serotype b, were recovered from invasive episodes, whereas only one of 30 isolates of ETs in clusters H1 and I1 (in division II), which are related to lines of serotype b ETs and serotype f ETs that are rarely recovered from invasive infections, was obtained from a normally sterile body fluid (table 8).

Serotype b Isolates from Native Human Populations and Patients of Known Ethnic Background

Three samples of isolates recovered from native North American populations were analyzed. A sample of 87 isolates from Alaska, which was collected as part of an efficacy trial of serotype b conjugate vaccine [41], included strains from Yupik and Inuit Eskimos, Amerinds, and individuals with mixed Native and non-Native American ancestry. Detailed epidemiologic data and strain characterization will be presented elsewhere (J. I. W. and J. M. M., manuscript in preparation). In brief it was found that: (1) among nonnatives in Alaska there were relatively minor differences in ET and OMP subtypes compared with strains recovered from elsewhere in North America, (2) strains of ET 1.9 were more frequently recovered from invasive disease in nonnative (predominantly Caucasian) individuals than in natives, (3) isolates of certain rare ETs were recovered from native populations, often only in isolated geographic areas, and (4) there is considerable variation among the various populations in the clones recovered from asymptomatic carriers and patients with invasive disease. The most striking observation was that 11 of the 12 isolates identified as ET 21.6/OMP 13L were carrier strains [42].

A sample of 141 isolates from Manitoba included 72 isolates from patients whose ethnic backgrounds were known: 52 strains were from Caucasians, 18 from Amerinds, and one each was from an Inuit and an "oriental" patient. Although 25% of these isolates were recovered from Amerinds, only 13% were identified as ET 1.9 (two isolates each from patients

Table 8. Genetic structure in relation to pathogenicity for serotype a isolates of *H. influenzae*.

	No. of isolates	
	Invasive	Noninvasive
Division I	10	12
Division II	1	29

NOTE. $P < .001$

with meningitis and cellulitis). Four of the six isolates of ET 25.6 were recovered from Amerindian patients with meningitis; each patient lived in a different part of the province (one isolate each from Garden Hill, Sandy Lake, St. Theresa Point, and Winnipeg).

Thirty-two isolates (21 invasive and 11 carrier strains) recovered from White Mountain Apache Indians living on the Fort Apache Indian Reservation in Arizona [43] were studied. Twelve invasive isolates and six carrier isolates were ET 1.9, and five invasive and five carrier isolates were ET 21.6; of the remaining isolates, all of which were recovered from invasive episodes, two were ET 122, and one each was ET 25.6 and ET 32. The *cap* region RFLP pattern was determined for 15 of the ET 1.9 isolates: 12 had the b(V) pattern (and, hence most certainly are not OMP pattern 1H, which is charcteristic of most United States ET 1.9 isolates), and three were typed as b(S).

The collection included 12 isolates from Japan: nine isolates were ET 12.8, two isolates were ET 12.9, and one isolate was assigned to ET 172.

Eighty-one isolates from patients in Malaysia were analyzed: 40 were obtained from Asian Indians, 24 from Malays, 15 from Chinese, and two from individuals whose ethnicity was not recorded. Twenty-seven of the isolates (33%) were obtained from blood or CSF, including 10 isolates from Malays, nine from Asian Indians, and eight from Chinese patients. Most of the remaining isolates were cultured from sputum, usually from patients with lower respiratory tract infection. For these individuals, the serotype b strains were recovered as a pure culture and are presumed to be the etiologic agent. The ethnic background of the patients and ETs of the blood and CSF isolates were as follows: ethnic Malays, six isolates of ET 12.8 and one isolate each of ET 60, ET 99, ET 132, and ET 25.6; Asian Indians, six isolates of ET 12.8 and one isolate each of ET 1.8, ET 60, and ET 25.6; and Chinese, six isolates of ET 12.8 and one isolate each of ET 60 and ET 21.8.

Among the 50 isolates associated with invasive episodes in the Republic of South Africa, 24 were obtained from blacks, two from "colored" individuals, and three from whites; the remainder were from patients of unknown ethnicity. The isolates from blacks included ET 8 (two isolates), ET 12.8 (two isolates), ET 105 (two isolates), ET 141 (four isolates), ET 152 (eight isolates), and one isolate each was ET 60, ET 68, ET 12.5, ET 109, ET 115, ET 130, and ET 153. The two isolates from colored patients were of ET 12.8 and ET 74, and the isolates from white patients each represented a unique ET, these being ET 35, ET 76, and ET 12.5.

Genotypic Variation in Relation to Geographic Source

Of the total of 1,975 serotype b isolates examined, 1,609 (81%) were assigned to one of the following nine ETs, all of which were represented by 20 or more isolates: ETs 1.9, 4, 12.5, 12.7, 12.8, 12.9, 21.6, 21.8, and 25.6 (table 9); the distribution by country of isolates of these nine ETs is shown in table 10.

At least three of these common ETs (ET 1.9, 12.5, and 12.8), and perhaps most of them, have very wide, even global, distributions. However, for each geographic region, our analysis revealed the presence of only one or two common ETs. For example, isolates of ET 1.9 have been recovered in 12 states in the continental United States, eight provinces of Canada, eight countries in western Europe, and in Mexico, the Dominican Republic, Hawaii, Papua New Guinea, the Republic of South Africa, and Australia. Of the 934 North American isolates, 379 (41%) were ET 1.9 and 151 (16%) were ET 12.8; no other single ET accounted for more than 10% of the isolates from this continent. Fifty-six percent of the 622 European isolates were ET 12.5, and 13% were ET 12.8. The commonest clone among Asian isolates was ET 12.8 (64%), and isolates of this ET were also the commonest multilocus genotype recovered in Africa, accounting for 31% of strains from that continent. Papua New Guinea was most unusual in that ET 25.6 was represented by 43% of isolates; ET 12.8 also was commonly recovered, accounting for 37% of isolates.

For serotype b strains, 68 ETs were represented by two or more isolates (mean 3.0). Thirty-four ETs (50%) were represented in two or more geographic source groups: 17 ETs from two areas, eight from three areas, five from four areas, and three from five areas; ET 12.8 was represented in all six geographic source groups. However, there is marked geographic variation in genotype representation (table 11). For example, 84% (n = 379) of strains of ET 1.9 were recovered in North America, and 70% (n = 348) of strains of ET 12.5 were from patients in Europe. Similarly, 30 (94%), 64 (84%), and 51 (89%) of the strains of ET 4, ET 12.7, and ET 21.6, respectively, were collected in North America (table 9).

Iceland was unusual compared with other European countries studied in that 82% of isolates were

Table 9. ETs of serotype b *H. influenzae* represented by 20 or more isolates, grouped by geographic source.

	No. (%) of isolates from indicated geographic source						
ET	North America (n = 934)	Europe* (n = 622)	Asia (n = 121)	Africa (n = 95)	Papua New Guinea (n = 60)	Other (n = 143)	Total (n = 1,975)
1.9	379 (41)	56 (9)	0	2 (2)	1 (2)	12 (8)	450 (23)
4	30 (3)	2 (<1)	0	0	0	0	32 (2)
12.5	86 (9)	348 (56)	0	3 (3)	2 (3)	58 (41)	497 (25)
12.7	64 (7)	8 (1)	1 (<1)	0	0	3 (2)	76 (4)
12.8	151 (16)	81 (13)	77 (64)	29 (31)	22 (37)	32 (22)	392 (20)
12.9	1 (<1)	14 (2)	7 (6)	0	0	1 (<1)	23 (1)
21.6	51 (5)	1 (<1)	2 (2)	0	1 (2)	2 (1)	57 (3)
21.8	4 (<1)	24 (4)	6 (5)	0	0	1 (<1)	35 (2)
25.6	16 (2)	0	5 (4)	0	26 (43)	0	47 (2)
Other	152 (16)	88 (14)	23 (19)	61 (64)	8 (13)	34 (24)	366 (18)

* If isolates from Iceland are excluded, the data for Europe are: ET 1.9 (4%), 4 (1%), 12.5 (60%), 12.7 (1%), 12.8 (13%), 12.9 (2%), 21.6 (1%), 21.8 (4%), 25.6 (0), and other (15%).

Table 10. Distribution by country of *H. influenzae* serotype b ETs represented by 20 or more isolates.

		Percentage of total sample represented by each ET									
Geographic source	n	1.9 (23)	4 (2)	12.5 (25)	12.7 (4)	12.8 (20)	12.9 (1)	21.6 (3)	21.8 (2)	25.6 (2)	Other (18)
Australia	40	7		60	2	5					26
Canada	375	38	1	17	6	19	<1	<1		2	24
Denmark	43	7		53	2	4		2			34
Dominican Republic	55	2		33		36	<1		2		26
England	21	14		57	5	5					19
Finland	100	2		23		44	3		19		9
France	77	1		78		6	1				14
The Gambia	30					60					40
Iceland	40	82	2	7					2		7
Malaysia	81				1	64			7	5	23
Norway	37	16		43	5	14	1				21
Papua New Guinea	60	2		3		37			2	43	13
Republic of South Africa	50	4		6		18					72
Scotland	20			65		5					30
Spain	71			97							3
Sweden	78	8	1	44		18	6		5		18
Switzerland	123	2		72	3	7	2				14
United States	472	44	4	4	6	16		8	<1	2	15

Table 11. Proportion of ETs shared between pairs of serotype b *H. influenzae* isolates from various geographic source groups.

	Source group					
Source group	North America	Europe	Asia	Africa	Papua New Guinea	Other
North America	–	0.155	0.061	0.048	0.056	0.118
Europe		–	0.116	0.130	0.078	0.156
Asia			–	0.073	0.130	0.150
Africa				–	0.091	0.098
Papua New Guinea					–	0.152
Other						–

ET 1.9; and, of these strains, 94% were OMP 2L. One strain of OMP 1H was identified. The *cap* region RFLP pattern was determined for three isolates classified as ET 1.9/OMP 2L and the single strain of ET 1.9/OMP 1H: two of the ET 1.9/2L isolates had the b(S) pattern, and the third isolate was typed as b(V); the ET 1.9/1H isolates had the b(G) *cap* region pattern. The ET 1.9/1H/b(G) strain was recovered from a 5-year-old American child with epiglottitis who lived on the U.S. naval base at Keflavik. The remaining isolates of ET 1.9 were cultured from patients with invasive disease living in communities throughout Iceland: Reykjavik (18 isolates), Akureyri

(three isolates), Kopavogur (two isolates), Keflavik town (two isolates), and one isolate each was from Husavik, Eskifjordur, Djupivogur, Egilsstadir, Budardalur, Vestmannaeyjar, and Gardabaer. Five additional strains were recovered from individuals in Reykjavik (two were ET 12.5, and one each was assigned to ET 4, ET 72, and ET 175), one strain was from Breiddalsvik (ET 12.5), and one isolate of ET 21.8 was from a patient in Hafnarfjorour.

Variation in the frequency of certain ETs also occurred between geographically contiguous countries that are closely allied historically, ethnically, and politically. One striking example, for serotype b isolates, was the occurrence of ET 100 (17 isolates) throughout Canada (Ontario, Manitoba, Alberta, and British Columbia) but its absence in the sample from the United States. Similarly, 17% of Canadian isolates, but only 4% of isolates from the United States, were ET 12.5.

Geographic variation in frequency was also recorded for ETs of isolates of other capsule serotypes; in many cases this was a result of the occurrence of one or a few "private" alleles [44] only in certain areas. For example, the two serotype e isolates from Kenya (ET 205) in the collection were distinguished from isolates assigned to the common serotype e genotype (ET 204) only by the occurrrence of a different allele at the PGM locus. Similarly, all serotype d isolates from Malaysia were represented by ET 162, which differs at two loci (PE1 and CAT) from the common clone (ET 157) of serotype d from the United Kingdom (table 2).

In summary, there is striking geographic variation in the frequency of recovery of certain multilocus genotypes. For serotype b organisms, ET 12.8 was frequently recovered in all regions, whereas isolates of ET 1.9, ET 12.5, and ET 25.6 were common only in single geographic source groups.

Serotype b Isolates from Asymptomatic Carriers

Isolates of seven distinctive multilocus genotypes were recovered from asymptomatic carriers. Of these isolates, seven were ET 1.9 (including five from Amerinds in Arizona), seven were ET 12.8 (United States, South Korea, and Canada), five were ET 12.9 (South Korea), two were ET 12.7 (Alaskan Natives), 47 were ET 12.5 (including 46 from associated individuals in three day care centers in Spain; [28]), five were ET 103 (Alaskan Natives), and 16 were ET 21.6 (11 from Alaskan Natives and five from White Mountain Apaches in Arizona).

Discussion

For purposes of reference and discussion, ETs are considered to mark clones, and isolates of a given ET with different OMP types, *cap* region RFLP patterns, biotypes, or LPS serotypes are regarded as subclones [20, 21].

Estimating Genetic Diversity and Relatedness Among Isolates

For *Escherichia coli* and *Shigella* species [45], *Legionella* species [46, 47], *Bordetella* species [48], *Salmonella* species (R. K. S., unpublished data), *Gluconobacter* species [49], the oral streptococci [50], *Listeria* species (J.-C. P., unpublished data), other bacteria [51, 52], and many higher organisms [53–55], estimates of genetic relatedness based on multilocus enzyme electrophoresis have been shown to be strongly correlated with measures of similarity in chromosomal nucleotide sequence derived from DNA hybridization experiments. Therefore, there is reason to believe that the 17 enzyme loci examined in this study are a representative sample of encapsulated *H. influenzae* genomes and, hence, provide a basis of estimating both overall genetic relationships among strains and levels of genetic diversity within populations.

Nature of the Sample Studied

Most of the strains analyzed were recovered in countries with modern bacteriology laboratories and strong affiliation with the United States or European countries. Samples from these countries were recovered from a very large number of different geographic sources (table 1) (and in some cases from two or more hospitals in the same community), and no consistent bias in isolate selection was identified. Hence, we are confident that the serotype b samples are generally representative of natural populations of medically important *H. influenzae*. This is especially true for the samples from Alaska and Finland, both of which were collected as part of population-based epidemiologic studies performed in conjunction with vaccine efficacy trials [41, 56] (a random sample of strains recovered from each area was analyzed), and from other regions represented by rela-

tively large samples (e.g., Canada, Switzerland, and Malaysia). Numerous attempts were made, without success, to obtain isolates from parts of the world not represented in the collection but of particular interest, such as the USSR, the People's Republic of China, India, and Japan (see beyond). Because the clonal composition of populations of *H. influenzae* varies geographically, additional ETs of encapsulated isolates undoubtedly exist in parts of the world that have yet to be sampled, but further sampling is not expected to affect the qualitative aspects of our findings.

Variation in Serotype in Relation to Population Structure

Although capsular polysaccharide serotypes clearly do not mark individual clones of *H. influenzae*, two lines of evidence demonstrate that capsule type has some relation to the phylogenetic structure of the species [21, 22]. First, none of the 280 ETs identified was represented by isolates of more than one serotype. Second, isolates expressing each of the six capsule polysaccharides belong to only one or a few major phylogenetic lineages, which in certain cases are strongly differentiated from the lineages of all other capsule types. This restriction accounts for the fact that mean genetic diversity per enzyme locus among ETs within serotypes was, on average, only two-thirds that recorded for all ETs. The latter observation, in conjunction with other data [24], is most parsimoniously interpreted as evidence that the common clones of each serotype were derived from a relatively small number of ancestral clones. In the case of serotype b isolates, this number is probably three or four, depending upon the number of horizontal gene transfer events that gave rise to serotype b isolates in phylogenetic line II (see beyond) [22].

Modal multilocus genotypes (ETs) of strains of the major lineages of each serotype are shown in table 12. These were constucted by taking the most frequent allele among ETs at each of the 17 enzyme loci assayed. (Because the modal multilocus genotypes were inferred on the basis of allele frequencies among ETs rather than isolates, they are not influenced by variation in the frequency of isolation of the ETs.) For serotype b, the modal ETs for the three major lineages of division I are the commonly recorded genotypes designated as ETs 1.9, 12.8, and 25.6, which may be interpreted as evidence that these genotypes are genetically similar to the three ancestral clones from which other clones of type b isolates have diverged.

In the laboratory, a type b strain can be transformed to express each of the other five capsule types with chromosomal DNA derived from isolates of the appropriate serotype [57]. Recently it has been shown by restriction enzyme mapping studies that the *cap* region of representative type a, b, c, and d strains is organized in a common fashion [58]. The *cap* region contains a central DNA segment coding for serotype-specific functions, and DNA probes derived from this region hybridize only to chromosomal DNA from isolates of the same capsule type: they are serotype specific [58]. The sizes of these regions are estimated to be (in kilobases) 3.2–5.4 (serotype a), 5.8–8.2 (serotype b), 5.3–7.7 (serotype c), and

Table 12. Modal ETs of common strains of encapsulated *H. influenzae*.

	Allele at indicated enzyme locus																
Serotype	CAK	NSP	PGI	MAE	MDH	G6P	GOT	ADK	6PG	PE2	PE1	LAP	PGM	CAT	GLD	G3P	FUM
a (division I)	6	5	5	3	5	6	3	2	2	4	5	2	1	2	4	2	3
a (division II)	7	5	4	1	5	1	3	2	5	2	3	2	4	3	1	2	2
b (lineage A1)	9	5	1	5	5	3	5	2	2	4	5	2	4	2	4	2	2
b (lineages A2-A5)	8	5	4	5	5	3	3	2	2	4	5	6	4	2	4	2	2
b (lineages B1-C1)	6	5	5	3	3	6	3	2	2	4	5	3	6	2	4	2	2
b (lineages J1b-J3)	7	5	4	2	5	1	3	1	4	3	5	2	1	3	4	2	2
c	6	5	10	5	1	6	5	2	2	4	1	2	6	2	4	2	2
d	6	5	5	3	5	6	3	2	2	4	5	3	6	2	4	2	2
e	9	5	9	2	5	3	3	2	2	2	3	3	7	2	5	2	2
f	7	5	2	2	5	1	3	1	5	2	2	3	4	3	4	2	2

NOTE. Modal ETs are combinations of the most common allele at each of the 17 enzyme loci. Enzyme abbreviations are defined in footnote to table 3.

5.3–8.7 (serotype d). The *cap* regions in serotype e and serotype f strains have not yet been examined in as much detail.

In areas flanking the serotype-specific region, there is great conservation of restriction endonuclease cleavage maps. Hence, the physical arrangement of chromosomal loci for capsulation functions in strains of these four serotypes is very uniform, and it has been suggested that the *cap* region is organized in such a way that exchange of central region DNA by horizontal gene transfer and recombination will lead to alteration of capsule structure [58]. It is highly probable that horizontal gene transfer of the central region of *cap* has been important in generating a small number of numerically unsuccessful cell lines of minimal clinical significance in phylogenetic line II [22]. However, because our analysis revealed that individual ETs are comprised solely of isolates expressing only a single capsule polysaccharide, we infer that alteration of capsule structure as a result of horizontal transfer of the serotype-specific region of the *cap* locus rarely occurs in natural populations. The observation that many type b isolates from invasive episodes fail to undergo DNA-mediated transformation in the laboratory [59] is consistent with this notion.

Variation in OMP in Relation to Population Structure

The OMP electrophoretic typing system was developed to classify serotype b and a isolates for epidemiologic studies [8, 18, 19]. For isolates of these serotypes, there are strong nonrandom associations between OMP type and ET, with little sharing of OMP subtypes between isolates of ETs in different phylogenetic clusters [20, 21]. However, as demonstrated here, individual OMP patterns may occur in genetically diverse clones. For the nine common serotype b OMP subtypes represented in the sample, mean genetic diversity was, on average, 53% of that in the total sample of serotype b isolates, with an average of nine ETs per OMP subtype. The practical significance of this finding for epidemiologic studies is that the grouping of strains solely on the basis of OMP subtype may lead to erroneous conclusions concerning strain relationships. This is especially true for subtypes 1L and 2L, which commonly occur in two different clone families. The occurrence of OMP subtypes 1L and 2L in genetically divergent clones means that these patterns do not consistently mark distinct multilocus genotypes. In contrast, there is strong statistical association of OMP subtypes 1H and 6U with certain multilocus enzyme genotypes, and isolates expressing OMP type 3L, although confined predominantly to lineage A2a (figure 1), are represented by 30 distinct ETs (table 6). Hence, until better methods are available, fine-structure epidemiologic studies should employ a combination of ET and OMP, ET and *cap* region RFLP, or ET, OMP, and *cap* region RFLP analyses.

Considering the abundant genetic diversity in enzyme genes identified among strains of certain OMP patterns, it is likely that the structural genes coding for individual OMPs of the same electrophoretic mobility are also polymorphic. Consistent with this notion is the recent observation [60] that OMPs of type b isolates with identical apparent molecular weights may possess different surface-exposed epitopes. It would clearly be of interest to perform comparative DNA sequencing on genes for individual OMPs selected with reference to the population genetic framework presented here. Studies of this type could answer the question of whether the occurrence of the same OMP pattern in strains belonging to different phylogenetic clusters of the dendrogram represents horizonal transfer events or the retention of an ancestral OMP pattern.

Takala et al. [61] recently reported that in Finland type b strains of OMP subtypes 1 and 1c (3L and 1L, respectively, in the system described by Barenkamp [8]) are nonrandomly associated with different types of invasive disease. The observation that strains of subtype 1c caused proportionally more meningitis and less epiglottitis than did those of subtype 1 was interpreted as evidence of a true difference in virulence between isolates expressing these OMP types, and it was hypothesized that the subtype 1c protein marks a clone with special virulence properties. The present study and others [39, 40] have shown that the OMP subtype 1 in Finland is commonly associated with ET 12.5, ET 12.8, and a small number of closely allied clones, whereas subtype 1c usually occurs in isolates of ET 21.8 (83% of isolates [37]), which differs from ET 12.5 and ET 12.8 at two and one loci, respectively, of the 17 enzyme loci assayed. Because all properties tend to be associated in clonal organisms [34, 62], the results are consistent with the interpretation that in Finland the type 1c outer membrane protein subtype frequently marks a subclone that is especially successful in invading the CSF. However, the occurrence in high frequency of this OMP subtype in strains recovered

from cases of meningitis does not mean that it, per se, has any role in virulence or organotropism.

Clonal Structure and Antibiotic Resistance

The strongly clonal population structure identified for encapsulated *H. influenzae* strains, coupled with an apparently relatively slow rate of clone dissemination over wide geographic areas, suggests that regional variation may also occur in antibiotic resistance elements. Although we made no systematic effort to identify antibiotic-resistant strains, the molecular epidemiology of some of the resistant serotype b isolates we examined has been studied by other investigators. Mendelman et al. [63] characterized 36 ampicillin-resistant type b strains recovered from native and nonnative individuals in four geographic regions of Alaska by OMP subtyping and restriction endonuclease plasmid profiling. Of the seven strains harboring detectable extrachromosomal DNA, four had a plasmid with a molecular mass of 40 megadaltons (MDa), and three contained a plasmid of 3 MDa. Isolates bearing the 3-MDa plasmid were identified only in the south-central region of the state; the 40-MDa plasmid from isolates originating in central Alaska was distinguishable from all other 40-MDa plasmids in Alaskan strains by restriction endonuclease analysis [63]. Similarly, the 40-MDa plasmids carried by isolates from individuals in northwest Alaska were identical to each other but differed in restriction endonuclease digestion pattern from 40-MDa resistance plasmids from other Alaskan regions. There were three distinct OMP subtypes [63] and nine ETs (data not shown) represented among the ampicillin-resistant strains.

Among the 66 ampicillin-resistant *H. influenzae* organisms from patients in the continental United States studied by Willard et al. [64], four plasmids of various sizes (3 MDa, 30 MDa, 36 MDa, and 50 MDa) were identified, but the Alaskan 40-mDa plasmid was not represented.

In a study of the molecular epidemiology of multiply resistant *H. influenzae* type b isolates from carriers and patients attending four day care centers in Spain, restriction endonuclease analysis revealed the presence of at least four distinct plasmids of molecular mass 45 MDa and 52 MDa and three different OMP subtypes [65].

The results from these studies, other reports [66], and unpublished data of the authors of this review indicate that extensive genetic diversity exists among the plasmids encoding antibiotic resistance genes and among the chromosomal genotypes of the resistant type b strains. Hence, unlike one documented case for *E. coli* [67], no single antibiotic-resistant clone has widespread geographic range: antibiotic-resistant determinants have been acquired by type b strains through multiple independent evolutionary events.

Rarity of Significant Associations of Clones and Clinical Syndromes

One situation was identified in which there was a statistically significant association of a clone with a particular clinical syndrome—ET 21.8 and meningitis in Finland—but no other obvious associations of clones and invasive disease types were revealed. For example, there was no unusual tendency for ET 21.8 to cause meningitis in countries other than Finland; and in Malaysia ET 21.8 may be more frequently associated with respiratory tract disease than meningitis (five of six isolates). However, only for the samples from Finland and Alaska were there available detailed clinical histories for all isolates.

The genetic distances separating the three clinically important major lineages of type b organisms are approximately equivalent to those dividing pig-, dog-, and fowl-specialist clones of *Bordetella bronchiseptica* [68] and separating clones of *Salmonella* species exhibiting host specialization [69]. It was therefore somewhat surprising that no strong disease-clone associations for the major serotype b phylogenetic lineages in division I were found. Although clones of all major serotype b lineages apparently are more or less equivalent in their ability to cause septicemia, meningitis, cellulitis, epiglottitis, or other invasive disease, it may well be that the failure to demonstrate clonal disease specificity was caused by a lack of detailed clinical information for the great majority of strains examined. In many areas sampled, a tendency was observed for association of clones and disease types, but these never reached statistical significance. Moreover, a clone or subclone tending to cause a proportionally greater amount of, for example, meningitis in one geographic area might cause less meningitis and more septicemia, epiglottitis, or other type of disease in another area with a different ethnic patient population. For example, in Pakistan and Papua New Guinea, an unusually high frequency of pneumonia with septicemia is

caused by strains of the ET 25.6 clone family (ET 141) [70–72], but, among Amerinds in Manitoba, this subclone has a tendency to cause meningitis.

Serotype b Asymptomatic Carrier Isolates

The demonstration that a number of different multilocus genotypes are associated with asymptomatic serotype b carriage is consistent with data on OMP patterns presented earlier [18]. In general, most ETs represented by carrier isolates were also commonly associated with invasive episodes, but isolates of ET 21.6/OMP 13L were rarely recovered from normally sterile body fluids. Of the several hypotheses that could explain these results, we favor the interpretation, advanced previously on the basis of OMP profiling data [18], that ET 21.6/OMP 13L isolates represent a subclone that is "less invasive." Clearly, additional epidemiologic and animal model studies will be required to test this hypothesis.

Clone and Subclone Distribution in Iceland

The finding that the frequency of common disease-causing type b clones in Iceland was highly unusual compared with other European countries is consistent with the results of an earlier analysis by van Alphen et al. [23] of OMP profiles, biotypes, and LPS serotypes in a smaller collection of strains. These investigators did not develop a cogent hypothesis to explain the difference in strain subtype distributions. One interpretation is that isolates of ET 1.9/OMP 2L represent an "old" clone in Iceland, perhaps introduced by Celts in the course of early settlement of the island and that ET 12.5/OMP 3L has been more recently introduced as a result of population migration and contact with other western European peoples. This hypothesis is based on two lines of evidence.

First, subclone ET 1.9/OMP 2L, which accounted for 77% of all Icelandic isolates, also was represented by 60% of isolates from patients in the Canadian Maritime Provinces, where the human population is predominantly of Celtic origin and, like that in Iceland, has experienced little immigration in recent centuries. (Strains from the Maritime Provinces were sought specifically to test this Celtic hypothesis.) Although it is not generally appreciated, genetic evidence based on the frequencies of A-B-O blood group antigens clearly demonstrates that Icelanders are closely allied to Scots and Irish and differ markedly from modern Norwegians and other Scandinavians [73]. The available data indicate that the early Icelandic population (A.D. 950) was a Celtic-Norse mixture, the exact proportions of which are unknown [74]. Germane to our hypothesis is the observation that the same globally uncommon subclone (ET 1.9/OMP 2L) occurs in high frequency among two human populations that share Celtic ancestry but that diverged relatively long ago and have experienced little immigration even to the present time.

Second, strains of ET 1.9/2L were recovered from patients residing in all parts of Iceland (Reykjavik and nine other communities), whereas ET 12.5/3L was confined to the Reykjavik area (two isolates) and Breiddalsvik (one isolate). Because of travel patterns in Iceland, and by analogy with the spread of measles virus on the island [75], it is most likely that the introduction of new clones of type b *H. influenzae* would occur first in Reykjavik and that subsequently they would be disseminated to smaller villages around the perimeter of the island.

We reject the alternative interpretation, that ET 12.5/3L is an "old" Icelandic clone and that ET 1.9/2L is "new," because of the rarity of ET 1.9 isolates in general, and ET 1.9/2L isolates in particular, in other European nations sampled, especially those whose populations would be the likely donors of strains with this ET/OMP combination. Intensive longitudinal sampling of invasive isolates from Icelanders, especially individuals living outside of the Reykjavik area, would help to resolve this issue.

Invasive Disease in Japan

Early in the study, the repeated recovery of isolates of ET 25.6 and closely allied clones from invasive episodes in Papua New Guinea and the virtual absence of these clones in collections from Europe and North America suggested that ET 25.6 occurs in appreciable frequency only in eastern Asia or the Pacific Basin. We attempted to test this hypothesis in two ways.

First, many requests for type b or other invasive isolates were sent to medical scientists in various countries in this region. An especially intensive effort was made to contact medical microbiologists in India, the People's Republic of China, the Republic of Singapore, and Japan. Unfortunately, however,

we were unsuccessful in obtaining isolates from India, the People's Republic of China, the Republic of Singapore, and a number of other countries. However, three investigators in Japan provided 12 type b isolates, including four recovered from normally sterile body fluids. Information obtained through correspondence with Drs. T. Konda, A. Wake, Y. Terawaki, Y. Sakata, and S. Uehara suggested that the frequency of invasive *H. influenzae* episodes in most areas in Japan is low and that the rarity of disease does not reflect merely an inability of Japanese medical microbiologists to culture *H. influenzae* [76, 77].

However, in Hong Kong, a city whose population is 98% ethnic Chinese, invasive serotype b disease apparently occurs with a frequency similar to that in Western countries (G. C. French, professor of Microbiology, Chinese University of Hong Kong, personal communication), and this observation is consistent with a report from Malaysia describing somewhat similar frequencies of invasive type b disease in ethnic Malays, Asian Indians, and Chinese living in Kuala Lumpur [78].

Because of the failure to obtain adequate samples of invasive isolates from Japan and other areas listed above, testing of the hypothesis was attempted in another way. The senior author contacted Dr. Arnold L. Smith, Children's Hospital of Seattle, Washington, a city with a very large component of people of oriental ancestry, and requested 40 invasive serotype b strains recovered from individuals with Japanese, Chinese, or other ethnically Asian surnames. Although Dr. Smith has for about 10 years received cultures of almost all isolates of *H. influenzae* recovered from sterile body fluids of patients in King County, Washington (population approximately 1.3 million) and maintains a collection of hundreds of invasive strains from the area, he was able to identify only 11 isolates that had originated from patients with Asian ancestry [79].

Although we fully recognize the dangers of formulating hypotheses on the basis of anecdotal information, we tentatively suggest that the Japanese are less prone to develop invasive *H. influenzae* disease than are Caucasians or certain other ethnic groups. Variation within and among human populations in immunologic response to serotype b capsule polysaccharide and susceptibility to other encapsulated bacterial pathogens has been reported [80, 81]. It also has been noted that individuals with the nonsecretor phenotype may be at increased risk of developing invasive disease caused by encapsulated bacterial pathogens, including *H. influenzae* [82, 83]. This hypothesis is advanced primarily to stimulate further study of the epidemiologic and immunologic aspects of invasive type b disease in regions of the world and ethnic populations for which currently there is little knowledge.

Evolution of Serotype b Clones

For reasons noted in Results, we suggest that ET 1.9, ET 12.8, and ET 25.6 are genetically similar to the three primordial clones of serotype b *H. influenzae* from which other extant clones producing this capsule type originated. According to this hypothesis, these three clones, through accumulation of point mutations, deletions and insertions, and, perhaps, occasional horizontal gene transfer and recombination of chromosomal genes, have generated most of the serotype b ETs represented in the collection. As discussed elsewhere [22], the occurrence of serotype b clones in primary division II represents a rare case of the transfer of b-specific segments of the *cap* region into evolutionarily old phylogenetic lines.

The combined evidence from estimates of genetic distance (shown in figure 1) and the pattern of sharing of restriction sites in the *cap* locus indicates that the divergence leading to type b clones in group B1b through cluster B5 from those type b clones in clusters A1 through A5 was an ancient event, occurring perhaps in conjunction with the divergence of Eurasians and Africans [84, 85]. Three lines of evidence suggest that the ET 12.8 clone is evolutionarily old. First, ET 12.8 is geographically the most widespread of any clone represented in the collection, being the only one recovered in most of the countries sampled in all six geographic source groups; and it has been identified in eight additional countries (J. M. Musser, unpublished data). Second, all ET 12.8 isolates except one have the b(S) *cap* region RFLP pattern, which is also characteristic of all tested strains of ET 25.6 and closely allied clones and which, based on the pattern of sharing of restriction enzyme recognition sites, probably represents the primordial *cap* b locus RFLP [22]. Third, there is an abundance of different OMP patterns among isolates of ET 12.8 (table 4), which indicates that this genotype has existed for a sufficient period to generate much sequence diversity in genes coding for OMPs. Taken together, the data are interpreted to mean that ET 12.8 has had a relatively long association with humans and has been dispersed world-

wide in the course of migrations of aboriginal populations and also during recorded history.

Strains of ET 25.6 and allied type b clones assigned to clusters B1 through C1 also apparently are evolutionarily old, as judged by the estimates of genetic distance shown in figure 1 and the presence of the b(S) *cap* region RFLP pattern. However, clones of this major type b phylogenetic line are less geographically widespread, are numerically less prominent, and are characterized by possession solely of the U electrophoretic mobility variant of the P1 heat-modifiable OMP.

Two explanations could account for the relative lack of diversity in enzyme-encoded genes and OMPs in conjunction with possession of the "old" *cap* b region RFLP. First, restricted genetic diversity could be caused by small, effective population size [34], ecologic niche specialization, or a low mutation rate. There is no reason to suspect that clones in this line are any less mutable than other serotype b *H. influenzae*. Evidence for some degree of niche specialization is provided by the recent discovery that strains of the ET 25.6 clone family were responsible for more than 95% of episodes of type b pneumonia with sepsis in Pakistan [70, 71], an observation suggesting possible viscerotropism for lung parenchyma. (Statistical association of ET 25.6 clones with pneumonia and sepsis was not recorded in other geographic areas [data not shown]; in Pakistan we have yet to examine a large sample of isolates from children with meningitis.) Small population size is suggested by the relative paucity of isolates of ET 25.6 in the collection and their restricted geographic distribution compared with those of ET 12.8 and many other clones.

A second possibility is that the length of time of association of the *cap* b locus with this phylogenetic line of *H. influenzae* actually has been much shorter than that implied by the depth of genetic distance shown in figure 1. According to this hypothesis, one or more progenitors (most probably unencapsulated) of phylogenetic lines B1 through C1 relatively recently received some part or all of the *cap* b region by horizontal transfer and recombination, and the resulting type b strains have not yet had an opportunity to spread globally. Clearly, with the data currently available, elimination of either of these alternative hypotheses is not possible. However, the interpretation that small population size accounts, in large part, for the restricted genetic diversity observed among isolates of ET 25.6 and associated lines is favored.

It was hypothesized above that ET 1.9 represents a third primordial clone of serotype b and is an "old" European clone that in relatively recent years has been supplanted by the ET 12.5/OMP 3L subclone. Two observations suggest that ET 12.5/OMP 3L has recently spread throughout Europe, most probably in an episode of periodic selection of a mutation affecting fitness [86–88]. First, almost all ET 12.5 isolates have the same biotype (BT I), and most are of a single LPS serotype (serotype 1) [23], which suggests that this subclone has not yet had time to accumulate significant recognizable diversity. Second, as noted, all isolates of ET 12.5 tested were OMP 3L and had the b(S) *cap* region RFLP pattern (table 4).

The probable displacement of ET 1.9 (most likely ET 1.9/OMP 2L) by subclone ET 12.5/OMP 3L represents an example of clonal replacement in populations of human pathogenic bacteria, first described by Caugant et al. [89] for the ET 5 complex of *Neisseria meningitidis*, also in Europe. Because of the relatively large genetic distance between ET 1.9 and ET 12.5 (these clones differ at four of 17 metabolic enzyme-encoding loci) and the differences in OMPs between 2L and 3L, the replacement episode is envisioned to be analogous to the phenomenon of antigenic shift identified in influenza virus epidemiology [90, 91]; by similar analogy, episodes involving temporal variation in OMP subclones represent antigenic drift. Many of the hypotheses presented above are readily testable by additional molecular analyses or by further geographic and longitudinal sampling of type b populations.

Hypothesis for the Global Population Stucture and Molecular Epidemiology of Serotype b Organisms

Can the data generated be used to develop a global hypothesis for serotype b *H. influenzae* epidemiology and evolution? Two questions are of special interest. Why is the type b population in the United States different in clonal composition and more diverse genetically than populations in northwestern Europe, Papua New Guinea, and certain other areas? Is there a causal relationship between the degree of ethnic or racial mixing of human populations and the extent of clonal diversity in bacterial populations? This question has received little consideration by medical microbiologists, a circumstance that is

but one aspect of a more general neglect of evolutionary theory by microbiologists [92].

For many species of pathogenic bacteria, including encapsulated *H. influenzae*, the population structure is basically clonal and the number of clones occurring in appreciable frequencies is small, being on the order of only hundreds, even in *E. coli* and other abundant species that are highly variable at the level of the individual gene locus. Aparently it is in major part the rarity of chromosomal recombination in conjunction with the periodic purging of variation through selection of fitness mutants that severely limits the number of multilocus gene combinations (clones) of serotype b *H. influenzae*, *E. coli*, and other bacteria, which would otherwise be astronomical [93]. With asexuality or very low rates of recombination, the appropriate primary unit of genetic analysis is the multilocus genotype or clone, rather than the individual gene; the units of differentiation of older genetic structures may be preserved even when there is extensive migration and mixing of human and bacterial populations. Moreover, the evolutionary information content of clones is much greater than that of alleles, since convergence to multilocus genotypes is extremely unlikely, whereas convergence to the same electromorph (allele) is a real possibility.

The four phylogenetic lineages of clones of serotype b isolates—A1; A2–A5; B1, B3–C1; and J1–J3 (figure 1)—are highly distinctive in multilocus genotype, which means that their phylogenetic relationships are relatively deep. Hence, the pattern of global geographic variation in clonal composition of populations discovered for serotype b *H. influenzae* does not represent merely a transitional state in the relatively recent spread of new minor mutant strains. Because studies of the *Enterobacteriaceae* indicate that nucleotide substitutions in the coding regions of bacterial genomes accumulate at a more or less constant rate of 1% per million years [94], this degree of genetic differentiation indicates that the several phylogenetic lineages expressing the type b capsule have had relatively long histories of differentiation; they are old.

It is suggested that the occurrence of greater clonal diversity in populations of serotype b isolates in the United States than those in northwestern Europe and Papua New Guinea reflects the much greater racial and ethnic admixture of the North American human populations. According to this hypothesis, a large component of the geographic variation in clonal composition of serotype b *H. influenzae* seen today reflects an older pattern of differentiation that evolved in relative geographic isolation before the Age of Exploration (beginning about 450 years ago) and has not yet been completely obscured by recent demographic changes. Since then some clones that were formerly confined to certain continents and segments of the human species, especially those of European and African origin, became worldwide or nearly so in distribution. In this category are placed clones of ET 1.9 and certain allied cell lines. Similarly, ET 12.8 probably had its origin in Africa or Asia, and ET 25.6 probably arose in eastern Asia or Oceania.

Two lines of evidence sugggest that ET 1.9 is an old European/Caucasian clone and ET 12.5 is a new clone that has rather recently spread through Europe. First, as discussed above, the frequencies of clones causing invasive disease in Iceland are essentially the inverse of those in mainland Europe: 82% are ET 1.9, whereas ET 12.5 accounts for <10% of serotype b disease on this island. In addition, on the basis of the geographic distribution of multilocus genotypes in Iceland, there is reason to believe that ET 12.5 has only recently been introduced. On the assumption that ET 1.9 is the old Icelandic strain, we reasoned that it should also occur in appreciable frequency in human populations that share a relatively high degree of ethnicity with primordial or ancient Icelandic populations and that have had relatively little ethnic admixture because of geographic isolation. Such a population exists in the Maritime Provinces of Canada, and strains obtained from a collection of serotype b isolates held by one of us (J. H.) revealed that ET 1.9 and ET 12.5 accounted for 66% and 22%, respectively, of isolates from Prince Edward Island and New Brunswick. Moreover, 32 of 33 ET 1.9 isolates from Iceland were typed as OMP pattern 2L (the exception was an isolate of ET 1.9/OMP 1H recovered from a 5-year-old American child at the U.S. naval base at Keflavik, Iceland, and all 19 ET 1.9 isolates tested from the Maritime Provinces were also subtype 2L (table 4). In contrast, the ET 1.9/OMP 2L subclone accounted for only 3% and 17% of isolates tested from mainland Europe, and non-Maritime provinces of Canada, respectively, and has never been recovered elsewhere in appreciable frequency.

As an alternative explanation for the observed pattern of geographic variation in clonal composition of serotype b *H. influenzae* populations, it may be

suggested that there are, in fact, no old human racial or ethnic correlates of the contemporary global pattern of distribution and abundance of the various clones or clone groups. The major phylogenetic groups of serotype b clones did not evolve in geographic isolation in conjunction with the racial differentiation of humans; rather, all the major clonal groups have since earliest times been present in all major geographic regions and human populations. This hypothesis postulates that the contemporary pattern of variation in the relative abundance of certain clones in northwestern Europe, North America, Papua New Guinea, and elsewhere can be attributed to differential susceptibility to various clones on the part of human populations. Thus, for example, the native Papua New Guineans may have had a long history of exposure to all the major clonal types, and the present predominance of ET 25.6 and closely related clones in Papua New Guinea may reflect a greater susceptibility to clones of that group, as opposed to ET 12.5 or other clones. Susceptibility could be genetically and/or environmentally determined and, as shown by research in genetic epidemiology [95], it would be difficult to distinguish the contributions of these two factors.

A weakness of this hypothesis is that it seems unlikely that all the clones of this organism could have achieved and maintained global distributions while the human species was undergoing geographic differentiation in strong geographic isolation in the past 50,000–100,000 years [84, 85]. Because there are no known nonhuman natural hosts of serotype b *H. influenzae* [38], dispersal apparently usually occurs via the close intermingling of children. Strains of type b occur mainly in children from birth to 10 years of age, in whom there is only generally (at most) a 5% colonization frequency of the throats of healthy individuals not associated with individuals with type b disease [96–99]. (The carriage rate for type b organisms may be much higher in contacts of children with invasive disease [45, 100–102].) Serotype b strains are even rarer in infants (1% colonization) and in healthy adults (<1% colonization).

Rigorous prospective testing of the two hypotheses is not possible for ethical reasons, but the demonstration that there is a greater chance of finding strongly differentiated clonal types of serotype b *H. influenzae* in the more isolated and genetically differentiated human populations would be more compatible with hypothesis 1 (the postulation of an older, human racial-associated component of bacterial population structure) than with hypothesis 2 (differential susceptibility of human racial populations). Additionally, hypothesis 1 predicts a stronger correlation between the clonal composition of *H. influenzae* populations and the geographic pattern of human racial differentiation and history of demographic movements than does hypothesis 2.

The most critical test of the hypotheses comes from an examination of isolates from native or other populations that have experienced very little immigration: hypothesis 1 explicitly predicts that these populations (aboriginals or others) should have a distinctive set of serotype b clones in relatively high frequency, whereas hypothesis 2 does not.

In this respect, six observations, all supporting hypothesis 1, are noteworthy. (1) As discussed elsewhere in this paper, serotype b isolates causing most invasive disease in Iceland, which has had no significant immigration in the last 800 years, are, as a population, genetically quite distinct from strains recovered from invasive episodes in geographically and culturally allied European nations. (2) The clone causing 43% of the serious type b disease in Papua New Guineans, a geographically, culturally, and genetically divergent and relatively isolated human population, was otherwise rarely encountered in the collection. (3) Sixty percent of invasive isolates from blacks in the Gambia, a group representing an aboriginal African population, were a single clone (ET 12.8); no other single ET accounted for more than 10% of cases, and isolates of most other ETs were closely allied to those of ET 12.8. Moreover, all six isolates of ET 12.8 tested for OMP pattern were subclone ET 12.8/OMP 2L, whereas only eight of 97 ET 12.8 isolates recovered in other countries represented this subclone ($P < .05$). (4) Nineteen percent of isolates recovered from invasive episodes in Finland were ET 21.8; in other regions of the world sampled, this clone caused, on average, <1% of disease ($P < .01$). (5) Analysis of 61 serotype b isolates from children in Pakistan with pneumonia and sepsis revealed that a single subclone, ET 25.6/OMP 6U, was responsible for 95% of cases [70, 71]. (6) Among Eskimos in Alaska, the analysis revealed cases in which ETs were uniquely confined to individuals living in single geographic regions ([42] and J. I. W., manuscript in preparation). Statistically significant differences in the frequency of occurrence of certain clones in native Alaskans compared to non-native Alaskans [42] have been identified.

In sum, there is extensive geographic variation—

on an intercontinental scale—in the clonal composition of populations of serotype b *H. influenzae*. The bulk of the data favors the hypothesis that a large component of this variation reflects an older pattern of differentiation that evolved in relative geographic isolation before the Age of Exploration approximately 450 years ago and that has not yet been completely obscured by recent human migration and population admixture. However, it is not possible to completely reject the alternative hypothesis that contemporary patterns of variation in the relative abundance of certain clones reflect differential susceptibility on the part of different human populations. Study of isolates from regions of the world not yet sampled, and especially strains recovered from additional human populations that have experienced little recent admixture, will provide critical data bearing on these hypotheses, as will more extensive analysis, at the nucleotide sequence level, of judiciously selected isolates from the sample described here.

Geographic Variation in Genetic Structure of Pathogenic Bacteria

Our study of encapsulated strains of *H. influenzae* is among the first to examine geographic variation in genetic structure of a bacterial pathogen on a very large scale in a systematic manner. The analysis revealed at least three distinct levels of geographic variation: intercontinental, intracontinental, and intracountry. For type b strains, the most striking aspect of geographic structure is the variation in frequency of recovery of clones from invasive episodes in patients on different continents; but intracontinental differences in clone recovery may also be extensive. For example, in North America, isolates of ET 100 accounted for 4.5% of strains from Canada, but this clone was not represented among isolates from invasive or carrier sources in the United States.

We cite two examples of intracountry variation. First, also in Canada, most isolates from the Maritime Provinces were ET 1.9, but isolates of this clone have been recovered in other provinces. Second, in the United States isolates of the ET 1.9/2L subclone are more frequently recovered from patients in Minnesota than in Texas [103]. It very well may be that geographic variation in the structure of type b populations extends to the level of the individual community (microgeographic variation [104]), but we made no attempt to test this hypothesis.

Although many ideas may be advanced to explain the occurrence of geographic variation in type b populations, we favor the notion that much of it is due to a relatively slow dispersal velocity for this pathogen. In this regard it is probably significant that geographic variation recorded in populations of *E. coli* [105], an organism that may survive in the environment and resides in the bowel of virtually every warm-blooded species, is far less dramatic than for type b organisms.

Mayr [104] noted in 1963 that the occurrence of genetic differences among spatially segregated populations of a species is most probably a universal phenomenon in the animal kingdom. Varying levels of geographic structure have been noted for small samples of other bacterial pathogens [48, 69, 106], and failure to detect it in all species may well be the consequence of investigation of insufficient sample sizes. Like many other aspects of prokaryotic evolution [92], the study of geographic variation and the factors accounting for it (molecular and other) is in its infancy.

Implications for the Control of Serotype b Disease

The restriction of isolates of most clones to one or a few parts of the world suggests that transmission of serotype b strains and those producing other capsular polysaccharides between geographically separate populations is occurring relatively slowly. Because humans are the only known reservoir of *H. influenzae* [38] and children are the primary carriers of serotype b strains, interpopulation spread must occur largely as a result of close contact between children. The implication of the discovery of the slow spread of clones is that in human populations that are relatively isolated for geographic, cultural, or other reasons, and for which no efficacious vaccine currently exists [107, 108], strategies designed to decrease the carrier rate might decrease the incidence of disease.

The hypothesis of slow spread of clones between populations is not incompatible with episodes of rapid turnover of clones or subclones within local [40, 103, 109] or distant [27] populations.

Utility of the Population Genetic Framework

One of the strong advantages of population genetic analysis of bacterial pathogens not widely appreciated by medical microbiologists and investigators interested in microbial pathogenesis is that it pro-

vides ready access to a wealth of genetic theory that can be exploited to provide insights and generate testable hypotheses regarding many aspects of pathogenic organisms. Four examples from studies with *H. influenzae* will suffice.

First, the extensive genetic differentiation of division I and II organisms, which approaches species-level divergence, suggests that strains assigned to each division should be quite different in the structure of putative virulence factors and natural history. This has been extensively confirmed by studies of the *cap* region [22, 58], human IgA1 protease (*iga*) gene [110], OMP patterns [20–22, 25], and virulence [22, 33] of isolates in each division. Perhaps the best indicator of basic biologic differences between isolates in the two primary divisions is the very infrequent recovery of division II isolates from invasive episodes. Based upon the degree of genetic differentiation, it is very possible that division I and division II serotype a and serotype b isolates differ in global regulation [111] of virulence genes.

Second, population genetic analysis revealed that all serotype d isolates examined are very closely related in overall chromosomal genotype and, thus, anticipated the finding that type d strains collected from worldwide sources over several decades have identical *cap* region RFLP patterns [58]. Three type d isolates recently have been shown to have closely similar genomic digest patterns revealed by orthogonal field-alternation gel electrophoresis [112].

Third, the occurrence of strong associations of multilocus enzyme genotypes, *cap* region RFLP patterns, and OMP subtypes and the recovery of isolates with identical genetic properties in widely separated geographic regions and over a 40-year period indicates that, despite the presence of a well-characterized DNA transformation system, the population structure of encapsulated *H. influenzae* is clonal and presaged a recent report of a lack of genetic competence among multiple independent natural isolates expressing serotype b capsule [59].

Fourth, the extensive knowledge of genetic relationships among the very large number of encapsulated strains described in this report led to the recognition that strains of *H. influenzae* biogroup aegyptius recovered from cases of Brazilian purpuric fever [113], a newly described invasive disease among children in certain rural areas in Brazil, are evolutionarily most closely allied with strains producing serotype c capsule (J. M. M. and R. K. S., "Brazilian Purpuric Fever: Evolutionary Genetic Relationships of the Case Clone of *Haemophilus influenzae* Biogroup aegyptius to Encapsulated Strains of *Haemophilus influenzae*," submitted).

The population genetic framework has revealed a number of observations regarding the evolutionary genetics of encapsulated *H. influenzae* strains that need explaining at the molecular level. First, by what evolutionary mechanism have serotype d strains arisen, most likely from either a serotype b progenitor strain in cluster B (figure 1)? Second, among serotype b organisms, why is the H variant of the heat-modifiable OMP P1 present in strains assigned to divisions I and II? Third, why are strains of ET 21.6 frequently associated with asymptomatic carriage in many different geographic areas and only rarely recovered from invasive disease episodes? Fourth, and perhaps most important from a medical standpoint, why do strains of only nine ETs account for 81% of all invasive episodes worldwide? Fifth, what processes are responsible for the dramatic geographic variation in population structure among serotype b clones? Examination of these and other questions will provide additional insights into the population biology of *H. influenzae*.

Concluding Comment

Olyhoek et al. [114] have recently analyzed electrophoretic variation in seven metabolic enzymes and two OMPs in 423 isolates of serogroup A *N. meningitidis* recovered from 23 epidemics or outbreaks occurring in 38 countries on six continents over a 70-year period. Caugant et al. [89, 115] have studied genetic diversity and relationships among 650 isolates of eight serogroups of *N. meningitidis* recovered from patients in 20 countries. Therefore, detailed information is now available on an intercontinental scale concerning the molecular population genetics and epidemiology of two other important human pathogenic bacterial species. This study of encapsulated *H. influenzae* is the first to identify possible correlations between contemporary patterns of geographic and temporal variation in clonal composition and diversity of the bacterial population and patterns of ethnic or racial differentiation and historical movement of the host human populations. The data suggest that a causal relation exists between degree of ethnic mixing of human populations and degree of diversity in clonal composition of serotype b *H. influenzae* populations.

Knowledge of the global genetic diversity and

clonal structure of type b strains will permit prospective assessment of the effect of vaccination on the genotypic population structure of the pathogen and may be important in the formulation of vaccination strategies.

A major goal of bacterial population genetic research is to exploit phylogenetic frameworks to contribute to an understanding of the molecular basis of pathogenesis through the analysis and interpretation of the distribution of proven and putative virulence genes and other factors that may be causally related to pathogenicity [116]. The framework presented here can be utilized to address a variety of questions that remain unresolved in *H. influenzae* research and to generate additional testable hypotheses regarding virulence, pathogenicity, and global epidemiology.

References

1. Turk DC, May JR. *Haemophilus influenzae*: its clinical importance. London: English University Press, **1967**
2. Sell SH, Wright PF, eds. *Haemophilus influenzae*: epidemiology, immunology, and prevention of disease. New York: Elsevier/North-Holland, **1982**
3. Turk DC. The pathogenicity of *Haemophilus influenzae*. J Med Microbiol **1984**;18:1-16
4. Pittman M. Variation and type specificity in the bacterial species *Hemophilus influenzae*. J Exp Med **1931**;53: 471-92
5. Kilian M. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. J Gen Microbiol **1976**;93:9-62
6. Anderson P, Flesher A, Shaw S, Harding AL, Smith DH. Phenotypic and genetic variation in the susceptibility of *Haemophilus influenzae* type b to antibodies to somatic antigens. J Clin Invest **1980**;65:885-91
7. Loeb MR, Smith DH. Outer membrane protein composition in disease isolates of *Haemophilus influenzae*: pathogenic and epidemiological implications. Infect Immun **1980**;30:709-17
8. Barenkamp SJ, Munson RS Jr., Granoff DM. Subtyping isolates of *Haemophilus influenzae* type b by outer-membrane protein profiles. J Infect Dis **1981**;143:668-76
9. van Alphen L, Riemens T, Poolman J, Hopman C, Zanen HC. Homogeneity of cell envelope protein subtypes, lipopolysaccharide serotypes, and biotypes among *Haemophilus influenzae* type b from patients with meningitis in the Netherlands. J Infect Dis **1983**;148:75-81
10. van Alphen L, Riemens T, Poolman J, Zanen HC. Characteristics of major outer membrane proteins of *Haemophilus influenzae*. J Bacteriol **1983**;155:878-85
11. Inzana TJ. Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of *Haemophilus influenzae*. J Infect Dis **1983**;148:492-9
12. Inzana TJ, Pichichero ME. Lipopolysaccharide subtypes of *Haemophilus influenzae* type b from an outbreak of invasive disease. J Clin Microbiol **1984**;20:145-50
13. Gulig PA, Frisch CF, Hansen EJ. A set of two monoclonal antibodies specific for the cell surface-exposed 39K major outer membrane protein of *Haemophilus influenzae* type b defines all strains of this pathogen. Infect Immun **1983**;42:516-24
14. Gulig PA, Hansen EJ. Coprecipitation of lipopolysacchride and the 39,000-molecular-weight major outer membrane protein of *Haemophilus influenzae* type b by lipopolysaccharide-directed monoclonal antibody. Infect Immun **1985**;49:819-27
15. Kilian M, Sørensen I, Frederiksen W. Biochemical characteristics of 130 recent isolates from *Haemophilus influenzae* meningitis. J Clin Microbiol **1979**;9:409-12
16. Tiller F-W. Biochemical differentiation of *Haemophilus influenzae*. Additional characterization of biotypes by carbohydrate fermentation patterns. Zentralbl Bakteriol Mikrobiol Hyg [A] **1982**;253:236-46
17. Tebbutt GM. A chemotyping scheme for clinical isolates of *Haemophilus influenzae*. J Med Microbiol **1984**;17:335-45
18. Hampton CM, Barenkamp SJ, Granoff DM. Comparison of outer membrane protein subtypes of *Haemophilus influenzae* type b isolates from healthy children in the general population and from diseased patients. J Clin Microbiol **1983**;18:596-600
19. Granoff DM, Barenkamp SJ, Munson RS Jr. Outer membrane protein subtypes for epidemiologic investigation of *Haemophilus influenzae* type b disease. In: Sell SH, Wright PF, eds. *Haemophilus influenzae*: epidemiology, immunology, and prevention of disease. New York: Elsevier/North-Holland, **1982**:43-55
20. Musser JM, Granoff DM, Pattison PE, Selander RK. A population genetic framework for the study of invasive diseases caused by serotype b strains of *Haemophilus influenzae*. Proc Natl Acad Sci USA **1985**;82:5078-82
21. Musser JM, Kroll JS, Moxon ER, Selander RK. Clonal population structure of encapsulated *Haemophilus influenzae*. Infect Immun **1988**;56:1837-45
22. Musser JM, Kroll JS, Moxon ER, Selander RK. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. Proc Natl Acad Sci USA **1988**;85:7758-62
23. van Alphen L, Geelen L, Jónsdóttir K, Takala AK, Kayhty H, Zanen HC. Distinct geographic distribution of subtypes of *Haemophilus influenzae* type b in western Europe. J Infect Dis **1987**;156:216-8
24. Allan I, Loeb MR, Moxon ER. Limited genetic diversity of *Haemophilus influenzae* (type b). Microb Pathog **1987**;2:139-45
25. Allan I, Kroll JS, Dhir A, Moxon ER. *Haemophilus influenzae* serotype a: outer membrane protein classification and correlation with DNA polymorphism at the *cap* locus. Infect Immun **1988**;56:529-31
26. Hoiseth SK, Connelly CJ, Moxon ER. Genetics of spontaneous, high-frequency loss of b capsule expression in *Haemophilus influenzae*. Infect Immun **1985**;49:389-95
27. Barenkamp SJ, Granoff DM, Pittman M. Outer membrane protein subtypes and biotypes of *Haemophilus influenzae* type b: relation between strains isolated in 1934-1954 and 1977-1980. J Infect Dis **1983**;148:1127

28. Campos J, Garcia-Tornel S, Musser JM, Selander RK, Smith AL. Molecular epidemiology of multiply resistant *Haemophilus influenzae* type b in day care centers. J Infect Dis **1987**;156:483–9
29. Musser JM, Barenkamp SJ, Granoff DM, Selander RK. Genetic relationships of serologically nontypable and serotype b strains of *Haemophilus influenzae*. Infect Immun **1986**;52:183–91
30. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol **1986**;51:873–4
31. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **1970**; 227:680–5
32. Lugtenberg B, Meijers J, Peters R, van der Hoek P, van Alphen L. Electrophoretic resolution of the 'major outer membrane protein' of *Escherichia coli* K12 into four bands. FEBS Lett **1975**;58:254–8
33. Kroll JS, Moxon ER. Capsulation and gene copy number at the *cap* locus of *Haemophilus influenzae* type b. J Bacteriol **1988**;170:859–64
34. Nei M. Molecular evolutionary genetics. New York: Columbia University Press, **1987**
35. Sneath PHA, Sokal RR. Numerical taxonomy: the principles and practice of numerical classification. San Francisco: Freeman, **1973**
36. Catlin BW, Bendler JW III, Goodgal SH. The type b capsulation locus of *Haemophilus influenzae*: map location and size. J Gen Microbiol **1972**;70:411–22
37. Kroll JS, Hopkins I, Moxon ER. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. Cell **1988**;53:347–56
38. Kilian M, Biberstein EL. Genus II. *Haemophilus* Winslow, Broadhurst, Buchanan, Krumwiede, Rogers, and Smith 1917. In: Kreig NR, Holt JG, eds. Bergey's manual of systematic bacteriology. Vol. 1. Baltimore: Williams and Wilkins, **1984**:558–69
39. Takala AK, Musser JM, Selander RK. High prevalence of *Haemophilus influenzae* serotype b of a rare electrophoretic type 21 among invasive isolates in Finland [abstract no. 786]. In: Programs and abstracts of the 27th Interscience Conference on Antimicrobial Agents and Chemotherapy. Washington DC: American Society for Microbiology, **1987**
40. Takala AK, van Alphen L, Musser JM, Geelen L, Selander RK, Eskola J, Makela PH. Bacteriologic epidemiology of *Hemophilus influenzae* type b strains causing invasive infections in Finland. J Infect Dis **1989**;160:237–42
41. Ward JI, Lum MKW, Hall DB, Silimperi DR, Bender TR. Invasive *Haemophilus influenzae* type b disease in Alaska: background epidemiology for a vaccine efficacy trial. J Infect Dis **1986**;153:17–26
42. Ward J, Burkart K, Christenson P, Fitzgerald M, Musser J, Selander R. Outer membrane protein (OMP) and isoenzyme analysis of *Haemophilus influenzae* type B (Hib) strains in Alaska [abstract no. 785]. In: Programs and abstracts of the 27th Interscience Conference on Antimicrobial Agents and Chemotherapy. Washington, DC: American Society for Microbiology, **1987**
43. Losonsky GA, Santosham M, Sehgal VM, Zwahlen A, Moxon ER. *Haemophilus influenzae* disease in the White Mountain Apaches: molecular epidemiology of a high risk population. Pediatr Infect Dis **1984**;3:539–47
44. Slatkin M. Rare alleles as indicators of gene flow. Evolution **1985**;39:53–65
45. Ochman H, Whittam TS, Caugant DA, Selander RK. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J Gen Microbiol **1983**; 129:2715–26
46. Selander RK, McKinney RM, Whittam TS, Bibb WF, Brenner DJ, Nolte FS, Pattison PE. Genetic structure of populations of *Legionella pneumophila*. J Bacteriol **1985**; 163:1021–37
47. Brenner DJ, Steigerwalt AG, Epple P, Bibb WF, McKinney RM, Starnes RWM, Colville JM, Selander RK, Edelstein PH, Moss CW. *Legionella pneumophila* serogroup Lansing 3 isolated from a patient with fatal pneumonia, and descriptions of *L. pneumophila* subsp. *pneumophila* subsp. nov., *L. pneumophila* subsp. *fraseri* subsp. nov., and *L. pneumophila* subsp. *pascullei* subsp. nov. J Clin Microbiol **1988**;26:1695–1703
48. Musser JM, Hewlett EL, Peppler MS, Selander RK. Genetic diversity and relationships in populations of *Bordetella* spp. J Bacteriol **1986**;166:230–7
49. Yamada Y, Akita M. An electrophoretic comparison of enzymes in strains of *Gluconobacter* species. J Gen Appl Microbiol **1984**;30:115–26
50. Gilmour MN, Whittam TS, Kilian M, Selander RK. Genetic relationships among the oral streptococci. J Bacteriol **1987**;169:5247–57
51. Chun PK, Sensabaugh GF, Vedros NA. Genetic relationships among *Neisseria* species assessed by comparative enzyme electrophoresis. J Gen Microbiol **1985**;131:3105–15
52. Selander RK, Musser JM, Caugant DA, Gilmour MN, Whittam TS. Population genetics of pathogenic bacteria. Microb Pathog **1987**;3:1–7
53. Tibayrenc M, Ayala FJ. Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical, and epidemiological significance. Evolution **1988**; 42:277–92
54. Beverley SM, Ismach RB, Pratt DM. Evolution of the genus *Leishmania* as revealed by comparisons of nuclear DNA restriction fragment patterns. Proc Natl Acad Sci USA **1987**;84:484–8
55. Caccone A, Powell JR. Molecular evolutionary divergence among North American cave crickets. II. DNA-DNA hybridization. Evolution **1987**;41:1215–38
56. Eskola J, Peltola H, Takala AK, Käyhty H, Hakulinen M, Karanko V, Kela E, Rekola P, Rönnberg P-R, Samuelson JS, Gordon LK, Mäkelä PH. Efficacy of *Haemophilus influenzae* type b polysaccharide-diphtheria toxoid conjugate vaccine in infancy. N Engl J Med **1987**;317:717–22
57. Zwahlen A, Rubin LG, Moxon ER. Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. Microb Pathog **1986**;1:465–73

58. Kroll JS, Zamze S, Loynds B, Moxon ER. Common organization of chromosomal loci for production of different capsular polysaccharides in *Haemophilus influenzae*. J Bacteriol **1989**;171:3343–7
59. Cope L, Edgley H, Hansen E. Lack of genetic competence among strains of *Haemophilus influenzae* type b [abstract no. D-36]. In: Abstracts of the 89th Annual Meeting of the American Society for Microbiology. Washington, DC: American Society for Microbiology, **1989**:88
60. Loeb MR, Woodin KA. Cross-reactivity of surface-exposed epitopes of outer membrane antigens of *Haemophilus influenzae* type b. Infect Immun **1987**;55:2977–83
61. Takala AK, van Alphen L, Eskola J, Palmgren J, Bol P, Mäkelä PH. *Haemophilus influenzae* type b strains of outer membrane subtypes 1 and 1c cause different types of invasive disease. Lancet **1987**;2:647–50
62. Selander RK, Caugant DA, Whittam TS. Genetic structure and variation in natural populations of *Escherichia coli*. In: Neihardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE, eds. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. Vol. 2. Washington, DC: American Society for Microbiology, **1987**:1625–48
63. Mendelman PM, Syriopoulou VP, Gandy SL, Ward JI, Smith AL. Molecular epidemiology of plasmid-mediated ampicillin resistance in *Haemophilus influenzae* type b isolates from Alaska. J Infect Dis **1985**;151:1061–72
64. Willard JE, Johnson EJ, Daum RS. Interrelationships between physical state, phenotypic stability and transferability of β-lactamase genes in *Haemophilus influenzae*. J Gen Microbiol **1982**;128:2353–60
65. Campos J, Garcia-Tornel S, Musser JM, Selander RK, Smith AL. Molecular epidemiology of multiply resistant *Haemophilus influenzae* type b in day care centers. J Infect Dis **1987**;156:483–9
66. Simasathien S, Duangmani C, Echeverria P. *Haemophilus influenzae* type b resistant to ampicillin and chloramphenicol in an orphanage in Thailand. Lancet **1980**;2:1214–7
67. O'Brien TF, Pla MP, Mayer KH, Kishi H, Gilleece E, Syvanen M, Hopkins JD. Intercontinental spread of a new antibiotic resistance gene on an epidemic plasmid. Science **1985**;230:87–8
68. Musser JM, Bemis DA, Ishikawa H, Selander RK. Clonal diversity and host distribution in *Bordetella bronchiseptica*. J Bacteriol **1987**;169:2793–803
69. Beltran P, Musser JM, Helmuth R, Farmer JJ III, Frerichs WM, Waschsmuth IK, Ferris K, McWhorter AC, Wells JG, Cravioto A, Selander RK. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S. choleraesuis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport*, and *S. typhimurium*. Proc Natl Acad Sci USA **1988**;85:7753–7
70. Ghafoor A, Weinberg GA, Ishaq Z, Anwar F, Khalid Nomani N, Kabeer M, Musser JM, Granoff DM. An epidemic of bacteremic nontypable *Haemophilus influenzae* (NTHI) respiratory infection (ARI) in children in Pakistan [abstract no. 895]. In: Programs and abstracts of the 28th Intersciences Conference on Antimicrobial Agents and Chemotherapy. Washington DC: American Society for Microbiology, **1988**
71. Weinberg GA, Ghafoor A, Ishaq Z, Nomani NK, Kabeer M, Anwar F, Burney MI, Qureshi AW, Musser JM, Selander JM, Granoff DM. Clonal analysis of *Hemophilus influenzae* strains isolated from children from Pakistan with lower respiratory tract infections. J Infect Dis **1989**;160:634–43
72. Munson RS Jr, Kabeer MH, Lenoir AA, Granoff DM. Epidemiology and prospects for prevention of disease due to *Haemophilus influenzae* in developing countries. Rev Infect Dis **1989**;11(Suppl):S588–97
73. Bjarnason O, Bjarnason V, Edwards JH, Fridriksson S, Magnusson M, Mourant AE, Tills D. The blood groups of Icelanders. Ann Hum Genet **1973**;36:425–58
74. Thompson EA. The Icelandic admixture problem. Ann Hum Genet **1973**;37:69–80
75. Cliff A, Haggett P. Island epidemics. Sci Am **1984**;250(5): 138–47
76. Kawakami Y, Okimura Y, Kanai M. Biochemical properties and antimicrobial susceptibility of recent clinical isolates of *Haemophilus* species. Microbiol Immunol **1981**;25:79–84
77. Uehara S, Terashima I, Nakamura A, Kurosaki T, Sugaya N, Okimoto Y, Himi K. The role of *Haemophilus influenzae* in lower respiratory infections in childhood. Acta Paediatra Jpn **1985**;89:94–101
78. Thong ML, Puthucheary SD, Omar A. *Haemophilus influenzae* meningitis. Asean J Clin Sci **1983**;4:47–56
79. Sherry B, Emanuel I, Kronmal RA, Smith AL, Char LF, Gale JL, Walkley E. Interannual variation of the incidence of *Haemophilus influenzae* type b meningitis. JAMA **1989**;261:1924–9
80. Petersen GM, Silimperi DR, Roter JI, Terasaki PI, Schanfield MS, Park MS, Ward JI. Genetic factors in *Haemophilus influenzae* type b disease susceptibility and antibody acquisition. J Pediatr **1987**;110:228–33
81. Siber GR, Santosham M, Priehs CM, Reid R, Letson W, Madore D, Eby R. Impaired antibody (ab) response to *H. influenzae* b (Hib) capsular polysaccharide (CP) in Native American children [abstract no. 326]. In: Programs and abstracts of the 27th Interscience Conference on Antimicrobial Agents and Chemotherapy. Washington, DC: American Society for Microbiology, **1987**
82. Blackwell CC, Jónsdóttir K, Hanson M, Todd WTA, Chaudhuri AKR, Mathew B, Brettle RP, Weir DM. Nonsecretion of ABO antigens predisposing to infection by *Neisseria meningitidis* and *Streptococcus pneumoniae* [letter]. Lancet **1986**;2:284–85
83. Blackwell CC, Jonsdottir K, Hanson MF, Weir DM. Nonsecretion of ABO blood group antigens predisposing to infection by *Haemophilus influenzae* [letter]. Lancet **1986**;2:687
84. Cavalli-Sforza LL, Bodmer WB. The genetics of human populations. San Franciso: W. H. Freeman, **1971**
85. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J. Reconstruction of human evolution: bringing together genetic, arachaeological, and linguistic data. Proc Natl Acad Sci USA **1988**;85:6002–6
86. Kubitschek HE. Operation of selection pressure on microbial populations. In: Carlile MJ, Skehel JJ, eds. Evolution in the microbial world. London: Cambridge University Press, **1974**:105–30

87. Koch AL. The pertinence of the periodic selection phenomenon to prokaryote evolution. Genetics **1974**;77:127–42
88. Levin BR. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics **1981**;99:1–23
89. Caugant DA, Frøholm LO, Bøvre K, Holten E, Frasch CE, Mocca LF, Zollinger WD, Selander RK. Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. Proc Natl Acad Sci USA **1986**;83:4927–31
90. Palese P, Kingsbury DW, eds. Genetics of influenza viruses. New York: Springer-Verlag, **1983**
91. Stuart-Harris C, Potter CW. The molecular virology and epidemiology of influenza. London: Academic Press. **1984**
92. Woese CR. Bacterial evolution. Microbiol Rev **1987**;51:221–71
93. Selander RK, Levin BR. Genetic diversity and structure in *Escherichia coli* populations. Science **1980**;210:545–7
94. Ochman H, Wilson AC. Evolutionary history of enteric bacteria. In: Neihardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE, eds. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. Vol. 2. Washington, DC: American Society for Microbiology, **1987**:1649–54
95. Chakraborty R, Szathmary EJE, eds. Progress in clinical and biological research. Vol. 194. Diseases of complex etiology in small populations. Ethnic differences and research approaches. New York: Alan R. Liss, **1985**
96. Dawson B, Zinnemann K. Incidence and type distribution of capsulated *H. influenzae* strains. Br Med J **1952**;1:740–2
97. Masters PL, Brumfitt W, Mendez RL, Likar M. Bacterial flora of the upper respiratory tract in Paddington families, 1952–4. Br Med J **1958**;1:1200–5
98. Turk DC. Naso-pharyngeal carriage of *Haemophilus influenzae* type b. J Hyg (Lond) **1963**;61:247–56
99. Straker EA. Serological typing of meningeal strains of *H. influenzae*. Lancet **1945**;1:817
100. Schiffer MS, MacLowry J, Schneerson R, Robbins JB, McReynolds JW, Thomas WJ, Bailey DW, Clarke EJ Jr, Mueller EJ, Escamilla J. Clinical, bacteriological, and immunological characterisation of ampicillin-resistant *Hemophilus influenzae* type b. Lancet **1974**;2:257–9
101. Ward JI, Gorman G, Phillips C, Fraser DW. *Haemophilus influenzae* type b disease in a day-care center. Report of an outbreak. J Pediatr **1978**;92:713–7
102. Yogev R, Lander HB, Davis AT. Effect of TMP-SMX on nasopharyngeal carriage of ampicillin-sensitive and ampicillin-resistant *Hemophilus influenzae* type b. J Pediatr **1978**;93:394–7
103. Granoff DM, Sheetz K, Pandey JP, Nahm MH, Rambeck JH, Jacobs JL, Musser J, Selander RK, Kabeer M, Murphy TV, Osterholm MT. Host and bacterial factors associated with *Haemophilus influenzae* type b disease in Minnesota children vaccinated with type b polysaccharide vaccine. J Infect Dis **1989**;159:908–16
104. Mayr E. Animal species and evolution. Cambridge, Mass: Belknap Press of Harvard University Press, **1963**
105. Whittam TS, Ochman H, Selander RK. Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. Mol Biol Evol **1983**;1:67–83
106. Musser JM, Rapp VJ, Selander RK. Clonal diversity in *Haemophilus pleuropneumoniae*. Infect Immun **1987**;55:1207–15
107. Gilsdorf JR. Bacterial meningitis in southwestern Alaska. Am J Epidemiol **1977**;106:388–91
108. Ward JI, Brenneman G, Letson G, Heyward W, the Alaska Vaccine Efficacy Trial Study Group. Limited protective efficacy of an *H. influenzae* type b conjugate vaccine PRP-D in Native Alaskan infants immunized at 2-, 4-, and 6-months of age [abstract no. 1127]. In: Programs and abstracts of the 28th Interscience Conference on Antimicrobial Agents and Chemotherapy. Washington, DC: American Society for Microbiology, **1988**
109. Crowe BA, Wall RA, Kusecek B, Neumann B, Olyhoek T, Abdillahi H, Hassan-King M, Greenwood BM, Poolman JT, Achtman M. Clonal and variable properties of *Neisseria meningitidis* isolated from cases and carriers during and after an epidemic in The Gambia, West Africa. J Infect Dis **1989**;159:686–700
110. Poulsen K, Hjorth JP, Kilian M. Limited diversity of the immunoglobulin A1 protease gene (*iga*) among *Haemophilus influenzae* serotype b strains. Infect Immun **1988**;56:987–92
111. Miller JF, Mekalanos JJ, Falkow S. Coordinate regulation and sensory transduction in the control of bacterial virulence. Science **1989**;243:916–22
112. Lee JJ, Smith HO, Redfield RJ. Organization of the *Haemophilus influenzae* Rd genome. J Bacteriol **1989**;171:3016–24
113. Brazilian Purpuric Fever Study Group. Brazilian purpuric fever: epidemic purpura fulminans associated with antecedent purulent conjuctivitis. Lancet **1987**;2:757–61
114. Olyhoek T, Crowe BA, Achtman M. Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983. Rev Infect Dis **1987**;9:665–92
115. Caugant DA, Mocca LF, Frasch CE, Frøholm LO, Zollinger WD, Selander RK. Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern. J Bacteriol **1987**;169:2781–92
116. Selander RK, Musser JM. The population genetics of bacterial pathogenesis. In: Iglewski BH, Clark VL, eds. Molecular basis of bacterial pathogenesis. Orlando: Academic Press, **1990** (in press)