

Diversity of PspA: Mosaic Genes and Evidence for Past Recombination in *Streptococcus pneumoniae*

SUSAN K. HOLLINGSHEAD,^{1*} ROBERT BECKER,² AND DAVID E. BRILES¹

*Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama,¹
and Aventis Pasteur Laboratories, Swiftwater, Pennsylvania²*

Received 15 November 1999/Returned for modification 31 March 2000/Accepted 26 April 2000

Pneumococcal surface protein A (PspA) is a serologically variable protein of *Streptococcus pneumoniae*. Twenty-four diverse alleles of the *pspA* gene were sequenced to investigate the genetic basis for serologic diversity and to evaluate the potential of diversity to have an impact on PspA's use in human vaccination. The 24 *pspA* gene sequences from unrelated strains revealed two major allelic types, termed "families," subdivided into clades. A highly mosaic gene structure was observed in which individual mosaic sequence blocks in PspAs diverged from each other by over 20% in many cases. This level of divergence exceeds that observed for blocks in the penicillin-binding proteins of *S. pneumoniae* or in many cross-species comparisons of gene loci. Conversely, because the mosaic pattern is so complex, each pair of *pspA* genes also has numerous shared blocks, but the position of conserved blocks differs from gene pair to gene pair. A central region of *pspA*, important for eliciting protective antibodies, was found in six clades, which each diverge from the other clades by >20%. Sequence relationships among the 24 alleles analyzed over three windows were discordant, indicating that intragenic recombination has occurred within this locus. The extensive recombination which generated the mosaic pattern seen in the *pspA* locus suggests that natural selection has operated in the history of this gene locus and underscores the likelihood that PspA may be important in the interaction between the pneumococcus and its human host.

Streptococcus pneumoniae is a formidable human pathogen responsible for a major portion of over 3 million deaths worldwide of children from pneumonia and meningitis (22). Recent increases in the rate of isolation of pneumococci resistant to antibiotics promote the expectation that morbidity and mortality caused by this pathogen will increase in the near future. The impact on health is greatest in very young children, elderly individuals, sickle cell patients, and immunocompromised persons of all ages (4). Currently available vaccines are based on immunity to capsular polysaccharides of the pneumococcus, of which 90 serotypes exist. A 23-valent polysaccharide vaccine, recommended for use in adults, is not effective for a large number of the at-risk individuals, including children less than 2 years of age, who are not yet capable of responding adequately to polysaccharide antigens (14, 29). A 7-valent conjugated polysaccharide vaccine was recently licensed for use in children (53), but nonvaccine serotypes will be likely to cause substantial pneumococcal disease even in vaccinated individuals (26, 27). Pneumococcal proteins either in addition to polysaccharides or as stand-alone vaccines have the potential to better protect those for whom the current vaccine is ineffective (5, 7, 45). PspA, which has been shown to effect antibody-mediated protection in mouse models of pneumococcal disease (38) and to be safe in administration to humans (43), is one such candidate.

PspA is an important virulence factor of the pneumococcus (63) that influences bacterium-host interactions through interference with the fixation of complement C3 (61). PspA also binds human lactoferrin (24). PspA is present on all pneumococcal strains and is serologically variable (15). Mouse models have shown that cross-reactive anti-PspA antisera are also

cross-protective (reviewed in references 8, 9, 60). Human antisera from a phase I vaccine trial were also competent for protecting mice from pneumococcal infection by challenge strains of various PspA types (5a).

Basic information about protein structural domains in PspA comes from the DNA sequence of *pspA/Rx1* (gene/strain name [described below]) and *pspA/EF5668* (28, 36, 62, 63). There are five domains (see Fig. 1), including (i) a signal peptide, (ii) an α -helical and charged domain that bears a strong 7-residue periodicity typical of coiled-coil proteins (amino acids 1 to 288), (iii) a proline-rich region (amino acids 289 to 370), (iv) a choline-binding domain consisting of 10 20-amino-acid repeats (amino acids 371 to 571), and (v) a C-terminal 17-amino-acid tail (amino acids 572 to 589). The large choline-binding repeat domain is required for the attachment of PspA on the pneumococcal cell surface via interaction with choline in membrane-associated lipoteichoic acid (64). This orientation results in the α -helical or charged domain of PspA being exposed on the surface and thus available to interact with the human host (8, 20).

The combinatorial serological diversity of monoclonal antibody (MAb)-detected epitopes on PspA proteins from different strains (15) coupled with the presence of one major chromosomal locus encoding PspA suggested that PspA proteins might be mosaics. Mosaic gene alleles in bacteria are formed by recombination following horizontal gene transfer (34, 35). The term "mosaic" derives from the pattern of interspersed blocks of nucleotide sequence which have different evolutionary histories, but are found combined in the resulting gene allele subsequent to recombination events (41). Recombination follows the horizontal transfer of gene segments mediated by transformation, transduction, conjugation, or other means; the recombined segments can be derived from other strains in the same species or from other more distant bacterial relatives (35, 40).

The most intensively studied mosaic alleles in *S. pneumoniae*

* Corresponding author. Mailing address: Department of Microbiology, BBRB654, University of Alabama at Birmingham, Birmingham, AL 35294. Phone: (205) 934-0570. Fax: (205) 975-5480. E-mail: hollings@uab.edu.

have been those encoding several penicillin-binding proteins (PBPs), including 1a, 2x, and 2b, in which the mosaic blocks have often been identified as resulting from an interspecies DNA transfer event (18, 23, 52, 54, 65; M. C. Maiden, B. Malorny, and M. Achtman, Letter, *Mol. Microbiol.* **21**:1297–1298, 1996). If pneumococci are capable of between-species horizontal transfer, they, in all likelihood, undergo even more frequent within-species horizontal gene transfer that could contribute to the development of mosaic alleles. The transfer of capsule cassettes (16) is one example of an intraspecies horizontal gene transfer, and these events have recently been documented in vivo (12, 13, 31). Both interspecies and intraspecies horizontal gene transfers are facilitated in the pneumococci because of the widespread capability for natural transformation within this species (49, 51).

In this study, the *pspA* alleles examined were from a group of 24 diverse clinical pneumococcal isolates. These alleles were sequenced, revealing that *pspA* genes and PspA proteins have a highly complex mosaic structure. This mosaic diversity was examined in the light of its potential cross-reactive immunogenicity for use in protein-based vaccines to protect children and adults from this pathogen.

MATERIALS AND METHODS

DNA isolation, primers, and gene-specific PCR. Pneumococcal strains (see Table 1 and Results for description) were inoculated from agar plates containing 5% sheep blood into 15-ml cultures of Todd-Hewitt broth with 0.5% added yeast extract, and cells were harvested after only a few hours of growth. Chromosomal DNA for each strain was isolated by a modification of the genomic DNA procedure of Promega.

By using two oligonucleotide primers, LSM13 (in the promoter region of Rx1) and SKH2 (in the first of the 10 C-terminal repeats), a DNA fragment of about 1,200 to 1,800 bp corresponding to the *pspA* gene was amplified from nearly all pneumococcal strains (Fig. 1). Because the site for LSM13 is in the region of DNA immediately upstream of *pspA*, which is highly conserved in all pneumococcal strains (59), this particular PCR-generated fragment derives from the *pspA* chromosomal locus. The sequence of primer LSM13 is 5'-GCAAGCTTAGATATAGAAATTTGTAAAC-3', and that of primer SKH2 is 5'-CCACATACCGTTTTCTTGTTCAGCC-3'.

PCRs were carried out in a standard PCR mixture of 50 μ l containing 2.5 mM MgCl₂, 200 μ M (each) deoxynucleoside triphosphates (dNTPs), 50 pmol of each primer, and 2.5 U of *Taq* DNA polymerase. Cycling consisted of 95°C for 1 min, 62°C for 1 min, and 72°C for 3 min, repeated 30 times. Before sequencing, PCR products were purified from agarose gels by using Microcon gel nebulizer filters or were treated with shrimp alkaline phosphatase and exonuclease I (25).

Automated DNA sequence analysis and DNA sequencing strategy. DNA sequencing was performed by directly sequencing the LSM13 and SKH2 PCR-generated DNA fragments. Automated sequencing reactions used dye terminator chemistry and were run on an Applied Biosystems model ABI Prism 377 sequencer. Initial sequence runs for each strain used one of the two primers from the initial PCR amplification of the *pspA* gene—either LSM13 or SKH2. The sequence was extended and also confirmed by sequence runs in the opposite direction through the use of additional primers that were designed based on the initial sequence data for each gene. Approximately 40 additional primers were necessary to fully assemble the sequence of the 24 distinct *pspA* genes (sequences of primers are available upon request). The total length of the sequence data examined for each gene was >1,100 bp.

DNA sequence alignment. Sequence data for each strain's *pspA* PCR-generated fragments were assembled and edited by using Sequencher (GeneCodes, Inc.). Further editing, alignment, and additional analysis were performed with MacVector DNA sequence analysis software (Oxford Molecular). The protein alignment presented was generated by the Clustal W algorithm by using the Blossum30 amino acid-scoring matrix. Distance calculations given in Fig. 2 and 4 were those calculated by MacVector, but equivocal distances were found for alternative alignments in other programs. Because variation is high, simple distances were used. The graph of the distribution of distances was generated by using Microsoft Excel.

Nucleotide sequence accession numbers. The *pspA* sequences have been submitted to GenBank and assigned the accession no. AF071802 to AF071827 as indicated beside each strain in Table 1.

RESULTS

Selection of strains and nomenclature. Twenty-four independent clinical isolates representing 13 different capsule and

TABLE 1. Strains chosen for *pspA* sequencing

Strain name ^a	Capsule type	PspA type ^b	Yr isolated	Site of origin	Accession no.
Clade 1					
BG9739	4	26	1988	Alabama	AF071804
DBL6A	6A	19	<1980	Unknown ^c	AF071805
L81905	4	23	1984	Alaska	AF071809
BG8743	23	ND ^e	1987	Alabama	AF071803
AC94	9L	0	1987	Canada	AF071802
BG6692	33	16	1984	Alabama	AF071808
BG8838	6	27	1987	Alabama	AF071807
DBL1	6B	0	<1980	Unknown ^c	AF071806
Clade 2					
Rx1	Rough	25	1940s	United States	M74122
E134	23	ND	1994	Alabama	AF071811
EF10197	3	18	<1985	Sweden	AF071812
EF6796	6A	1	<1985	Sweden	AF071813
BG9163	6B	21	1987	Alabama	AF071815
DBL5	5	33	<1980	Unknown ^d	AF071810
WU2	3	1	1966	Missouri	AF071814
Clade 3					
EF3296	4	20	<1985	Sweden	AF071816
BG8090	19	16	1986	Alabama	AF071817
AC122	9V	18	1984	Canada	AF071818
Clade 4					
EF5668	4	12	<1985	Sweden	U89711
BG7561	15	ND	1985	Alabama	AF071824
BG7817	12	ND	1986	Alabama	AF071826
BG11703	18	ND	1992	Alabama	AF071821
Clade 5					
ATCC63033		7	1940s	Unknown	AF071820
Clade 6					
BG6380	37	ND	1984	Alabama	AF071823

^a PspA clade designations are from this paper.

^b Based on typing with MAb (15).

^c From B. Gray, origin unknown.

^d From S. Szu, origin unknown.

^e ND, not determined.

17 serologic PspA types (based on an earlier typing scheme using a MAb [15]) were chosen for sequencing in order to evaluate the patterns of diversity present at the *pspA* chromosomal locus in pneumococcal strains. An effort was made to choose strains which would span the range of diversity among

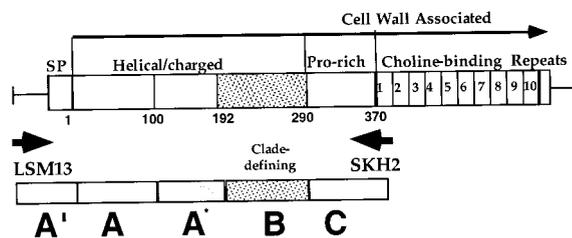


FIG. 1. Modular *pspA* gene showing windows of the sequence analyzed. Major domains of PspA are indicated above the line drawing. Within the α -helical or charged domain, a clade-defining region of the PspA molecule is indicated by the stippled box. Choline-binding repeats are numbered 1 to 10. The small numbers directly below the gene are equivalent to amino acid residue numbers based on PspA/Rx1, the prototype PspA molecule previously sequenced (62). Arrows LSM13 and SKH2 indicate two primers used for PCR. Boxes A', A, A*, B, and C indicate windows of sequence for analysis.

Family		Clade		1										2					3										
				1					2					3		4			5	6									
Strain		BG8743	BG9739	BG6692	DBL6A	L81905	AC94	DBL1	BG8838	WU2	DBL5	EF10197	E134	BG9163	EF6796	Rx1	EF3296	AC122	BG8090	EF5668	BG7561	BG7817	BG11703	ATCC6303	BG6380				
1	1	BG8743	76	71	88	77	81	85	71	40	59	47	77	70	48	51	43	53	48	50	47	45	46	50	38				
		BG9739	76	76	89	87	96	76	73	68	63	64	76	56	64	63	69	40	50	40	47	45	45	45	42	37			
		BG6692	74	72	76	73	87	74	74	86	75	59	73	60	60	59	78	40	49	42	54	53	52	53	54	38			
		DBL6A	70	89	79	76	85	76	74	73	59	58	81	53	58	35	60	39	48	38	47	44	44	44	44	46	38		
		L81905	75	98	71	88	74	72	86	62	62	62	75	54	63	61	63	40	50	40	47	44	43	44	42	37			
		AC94	70	80	76	81	79	76	82	75	69	61	73	50	80	60	69	41	51	42	55	52	51	53	54	41			
		DBL1	73	80	77	82	79	86	75	67	67	76	72	55	37	37	69	41	49	39	53	52	50	51	53	40			
		BG8838	72	78	88	88	73	79	80	76	76	64	70	60	59	47	76	40	50	41	54	54	53	54	51	39			
		WU2	71	67	78	73	88	73	73	79	60	60	78	68	66	67	86	40	50	40	59	54	55	55	53	39			
		DBL5	68	72	67	68	71	69	69	73	69	66	61	64	71	68	59	39	47	39	41	41	40	40	40	40	33		
2	2	EF10197	64	81	69	83	81	74	74	71	81	70	60	69	67	74	38	47	38	52	50	50	51	49	40				
		E134	61	64	66	82	83	81	62	65	69	70	63	66	64	62	39	47	41	44	42	41	41	45	33				
		BG9163	77	71	70	67	70	67	89	76	73	76	68	73	69	95	40	51	40	45	44	44	44	44	42	31			
		EF6796	73	69	65	62	69	65	85	65	69	75	58	68	90	91	43	52	43	46	44	43	43	46	29				
		Rx1	70	66	60	69	66	70	76	75	86	69	58	67	71	88	40	49	39	57	54	54	54	54	54	38			
		3	3	EF3296	42	39	42	41	39	41	40	42	43	43	42	43	42	42	47	42	43	42	42	42	42	42	42	33	
				AC122	53	51	54	52	50	54	52	54	56	54	51	52	54	50	57	43	43	43	43	43	43	43	43	32	
				BG8090	48	42	46	44	42	45	44	45	36	47	44	48	45	45	48	40	40	40	40	40	40	40	40	31	
				EF5668	55	54	62	58	53	62	59	62	58	54	57	54	56	54	64	45	45	45	45	45	45	45	45	45	45
				BG7561	49	48	52	49	47	49	49	52	62	53	50	52	53	50	60	45	45	45	45	45	45	45	45	45	45
4	4	BG7817	49	47	51	48	46	48	49	51	61	53	49	51	53	50	59	45	45	45	45	45	45	45	45	45			
		BG11703	49	48	52	49	47	48	49	52	62	53	50	51	54	50	59	45	45	45	45	45	45	45	45	45			
		ATCC6303	51	50	58	56	49	54	56	59	53	50	51	51	49	48	62	45	45	45	45	45	45	45	45	45			
6	6	BG6380	42	41	49	46	41	45	45	48	47	44	46	43	42	39	52	44	45	45	52	55	55	55	55	55			

FIG. 2. Pairwise comparisons among *pspA* genes and PspA proteins. Below the diagonal, values represent percent DNA identity. DNA comparisons are highlighted white for values less than 60%, light gray for values between 60 and 75%, and dark gray for values greater than 75%. Above the diagonal, values represent percent protein identity. Protein comparisons are highlighted white for values less than 55%, light gray for values between 55 and 73%, and dark gray for values greater than 73%. Black boxes along the diagonal indicate the 100% identity for each gene when compared with itself. Indications of the groups forming clades 1 to 6 and families 1 to 3 as described in the text are given along the left axis and across the top.

strains available to us. No two of the *pspA* genes examined came from strains either suspected or known to be clonally related. An initial group of 19 strains were chosen from strains of the most diverse PspA types we could find based on the serologic typing with seven MAbs that detected different combinations of epitopes in the α -helical or charged region of PspA (Fig. 1). In reviewing additional data on the reactivity of over 50 MAbs (R. Becker, unpublished data) with a panel of >50 additional strains, we identified 5 additional PspAs for sequencing that failed to react with any of the 7 original MAbs and that had unusual patterns of reactivity with the larger group of 50 MAbs. Thus, these 24 isolates are expected to exhibit a greater diversity than the PspAs in any random selection of pneumococcal isolates. The serologic properties, dates, and places of origin of the *S. pneumoniae* strains are given in Table 1. These isolates are from four main geographic sites: Alabama, Sweden, Alaska, and Canada. Because each gene sequenced differs from every other gene at multiple positions, each *pspA* gene was designated *pspA*/strain name and, similarly, the protein encoded was designated PspA/strain name (e.g., PspA/Rx1 and *pspA*/Rx1 from strain Rx1, an unencapsulated derivative of D39).

Identification of PspA families. Using the specific PCR primers LSM13 and SKH2, we have found that virtually every isolate of *S. pneumoniae* has a *pspA* gene (S. K. Hollingshead and D. E. Briles, unpublished data). We amplified and se-

quenced the complete α -helical portion of each of the 24 genes from strains in Table 1. The repeat region encoding the carboxy-terminal choline-binding domain was not sequenced, since previous studies have indicated that it is relatively invariant (10, 36). In pairwise comparisons, the genes and the proteins were exceptionally diverse in their α -helical regions. Figure 2 gives the percent identity values for nucleotide-nucleotide and protein-protein comparisons. When the pairwise identity was $\geq 60\%$, the sequences could be aligned with minimal gaps. When a pairwise comparison showed less than 60% identity, the alignment required the introduction of multiple gaps. By using the 40% nucleotide divergence as a cutoff value, two major families of PspA proteins were identified, with a single PspA falling into a third family. The requirement for multiple gaps in the alignments of proteins in different families, but not in the alignment of proteins within families, indicates a much weaker phylogenetic relationship between PspAs in different families than between the PspAs within a single family. Thus, all 24 *pspA* genes may not descend from a common ancestral gene over their length, and each PspA family may have a distinct ancestor. For the data set of 24 genes, base substitutions, replacement blocks, and small insertions or deletions were not clustered in one or more gene regions, but were distributed throughout the entire gene. Sample dot plots showing both same-family and cross-family comparisons indicate the variation throughout the genes (Fig. 3).

Several alignments of the 24 *pspA* genes were considered. One generated by the Clustal W algorithm and later adjusted to align portions of the proteins that are structurally the same and thus more likely to be homologous is presented in Fig. 4. These anchor regions included the mature N terminus of the protein, the transition zone between the α -helical region and the proline-rich regions of PspA, and one or more breakpoints in the coiled-coil structure as noted previously (28, 36, 62). The breakpoints serve to anchor the alignment even as diversity increases, necessitating the insertion of the many gaps in the overall alignment.

The low sequence identities between *pspA* genes in different families suggest that at least two different parental ancestral genes contributed to the current diversity in the *pspA* locus. This being the case, nonsynonymous sites tend to be saturated for some genes and not for others, and their examination is not valid for the cases in which ancestral sequences differ. The comparison of divergent sequences potentiates a high risk for nonhomologous alignment, especially when the overall lengths of the sequences compared differ. For this reason, we focused the initial examination of this diverse group of *pspA* genes on several windows within the gene that had the maximum potential to be correctly aligned. The three windows, A, B, and C, used for analysis are depicted in Fig. 1.

Alignment. The aligned amino acid sequences show the diversity of PspAs over the different windows (Fig. 4). The sequence differences in windows A to C are analyzed further in Fig. 5. Window A' includes approximately 200 nucleotides of DNA sequence upstream of each gene, including the region encoding the signal peptide. A' was highly conserved, ranging from 95 to 100% nucleotide identity (data not shown). The conservation of the DNA sequence in A' provided confirmation that the DNA fragment amplified and sequenced came from the *pspA* gene locus and not a related paralogous gene, such as that of *pspC*. This A' window is not included in the Fig. 5 analyses because it was so highly conserved and because it contains both coding and noncoding sequences.

Window A encodes the first 100 amino acids (300 nucleotides) of each PspA molecule, beginning with the first amino acid of the mature protein. The conserved region of signal peptides extends into the first 50 amino acids, with the majority of comparisons exhibiting greater than 50% amino acid identity in A. The PspA molecules are much less conserved over the second half of window A, where the sequences begin to diverge and fall into groups (Fig. 4).

Window A* is highly diverse (Fig. 4). The distribution values for this region are also not included in Fig. 5, because the lengths of the sequences in this region are so variable that meaningful alignments are not possible in this context. Inspection of this region, however, shows it to be a transition zone between the picture shown by window A and that shown by window B. The N-terminal end of A* is hypervariable, and the C-terminal end is characterized for the most part by the same sequence groups that are identified in window B.

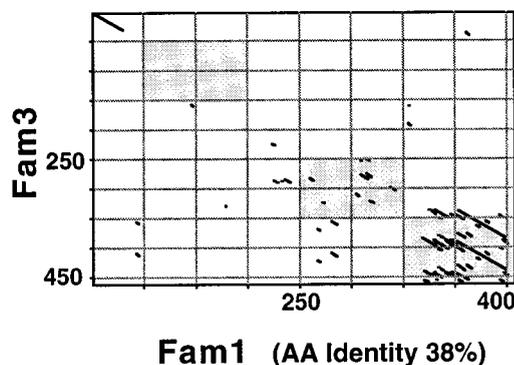
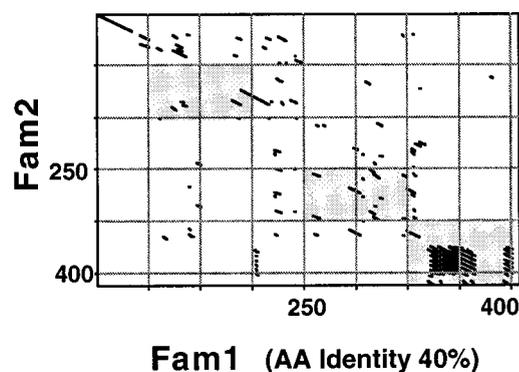
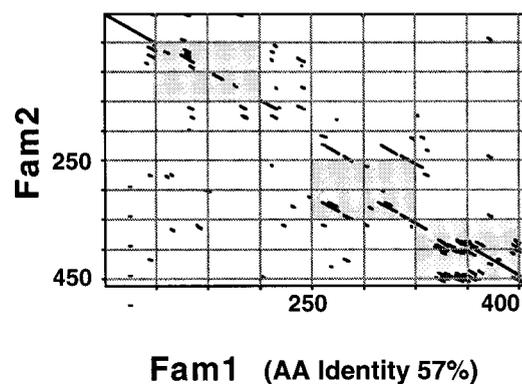
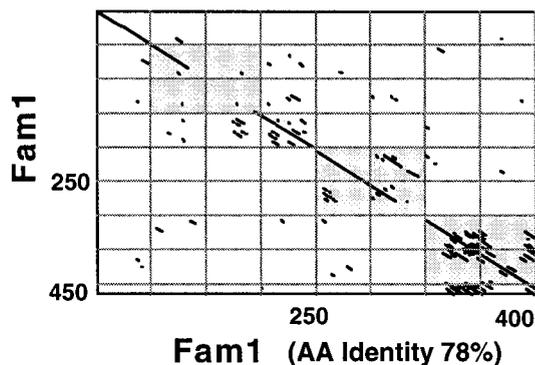


FIG. 3. Dot matrix analysis of representative PspA pairs indicating the regions where variance is found. In each case, the x axis is family 1 (Fam1) and the y axis is either family 1, family 2, or family 3. The protein comparison begins with the signal peptide and continues to around amino acid 400. Comparisons were made by the method of Pustell and Kafatos (50) with the default values of a window size of 8 and a minimum percent score of 60 by using the pam250 matrix and a hash value of 2. Specific proteins used in this analysis were PspA/Rx1 (family 1, x axis), PspA/BG6692 (family 1, y axis), PspA/EF5668 (family 2, first), PspA/EF3296 (family 2, second), and PspA/BG6380 (family 3). Three gray boxes in each graph represent windows A, B, and C. The percent amino acid (AA) identity values are given for each case.

Window B shows six sequence groups that are quite distinct (Fig. 4). These will be discussed further below.

Window C contains the proline-rich region of the PspA molecules. This region of PspA is quite repetitive, with many imperfect repeats of the sequence PAPAP (Fig. 4). The sequence-to-sequence distances in this window largely reflect the variation in the iterations of this repetitive sequence. A second factor influencing the distribution of sequence distances in this window is the presence or the absence of a 27-amino-acid non-proline-rich block of the consensus amino acid sequence EKS/TADQQAEDYARRSEEEYNRLTQQQ, which is present in this region in 14 of 24 of the sequences and absent from the others (Fig. 4 and 5).

Distribution of pairwise comparisons. The A-to-C windows are of approximately equal lengths, and each is proximal to the transition between one domain of the protein and another (Fig. 1). As was observed in the overall comparison of genes (Fig. 2), when individual sequence windows A to C were analyzed, any two PspA molecules were found to share as few as 14% or as many as 100% of the amino acids in a given window (data not shown). Although a similar range of distance comparisons was observed for each of the three windows, a closer examination reveals that the pattern of *pspA* diversity differs within each window (Fig. 5).

The 24×24 gene comparisons per window resulted in 276 distance calculations. The distribution of the 276 distances by percentile was plotted for each window. Figure 5 shows the distribution of these pairwise comparisons and indicates some striking differences between the variations seen at windows A, B, and C. Window A shows a normal distribution around the median of 70% amino acid identity. The pairwise comparisons at window B exhibit two modes. One peak in the distribution reflected pairwise comparisons of quite similar proteins (60 to 70% amino acid identity), and the second peak reflected pairwise comparisons of quite divergent proteins (20 to 30% identity). This biphasic distribution represents the distance profiles between genes of the same PspA sequence “type” versus distances between PspAs of differing sequence “types.” The two major types here correspond to PspA families as defined previously. The deviation of this profile from a normal distribution is significant ($P < 0.0001$; χ^2 test). Window C showed yet a third type of distribution in which the frequencies of pairwise comparisons were nearly equivalent over each distribution range sampled, giving a relatively broad flat curve with more spread. Hence, almost the same number of pairwise comparisons yielded identities in the 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80, 81 to 90, and 91 to 100% ranges, with only a slight peak in the 51 to 60% range.

The normal distribution for sequence comparisons in the A region results in a deep branching dendrogram for the A region in Fig. 5, which reflects the equidistant relationship among the proteins in this window. The strong group structure (B region) forming the two major families drives the biphasic distribution of the pairwise comparisons (Fig. 4 and 5) and the relatively tight branches or clusters seen in the dendrogram in the B window. The six groups seen in this window represent major PspA lineages and are discussed further below. The dendrogram for window C is mainly explained by the presence or the absence of the non-proline-rich block.

Clade-defining region of PspA. The region of sequence including window B was previously found to be an important protection-eliciting region in PspA/Rx1, making it critical to the development of PspA as a protein-based vaccine (37). Among the 24 PspAs, there are six groups that are mutually distinct at differences of >20% of amino acid positions in this B region. We are defining these groups as PspA clades, a term

coming from the science of cladistics and meaning monophyletic group. On the phenogram generated from the amino acid sequence in this region, each monophyletic group that was separated from the others with bootstrap values of 100% was considered a clade (Fig. 5B and 6C). The clades were numbered from 1 to 6. The assignment of the individual *pspA* genes to clade groups is given in Table 1, Fig. 5B, and Fig. 6C.

The clades of the 24 sequences are depicted in a tree that represents the different groupings as clusters or branch points on the tree (Fig. 6C, protein; 6D, DNA). Proteins within the same clade are greater than 90% identical in this B region. A higher branching division, which corresponds to the family division described earlier, is also shown on the tree (Fig. 2 and 6). Each of the two major families contains more than one clade, and family 3 has only a single clade (Fig. 6). The three families diverge from each other at >45% of amino acid positions in this B region. The family structure is also evident in distance trees constructed from the whole sequences (Fig. 6A, protein; 6B, DNA). The clade structure is blurred somewhat in whole gene trees due to the discordant contribution of main recombinant blocks within the A and C regions.

PspA proteins in the same clade appear to share a recent ancestral sequence for the B window or “clade-defining” region because they share their monophyletic position. Variation in this case is found to be restricted to single-amino-acid substitutions (Fig. 4). PspA proteins in the same family are also related in ancestry, but may represent products of parallel lineages from each other over the B window if they are in different clades. Variation in this case sometimes involves small blocks of divergent amino acid substitutions (Fig. 4). PspA proteins in different families are quite divergent, indicating a significant separation in ancestry of the respective *pspA* genes over the B window. Variation, in this third case, is the norm, and few identifiable blocks of conserved sequence are present.

Evidence for recombination in *pspA* genes. Strong evidence for past recombination events is apparent from the discordance of the trees for windows A, B, and C. Figure 5 shows the dendrograms generated based on sequence similarities for each of the three windows. The three dendrograms are non-congruent; the relationship for a strain-to-strain comparison is dependent upon the window or block of sequence used for assigning the relationship. For example, in window B, PspA/EF6796 is only 15% diverged from PspA/Rx1, and they both fall into clade 2; PspA/Rx1 is 74% diverged from PspA/EF3296. In window A, the situation is reversed; PspA/EF6796 is 65% diverged from PspA/Rx1 in this window and only 33% diverged from PspA/EF3296. Numerous examples of such crossovers exist within this limited data set of 24 PspA molecules. Evidence for recombination may also be viewed in Fig. 2 in the isolated shaded blocks as indications of the uneven distribution of homology and in numerous blocks along the alignment in Fig. 4.

DISCUSSION

The alignment of 24 diverse PspA proteins provides a snapshot of a highly complex mosaic gene pattern (Fig. 4). Mosaic genes are formed by recombination. In this data set, there are numerous instances in which individual *pspA* sequences may be very similar in sequence to certain *pspA* sequences in one location and very similar in sequence to other *pspA* sequences in another location within the alignment. Examples are present throughout both the DNA and protein alignments (Fig. 4). At the level of the encoded proteins, the impression is formed that nature has been shuffling the deck of epitopes through repeated recombination events over the *pspA* locus. This may

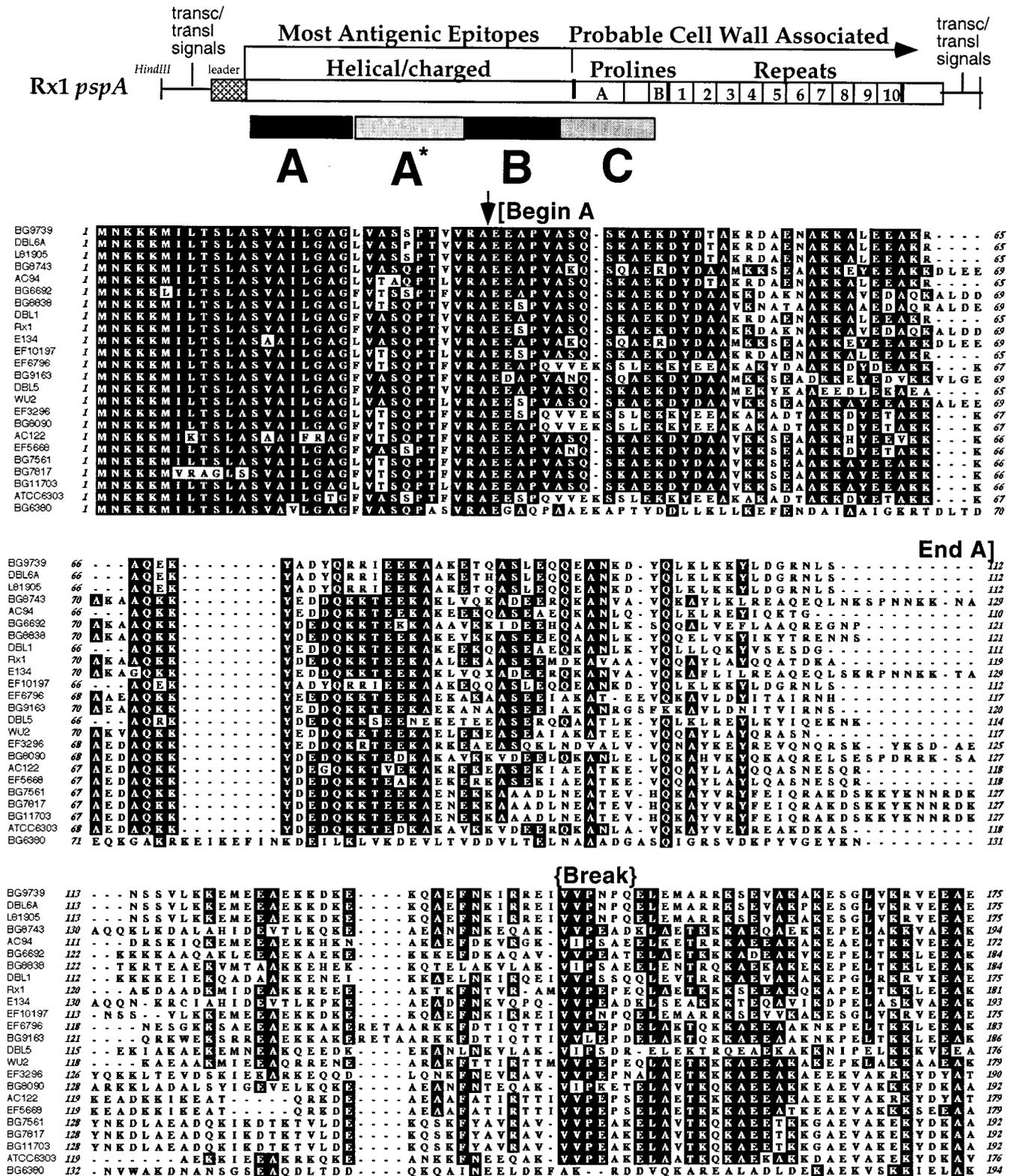


FIG. 4. Alignment of all 24 PspAs by the Clustal W algorithm and the Blosum30 amino acid scoring matrix in MacVector. The printed output shows amino acids common to over 51% of the group as darkened boxes. Regions used for window A, B, and C analyses are marked on the figure by brackets at the beginning and end. The alignment is ordered based on the clade groupings identified in the B window. Line breaks under strain names indicate clades 1 to 6. Indications above the sequence mark the start of the mature protein (amino acid [AA] 32), regions of break in coiled-coil (approximately from AA 143 to 148), and, in window C, the location of a non-proline sequence block (approximately from AA 368 to 395) that is optionally present or absent in this region of PspA molecules (Fig. 5).

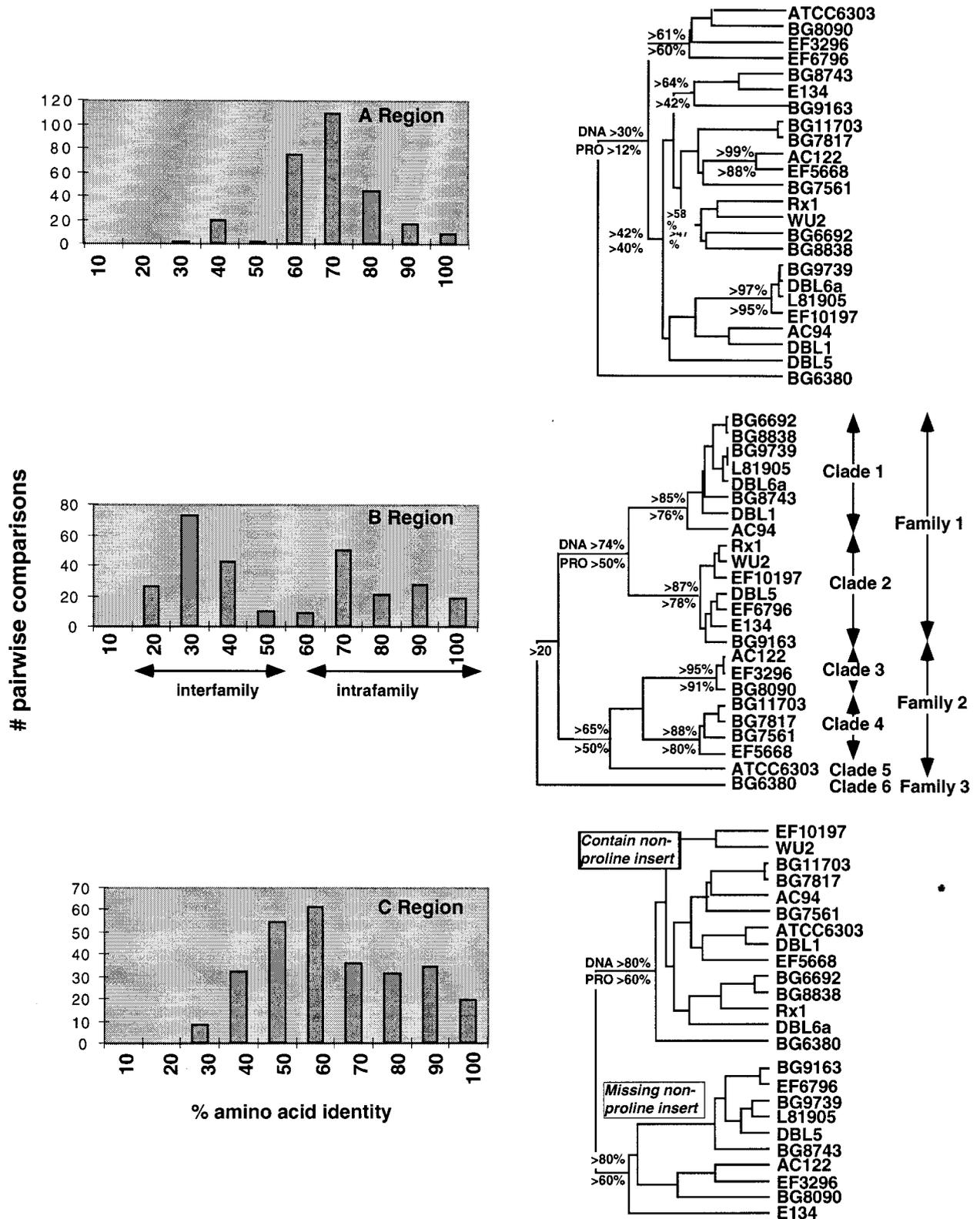


FIG. 5. Distribution of pairwise comparisons of sequence distance among PspA proteins in windows A, B, and C. The y axis in each graph represents the number of pairwise comparisons out of 276 total comparisons which fell within a range of the percent amino acid identity given on the x axis. To the right of each graph is a dendrogram representing the relationships between the PspAs for the sequence obtained in window A, B, or C as described in the text. The numbers at the indicated nodes indicate the cutoff values for DNA (above) or amino acid (below) identity that define the branches extending from that node. Clade groups defined in the text based on region B are indicated by arrows in the center of the figure.

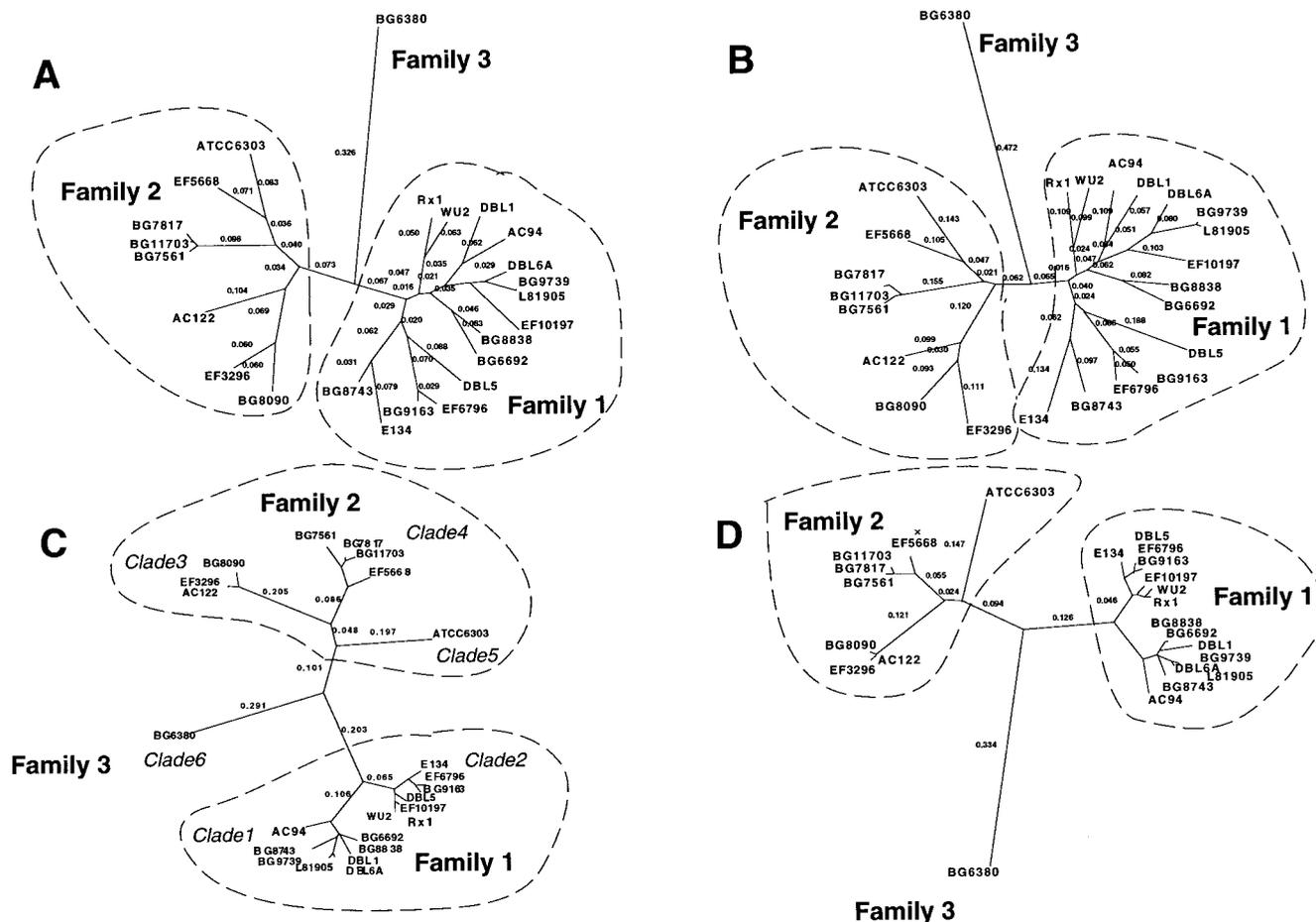


FIG. 6. Relationship between PspA sequences over window B and over the entire sequence. Shown are four unrooted phylogenetic trees generated from mean distances by the neighbor-joining method in the program PAUP 4.0B, as follows: A, proteins, all; B, DNA, all; C, proteins, B region; D, DNA, B region. The numbers on the tree indicate the distances along the branch lengths as calculated by PAUP with the terminal branch distances suppressed. Clades and families as defined in the text are listed near the clusters or branches that they encompass.

diverges from the IgA2 protease gene in *Streptococcus sanguis* at 38% of nucleotide sites and 41% of amino acid sites (48).

Source of mosaic blocks. Recent studies in a number of laboratories investigating both *Neisseria* and streptococci have suggested the possibility of “global” gene pools (19, 52, 65; Maiden et al., Letter) based on documented interspecies transfer of gene segments (18, 32, 56, 57). For the most part, these global gene pools operate at the genus-wide level. For *S. pneumoniae*, the closest relatives among other streptococci appear to be *Streptococcus oralis* and *Streptococcus mitis* based on 16S rRNA, 23S rRNA, and other factors (30, 58). Indeed, transfers in *pbp* alleles of *S. pneumoniae* have most frequently been traced to these very closely related oral commensal species (11, 17, 52). IgA1 protease alleles of *S. oralis* and *S. mitis* are also shared (47). Although not extensively studied, there are gene loci whose alleles are found to be species restricted as well. The presence of the *pspA* gene is not demonstrable in the nearest relatives of *S. pneumoniae*, with the possible exception of a few *S. mitis*-like organisms (46). Thus, a source for interspecies transfer of *pspA* blocks is missing. Second, both family 1 and family 2 PspA proteins are prevalent among recent clinical isolates of pneumococci of all capsular serotypes, a finding which argues that they are not recent acquisitions from another

streptococcal species (Hollingshead and Briles, unpublished data).

Although interspecies horizontal transfer of *pspA* is possible, it seems likely that the vast majority of recombination in the *pspA* locus comes from intraspecies horizontal transfer. The pneumococcus is carried in the nasopharynx, and frequently more than one strain can be carried at the same time (21, 55). *S. pneumoniae* is well known for its special capacity to take up DNA from its environment and incorporate it into its chromosome. Because other pneumococci share variant *pspA* loci that still have enough homology to allow efficient recombination, the neighboring strains of pneumococci are likely to serve as the most common donors.

Implications for vaccine development. One aim of sequencing this many *pspA* genes was to address the breadth and depth of the genetic diversity that was present at this locus. It was felt that this information might be an aid in understanding the serological diversity of *pspA* and thus shed light on the appropriate composition of a PspA-based pneumococcal vaccine for humans. Although at first glance the level of diversity appears daunting for this purpose, PspA is remarkably immunogenic and cross-reactive. The data from cross-protective studies in mice (36, 39) show that a PspA molecule in clades 1 to 4 can

elicit protective antibody responses to pneumococci which differ from the immunogen at >50% of their amino acid positions in the α -helical portion of the molecule (5, 60). Immunization of healthy adults with a single recombinant PspA stimulated cross-reactive antibodies to heterologous PspAs in different clades or families which were cross-protective in mice (43; Briles et al., submitted). In producing a broadly protective vaccine, the cross-reactive responses that have been documented could easily be further encouraged by the inclusion of PspA molecules from each of the families and clades in the vaccine.

The verification from sequence data that virtually all PspAs share significant short stretches of amino acids helps our understanding of the nature of PspA cross-reactivity. Cross-reactive antibodies may recognize the amino acid identities that exist in small stretches depicted as short lines (Fig. 3) and indicating a region scoring above the cutoff value in a dot plot of *pspA* genes or PspA proteins. These short stretches of matching amino acids are present in comparing any two PspA proteins when using default cutoff values with a window size of 8, minimal percentage score of 60%, and hash value of 2. If the short conserved sequences are of sufficient length and structure to comprise an epitope, then their presence along the α -helical or charged region of PspA may mark encoded cross-reactive sites. The cross-reactive sites are the residues of the repeated recombination that has exchanged sequence blocks and generated new chimeric sites within the *pspA* gene. The scattered distribution of cross-reactive sites explains in part the different combinations of epitopes in PspA that were previously detected by MAbs (15).

Paradoxically, the extraordinary degree of mosaicism exhibited by *pspA* genes may indicate the importance of this surface protein as a natural target for host defense against the pneumococcus. Surface proteins in bacteria and antigen receptor proteins of T and B cells in eukaryotes are often those proteins which exhibit the greatest divergence in species-to-species comparisons (42). This divergence often results from the processes that create mosaic proteins. The presence of mosaicism and, in particular, the complex pattern of mosaicism seen in the *pspA* gene locus could indicate that *pspA* is often the target of positive or negative selection in the interaction between the pneumococcus and its human host. The antibodies that PspA elicits clearly play an important role in protection in mouse models (8), and the protein appears to interfere with complement fixation *in vivo* in mice and is likely to do so in humans as well (1, 6, 44, 61). Boosting the level of natural antibodies to PspA may bolster the host's ability to resist the pneumococcus.

ACKNOWLEDGMENTS

We gratefully acknowledge the technical support of Xiping Wu and Terri Readdy and are thankful to Alexis Brooks-Walter, Elliot Lefkowitz, and D. Ashley Robinson for suggestions throughout the course of this work and Sylvie Rodriguez for reading the manuscript.

The work was supported in part by grants AI21548, AI40645, and HL54818 from the National Institutes of Health (NIH) and by a contract from Aventis Pasteur. The DNA Sequencing Core Facilities were supported by a grant from NIH to the Center for Aids Research (AI27767), by a grant from the Tennessee Valley Authority to the Department of Microbiology at the University of Alabama at Birmingham, and by the Howard Hughes Foundation to the University of Alabama at Birmingham Medical School.

REFERENCES

- Abeyta, M. 1999. Ph.D. dissertation. University of Alabama at Birmingham, Birmingham.
- Atherton, J. C., P. Cao, R. M. Peek, Jr., M. K. Tummuru, M. J. Blaser, and T. L. Cover. 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter*

pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. *J. Biol. Chem.* **270**:17771–17777.

- Atherton, J. C., P. M. Sharp, T. L. Cover, G. Gonzalez-Valencia, R. M. Peek, Jr., S. A. Thompson, C. J. Hawkey, and M. J. Blaser. 1999. Vacuolating cytotoxin (*vacA*) alleles of *Helicobacter pylori* comprise two geographically widespread types, m1 and m2, and have evolved through limited recombination. *Curr. Microbiol.* **39**:211–218.
- Breiman, R., J. C. Butler, F. C. Tenover, J. Elliot, and R. R. Facklam. 1994. Emergence of drug-resistant pneumococcal infections in the United States. *JAMA* **271**:1831–1835.
- Briles, D. E., S. Hollingshead, A. Brooks-Walter, G. S. Nabors, L. Ferguson, M. Schilling, S. Gravenstein, P. Braun, J. King, and A. Swift. 2000. The potential to use PspA and other pneumococcal proteins to elicit protection against pneumococcal infection. *Vaccine* **18**:1707–1711.
- Briles, D. E., S. K. Hollingshead, J. E. King, A. Swift, P. Braun, L. M. Ferguson, M. Nahm, and G. S. Nabors. Immunization of humans with rPspA elicits antibodies which passively protect mice from fatal infection with *Streptococcus pneumoniae* expressing heterologous PspA molecules. *J. Infect. Dis.*, in press.
- Briles, D. E., S. K. Hollingshead, E. Swiatlo, A. Brooks-Walter, A. Szalai, A. Virolainen, L. S. McDaniel, K. A. Benton, P. White, K. Prellner, A. Hermanson, C. Needleman, H. Van Dijk, and M. J. Crain. 1997. PspA and PspC: their potential for use as pneumococcal vaccines. *Microb. Drug Resist.* **3**:401–408.
- Briles, D. E., E. Swiatlo, and K. Edwards. 2000. Vaccine strategies for *S. pneumoniae*, p. 419–433. In D. L. Stevens (ed.), *Streptococci*. Oxford University Press, New York, N.Y.
- Briles, D. E., R. C. Tart, E. Swiatlo, J. P. Dillard, P. Smith, K. A. Benton, B. A. Ralph, A. Brooks-Walter, M. J. Crain, S. K. Hollingshead, and L. S. McDaniel. 1998. Pneumococcal diversity: considerations for new vaccine strategies with emphasis on pneumococcal surface protein A (PspA). *Clin. Microbiol. Rev.* **11**:645–657.
- Briles, D. E., R. C. Tart, H.-Y. Wu, B. A. Ralph, M. W. Russell, and L. S. McDaniel. 1996. Systemic and mucosal protective immunity to pneumococcal surface protein A. *Ann. N. Y. Acad. Sci.* **797**:118–126.
- Brooks-Walter, A., D. E. Briles, and S. K. Hollingshead. 1999. The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect. Immun.* **67**:6533–6542.
- Coffey, T. J., C. G. Dowson, M. Daniels, and B. G. Spratt. 1993. Horizontal spread of an altered penicillin-binding protein 2B gene between *Streptococcus pneumoniae* and *Streptococcus oralis*. *FEMS Microbiol. Lett.* **110**:335–339.
- Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt, and J. M. Musser. 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsule biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5**:2255–2260.
- Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, J. C. Paton, and B. G. Spratt. 1998. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol. Microbiol.* **27**:73–83.
- Cowan, M. J., A. J. Ammann, D. W. Wara, V. M. Howie, L. Schultz, N. Doyle, and M. Kaplan. 1978. Pneumococcal polysaccharide immunization in infants and children. *Pediatrics* **62**:721–727.
- Crain, M. J., W. D. Waltman II, J. S. Turner, J. Yother, D. E. Talkington, L. S. McDaniel, B. M. Gray, and D. E. Briles. 1990. Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infect. Immun.* **58**:3293–3299.
- Dillard, J. P., M. W. Vandersea, and J. Yother. 1995. Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J. Exp. Med.* **181**:973–983.
- Dowson, C., A. Hutchison, N. Woodford, A. Johnson, R. George, and B. G. Spratt. 1990. Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **87**:5858–5862.
- Dowson, C. G., A. Hutchison, J. A. Brannigan, R. C. George, D. Hansman, J. Linares, A. Tomasz, J. Maynard, and B. G. Spratt. 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **86**:8842–8846.
- Fussenegger, M., T. Rudel, R. Barten, R. Ryll, and T. F. Meyer. 1997. Transformation competence and type-4 pilus biogenesis in *Neisseria gonorrhoeae*—a review. *Gene* **192**:125–134.
- Gray, B. M. 1995. Pneumococcal infections in an era of multiple antibiotic resistance. *Adv. Pediatr. Infect. Dis.* **11**:55–100.
- Gray, B. M., G. M. Converse III, and H. C. Dillon. 1980. Epidemiologic studies of *Streptococcus pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life. *J. Infect. Dis.* **142**:923–933.
- Greenwood, B. 1999. The epidemiology of pneumococcal infection in chil-

- dren in the developing world. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **354**:777–785.
23. Hakenbeck, R., A. König, I. Kern, M. van der Linden, W. Keck, D. Billot-Klein, R. Legrand, B. Schoot, and L. Gutmann. 1998. Acquisition of five high- M_r penicillin-binding protein variants during transfer of high-level β -lactam resistance from *Streptococcus mitis* to *Streptococcus pneumoniae*. *J. Bacteriol.* **180**:1831–1840.
 24. Hammerschmidt, S., G. Bethe, P. H. Remone, and G. S. Chhatwal. 1999. Identification of pneumococcal surface protein A as a lactoferrin-binding protein of *Streptococcus pneumoniae*. *Infect. Immun.* **67**:1683–1687.
 25. Hanke, M., and M. Wink. 1994. Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs. *Bio-Techniques* **17**:858–860.
 26. Hausdorff, W. P., J. Bryant, C. Kloek, P. R. Paradiso, and G. R. Siber. 2000. The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin. Infect. Dis.* **30**:122–140.
 27. Hausdorff, W. P., J. Bryant, P. R. Paradiso, and G. R. Siber. 2000. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin. Infect. Dis.* **30**:100–121.
 28. Jedrzejewski, M. J., S. K. Hollingshead, J. Lebowitz, L. Chantalat, D. E. Briles, and E. Lamani. 2000. Production and characterization of the functional fragment of pneumococcal surface protein A. *Arch. Biochem. Biophys.* **373**:116–125.
 29. Jernigan, D. B., M. S. Cetron, and R. F. Breiman. 1996. Defining the public health impact of drug resistant *Streptococcus pneumoniae*: report of a working group. *Morb. Mortal. Wkly. Rep.* **45**:1–20.
 30. Kawamura, Y., X.-G. Hou, F. Sultana, H. Miura, and T. Ezaki. 1995. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int. J. Syst. Bacteriol.* **45**:406–408. (Erratum, 45:882.)
 31. Kell, C. M., J. Z. Jordens, M. Daniels, T. J. Coffey, J. Bates, J. Paul, C. Gilks, and B. G. Spratt. 1993. Molecular epidemiology of penicillin-resistant pneumococci isolated in Nairobi, Kenya. *Infect. Immun.* **61**:4382–4391.
 32. Laible, G., B. G. Spratt, and R. Hakenbeck. 1991. Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5**:1993–2002.
 33. Lin, B., S. K. Hollingshead, J. E. Coligan, M. L. Egan, J. R. Baker, and D. E. Pritchard. 1994. Cloning and expression of the gene for group B streptococcal hyaluronate lyase. *J. Biol. Chem.* **269**:30113–30116.
 34. Maynard Smith, J. 1996. Population genetics: an introduction, p. 2685–2690. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 2. American Society of Microbiology, Washington, D.C.
 35. Maynard Smith, J., C. G. Dowson, and B. G. Spratt. 1991. Localized sex in bacteria. *Nature* **349**:29–31.
 36. McDaniel, L. S., D. O. McDaniel, S. K. Hollingshead, and D. E. Briles. 1998. Comparison of the PspA sequence from *Streptococcus pneumoniae* EF5668 to the previously identified PspA sequence from strain Rx1 and ability of PspA from EF5668 to elicit protection against pneumococci of different capsular types. *Infect. Immun.* **66**:4748–4754.
 37. McDaniel, L. S., B. A. Ralph, D. O. McDaniel, and D. E. Briles. 1994. Localization of protection-eliciting epitopes on PspA of *Streptococcus pneumoniae* between amino acid residues 192 and 260. *Microb. Pathog.* **17**:323–337.
 38. McDaniel, L. S., J. S. Sheffield, P. Delucchi, and D. E. Briles. 1991. PspA, a surface protein of *Streptococcus pneumoniae*, is capable of eliciting protection against pneumococci of more than one capsular type. *Infect. Immun.* **59**:222–228.
 39. McGraw, E. A., J. Li, R. K. Selander, and T. S. Whittam. 1999. Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic *Escherichia coli*. *Mol. Biol. Evol.* **16**:12–22.
 40. Milkman, R. 1997. Recombination and population structure in *Escherichia coli*. *Genetics* **146**:745–750.
 41. Milkman, R., and A. Stoltzfus. 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. *Genetics* **120**:359–366.
 42. Murphy, P. M. 1993. Molecular mimicry and the generation of host defense protein diversity. *Cell* **72**:823–826.
 43. Nabors, G. S., P. A. Braun, D. J. Herrmann, M. L. Heise, D. J. Pyle, S. Gravenstein, M. Schilling, L. M. Ferguson, S. K. Hollingshead, D. E. Briles, and R. S. Becker. 2000. Immunization of healthy adults with a single recombinant pneumococcal surface protein A (PspA) stimulated cross-reactive antibodies to heterologous PspA molecules. *Vaccine* **18**:1743–1754.
 44. Neeleman, C., S. P. M. Geelen, P. C. Aerts, M. R. Daha, T. E. Mollnes, J. J. Roord, G. Posthuma, H. van Dijk, and A. Fleer. 1999. Resistance to both complement activation and phagocytosis in type 3 pneumococci is mediated by binding of complement regulatory protein factor H. *Infect. Immun.* **67**:4517–4524.
 45. Ogunniyi, A. D., R. L. Folland, D. E. Briles, S. K. Hollingshead, and J. C. Paton. 2000. Immunization of mice with combinations of pneumococcal virulence proteins elicits enhanced protection against challenge with *Streptococcus pneumoniae*. *Infect. Immun.* **68**:3028–3033.
 46. Poulsen, K., and M. Kilian. 1998. Genetic relationships between *Streptococcus pneumoniae*, *Streptococcus mitis*, and *Streptococcus oralis*, p. 81. *In* Proceedings of the ASM Conference on Streptococcal Genetics. American Society for Microbiology, Washington, D.C.
 47. Poulsen, K., J. Reinholdt, C. Jespersgaard, K. Boye, T. A. Brown, M. Hauge, and M. Kilian. 1998. A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: evidence for recombination within and between species. *Infect. Immun.* **66**:181–190.
 48. Poulsen, K., J. Reinholdt, and M. Kilian. 1996. Characterization of the *Streptococcus pneumoniae* immunoglobulin A1 protease gene (*iga*) and its translation product. *Infect. Immun.* **64**:3957–3966.
 49. Pozzi, G., L. Masala, F. Iannelli, R. Manganeli, L. S. Håvarstein, L. Piccoli, D. Simon, and D. A. Morrison. 1996. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *J. Bacteriol.* **178**:6087–6090.
 50. Pustell, J., and F. C. Kafatos. 1982. A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Res.* **10**:4765–4782.
 51. Ramirez, M., D. A. Morrison, and A. Tomasz. 1997. Ubiquitous distribution of the competence related genes *comA* and *comC* among isolates of *Streptococcus pneumoniae*. *Microb. Drug Resist.* **3**:39–52.
 52. Reichmann, P., A. König, J. Linares, F. Alcaide, F. C. Tenover, L. McDougal, S. Swidsinski, and R. Hakenbeck. 1997. A global gene pool for high-level cephalosporin resistance in commensal *Streptococcus* species and *Streptococcus pneumoniae*. *J. Infect. Dis.* **176**:1001–1012.
 53. Shinefield, H. R., S. Black, P. Ray, I. Chang, N. Lewis, B. Fireman, J. Hackell, P. R. Paradiso, G. Siber, R. Kohberger, D. V. Madore, F. J. Malinowski, A. Kimura, C. Le, I. Landaw, J. Aguilar, and J. Hansen. 1999. Safety and immunogenicity of heptavalent pneumococcal CRM197 conjugate vaccine in infants and toddlers. *Pediatr. Infect. Dis. J.* **18**:757–763.
 54. Sibold, C., J. Henrichsen, A. König, C. Martin, L. Chalkley, and R. Hakenbeck. 1994. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. *Mol. Microbiol.* **12**:1013–1023.
 55. Sluijter, M., H. Faden, R. de Groot, N. Lemmens, W. H. F. Goessens, A. van Belkum, and P. W. M. Hermans. 1998. Molecular characterization of pneumococcal nasopharynx isolates collected from children during their first 2 years of life. *J. Clin. Microbiol.* **36**:2248–2253.
 56. Spratt, B. G. 1988. Hybrid penicillin-binding proteins in penicillin-resistant strains of *Neisseria gonorrhoeae*. *Nature* **332**:173–176.
 57. Spratt, B. G. 1994. Resistance to antibiotics mediated by target alterations. *Science* **264**:388–393.
 58. Sultana, F., Y. Kawamura, X. G. Hou, S. E. Shu, and T. Ezaki. 1998. Determination of 23S rRNA sequences from members of the genus *Streptococcus* and characterization of genetically distinct organisms previously identified as members of the *Streptococcus anginosus* group. *FEMS Microbiol. Lett.* **158**:223–230.
 59. Swiatlo, E., A. Brooks-Walter, D. E. Briles, and L. S. McDaniel. 1997. Oligonucleotides identify conserved and variable regions of *pspA* and *pspA*-like sequences of *Streptococcus pneumoniae*. *Gene* **188**:279–284.
 60. Tart, R. C., L. S. McDaniel, B. A. Ralph, and D. E. Briles. 1996. Truncated *Streptococcus pneumoniae* PspA molecules elicit cross-protective immunity against pneumococcal challenge in mice. *J. Infect. Dis.* **173**:380–386.
 61. Tu, A.-H., R. L. Fulgham, M. A. McCrory, D. E. Briles, and A. J. Szalai. 1999. Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect. Immun.* **67**:4720–4724.
 62. Yother, J., and D. E. Briles. 1992. Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *J. Bacteriol.* **174**:601–609.
 63. Yother, J., G. L. Handsome, and D. E. Briles. 1992. Truncated forms of PspA that are secreted from *Streptococcus pneumoniae* and their use in functional studies and cloning of the *pspA* gene. *J. Bacteriol.* **174**:610–618.
 64. Yother, J., and J. M. White. 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. *J. Bacteriol.* **176**:2976–2985.
 65. Zhou, J., L. D. Bowler, and B. G. Spratt. 1997. Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. *Mol. Microbiol.* **23**:799–812.