Human SP-A Locus: Allele Frequencies and Linkage Disequilibrium between the Two Surfactant Protein A Genes

Joanna Floros, Susan DiAngelo, Michael Koptides, Anne M. Karinch, Peter K. Rogan, Heber Nielsen, Roger G. Spragg, Kristi Watterberg, and Gina Deiter

Departments of Cellular and Molecular Physiology and Pediatrics, The Milton S. Hershey Medical Center, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania; Department of Pediatrics, Tufts University School of Medicine, Boston, Massachusetts; and Medicine Service La Jolla Veteran's Affairs Medical Center, San Diego, California

Two surfactant protein A (SP-A) genes and several alleles for each SP-A locus have been previously described. In this report we investigate the potential usefulness of the SP-A loci as markers for genetic studies. We establish conditions that allow the identification of alleles with very similar sequences; We also determine the degree of polymorphism for each SP-A locus: The heterozygosity and polymorphism information content (PIC) values for the SP-A1 locus are 0.63 and 0.55, respectively, and for the SP-A2 locus are 0.50 and 0.56. In the course of these studies, we identify one new SP-A2 allele and show that the SP-A1 and SP-A2 loci are in linkage disequilibrium ($P < 0.000001$). We also identify 19 of the 20 possible haplotypes in a population of $n = 239$. Nine of the observed haplotypes reach statistical significance ($P < 0.01$) in this population, and the segregation of two haplotypes ($6A^2/1A^0$ and $6A^4/1A$) without recombination is verified in a family pedigree. These data together indicate that both SP-A loci are sufficiently polymorphic to be good markers for use in genetic studies. Furthermore, the finding of one novel allele suggests that additional unknown SP-A alleles are yet to be found. **Floros, J., S. DiAngelo, M. Koptides, A. M. Karinch, P. K. Rogan, H. Nielsen, R. G. Spragg, K. Watterberg, and G. Deiter. 1996. Human SP-A Locus: allele frequencies and Linkage disequilibrium between the two surfactant protein A genes. Am. J. Respir. Cell Mol. Biol. 15:489–498.**

Pulmonary surfactant, a lipoprotein complex, is essential for normal lung function. It consists of a variety of lipids, especially phospholipids, and a number of proteins. The surfactant-associated proteins known to contribute to surfactant function, structure, and metabolism are surfactant protein A (SP-A), SP-B, and SP-C (1). The genes encoding SP-A are located on chromosome bands 10q21 through q24 (2). There are two functional SP-A genes (SP-A1 and SP-A2) (3, 4) and an SP-A pseudogene (5).

(Received in original form November 18, 1995 and in revised form April 1, 1996)

Dr. Rogan's present address is the Department of Human Genetics, Allegheny-Singer Research Institute, Pittsburgh, PA.

Address correspondence to: Joanna Floros, Ph.D., Dept of Cellular and Molecular Physiology, The Milton S. Hershey Medical Center, The Pennsylvania State University College of Medicine, 500 University Drive, Hershey, PA 17033-0850.

Abbreviations: adult respiratory distress syndrome, ARDS; adenosine triphosphate, ATP; polymerase chain reaction, PCR; polymorphism information content, PIC; respiratory distress syndrome, RDS; sodium dodecyl sulfate, SDS; surfactant protein, SP.

The DNA sequences of the two functional SP-A genes are very similar to one another. The genomic structural organization of each SP-A gene consists of four coding exons and spans a region of < 5 kb (3, 4). Each SP-A gene contains a number of untranslated exons at the 5′ end (denoted A, B, C, and D) that generate several alternatively spliced transcripts (6–8). Each functional SP-A gene encodes a precursor molecule (9, 10) that undergoes a number of post-translational modifications resulting in a sialoglycoprotein. The *in vitro* translated protein from each functional SP-A gene is required for the formation of a structure similar to that observed in native SP-A (11). Both products appear to be necessary for a fully functional and stable mature SP-A (12). cDNA sequences that correspond to the SP-A1 and SP-A2 genes have been published and denoted 6A and 1A, respectively (13). A number of alleles have been recently identified and characterized for each SP-A gene (6, 7, 14). The naming of alleles is based on the nomenclature of the cDNAs (6A and 1A), and thus the alleles for the SP-A1 gene are denoted $6A^n$, and the alleles of the SP-A2 gene are denoted $1A^n$. The sequences of the SP-A alleles are very similar within the coding regions (≥ 99%), making identification of each allele challenging.

Because SP-A has been shown to be involved in many

aspects of surfactant physiology and biology, we wished to determine whether the SP-A locus would be a useful marker for genetic studies of specific pulmonary diseases. Toward this goal, in this report we (*1*) describe conditions that permit SP-A gene-specific amplification and hybridization, and allele-specific hybridizations; (*2*) describe one additional allele for the SP-A2 gene; (*3*) determine the degree of polymorphism of the SP-A1 and SP-A2 loci in several human populations by assessing their heterozygosity and polymorphism information content (PIC) values in these populations; and (*4*) show that the two SP-A genes (SP-A1 and SP-A2) are in linkage disequilibrium.

Materials and Methods

Molecular Reagents

The SP-A cDNAs that we use as hybridization controls are our previously published sequences: 6A and 1A (13); $6A^2$, $6A^3$, $6A^4$, and $1A^0$, $1A^1$, $1A^2$ (6, 7). The genomic SP-A1 and SP-A2 clones that we use as controls in PCR are from Genome Systems (St. Louis, MO). The genomic SP-A1 clone of allele $6A^0$ (6) is a kind gift of Dr. Tyler White (Fcios, Inc., Mountain View, CA). The genomic SP-A1 clone of allele $6A^1$ (6) is the clone previously described by Rishi and colleagues (14).

Oligonucleotides shown in Figure 1 and Table 1 are synthesized by the Macromolecular Core at the Pennsylvania State University College of Medicine (Hershey, PA). The oligonucleotide number identifies the order of the oligonucleotide in our logbook. Reagents used in polymerase chain reaction (PCR), including Taq polymerase, are purchased from Perkin-Elmer, and γ^{32}P-adenosine triphosphate (ATP) (3,000 Ci/mmol) from New England Nuclear (Boston, MA).

Specimens

DNA is extracted from either cord or venous blood, lung tissue, or cell pellets from bronchoalveolar lavage. These specimens are obtained from a variety of populations as shown in RESULTS. The groups "RDS" and "Preterm" represent specimens obtained from prematurely born infants who were either diagnosed with respiratory distress syndrome (RDS) or were free of RDS. Specimens denoted "random" are obtained from randomly selected adult individuals. The "mixed" group consisted of samples from patients diagnosed with adult respiratory distress syndrome (ARDS) or at risk for ARDS, from patients with bronchopulmonary dysplasia, and from patients with lung adenocarcinoma. The following DNA samples are kindly provided by the Colorado Lung Spore Tissue Bank Core Laboratory: 15 DNA samples from lung carcinomas, 16 from bronchoalveolar lavage from normal nonsmokers, and four from smokers. More than 80% of the Nigerian samples are obtained from individuals of the Yoruba and Ibo tribes (south and southeast part of Nigeria). Paternity in the family pedigree is confirmed by PCR with two sets (D21S1436 and D21S1437) of microsatellite polymorphic markers.

PCR and Hybridization Conditions for the SP-A1 Gene and SP-A1 Alleles

PCR. SP-A1 is amplified with 50 ng each of oligonucleotides 326 (SP-A1–specific) and 68 (Figure 1 and Table 1). The reaction consists of 200 ng genomic DNA in 1.5 mM $MgCl_2$, 10 mM Tris, pH 8.3, 50 mM KCl, and 0.3 mM dNTPs (Perkin Elmer) in a total volume of 100 μl, with the following cycling conditions: initial denaturation at 95°C 5′, addition of 2.5 U of AmpliTaq, followed by 30 cycles at 94°C 30″, 58°C 30″, 72°C 1′ 30″, and a final 5′ extension step at 72°C.

Slot-blot preparation. An aliquot of the PCR product is used to evaluate the success of the PCR reaction by agarose gel electrophoresis, and the remaining PCR products are extracted with phenol:chloroform:isoamyl alcohol (25:24:1). An attempt was made to load equivalent amounts (as determined by the intensity of the ethidium bromide–stained bands on the agarose gel) of each DNA sample onto triplicate slot-blots. DNA loaded onto the blots was prepared as follows. DNA is diluted in H_2O to a final volume of 100 μl and a final concentration of 0.1 N NaOH. The sample is first incubated at room temperature for 10′, and then placed on ice, at which time 900 μl of 10× SSC is added. Samples are held on ice until slot-blot preparation is completed. Each well is rinsed once with 300 μl of 10× SSC, then 300 μl of DNA sample is applied and the well is rinsed again twice with 300 μl of 10× SSC.

The membrane used for slot blots is Genescreen hybridization transfer membrane (Dupont, Boston, MA). The DNA is fixed to the membrane by baking at 80°C under vacuum for 2 h. To assess equivalency of loading for all alleles of each gene, hybridization to a common oligonucleotide was performed. (The common specific oligonucleotide for SP-A1 is 293 and for SP-A2 is 292.)

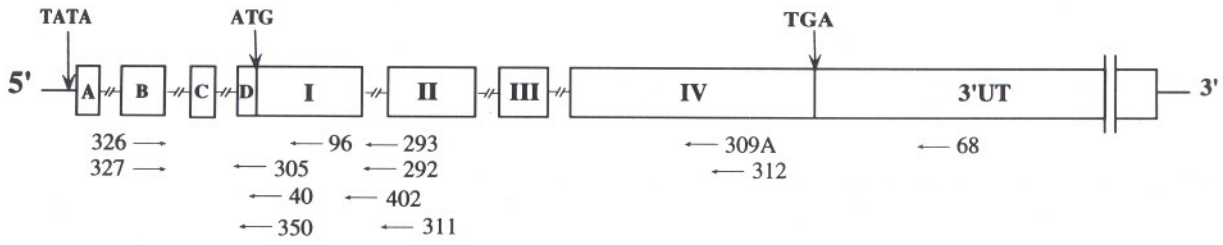

Figure 1. Structural organization of the SP-A gene and location of the gene- and allele-specific oligonucleotides. The gene-specific oligonucleotides for SP-A1 are 326 and 293, and for SP-A2 are 327 and 292. Oligonucleotide 68 is common to both genes. The remaining oligonucleotides are used as probes in hybridizations to determine the identity of each allele (*see* Table 1).

TABLE 1

Sequence, orientation, nucleotide position, gene location, and specificity of oligonucleotides used in the present study

Oligo Number	Sequence	Orientation	Nucleotide Position*	Gene Location	Specificity
40	5′-GATGAGGTTGAGGGCCAGAG-3′	AS	1059–1078	Exon I	$1A^0$, $1A^3$, SP-A1
68	5′-TGCCACAGAGACCTCAGAGT-3′	AS	3845–3864	3′UT	SP-A1, SP-A2
96	5′-TCCTTTGACACCATCTC-3′	AS	1185–1201	Exon I	$6A^2$, SP-A2
292	5′-CCATTATTCCCAGGAGGACATGGTG-3′	AS	1555–1579	Exon II	SP-A2
293	5′-CCATCATTTCCAGGAGGACATGGCA-3′	AS	1555–1579	Exon II	SP-A1
305	5′-CTTCGCACGCAGCACCA-3′	AS	1093–1109	Exon I	6A, SP-A2
309A	5′-TTTTCCCCAACCTGCGGA-3′	AS	3233–3249	Exon IV	$6A^4$
311	5′-CACGCCAGGCTCCCCCT-3′	AS	1641–1657	Exon II	1A, SP-A1
312	5′-CACACACTTCTCTTTTC-3′	AS	3246–3261	Exon IV	$1A^1$, $1A^3$
326	5′-ACTCCATGACTGACCACCTT-3′	S	469–488	5′UT	SP-A1
327	5′-ATCACTGACTGTGAGAGGGT-3′	S	472–491	5′UT	SP-A2
350	5′-GATGAGGGTGAGGGCCAGAG-3′	AS	1059–1078	Exon I	1A, $1A^1$, $1A^2$
402	5′-TTTCTCCAGGTGGACCCATG-3′	AS	1536–1555	Exon II	$6A^2$, $6A^3$

* Katyal and colleagues (3). AS = antisense; S = sense.

Hybridizations. The conditions for gene-specific and allele-specific hybridizations were determined empirically using cloned cDNAs. The filters are prehybridized for 1 h at 65°C in 5× Denhardts and 100 μg/ml denatured fish sperm DNA, and then hybridized at the appropriate temperature for 2 h in a solution containing 1× Denhardts, 60 mM Tris, pH 8.0, 300 mM NaCl, 2 mM EDTA, 1% sodium dodecyl sulfate (SDS), 100 μg/ml denatured fish sperm DNA, and 5×10^5 counts per minute (cpm)/ml labeled oligonucleotide. The filters are first washed 2 × 5′ at room temperature in solution I (60 mM Tris, pH 8.0, 300 mM NaCl, and 2 mM EDTA), and then 1 × 60′ at the appropriate temperature in solution II (60 mM Tris, pH 8.0, 300 mM NaCl, 2 mM EDTA, and 0.1% SDS). The hybridization and wash temperatures, respectively, in solution II for each oligonucleotide are as follows: for oligonucleotide 293 (SP-A1 specific), 50°C and 55°C; for allele-specific oligonucleotides 96 and 309A, 45°C and 45°C; for 305, 45°C and 55°C; and for 402, 53°C and 55°C. All oligonucleotide probes are end-labeled using [γ^{32}P-ATP] (3,000 Ci/mmol). The characteristics of the oligonucleotide probes are shown in Table 1 and Figure 1 (sequence and location) and in Table 2 (hybridization pattern). Membranes are stripped in wash solution II at 75°C for 1 to 2 h, exposed to X-ray film to assess completion of stripping, and then rehybridized with additional oligonucleotides.

Scoring of alleles. The genotype of each sample is determined by the characteristic pattern of oligonucleotide hybridizations shown in Table 2, and in certain cases the intensity of the band in the autoradiogram is assessed to determine the homozygous versus the heterozygous status of the particular sample. Specifically, samples were scored positive or negative for all allele-specific hybridizations (oligonucleotides 96, 305, 402, and 309A for SP-A1; and oligonucleotides 40, 311, 312, and 350 for SP-A2). For some genotypes, this scoring provided sufficient information to assign genotypes by a process of elimination (Table 2). For example, a sample positive with only oligonucleotide 40 or oligonucleotide 350

TABLE 2

Characteristic oligonucleotide hybridization patterns for SP-A1 and SP-A2 alleles

		SP-A2 Alleles					
Oligo Number	Codon	1A	$1A^0$	$1A^1$	$1A^2$	$1A^3$	SP-A1
40	9	−	+	−	−	+	+
311	91	+	−	−	−	−	+
312	223	−	−	+	−	+	−
350	9	+	−	+	+	−	−
		SP-A1 Alleles					
Oligo Number	Codon	6A	$6A^2$	$6A^3$	$6A^4$	SP-A2	
96	50	−	+	−	−	+	
305	19	+	−	−	−	+	
309A	219	−	−	−	+	−	
402	62	−	+	+	−	−	

+ = hybridization; − = absence of hybridization.

would have the $1A^0 1A^0$ or $1A^2 1A^2$ genotype, respectively. Because each individual is limited to two alleles, a sample that was positive with oligos 40, 311, and 350 would have the $1A^0 1A$ genotype. However, for some samples (for both SP-A1 and SP-A2) this simple process of elimination could not be used. In these cases it was necessary to determine the homozygosity or heterozygosity of the sample. For example, a sample positive with oligonucleotides 311 and 350 could have the 1A1A or $1A1A^2$ genotype. Determination of heterozygosity in this case was made by comparing the 311 hybridization intensity (relative to other samples) with their relative loads (based on hybridization to a common oligonucleotide; *see above* slot-blot hybridization). By comparing their relative loads we could assess whether the oligonucleotide 311 hybridization intensity represented one or two copies of the 1A allele. An additional or alternative and confirming method was used where we compared the intensity of the hybridization signal with the signal of other known heterozygous samples on the same blot. For instance, comparison with a previously determined $1A^0 1A$ sample allowed us to visualize the intensity represented by one copy of 1A. This method of scoring was verified in a few cases with cloning and sequencing of the scored alleles.

PCR and Hybridization Conditions for the SP-A2 Gene and SP-A2 Alleles

For amplification of SP-A2 we use oligonucleotides 327 (SP-A2 specific) and 68 (100 ng each). The PCR conditions are the same as those described above for the SP-A1 gene. Slot-blot preparation and allele-specific hybridization are also the same as those described for the SP-A1 alleles, except that the temperatures for the hybridization and for washes in solution II, for each oligonucleotide, are: for 292 (SP-A2 specific), 50°C and 60°C; for the allele-specific oligonucleotides 311, 52°C and 53.5°C; for 312, 40°C and 40°C; and for 40 and 350, 55°C and 60°C. For oligonucleotide 350 we found that 62°C wash temperature also maintains specificity, and are currently using this higher stringency condition.

Heterozygosity and PIC Values: Linkage Disequilibrium

The observed heterozygosity (h) value is the fraction of heterozygotes in the total population. The expected heterozygosity (Hu) is based on the expected distribution of different alleles in a given population assuming random mating. The 95% confidence intervals circumscribing the observed heterozygosity are determined from the binomial distribution using the computer program Binom (15). The polymorphism information content (PIC) value is determined by the PIC computer program based on counts of alleles at each locus (15). All of the above values for each SP-A locus are shown in RESULTS (Tables 3 and 4).

Tests of linkage disequilibrium are performed using the computer program EH (16), which is designed to test and estimate linkage disequilibrium between alleles of different marker loci. In the present study, we first test whether the alleles of each SP-A locus (gene) are in linkage disequilibrium. Then we identify, in 239 samples, SP-A1/SP-A2 haplotypes by χ^2 analysis (with one degree of freedom). This analysis establishes which alleles of the SP-A1 locus are in linkage disequilibrium with corresponding alleles of the SP-A2 locus. The haplotypes and the associated probability against linkage disequilibrium ($P < 0.01$) are shown in Table 5.

TABLE 3

*Heterozygosity and PIC values for the SP-A1 locus**

		Heterozygosity			
Group	*n*	h	CI	Hu	PIC
Random	53	0.51	0.37–0.65	0.62	0.55
RDS	62	0.64	0.51–0.76	0.60	0.54
Preterm	60	0.60	0.46–0.72	0.61	0.54
Nigerian	40	0.67	0.51–0.81	0.64	0.57
Mixed	48	0.62	0.47–0.76	0.61	0.55
All	263	0.63	0.57–0.69	0.61	0.55

* *n* = number of individuals in each group; h = observed heterozygosity. The h value is the fraction of heterozygotes in the total population; Hu = expected (unbiased) heterozygosity. The Hu value is based on the expected distribution of different alleles in a given population assuming random segregation; CI = confidence intervals. The 95% CI of the h value were determined from the binomial distribution using the computer program Bionom; PIC = polymorphism information content. PIC was determined by the PIC computer program based on counts of alleles at each locus.

Results

Background, Strategy, and Initial Observations

Based on the availability of published complete coding sequence (3, 4, 13, 14) and recently described alleles (6), six allelic variants of the SP-A1 gene (6A, $6A^0$, $6A^1$, $6A^2$, $6A^3$, and $6A^4$), and four allelic variants of the SP-A2 gene (1A, $1A^0$, $1A^1$, and $1A^2$) have been described to date (6, 7). Although the DNA sequences of the alleles are very similar to one another, analysis of the sequence of these variants as well as of the published genomic sequences (3, 4) allowed identification of both gene- and allele-specific differences. We designed gene- and allele-specific oligonucleotides based on these sequence differences. These oligonucleotides are used as primers in the PCR to amplify gene-specific segments, and/or as probes in slot-blot hybridizations as additional verification that each SP-A1 or SP-A2 gene is specifically amplified. In addition, certain oligonucleotides are used as probes to identify each allele of each SP-A gene. The characteristics of these oligonucleotides are shown in Table 1, and their location on the SP-A gene is shown diagramatically in Figure 1. Oligonucleotides 293 and 326 are SP-A1 specific and oligonucleotides 292 and 327 are SP-A2 specific. Oligonucleotide 68 is common to both SP-A genes. The remaining oligonucleotides recognize specific alleles and the charac-

TABLE 4

*Heterozygosity and PIC values for the SP-A2 locus**

		Heterozygosity			
Group	*n*	h	CI	Hu	PIC
Random	48	0.50	0.35–0.65	0.53	0.48
RDS	68	0.55	0.43–0.68	0.62	0.57
Preterm	56	0.57	0.43–0.70	0.67	0.62
Nigerian	29	0.34	0.18–0.54	0.65	0.57
Mixed	44	0.45	0.30–0.61	0.54	0.50
All	245	0.50	0.44–0.57	0.61	0.56

* Abbreviations as in Table 3.

teristic hybridization pattern for each allele is shown in Table 2. For example, oligonucleotide 40 distinguishes differences at codon 9, and hybridizes only to SP-A2 alleles $1A^0$ and $1A^3$ and to all SP-A1 alleles. Oligonucleotide 311 detects only 1A, and all SP-A1 alleles. Oligonucleotide 312 detects only $1A^1$ and $1A^3$ alleles and oligonucleotide 350 detects only 1A, $1A^1$, and $1A^2$ alleles. The corresponding hybridization patterns for each SP-A1 allele are also shown in Table 2.

In our initial studies, when allele-specific oligonucleotides were used as hybridization probes with cloned·SP-A genomic DNA, we observed that two of the previously characterized SP-A1 alleles, namely $6A^0$ and $6A^1$ (6), did not show the expected results. Reanalysis of these DNA sequences showed that the $6A^0$ allele is identical to $6A^3$, and $6A^1$ is identical to $6A^2$ (6). Codons 45 and 54 of allele $6A^0$ are previously reported (4) to be CAC and CTG, respectively; however, the present study found them to be GAC and CCT, (the sequence of the $6A^3$ allele). Similarly, codon 4 of the $6A^1$ allele is reported (14) as TGT but was found to be TGC in the present study (the sequence of the $6A^2$ allele). With these findings the SP-A1 alleles are: 6A, $6A^2$, $6A^3$, and $6A^4$. When genomic DNAs were amplified with gene-specific oligonucleotides and used in slot-blot analysis with allele-specific oligonucleotide probes, a few of the DNA samples showed hybridization patterns that are not consistent with the known alleles. This observation suggests the presence of one additional, previously unknown allele. The new allele, now named $1A^3$, hybridizes with both oligonucleotides 40 (codon 9) and 312 (codon 223), as shown in Table 2. This allele is cloned and the hybridization pattern is confirmed by sequencing. The coding region of allele $1A^3$ also contains a nucleotide change for codon 140; TCC is changed to TCT. The change in codon 140 is confirmed by sequencing in five individuals, four of whom were members of the same three generation family (paternal grandfather, father, mother, and child), and three of the five individuals were unrelated (father and mother from the above family and a third individual, who was not part of the family). Allele $1A^3$ occurs with low frequency in the population (*see below*). In addition, we note here, as assessed by both sequencing and oligonucleotide hybridization that codon 140 of $1A^1$ allele is also TCT.

Gene-specific Amplifications

The conditions for amplifying each SP-A locus are established using cloned SP-A1 and SP-A2 genomic DNA as a template and oligonucleotide 68 and 326 (SP-A1 specific), or oligonucleotides 68 and 327 (SP-A2 specific), as described in MATERIALS AND METHODS. An example of these amplifications is depicted in Figure 2A. Lanes 1 and 2 depict the PCR products when cloned SP-A2 genomic DNA is used as template with oligonucleotides 327/68 (positive control) and 326/68 (negative control). The PCR products with cloned SP-A1 genomic DNA as template are shown in lanes 3 (positive control) and 4 (negative control). When genomic DNA is used as template, a PCR product (lanes 5 and 6) is obtained with size similar to that obtained from cloned templates (lanes 1 and 3). To verify that the amplified products shown in lanes 5 and 6 are specific for the SP-A1 and SP-A2 genes, respectively, we hybridized each amplified product with gene-specific oligonucleotides (Figure 2B). The slots in the top line (*asterisks*) of both blots contain products similar to those shown in Figure 2A, lane 5 (i.e., PCR products specific to SP-A1), and the slots in the bottom line (*closed triangles*) contain products similar to those in lane 6, Figure 2A (i.e., PCR products specific to the SP-A2 gene). The PCR products in the top line (*asterisks*) hybridize only with oligonucleotide 293 which detects SP-A1 sequences, whereas the products in the bottom line (*closed triangles*) hybridize only to oligonucleotide 292. These results indicate that the amplification conditions used in Figure 2A are appropriate for the amplification of each SP-A gene from genomic DNA.

Allele-specific Hybridization and Allele Frequencies

PCR products from genomic DNA of individuals in the study populations are placed on slot blots (as described in MATERIALS AND METHODS), along with control cloned allele-specific cDNAs. The blots are then hybridized with various ^{32}P-labeled oligonucleotides, and the hybridization pattern (Table 2) is used to determine the genotype of each genomic DNA. The signal intensity varied among the 10 specific oligonucleotides regardless of the specific activity of the hybridizing oligonucleotides. Examples of control hybridizations specific for SP-A1 and SP-A2 alleles are shown in Figures 3 and 4, respectively. For example, as indicated diagramatically in Table 2 and shown in Figure 3, oligonucleotides 305 and 96 hybridize to 6A and $6A^2$ alleles, respectively, and to SP-A2 alleles, whereas oligonucleotide 402 hybridizes to both $6A^2$ and $6A^3$ sequences and oligonucleotide 309A detects only $6A^4$ sequences. Figure 4 shows the specificity of the corresponding hybridizations for the SP-A2 alleles. The SP-A2 control DNA in Figure 3 and the SP-A1 control DNA in Figure 4 are the original 1A and 6A cDNAs (13). The frequency of the alleles for the SP-A1 and SP-A2 genes varies, with some occurring considerably more frequently than others. The specific frequencies for the alleles of the SP-A1 gene in the study population (n = 526 alleles) are: 6A = 9.3%, $6A^2$ = 54.4%, $6A^3$ = 27.75%, and $6A^4$ = 8.5%; and for the alleles of the SP-A2 gene in the study population (n = 490 alleles) are: 1A = 12%, $1A^0$ = 56.5%, $1A^1$ = 21.6%, $1A^2$ = 9%, and $1A^3$ = 0.8%.

Evaluation of the Degree of Polymorphism of Each SP-A Locus

Marker loci have played a central role in numerous genetic studies of disease. The availability of highly informative markers is key to identifying the genetic locus associated with a specific disease or to identifying a subgroup of patients with the disease. The usefulness of a marker in genetic studies varies according to the marker's degree of informativeness. By definition, the most useful genetic markers are those that display a high degree of polymorphism (i.e., a large number of alleles for a given marker locus). To determine the degree of polymorphism at the SP-A gene loci, and thus assess their informativeness, we determined the heterozygosity and PIC values (*see* MATERIALS AND METHODS) in a large diverse group of individuals and in smaller subgroups of individuals. To further assess whether the SP-A alleles segregate randomly in the study populations (Hardy–Wein-

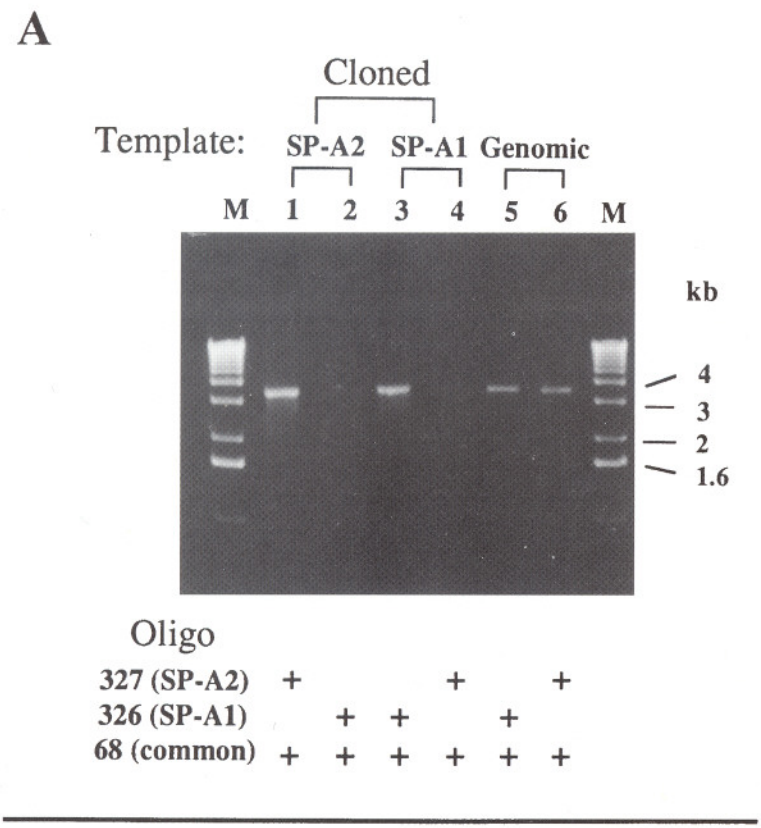

Figure 2. Gene-specific amplifications and hybridizations. *Panel A* depicts the 327/68 (SP-A2–specific) and the 326/68 (SP-A1–specific) amplifications using, as template, cloned SP-A1 and SP-A2 DNA (lanes 1 through 4) and genomic DNA (lanes 5 and 6). The sequences of oligonucleotides 327 and 326 are specific for genes SP-A2 and SP-A1, respectively. The sequence of oligonucleotide 68 is common to both SP-A genes. SP-A2–specific PCR products are shown in lanes 1 and 6 and SP-A1–specific PCR products in lanes 3 and 5. The specificity of these amplifications is further verified by the lack of PCR products in lanes 2 and 4. The SP-A1–specific oligonucleotide pair (326/68) did not amplify the cloned SP-A2 gene (lane 2) and the SP-A2 specific oligonucleotide pair (327/68) did not amplify the cloned SP-A1 gene (lane 4). M denotes 1-kb ladder markers (Gibco-BRL, Grand Island, NY). *Panel B* depicts hybridizations of SP-A1– and SP-A2–specific PCR products with oligonucleotides that hybridize to SP-A1 (293) or SP-A2 (292) sequences, further verifying the gene-specific amplifications. The slots in the top line (*asterisks*) in both blots contain SP-A1 PCR products similar to those shown in lane 5, *panel A*; the slots in the bottom line (*closed triangles*) contain SP-A2 PCR products similar to those shown in lane 6, *panel A*.

berg equilibrium), we compared the observed heterozygosities in the populations with the expected heterozygosities. The latter depends on the expected distribution of different alleles in a given population assuming random segregation of alleles.

For the SP-A1 locus, the observed (h) and expected (Hu) heterozygosities for the entire population (n = 253 individuals) are 0.63 and 0.61, respectively (Table 3). At the SP-A2 locus, h = 0.50 and Hu = 0.61 (n = 245 individuals) (Table 4). For the SP-A1 locus, the observed and expected heterozygosities of the alleles in both mixed and individual subpopulations are statistically indistinguishable, since the expected heterozygosity falls within the confidence interval of the corresponding observed value (Table 3). Similar findings are observed for the SP-A2 locus (Table 4), except in one of the subpopulations (Nigerian), which had the lowest number of individuals (n = 29). The expected heterozygosity value (Hu = 0.61) for all groups combined does not fall within the confidence intervals (0.44 to 0.57) of the h value. This observation may reflect the findings of the Nigerian population. We recalculated the h and the Hu values for the combined group without including the Nigerian population. In this circumstance (n = 216 individuals) we found the Hu value (0.60) to be within the confidence intervals (0.46 to 0.60) of the h value (0.53), suggesting that the findings of the Nigerian population influenced the outcome in the combined group. The PIC value remains the same. These observations suggest that the SP-A marker loci are in Hardy–Weinberg equilibrium. The heterozygosity values for the SP-A1 (Table 3) and SP-A2 (Table 4) marker loci indicate that each marker

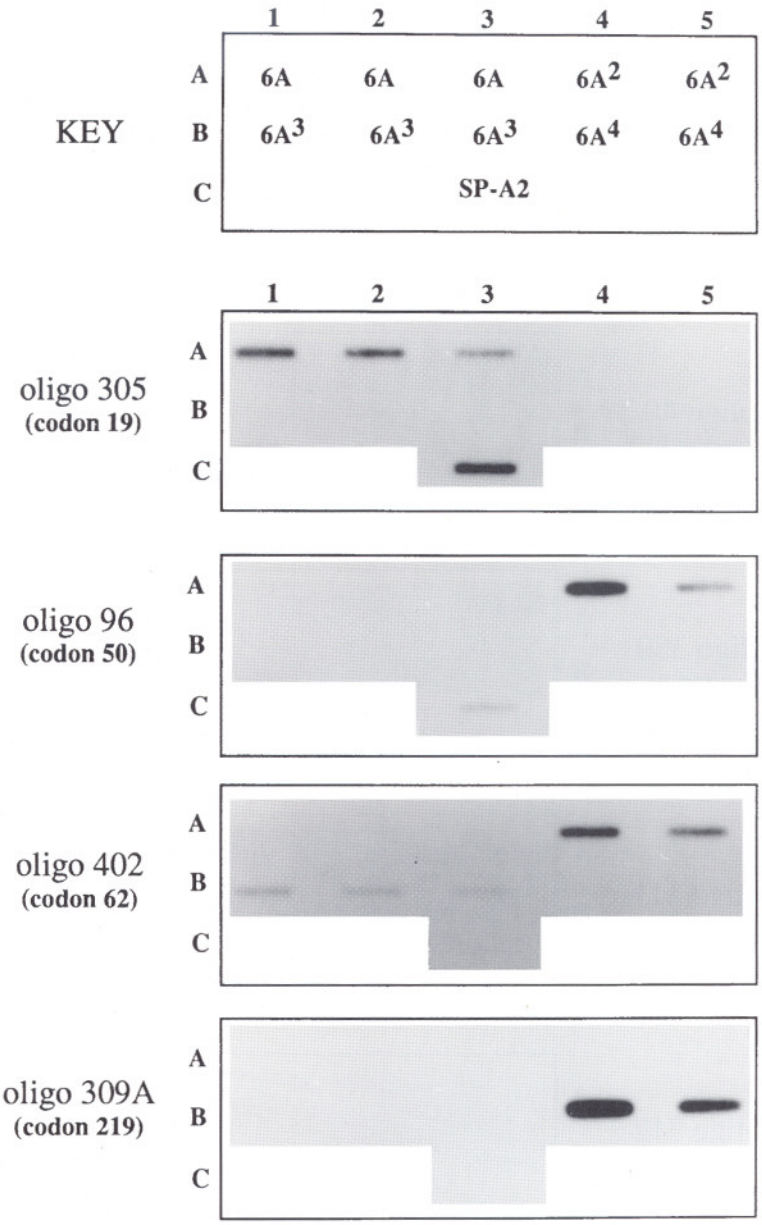

Figure 3. SP-A1-allele–specific hybridizations. Cloned DNA for each SP-A1 allele and from one SP-A2 allele (1A) was placed on a slot blot as shown in the key. Specific oligonucleotides were used as probes. The pattern of hybridization for each oligonucleotide is shown diagramatically in Table 2 and the actual hybridizations are shown here.

locus is sufficiently informative for use in disease-related genetic studies.

Linkage Disequilibrium

The SP-A locus has been previously localized by chromosomal *in situ* hybridization to chromosome 10, bands 10q21 to q24 (2). However, the physical distance between the two SP-A genes is currently unknown. To determine whether the SP-A genes are close to one another on the chromosome, we performed a test for linkage disequilibrium of SP-A1 and SP-A2 genes. Evidence of linkage disequilibrium between two genes suggests that the genes lie in proximity to one another (i.e., < 2 million bases apart) (17, 18). We used the linkage EH program to test for linkage disequilibrium between SP-A1 and SP-A2. The probability that the SP-A1 and SP-A2 loci are unlinked is very low ($P < 0.000001$), indicating that the two loci are in close physical proximity. If two genes are in linkage disequilibrium, then alleles at the two loci will cosegregate as a unit (haplotype). Of the 20 possible haplotypes (4 alleles of SP-A1 × 5 alleles for SP-A2 = 20 permutations) we identified 19 haplotypes in the study population; nine haplotypes had P values < 0.01 (Table 5), whereas the remaining haplotypes did not reach statistical significance in a population of this size. The only haplotype that was not observed is $6A^4/1A^3$. The data from the three generation family pedigree (Figure 5) indicate that segregation of haplotypes ($6A^2/1A^0$ and $6A^4/1A$) occurs without recombination, supporting the linkage analysis data

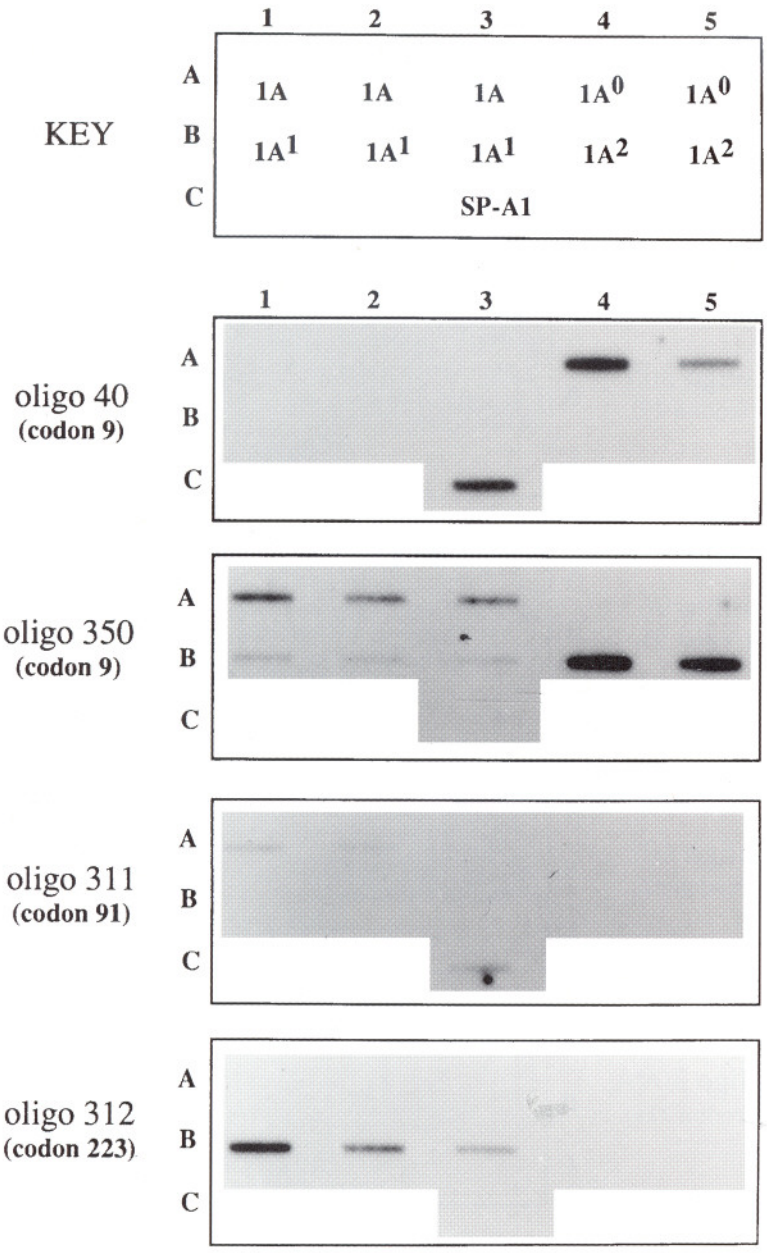

Figure 4. SP-A2-allele-specific hybridizations. Cloned DNA for each SP-A2 allele and from one SP-A1 allele (6A) was placed on a slot blot as shown. Specific oligonucleotides were used as probes. The pattern of hybridization for each oligonucleotide is shown schematically in Table 2 and the actual hybridizations are shown here.

obtained from unrelated individuals. In addition, the information obtained from these data indicates that these alleles are meiotically stable.

Discussion

Recently, a number of reports have shown that the human SP-A locus is highly polymorphic (6, 8). Since highly polymorphic loci (i.e., highly informative marker loci) have played a central role in numerous genetic studies of disease, we wished to evaluate the SP-A loci for their use as markers in genetic studies. We have previously observed that the levels of SP-A mRNA vary significantly among humans, whereas the levels of SP-C mRNA are considerably less variable (19). The SP-C locus was found subsequently to be quite conserved among humans (20). We became intrigued by these observations and hope to investigate correlations between genotype and the level of expression of SP-A alleles. As a prelude to the above goals, in this report we (*1*) describe the experimental conditions that allow us to distinguish among similar SP-A sequences; (*2*) determine the informativeness of the SP-A loci; and (*3*) show that the SP-A1 and SP-A2 loci are in linkage disequilibrium. We also identify 19 haplotypes, nine of which reach statistical significance in this size population ($n = 239$). Furthermore, we identify one additional allele for the SP-A2 gene.

Genetic markers can be the key that allows the genetic complexities of a given disease or population heterogeneity to be revealed. For example, mutations or polymorphisms

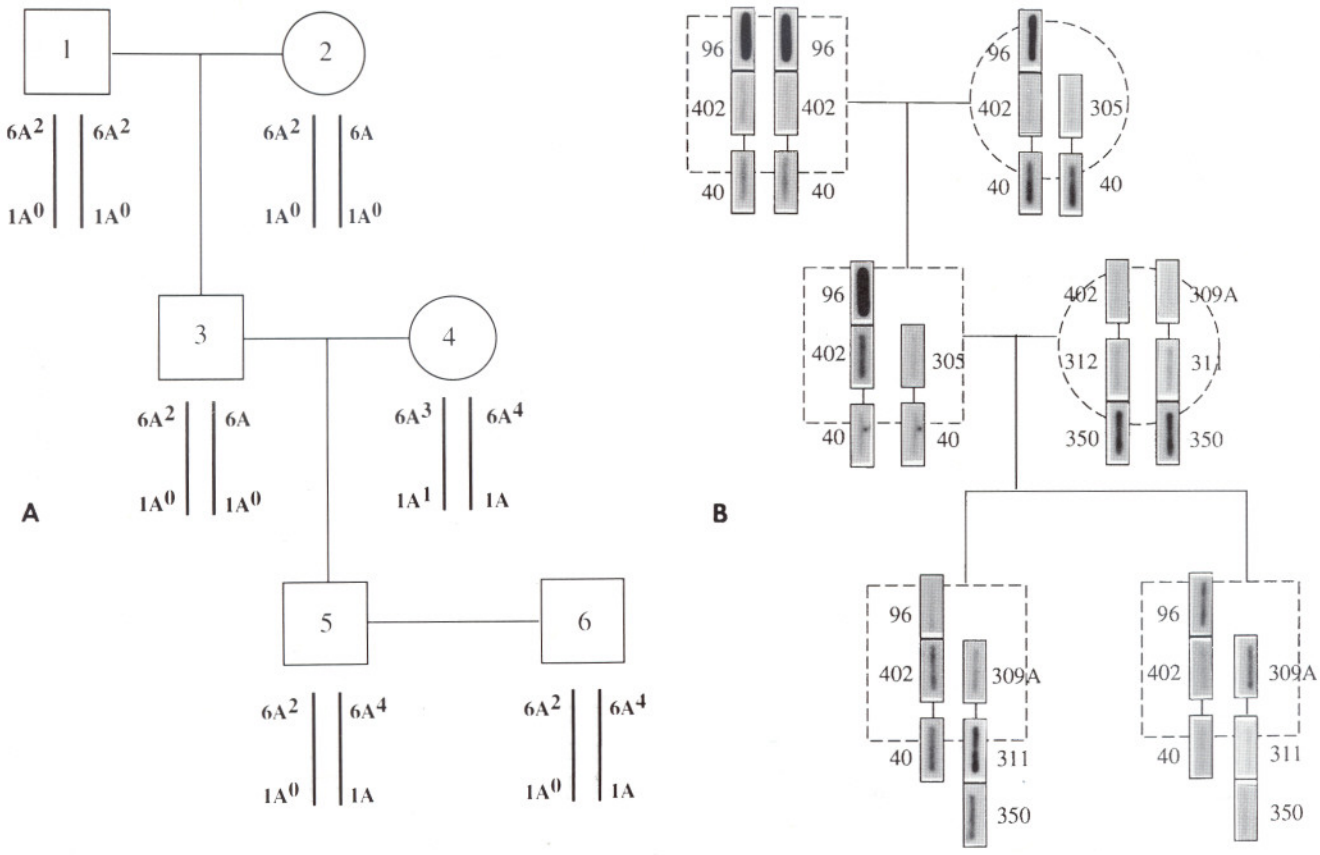

Figure 5. Segregation of linked polymorphisms at the SP-A1 and SP-A2 loci. Coupled polymorphisms of the two SP-A loci (haplotypes) are shown below each individual in *panel A*. Actual hybridizations of oligonucleotides that detect specific allelic SP-A1 or SP-A2 sequences are shown in *panel B*. The number of the particular oligonucleotide used as probe is shown next to the hybridizing slot. For example, individual 1 in *panel A* is shown to have a haplotype $6A^2/1A^0$. For this individual to have such a haplotype, PCR products from genomic DNA from this individual must hybridize to oligonucleotide probes 96 and 402 ($6A^2$), and to 40 ($1A^0$). The actual hybridization pattern is shown in *panel B* and the characteristic hybridization pattern that identifies each SP-A1 and each SP-A2 allele is shown in Table 2. The specificity of hybridization by these oligonucleotides is depicted in Figures 3 and 4.

have been shown to associate with the severity of a disease in a number of diseases [e.g., phenylketonuria (21) and cystic fibrosis (22)], as well as with diseases caused by a single point mutation [e.g., sickle cell anemia (23)]. The usefulness of a marker locus varies according to the degree of its informativeness as assessed by its heterozygosity and PIC indices.

In the present study, we assessed the degree of polymorphism of the SP-A loci by determining heterozygosity in a large diverse population (h = 0.63, PIC 0.55 for SP-A1 and h = 0.50, PIC 0.56 for SP-A2) and in small subpopulations (Tables 3 and 4). The findings from these analyses indicate that both SP-A loci are highly polymorphic, and thus would be appropriate for use in genetic studies. As a point of reference, the average PIC value of restriction fragment-length polymorphisms, which have been used successfully in genetic association studies, is about 0.3 (24). When the major subpopulations ($n > 29$) were evaluated further, we found the confidence intervals of the observed heterozygosity of one subpopulation to overlap the confidence intervals of the other, indicating that there is no significant difference in the degree of polymorphism among the various populations (i.e., each SP-A marker locus is equally polymorphic within the various populations). This observation validates the appropriateness of these marker loci for use in studies where one wishes to evaluate associations between this locus and other loci (be they a disease locus or other marker loci) in

TABLE 5

Haplotypes of the SP-A locus

Haplotypes	*P* Value
6A/1A	0.000002
$6A^2/1A$	0.008750
$6A^2/1A^0$	0.000069
$6A^2/1A^1$	0.001500
$6A^3/1A^0$	0.003759
$6A^3/1A^1$	0.000000
$6A^4/1A$	0.000030
$6A^4/1A^0$*	0.002372
$6A^4/1A^2$	0.000003

* The presence of this haplotype cannot be made with great certainty because the number of observations was < 5 in the population under study (n = 239 individuals). The *P* value indicates the probability that the particular alleles are unlinked.

a variety of populations. In such studies, if the frequency of a particular genotype differs between the case and control populations under study, it would suggest that the specific genotype contributes to, or is linked to, differences between those groups. In addition, the expected and observed heterozygosity values for each study population (Tables 3 and 4) are indistinguishable from one another (both fall within the confidence interval of the observed value). This observation indicates that the SP-A alleles segregate randomly in these populations (i.e., are in Hardy–Weinberg equilibrium). However, in one of the subgroups (Nigerian), the expected heterozygosity value for the SP-A2 locus was not found within the confidence interval of the corresponding observed value. This observation may reflect the small number of specimens analyzed (Nigerian, $n = 29$) or may represent selection for, or against, one of the alleles in these population. The data from a recent study (25) suggest that the above observation is the result of the small number of Nigerian specimens analyzed, rather than the result of a genetically isolated population. The distribution of SP-B alleles between African Americans and Nigerians was found to be similar between these two populations, although differences were observed between Caucasians and African Americans, or between Caucasians and Nigerians.

Linkage disequilibrium analysis indicates that the two SP-A loci are in linkage disequilibrium, and thus cosegregate as a unit (haplotype). This finding means that the resolution of the SP-A locus, when used as a marker for genetic studies, is increased. If the SP-A1 and SP-A2 genes were not linked, each would be used independently in linkage analysis. Because they are linked they can also be used as a unit (haplotype) in linkage analysis. Their combined polymorphism results in 20 haplotypes (4 alleles of SP-A1 $\times$ 5 alleles of SP-A2), thus increasing the resolution. Therefore, identification of SP-A1/SP-A2 haplotypes present in the population can increase the informativeness of the SP-A locus as a marker in studies designed to examine potential associations between SP-A genotype and pulmonary phenotype, since alleles at both loci can be scored.

In summary, we show a high degree of polymorphism at the SP-A loci, making these loci good markers for genetic studies in a variety of populations. We further show that the two loci are in linkage disequilibrium, which in turn increases their sensitivity for detection of differences between populations (be they, for example, genetic heterogeneity of disease or of populations). We would also like to make the following additional observations about the SP-A alleles. (*1*) The finding of one novel allele suggests that other additional yet unknown SP-A alleles are yet to be found. (*2*) Though the classification of the SP-A alleles is based upon changes found within the coding sequence, additional polymorphisms may exist within noncoding sequences (3, 26) that divide "coding" alleles into subclasses with a different phenotype. This "noncoding" class of polymorphism may be detected if individuals of a known single genotype fall into distinct phenotypic subgroups. In that circumstance, subsequent investigation of noncoding regions may be required to identify the genetic basis of the distinct subgroups.

Acknowledgments: The writers thank Drs. D. S. Phelps, M. S. Dunn, B. A. Nagourney, and S. A. Olowe; Andreas and Nikos Floros Phelps; Don and Phyllis Phelps; Mr. and Mrs. Weaver; and Margie Welsh for their contributions; and Beth Ditzler for typing of the manuscript. This work was supported by Grant HL49823 from the National Institutes of Health.

References

1. Possmayer, F. 1988. Pulmonary perspective: a proposed nomenclature for pulmonary surfactant-associated proteins. *Am. Rev. Respir. Dis.* 138: 990–998.
2. Bruns, G., H. Stroh, G. M. Veldman, S. A. Latt, and J. Floros. 1987. The 35 kd pulmonary surfactant-associated protein is encoded on chromosome 10. *Hum. Genet.* 76:58–62.
3. Katyal, S. L., G. Singh, and J. Locker. 1992. Characterization of a second human pulmonary surfactant-associated protein SP-A gene. *Am. J. Respir. Cell Mol. Biol.* 6:446–452.
4. White, R. T., D. Damm, J. Miller, K. Spratt, J. Schilling, S. Hawgood, B. Benson, and B. Cordell. 1985. Isolation and characterization of the human pulmonary surfactant apoprotein gene. *Nature* 317:361–363.
5. Korfhagen, T. R., S. W. Glasser, M. D. Bruno, M. J. McMahan, and J. A. Whitsett. 1991. A portion of the human surfactant protein A (SP-A) gene locus consists of a pseudogene. *Am. J. Respir. Cell Mol. Biol.* 4:463–469.
6. Karinch, A. M., and J. Floros. 1995. 5′ splicing and allelic variants of the human pulmonary surfactant protein A genes. *Am. J. Respir. Cell Mol. Biol.* 12:77–88.
7. Floros, J., and A. M. Karinch. 1995. Human SP-A: then and now. *Am. J. Physiol.* 268(*Lung Cell. Mol. Physiol.* 12):L162–L165.
8. McCormick, S. M., V. Boggaram, and C. R. Mendelson. 1994. Characterization of mRNA transcripts and organization of human SP-A1 and SP-A2 genes. *Am. J. Physiol.* 266(*Lung Cell. Mol. Physiol.* 10):L354–L366.
9. Floros, J., D. S. Phelps, and H. W. Taeusch. (1985) Biosynthesis and *in vitro* translation of the major surfactant-associated protein from human lung. *J. Biol. Chem.* 260:495–500.
10. Phelps, D. S., J. Floros, and H. W. Taeusch, Jr. 1986. Post-translational modification of the major human surfactant-associated proteins. *Biochem. J.* 237:373–377.
11. Voss, T., K. Melchers, G. Scheirle, and K. P. Schafer. 1991. Structural comparison of recombinant pulmonary surfactant protein SP-A derived from two human coding sequences: implications for the chain composition of natural human SP-A. *Am. J. Respir. Cell Mol. Biol.* 4:88–94.
12. Haagsman, H. P., R. T. White, J. Schilling, K. Lau, B. J. Benson, J. Golden, S. Hawgood, and J. A. Clemens. 1989. Studies of the structure of lung surfactant protein SP-A. *Am. J. Physiol.* 257(*Lung Cell Mol. Physiol.* 1):L421–L429.
13. Floros, J., R. Steinbrink, K. Jacobs, D. Phelps, R. Kriz, L. Sultzman, M. Recny, S. Jones, H. W. Taeusch, H. A. Frank, and E. F. Fritsch. 1986. Isolation and characterization of cDNA clones for the 35kDa pulmonary surfactant associated protein (PSP-A). *J. Biol. Chem.* 261:9029–9033.
14. Rishi, A., D. Hatzis, K. McAlmon, and J. Floros. 1992. An allelic variant of the 6A gene for human surfactant protein A. *Am. J. Physiol.* 262(*Lung Cell. Mol. Physiol.* 6):L566–L573.
15. Ott, J. 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. *Am. J. Hum. Genet.* 51:283–290.
16. Terwillinger, J. D., and J. Ott. 1994. Handbook of Human Genetic Linkage. The Johns Hopkins University Press. Baltimore, MD. 188–210.
17. Beaudet, A. L., G. L. Feldman, S. D. Fernbach, G. J. Buffone, and W. E. O'Brien. 1989. Linkage disequilibrium, cystic fibrosis and genetic counseling. *Am. J. Hum. Genet.* 44:319–326.
18. Botstein, D., R. L. White, M. H. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–331.
19. Floros, J., D. S. Phelps, D. E. deMello, J. Longmate, H. Harding, B. Benson, and T. White. 1991. The utility of postmortem lung for RNA studies: variability and correlation of the expression of surfactant proteins in human lung. *Exp. Lung Res.* 17:91–104.
20. Hatzis, D., G. Deiter, D. E. deMello, and J. Floros. 1994. Human surfactant protein-C: genetic homogeneity and expression in RDS; comparison with other species. *Exp. Lung Res.* 20:57–72.
21. Okano, Y., R. C. Eisensmith, F. Guttler, U. Lichter-Konecki, D. S. Koneck, F. K. Trefz, M. Dasovich, T. Wnag, K. Henricksen, H. Lou, and S. L. C. Woo. 1991. Molecular basis of pheontypic heterogeneity in phenylketonuria. *N. Engl. J. Med.* 324:1232–1238.
22. Collins, F. S. 1992. Cystic fibrosis: Molecular biology and therapeutic implications. *Science* 256:774–779.
23. Powars, D., and A. Hiti. 1993. Sickle cell anemia, b5 gene cluster haplotypes as genetic markers for severe disease expression. *Am. J. Dis. Child.* 147:1197–1202.
24. Weber, J. L. 1990. Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. *In* Genome Analysis, Vol. 1: Genetic and Physical Mapping. K. E. Davies and S. M. Tilghman, editors. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 159–181.
25. Veletza, S. V., P. K. Rogan, T. TenHave, S. A. Olowe, and J. Floros. 1996. Racial differences in allelic distribution at the human pulmonary surfactant protein B locus (SP-B). *Exp. Lung Res.* 22:489–494.
26. Krizkova, L., R. Sakthivel, S. A. Olowe, P. Rogan, and J. Floros. 1994. Human SP-A: genotype and single strand conformation polymorphism analysis. *Am. J. Physiol.* 266(*Lung Cell. Mol. Physiol.* 10):L519–L527.