Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 574 – 581

GENOMICS

www.elsevier.com/locate/ygeno

Distribution of human β-defensin polymorphisms in various control and cystic fibrosis populations

Anne Vankeerberghen[a,1], Olga Scudiero[b,1], Kris De Boeck[c], Milan Macek Jr.[d], Pier Franco Pignatti[e], Noémi Van Hul[a], Hilde Nuytten[a], Francesco Salvatore[b], Giuseppe Castaldo[b], Daniela Zemkova[d], Vera Vavrova[d], Jean-Jacques Cassiman[a], Harry Cuppens[a,*]

[a]*Department of Human Genetics, KULeuven, Herestraat 49, O&N6, 3000 Louvain, Belgium*
[b]*Department of Biochemistry and Medical Biotechnology–CEINGE, University of Naples Federico II, Naples, Italy*
[c]*Department of Pediatrics, UZ Gasthuisberg, Herestraat 49, 3000 Louvain, Belgium*
[d]*Institute of Biology and Medical Genetics and Department of Pediatrics, Charles University Prague and University Hospital Motol, Cystic Fibrosis Center, V Uvalu 84, CZ 15006 Prague, Czech Republic*
[e]*Section of Biology and Genetics, Department of Mother and Child, Biology and Genetics, University of Verona, Verona, Italy*

Received 12 December 2004; accepted 7 February 2005

Abstract

Human β defensins contribute to the first line of defense against infection of the lung. Polymorphisms in these genes are therefore potential modifiers of the severity of lung disease in cystic fibrosis. Polymorphisms were sought in the human β-defensin genes *DEFB1, DEFB4, DEFB103A*, and *DEFB104* in healthy individuals and cystic fibrosis (CF) patients living in various European countries. *DEFB1, DEFB4*, and *DEFB104* were very polymorphic, but *DEFB103A* was not. Within Europe, differences between control populations were found for some of the frequent polymorphisms in *DEFB1*, with significant differences between South-Italian and Czech populations. Moreover, frequent polymorphisms located in *DEFB4* and *DEFB104* were not in Hardy Weinberg equilibrium in all populations studied, while those in *DEFB1* were in Hardy Weinberg equilibrium. Sequencing of a monochromosomal chromosome 8 mouse–human hybrid cell line revealed signals for multiple alleles for some loci in *DEFB4* and *DEFB104*, but not for *DEFB1*. This indicated that more than one *DEFB4* and *DEFB104* gene was present on this chromosome 8, in agreement with recent findings that *DEFB4* and *DEFB104* are part of a repeat region. Individual *DEFB4* and *DEFB104* PCR amplification products of various samples were cloned and sequenced. The results showed that one DNA sample could contain more than two haplotypes, indicating that the various repeats on one chromosome were not identical. Given the higher complexity found in the genomic organization of the *DEFB4* and *DEFB104* genes, association studies with CF lung disease severity were performed only for frequent polymorphisms located in *DEFB1*. No association with the age of first infection by *Pseudomonas aeruginosa* or with the FEV1 percentage at the age of 11–13 years could be found.

Keywords: β-Defensin; SNP; Polymorphism; DNA repeat; Cystic fibrosis; Modulator

Defensins are small cationic antimicrobial peptides that contribute to host defense in multicellular organisms. Humans have two classes of defensins, named α and β, based on the arrangements of cysteines within their sequences. α Defensins are expressed mainly in cells of the immune system, like Paneth cells and neutrophils. The expression of β defensins is restricted mainly to epithelial cells that line the body and can be induced by microorganisms and inflammatory factors. Their function

* Corresponding author. Fax: +32 16 34 59 97.
E-mail address: harry.cuppens@med.kuleuven.ac.be (H. Cuppens).
[1] These authors equally contributed to the study.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.02.003

in the immune system is dual. First, due to their positive charge they can interact with and integrate into the membrane of pathogens, including bacteria, fungi, and enveloped viruses, to kill them. Second, they can attract cells of the acquired immune system. DEFB4 can even induce the maturation of dendritic cells and act as an adjuvant. More than 30 potential β-defensin genes have been identified in humans [2], but only 4 have been studied in detail, *DEFB1, DEFB4, DEFB103A*, and *DEFB104*. These genes are composed of two exons and are located on chromosome 8p22–p23. *DEFB4, DEFB103A*, and *DEFB104* are clustered together at a distance of ~600 kb from *DEFB1*. Although the peptides have similar structures, they seem to have specific antimicrobial spectra and induction pathways [3]. An upregulation of *DEFB4* and *DEFB103A* is often seen in inflamed tissues. High DEFB103A concentrations are found in psoriasis plaques and increased concentrations of DEFB4 peptides are found in plasma and bronchoalveolar lavage fluid of patients with diffuse panbronchiolitis [4]. DEFB1 and DEFB4 peptides are detected in the lungs [5,6].

Cystic fibrosis (CF) is caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, which encodes a chloride channel expressed in epithelial cells lining most exocrine glands. Defects in this protein are associated with recurrent lung infections and pancreatic insufficiency. The severity of lung disease does not always correlate well with the *CFTR* genotype. Other environmental and genetic factors affect the lung phenotype. Genes coding for mannose-binding lectin, tumor necrosis factor α, and transforming growth factor β_1 have been reported as possible modifier genes [7]. Since human β defensins form part of our first line of defense in the lung, they are potential modifier genes for cystic fibrosis lung disease. At the start of this study, however, only four polymorphic loci, all located in *DEFB1*, had been reported [8]. We therefore first sequenced the genomic region harboring the *DEFB1, DEFB4, DEFB103A*, and *DEFB104* genes in detail in individuals derived from various populations.

Frequent polymorphisms located in *DEFB4* and *DEFB104* were not in Hardy Weinberg equilibrium in all populations studied, while those in *DEFB1* were all in Hardy Weinberg equilibrium. To investigate this further, the β-defensin region of a monochromosomal chromosome 8 hybrid was sequenced. In addition, individual *DEFB4* and *DEFB104* PCR amplification products were cloned and sequenced. The results will confirm the results of Hollox et al. [1], that the *DEFB4, DEFB103A*, and *DEFB104* genes are part of a repeated genomic region.

Finally, association studies of the frequent polymorphisms located in *DEFB1* with the severity of the lung disease were performed in cystic fibrosis patients.

Results

Identification of nucleotide changes in DEFB1, DEFB4, DEFB103A, and DEFB104 exons and their exon/intron boundaries

The four β-defensin genes are each composed of two exons. To identify nucleotide changes in these genes, their exon and intron boundaries were amplified by PCR and sequenced. Initially, four populations were included in the study: a Belgian and South-Italian control population and a Belgian and South-Italian CF population composed only of patients homozygous for the F508del *CFTR* mutation. The β-defensin mutations occurring at a frequency higher than 5% are shown in Table 1. The nucleotides are numbered at the transcript level according to the published sequences *DEFB1*, GenBank U50930; *DEFB4*, GenBank AF040153; *DEFB103A*, GenBank AF217245; and *DEFB104* [9], with nucleotide 1 being the first nucleotide of exon 1.

DEFB1, DEFB4, and *DEFB104* were found to be very polymorphic (Table 1). *DEFB103A*, in contrast, was not polymorphic at all: only 1 of the 350 DNA samples harbored the nucleotide change 16T → C (Table 2). The most frequent nucleotide changes were located in the introns and the 5′ and 3′UTRs. The intron of *DEFB4* harbored frequent insertion polymorphisms: 81+160 ins T prolongs a tract of 6 T to 7 T and 81+261 ins GTAGACTGA resulted in a duplication of TAGACTGA separated by 1 G.

Defensin peptides are composed of a hydrophobic prepro sequence, which is thought to regulate the processing of the peptide, and a mature cationic sequence that encodes the antimicrobial and chemotactic active form of the protein. Most of the DEFB1 amino acid changes are located in the mature part of the peptide (Table 2). In *DEFB4*, only two missense mutations were found, which were located in the prepro sequence (Table 2). Contrary to *DEFB1* and *DEFB4*, the missense mutations located in *DEFB104* were detected frequently. Moreover, R36C and R38Q showed population differences in distribution: they were found only once in the South-Italian control and CF populations but at higher frequency in the Belgian and Czech populations.

Population differences in haplotype and genotype frequencies of DEFB1 polymorphisms

The distribution of the genotypes of all the frequent *DEFB1* polymorphisms were in Hardy Weinberg equilibrium. Three polymorphisms located in the 3′UTR of *DEFB1* (at the polymorphic loci 16G/A, 24C/G, and 48A/G) have been reported to be observed on only three possible haplotypes [8]. We could confirm this by segregation analysis in 50 nuclear families (data not shown). Moreover, we also found that an allele at a fourth polymorphic locus was linked to this haplotype:

Table 1
Frequent nucleotide changes (frequency >5%) in the human β-defensin genes

Location	Polymorphism		Control populations			CF populations		
	Polymorphic locus[a]	Allele	Belgian	S. Italian	Czech	Belgian	S. Italian	Czech
DEFB1 5′UTR	16G/A	16A	33.0%	44.5%	27.8%	40.3%	44.6%	39.1%
DEFB1 5′UTR	24C/G	24G	20.5%	18.5%	22.7%	19.4%	16.3%	22.5%
DEFB1 5′UTR	48A/G	48G	53.5%	63.0%	50.5%	59.7%	60.8%	61.6%
DEFB1 intron 1	128+19T/A	128+19A	19.5%	18.5%	22.7%	19.4%	16.3%	22.5%
DEFB1 intron 1	129–226T/C	129–226C	29.5%	34.5%	21.6%	34.4%	36.1%	30.4%
DEFB1 intron 1	129–170C/G	129–170G	29.0%	34.5%	21.6%	34.4%	35.5%	30.4%
DEFB1 3′UTR	279A/G	279G	80.0%	88.5%	74.7%	92.6%	91.0%	83.8%
DEFB4 promoter	1–85T/C	1–85C	**39.2%**	**34.0%**	na	**39.3%**	**32.1%**	na
DEFB4 intron 1	81+160 ins T 81+261 ins	81+160 ins T 81+261 ins	85.1%	83.7%	na	**76.2%**	80.4%	na
DEFB4 intron 1	GTAGACTGA	GTAGACTGA	**30.9%**	**35.7%**	na	**31.8%**	**34.6%**	na
DEFB4 prepro sequence	107T/C (P28P)	107C	**29.3%**	28.0%	**43.1%**	**36.9%**	**29.9%**	**42.9%**
DEFB4 3′UTR	262C/T	262T	**27.8%**	**24.5%**	**31.4%**	**29.5%**	**23.5%**	**29.3%**
DEFB4 intron 2	319+3C/G	319+3G	**23.7%**	18.5%	**31.4%**	**23.0%**	17.9%	**29.3%**
DEFB4 intron 2	319+53G/A	319+53A	**28.3%**	24.0%	**43.1%**	**37.7%**	**27.2%**	**42.1%**
DEFB104 prepro sequence	42G/A (V10I)	42A (10I)	**60.0%**	**53.0%**	na	**54.5%**	**56.5%**	na
DEFB104 intron 1	73–192T/C	73–192C	**37.5%**	**36.5%**	**40.4%**	**28.4%**	**39.4%**	**42.0%**
DEFB104 intron 1	73–158A/G	73–158G	**62.5%**	**63.0%**	**59.0%**	**60.5%**	**54.7%**	**57.9%**
DEFB104 intron 1	73–114C/A	73–114A	**43.0%**	**43.5%**	**39.4%**	**44.0%**	**39.4%**	**39.3%**
DEFB104 intron 1	73–77C/T	73–77T	13.5%	11.5%	13.7%	9.7%	**28.8%**	12.0%
DEFB104 intron 1	73–29G/T	73–29T	**52.5%**	**50.5%**	**50.0%**	**51.5%**	**51.2%**	**47.9%**
DEFB104 mature sequence	120C/T (R36C)	120T (36C)	7.0%	1.5%	9.0%	9.0%	0.0%	12.7%
DEFB104 mature sequence	127G/A (R38Q)	127A (38Q)	9.5%	1.5%	18.0%	16.0%	0.0%	**21.8%**
DEFB104 intron 2	285+13C/T	285+13T	5.5%	10.5%	**24.2%**	3.0%	1.8%	**28.9%**
DEFB104 intron 2	285+26A/G	285+26G	**57.0%**	**60.0%**	**60.0%**	**55.2%**	**60.0%**	**58.5%**
DEFB104 intron 2	285+99C/T	285+99T	**57.0%**	**60.0%**	**59.5%**	**55.2%**	**60.0%**	**58.5%**

The frequencies of the polymorphic alleles are shown for the various populations (na, not analyzed). The amino acid changes are indicated in parentheses. The polymorphic loci that are not in Hardy Weinberg equilibrium in the population are shown in bold.
[a] The nucleotides are numbered at the transcript level according to the published sequences *DEFB1*, GenBank U50930; *DEFB4*, GenBank AF040153; *DEFB103A*, GenBank AF217245; and *DEFB104* [9], with nucleotide 1 being the first nucleotide of exon 1. For mutations located in introns, the affected nucleotide position respective to the first (minus sign) or last nucleotide (plus sign) of the closest exon is given. The possible nucleotides that can be found are then given and separated by a slash.

Table 2
Rare nucleotide changes (frequency <5%) in the human β-defensin genes

Location	Polymorphic locus[a]	Location	Polymorphic locus[a]
DEFB1 promoter	1–127A/T	*DEFB4* promoter	1–180A/C
DEFB1 promoter	1–113C/T	*DEFB4* promoter	1–179G/C
DEFB1 promoter	1–112G/A	*DEFB4* prepro sequence	39C/T (L6F)
DEFB1 promoter	1–69G/A	*DEFB4* prepro sequence	44C/T (L7L)
DEFB1 promoter	1–60G/A	*DEFB4* prepro sequence	65C/T (F14F)
DEFB1 intron 1	128+5G/A	*DEFB4* prepro sequence	71G/A (M16I)
DEFB1 intron 1	129–69C/T	*DEFB4* intron 1	81+145C/A
DEFB1 prepro sequence	131G/T (G22C)	*DEFB4* 3′UTR	247C/G
DEFB1 mature sequence	168A/G (H34R)	*DEFB4* 3′UTR	269A/G
DEFB1 mature sequence	171A/G (Y35C)	*DEFB4* 3′UTR	304A/T
DEFB1 mature sequence	178C/T (C37C)	*DEFB4* intron 2	319+26G/A
DEFB1 mature sequence	179G/A (V38I)		
DEFB1 mature sequence	217G/A (P50P)	*DEFB103A* prepro sequence	16T/C (L6F)
DEFB1 mature sequence	266T/A (C67S)		
DEFB1 3′UTR	300G/A	*DEFB104* promoter	1–23C/T
DEFB1 3′UTR	300G/C	*DEFB104* intron 1	72+2T/C
DEFB1 3′UTR	306G/C	*DEFB104* intron 1	72+147C/A
DEFB1 3′UTR	361A/G	*DEFB104* intron 2	285+57C/G

The amino acid changes are indicated in parentheses.
[a] The nucleotides are numbered at the transcript level according to the published sequences *DEFB1*, GenBank U50930; *DEFB4*, GenBank AF040153; *DEFB103A*, GenBank AF217245; and *DEFB104* [9], with nucleotide 1 being the first nucleotide of exon 1. For mutations located in introns, the affected nucleotide position respective to the first (minus sign) or last nucleotide (plus sign) of the closest exon is given. The possible nucleotides that can be found are then given and separated by a slash.

128+19T/A located in intron 1. The three haplotypes were named A (16G-24C-48A-128+19T), B (16A-24C-48G-128+19T), and C (16G-24G-48G-128+19A). When the frequencies of the A, B, and C haplotypes were compared between the Belgian and the South-Italian control and CF populations, a distribution difference was observed between the two control populations: the A haplotype was more frequent in the Belgians and the B haplotype more frequent in the South Italians. The two CF populations, however, showed no significant differences (Table 3). To analyze this difference in distribution further, three additional populations were analyzed: a Czech control and a Czech F508del/F508del CF population and a North-Italian control population composed of people living in the Northeast of Italy for at least three generations. The CF populations again showed a similar distribution. The North-Italian, Belgian, and Czech control samples were similar to one another but different from the South-Italian controls, with the A haplotype being more frequent in the more northern European populations and the B haplotype being more frequent in the south of Italy. The difference was the largest between the Czechs and the South Italians ($p = 0.0023$). Polymorphisms located in the second exon of *DEFB1* also showed a difference in distribution in South-Italian and Czech control populations (Table 4): 129−170G, 129−226C, and 279G were more frequent in the south of Italy.

Table 4
Distribution of the frequent *DEFB1* intron 1 and exon 2 polymorphic loci

	Genotype			Haplotype	
	TT	TC	CC	T	C
DEFB1 intron 1 129−226T/C					
South-Italian controls ($n = 100$)	43.00%	45.00%	12.00%	65.50%	34.50%
Czech controls ($n = 99$)	60.00%	36.84%	3.16%	78.42%	21.58%
p value χ^2 test Czech/S Italian		0.014		0.005	
	CC	CG	GG	C	G
DEFB1 intron 1 129−170C/G					
South-Italian controls ($n = 100$)	42.00%	44.00%	14.00%	64.00%	36.00%
Czech controls ($n = 99$)	60.00%	36.84%	3.16%	78.42%	21.58%
p value χ^2 test Czech/S Italian		0.006		0.002	
	AA	AG	GG	A	G
DEFB1 exon 2 279A/G					
South-Italian controls ($n = 100$)	1.00%	21.00%	78.00%	11.50%	88.50%
Czech controls ($n = 99$)	8.42%	33.68%	57.89%	25.26%	74.74%
p value χ^2 test Czech/S Italian		0.003		0.0004	

Hardy Weinberg disequilibrium for polymorphisms in DEFB4 and DEFB104

Most polymorphisms located in *DEFB4* and *DEFB104* and occurring at a frequency higher than 20% were not in Hardy Weinberg equilibrium (Table 1). The observed distribution and the expected Hardy Weinberg genotype distributions for some of these polymorphisms in the Belgian control population are shown in Fig. 1. The same observation was made in all other populations. In essence, the number of heterozygotes was higher and the number of homozygotes was lower than expected. For some of the polymorphisms, like 81+261 ins GTAGACTGA or 73−192T/C, no homozygous variants could be observed.

Table 3
Distribution of exon 1 genotypes and haplotypes in *DEFB1* in control and CF patients

	DEFB1 exon 1 ABC genotype						DEFB1 exon 1 ABC haplotype		
	AA	AB	AC	BB	BC	CC	A	B	C
Belgian controls ($n = 100$)	17.00%	36.00%	23.00%	9.00%	12.00%	3.00%	46.50%	33.00%	20.50%
South-Italian controls ($n = 100$)	18.00%	26.00%	12.00%	22.00%	19.00%	3.00%	37.00%	44.50%	18.50%
North-Italian controls ($n = 99$)	22.22%	32.32%	19.19%	12.12%	14.14%	0.00%	47.98%	35.35%	16.67%
Czech controls ($n = 99$)	23.23%	29.29%	23.23%	6.06%	14.14%	4.04%	49.49%	27.78%	22.73%
p value χ^2 test Belg/S Italian			**0.031**					**0.056**	
p value χ^2 test Belg/Czech			0.77					0.52	
p value χ^2 test Czech/S Italian			**0.014**					**0.0023**	
p value χ^2 test Belg/N Italian			0.436					0.608	
p value χ^2 test N Italian/S Italian			0.098					**0.078**	
p value χ^2 test Czech/N Italian			0.256					0.158	
Belgian F508del patients ($n = 61$)	12.90%	41.94%	12.90%	11.29%	16.13%	4.84%	40.32%	40.32%	19.35%
South-Italian F508del patients ($n = 82$)	10.84%	44.58%	12.05%	13.25%	18.07%	1.20%	39.16%	44.58%	16.27%
Czech F508del patients ($n = 69$)	17.39%	27.54%	14.49%	17.39%	15.94%	7.25%	38.41%	39.13%	22.46%
p value χ^2 test Belg/S Italian			0.814					0.739	
p value χ^2 test Belg/Czech			0.601					0.766	
p value χ^2 test Czech/S Italian			0.159					0.381	

A, 16G-24C-48A-128+19T; B, 16A-24C-48G-128+19T; C, 16G-24G-48G-128+19A. Significant values are shown in bold.

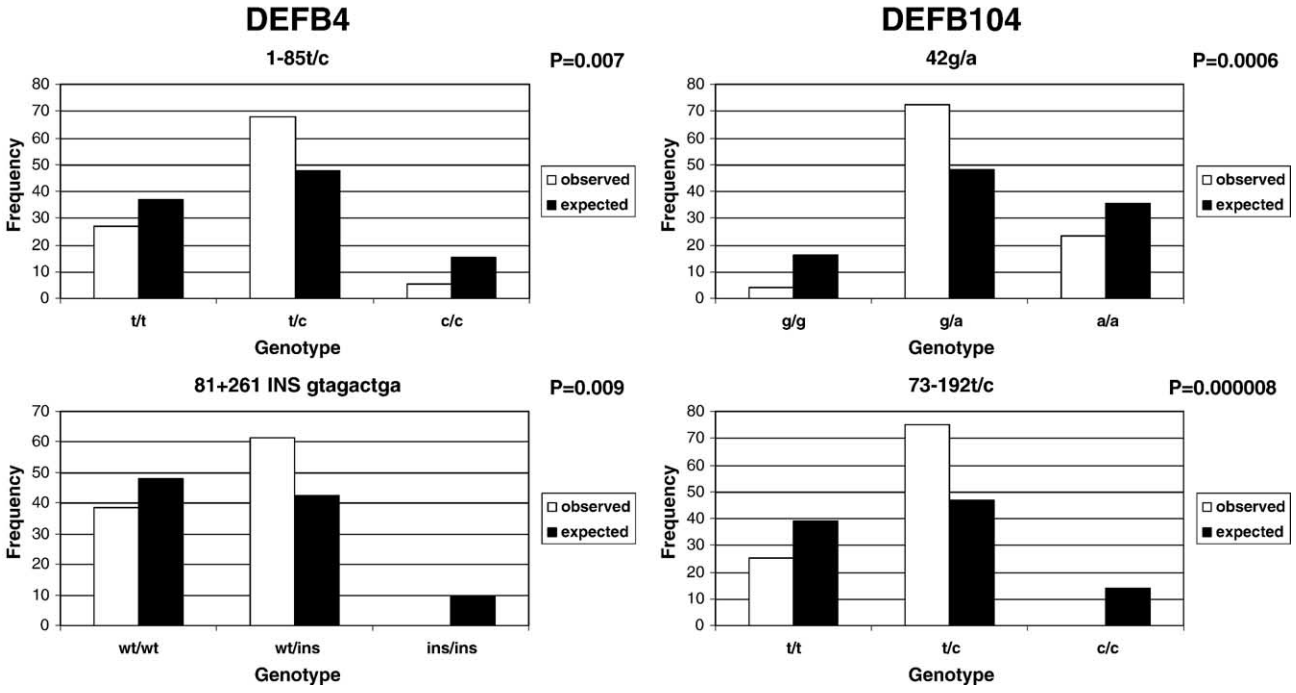

Fig. 1. Hardy Weinberg disequilibrium for some frequent polymorphisms located in *DEFB4* (left) and *DEFB104* (right). The genotype distribution of the Belgian control samples is depicted with white bars; the expected Hardy Weinberg genotype distribution is depicted with black bars. The observed and expected genotype distributions of each polymorphic locus were compared using χ^2 testing; the p values are indicated.

Hollox et al. found that the *DEFB4, DEFB103A*, and *DEFB104* genes, but not *DEFB1*, are part of a 240-kb repeat unit found in 2 to 12 copies in a diploid genome [1]. If some of the different copies, present on one chromosome 8, harbor at least one polymorphism, then that individual would be characterized as being heterozygous in a sequencing assay. To analyze this, the DNA of a monochromosomal hybrid containing only one chromosome 8 (hybrid panel Eurogen) was amplified for *DEFB1, DEFB4*, and *DEFB104* by PCR and sequenced. The hybrid cell line contained only one sequence for *DEFB1* but heterozygous signals were observed for one polymorphism in *DEFB4* and six polymorphisms in *DEFB104*. These results indicate that multiple different copies of *DEFB4* and *DEFB104* can be present on one chromosome 8. *DEFB1*, which is located at a distance of ~600 kb from the repeat region, did not show more than one signal for any of the *DEFB1* polymorphisms.

To determine whether a DNA sample derived from an individual could indeed contain more than two *DEFB4* and *DEFB104* haplotypes, the *DEFB4* and *DEFB104* genes were amplified by PCR with *Pwo* DNA polymerase for eight individuals. This enzyme has proofreading activity such that misincorporation of nucleotides during PCR remains very low. Individual PCR fragments were then cloned and sequenced. For each exon, 10 cloned PCR fragments were sequenced and the haplotype of the exon was determined. No polymorphisms other than those found in the original DNA sample were found, indicating that *Pwo* DNA polymerase did not introduce mistakes into the PCR fragments analyzed. The eight samples contained two to six different *DEFB4* exon 1 and exon 2 haplotypes and three to seven different *DEFB104* exon 2 haplotypes. An example of the haplotype analysis of the cloned fragments for sample 6 is given in Table 5. Seven haplotypes for *DEFB104* exon 2 were found, indicating that the *DEFB104* region was present in at least seven different copies on the two chromosomes 8.

Association of DEFB1 polymorphisms with lung disease severity in cystic fibrosis

Initially, this study was set up to analyze the influence of polymorphisms in the human β defensins on the cystic fibrosis lung phenotype. Given the more complex nature that is now found for *DEFB4, DEFB103A*, and *DEFB104*, and since we do not know the exact number of repeats in the individuals in this study, association studies could not be performed for *DEFB4, DEFB103A*, and *DEFB104*.

Association studies of *DEFB1* polymorphisms with CF lung disease severity could be performed since *DEFB1* is not part of this repeat region. Recurrent infections with *Pseudomonas aeruginosa* result in increasing lung deterioration. A lower age of first infection is generally associated with worse outcome. Therefore, the different frequent polymorphisms (>20%) located in *DEFB1* were analyzed for a possible correlation with an earlier, or later, age of first infection by *P. aeruginosa*. The difference in the mean age of first colonization of the different genotypes of the polymorphisms was analyzed using ANOVA. No association was

Table 5
Composition of the different repeat units of *DEFB4* exon 1, *DEFB4* exon 2, and *DEFB104* exon 2 found in one DNA sample

	Haplotypes identified						
DEFB4 exon 1							
1–180A/C	C	A	A	A	C		
1–85T/C	C	T	T	T	C		
81+160 ins T	Ins	Ins	No ins	Ins	No ins		
81+261 ins GTAGACTGA	No ins	No ins	No ins	Ins	No ins		
DEFB4 exon 2							
107T/C	T	C	T	T			
319+26G/A	A	G	G	G			
319+53G/A	G	A	G	A			
DEFB104 exon 2							
73–192T/C	C	T	T	T	T	C	T
73–158A/G	G	A	A	G	A	G	G
73–77C/T	T	C	C	C	C	C	C
73–29G/T	T	G	G	T	G	G	G
120C/T	C	C	T	C	C	C	T
127G/A	G	A	G	G	G	A	G
285+13C/T	C	T	C	C	C	T	C
285+26A/G	G	G	A	G	G	G	A
285+99C/T	T	T	C	T	T	T	C

The polymorphic loci in this sample are depicted in the left column. The identities of the different alleles found in the various single amplification products derived from that DNA sample are indicated in the remaining columns. Five different haplotypes were found for *DEFB4* exon 1, four for *DEFB4* exon 2, and seven for *DEFB104* exon 2.

found with any of the frequent *DEFB1* polymorphisms and the age of presumed first *P. aeruginosa* infection.

The severity of lung disease is measured by the forced expiratory volume in 1 s (FEV1). The FEV1 values at the age of 12–14 were available for most patients and they were calculated as FEV1% predicted according to Knudson et al. [16]. No association was found between the frequent polymorphisms (>20%) located in *DEFB1* and FEV1% for these patients.

Discussion

During evolution, defensins have evolved by rounds of duplication and selection, with rapid divergence of sequences encoding the mature protein, presumably to create antimicrobial compounds that are more active and/or better adapted, thereby strengthening the innate defenses of an individual [10]. At the start of this study only four polymorphic loci, all located in *DEFB1*, were known [8]. The sequencing results reported in this study show that *DEFB1, DEFB4*, and *DEFB104* are very polymorphic and that *DEFB103A* is not polymorphic. In *DEFB103A*, only one nucleotide change was found and this in only 1 of the 350 samples analyzed. *DEFB103A* has the highest antimicrobial activity, in comparison to the other defensins analyzed in this study [11]. *DEFB103A* also shows a marked degree of conservation in humans, great apes, and New and Old World monkeys [12]. These data might indicate that this very active, and even salt-insensitive, antimicrobial agent is very crucial in our defenses and that mutations are not tolerated. This hypothesis requires, however, further investigation.

The promoters of *DEFB1* and *DEFB4* harbored multiple rare mutations. None of the changes found in the promoter of *DEFB4* affect NF-κB binding sites. NF-κB is a known mediator of bacteria- and inflammation-induced *DEFB4* expression. Multiple missense mutations were found at low frequency in *DEFB1*. Only one was found in the pro sequence and six were found in the mature sequence. V38I has been detected in 2.8% of healthy Japanese individuals and in 15% of Japanese COPD patients [13]. We observed this mutation only once in a CF patient.

Population differences in the frequency of polymorphisms have been observed throughout the whole *DEFB1* gene. Jurevic et al. [14] described differences in the distribution of 16G/A, 48A/G, 128+19T/A, 279A/G, and 361A/G comparing Northern European, Afro-American, Middle Eastern, Mexican, and Senegalese individuals. We observed differences within European countries: the distribution of these polymorphisms was similar between Czech, Belgian, and North-Italian populations but differed significantly from that of South Italians.

Most of the frequent *DEFB4* and *DEFB104* polymorphisms were not in Hardy Weinberg equilibrium in the populations analyzed, while the *DEFB1* polymorphisms were all in Hardy Weinberg equilibrium. All these genes are located on the short arm of chromosome 8, but *DEFB4, DEFB103A*, and *DEFB104* are clustered together at a distance of ~600 kb from *DEFB1*. Since the start of this study, it was found that *DEFB4, DEFB103A*, and *DEFB104* are part of a repeat unit of >240 kb that can be found at a copy number of 2 to 12 in a diploid genome [1]. Sequencing of a monochromosomal hybrid containing one chromosome 8 indeed showed 2 nucleotides for some polymorphic loci, supporting this observation. Moreover, cloning and sequencing of individual PCR fragments of *DEFB4* exons 1 and 2 and of *DEFB104* exon 2 showed that more than two haplotypes can be present for the different exons in an individual and that the repeats on one chromosome may thus not be identical. In one DNA sample, even seven different haplotypes for *DEFB104* exon 2 were found. When the frequency of these frequent polymorphisms was analyzed in the different populations studied, Hardy Weinberg disequilibrium was found due to the fact that the number of homozygotes was too low and the number of heterozygotes was too high. This might be explained by the fact that the various copies of the *DEFB4–DEFB103A–DEFB104* region present on one chromosome 8 might not be identical. Indeed, from the moment that one of the various repeats carries a different allele at a particular polymorphic locus, the individual will be typed as heterozygous for that polymorphic marker with a sequencing assay. An individual will be typed as homozygous for a

polymorphic marker only if the same allele is found at that polymorphic marker in all repeats.

β Defensins are part of our innate immune system. They can kill microbes by integrating into the membrane of bacteria, fungi, and enveloped viruses. Through their chemotactic activity they form the link between the innate and the acquired immune system. Polymorphisms in these genes might therefore influence the clinical phenotype of patients affected by lung diseases, as has been observed for V38I in Japanese COPD patients. The *DEFB1* 24C/G polymorphism has also been associated with low levels of oral *Candida* [15]. The different frequent polymorphisms (>20%) found in *DEFB1* were analyzed for a possible modulating role on CF lung disease severity. No association was found between the tested *DEFB1* polymorphisms with the age of first infection by *P. aeruginosa* or with FEV1%. Strangely enough the significant differences in *DEFB1* polymorphisms observed between South Italians and other European populations in the controls could not be confirmed in the CF patients. This suggests that the CF patients of the south of Italy represent a biased sample of the population. Whether this is in any way related to the disease remains to be determined.

Most of the frequent *DEFB4* and *DEFB104* polymorphisms were not in Hardy Weinberg equilibrium. Nevertheless the distribution of these polymorphisms was significantly different between the CF patients and the controls (data not shown). The interpretation of these differences, however, will remain hazardous if no haplotype analysis can be done. Complete characterization of the number of repeats and of the polymorphisms they each carry will be required before these data can be examined with confidence.

Materials and methods

Populations

For each population, i.e., Belgian, South Italian, North Italian, and Czech, 100 DNA samples of individuals not related to the cystic fibrosis patients were studied. For association studies, 61 Belgian, 82 South-Italian, and 67 Czech CF patients were investigated. Only CF patients homozygous for the F508del mutation were included. FEV1 values and age of first detected infection by *P. aeruginosa* were taken from their clinical records. The FEV1 values were normalized to FEV1% according to Knudson et al. [16].

DNA extraction and PCR

DNA was isolated from whole blood samples. The two exons, and their exon/intron boundaries, of each β-defensin gene were amplified in two PCRs (PCR primer sequences available on request). For *DEFB1* and *DEFB4*, 200 ng of genomic DNA was used for PCR in a total volume of 50 μl containing 1× PCR buffer (Roche), 1.5 mM $MgCl_2$, 200 μM dNTPs (Amersham Pharmacia), 10 pmol of each primer (forward and reverse), and 1.75 U of AmpliTaq DNA polymerase (Roche). Amplification was performed using the following temperature profile: 5 min at 94°C; 35 cycles composed of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; 7 min at 72°C; and 15 min at 4°C. The PCR mixtures for *DEFB103A* and *DEFB104* included (in a final volume of 50 μl) 200 ng of DNA, 1× PCR buffer II (Applied Biosystems), 1.5 mM $MgCl_2$, 250 μM dNTPs (Amersham Pharmacia), 15 μM primers (forward and reverse), and 2 U of *Taq* DNA polymerase (Applied Biosystems). A touchdown PCR protocol was used: initial denaturation, 94°C for 3 min; amplification, 14 cycles of 94°C for 20 s, 62°C for 40 s, decreasing 0.5°C each cycle, 72°C for 45 s, and then 25 further cycles of 94°C for 20 s, 55°C for 40 s, 72°C for 45 s; final extension, 72°C for 7 min. All PCR amplifications were performed on a GeneAmp PCR System 9700 (Applied Biosystems). PCR amplification was checked by 2% agarose gel electrophoresis in 1× TBE.

Sequencing analysis

Sequence analysis of the PCR products, purified according to Werle et al. [17], was performed on an automated DNA sequencer (Applied Biosystems; Model 3100 genetic analyzer). Cycle sequencing was performed directly on the purified PCR product by BigDye terminator chemistry according to the manufacturer's recommendations (Applied Biosystems; BigDye Terminator v3.0 Cycle Sequencing) (sequencing primers available on request). All sequence changes were confirmed on both strands. Sequencing of *DEFB4* exon 1 + intron 1 revealed an insertion polymorphism of 8 nucleotides, located near the 3′ end of the PCR product. To identify this polymorphism in the DNA samples, an extra sequencing reaction was performed with a forward internal primer.

Cloning of single PCR fragments

DEFB4 exon 1, *DEFB4* exon 2, and *DEFB104* exon 2 were amplified as described. The proofreading enzyme *Pwo* DNA polymerase (Roche) was used instead of *Taq* DNA polymerase. Amplified fragments were separated on a 1% TBE–agarose gel, purified from the gel with the Qiaquick gel extraction kit (Qiagen), ligated in pUC18 with the TOPO cloning kit (Invitrogen), and transfected into *Escherichia coli* cells as described by the manufacturer. The β-defensin-plasmid from individual clones was amplified by colony PCR and the β-defensin polymorphisms were identified by sequencing of the PCR fragments.

Statistics

Statistical analysis was performed using ANOVA (SystatW5 program) to compare the means of groups of data and

$\chi 2$ analysis or Fisher's exact test (Statcalc program) to compare frequency data.

Acknowledgments

The DNA of the monochromosomal hybrids was kindly provided by Peter Marynen (University of Leuven, Louvain, Belgium). Anne Vankeerberghen is a fellow of the Fonds voor Wetenschappelijk Onderzoek Vlaanderen. These investigations have been supported by an FWO grant (G.0389.02N), the CF-PRONET grant (QLG1-CT-2001-01005) from the European Commission, an Interuniversity Poles of Attraction Program grant (P5/25-H), a grant (GOA 99/07) from the Onderzoeksraad KU Leuven, a grant from the Alphonse and Jean Forton–Koning Boudewijn Stichting (2000 14 R7115 B0), the Italian Ministry of University and Research (FIRB 2001-2003), and MZ CR-00000064203, 1A/8336-3, MSMT CR-LN00A079, 111300003, EU CF Chip, and CRMGEN. J.-J. Cassiman is the holder of the Arthur Bax and Anna Vanluffelen Chair of Human Genetics.

References

[1] E.J. Hollox, J.A. Armour, J.C. Barber, Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster, Am. J. Hum. Genet. 73 (2003) 591–600.
[2] B.C. Schutte, et al., Discovery of five conserved beta-defensin gene clusters using a computational search strategy, Proc. Natl. Acad. Sci. USA 99 (2002) 2129–2133.
[3] T. Ganz, Defensins: antimicrobial peptides of innate immunity, Nat. Rev. Immunol. 3 (2003) 710–720.
[4] T. Hiratsuka, et al., Increased concentrations of human beta-defensins in plasma and bronchoalveolar lavage fluid of patients with diffuse panbronchiolitis, Thorax 58 (2003) 425–430.
[5] M.J. Goldman, et al., Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis, Cell 88 (1997) 553–560.
[6] R. Bals, et al., Human beta-defensin 2 is a salt-sensitive peptide antibiotic expressed in human lung, J. Clin. Invest. 102 (1998) 874–880.
[7] F.J. Accurso, M.K. Sontag, Seeking modifier genes in cystic fibrosis, Am. J. Respir. Crit. Care Med. 167 (2003) 289–290.
[8] T. Dork, M. Stuhrmann, Polymorphisms of the human beta-defensin-1 gene, Mol. Cell. Probes 12 (1998) 171–173.
[9] J.R. Garcia, et al., Human beta-defensin 4: a novel inducible peptide with a specific salt-sensitive spectrum of antimicrobial activity, FASEB J. 15 (2001) 1819–1821.
[10] C.A. Semple, M. Rolfe, J.R. Dorin, Duplication and selection in the evolution of primate beta-defensin genes, Genome Biol. 4 (2003) R31.
[11] J. Harder, J. Bartels, E. Christophers, J.M. Schroder, Isolation and characterization of human beta -defensin-3, a novel human inducible peptide antibiotic, J. Biol. Chem. 276 (2001) 5707–5713.
[12] M. Boniotto, et al., A study of host defence peptide beta-defensin 3 in primates, Biochem. J. 374 (2003) 707–714.
[13] I. Matsushita, et al., Genetic variants of human beta-defensin-1 and chronic obstructive pulmonary disease, Biochem. Biophys. Res. Commun. 291 (2002) 17–22.
[14] R.J. Jurevic, P. Chrisman, L. Mancl, R. Livingston, B.A. Dale, Single-nucleotide polymorphisms and haplotype analysis in beta-defensin genes in different ethnic populations, Genet. Test. 6 (2002) 261–269.
[15] R.J. Jurevic, M. Bai, R.B. Chadwick, T.C. White, B.A. Dale, Single-nucleotide polymorphisms (SNPs) in human beta-defensin 1: high-throughput SNP assays and association with Candida carriage in type I diabetics and nondiabetic controls, J. Clin. Microbiol. 41 (2003) 90–96.
[16] R.J. Knudson, M.D. Lebowitz, C.J. Holberg, B. Burrows, Changes in the normal maximal expiratory flow-volume curve with growth and aging, Am. Rev. Respir. Dis. 127 (1983) 725–734.
[17] E. Werle, C. Schneider, M. Renner, M. Volker, W. Fiehn, Convenient single-step, one tube purification of PCR products for direct sequencing, Nucleic Acids Res. 22 (1994) 4354–4355.