

Original article

CLC – a novel susceptibility gene for allergic rhinitis?

Background: Studies of the nasal lavage fluid proteome have previously identified proteins differently expressed in patients with symptomatic allergic rhinitis, e.g. S100A7, prolactin-inducible protein (PIP), wntless-type MMTV integration site family, member 2B (WNT2B), Charcot-Leyden crystal protein (CLC) and palate lung nasal epithelial clone (PLUNC). The aim of the present study was to investigate if genetic variation associated with allergic rhinitis can be found in these genes.

Methods: Peripheral blood was collected from 251 patients with birch and/or grass pollen-induced allergic rhinitis and 386 nonatopic healthy controls. A total of 39 single nucleotide polymorphisms (SNPs) distributed over the genes *PIP*, *WNT2B*, *CLC* and *PLUNC* were selected from dbSNP, genotyped and investigated for associations with allergic rhinitis. Twelve additional SNPs were subsequently analysed for *CLC*.

Results: All 22 investigated SNPs in *CLC* were polymorphic. Ten SNPs yielded significant differences between cases and controls with respect to genotype frequencies. Homozygotes for the minor allele were more common in allergic individuals compared to healthy controls. The minor alleles of these SNPs were all located on the same haplotype. Furthermore, homozygotes for the minor allele of two of the promoter SNPs had higher average scores for birch in skin prick test. In contrast, for seven SNPs within the gene, heterozygotes and homozygotes for the major allele had higher average scores for grass. None of the other three genes showed association.

Conclusion: Genetic variation in *CLC* was found to be associated with allergic rhinitis. The pattern of variation is compatible with a recessive inheritance model and the previously observed altered protein levels detected in patients with allergic rhinitis.

M. Bryborn¹, C. Halldén², T. Säll³,
L. O. Cardell¹

¹Department of Otorhinolaryngology, Karolinska Institutet, Huddinge; ²Department of Clinical Chemistry, Malmö University Hospital, Malmö; ³Department of Cell and Organism Biology, Lund University, Lund, Sweden

Key words: allergic rhinitis; association study; Charcot-Leyden crystal protein; genetic variation.

L. Olaf Cardell
Professor and Head of Division
Division for Otorhinolaryngology
Karolinska Institutet
SE-141 57 Huddinge
Sweden

Accepted for publication 4 June 2009

Allergic rhinitis is defined as allergic inflammation in the nasal mucosa and during ongoing inflammation a large number of different pro- and anti-inflammatory mediators are emitted into the nose, either through plasma extravasation or by secretion from serous and mucous glands (1). Studies have previously been conducted where two-dimensional gel electrophoresis in combination with matrix-assisted laser desorption/ionisation-time-of-flight mass spectrometry (MALDI-TOF) analysis has been used in an attempt to establish biomarkers in nasal lavage fluid, that can be of importance for the development and expression of allergic rhinitis (2, 3). This has resulted in the identification of a number of proteins differently expressed in patients with symptomatic allergic rhinitis against birch and/or grass pollen, e.g. S100A7/psoriasin, wntless-type MMTV integration site family, member 2B (WNT2B) (isoform 1), prolactin-inducible protein (PIP), palate lung nasal epithelial clone (PLUNC) and Charcot-Leyden crystal protein (CLC)/

galectin-10 (2, 3). Proteins involved in innate and acquired immunity together constitute 48% of the total protein content in nasal lavage fluid (4) and it is thus not surprising that the above mentioned proteins all in some way are connected to the immune system; S100A7 has chemotactic and antimicrobial properties and it has recently been shown to activate neutrophils to produce various cytokines and α -defensins (5–7), WNT2B is involved in haematopoiesis (8), PIP inhibits T-lymphocyte apoptosis (9), PLUNC is involved in bacterial defense (10) and CLC/galectin-10 is abundantly expressed in eosinophils and more recently shown also to be expressed in CD4⁺ CD25⁺ regulatory T cells (11, 12). CLC mRNA has been found to be over-expressed in asthma (13). Their functional connection and clinical relevance for the development and manifestation of allergic rhinitis has not yet been established but they should all be regarded as potential biomarkers for allergic rhinitis and thus need further attention.

It is well known that the development of allergic diseases involves both genetic and environmental factors. The genetic factors play a major role and a heritability of *c.* 70% has been reported for allergic rhinitis (14). The genetics behind allergic rhinitis has not been as well investigated as for asthma and atopic dermatitis (15, 16), but it is apparent that many different genes, all with little to modest effect are involved. *S100A7*, *PIP*, *WNT2B*, *PLUNC* and *CLC/galectin-10* are potential effector proteins in allergic rhinitis and it is possible that their expression is modified by genetic variation. Genetic variation in *S100A7* has previously been explored and the *S100A7* gene contains a nonsynonymous SNP (rs3014837) that is associated with the occurrence of allergic rhinitis (17). The present study was designed to investigate if there is genetic variation associated with allergic rhinitis in the genes coding for *PIP*, *WNT2B*, *PLUNC* and *CLC/galectin-10*.

Methods

Subjects

Peripheral blood samples were collected from 251 patients (116 female, 135 male) with symptomatic birch and/or grass pollen-induced intermittent allergic rhinitis and 386 healthy individuals (167 female, 219 male), serving as controls. The median (range) age of patients and controls was 32 (18–64) and 46 (27–69). The diagnosis of birch and/or grass pollen-induced allergic rhinitis was based on a positive history of intermittent allergic rhinitis for at least 2 years and a positive skin prick test (SPT) or Phadiatop test (Pharmacia Upjohn, Uppsala, Sweden) to birch and/or grass.

All patients were classified according to the ARIA criteria (18), as having severe symptoms (itchy nose and eyes, sneezing, nasal secretion and nasal blockage) during pollen season and they had all been treated with antihistamines and nasal steroids during pollen seasons previous years. All patients presented a wheal reaction diameter >3 mm towards birch or timothy (grass) in SPT or a positive Phadiatop test, with at least class 2 in subsequent test with specific allergens. Controls had no history of allergic rhinitis and a negative SPT or Phadiatop test. Both patients and controls were of European

origin, with both parents born in Sweden. The study was approved by the Ethics Committee of the Medical Faculty, Lund University and written informed consent was obtained from all subjects.

SNP selection

To provide a comprehensive overview of the genetic variation in the investigated genes, we aimed to select SNPs distributed over the whole genes, including the putative promoter region, coding regions and flanking intronic sequences. The following criteria were used to select the SNPs: (i) promoter SNPs that might affect the transcriptional activity, as quantitative differences have been reported for all the proteins; (ii) nonsynonymous SNPs and (iii) SNPs in intronic and 3'-UTR regions, with reported frequency in HapMap or dbSNP. Selected SNPs are depicted in Table 1.

DNA sequencing

Genomic DNA was extracted from whole blood using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). The concentrations were determined by fluorimetry using PicoGreen (Molecular Probes, Invitrogen, Eugene, OR, USA). A subset of 96 randomly selected individuals (57 healthy and 39 allergic) were used for sequencing of the putative promoter region and exon 1, 2, 3 and 4 of *CLC*. Primers were designed using OLIGO version 6.69 using M13-derived tails. The samples were sequenced using BigDye Terminator chemistry, version 3.1 on an ABI 3130XL sequencer (Applied Biosystems, Foster City, CA, USA). In brief, primary PCR was performed using 1 ng/μl of template DNA, 1× Amplitaq Gold Master Mix (Applied Biosystems) and 1.5 μM primer (DNA Technology, Risskov, Denmark). PCR amplifications were performed using GeneAmp 9700 machines (Applied Biosystems) as follows: 96°C for 5 min, 40 cycles of 94°C for 30 s, 60 or 63°C for 45 s, 72°C for 45 s and finally 72°C for 10 min. The primary PCR products were purified using size exclusion separation from Millipore (Multiscreen PCR₉₆; Millipore Corporation, Billerica, MA, USA) and a Multimek robotic workstation. Sequencing reactions were then performed using 1 ng/μl of template DNA, 0.25× BigDye Ready Reaction Mix version 3.1, 0.75× BigDye Sequencing buffer, 3.0 μM M13-derived sequencing primer (DNA Technology) and water. PCR amplifications were subsequently performed using GeneAmp 9700 machines as follows: 96°C for 1 min and 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. Reactions were purified using the Montage SEQ₉₆ kit from Millipore and a

Table 1. Investigated SNPs in *PIP*, *WNT2B*, *CLC* and *PLUNC*

<i>PIP</i>		<i>WNT2B</i>		<i>CLC</i>		<i>PLUNC</i>	
SNP	Location	SNP	Location	SNP	Location	SNP	Location
rs315284	Promoter	rs1175646	Promoter	rs420297	Promoter	rs1998149	Promoter
rs10232903	Promoter	rs17490336	Promoter	rs413216	Promoter	rs6059176	Promoter
rs315283	Intron 1	rs12097320	Promoter	rs375688	Promoter	rs2752902	Promoter
rs10234174	Intron 1	rs1175650	Intron 1	rs2074928	Promoter	rs11907039	Promoter
rs315290	Intron 2	rs880727	Intron 3	rs367529	Intron 1	rs750064	Promoter
rs9986765	Intron 2	rs351371	Exon 4	rs17608	Exon 2	rs3787144	Intron 1
rs6959713	Intron 3	rs3737136	Intron 5	rs385520	Intron 2	rs6120177	Exon 2
rs17850704	Exon 4	rs17030446	Intron 5	rs384138	Exon 3	rs2752909	Intron 2
		rs11102489	3'UTR	rs440191	3'UTR	rs1321418	Intron 3
				rs10407799	3'UTR	rs6120180	Exon 4
						rs6120181	Exon 4
						rs6119348	Intron 5

Beckman Multimek robotic workstation according to supplier's instructions. The sequence data was assembled and compared using SEQSCAPE version 2.5 (Applied Biosystems). All automatically identified candidate heterozygotes were confirmed manually.

Genotyping

Genotypes were determined using the Sequenom MassARRAY MALDI-TOF system. The system analyses allele-specific primer extension products using mass-spectrometry. Assay design was made using the MassARRAY Assay Design version 2.0 software (Sequenom, San Diego, CA, USA). Primers were obtained from Metabion GmbH (Martinsried, Germany) and all reactions were run under the same conditions except for the primer annealing temperature of the primary PCR. PCR reactions were performed in a total volume of 6 µl containing 2.5 ng of template DNA, 1.25× HotStar Taq PCR buffer, 0.15 units of HotStar Taq polymerase, 3.5 mM MgCl₂, 0.5 mM dNTPs and 100 nM of each primer. Amplifications were performed using GeneAmp 9700 machines with dual-384 heads as follows: 95°C for 15 min, 45 cycles of 95°C for 20 s, 56, 60 or 64°C for 30 s, 72°C for 60 s and finally 72°C for 3 min. Dephosphorylation of unincorporated dNTPs was achieved using shrimp alkaline phosphatase. Concentrations of individual hME primer pairs were adjusted to even out peak heights in the mass spectrum. The extension reactions were then made by mixing the adjusted MassEXTEND primer mix (containing approximately 1 µM of each primer) with hME EXTEND mix containing buffer and 50 µM of each d/ddNTP mix and 1.25 units of Thermo Sequenase. PCR amplification of hME reactions was performed as follows: 94°C for 2 min and 99 cycles of 94°C for 5 s, 52°C for 5 s and 72°C for 5 s. The samples were then manually desalted by using 6 mg of Clean Resin and a dimple plate and subsequently transferred to a 384-well SpectroCHIP (Sequenom, San Diego, CA, USA) using a nanodispenser.

Genetic analysis

Single nucleotide polymorphisms that produced good quality data were used to analyse 251 patients and 386 controls for associations between genetic variation in *PIP*, *WNT2B*, *CLC* and *PLUNC* and allergic rhinitis. First, the genotype frequencies were calculated and tested for Hardy–Weinberg equilibrium (HWE). Due to the high prevalence of allergy, the test for HW proportions was performed in both patients and controls. Accordingly, a deviation from HWE in controls was not taken as an indication of technical problems with the SNP, but as a potential effect of association. Next, chi-squared homogeneity tests were used to test for associations between allergic rhinitis and genotype- and allele-frequencies at each SNP. Later, when a detailed analysis of the variation in *CLC* was made, it was found that some genotypes were quite rare, causing expected values much lower than 5, which is a standard criterion. Because of this, a permutation test was used instead. First, the difference in allele frequency was calculated from the data set. In order to test for differences between genotype distributions in cases and controls, we calculated the deviation from perfect HW proportions through the *F*-value. For a SNP where *p* is the allele frequency of the major allele and *q* = 1 – *p* is the frequency of the minor allele and *H*_{obs} is the observed frequency of heterozygotes, the *F*-value was calculated as $F = 1 - H_{obs}/(2pq)$. This was done separately for cases and controls and the difference between the *F*-values was calculated. Next, from the total dataset including 635 individuals (both cases and controls), 251 individuals were randomly selected without replacement. The differences in allele frequencies and in *F*-values were then calculated and compared to the observed differences. The

procedure was repeated 10000 times and the *P*-value was taken as the proportion of permutations where the simulated difference was larger than the observed. Using the SNP data, we also investigated the level and pattern of linkage disequilibrium (LD) and haplotype frequencies. For any two loci (sites) A and B with alleles A₁/A₂ and B₁/B₂, respectively, LD was quantified through $R^2 = D^2/(p_{A1}p_{A2}p_{B1}p_{B2})$, where *p*_{A1} is the allele frequency of allele 1 at locus A, etc. and $D = P_{A1B1} - p_{A1}p_{B1}$, where *P*_{A1B1} is the gamete (haplotype) frequency of A₁B₁. Haplotypes for *CLC* were estimated separately for patients and controls using the program PHASE (19).

Results

Association with allergic rhinitis

A total of 39 SNPs distributed over the genes *PIP*, *WNT2B*, *CLC* and *PLUNC* were initially genotyped (Table 1). Eight SNPs were genotyped in *PIP* (7.7 kb) and 9, 10 and 12 SNPs were genotyped in *WNT2B* (53.9 kb), *CLC* (6.8 kb) and *PLUNC* (7.3 kb) respectively. Of the eight analysed SNPs in *PIP*, one (rs17850704) was not polymorphic in our population. The same was true for five of the SNPs in *WNT2B* (rs1175646, rs12097320, rs880727, rs351371 and rs17030446) and five SNPs in *PLUNC* (rs6059176, rs2752909, rs1321418, rs6120180 and rs6120181). Consequently, these SNPs were not further analysed. In *CLC*, all investigated SNPs were polymorphic. This resulted in a SNP density of *c.* 1 SNP/13 kb for *WNT2B* and a density of *c.* 1 SNP/kb for the other three genes. Genotype frequencies were calculated and subsequently tested for HWE. All SNPs in *WNT2B* and *PLUNC* were in HWE, whereas rs6959713 ($\chi^2 = 7.13$, d.f. = 1, *P* = 0.008) in *PIP* and rs420297 ($\chi^2 = 5.24$, d.f. = 1, *P* = 0.022), rs413216 ($\chi^2 = 5.11$, d.f. = 1, *P* = 0.024), rs375688 ($\chi^2 = 5.11$, d.f. = 1, *P* = 0.024) and rs10407799 ($\chi^2 = 4.67$, d.f. = 1, *P* = 0.031) in *CLC* were not. These deviations all occurred in the controls but were not observed in the patients. In all cases, there was an excess of heterozygotes.

All polymorphic SNPs were analysed for association with allergic rhinitis using a chi-squared homogeneity test, both at the allele and genotype level. In *CLC*, no differences were found between the allergic and healthy groups when analysing the allele frequencies. However, when analysing the genotype frequencies it appeared that three SNPs showed increased chi-squared values. One SNP, rs420297, was statistically significant ($\chi^2 = 7.68$, d.f. = 1, *P* = 0.006). None of the SNPs in *PIP*, *WNT2B* and *PLUNC* were found to be associated with allergic rhinitis, neither at allele nor at genotype level (data not shown). These results encouraged us to analyse more SNPs in *CLC*, in an attempt to further strengthen the results. In a second round of genotyping, five additional SNPs in the promoter region (rs2108686, rs1034995, rs412211, rs12611334 and rs390406), three SNPs in intron 1 (rs395969, rs17709471 and rs1985359), one SNP in

intron 2 (rs399641) and three SNPs in intron 3 (rs389708, rs367788 and rs2074926) were analysed. These were tested for HWE in patients and controls, as well as for association at the genotype and allele level. Table 2 summarizes the results of the HWE tests. Out of the 22 SNPs in the CLC gene region, eight showed a significant deviation for HWE among controls. It also shows the *P*-values for the association tests at the allele and genotype levels. No differences were observed with respect to the allele frequencies in the cases and the controls. However, for the test of the genotype distribution, two of the SNPs showed a *P*-value < 0.01 and eight showed a *P*-value < 0.05. These are uncorrected *P*-values. Bon-Ferroni correction based on the 22 SNPs identifies one SNP which remains significant at the 5% level. However, more importantly, in the case of a complete absence of a difference between cases and controls, the expected number of significant SNPs (*P* < 0.05) would be *c.* 1 or 2, if both tests are considered, which is much lower than the observed 10 significant SNPs. It appears that homozygotes for the minor allele are more common in the allergic group compared to the control group, in the putative promoter region (Table 3). Only one homozygote was present in the control group (0.3%), whereas four to five homozygotes were present in the allergic group (1.6–2.2%). This effect was quantified using odds ratios (OR). It is seen in

Table 3 that the OR of the homozygote of the minor allele relative to the homozygote of the major allele is very high. At the same time, the heterozygotes are underrepresented (OR < 1). The latter effect is the cause of the heterozygote excess in the controls. However, although numbers are low for the rare homozygote, the data indicate a much stronger effect of the homozygote.

Association between SPT results and genotype in patients

The Kruskal–Wallis nonparametric test was used to test for effect of patient genotype on the level of allergy, as scored in SPTs. The responses to a number of allergens were measured in the patients. However, several of these had a large proportion of missing data and several showed a skewed distribution. The variables with skewed distributions all had a large proportion of observations with the score zero, i.e. the patient showed no allergy to the allergen in question. Thus, these variables had very low information content and because of this, we made a selection of three allergens; birch, timothy grass and dog which all had a low proportion of missing data and a low proportion of zeros. These three variables were tested against all SNPs. The result is shown in Table 4. Of the nine SNPs in the promoter region, six show a similar pattern. They have relatively low *P*-values for 'birch' and for SNPs rs413216 and rs375688 the *P*-value is exactly

Table 2. HWE test and association test results for all investigated SNPs in *CLC*

SNP	Location	HWE, controls (<i>P</i> -value)	HWE, patients (<i>P</i> -value)	Association test at genotype level (<i>P</i> -value)	Association test at allele level (<i>P</i> -value)	Alleles in Homo sapiens	Major allele frequency, HapMap, CEU/CHB/JPT/YRI†	Allele in Pan Trogodytes
rs2108686	Promoter	0.016*	0.708	0.018*	0.877	C/T	0.858, 1.000, 1.000, 0.408	G‡
rs420297	Promoter	0.022*	0.202	0.001*	0.844	C/T	0.858, 1.000, 1.000, 0.408	T
rs1034995	Promoter	0.216	0.091	0.029*	0.995	T/C	0.867, 1.000, 1.000, 0.817	T
rs413216	Promoter	0.024*	0.566	0.013*	0.860	C/T	0.867, 1.000, 1.000, 0.950	T
rs412211	Promoter	0.035*	0.603	0.024*	0.937	A/T	0.867, 1.000, 1.000, 0.810	ND
rs12611334	Promoter	0.689	0.920	0.824	0.717	C/G	0.658, 0.822, 0.822, 0.842	ND
rs375688	Promoter	0.024*	0.617	0.016*	0.870	T/C	0.867, ND, ND, 0.810	A‡
rs390406	Promoter	0.018*	0.313	0.008*	0.952	G/A	0.867, 1.000, 1.000, 0.517	A
rs2074928	Promoter	0.115	0.680	0.498	0.916	C/T	0.950, 0.667, 0.544, 1.000	T
rs395969	Intron 1	0.029*	0.639	0.021*	0.927	C/T	1.000, 1.000, 1.000, 0.981	T
rs17709471	Intron 1	0.439	0.888	0.567	0.810	G/A	0.767, 0.811, 0.791, 0.991	G
rs367529	Intron 1	0.277	0.435	0.991	0.951	A/G	ND	ND
rs1985359	Intron 1	0.254	0.740	0.712	0.684	G/A	0.667, 0.830, 0.822, 0.898	ND
rs17608	Exon 2	0.299	0.435	0.985	0.950	T/C	0.708, 0.375, 0.409, 0.767	A‡
rs385520	Intron 2	0.277	0.435	0.991	0.951	T/A	0.708, 0.372, 0.409, 0.771	A
rs399641	Intron 2	1.64	0.320	0.790	0.998	A/T	0.708, 0.367, 0.400, 0.758	Indel‡
rs384138	Exon 3	0.200	0.522	0.772	0.962	G/A	ND	A
rs389708	Intron 3	0.192	0.590	0.019*	0.889	T/A	0.877, 0.989, 1.000, 0.818	ND
rs367788	Intron 3	0.320	0.180	0.688	0.955	C/T	ND	T
rs2074926	Intron 3	0.740	1.000	0.895	0.889	T/C	0.607, 0.381, 0.402, 0.309	C
rs440191	3'UTR	0.277	0.435	0.905	0.915	G/A	0.708, 0.367, 0.409, 0.788	ND
rs10407799	3'UTR	0.031*	0.888	0.040*	0.838	C/T	0.867, 1.000, 1.000, 0.867	ND

ND, not determined.

**P* < 0.05, HWE tests and association tests for allele frequencies were performed with 1 d.f., whereas the association tests for genotype frequencies were performed with 2 d.f.

†Human populations analysed in the HapMap project; CEU [CEPH (Utah residents with ancestry from northern and western Europe)], CHB (Han Chinese in Beijing, China), JPT (Japanese in Tokyo, Japan) and YRI (Yoruba in Ibadan, Nigeria).

‡Allele not in human sample.

Table 3. Genotype frequencies for significant SNPs in the putative promoter region of *CLC*

SNP	Genotype	Controls, <i>n</i> (%)	Patients, <i>n</i> (%)	Association test at genotype level (<i>P</i> -value*)	Odds ratio: 11 vs 12, 11 vs 22
rs2108686	CC	288 (74.8)	194 (78.2)	0.02	
	CT	96 (24.9)	50 (20.2)		0.78
	TT	1 (0.3)	4 (1.6)		6.00
rs420297	CC	285 (75.4)	147 (79.9)	0.001	
	CT	92 (24.3)	33 (17.9)		0.68
	TT	1 (0.3)	4 (2.2)		7.80
rs1034995	TT	269 (81.5)	175 (82.9)	0.03	
	CT	60 (18.2)	32 (15.2)		0.80
	CC	1 (0.3)	4 (1.9)		6.18
rs413216	CC	284 (75.5)	193 (79.1)	0.01	
	CT	91 (24.2)	47 (19.3)		0.75
	TT	1 (0.3)	4 (1.6)		5.92
rs412211	AA	294 (76.8)	195 (78.9)	0.02	
	TA	88 (23.0)	48 (19.4)		0.81
	TT	1 (0.3)	4 (1.6)		6.10
rs375688	TT	284 (75.5)	193 (78.8)	0.02	
	TC	91 (24.2)	48 (19.6)		0.76
	CC	1 (0.3)	4 (1.6)		5.92
rs390406	GG	289 (75.1)	197 (78.8)	0.001	
	GA	95 (24.7)	48 (19.2)		0.80
	AA	1 (0.3)	5 (2.0)		7.50
rs395969	CC	294 (76.4)	197 (78.8)	0.02	
	TC	90 (23.4)	49 (19.6)		0.80
	TT	1 (0.3)	4 (1.6)		6.00

Odds ratio: 11 – homozygote for major allele; 12 – heterozygote; 22 – homozygote for minor allele.

*2 degrees of freedom.

0.05. An inspection of the means for the genotypes show that it is the homozygotes for the rare allele, identified above as risk genotypes, which have the highest average scores for the variable. As an illustration, in rs375688 the average mean for genotype TT is 2.10, for TC 2.70 and for CC 4.0 where C is the minor allele. Three other SNPs; rs395969, rs389708 and rs2074926, also showed low *P*-values for birch. When tested for association to 'timothy grass', another group of seven SNPs located within the gene showed *P*-values between 0.009 and 0.012. Here, the homozygote for the major allele and the heterozygote show higher scores than the homozygote for the minor allele. As an illustration, in rs367529 the average mean for genotype AA is 3.32, for AG 3.66 and for GG 2.12 where G is the minor allele.

Linkage disequilibrium and haplotype frequencies

The pattern of LD, quantified as R^2 , across *CLC* is shown in Table 5. Two major groups of SNPs with very high LD within groups and virtually no LD between groups appear. The first group contains 10 SNPs, including

Table 4. Quantitative analysis of genotype effects in patients

SNP	Birch (<i>P</i> -value)	Grass (<i>P</i> -value)	Dog (<i>P</i> -value)
rs2108686	0.274	0.166	0.165
rs420297	0.052	0.301	0.506
rs1034995	0.065	0.474	0.270
rs413216	0.050	0.306	0.521
rs412211	0.077	0.258	0.421
rs12611334	0.163	0.380	0.536
rs375688	0.050	0.306	0.521
rs390406	0.070	0.266	0.435
rs2074928	0.500	0.772	0.607
rs395969	0.070	0.266	0.435
rs17709471	0.134	0.121	0.605
rs367529	0.614	0.011	0.804
rs1985359	0.176	0.327	0.631
rs17608	0.614	0.011	0.805
rs385520	0.614	0.011	0.805
rs399641	0.656	0.009	0.760
rs384138	0.552	0.012	0.824
rs389708	0.084	0.178	0.555
rs367788	0.689	0.010	0.803
rs2074926	0.045	0.601	0.863
rs440191	0.614	0.011	0.804
rs10407799	0.185	0.306	0.156

P-values < 0.05 are marked in bold.

eight SNPs in the beginning of the gene and two SNPs within the gene, rs389708 and rs10407799. The other major group consists of the seven SNPs within the gene that all show significant genotype effects among the patients for 'timothy grass' in Table 4. Finally, rs12611334 and rs1985359 show strong LD but are unassociated to all other SNPs. In conclusion, the LD pattern is in accordance both with the case-control test results and the quantitative test results.

The results above prompted us to estimate the haplotypes of the complete region. Table 6 shows the results for the SNPs in the beginning of the gene, called the putative promoter region. The results reveal a relatively simple haplotype pattern in this part of the *CLC* gene, encompassing approximately 4 kbp from marker rs2108686 to rs395969. Two very similar haplotypes differing in one SNP only (no. 1 and 2), constitute approximately 50% and 30% of all chromosomes respectively. A third much less common haplotype (no. 4), is also very similar to the two common haplotypes and constitute approximately 3%. Finally, one haplotype which almost entirely is composed of the minor allele for each SNP represents slightly more than 12% (no. 3). This haplotype differs in 8–9 SNPs from the other three (Table 6). The pattern is very similar for both patients and controls, reflecting the fact that deviation was observed at the genotype level but not when allele frequencies were tested. The fact that the percentages do not sum up to 100% is due to a number of very rare haplotypes not being included in the table. Thus, average haplotype frequencies are given in Table 6. The most important implication of the haplotype frequencies is that

Table 5. Linkage disequilibrium of the *CLC* gene region

SNP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1.0	0.93	0.90	0.93	0.93	0.07	0.93	1.0	0.01	0.93	0.04	0.07	0.07	0.07	0.07	0.07	0.07	0.93	0.07	0.18	0.07	0.93
2		1.0		1.0	0.98	0.07	1.0	0.93	0.01	0.98	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.98	0.07	0.18	0.06	1.0
3			1.0	0.96	1.0	0.05	0.96	0.90	0.01	1.0	0.03	0.05	0.05	0.05	0.05	0.05	0.05	1.0	0.05	0.14	0.05	0.96
4				1.0	0.98	0.07	1.0	0.93	0.01	0.98	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.98	0.06	0.18	0.06	1.0
5					1.0	0.06	0.98	0.93	0.01	1.0	0.04	0.06	0.06	0.06	0.06	0.06	0.06	1.0	0.06	0.17	0.06	0.97
6						1.0	0.07	0.07	0.02	0.06	0.14	0.21	0.99	0.21	0.21	0.21	0.21	0.06	0.21	0.36	0.21	0.06
7							1.0	0.93	0.01	0.98	0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.98	0.06	0.18	0.06	1.0
8								1.0	0.01	0.93	0.05	0.07	0.07	0.07	0.07	0.07	0.07	0.93	0.07	0.18	0.07	0.93
9									1.0	0.01	0.01	0.10	0.02	0.10	0.10	0.09	0.10	0.01	0.10	0.06	0.09	0.01
10										1.0	0.04	0.06	0.06	0.06	0.06	0.06	0.06	1.0	0.06	0.17	0.06	0.98
11											1.0	0.14	0.14	0.14	0.14	0.14	0.14	0.04	0.14	0.25	0.14	0.04
12												1.0	0.20	1.0	1.0	0.98	1.0	0.06	0.99	0.56	0.98	0.06
13													1.0	0.20	0.20	0.20	0.20	0.06	0.20	0.35	0.20	0.06
14														1.0	1.0	0.98	1.0	0.06	0.99	0.56	0.98	0.06
15															1.0	0.98	1.0	0.06	0.99	0.56	0.98	0.06
16																1.0	0.98	0.06	0.99	0.57	0.96	0.06
17																	1.0	0.06	0.99	0.56	0.98	0.06
18																		1.0	0.06	0.17	0.06	0.98
19																			1.0	0.57	0.97	0.06
20																				1.0	0.55	0.17
21																					1.0	0.06
22																						1.0

SNP order: 1 – rs2108686; 2 – rs420297; 3 – rs1034995; 4 – rs413216; 5 – rs412211; 6 – rs12611334; 7 – rs375688; 8 – rs390406; 9 – rs2074928; 10 – rs395969; 11 – rs17709471; 12 – rs367529; 13 – rs1985359; 14 – rs17608; 15 – rs385520; 16 – rs399641; 17 – rs384138; 18 – rs389708; 19 – rs367788; 20 – rs2074926; 21 – rs440191; 22 – rs10407799. Linkage disequilibrium, quantified as R^2 of the *CLC* gene region. R^2 -values > 0.90 are marked in dark grey and R^2 -values < 0.90 > 0.10 are marked in light grey.

the association observed at the 8 SNPs is due to a single haplotype, haplotype no. 3.

The haplotypes of the seven SNPs that showed an effect for 'timothy grass' was investigated separately. Only two haplotypes had population frequency estimates over 0.005. The common haplotype had an estimated frequency of 0.69 and had the alleles A, T, T, A, G, C, C at rs367529, rs17608, rs385520, rs399641, rs384138, rs367788 and rs440191 respectively. The less common haplotype had a population frequency of 0.31 and had G, C, A, T, A, T, A at the corresponding sites. This region is approximately 5 kbp and also in this case, it is the complete haplotype that is responsible for the effects seen among the patients.

Genetic variation in *CLC*

In order to investigate the presence of additional nonsynonymous and promoter polymorphisms in *CLC*, possibly involved in the development of allergic rhinitis, we resequenced the putative promoter region and exon 1, 2, 3 and 4 of this gene in a total of 96 individuals. Six

previously known polymorphisms were identified. All of them have minor allele frequencies in excess of 5% and all represent annotated SNPs with frequencies in the sequenced regions. Thus, this investigation did not identify additional rare polymorphisms. However, comparison of the genotyped SNPs in our study with HapMap data and Chimpanzee data (Pan troglodytes) did reveal a number of interesting features. When comparing the SNPs in the promoter region (Table 2), it is seen that the CEU population (Thirty sets of samples with two parents and one child collected by Centre d'Étude du Polymorphisme Humain from Utah residents with ancestry from northern and western Europe) used in HapMap is rather similar to our Swedish sample, where the minor allele has a frequency of approximately 12% throughout the region. What is also notable is that in the CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) populations, the major alleles are fixed for six out of eight informative SNPs, while the YRI (Yoruba in Ibadan, Nigeria) population shows the usual pattern of being more variable. The most striking observation in the promoter region is that the Chimpan-

Table 6. Estimated haplotype frequencies in the putative promoter region of *CLC*

SNP haplotype	rs2108686	rs420297	rs1034995	rs413216	rs412211	rs12611334	rs375688	rs390406	rs2074928	rs395969	Average haplotype frequency
1	C	C	T	C	A	C	T	G	C	C	0.51
2	C	C	T	C	A	G	T	G	C	C	0.32
3	T	T	C	T	T	C	C	A	C	T	0.12
4	C	C	T	C	A	C	T	G	T	C	0.03

Minor alleles are given in bold letter.

zee carries the minor allele in four out of seven informative positions. Thus, the minor haplotype in our population sample (haplotype no. 3, Table 6) might to a large extent be the ancestral haplotype. The remaining SNPs covering the rest of the gene display a completely different pattern in the HapMap populations. The overall level of variation is higher and the CHB and JPT populations are the most variable. In this region, the chimpanzee carries the minor allele in five out of eight of the informative SNPs, further suggesting that the minor haplotype is the ancestral haplotype.

Discussion

The present study aimed to investigate, if there is genetic variation associated with allergic rhinitis in the genes coding for PIP, WNT2B, PLUNC and CLC/galectin-10. These proteins have all been previously identified as being differently expressed in nasal lavage fluid during symptomatic allergic rhinitis. Genetic variation in the putative promoter region of *CLC* was found to be associated with allergic rhinitis in our population, whereas none of the other three genes showed association. Homozygotes for the minor allele of 8 SNPs in the *CLC* promoter region were more common in the allergic group. Despite the fact that the healthy nonatopic group contained 1.5-fold more individuals than the allergic group, only one healthy individual (0.3%) was homozygous for the minor allele, whereas four to five homozygotes (1.6–2.2%) were present in the allergic group. The 8 SNPs in the putative promoter region were all located on the same haplotype, thus this haplotype appears to act as a recessive risk factor for allergic rhinitis. What is also notable is that this haplotype can be the ancestral haplotype, present in our closest relative among the apes, the Chimpanzee, making the possibility of a selective history of the region more likely. However, further investigations are needed before this can be concluded.

CLC/galectin-10 belongs to the β -galactoside-binding lectin family. Several members of this family are suggested to be implicated in the allergic response. One of the best characterized galectins, galectin-3, is involved in various inflammatory processes, including chemoattraction (20) and cell activation (21). For instance, it has the ability to bind to IgE and Fc ϵ R1, thereby inducing mast cell activation (22). Furthermore, galectin-9 is a chemoattractant and activator for eosinophils (23, 24) and galectin-14 is found in bronchoalveolar lavage after allergen challenge (25). CLC/galectin-10 has long been regarded as an eosinophil marker due to the abundant expression in this cell type and high levels of CLC/galectin-10 are often found in various tissues, body fluids and secretions in association with allergic inflammation (26–28). In nasal lavage fluid, CLC/galectin-10 is detected in allergic individuals during pollen season, at considerable higher levels compared to healthy controls, but not outside season (3). The infiltration of eosinophils in the nasal mucosa is more profound in

allergic individuals and the fact that CLC/galectin-10 is detected only during pollen season suggests that eosinophils release this protein during pollen season as a result of activation. The functions of CLC/galectin-10 have not been fully established. It was long regarded as the lysophospholipase of eosinophils but recent data show that this is not the case (29). Current data suggest that intracellular CLC is essential for the functional properties of CD⁺ CD25⁺ regulatory T cells (12). The results in the present study indicate that the promoter region of the *CLC* gene is associated with allergic rhinitis. Genetic variation in a promoter region may affect the gene transcription level but additional studies are essential to distinguish if this in fact affects the CLC/galectin-10 protein levels in allergic individuals.

A complication in the overall analysis of genotype frequencies is the puzzling observation that there is a deviation from HWE in the controls but not in the allergic group. As deviations from HWE in many cases indicate technical problems in genotype determination, we would like to point out that: (i) very few missing values are present in the data set; (ii) all SNPs gave rise to high quality calls with very good discrimination between the different alleles and (iii) all seven markers in the putative promoter region show a similar deviation from HWE and are independent in the meaning that their PCR assays are not overlapping. We therefore rule out technical problems as an explanation to these results. Because of the high prevalence of allergy, the control group cannot be viewed as a sample of the entire population, as they are selected conditionally on the absence of allergy. Instead, an associated SNP may very well show a deviation from HWE in both groups. However, if we estimate the genotype frequencies in the whole population under the assumption of a prevalence of allergic rhinitis of 23% (30–32), we still get an overrepresentation of heterozygotes. To illustrate, in SNP rs2108686 the genotype frequencies will be 0.756, 0.238 and 0.006, which corresponds to a relative excess of heterozygotes of approximately 8%. Whether this deviation is true remains to be investigated but we would like to emphasize that the most pronounced difference between patients and controls is the effect of the homozygotes of the minor allele, an observation that renders *CLC* a strong candidate for further investigation regarding its association with allergic rhinitis.

Acknowledgments

This study was supported by the Swedish Medical Research Council, the Swedish Heart Lung Foundation, the Swedish Association for Allergology and the Swedish Foundation for Health Care Science and Allergic Research. The authors would also like to thank Ingegerd Larsson, Ann Reutherborg, Anna Karin Bastos, Josefine P. Riikonen and Eva Thylander for generous help with collecting the blood samples and Agneta Östensson, Agneta Sterner, Liselotte Hall and Maria Sterner for DNA preparation and SNP genotyping.

References

1. Casado B, Pannell LK, Viglio S, Iadarola P, Baraniuk JN. Analysis of the sinusitis nasal lavage fluid proteome using capillary liquid chromatography interfaced to electrospray ionization-quadrupole time of flight – tandem mass spectrometry. *Electrophoresis* 2004;**25**:1386–1393.
2. Bryborn M, Adner M, Cardell LO. Psoriasin, one of several new proteins identified in nasal lavage fluid from allergic and non-allergic individuals using 2-dimensional gel electrophoresis and mass spectrometry. *Respir Res* 2005;**6**:118.
3. Ghafouri B, Irander K, Lindbom J, Tagesson C, Lindahl M. Comparative proteomics of nasal fluid in seasonal allergic rhinitis. *J Proteome Res* 2006;**5**:330–338.
4. Casado B, Pannell LK, Iadarola P, Baraniuk JN. Identification of human nasal mucous proteins using proteomics. *Proteomics* 2005;**5**:2949–2959.
5. Jinquan T, Vorum H, Larsen CG, Madsen P, Rasmussen HH, Gesser B et al. Psoriasin: a novel chemotactic protein. *J Invest Dermatol* 1996;**107**:5–10.
6. Glaser R, Harder J, Lange H, Bartels J, Christophers E, Schroder JM. Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. *Nat Immunol* 2005;**6**:57–64.
7. Zheng Y, Niyonsaba F, Ushio H, Ikeda S, Nagaoka I, Okumura K et al. Microbicidal protein psoriasin is a multifunctional modulator of neutrophil activation. *Immunology* 2008;**124**:357–367.
8. Van Den Berg DJ, Sharma AK, Bruno E, Hoffman R. Role of members of the Wnt gene family in human hematopoiesis. *Blood* 1998;**92**:3189–3202.
9. Gaubin M, Autiero M, Basmaciogullari S, Metivier D, Mis hal Z, Culerrier R et al. Potent inhibition of CD4/TCR-mediated T cell apoptosis by a CD4-binding glycoprotein secreted from breast tumor and seminal vesicle cells. *J Immunol* 1999;**162**:2631–2638.
10. Bingle CD, Gorr SU. Host defense in oral and airway epithelia: chromosome 20 contributes a new protein family. *Int J Biochem Cell Biol* 2004;**36**:2144–2152.
11. Giembycz MA, Lindsay MA. Pharmacology of the eosinophil. *Pharmacol Rev* 1999;**51**:213–340.
12. Kubach J, Lutter P, Bopp T, Stoll S, Becker C, Huter E et al. Human CD4+CD25+ regulatory T cells: proteome analysis identifies galectin-10 as a novel marker essential for their anergy and suppressive function. *Blood* 2007;**110**:1550–1558.
13. Devouassoux G, Pachot A, Laforest L, Diasparra J, Freymond N, Van Ganse E et al. Galectin-10 mRNA is overexpressed in peripheral blood of aspirin-induced asthma. *Allergy* 2008;**63**:125–131.
14. van Beijsterveldt CE, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* 2007;**29**:516–521.
15. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 2006;**7**:95–100.
16. Morar N, Willis-Owen SA, Moffatt MF, Cookson WO. The genetics of atopic dermatitis. *J Allergy Clin Immunol* 2006;**118**:24–34; quiz 35–36.
17. Bryborn M, Hallden C, Sall T, Adner M, Cardell LO. Comprehensive evaluation of genetic variation in S100A7 suggests an association with the occurrence of allergic rhinitis. *Respir Res* 2008;**9**:29.
18. Bousquet J, Van Cauwenberge P, Khaltaev N. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol* 2001;**108**:S147–S334.
19. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;**68**:978–989.
20. Sano H, Hsu DK, Yu L, Apgar JR, Kuwabara I, Yamanaka T et al. Human galectin-3 is a novel chemoattractant for monocytes and macrophages. *J Immunol* 2000;**165**:2156–2164.
21. Yamaoka A, Kuwabara I, Frigeri LG, Liu FT. A human lectin, galectin-3 (epsilon bp/Mac-2), stimulates superoxide production by neutrophils. *J Immunol* 1995;**154**:3479–3487.
22. Frigeri LG, Zuberi RI, Liu FT. Epsilon BP, a beta-galactoside-binding animal lectin, recognizes IgE receptor (Fc epsilon RI) and activates mast cells. *Biochemistry* 1993;**32**:7644–7649.
23. Matsumoto R, Matsumoto H, Seki M, Hata M, Asano Y, Kanegasaki S et al. Human ecalectin, a variant of human galectin-9, is a novel eosinophil chemoattractant produced by T lymphocytes. *J Biol Chem* 1998;**273**:16976–16984.
24. Matsumoto R, Hirashima M, Kita H, Gleich GJ. Biological activities of ecalectin: a novel eosinophil-activating factor. *J Immunol* 2002;**168**:1961–1967.
25. Dunphy JL, Barcham GJ, Bischof RJ, Young AR, Nash A, Meeusen EN. Isolation and characterization of a novel eosinophil-specific galectin released into the lungs in response to allergen challenge. *J Biol Chem* 2002;**277**:14916–14924.
26. Benson M, Langston MA, Adner M, Andersson B, Torinsson-Naluai A, Cardell LO. A network-based analysis of the late-phase reaction of the skin. *J Allergy Clin Immunol* 2006;**118**:220–225.
27. Pantanowitz L, Balogh K. Charcot-Leyden crystals: pathology and diagnostic utility. *Ear Nose Throat J* 2004;**83**:489–490.
28. Dor PJ, Ackerman SJ, Gleich GJ. Charcot-Leyden crystal protein and eosinophil granule major basic protein in sputum of patients with respiratory diseases. *Am Rev Respir Dis* 1984;**130**:1072–1077.
29. Ackerman SJ, Liu L, Kwatia MA, Savage MP, Leonidas DD, Swaminathan GJ et al. Charcot-Leyden crystal protein (galectin-10) is not a dual function galectin with lysophospholipase activity but binds a lysophospholipase inhibitor in a novel structural fashion. *J Biol Chem* 2002;**277**:14859–14868.
30. Janson C, Anto J, Burney P, Chinn S, de Marco R, Heinrich J et al. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. *Eur Respir J* 2001;**18**:598–611.
31. Bauchau V, Durham SR. Prevalence and rate of diagnosis of allergic rhinitis in Europe. *Eur Respir J* 2004;**24**:758–764.
32. Olsson P, Berglund N, Bellander T, Stjarne P. Prevalence of self-reported allergic and non-allergic rhinitis symptoms in Stockholm: relation to age, gender, olfactory sense and smoking. *Acta Otolaryngol* 2003;**123**:75–80.