Sequencing the *IL4* locus in African Americans implicates rare noncoding variants in asthma susceptibility

Gabe Haller, BA,* Dara G. Torgerson, PhD, Carole Ober, PhD, and Emma E. Thompson, PhD *Chicago, Ill*

Background: Common genetic variations in the *IL4* gene have been associated with asthma and atopy in European and Asian populations, but not in African Americans.
Objective: Because populations of African descent have increased levels of genetic variation compared with other populations, particularly with respect to low frequency or rare variants, we hypothesized that rare variants in the *IL4* gene contribute to the development of asthma in African Americans.
Methods: To test this hypothesis, we sequenced the *IL4* locus in 72 African Americans with asthma and 70 African American controls without asthma to identify novel and rare polymorphisms in the *IL4* gene that may be contributing to asthma susceptibility.
Results: We report an excess of private noncoding single nucleotide polymorphisms (SNPs) in the subjects with asthma compared with control subjects without asthma ($P = .031$). Tajima's D is significantly more negative in subjects with asthma (–0.375) than controls (–0.073; $P = .04$), reflecting an excess of rare variants in the subjects with asthma.
Conclusion: Our findings indicate that SNPs at the *IL4* locus that are potentially exclusive to African Americans are associated with susceptibility to asthma. Only 3 of the 26 private SNPs (ie, SNPs present only in the subjects with asthma or only in the controls) are tagged by single SNPs on one of the common genotyping platforms used in genome-wide association studies. We also find that most of the private SNPs cannot be reliably imputed, highlighting the importance of sequencing to identify genetic variants contributing to common diseases in African Americans. (J Allergy Clin Immunol 2009;124:1204-9.)

Key words: *Rare variants, private alleles, asthma, IL4, IgE, African Americans*

Abbreviations used
MAF: Minor allele frequency
SNP: Single nucleotide polymorphism

IL-4 is the major T_H2 cytokine that induces the isotype switch to IgE in B lymphocytes[1,2] and is involved in host response to both parasitic infection and allergens.[3] Genetic variations in the *IL4* gene have been associated with expression levels of IL-4[4] as well as susceptibility to atopic diseases, including asthma (see review[5]), and to parasitic infections.[6,7] Even though *IL4* is among the most replicated asthma/atopy susceptibility loci,[5] only 2 studies have included individuals of African descent.[8,9] To date, all or part of the *IL4* gene has been sequenced in 78 European and 50 non-European individuals (38 African American, 6 Asian, 6 Hispanic).[8,10,11] Together, 5 single nucleotide polymorphisms (SNPs) have been identified in the coding region of the gene (3 of which are rare nonsynonymous SNPs).[12] On the other hand, an abundance of noncoding SNPs have been identified at the *IL4* locus, several of which have been implicated in disease risk. For example, 2 common SNPs have been associated with allergic disease (including asthma) in many studies.[5] These polymorphisms, a functional promoter SNP (-589C/T; rs2243250) and an SNP of unknown function in the 5′ untranslated region (-33C/T; rs2070874), are in linkage disequilibrium in European and Asian HapMap samples (Europeans, $r^2 = 1$; Asians, $r^2 = 1$; Africans, $r^2 = 0.18$). Other variations in intron 2 (3017 T/G, rs2227284, and 2 repeat polymorphisms) have also been associated with asthma and IgE in diverse populations. However, none of these common polymorphisms or any of the other tested variants in the *IL4* gene has been associated with disease risk in African Americans when *P* values are corrected for multiple testing.[8,9]

Because populations of African descent have increased levels of genetic variation compared with other populations, particularly with respect to low-frequency or rare variants,[13,14] we hypothesized that rare variants in the *IL4* gene contribute to the development of asthma in African Americans. To test this hypothesis, we sequenced nearly the entire *IL4* gene in 72 African American subjects with asthma and 70 African American control subjects. We report a significant excess of private SNPs (ie, SNPs present only in subjects with asthma or only in controls) and rare noncoding SNPs (minor allele frequency [MAF] < 5%) in subjects with asthma compared with control subjects, supporting the hypothesis that rare variants in the *IL4* gene play an important role in disease susceptibility in African Americans.

METHODS

Study samples

DNA from 142 unrelated African Americans who participate in the Chicago Asthma Genetics study were included in this study. The 72 subjects with asthma met the following criteria: (1) a physician's diagnosis of asthma (with no conflicting diagnosis); (2) the presence of at least 2 self-reported symptoms (cough, wheeze, shortness of breath); (3) current use of asthma medications; (4) either bronchial hyperresponsiveness, defined as a ≥20% decrease in FEV_1 after inhalation of ≤25 mg/mL methacholine, or

From the Department of Human Genetics, University of Chicago.
*G. Haller is currently affiliated with Washington University, Division of Biology and Biomedical Sciences, St Louis, Mo.
Supported by R01 HL72414, U01 HL049596, and M01 RR00055. D.G.T. is supported by R01 HL087665 and U10 HL064313. E.E.T. is supported by T32 HL007605.
Disclosure of potential conflict of interest: The authors have declared that they have no conflict of interest.
Received for publication July 20, 2009; revised September 9, 2009; accepted for publication September 10, 2009.
Available online November 12, 2009.
Reprint requests: Emma E. Thompson, PhD, University of Chicago, Department of Human Genetics, 920 E. 58th Street, Chicago, IL 60637. E-mail: eethomps@uchicago.edu.
0091-6749/$36.00

doi:10.1016/j.jaci.2009.09.013

reversibility to inhaled bronchodilator, defined as a ≥15% increase in baseline FEV_1 after inhalation of a bronchodilator (albuterol) or after treatment; (5) birth weight <4.4 lb; and (6) <3 pack-years of cigarette smoking. The 70 control subjects had no personal or family history of asthma among first-degree relatives and were ≥18 years of age. All subjects with asthma and control subjects reported at least 3 grandparents of African or African American descent. Total serum IgE measurements were available for all but 19 subjects with asthma and 11 controls without asthma; allergen skin prick testing was not performed in the control subjects, and those data therefore were not included in this study.

Sequencing studies

The consensus sequence AF395008 (Genbank) was used to design overlapping primer sets to cover the entire *IL4* gene, including 1743 bp upstream of the ATG start site, and 1724 bp downstream of the last exon (exon 4; see this article's Table E1 in the Online Repository at www.jacionline.org). Nucleotide positions throughout this article are given with respect to the ATG start site unless otherwise noted. PCR was performed in a total volume of 23 μL, with 5 mmol/L deoxyribonucleotide triphosphates (dNTPs), 37.5 mmol/L $MgCl_2$, 12.5 μmol/L each primer, 5 μL 5x Taq Buffer, 0.2 μL GoTaq Flexi DNA polymerase (Promega, Madison, Wis), and 20 ng genomic DNA. Unincorporated nucleotides and excess primers were removed from PCR products by using Exonuclease (New England Biolabs, Ipswich, Mass)/Shrimp Alkaline Phosphatase (USBio, Marblehead, Mass). All amplifications were sequenced in both directions by using BigDye Terminator Sequencing Kits (Applied Biosystems, Foster City, Calif). The chimpanzee consensus sequence (GeneID: 449565) was aligned to the human reference sequence with ClustalW[15] and used to determine ancestral alleles.

SNP identification

The Phred-Phrap-Consed-PolyPhred package was used to assemble the sequences and identify SNPs.[16] All sequences were visually inspected. Because of the sequence overlap, more than 1 call for each genotype was often obtained for each position in a sample.

Genotyping a variable element in intron 3

The genotypes for a variable element in intron 3 (VE6566),[10] a 70-bp copy number variant, were determined by size separation on 3% agarose gels (1-3 copies). DNA was amplified by using PCR primers (Table E1) that flanked the variable element. Genotypes at a TG dinucleotide repeat in the second intron of *IL4*[17,18] were not included in this study because we could not discern genotypes by sequencing or by an electrophoretic gel assay.

Data analysis

Polymorphisms were tested for Hardy-Weinberg equilibrium in the subjects with asthma, the controls, and the combined sample by using Haploview.[19] Single SNPs were tested for association with asthma status using a χ^2 test as implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/).[20] The proportion of private SNPs in subjects with asthma compared with controls was evaluated by permutation test, in which case/control status was permuted 100,000 times in the subjects with asthma and controls (combined), holding the genotypes constant and preserving the pattern of missing data, to build an empirical distribution of the differences in the proportion of private SNPs. As a second test, a weighted-sum statistic was calculated to test for association of the locus as a whole with case status, using the method of Madsen and Browning.[21] In this method, each variant was weighted by its MAF in unaffected individuals, and then individuals were given a score consisting of the sum of their weighted alleles (with rarer alleles given greater weight). Individuals were ranked on the basis of their score, and a sum of ranks for affected individuals was calculated. Significance was determined by permuting case/control status 100,000 times to produce an empirical distribution of summed ranks, and to preserve the pattern of missing data. For all permutation tests, a threshold for statistical significance was set at P <.05 (ie, less than 5000 of the 100,000 permutations were greater than the observed sum of ranks for affected individuals).

Two measures of nucleotide diversity are commonly used to compare SNP frequencies among samples of varying size and DNA fragment length: the average number of pairwise differences in a given set of chromosomes (π),[22] and nucleotide diversity estimated from the allele frequency of the polymorphic sites (θ_w).[23] The difference between these 2 estimates (relative to their SE) is expressed as Tajima's D, which is expected to be 0 under neutrality. A positive Tajima's D reflects a larger π than θ_w and indicates an excess of intermediate-frequency variants, whereas a negative value (larger θ_w than π) indicates an excess of rare variants. Significant deviations from 0 in either direction indicate a skew in the allele frequency spectrum, which can be a sign of nonneutral evolution (ie, selection) or demographic events (ie, population history). Permutation tests were used to assess whether Tajima's D was significantly more negative in subjects with asthma compared with controls. A total of 100,000 permutations were conducted by randomly sampling 72 individuals without replacement to represent subjects with asthma from the pooled set of cases and controls. The remaining sample of 70 individuals was then taken to represent controls. For each permutation, Tajima's D was calculated for the sampled subjects with asthma and controls, and the difference between their values was calculated to assess the probability that we have observed a larger difference between the values of Tajima's D for subjects with asthma and controls than expected by chance.

Admixture estimates

European admixture in 112 African American subjects (54 cases, 58 controls) was estimated at both the genomic and local scales by using genotypes from more than 1 million SNPs on the Illumina Human 1 M array. Genomic European admixture was estimated from the first principal component in a principal component analysis including the HapMap CEU and YRI as reference populations in EIGENSTRAT.[24] Local European admixture at the *IL4* locus was estimated using 63,598 SNPs on chromosome 5 with the program Local Ancestry in adMixed Populations (LAMP)[25] by assuming 20 generations of European admixture, a constant recombination rate of 10^{-7}, and a population admixture rate of 81% obtained from genomic admixture estimates. Wilcoxon rank sum tests (WRSTs) with continuity corrections were used to compare percent European ancestry in subjects with asthma and controls, and between individuals who harbored private SNPs and those who did not. Genomic European admixture did not differ between subjects with asthma and controls (P = .64; see this article's Fig E1 in the Online Repository at www.jacionline.org) or between individuals who carried private polymorphisms and those who did not (Wilcoxon rank-sum test; P = .83). At the local scale, there was also no significant difference in European admixture between subjects with asthma and controls (P = .28; see this article's Fig E2 in the Online Repository at www.jacionline.org), or between individuals who carried private polymorphisms and those who did not (P = .49).

Estimates of linkage disequilibrium and SNP imputation

Linkage disequilibrium between SNPs identified from the sequencing of *IL4* and SNPs genotyped on the Illumina 1 M genotyping platform in 112 African American individuals (54 cases, 58 controls), was estimated by using Haploview 4.1.[19] The *IL4* SNPs from sequencing were considered to be tagged by SNPs on the Illumina 1 M if they demonstrated an r^2 value >0.5 and were within 500 kb of the *IL4* locus. To evaluate the potential for imputing the *IL4* SNPs identified by sequencing, we generated 2 datasets (a query and reference panel) including the same 112 African American individuals who were typed on the Illumina 1 M platform. The query dataset included only the 588 SNPs typed on the Illumina 1 M platform that were within 500 kb of the *IL4* transcription start and stop site. The reference panel included both the 588 SNPs from the Illumina 1 M and the *IL4* SNPs from the sequencing of the same individuals. The *IL4* SNPs in the query dataset were imputed with Mach 1.0[26] by using 50 iterations of the Markov sampler and considering 200 haplotype states when updating each individual. The accuracy in imputing