Evolutionary and Functional Evidence for Positive Selection at the Human *CD5* Immune Receptor Gene

Elena Carnero-Montoro,[1,†] Lizette Bonet,[2,†] Johannes Engelken,[1,3] Torsten Bielig,[2] Mario Martínez-Florensa,[2] Francisco Lozano,[2,4,5,‡] and Elena Bosch[1,‡,*]

[1]Institut de Biologia Evolutiva (UPF-CSIC), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, 08003 Barcelona, Spain

[2]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centre Esther Koplowitz, 08036 Barcelona, Spain

[3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

[4]Servei d'Immunologia, Hospital Clínic, 08036 Barcelona, Spain

[5]Departament de Biologia Cel·lular, Immunologia i Neurociències, Facultat de Medicina, Universitat de Barcelona, 08036 Barcelona, Spain

†These authors contributed equally to the work.

‡These authors should be considered as senior coauthors of this work.

***Corresponding author:** E-mail: elena.bosch@upf.edu.

Associate editor: Anne Stone

Abstract

CD5 is a lymphocyte surface coreceptor of still incompletely understood function. Currently available information indicates that CD5 participates not only in cell-to-cell immune interactions through still poorly defined endogenous ligands expressed on hemopoietic and nonhemopoietic cells but also in recognition of exogenous and highly conserved microbial structures such as fungal β-glucans. Preceding single nucleotide polymorphism (SNP) data analysis provided evidence for a recent selective sweep in East Asia and suggested a nonsynonymous substitution at position 471 (A471V; rs2229177) of the cytoplasmatic region of the CD5 receptor as the most plausible target of selection. The present report further investigates the role of natural selection in the *CD5* gene by a resequencing approach in 60 individuals representing populations from 3 different continents (20 Africans, 20 Europeans and 20 East Asians) and by functionally assaying the relevance of the A471V replacement on CD5 signaling. The high differentiation pattern found at the nonsynonymous A471V site together with the low diversity, most of the performed neutrality tests (Tajima's *D*, Fu and Li's *F** and *D**, and Fu's *Fs*) and the predominance of a major haplotype in East Asians strongly argue in favor of positive selection for the A471V site. Importantly, anti-CD5 monoclonal antibody cross-linking unveiled significant differences among A471V variants regarding the mitogen-activated protein kinase (MAPK) cascade activation on COS7 and on human peripheral blood mononuclear cells. Similar differences on MAPK activation and IL-8 cytokine release were also observed upon exposure of HEK293 cell transfectants expressing the A471V variants to Zymosan, a β-glucan–rich fungal particle. Taken together, the results provide evidence for the hypothesis of an adaptive role of the A471V substitution to environmental challenges, most likely infectious pathogens, in East Asian populations.

Key words: CD5 receptor, human adaptation, positive selection, resequencing data.

Introduction

CD5 is a monomeric type I transmembrane glycoprotein of 67-kDa belonging to group B of the scavenger receptor cysteine-rich (SRCR) superfamily (Sarrias et al. 2004). It is constitutively expressed on thymocytes from early stages of development, mature peripheral T cells, and mature B-1a cells, as well as on activated conventional B2 cells (Youinou et al. 1999). CD5 associates with the antigen-specific receptors present on T and B cells (Osman et al. 1992; Lankester et al. 1994), and it is therefore well-positioned to modulate signaling responses (Tarakhovsky et al. 1995; Bikah et al. 1996) necessary for proper T- and B-cell activation and differentiation.

CD5 is composed of three extracellular SRCR domains, a hydrophobic transmembrane domain and a large cytoplasmic region that lacks intrinsic catalytic activity but contains several conserved motifs involved in signal transduction and likely mediating protein–protein interactions (Lozano et al. 2000). In the fully-processed mature protein, important signaling motifs are located around tyrosine residues Y429 and Y463, which were both shown to be targets for protein tyrosine kinase Lck (Vila et al. 2001), and serine residues S459 and S461, which are targets for casein kinase II (CKII) (Calvo et al. 1999). The main function of CD5 has been established as the negative regulation of T-cell receptor (TCR) and B-cell receptor (BCR) signaling through activation of negative regulators such as CKII (Raman et al. 1998) or Src homology domain-bearing protein phosphatase-1 (Perez-Villar et al. 1999). However, recent studies reveal a broader function for CD5 that includes its role as regulator of lymphocyte survival, regulator of peripheral tolerance, and as receptor for pathogen-associated

molecular patterns (Soldevila et al. 2011). Evidence coming from experimental autoimmune and tumor mouse models indicates that CD5 promotes T-cell survival by downmodulating activation-induced cell death (AICD) (Axtell et al. 2004; Friedlein et al. 2007). Upregulated CD5 expression has also been reported a characteristic of regulatory lymphocytes (Treg and B10 cells) (Yanaba et al. 2008; Ordonez-Rueda et al. 2009) as well as anergized T and B cells (Hippen et al. 2000; Stamou et al. 2003). As reported for some members of the SRCR superfamily, it has been recently shown that CD5 works as a pattern recognition receptor by sensing the presence of conserved fungal (namely β-glucans) but not bacterial cell wall components (Vera et al. 2009). All these reported properties make CD5 a relevant immune receptor both under physiologic and pathologic conditions of autoimmune, neoplastic, or infectious origin (Soldevila et al. 2011).

A search for recent signatures of selection in 11 genomic regions centered on human fast-evolving genes in the Human Genome Diversity Panel (HGDP-CEPH) (Cann et al. 2002) revealed strong evidence for a recent selective sweep in the *VPS37C* region in most East Asian populations (Moreno-Estrada et al. 2009). In that study, SNP genotyping data were analyzed by population differentiation, allele frequency threshold analyses, and two complementary linkage disequilibrium (LD)–based approaches. The same genomic region had already been reported to show evidence for a recent complete selective sweep in Chinese samples when using a composite likelihood ratio approach on genome-wide SNP data (Williamson et al. 2007). Although *VPS37C* was the nearest gene to the composite likelihood estimate of the sweep position, the implication of the *CD5* gene (just 2.4-kb upstream from *VPS37C*) or other nearby genes could not be discarded. Despite the limitations to accurately locating the actual target of selection, the exhaustive search for functional variants putatively linked to the signal of selection from publicly available SNP data led us to suggest a nonsynonymous SNP (rs2229177) as the most plausible candidate (Moreno-Estrada et al. 2009). This nonsynonymous SNP implies a Val to Ala substitution in the cytoplasmatic region of the CD5 receptor (A471V, which corresponds to amino acid position 447 in the fully-processed mature protein). Moreover, two different reviews (Akey 2009; Barreiro and Quintana-Murci 2010) on available genome-wide scans for selection using HapMap or Perlegen SNP data consistently replicated significant LD and population differentiation patterns along the aforementioned genomic region in their respective samples of East Asian ancestry. A common feature of all these studies is that all are based on SNP data with a possible ascertainment bias (Clark et al. 2005) toward nonrare variants.

In the present study, the *CD5* gene was resequenced in Yorubas, Europeans and East Asians with the specific aim of confirming previous evidence for selection using the complete allele spectrum and therefore overcoming the possible ascertainment biases obtained from SNP data. Next, the amino acid variability of the complete CD5 lymphocyte receptor was examined and the functional relevance of the A471V replacement that was putatively linked to the signature of selection found in East Asia was explored. We hypothesize that the distinct signaling capabilities of the variants (A471V) of the CD5 immune receptor may have driven the signal of selection in the East Asian population.

Materials and Methods

Population Samples

Sixty HapMap DNA samples were obtained from Coriell Cell Repositories (Camden, NJ) for resequencing. These consisted of 20 Yorubas (from Ibadan in Nigeria), 20 Europeans (Utah residents with ancestry from Northern and Western Europe), and 20 Asians (10 Japanese from Tokyo and 10 Chinese from Beijing). Coriell Repository numbers for these samples are as follows: Yoruban (NA18501, NA18502, NA18507, NA18508, NA18855, NA18856, NA18861, NA18862, NA19127, NA19128, NA19137, NA19138, NA19171, NA19172, NA19203, NA19204, NA19206, NA19207, NA19209, NA19210), European (NA06994, NA07000, NA07345, NA07357, NA11829, NA11830, NA11839, NA11840, NA11992, NA11993, NA12003, NA12004, NA12043, NA12044, NA12056, NA12057, NA12750, NA12751, NA12812, NA12813), Han Chinese (NA18576, NA18577, NA18579, NA18582, NA18593, NA18623, NA18624, NA18632, NA18635, NA18636), and Japanese (NA18940, NA18942, NA18943, NA18944, NA18948, NA18949, NA18951, NA18956, NA18970, NA18973).

The H971 subset (Rosenberg 2006) of the HGDP-CEPH, which represents 51 globally distributed populations (Cann et al. 2002) was genotyped for rs2229177. In order to maximize sample sizes, individuals were regrouped into 39 populations based on geographic and ethnic criteria as in Gardner et al. (2006). For part of the analysis, populations were further grouped into seven geographical regions (see supplementary table S1, Supplementary Material online): Sub-Saharan Africa, Middle East-North Africa, Europe, Central-South Asia, East Asia, Oceania and America.

All samples used were originally collected with proper informed consent.

Sequencing and Genotyping

Three nonoverlapping regions of 1,740, 1,754, and 2,216 bp, covering 8 out of 10 coding exons of *CD5* (supplementary fig. S1, Supplementary Material online), were amplified by using the primers shown in supplementary table S2, Supplementary Material online. Resequencing was directed toward the exonic regions found in the fully processed form of the protein, although flanking intronic regions were also included. Polymerase chain reactions (PCRs) were performed in a total volume of 25 μl, containing 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.5 μM of each primer, 1× buffer, 0.05 U Taq polymerase (Ecogen), and 10-ng genomic DNA. PCR conditions were as follows: 3 min at 94 °C, 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 3 min; and a final step of extension of 5 min at 72 °C. PCR products were

purified using MultiScreen PCRμ96 filter plates (Millipore) according to the manufacturer's protocol. Purified PCR products were sequenced using Big Dye Terminator chemistry version 3.1 (Applied Biosystems) with sequencing and amplification primers (supplementary table S2, Supplementary Material online). Sequencing products were purified using the Montage SEQ96 Cleanup kit (Millipore) and run on an ABI 3730 XL sequencer. Sequence assembly, visualization, and editing were performed with the Seqman module within the DNASTAR Lasergene software version 7.1.0. In each final individual assembly, we used as a reference a concatenated sequence based on the human hg19/CRGh37 built, which contained exclusively those regions sequenced in all the individuals (supplementary figs S1 and S2, Supplementary Material online). Individual consensus sequences were visually inspected at least twice, and all detected polymorphic sites were manually checked. Additional curation of the obtained consensus sequences included comparison with the corresponding HapMap II genotypes (www.hapmap.org) (Frazer et al. 2007). Nucleotide sequences determined are available from GenBank database under accession numbers JN159734–JN159853.

Rs2229177 was genotyped using a TaqMan SNP Genotyping assay (Applied Biosystems). The assay was performed on an Applied Biosystems 7900HR system in a 384-well format using the TaqMan Universal PCR Master Mix and the corresponding predesigned TaqMan probes (assay ID C__3237272_10) following the manufacturer's recommendations.

Statistical Analysis

Genotype data for all detected polymorphic sites were collected and stored within the SNPator web environment (http://bioinformatica.cegen.upf.es) (Morcillo-Suarez et al. 2008), where part of the analysis such as genotype and allele frequencies and Hardy–Weinberg equilibrium tests were performed. Haplotypes were inferred from genotype data with PHASE 2.1 (Stephens et al. 2001) using the default parameter set with 1,000 interactions. Nucleotide diversity statistics, analysis of population polymorphic sites, Tajima's *D*, Fu and Li's *F** and *D**, Fu's *Fs* and Fay and Wu's *H* tests were carried out using custom Java scripts (Ramirez-Soriano et al. 2008). An orthologous concatenated sequence was built from the chimpanzee genome (panTro2, March 2006 assembly) and used in the tests requiring an outgroup species.

The significance for each neutrality test was evaluated by means of 10,000 coalescence simulations using Cosi version 1.1 and considering the ad hoc, continent-specific human demographic calibration as described in Schaffner et al. (2005). Coalescent simulations were conditioned on the actual number of base pairs analyzed, sample size and number of segregating sites. The local recombination rate of 0.69 cM/Mb for the genomic region was obtained from HapMap Phase II data (www.hapmap.org). Given that the actual resequenced region is shorter (4,104 bp) than the total genomic region (7,840 bp), a proportionally increased rate of 1.32cM/Mb was used in simulations to obtain the same expected total number of recombinations. We evaluated the significance of neutrality tests based on the 95% and 99% empirical quantiles simulated for each population. Significance for the whole human sample was evaluated by means of coalescence neutral simulations considering constant population size using the DnaSP software version 5.00 (Librado and Rozas 2009).

Arlequin (Excoffier et al. 2005) was used to calculate global F_{ST} values among the three analyzed populations with a locus-by-locus analysis of molecular variance (Excoffier et al. 1992). Relationships among nonunique haplotypes were investigated using the median-joining network algorithm (Bandelt et al. 1995; Bandelt et al. 1999) within the Network 4.5.0.1 software package (http://www.fluxus-engineering.com). We considered as ancestral alleles those recovered from the chimpanzee genome (panTro2, March 2006 assembly). In silico prediction of the possible impact on the structure and function of the CD5 receptor was explored for each nonsynonymous variant using the PolyPhen tool (http://genetics.bwh.harvard.edu/pph/) (Ramensky et al. 2002).

Generation and Expression of cDNA Constructs for CD5 Variants A471V

The construct coding for the Val variant (V471) was generated as previously described (Simarro et al. 1997). The construct carrying the Ala variant (A471) was generated by gene synthesis (Genscript Co) and was identical in size and sequence except for the 471 codon (GTT → GCT). Both cDNAs were cloned into the pHβApr-1-*neo* mammalian expression vector as *Eco*RI restricted fragments.

For transient transfection purposes, 2×10^5 COS7 (American Tissue Culture Collection) or HEK293 cells (a kind gift of Dr Golenbock, University of Massachusetts Medical School, Worcester, MA) were seeded onto six-well plates (Techno Plastic Products) 24 h before transfection and grown at 37°C in a 5% CO_2 atmosphere in Dulbecco's-Modified Eagle Medium Nutrient Mixture F-12 (DMEM/F12) supplemented with Glutamax (Gibco), 10% fetal calf serum (BioWhitaker), 100 U/ml penicillin, and 50 μg/ml streptomycin. Cell transfection was performed by using Lipofectamine 2000 (Invitrogen Life Technologies) according to manufacturer's instructions. Briefly, 2.5 μl of Lipofectamine 2000 was mixed with 1 μg of the mammalian expression vectors coding for the A471V variants and added to each well in the absence of antibiotics. After 5-h incubation at 37°C, the mix was removed and replaced by the above-mentioned culture medium and left in culture for 48 h until assayed.

Assays for Mitogen-Activated Protein Kinase Activity

For testing mitogen-activated protein kinase (MAPK) activity, cells were serum-starved for 4 h in serum-free RPMI 1640 medium supplemented with L-glutamine (BioWhittaker) or DMEM/F12 prior the assay. The cells used were either fresh isolated peripheral blood mononuclear cells (PBMC) obtained from healthy volunteers by standard

density gradient centrifugation over Ficoll (density 1.077 g/cm^3) (Linfosep; Biomedics) or COS7 and HEK293 cell transfectants at 48-h post-transfection. Cells were incubated for 15–60 min at 37°C with different stimuli: cross-linking with 10 μg/ml of mouse antihuman CD5 83-C4 monoclonal antibody (mAb) (Calvo et al. 1999), 60 ng/ml phorbol 12-myristate 13-acetate (PMA) (Sigma-Aldrich), and 40 μg/ml of Zymosan A from *Saccharomyces cerevisiae* (Sigma-Aldrich). After stimulation, cells were washed once with cold phosphate-buffered saline (PBS), solubilized with lysis buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid, 1 mM ethylene glycol tetraacetic acid, 1% Nonidet P-40, 1 mM Na_3VO_4, 1 mM phenylmethylsulfonyl fluoride) containing complete protease inhibitor cocktail (Roche) and kept at -80°C for further processing. The nuclei and cell debris were removed by centrifugation at 14,000 rpm for 10 min at 4°C, and protein concentration in supernatants was measured by using the BCA Protein Assay Reagent Kit (Pierce) following the manufacturer's instructions. Lysate samples (3-μg protein) were separated by 8% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) under reducing conditions and analyzed by Western blot with rabbit antisera against phosphorylated (Cell Signaling) and total extracellular signal-regulated kinases (ERK)1/2 (Sigma-Aldrich) following manufacturer's instructions. Briefly, nitrocellulose membranes were blocked with 5% bovine serum albumin (BSA, Sigma-Aldrich) in tris-buffered saline and Tween 20 (TBS-T) (10 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.5) for 1 h at room temperature and then incubated overnight at 4°C with diluted antisera in the same blocking solution. After extensive washing in TBS-T, blots were developed with antirabbit IgG horseradish peroxidase–linked whole antibody (GE Healthcare) and Super Signal West Dura Extended Duration Substrate (Pierce). The blots were analyzed with a Luminescent Image Analyzer LAS 3000 (Fujifilm), and densitometric analysis was performed using the Image J program (National Institutes of Health, http://rsb.info.nih.gov/ij/). The results are expressed as the ratio of the intensity of the bands corresponding to phosphorylated (p) and total (t) ERK. This value was always corrected by intensity of the bands corresponding to the CD5 surface expression in each cell population, assessed as indicate below.

IL-8 Cytokine Release Assay

For the IL-8 release assay, transient HEK293 transfectants expressing any of the two A471V variants were serum-starved for 4 h after 48 h of transfection and then incubated with 20 μg/ml of Zymosan for 24 h. Next, supernatant samples (100 μl) were collected and assayed for human IL-8 by enzyme-linked immunosorbent assay (ELISA) following the manufacturer's instructions (BD OptEIA, Human IL8 ELISA Set, BD Biosciences, San Diego, CA).

Assessment of Cell Surface CD5 Expression

The level of surface CD5 expression in PBMC and cell transfectants was assessed by cell surface labeling of intact cells with biotin (EZ-Link Sulfo-NHS-LC-LC–biotin, Thermo Scientific) and further Western blot analysis of CD5 immunoprecipitates. Biotinylation of COS7 and HEK293 transfectants (confluent monolayers of a six-well plate) or PMBC (2×10^6) cell suspensions washed with ice-cold PBS was done by incubation with 1 ml of biotin solution (0.25 mg/ml in PBS) for 20–25 min on ice under gentle shaking. After that, the solution was discarded, and the cells were resuspended in 1 ml of RPMI 1640 or DMEM/F12 for 5 min on ice, then washed twice with ice-cold PBS and disrupted in 600-μl lysis buffer. Following removal of nuclei and cell debris by centrifugation at 14,000 rpm for 10 min at 4°C, detergent solubilizates were immunoprecipitated overnight at 4°C using 1.5 μg of mouse antihuman CD5 Cris-1 mAb (a kind gift of R. Vilella, Immunology Department, Hospital Clínic of Barcelona) plus 20 μl of 50% slurry of Protein A Sepharose 4 Fast Flow (GE Healthcare Bio-Sciences AB). Subsequently, the immune complexes were washed several times with lysis buffer and 50 mM Tris (pH 7.5) and then eluted in nonreducing loading buffer (for 10 min at 100°C) for further separation in 8% SDS–PAGE and transfer to nitrocellulose (Bio-Rad). Membranes were blocked overnight at 4°C with 5% BSA in TBS-T and then developed with peroxidase-labeled Streptavidin (Roche) and Super Signal West Dura Extended Duration Substrate (Pierce). The intensity of the bands was analyzed with a Luminescent Image Analyzer LAS 3000 and the Image J program.

Results

Pattern of CD5 Sequence Variation

A noncontiguous 4,104-bp region encompassing exons 3–10 and their immediately intronic flanking regions of the *CD5* gene (fig. 1 and supplementary fig. S1, Supplementary Material online) were resequenced in 20 Yorubas, 20 Europeans, and 20 Asians. A total of 27 polymorphic sites, comprising 17 intronic, 3 synonymous, and 7 nonsynonymous substitutions, were detected (table 1). After Bonferroni correction, no polymorphic site was significantly out of Hardy–Weinberg equilibrium. Only 5 of the 27 polymorphic loci found here had not been previously reported in the National Center for Biotechnology Information (NCBI) Single Nucleotide Polymorphism database (dbSNP) Build 132. Ten substitutions were singletons: four were found to be specific to Yorubas, five to Europeans, and one to East Asians. From all nonsynonymous substitutions, only SNP7 (rs2241002) and SNP26 (rs2229177) showed intermediate-high allele frequencies. For each polymorphic site, global F_{ST} values among the three analyzed populations are shown in table 1. Two polymorphisms (including the nonsynonymous SNP26) showed F_{ST} values above the 95th percentile of a global genome-wide F_{ST} distribution obtained across the same three populations using the HapMap2.1 SNP data release (Barreiro et al. 2008).

Inferred haplotypes and their absolute frequencies are shown in table 2. Half of the chromosomes analyzed belong to haplotype H12, which is the only haplotype shared among the three populations studied. Thirteen haplotypes

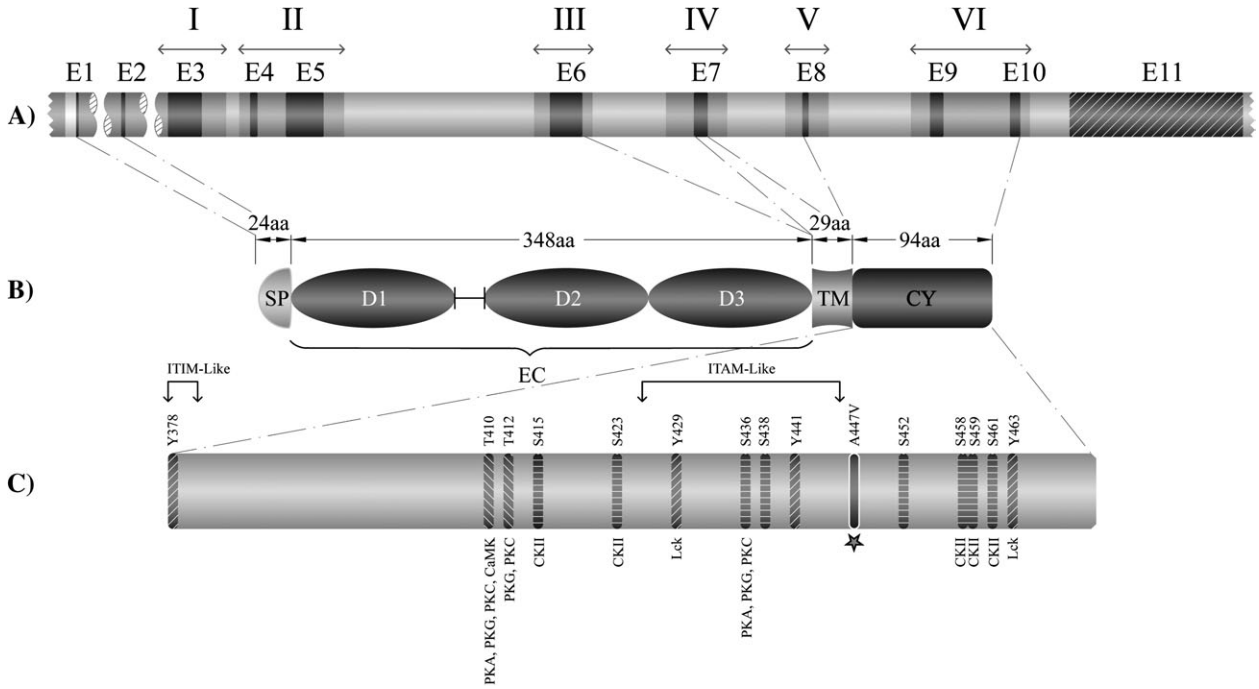

Fig. 1. Gene and protein structure of the human CD5 receptor. (*A*) *CD5* gene structure and resequenced regions. Double arrows from I to VI represent each of the resequenced genomic segments. (*B*) Protein structure of the CD5 receptor. SP, signal peptide; EC, extracellular region; D1–D2–D3, EC SRCR domains; TM, transmembrane region; and CY, cytoplasmic region. (*C*) Cytoplasmatic region showing conserved signaling motifs and the location of the A471V substitution.

were specific to Yorubas, six to Europeans, and four to East Asians. Phylogenetic relationships among the identified haplotypes were inferred through a median-joining network analysis (fig. 2). Although a reticulated pattern in some regions of the network points to the action of recurrent mutation or recombination, an evolutionary pathway among the different haplotypes can easily be visualized. The ancestral haplotype considering the chimpanzee allele in each human polymorphic site divides the network in two main branches, one of which contains only African haplotypes. While Yoruba and European haplotypes are widely distributed in the rest of the network, all East Asians cluster together around H12. In fact, H12 contains all but four East Asian chromosomes, which are one or two mutation steps away from it. Moreover, the derived allele of SNP26 (rs2229177) and the ancestral allele of SNP7 (rs2241002), which are two nonsynonymous sites, characterize all but one *CD5* haplotypes in East Asia.

Summary statistics of genetic diversity and neutrality tests for the *CD5* gene are presented in table 3. The numbers of segregating sites and haplotypes, as well as the nucleotide and haplotype diversities, were higher in Yorubas than in the other populations. Furthermore, all diversity parameters were extremely reduced in East Asia. To test whether patterns of DNA sequence variation in *CD5* fit expectation under neutrality, we analyzed the sequences in each of the three populations, as well as in the whole human sample with Tajima's *D*, Fu and Li's *F** and *D**, Fu's *Fs*, and Fay and Wu's *H* test statistics. Demographic factors such as bottlenecks and population growth can produce patterns of nucleotide variation resembling those of selective sweeps (Ramirez-Soriano et al. 2008). Given that several population demographic parameters have been previously characterized for the three HapMap populations studied here to best fit their genomic patterns of variation (Schaffner et al. 2005), we evaluated neutrality test significance through coalescent simulations considering each particular population demographic history. While the diversity pattern found for Yorubans and Europeans is compatible with neutrality, East Asians displayed significant departures in all but one of the tests performed ($P < 0.005$). The exception was Fay and Wu's *H* test, which compares the fit of the observed derived allele frequency spectrum with that expected under neutrality (Fay and Wu 2000). However, the fact that the most common *CD5* haplotype in East Asians (H12) presents derived alleles for only 5 out of the 27 human polymorphic sites found (table 2) could explain why a significant skew toward high-frequency derived alleles is not detected.

Nonsynonymous Variants in the *CD5* Gene

In silico functional characterization for nonsynonymous variants in the *CD5* gene region is presented in table 4. In addition to all nonsynonymous changes detected in the resequencing analysis, those nonsynonymous SNPs reported in NCBI dbSNP Build 132 for the *CD5* gene were also included. Among the nonbenign SNPs, SNP19 and SNP22 (rs34209302) were only found at low frequency in Africans and thus they are not expected to have caused the apparent selective sweep in Asians. Similarly, there was only one minor allele count for rs117045863 in a sample of 120 chromosomes of European ancestry. On the contrary, SNP7

Table 1. Polymorphisms Detected in the *CD5* Gene.

SNP	Nucleotide position[a]	dbSNP code	SNP alleles[b]	Type of change[c]	Derived allele frequencies			F_{ST}
					Yoruban $2N = 40$	European $2N = 40$	East Asian $2N = 40$	
1	60885986	rs114893974	G/A	I	0.100	0	0.025	0.042
2	60886412		G/A	Syn	0	0.025	0	0
3	60886519	rs79811054	T/C	I	0.100	0.100	0.025	0.002
4	60886566		C/A	I	0	0.025	0	0
5	60886683	rs76108602	C/G	I	0.075	0	0.025	0.020
6	60886850		A/G	NS (N/S)	0	0	0.025	0
7	60886913	rs2241002	C/T	NS (P/L)	0.325	0.125	0.025	0.144
8	60889042	rs17155684	G/T	I	0.675	1	1	0.308
9	60889159	rs35979264	G/A	Syn	0.325	0	0	0.308
10	60889246	rs34821010	C/T	Syn	0.025	0	0	0
11	60889398	rs7108787	G/A	I	0.325	0	0	0.308
12	60890292	rs115588172	G/A	I	0.025	0	0	0
13	60890408	rs117646548	G/A	NS (A/T)	0	0.025	0	0
14	60890563	rs4245224	C/T	I	0.475	0.475	1	0.339
15	60890608	rs11230602	C/T	I	0.250	0	0	0.231
16	60891234		T/C	I	0.025	0	0.025	−0.013
17	60891283	rs7103357	G/A	I	0.525	0.800	1	0.276
18	60891310	rs7106643	C/T	I	0.025	0.275	0	0.217
19	60891358		C/T	NS (R/C)	0.025	0	0	0
20	60891421	rs12364279	C/T	I	0.125	0.200	0	0.079
21	60891530	rs111549923	C/T	I	0.025	0	0	0
22	60892605	rs34209302	C/T	NS (R/C)	0.075	0	0	0.051
23	60892606	rs637186	G/A	NS (R/H)	0	0.025	0	0
24	60892674	rs75571803	G/A	I	0.075	0	0	0.051
25	60892747	rs78926081	G/A	I	0.975	1	1	0
26	60893235	rs2229177	C/T	NS (A/V)	0.475	0.450	1	0.352
27	60893322	rs574843	G/C	I	0	0.025	0	0

[a] SNP positions are based on NCBI Build 37.1.
[b] Observed alleles are indicated as ancestral/derived.
[c] Type of change: I, intronic; Syn, synonymous substitution; and NS, nonsynonymous substitution (residue change).

(rs2241002) and SNP26 (rs2229177) are much more polymorphic, with SNP26 showing both an extreme F_{ST} value across the three analyzed populations and a higher population differentiation between Asians and non-Asians. Interestingly, SNP26 (rs2229177) leads to an Ala → Val substitution at amino acid 471 (A471V) in the near proximity of cytoplasmic motifs critical for signal transduction via CD5 (namely Y487, S483, and S485, these corresponding to positions Y463, S459, and S461 in the fully processed mature protein) (Raman et al. 1998; Vila et al. 2001).

The high population differentiation of this nonsynonymous A471V substitution, its involvement in defining the putative selected haplotype in East Asia (H12), together with its localization in the cytoplasmatic portion of the receptor led us to further explore its geographical distribution in the HGDP-CEPH diversity panel (Cann et al. 2002) as well as its functional relevance by means of several in vitro approaches. Allele frequencies for the A471V site in the HGDP-CEPH diversity panel (fig. 3 and supplementary table S1, Supplementary Material online) displayed a geographical pattern similar to that previously obtained from the three HapMap populations: all East Asian populations (comprising Northeast China with $2N = 90$, South China with $2N = 134$, Han with $2N = 90$, Yakut with $2N = 48$, Cambodians with $2N = 20$, and Japanese with $2N = 60$) presented derived allele frequencies ≥ 0.922, whereas in Europe and Africa (without considering the small San sample) derived allele frequencies ranged from 0.385 in the Biaka Pygmies ($2N = 52$) to 0.66 in Russians ($2N = 50$). Furthermore, allele frequencies in populations from Middle East and North Africa, Central South Asia, Oceania, and America were intermediate to their immediate geographical neighboring regions.

The mAb-Induced CD5 Cross-Linking Reveals MAPK Activation Differences among A471V Variants

CD5 cross-linking by specific mAbs is known to induce both TCR-dependent and -independent CD5-mediated early intracellular signaling events, which include activation of integral components of the MAPK cascade (Simarro et al. 1999). Thus, the possible existence of different signaling capabilities among the A471V variants was first analyzed on transiently transfected COS7 cells expressing CD5 forms differing at this single amino acid position but presenting the remaining amino acid composition common to haplotypes H11 and H12. Cell lysates obtained at different time points (0, 30, 60 min) poststimulation with a mouse antihuman CD5 mAb (83C4, 10 μg/ml) or PMA (60 ng/ml) were assayed by Western blot for levels of phosphorylated and total ERK1/2 (pERK/tERK). As illustrated by figure 4A, COS7 transfectants expressing the V471 variant showed a significant increase in the pERK/tERK ratio at 30 min post-mAb–induced CD5 cross-linking, which declined

Table 2. *CD5* Haplotypes and Their Absolute Frequencies

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Africa	Europe	East Asia	Total
Anc	G	G	T	C	C	A	C	G	G	C	G	G	G	C	C	T	G	C	C	C	C	C	G	G	G	C	G				
H1	·	·	·	·	·	·	·	·	A	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	—	—	1
H2	·	·	·	·	·	·	·	·	A	·	A	·	·	T	·	·	·	·	·	·	·	·	·	·	A	·	·	2	—	—	2
H3	·	·	·	·	·	·	·	·	A	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	A	·	·	7	—	—	7
H4	·	·	·	·	·	·	·	·	A	·	A	·	·	·	T	C	·	·	·	·	·	·	·	·	A	·	·	1	—	—	1
H5	·	·	·	·	·	·	·	·	A	·	A	A	·	·	T	·	·	·	·	·	·	·	·	·	A	·	·	1	—	—	1
H6	·	·	·	·	·	·	·	·	A	·	A	·	·	·	T	·	·	·	T	·	·	·	·	·	A	·	·	1	—	—	1
H7	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	A	·	·	—	1	—	1
H8	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	A	T	·	·	·	·	·	·	A	·	·	—	10	—	10
H9	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	A	T	·	·	·	·	A	·	A	·	C	—	1	—	1
H10	·	·	·	·	·	·	·	T	·	·	·	·	·	T	·	·	A	T	·	·	·	·	·	·	A	·	·	1	—	—	1
H11	·	·	·	·	·	·	·	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	·	·	—	1	—	1
H12	·	·	·	·	·	·	·	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	7	17	36	60
H13	·	·	·	·	·	G	·	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	—	—	1	1
H14	·	·	·	·	·	·	·	T	·	·	·	·	·	T	·	C	A	·	·	·	·	·	·	·	A	T	·	—	—	1	1
H15	A	·	·	·	·	·	·	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	1	—	—	1
H16	A	·	·	·	G	·	·	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	—	—	1	1
H17	·	·	·	·	·	·	T	T	·	·	·	·	·	T	·	·	A	·	·	·	T	·	·	·	A	T	·	1	—	—	1
H18	·	·	·	·	·	·	T	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	7	1	—	8
H19	·	·	C	·	·	·	T	T	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	A	T	·	—	—	1	1
H20	·	·	C	·	·	·	T	T	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	A	T	·	3	—	—	3
H21	·	·	C	·	·	·	T	T	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	A	·	·	—	1	—	1
H22	·	·	·	·	·	·	T	T	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	A	·	·	1	—	—	1
H23	·	·	C	·	·	·	T	T	·	·	·	·	·	·	·	·	·	·	·	T	·	·	·	·	A	·	·	1	2	—	3
H24	·	A	C	A	·	·	T	T	·	·	·	·	A	·	·	·	·	·	·	T	·	·	·	·	A	·	·	—	1	—	1
H25	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	T	·	·	·	·	A	·	·	1	5	—	6
H26	A	·	·	·	G	·	·	T	·	·	·	·	·	·	·	·	·	·	·	T	·	T	·	A	A	·	·	3	—	—	3
H27	·	·	·	·	·	·	·	T	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	·	·	1	—	—	1

NOTE.—Each polymorphic variant is displayed below the corresponding ancestral position. Ancestral-like alleles are indicated with dots.

to basal levels by 60 min. By contrast, transfectants expressing the ancestral A471 variant did not show significant differences among the pERK/tERK ratio at any time point tested. Interestingly, when transfectants expressing either A471 or V471 variants were stimulated with PMA, a well-known and potent MAPK activator used here as a positive control, they both showed similar changes in the pERK/tERK ratio at all time points tested that peaked at 30 min and were sustained for 60 min. This indicates that the differences in MAPK activation observed among the A471V variants are indeed intrinsic to the CD5 signaling pathway and not a general defect in the MAPK pathway among transfectants.

A similar set of experiments was also performed on PBMC from healthy volunteers known to be homozygous for each of the A471V variants. As shown by figure 4*B*, following 30-min stimulation with an anti-CD5 mAb, PBMC from V471 homozygotes disclosed a significant increase in the pERK/tERK ratio ($P = 0.01$) as compared with basal levels. By contrast, homozygotes for the ancestral A471 variant did not exhibit such a significant increase ($P = 0.22$). These results are in full agreement with those above reported for COS7 transfectants, thus supporting the higher signaling capability for the more recently derived V471 variant. This differential ability is evident in spite of the higher heterogeneity of the pERK/tERK ratios observed for PBMC compared with COS7 transfectants, which likely result from the different genetic background among the individuals analyzed, as other genetic factors not controlled in this experimental setting may easily influence the status of the CD5 signaling pathway.

Exposure to the Fungal Cell Particle Zymosan Reveals Differential MAPK Activation and IL-8 Release among A471V Variants

Zymosan, a β-glucan–rich particle from the cell wall of *S. cerevisiae*, is known to bind to the CD5 receptor and subsequently induce both MAPK activation and cytokine release (Vera et al. 2009). Thus, it was next explored whether the A471V variants showed differentiated responses when the receptor was exposed to this particular exogenous CD5 ligand. To this, transiently transfected HEK293 cells expressing the A471 or V471 variants previously expressed in COS7 were stimulated with 40 μg/ml of Zymosan for different periods of time and the pERK/tERK ratio determined. As shown by figure 5*A*, HEK 293 transfectants expressing the V471 variant showed a significant increase in pERK/tERK ratio, this peaking at 15 min and slightly declining by 30 min. By contrast, HEK 293 cells transfected with the ancestral A471 variant showed a much lower and fairly significant increment at 15 min that declined to basal levels by 30 min. These results are closely similar to those previously shown in COS7 cells upon CD5 mAb-induced cross-linking and support the existence of different early signaling capabilities among the A471V variants also when cross-linked by conserved fungal cell-wall components.

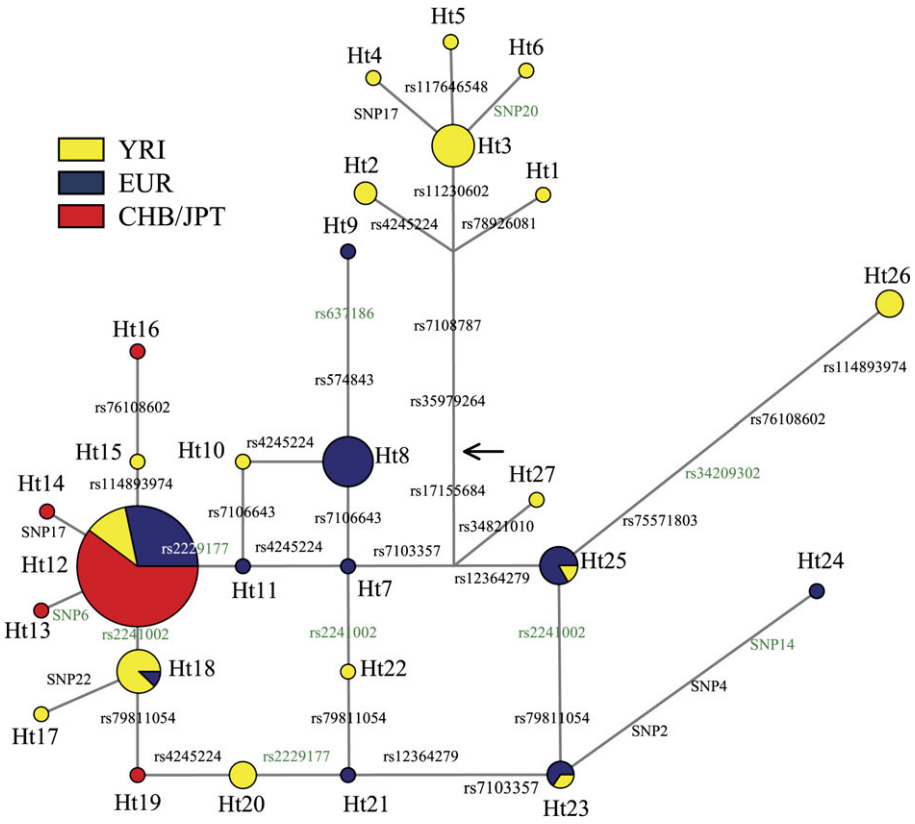

FIG. 2. Median joining network of *CD5* haplotypes. Node areas are proportional to haplotype frequencies and branch lengths to the number of polymorphic substitutions, with nonsynonymous changes in green. The ancestral haplotype considering the chimpanzee allele in each human polymorphic site lies at the position indicated with an arrow. SNPs and haplotypes are numbered as in tables 1 and 2, respectively.

In an independent set of experiments, HEK 293 transfectants expressing the A471 or V471 variants were also assayed for later signaling events such as cytokine gene transcription and secretion. Thus, their ability to release IL-8 upon stimulation with 20 μg/ml of Zymosan for 24 h was explored. According to previously reported results (Vera et al. 2009), the ratio of IL-8 release from stimulated versus nonstimulated cells was found to be higher in CD5-transfected than in nontransfected HEK293 cells (fig. 5*B*). This was especially true for transfectants expressing the V471 variant, which showed significantly higher IL-8 release ratios than those observed in A471 transfectant cells ($P < 0.001$). Taken together, these results indicate that the A471V variants differ in their signaling capabilities, not only at early (MAPK activity) but also at later signaling events (cytokine release), upon receptor CD5 ligation by microbial structures of fungal origin.

Discussion

Convincing evidence for natural selection and a true understanding of adaptative evolution in humans requires the integration of biological function in a population genetics framework. A number of studies searching for signatures of selection based on SNP data, both in particular candidate regions (Moreno-Estrada et al. 2009) and at a genome-wide level (Williamson et al. 2007; Akey 2009; Barreiro and

Table 3. Summary Statistics and Neutrality Tests for the *CD5* Gene.

Population	$2N^a$	S^b	π^c	K^d	H^e	Tajima's *D*	Fu and Li's *D**	Fu and Li's *F**	Fu's *Fs*	Fay and Wu's *H*
Yoruban	40	21	0.00124	17	0.910 ± 0.024	0.103	−0.642	−0.464	−3.480	−0.228
European	40	12	0.00067	10	0.754 ± 0.051	−0.085	−1.074	−0.887	−1.120	0.256
East Asian	40	6	0.00007	5	0.192 ± 0.083	−2.101**	−3.752*	−3.791**	−3.721*	0.292
Total	120	27	0.00079	27	0.735 ± 0.042	−1.040	−2.225	−2.109	−11.539**	−3.417

[a] Number of chromosomes.
[b] Number of segregating sites.
[c] Nucleotide diversity per base pairs.
[d] Total number of haplotypes.
[e] Haplotype diversity ± standard deviation.
* Significant values $P < 0.05$ and **significant values $P < 0.01$.

Table 4. In Silico Functional Characterization for Coding Nonsynonymous SNPs in the *CD5* Gene.

SNP	dbSNP code	Alleles[a]	Residue change	MAF[b]	Granthan distance[c]	PolyPhen prediction
—	rs73484395	T/G	L12V	0.075	32	Benign
—	rs62641694	G/A	R25Q	0.017	43	Benign
6	—	A/G	N203S	—	46	Benign
7	rs2241002	C/T	P224L	0.173	98	Probably damaging
—	rs62641695	C/T	R338W	0.006	101	Benign
13	rs117646548	G/A	A377T	0.005	58	Benign
—	rs117045863	G/C	K407N	0.002	94	Possibly damaging
19	—	C/T	R410C	—	180	Probably damaging
22	rs34209302	C/T	R461C	0.028	180	Probably damaging
23	rs637186	G/A	R461H	0.040	29	Benign
26	rs2229177	C/T	A471V	0.344	64	Possibly damaging

[a] Observed alleles are indicated as ancestral/derived.
[b] Minor allele frequency as reported in NCBI dbSNP Build 132.
[c] Mean chemical distance for the corresponding amino acid pair.

Quintana-Murci 2010), show high levels of population differentiation (extreme F_{ST} values), extended haplotype homozygosity, and/or departure from neutral allele frequency patterns in East Asia along a ~420-kb genomic region on chromosome 11 containing several genes. In silico analysis of the publicly available data linked to the putatively selected haplotype in East Asians led us to previously suggest a nonsynonymous SNP on the *CD5* gene (rs2229177) as the most plausible candidate behind such a selection sweep (Moreno-Estrada et al. 2009). This SNP implies an Ala → Val substitution at amino acid position 471 (A471V; numbering according to the immature non-processed protein isoform) in the cytoplasmatic domain of the CD5 receptor. Here, further evidence is provided for the adaptive evolution in the *CD5* gene by a resequencing approach in 20 Yorubans, 20 Europeans, and 20 East Asians, analyzing the complete amino acid variability of the CD5 lymphocyte surface receptor and functionally assaying the possible relevance of the A471V substitution, putatively linked to the signature of selection, on lymphocyte cell signaling.

Most of the neutrality tests performed (Tajima's *D*, Fu and Li's *F** and *D**, and Fu's *Fs*) and the predominance of a major haplotype in East Asia carrying the derived allele at the A471V site confirm previous evidence for selection obtained from SNP data. Our sequencing data complemented by publicly available data on polymorphisms covering the entire *CD5* gene led us to maintain this nonsynonymous substitution as the most plausible target of selection in the gene. Furthermore, the worldwide frequency distribution obtained here for the derived V471 variant shows its fixation in the Han ($2N = 90$) and values close to fixation in the remaining East Asian populations from the HGDP-CEPH diversity panel. This pattern could imply that the derived V471 variant facilitated the evolutionary adaptation to specific environmental

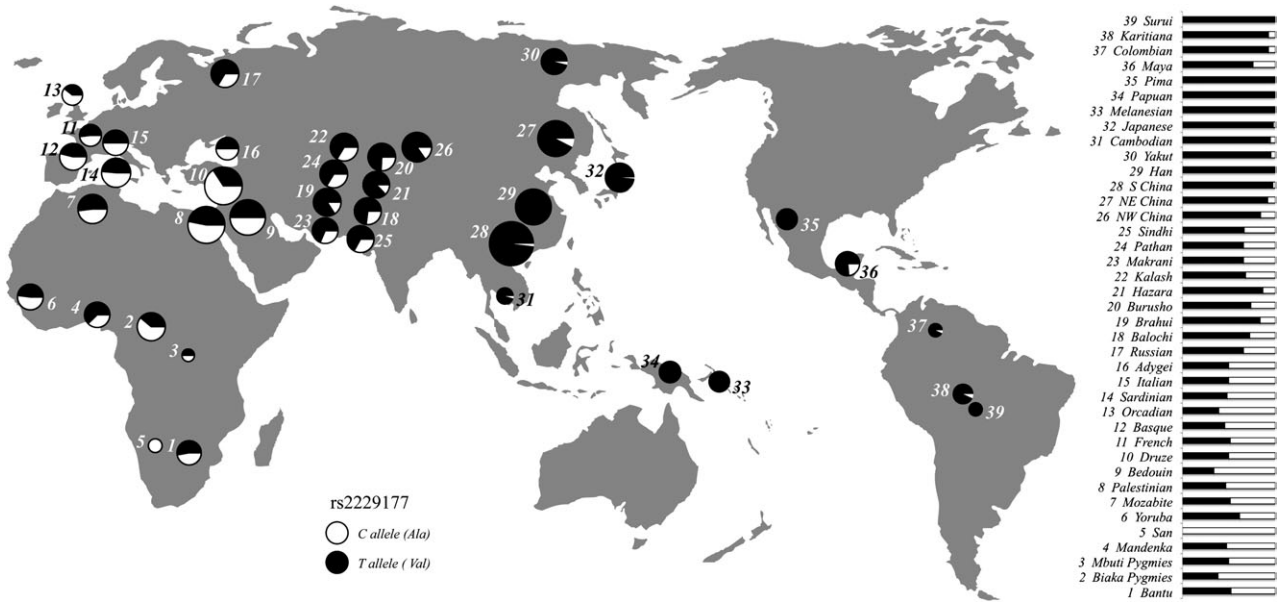

Fig. 3. Geographical distribution of the ancestral (white) and derived (black) alleles for rs2229177 in the HGDP-CEPH. The area of each pie chart is proportional to the sample size in a given population. Populations are numbered from 1 to 39 as represented in the allele frequency bar plot on the right.

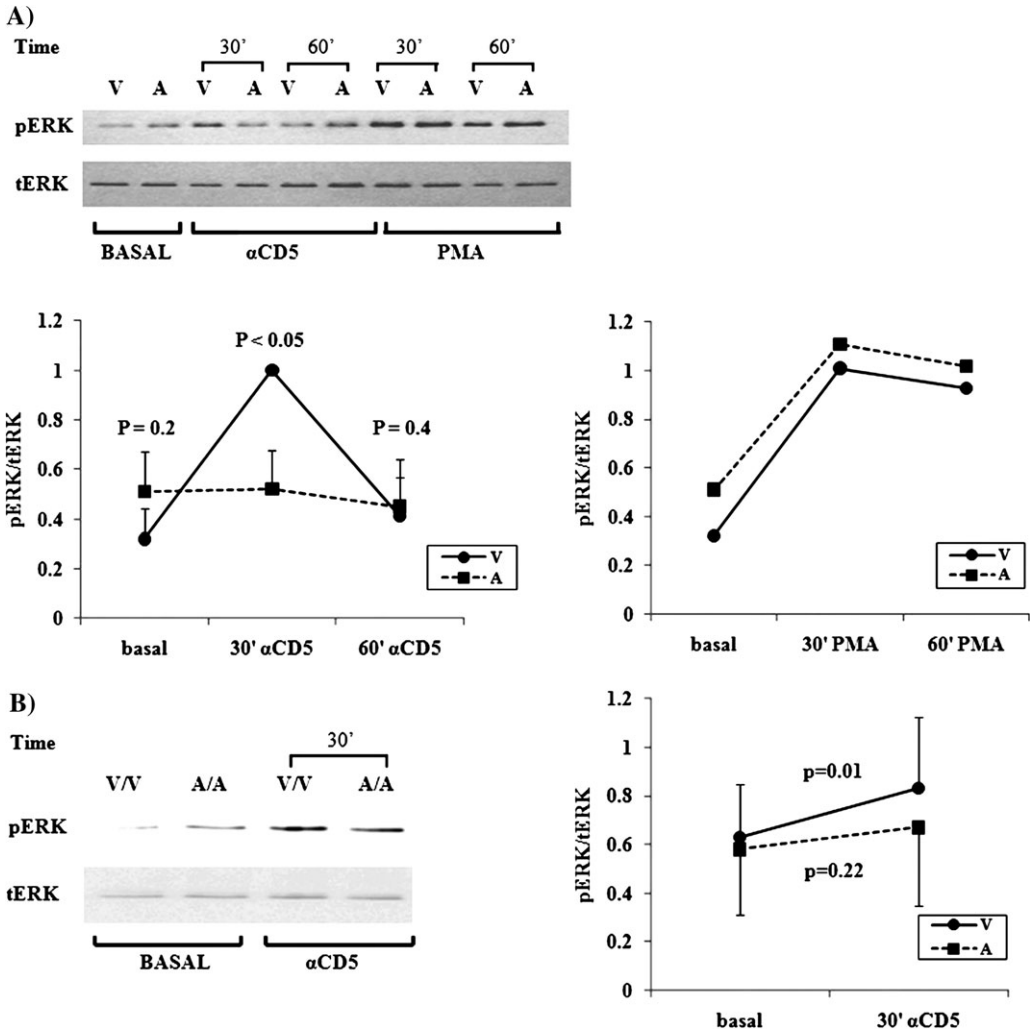

Fig. 4. Differential MAPK activation by A471V variants following mAb-induced CD5 cross-linking. (*A*) Transient COS7 transfectants expressing either the A471 or the V471 variants were serum-starved for 4 h and then stimulated for 30 and 60 min at 37 °C with 10 μg/ml of mouse antihuman CD5 mAb (83C4) or PMA (60 ng/ml) in DMEM. Nonstimulated transfected cells were used as a negative control (basal condition). pERK and tERK were assayed by Western blot. To this, 3 μg of proteins from cell lysates were resolved by 8% SDS–PAGE, transferred to nitrocellulose membrane, and subjected to immunoblotting with anti-pERK1/2 and anti-tERK1/2. The top panel shows a representative Western blot analysis of two performed. The bottom left panel shows the mean ± standard deviation of the pERK/tERK ratio (corrected by the surface CD5 expression level) from the two experiments performed after stimulation with mouse antihuman CD5 mAb. To homogenize the results of the two experiments, poststimulus values at 30 min for V471 transfectants were set to 1, and the rest of the values at all time points were scaled accordingly. *P* values by Student's *t*-test compare the two sets of transfectants at each time point. The bottom right panel shows the pERK/tERK ratio after stimulation with PMA from the experiment shown at the top panel. (*B*) PBMC from healthy volunteers homozygous for each of the A471V variants (A/A or V/V) were serum-starved for 4 h and stimulated for 30 min at 37 °C with 10 μg/ml of anti-CD5 mAb (83C4) in RPMI 1640. The left panel shows the Western blot analysis of pERK and tERK levels in two representative individuals, each homozygous for the A471V variants. The right panel shows the mean ± SD of the pERK/tERK ratio (corrected by the CD5 surface expression level) in three homozygote individuals for each of the A471V variants. Given the heterogeneity in the pERK/tERK ratio found among the analyzed individuals, poststimulus value for one of the V471 homozygote individuals in each experiment was set to 1, and the rest of the values were scaled accordingly. *P* values by Student's *t*-test compare stimulated to baseline pERK/tERK ratios.

circumstances affecting East Asia but not other geographical regions. This assumption is supported by functional assays proving the existence of differential signaling capabilities among the A471V variants exposed to both supraphysiological and more physiological stimulation conditions (specific mAb cross-linking or binding to fungal particles, respectively). Accordingly, when assessing the CD5-mediated activation of the MAPK signaling cascade in transient COS7 and HEK293 transfectants with anti-CD5 mAb and/or Zymosan, it was found that the two allele variants at the A471V site had clearly differentiated signaling kinetics. Interestingly, human peripheral blood lymphocytes from homozygous individuals at the A471V site reproduced the same differentiated intracellular activation pattern. Additionally, shortly after inducing CD5-mediated activation in HEK293 transfectants, the derived V471 form of the receptor not only produced increased activation of the MAPK cascade but also higher release of IL-8 when compared with the ancestral A471 form. Interestingly, IL-8 is a potent chemoattractant for neutrophil

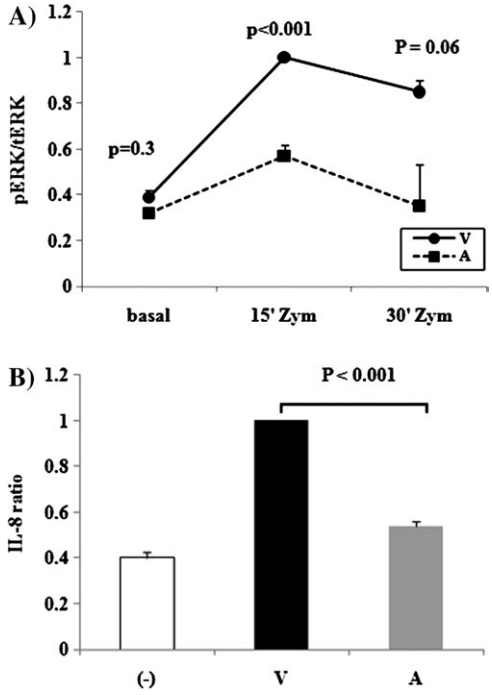

Fig. 5. Differential MAPK activation and IL-8 release by A471V variants following stimulation of HEK293 cell transfectants with the fungal particle Zymosan. (*A*) Transient HEK293 transfectants expressing either the A471 or the V471 variants were serum-starved for 4 h and then stimulated with 40 μg/ml of Zymosan for 15 and 30 min at 37 °C. Cell lysate samples (20 μg of total proteins) were resolved by 8% SDS–PAGE, transferred to nitrocellulose membrane, and subjected to immunoblotting for phosphorylated (p) and total (t) ERK1/2. The top panel shows the mean ± SD of the pERK/tERK ratio from the two experiments performed. To homogenize the results of the two experiments, poststimulus values obtained at 15 min for one of the V471 transfectants were set to 1, and the rest of the values at all time points were scaled accordingly. *P* values by Student's *t*-test compare the two sets of transfectants at each time point. (*B*) Untransfected (−) or transfected HEK293 cells expressing either the A471 or the V471 variants were serum-starved for 4 h and then incubated for 24 h at 37 °C with 20 μg/ml of Zymosan. Supernatant samples (100 μl) were assayed for IL-8 secretion by ELISA BD (OptEIA, Human IL-8 ELISA Set). The optical density at 450 nm was measured, and the IL-8 concentration was determined in picogram per milliliter. The ratio of IL-8 values between stimulated and nonstimulated cells is represented as the mean ± SD of the two experiments performed, corrected by the CD5 surface expression level. To homogenize the results of the two experiments, poststimulus values obtained for V471 transfectants were set to 1, and the rest of the values at all time points were scaled accordingly. *P* value by Student's *t*-test compares the ratio values for the two sets of transfectants.

granulocytes, which are critical cell effectors for fighting several microbial infections, mainly those of fungal origin. Therefore, it is tempting to speculate that differential immunological responses to environmental pathogens are associated with the two allele variants of the A471V site.

The extreme population differentiation of the A471V site together with its involvement in defining the putatively selected haplotype and the higher activation response associated to the V417 variant reinforces our initial hypothesis that this nonsynonymous site is the most likely target of selection. Given the complex function of the CD5 receptor and its ability to regulate multiple immunological processes, it is difficult to precisely define how such a differentiated response could translate into a phenotypical advantage. The fact that homozygosity for the derived V471 allele has been associated with better prognosis in chronic lymphocytic leukemia (Sellick et al. 2008) provides further evidence for the functional relevance of this nonsynonymous site. However, with the data analyzed here, we cannot completely discard the possibility that other unknown functional variants not related to the *CD5* gene might also contribute to the detected signatures of selection.

The exact selective pressure behind the detected signatures of selection in the *CD5* gene is unknown. However, the deep involvement of the CD5 receptor in the regulation of T- and B-cell immune responses (Soldevila et al. 2011) allows the assumption that a coupled pathogen or environmental–host interactive force may have mediated such a selective sweep in East Asia. It has been previously shown that the CD5 ectodomain binds to conserved fungal cell-wall components, namely β-glucans, and that such interaction induces both MAPK activation and cytokine release (Vera et al. 2009). Given its cytoplasmatic location, the A471V site cannot compromise such direct binding to fungi but more likely modify the subsequent lymphocyte activation and/or differentiation to a more adaptive immune response. Unfortunately, we are not aware of any fungal pathogen or any other β-glucan carrier that could have driven such particular geographical adaptation in East Asia. In this regard, it should be noted that, although little or no binding to bacterial PAMPs has been reported for CD5, the analysis of its interaction with alternative conserved microbial structures present on virus or parasites is still pending. Alternatively, it could be argued that the adaptive role of the A471V site might have arisen not from regulation of direct CD5-pathogen interactions but indirectly from differences in the CD5-mediated regulation of particular BCR or TCR signaling responses following specific antigen presentation. The downmodulation of BCR or TCR signaling by CD5 is known to control the fate of antigen-activated lymphocytes, driving them to apoptosis (AICD), anergy, or differentiation into effector cells (Soldevila et al. 2011). Thus, variants of CD5 differing in their signaling capabilities might well have contributed to better adapt the specific lymphocyte responses against environmental pathogens.

The footprint of natural selection on haplotype length tends to fade faster than other types of signals, and it has been suggested that it would not persist longer than 25,000 years (Sabeti et al. 2006). Because different LD-based methods were significant for this region (Moreno-Estrada et al. 2009), a recent time frame for the action of selection on *CD5* can be envisaged, probably long after the first colonization of East Asia by anatomically modern humans.

In conclusion, the present study provides evidence of signatures of positive selection on the nucleotide sequence

variability of the *CD5* gene and demonstrates differential activation patterns in immune signaling for the derived V471 allele as compared with the ancestral residue. However, further investigations on pathogenic agents in East Asia and a greater understanding of the molecular mechanisms underlying the wide range of immune responses mediated by CD5 are needed before a more specific adaptive role for the A471V variants can be properly elucidated.

Supplementary Material

Supplementary figures S1 and S2 and supplementary tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgment

We specially thank M. Vallès and M. Torres-Ciuro for technical support. This research was supported by grants from the Spanish Ministerio de Ciencia e Innovación (BFU2008-01046/BMC to E.B. and SAF2007-62197 and SAF2010-19717 to F.L.); Generalitat de Catalunya (2009SGR1101); Spanish National Institute for Bioinformatics (www.inb.org); Fundación de Investigación y Prevención del SIDA en España (FIPSE 36-0773-09 to F.L.); Instituto de Salud Carlos III (Red Española de Investigación en Patología Infecciosa, RD06/0008/1013 to F.L.); Volkswagenstiftung Scholarship (I/85 198 to J.E.).

References

Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* 19:711–722.

Axtell RC, Webb MS, Barnum SR, Raman C. 2004. Cutting edge: critical role for CD5 in experimental autoimmune encephalomyelitis: inhibition of engagement reverses disease in mice. *J Immunol.* 173:2928–2932.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet.* 40:340–345.

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 11:17–30.

Bikah G, Carey J, Ciallella JR, Tarakhovsky A, Bondada S. 1996. CD5-mediated negative regulation of antigen receptor-induced growth signals in B-1 B cells. *Science* 274:1906–1909.

Calvo J, Padilla O, Places L, Vigorito E, Vila JM, Vilella R, Mila J, Vives J, Bowen MA, Lozano F. 1999. Relevance of individual CD5 extracellular domains on antibody recognition, glycosylation and co-mitogenic signalling. *Tissue Antigens.* 54:16–26.

Cann HM, de Toma C, Cazes L, et al. (41 co-authors). 2002. A human genome diversity cell line panel. *Science* 296:261–262.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 15:1496–1502.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

Frazer KA, Ballinger DG, Cox DR, et al. (250 co-authors). 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.

Friedlein G, El Hage F, Vergnon I, et al. (11 co-authors). 2007. Human CD5 protects circulating tumor antigen-specific CTL from tumor-mediated activation-induced cell death. *J Immunol.* 178:6821–6827.

Gardner M, Gonzalez-Neira A, Lao O, Calafell F, Bertranpetit J, Comas D. 2006. Extreme population differences across Neuregulin 1 gene, with implications for association studies. *Mol Psychiatry.* 11:66–75.

Hippen KL, Tze LE, Behrens TW. 2000. CD5 maintains tolerance in anergic B cells. *J Exp Med.* 191:883–890.

Lankester AC, van Schijndel GM, Cordell JL, van Noesel CJ, van Lier RA. 1994. CD5 is associated with the human B cell antigen receptor complex. *Eur J Immunol.* 24:812–816.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Lozano F, Simarro M, Calvo J, Vila JM, Padilla O, Bowen MA, Campbell KS. 2000. CD5 signal transduction: positive or negative modulation of antigen receptor signaling. *Crit Rev Immunol.* 20:347–358.

Morcillo-Suarez C, Alegre J, Sangros R, et al. (17 co-authors). 2008. SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. *Bioinformatics* 24:1643–1644.

Moreno-Estrada A, Tang K, Sikora M, Marques-Bonet T, Casals F, Navarro A, Calafell F, Bertranpetit J, Stoneking M, Bosch E. 2009. Interrogating 11 fast-evolving genes for signatures of recent positive selection in worldwide human populations. *Mol Biol Evol.* 26:2285–2297.

Ordonez-Rueda D, Lozano F, Sarukhan A, Raman C, Garcia-Zepeda EA, Soldevila G. 2009. Increased numbers of thymic and peripheral CD4+ CD25+Foxp3+ cells in the absence of CD5 signaling. *Eur J Immunol.* 39:2233–2247.

Osman N, Ley SC, Crumpton MJ. 1992. Evidence for an association between the T cell receptor/CD3 antigen complex and the CD5 antigen in human T lymphocytes. *Eur J Immunol.* 22:2995–3000.

Perez-Villar JJ, Whitney GS, Bowen MA, Hewgill DH, Aruffo AA, Kanner SB. 1999. CD5 negatively regulates the T-cell antigen receptor signal transduction pathway: involvement of SH2-containing phosphotyrosine phosphatase SHP-1. *Mol Cell Biol.* 19:2903–2912.

Raman C, Kuo A, Deshane J, Litchfield DW, Kimberly RP. 1998. Regulation of casein kinase 2 by direct interaction with cell surface receptor CD5. *J Biol Chem.* 273:19183–19189.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 30:3894–3900.

Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* 179:555–567.

Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet.* 70:841–847.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.

Sarrias MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F. 2004. The scavenger receptor cysteine-rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit Rev Immunol.* 24:1–37.

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. 2005. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 15:1576–1583.

Sellick GS, Wade R, Richards S, Oscier DG, Catovsky D, Houlston RS. 2008. Scan of 977 nonsynonymous SNPs in CLL4 trial patients for the identification of genetic variants influencing prognosis. *Blood* 111:1625–1633.

Simarro M, Calvo J, Vila JM, Places L, Padilla O, Alberola-Ila J, Vives J, Lozano F. 1999. Signaling through CD5 involves acidic sphingomyelinase, protein kinase C-zeta, mitogen-activated protein kinase kinase, and c-Jun NH2-terminal kinase. *J Immunol.* 162:5149–5155.

Simarro M, Pelassy C, Calvo J, Places L, Aussel C, Lozano F. 1997. The cytoplasmic domain of CD5 mediates both TCR/CD3-dependent and -independent diacylglycerol production. *J Immunol.* 159:4307–4315.

Soldevila G, Raman C, Lozano F. 2011. The immunomodulatory properties of the CD5 lymphocyte receptor in health and disease. *Curr Opin Immunol.* 23:310–318.

Stamou P, de Jersey J, Carmignac D, Mamalaki C, Kioussis D, Stockinger B. 2003. Chronic exposure to low levels of antigen in the periphery causes reversible functional impairment correlating with changes in CD5 levels in monoclonal CD8 T cells. *J Immunol.* 171:1278–1284.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 68:978–989.

Tarakhovsky A, Kanner SB, Hombach J, Ledbetter JA, Muller W, Killeen N, Rajewsky K. 1995. A role for CD5 in TCR-mediated signal transduction and thymocyte selection. *Science.* 269:535–537.

Vera J, Fenutria R, Canadas O, Figueras M, Mota R, Sarrias MR, Williams DL, Casals C, Yelamos J, Lozano F. 2009. The CD5 ectodomain interacts with conserved fungal cell wall components and protects from zymosan-induced septic shock-like syndrome. *Proc Natl Acad Sci U S A.* 106:1506–1511.

Vila JM, Gimferrer I, Padilla O, Arman M, Places L, Simarro M, Vives J, Lozano F. 2001. Residues Y429 and Y463 of the human CD5 are targeted by protein tyrosine kinases. *Eur J Immunol.* 31:1191–1198.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3:e90.

Yanaba K, Bouaziz JD, Haas KM, Poe JC, Fujimoto M, Tedder TF. 2008. A regulatory B cell subset with a unique CD1dhiCD5+ phenotype controls T cell-dependent inflammatory responses. *Immunity* 28:639–650.

Youinou P, Jamin C, Lydyard PM. 1999. CD5 expression in human B-cell populations. *Immunol Today.* 20:312–316.