HHS Public Access

Author manuscript

Sci Transl Med. Author manuscript; available in PMC 2015 March 20.

Published in final edited form as:

Sci Transl Med. 2013 July 3; 5(192): 192ra86. doi:10.1126/scitranslmed.3006338.

Natural Selection in a Bangladeshi Population from the Cholera-Endemic Ganges River Delta

Elinor K. Karlsson[1,2,*], **Jason B. Harris**[3,4], **Shervin Tabrizi**[1,2], **Atiqur Rahman**[3,5,6], **Ilya Shlyakhter**[1,2], **Nick Patterson**[2], **Colm O'Dushlaine**[2], **Stephen F. Schaffner**[2,7], **Sameer Gupta**[8], **Fahima Chowdhury**[5], **Alaullah Sheikh**[5,9], **Ok Sarah Shin**[3,10], **Crystal Ellis**[3], **Christine E. Becker**[11], **Lynda M. Stuart**[2,11], **Stephen B. Calderwood**[3,12,13], **Edward T. Ryan**[3,7,12], **Firdausi Qadri**[5,†], **Pardis C. Sabeti**[1,2,7,*,†], and **Regina C. LaRocque**[3,12,*,†]

[1]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA

[4]Department of Pediatrics, Harvard Medical School, Boston, MA

[5]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

[7]Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA

[*]To whom correspondence should be addressed: elinor@broadinstitute.org (E.K.K.); psabeti@oeb.harvard.edu (P.C.S.); rclarocque@partners.org (R.C.L.).

[6]Department of Biochemistry and Molecular Biology, Faculty of Biological Sciences, University of Dhaka, Bangladesh (current)

[9]Molecular Microbiology and Microbial Pathogenesis Program, Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO (current)

[10]Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea (current)

[†]These authors contributed equally

Supplementary Materials

Author contributions: R.C.L., P.C.S., J.B.H., F.Q. and E.K.K. conceived the study and designed the experiments. R.C.L., F.Q., J.B.H., E.T.R., S.B.C., A.S. and F.C. collected, prepared and characterized diagnoses for human samples. E.K.K., S.T., I.S. N.P. C.O., P.C.S. and S.S. designed analysis methods and statistical tests. E.K.K., R.C.L., A.R., C.E., S.T, I.S. N.P. and S.G. performed and interpreted population genetic, selection, enrichment and association analyses. R.C.L., J.B.H., L.M.S., C.B. and O.S.S. designed and performed the in vitro experiments. E.K.K, R.C.L., P.C.S., J.B.H. and F.Q. wrote the paper with input from S.S. and other authors.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: The CMS code is available at http://www.broadinstitute.org/mpg/cms.

[8]Harvard Medical School, Boston, MA

[11]Developmental Immunology and Center for Computational and Integrative Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA

[12]Department of Medicine, Harvard Medical School, Boston, MA

[13]Department of Microbiology and Immunobiology, Harvard Medical School, Boston MA

Abstract

As an ancient disease with high fatality, cholera has likely exerted strong selective pressure on affected human populations. We performed a genome-wide study of natural selection in a population from the Ganges River Delta, the historic geographic epicenter of cholera. We identified 305 candidate selected regions using the Composite of Multiple Signals (CMS) method. The regions were enriched for potassium channel genes involved in cyclic AMP-mediated chloride secretion and for components of the innate immune system involved in NF-κB signaling. We demonstrate that a number of these strongly selected genes are associated with cholera susceptibility in two separate cohorts. We further identify repeated examples of selection and association in an NF-kB / inflammasome-dependent pathway that is activated *in vitro* by *Vibrio cholerae*. Our findings shed light on the genetic basis of cholera resistance in a population from the Ganges River Delta and present a promising approach for identifying genetic factors influencing susceptibility to infectious diseases.

Introduction

Several lines of evidence suggest that cholera exerted a potent selective pressure on the human population in Bangladesh. Cholera is an ancient disease in the Ganges River Delta(1) and remains highly prevalent in Bangladesh today, with up to 4 per 1000 affected annually in urban slums of Dhaka(2). By the age of 15 years, over half the population of Bangladesh has serological evidence of previous cholera infection(3, 4). Historically, the mortality rate of cholera exceeded 50% and remains as high as 5 – 10% in recent outbreaks(5–7). Cholera's ancient origins, high prevalence, and high fatality rates suggest strong evolutionary pressure, and observational data support this. In particular, the Ganges River Delta has the world's lowest prevalence of blood group O, which is associated with an increased risk of severe cholera(8, 9). In addition, first degree relatives living in the household of cholera patients have nearly three-fold higher risk of cholera than unrelated household contacts(4, 8, 9), suggesting susceptibility to cholera has a heritable component.

A history of selective pressure should make it easier to identify host variants associated with cholera susceptibility and severity, an approach that can give insight into disease pathogenesis(10–12). Natural selection drives beneficial mutations to rise in prevalence, generating common mutations of strong effect that are detectable with small samples. In addition, the process of selection leaves a distinctive signature in the genome that can add power to a genome-wide association study (GWAS)(13–15).

The virulence factors of the human-specific pathogen *Vibrio cholerae*, which causes cholera, are well characterized(16–20). Toxigenic strains of *V. cholerae* colonize the surface of the

small intestine and express cholera toxin, an AB_5 bacterial toxin that induces profound diarrhea mediated by secretion of chloride through apical chloride channels(21, 22). Less is known about host factors that contribute to cholera susceptibility and about the interaction between *V. cholerae* and the human intestine. Pathways related to mucosal immunity and intestinal homeostasis likely play an important role(20).

We sought to identify host genetic factors associated with susceptibility to cholera by combining selection, association, and functional studies. We previously developed the Composite of Multiple Signals (CMS) method to pinpoint regions of positive selection in the genome with more power and specificity than any single test(23, 24). Here, we used CMS to identify targets of positive selection in a population from the Ganges River Delta, and then searched for association with cholera resistance among the identified targets.

Results

Genetics of the study population

We genotyped 42 mother-father-child trios (126 individuals) of Bengali ethnicity from Dhaka, Bangladesh (abbreviated “BEB” for BEngali of Bangladesh) on the Illumina 1M array, yielding 36 complete trios and 1,112,946 single nucleotide polymorphisms (SNPs) after quality filtering. We compared this population with publically available data for 16 populations: eight from the International HapMap3 Project (HapMap3), three from the Singapore Genome Variation Project and five from the Human Genome Diversity Project (Table S1)(25–27).

The BEB population is a distinct genetic grouping with no evidence of substantial genetic structure; principle component analysis reveals the population to be tightly clustered with little overlap with any other HapMap3 population (fig. 1A, fig. S1). Among the eight HapMap3 populations, the closest relatives of the BEB are the Indian Gujarati (F_{ST}=0.0047) (Table S2). The BEB are closer than the Gujarati to the Japanese and Han Chinese, potentially reflecting admixture. We tested seven populations (21 pairs) as ancestors using the *3 Population Test* (28) and found that admixture between Indian and East Asian populations best explains the observed genetic architecture of the BEB (table S3). While all pairs of East Asian and Indian populations fit well, of those included, the Singapore Indians and the Singapore Chinese most closely match the ancestors of the BEB population (Z= −48.4).

We estimate that the admixture between Indian and East Asian populations occurred 52 +/− 2 generations ago (generation=29 years)(29), or around 500AD,based on the exponential decline of linkage disequilibrium (LD) with distance analyzed using *ROLLOFF*(30, 31)(fig. 1B). This remarkably close-fitted age estimate roughly corresponds to the collapse of the Indian Gupta Empire, the rise of the Chinese Tang dynasty and the brief unification of Bengal under a single ruler (590 AD–625 AD). While alternative histories, such as continuous admixture or multiple admixture events, are possible, the single event model shows excellent fit to our data and we found no statistical support for very ancient flow (fig. S2). Using the maximum likelihood-based ancestry estimation software ADMIXTURE(32), we found 9.3% +/− 2.6% East Asian ancestry in the BEB (fig. 1C, fig. S3, table S4).

Natural Selection in the BEB Population

The CMS test combines three indicators of positive selection – long haplotypes, high-frequency derived alleles, and highly differentiated alleles – to identify narrow candidate regions, in many cases pinpointing a single gene or genomic element as the probable target of selection(23). We measured natural selection in the autosomal genome of the BEB population with CMS_{GW}, a genome-wide test(24), using for comparison three sequenced populations from the 1000 Genomes Project: north/west Europeans, Han Chinese/Japanese and Yoruba (795, 517 overlapping SNPs)(table S1)(33).We identified 305 regions with signals of natural selection (normalized CMS_{GW} score >3; false positive rate (FPR)<0.01%, fig. 2, table S5), with a median size of 149 kb and encompassing 2% of the autosomal genome. On average, the selected regions contain 2.3 +/− 3.6 genes; 95 regions contain one gene and 72 have no genes. 91 regions contain long intergenic non-coding (LINC) RNAs, including 29 non-genic regions (table S6).

Enrichment of gene sets in selected genomic regions in the BEB population

Enrichment analysis of gene categories is used to identify common functions among candidate regions identified in GWAS(34). It can similarly be used with selection candidates, providing clues to the evolutionary forces shaping a population. However, given that selection acts on many distinct traits, we expect to see clear evidence of gene enrichment only in extreme cases, when multiple genes in a particular pathway are targeted by strong selection for a single trait (e.g. skin pigmentation in Europeans).

We looked for evidence of enrichment in the 305 selected regions in the BEB population using INRICH, a permutation-based genome-wide analysis tool that is robust to confounding factors like marker density and gene size(35). INRICH reports the empirical significance of enrichment for each gene set (P_{set}), and a corrected experiment-wide empirical significance (P_{exp}). We considered $P_{\text{exp}}<0.05$ to represent significant enrichment. We tested 851 sets of 10 or more genes from the Molecular Signatures Database (MSig-c4), which identifies sets of genes with shared expression patterns(36). We also tested 2,430 manually curated sets of 10 or more genes from the Gene Ontology (GO)(37).

One gene set was significantly enriched in the BEB population - a module of genes in the expression neighborhood of *IKBKG* (MORF_IKBKG, $P_{set}=5.2\times10^{-5}$, $P_{exp}=0.017$)(table 1). MORF_IKBKG was one of just two gene sets significantly enriched in any population (table S7). *IKBKG* encodes the protein NEMO (IKK-γ), a subunit of the IκB kinase that activates canonical signaling of nuclear factor (NF)-κB. Sixteen of the 110 genes in the MORF_IKBKG gene set lie in 15 distinct genomic regions that show evidence of natural selection in the BEB population (fig. 2). We observed no enrichment for this gene set in the East Asians, the Europeans or the Yoruba (P_{set}=0.56, 0.44, 0.17, respectively).

Many genes in the top selected regions in the BEB population, including 3 of the top 10 regions, were classified as potassium ion transport genes (table S5). Enrichment analysis of GO sets in strongly selected regions (CMS_{GW} score >5) ranks potassium ion transport (GO: 0006813) third ($P_{set}=3.8\times10^{-3}$)(table S7). To further investigate, we tested 19 categories of potassium channel genes, defined by their physiological function in gastrointestinal

epithelial cells(38), and found significant enrichment for voltage gated potassium channel genes (5/47 genes, P_{set}=1.6×10^{-3},P_{exp}=0.008) and in particular for eag-related genes (3/8 genes, P_{set}=5.2×10^{-3},P_{exp}=0.03)(fig. 2, table 1, table S8).

We also tested for enrichment in 31 blood group systems (table S8) because of the association between blood group O and severe cholera. Genes related to the Kell blood group system were significantly enriched in BEB selected regions (3/8 genes, P_{set}=4.4×10^{-3}, P_{exp}=0.007)(table 1, fig. 2) and in East Asians (2/8 genes, P_{set}=0.025, P_{exp}=0.03)(table S7). We detected no selection in ABO blood group genes, possibly because CMS_{GW} is designed to detect positive selection on novel variants, not the negative selection against long-standing genetic variation (i.e. the O blood group) expected at the ABO locus.

Testing of Top Regions for Association with Cholera Susceptibility

To test the hypothesis that resistance to cholera drove selection in the BEB population, we performed an association study in a separate, well-phenotyped cohort of Bangladeshis of Bengali origin from Dhaka comprised of 105 cholera patients and 167 unaffected individuals exposed to a case of cholera within their household(39). This cohort included 38 parent-affected child trios. We genotyped 536 SNPs in 28 selected regions; these regions included the top ten selected regions in the BEB genome and regions that contained genes in enriched gene sets or that were biologically plausible candidates for cholera resistance (table S9). 370 polymorphic SNPs (minor allele frequency >0.01) passed quality filters, including 19 SNPs randomly chosen from regions without evidence of natural selection. We tested for association using the DFAM method in *PLINK* (40), which combines related and unrelated individuals into a single analysis. The region with strongest association to cholera is within the top signal of natural selection in the genome and encompasses five genes --*SNRNP200*, *CIAO1*, *ITPRIPL1*, *NCAPH*, and *TMEM127* (fig. 3A,B; table S10). The most associated SNP is between *SNRNP200* and *ITPRIPL1* (rs62153901; p_{DFAM}=0.0015, p=0.042 after Bonferroni correction for 28 independent loci tested), 31kb from the top SNP in the selection analysis (rs3171927; in *SNRNP200*).

This association is unlikely to reflect population stratification. The top associations persist (Spearman correlation of 0.76 for SNPs with p<0.01) when we restrict our analysis to the 38 parent-affected child trios in our dataset using the transmission disequilibrium test, a method that internally controls for population stratification(41). Allele sharing across the 19 random markers, measured as the identity-by-state distance(40), is similar within cases (0.734 +/− 0.068), within controls (0.735 +/− 0.070) and between cases and controls (0.734 +/− 0.69), consistent with the lack of substructure in the population (fig. 1A).

Our association cohort contains individuals with a range of severity of cholera, including cases with severe dehydration requiring hospitalization. Restricting our analysis to these severe cases reduces our statistical power, but may also increase our sensitivity through a more rigorous phenotype. We indeed found that analyzing the most severe cholera cases alone resulted in the identification of additional associations between cholera and SNPs in genes located in three selected regions -- the potassium ion transport genes *KCNH7* (p=0.019) and *KCNH5* (p=0.036), and the ribosomal protein kinase gene *RPS6KB2* (p=0.049)(fig. 3A, table S10).

We performed a replication study on the top 12 associated regions to confirm the association to cholera (44 SNPs; 204 SNPs after imputation). We genotyped 124 unrelated, severe cholera cases and performed a case-control analysis with 72 parents from the trio population cohort as unphenotyped controls, measuring association at 49 haplotype blocks of SNPs in strong LD(42). We found that the top region in the original association study was among the four regions most associated with cholera ($p<0.01$) in this separate analysis (Fig3A, table S10). We also found a new association on chromosome 16, a gene dense region that includes the gene *PYCARD* and ranked 5th in the selection study.

Biological association between genes under natural selection and cholera

Our selection scan and enrichment analysis indicate that recent, positive natural selection in the BEB population has repeatedly targeted genes related to *IKBKG*, a key component of NF-κB signaling. Our studies further indicate that certain genes in the top regions of selection are specifically associated with cholera. We and others have recently shown that *V. cholerae* lipopolysaccharide (LPS) incites pro-inflammatory innate immune responses through activation of NF-kB signaling pathways(43, 44), suggesting that cholera may represent a selective pressure on this pathway.

We also noted that the selected regions in the BEB population contain genes related to activation of the inflammasome, including *PYCARD* (also known as *Asc*)(table S5). To further investigate this, we evaluated the effect of cholera toxin on inflammasome activation in mouse macrophages *in vitro* and found that it induced robust IL-1β secretion in LPS-primed mouse macrophages (fig. 4). This effect was abrogated in macrophages from mice deficient in caspase-1 (*Ice*−/−) and *PYCARD* (*Asc*−/−), two key components of the inflammasome signaling pathway. Thus, cholera toxin leads to caspase-dependent induction of pro-inflammatory cytokines, consistent with *V. cholerae* infection resulting in activation of the inflammasome.

Discussion

Using Ingenuity's IPA software (45), we developed a model of the human innate immune signaling pathways that respond to *V. cholerae* infection and have been selected in the BEB population (fig. 5). In this model, inflammasome activation and the NF-kB signaling pathway play an integrated role in TLR4-mediated sensing of *V. cholerae*. This model is consistent with recent *in vitro* data regarding innate immune sensing of Gram-negative bacteria (46) and is supported by *in vitro* findings of NF-kB signaling (43) and inflammasome activation(47) by *V. cholera* antigens.

Our results suggest that natural selection in the BEB population has repeatedly acted on key regulators of the proposed *V. cholerae* response pathway. In particular, genes expressed in concert with *IKBKG*, which encodes the protein NEMO, are the most strongly enriched MSigDB gene set in the BEB population. NEMO plays an essential role in the canonical activation and regulation of NF-κB, the master regulator of inflammation, immunity and cell survival (48).The MSigDB IKBKG module also includes *RPS6KB2*, a ribosomal protein S6 kinase gene that is both under selection in the BEB population and associated with severe cholera in our disease association study. While the function of *RPS6KB2* is poorly

understood, other members of this family are known to regulate NF-κB and factors associated with gut inflammation and apoptosis (49). We found the most significant overlap of our selection and cholera association data near the novel gene *SNRNP200*, which encodes a previously uncharacterized RNA splicing, ATP-dependent helicase shown to bind caspase-4(50). The murine homolog of caspase-4, caspase-11, is a key regulator of caspase-1 activation in response to Gram negative bacteria, including *V. cholerae*, and subsequent inflammasome activation(47). *PYCARD* (*Asc*), located in the fifth ranked selected region, mediates the formation of the multi-protein inflammasome complex. Several downstream effectors of the proposed innate immune signaling pathways, including ITGAM and IL-1β, are also under selection in the BEB population or are highly expressed in the duodenum of cholera patients(51).

Cholera in Bangladesh is ancient, prevalent, and often fatal. There is compelling evidence that heritable factors influence susceptibility to the disease, yet little is known about the specific genetic factors that are involved. Our scan for natural selection in a historically cholera-affected population from Bangladesh, combined with disease association and biological data, elucidates new host factors involved in cholera. We note that, while our analysis found no evidence that the association reflects population stratification, this potential confounder cannot be completely excluded because of the small size and targeted design of the association study.

Our results suggest that cholera has exerted strong selective pressure on key pathways of innate immunity, including NF-κB and inflammasome signaling, in this population from Bangladesh. Our candidate regions commonly include regulatory components of these pathways rather than the central mediators. This suggests a biologically sensible hypothesis – namely, that selection acts to modulate these fundamental pathways instead of affecting the key components.

Functional studies increasingly support the role of innate immunity in the response to *V. cholerae* infection. In a recent screen of *Drosophila* insertion mutants, *V. cholerae* susceptibility was associated with mutations in innate immunity genes related to NF-kB signaling(52, 53). Animal models also support a central role for inflammasome activation in response to infection with *V. cholerae* and other Gram negative organisms(46, 47). Since the innate immune system provides the first line of host defense against infection, it is possible that other microbial pathogens exerted selective pressure on the same pathways (54–57). Association studies for other diseases historically prevalent in the Ganges River Delta, focused on the regions of natural selection that we have identified, would consequently be of interest.

The NF-κB signaling pathway we identified as under selection in the BEB population modulates gut epithelial integrity and the interaction between the mucosal immune system and gut microflora(58). Ablation of *IKBKG* in the intestinal epithelia of mice causes severe, chronic intestinal inflammation, much like human inflammatory bowel disease (IBD)(59). We noted a significant overlap between genes that are highly selected in the BEB population and loci that are strongly associated ($p_{GWAS}<1\times10^{-10}$) with an increased risk of ulcerative colitis (P_{set}=0.017,P_{exp}=0.049, 3/17 GWAS loci)(table S11)(60). Thus, pathways under

selection in the BEB population, and potentially involved in susceptibility to mucosal infection with *V. cholerae*, may also be relevant to understanding a common autoimmune disease occurring at the mucosal surface. Notably, epidemiological studies suggest an increased risk for UC and IBD in South Asian populations (61, 62).

A role for ion channels, particularly the cystic fibrosis transmembrane conductance regulator (CFTR), in susceptibility to cholera has long been hypothesized (63, 64). We found no evidence of selection at *CFTR* (fig. S4). We did find significant enrichment in top selected regions for voltage-gated potassium channel genes; one of these potassium channel genes, *KCND2*, is approximately 2.5Mb downstream of CFTR. We identified a specific association between two other voltage-gated potassium channel genes, *KCNH7* and *KCNH5*, and severe *V. cholerae* infection. *KCNH7* and *KCNH5* are in a class of potassium channels that is expressed in excitable cells but is also found in gastrointestinal tract epithelia (65). These channels appear to control cyclic AMP-mediated chloride secretion, the mechanism by which cholera toxin causes secretory diarrhea (21, 22, 66–70).

Beyond insights on cholera, our genome-wide scan for selection provides clues to other important traits in the BEB population. For example, we found the Kell blood group genes to be highly selected both in the BEB population and in East Asians. The expression of Kell blood group genes increases in arsenite-treated cells(71). High levels of carcinogenic arsenic groundwater contamination are found in Bangladesh and parts of China(72), suggesting a possible selective pressure for this finding.

In this study, we leveraged the shared history between a human population and a powerfully selective pathogen in order to identify specific host genetic factors influencing disease susceptibility. By identifying areas of overlap between targets of natural selection and genes associated or biologically linked with cholera, we were able to propose a testable, biological hypothesis regarding the interaction between *V. cholerae* and the inflammasome. We were also able to identify an innate immune response pathway under selection by cholera; this pathway includes genes not found using other methods, such as animal models and human expression studies. Further studies of this pathway may shed light on immune responses to enteric pathogens or the mechanisms of intestinal homeostasis (73). Our approach is applicable to other historically prevalent infectious diseases, such as Lassa fever, tuberculosis, leishmaniasis and malaria, and to complex, common diseases for which the associated genes may have been historically selected, such as IBD. With the number of publically accessible genetic datasets for human populations growing rapidly, incorporating tests for natural selection into existing and future GWAS will add a new dimension to our understanding of human genetics and disease.

Materials and Methods

Selection study design

We designed the selection study using the same mother-father-child trio design as the Haplotype Map, which we have previously shown effective for CMS analysis(23, 24). We enrolled parent-child trios of Bengali ethnicity at a field site of the International Centre for Diarrheal Disease Research (icddr,b) in Dhaka, Bangladesh. We collected buccal swabs

from study participants for extraction of genomic DNA (Qiagen). Informed consent was obtained from participants or adult guardians. The study was approved by the Ethical and Research Review Committees of icddr,b, Partners Healthcare Human Research Committee, and the institutional review board of Broad Institute/MIT. We genotyped 42 trios (126 individuals) on the Illumina 1M-Duo array (1, 199, 026 SNPs). We excluded male heterozygous genotypes on nonpseudoautosomal X, removed SNPs (n=79, 961) and individuals (n=3) with call rates <90%, families with >1,000 Mendelian errors (n=3), and SNPs with Hardy-Weinberg equilibrium $p<0.001$ (n=597) or >1 Mendelian errors (n=4497) (40). We phased and imputed the final dataset of 36 trios and 1,112,946 SNPs (average call rate 99.5%) with BEAGLE 3.2.0(default parameters)(74) on genome build hg18. We identified regions of natural selection using CMS_{GW}.

Association study design

We tested our top selected variants for association with the phenotype of cholera susceptibility in a targeted association study. We compared individuals with and without a diagnosis of cholera from an observational cohort of index cases (presenting at icddr,b's Dhaka hospital between January 2004 and November 2005) and family members sharing household and cooking facilities(39). A diagnosis of cholera in the index cases required acute watery diarrhea and a stool culture positive for *V. cholerae* O1, and in the family members, a positive rectal swab culture for O1 or symptoms of diarrhea during 21 days of follow-up. Hospitalized individuals were severe cases. We had 105 cases (69 severe) and 167 controls, including 38 parent-affected offspring trios (16 severe). The study was approved by the Ethical and Research Review Committees of the icddr,b and by the Partners Healthcare Human Research Committee.

We genotyped 517 SNPs tiled across 28 top selected regions, including 9 of the top 10 (excluding one containing only olfactory genes)(table S9), and 19 randomly chosen SNPs in the cases, controls and the 42 trios from the BEB selection study using the Sequenom MassArray iPlex platform. We excluded SNPs and individuals with missing call rates >25%, yielding 496 SNPs (genotype rate 98.2%), with 371 polymorphic (minor allele frequency >1%). We estimated allele sharing as IBS distance (number shared/total number of alleles). We tested for association using the TDT and DFAM tests implemented in PLINK(40).

We designed the replication study to validate the top associations by genotyping 57 SNPs in an independent cohort of 124 unrelated, severe cases compared to the 72 unrelated parents from the BEB trios (unphenotyped controls). We removed 2 SNPs with call rates < 75%, leaving 44 SNPs from the top 12 candidate regions and 11 randomly selected SNPs (average genotyping rate 95.9%). We used BEAGLE to impute calls at SNPs in LD with a genotyped SNP ($r^2>0.9$), using the Illumina 1M and Sequenom data from the BEB trios as a reference, yielding a final dataset of 204 SNPs(74). We tested for association to 49 haplotype blocks of SNPs in strong LD, defined using Haploview's "Gabriel et al" method(40, 42).

Population Genetics and CMS analysis

Principal components analysis was done with GCTA(75) and maximum likelihood estimation of individual ancestries with ADMIXTURE 1.22 on a thinned marker set with

reduced LD (PLINK; 50 SNP windows, r^2<0.1)(32). We tested for admixture using the 3 Population Test (f3), and estimated the date of admixture using *ROLLOFF*(28, 30, 31). CMS was run on the BEB and the phased SNP data from the 1000 Genomes Project (Supp methods). We tested for signals of natural selection in the BEB, Yoruba, North/West Europeans, and East Asians by calculating the CMS_{GW} score for SNPs genotyped in all populations (24)(supp. methods). We defined selected regions as >100kb regions where >30% SNPs have normalized CMS score >3 (0.1% FPR in simulations), merged overlapping regions, and identify genes in regions using the RefSeqGene catalog and the LINC RNA catalog (76–78).

Gene set enrichment testing

We tested regions for enriched gene sets using INRICH, calculating significance empirically through 1,000,000 permutations matched for region size, SNP density, and gene number(35). Reference gene sets detailed in supplementary methods.

Activation of the inflammasome by cholera toxin

Immortalized bone marrow derived macrophages (BMDM) from wild-type, caspase-1 knockout (*Ice*−/−) and *Asc* knockout (*Asc*−/−) mice were provided by D. Golenbock (University of Massachusetts Medical School). Cells were plated in 96-well flat bottom plates at 2×10^5 cells/well in DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells were primed with LPS (200 ng/ml) for 3 hours and stimulated with cholera toxin (1ug/ml) for 22 hours. Supernatants were collected and IL-1β concentration determined by IL-1β sandwich ELISA using the DuoSet ELISA Development Kit (R&D System).

Selection on innate immunity pathways related to V. cholerae infection

We developed the model of selection on the human innate immune signaling pathways that respond to *V. cholerae* infection using Ingenuity's IPA software(45). We first compiled a core gene set from experimentally derived pathways linked to cholera(46, 47) and genes up or down regulated in duodenal biopsies obtained from patients with acute cholera(51). We used IPA to agnostically identify all interactions annotated as direct and experimentally validated between this gene set and two sets of candidate selected genes: (1) those in the top ten regions of selection in the BEB population; (2) *IKBKG* and the selected genes in the MORF_IKBKG gene set. We included one additional gene, *CD86*, as it is known to respond to cholera toxin B subunit(79) and is one of just two known interactions in the Ingenuity Knowledge Base for *MARCH8*, the only gene in the 4th highest region of selection in the BEB (CMS_{GW}=9.02)(80, 81).

Statistics

For CMS_{GW}, we considered a normalized score above 3 significant, corresponding to a 0.1% false positive rate (FPR) in simulations(24). For gene set enrichment with INRICH (35) and trait association (TDT, DFAM and chi-squared tests in PLINK) (40), $p < 0.05$ (one-sided) was considered significant. See methods for details.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank P. Moorjani, S. Grossman, P. Lee, D. Reich, S. Purcell and C. Edwards for invaluable support and critical discussions throughout.

Funding: Supported by a Physician-Scientist Early Career Award from the Howard Hughes Medical Institute (R.C.L.), a Claflin Distinguished Scholar Award from the Massachusetts General Hospital (R.C.L.), an International Research Scientist Development Award from the NIH K01TW007409 (J.B.H.), an NIH grant AI058935 (S.B.C, E.T.R, and F.Q), an NIH RO1 AI079198 (C.B. and L.M.S.), a FIC-NIH training grant D43 TW005572 (A.S.), an American Cancer Society Postdoctoral Fellowship (E.K.K.), the Packard Foundation Fellowship in Science and Engineering (P.C.S. and E.K.K.) and NIH Innovator award 1DP2OD006514-01 (P.C.S. and E.K.K.).

References

1. Lee K. The global dimensions of cholera. Global Change & Human Health. 2001; 2:6.
2. Chowdhury F, et al. Impact of Rapid Urbanization on the Rates of Infection by Vibrio cholerae O1 and Enterotoxigenic Escherichia coli in Dhaka, Bangladesh. PLoS Neglected Tropical Diseases. 2011; 5:e999. [PubMed: 21483709]
3. Mosley WH, McCormack WM, Ahmed A, Chowdhury AK, Barui RK. Report of the 1966–67 cholera vaccine field trial in rural East Pakistan. 2. Results of the serological surveys in the study population--the relationship of case rate to antibody titre and an estimate of the inapparent infection rate with Vibrio cholerae. Bulletin of the World Health Organization. 1969; 40:187. [PubMed: 5306539]
4. Glass RI, et al. Seroepidemiological studies of El Tor cholera in Bangladesh: association of serum antibody levels with protection. J Infect Dis. 1985; 151:236. [PubMed: 3968450]
5. Harris JB, et al. Cholera's western front. Lancet. 2010; 376:1961. [PubMed: 21112083]
6. Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. Lancet. 2004; 363:223. [PubMed: 14738797]
7. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. Cholera. Lancet. 2012; 379:2466. [PubMed: 22748592]
8. Harris JB, et al. Susceptibility to Vibrio cholerae Infection in a Cohort of Household Contacts of Patients with Cholera in Bangladesh. PLoS Neglected Tropical Diseases. 2008; 2:e221. [PubMed: 18398491]
9. Barua D, Paguio AS. ABO blood groups and cholera. Annals of human biology. 1977; 4:489. [PubMed: 603230]
10. Fellay J, et al. A whole-genome association study of major determinants for host control of HIV-1. Science. 2007; 317:944. [PubMed: 17641165]
11. Thomas DL, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009; 461:798. [PubMed: 19759533]
12. Tanaka Y, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nature genetics. 2009; 41:1105. [PubMed: 19749757]
13. Sabeti PC, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449:913. [PubMed: 17943131]
14. Park DJ, et al. Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:13052. [PubMed: 22826220]
15. Schwarzenbacher H, et al. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. BMC genomics. 2012; 13:48. [PubMed: 22289501]

16. Matson JS, Withey JH, DiRita VJ. Regulatory networks controlling Vibrio cholerae virulence gene expression. Infect Immun. 2007; 75:5542. [PubMed: 17875629]
17. Heidelberg JF, et al. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature. 2000; 406:477. [PubMed: 10952301]
18. Faruque SM, Mekalanos JJ. Pathogenicity islands and phages in Vibrio cholerae evolution. Trends Microbiol. 2003; 11:505. [PubMed: 14607067]
19. Mekalanos JJ, Rubin EJ, Waldor MK. Cholera: molecular basis for emergence and pathogenesis. FEMS Immunol Med Microbiol. 1997; 18:241. [PubMed: 9348159]
20. Nelson EJ, Harris JB, Morris JG Jr, Calderwood SB, Camilli A. Cholera transmission: the host, pathogen and bacteriophage dynamic. Nature reviews. Microbiology. 2009; 7:693.
21. Cassel D, Pfeuffer T. Mechanism of cholera toxin action: covalent modification of the guanyl nucleotide-binding protein of the adenylate cyclase system. Proceedings of the National Academy of Sciences of the United States of America. 1978; 75:2669. [PubMed: 208069]
22. Gill DM, Meren R. ADP-ribosylation of membrane proteins catalyzed by cholera toxin: basis of the activation of adenylate cyclase. Proceedings of the National Academy of Sciences of the United States of America. 1978; 75:3050. [PubMed: 210449]
23. Grossman SR, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. 2010; 327:883. [PubMed: 20056855]
24. Grossman SR, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013; 152:703. [PubMed: 23415221]
25. Frazer K, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851. [PubMed: 17943122]
26. Li JZ, et al. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science (New York, NY). 2008; 319:1100.
27. Teo YY, et al. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome research. 2009; 19:2154. [PubMed: 19700652]
28. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009; 461:489. [PubMed: 19779445]
29. Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol. 2005; 128:415. [PubMed: 15795887]
30. Moorjani P, et al. The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genetics. 2011; 7:e1001373. [PubMed: 21533020]
31. Patterson N, et al. Ancient admixture in human history. Genetics. 2012; 192:1065. [PubMed: 22960212]
32. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009; 19:1655. [PubMed: 19648217]
33. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061. [PubMed: 20981092]
34. Sklar P, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics. 2011; 43:977. [PubMed: 21926972]
35. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics. 2012; 28:1797. [PubMed: 22513993]
36. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15545. [PubMed: 16199517]
37. Harris MA, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 2004; 32:D258. [PubMed: 14681407]
38. Heitzmann D, Warth R. Physiology and pathophysiology of potassium channels in gastrointestinal epithelia. Physiol Rev. 2008; 88:1119. [PubMed: 18626068]
39. Saha D, et al. Incomplete correlation of serum vibriocidal antibody titer with protection from Vibrio cholerae infection in urban Bangladesh. J Infect Dis. 2004; 189:2318. [PubMed: 15181581]
40. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81:559. [PubMed: 17701901]

41. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American journal of human genetics. 1993; 52:506. [PubMed: 8447318]
42. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263. [PubMed: 15297300]
43. Shin OS, et al. LPLUNC1 modulates innate immune responses to Vibrio cholerae. J Infect Dis. 2011; 204:1349. [PubMed: 21900486]
44. Thanawastien A, Montor WR, Labaer J, Mekalanos JJ, Yoon SS. Vibrio cholerae proteome-wide screen for immunostimulatory proteins identifies phosphatidylserine decarboxylase as a novel Toll-like receptor 4 agonist. PLoS Pathog. 2009; 5
45. IPA from Ingenuity® Systems. 2012 http://www.ingenuity.com.
46. Rathinam VA, et al. TRIF Licenses Caspase-11-Dependent NLRP3 Inflammasome Activation by Gram-Negative Bacteria. Cell. 2012; 150:606. [PubMed: 22819539]
47. Kayagaki N, et al. Non-canonical inflammasome activation targets caspase-11. Nature. 2011; 479:117. [PubMed: 22002608]
48. Hayden MS, Ghosh S. NF-kappaB, the first quarter-century: remarkable progress and outstanding questions. Genes & development. 2012; 26:203. [PubMed: 22302935]
49. Xu S, Bayat H, Hou X, Jiang B. Ribosomal S6 kinase-1 modulates interleukin-1betainduced persistent activation of NF-kappaB through phosphorylation of IkappaBbeta. Am J Physiol Cell Physiol. 2006; 291:1336.
50. Ewing RM, et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. Molecular systems biology. 2007; 3:89. [PubMed: 17353931]
51. Flach CF, et al. Broad up-regulation of innate defense factors during acute cholera. Infect Immun. 2007; 75:2343. [PubMed: 17307946]
52. Berkey CD, Blow N, Watnick PI. Genetic analysis of Drosophila melanogaster susceptibility to intestinal Vibrio cholerae infection. Cell Microbiol. 2009; 11:461. [PubMed: 19046341]
53. Wang Z, Berkey CD, Watnick PI. The Drosophila protein mustard tailors the innate immune response activated by the immune deficiency pathway. J Immunol. 2012; 188:3993. [PubMed: 22427641]
54. Rahman MM, McFadden G. Modulation of NF-κB signalling by microbial pathogens. Nature Publishing Group. 2011; 9:291.
55. Chapman SJ, et al. NFKBIZ polymorphisms and susceptibility to pneumococcal disease in European and African populations. Genes Immun. 2010; 11:319. [PubMed: 19798075]
56. Chapman SJ, et al. Common NFKBIL2 polymorphisms and susceptibility to pneumococcal disease: a genetic association study. Crit Care. 2010; 14
57. Chapman SJ, et al. IkappaB genetic polymorphisms and invasive pneumococcal disease. Am J Respir Crit Care Med. 2007; 176:181. [PubMed: 17463416]
58. Smith PD, et al. Intestinal macrophages and response to microbial encroachment. Mucosal Immunol. 2011; 4:31. [PubMed: 20962772]
59. Nenci A, et al. Epithelial NEMO links innate immunity to chronic intestinal inflammation. Nature. 2007; 446:557. [PubMed: 17361131]
60. Anderson CA, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nature genetics. 2011; 43:246. [PubMed: 21297633]
61. Pinsk V, et al. Inflammatory bowel disease in the South Asian pediatric population of British Columbia. The American journal of gastroenterology. 2007; 102:1077. [PubMed: 17378907]
62. Carr I, Mayberry JF. The effects of migration on ulcerative colitis: a three-year prospective study among Europeans and first- and second- generation South Asians in Leicester (1991–1994). The American journal of gastroenterology. 1999; 94:2918. [PubMed: 10520845]
63. Baxter PS, Goldhill J, Hardcastle J, Hardcastle PT, Taylor CJ. Accounting for cystic fibrosis. Nature. 1988; 335:211. [PubMed: 3412484]
64. Rodman DM, Zamudio S. The cystic fibrosis heterozygote--advantage in surviving cholera? Med Hypotheses. 1991; 36:253. [PubMed: 1724059]

65. O'Grady SM, Lee SY. Molecular diversity and function of voltage-gated (Kv) potassium channels in epithelial cells. The international journal of biochemistry & cell biology. 2005; 37:1578. [PubMed: 15882958]
66. Kunzelmann K, et al. Expression and function of colonic epithelial KvLQT1 K+ channels. Clinical and experimental pharmacology & physiology. 2001; 28:79. [PubMed: 11153543]
67. Lohrmann E, et al. A new class of inhibitors of cAMP-mediated Cl- secretion in rabbit colon, acting by the reduction of cAMP-activated K+ conductance. Pflugers Arch. 1995; 429:517. [PubMed: 7617442]
68. Vallon V, et al. KCNQ1-dependent transport in renal and gastrointestinal epithelia. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:17864. [PubMed: 16314573]
69. MacVinish LJ, Hickman ME, Mufti DA, Durrington HJ, Cuthbert AW. Importance of basolateral K+ conductance in maintaining Cl- secretion in murine nasal and colonic epithelia. The Journal of physiology. 1998; 510(Pt 1):237. [PubMed: 9625880]
70. Mall M, et al. Cholinergic ion secretion in human colon requires coactivation by cAMP. Am J Physiol. 1998; 275:1274.
71. Yih LH, Peck K, Lee TC. Changes in gene expression profiles of human fibroblasts in response to sodium arsenite treatment. Carcinogenesis. 2002; 23:867. [PubMed: 12016162]
72. Mukherjee A, et al. Arsenic contamination in groundwater: a global perspective with emphasis on the Asian scenario. J Health Popul Nutr. 2006; 24:142. [PubMed: 17195556]
73. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207. [PubMed: 22699609]
74. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American journal of human genetics. 2009; 84:210. [PubMed: 19200528]
75. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011; 88:76. [PubMed: 21167468]
76. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & development. 2011; 25:1915. [PubMed: 21890647]
77. Karolchik D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32:D493. [PubMed: 14681465]
78. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic acids research. 2012; 40:D130. [PubMed: 22121212]
79. George-Chandy A, et al. Cholera toxin B subunit as a carrier molecule promotes antigen presentation and increases CD40 and CD86 expression on antigen-presenting cells. Infection and Immunity. 2001; 69:5716. [PubMed: 11500448]
80. Bourgeois-Daigneault MC, Thibodeau J. Autoregulation of MARCH1 expression by dimerization and autoubiquitination. J Immunol. 2012; 188:4959. [PubMed: 22508929]
81. Goto E, et al. c-MIR, a human E3 ubiquitin ligase, is a functional homolog of herpesvirus proteins MIR1 and MIR2 and has similar activity. The Journal of biological chemistry. 2003; 278:14657. [PubMed: 12582153]
82. Roy LD, et al. MUC1 enhances invasiveness of pancreatic cancer cells by inducing epithelial to mesenchymal transition. Oncogene. 2011; 30:1449. [PubMed: 21102519]
83. Fenner BJ, Scannell M, Prehn JH. Expanding the substantial interactome of NEMO using protein microarrays. PloS one. 2010; 5:e8799. [PubMed: 20098747]
84. Bouwmeester T, et al. A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. Nature Cell Biology. 2004; 6:97.
85. Garcia-Arnandis I, et al. High mobility group box 1 potentiates the pro-inflammatory effects of interleukin-1beta in osteoarthritic synoviocytes. Arthritis research & therapy. 2010; 12:R165. [PubMed: 20799933]
86. Stefanidakis M, Bjorklund M, Ihanus E, Gahmberg CG, Koivunen E. Identification of a negatively charged peptide motif within the catalytic domain of progelatinases that mediates binding to

leukocyte beta 2 integrins. The Journal of biological chemistry. 2003; 278:34674. [PubMed: 12824186]

87. Dumitriu IE, et al. Release of high mobility group box 1 by dendritic cells controls T cell activation via the receptor for advanced glycation end products. J Immunol. 2005; 174:7506. [PubMed: 15944249]

88. Popovic PJ, et al. High mobility group B1 protein suppresses the human plasmacytoid dendritic cell response to TLR9 agonists. J Immunol. 2006; 177:8701. [PubMed: 17142771]

Accessible Summary

Modern lessons from an ancient disease

A history of natural selection favoring resistance to an infectious disease should drive the emergence of underlying genetic variants that can be readily detected. We show this for cholera, an ancient, often fatal disease that likely exerted selection pressure on Bangladeshi populations. We combine a selection scan with an association study of cholera susceptibility, and translate genetic discoveries into clinically relevant biology.

We performed whole-genome scans of Bangladeshi families to identify 305 genomic regions of selection. These regions are highly enriched for potassium channel genes and genes in the NF-κB pathway, a master regulator of inflammation and immunity that is also involved in protecting the lining of the gut. We show, by comparing cholera-affected and healthy individuals, that top selected genes correlate with cholera susceptibility. These genes regulate an innate immune signaling pathway activated by *V. cholerae* and repeatedly targeted by selection.

Our combined selection and association approach identifies genes not previously implicated in cholera host response and highlights the role of innate immunity and intestinal homeostasis in disease pathogenesis. This approach of leveraging ancient history in genetic studies is applicable to many other ancient infectious diseases still circulating in the population today.

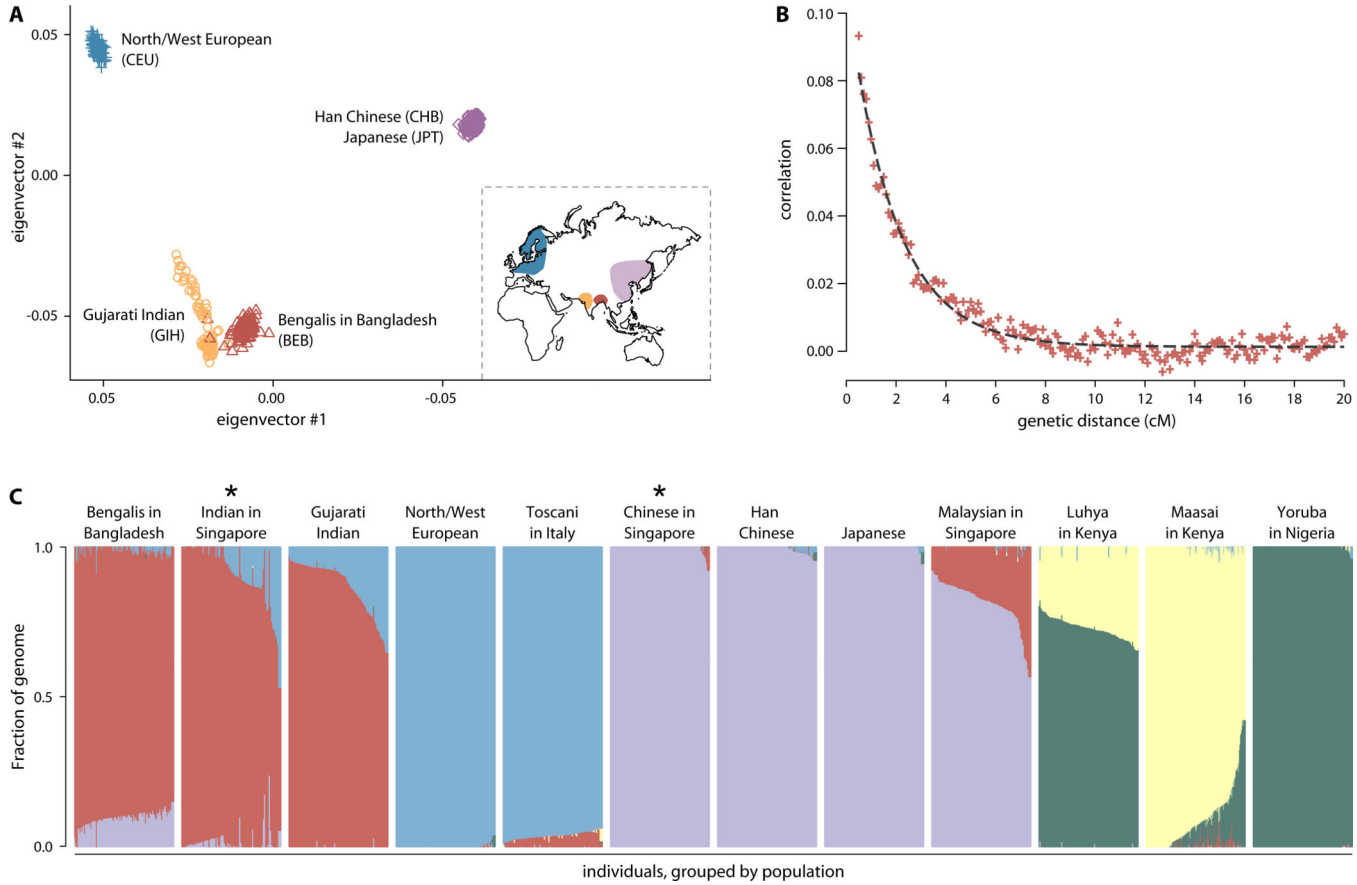

Figure 1. The Bengali population from Dhaka, Bangladesh has no evident structure and descends from ancient admixture between India and East Asia

(A) The first two principal components: Bangladeshis (red) cluster closest to the Gujarati Indians (orange), which is consistent with geography (inset map), and are shifted slightly towards the East Asian populations (purple). (B) The decline of admixture LD in the Bangladeshis suggests a founder event 52 +/− 2 generations ago (~500AD). History was estimated using the pulse admixture model based *ROLLOFF* method, which is robust to inaccurate parental populations and other modeling issues(31). (C) ADMIXTURE maximum likelihood estimation of ancestries (5 clusters, CV error=0.375) estimates the BEB are 89.7% +/− 2.4% “Indian” ancestry (red) and 9.3% +/− 2.6% “East Asian” ancestry (purple). The starred populations most closely fit the ancestral admixed populations.

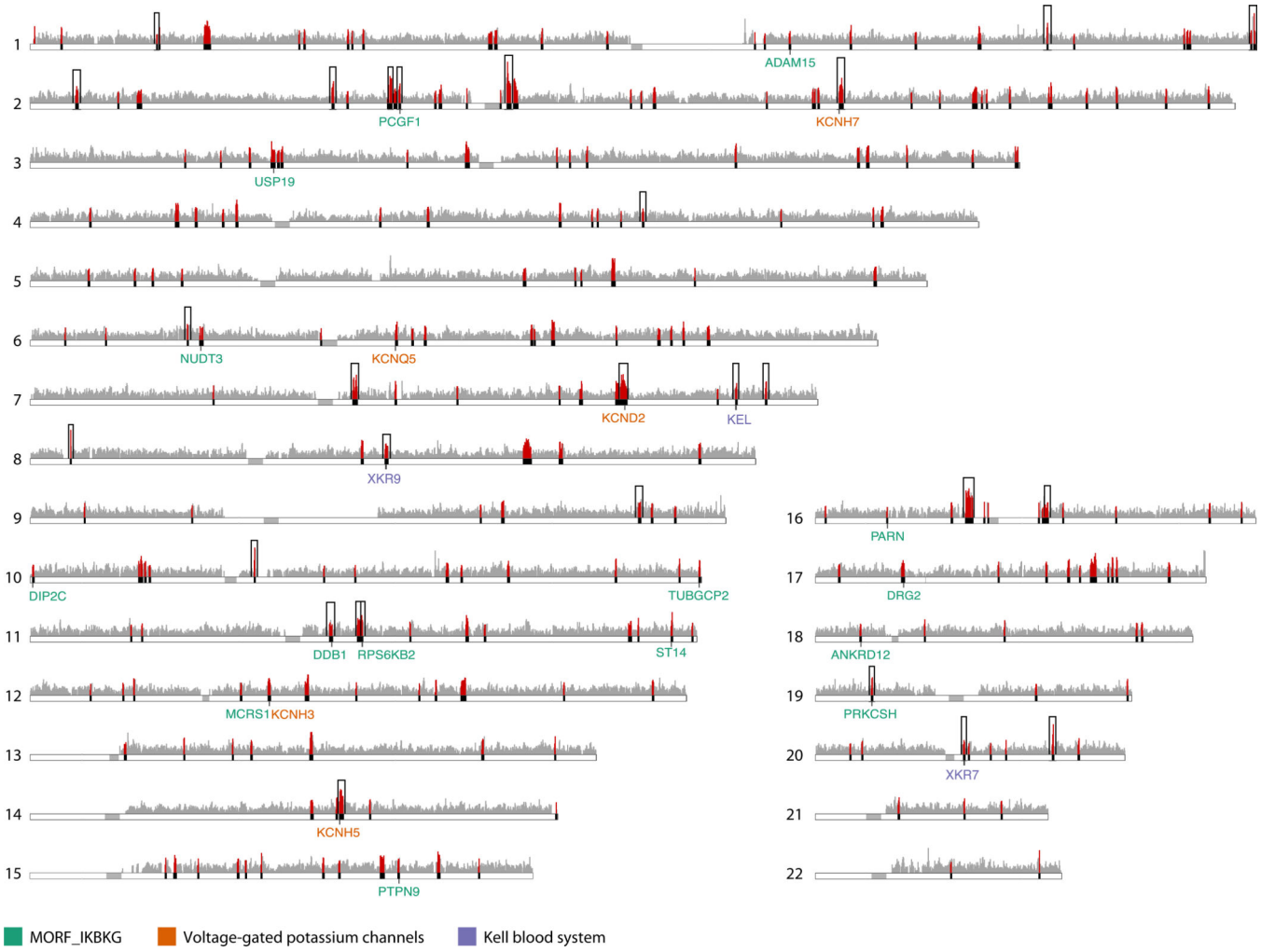

Figure 2. Signals of natural selection in the BEB population are enriched for genes from three distinct gene sets

CMS scan across the human autosomes identifies 305 selected regions (solid black bars with scores as red vertical bars; grey vertical bars show scores for all other regions). Genes in three enriched gene sets are shown with colored labels (see figure key). Candidate genes in 28 regions (black boxes) were tested for association with cholera susceptibility.

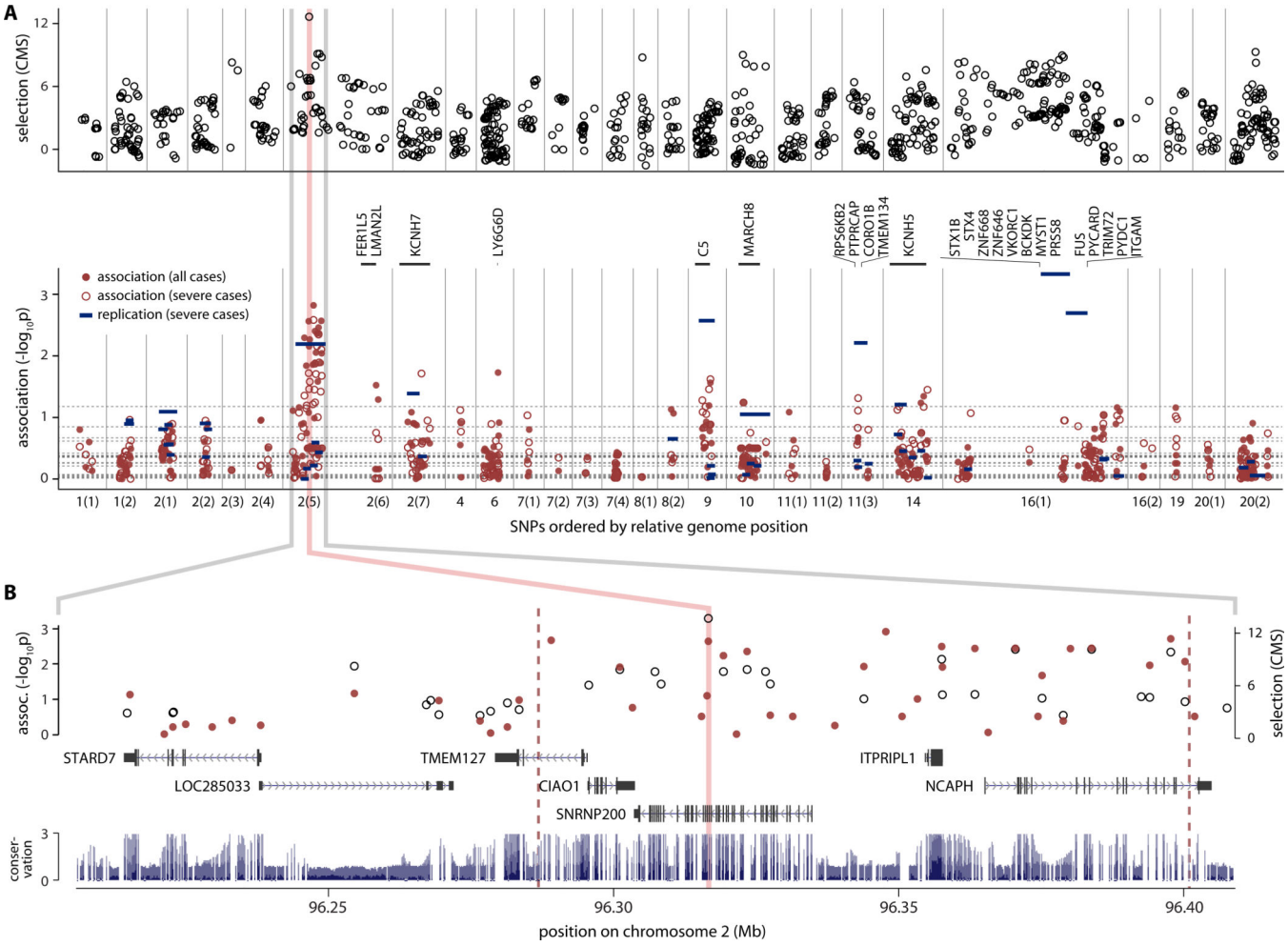

Figure 3. Top signal of selection in Bangladeshi population is associated with cholera susceptibility

(**A**) Association testing of 536 SNPs genotyped in 104 cholera cases and 167 controls in 28 candidate selected regions (empty black circles); the strongest selection signal in the genome corresponds to the region with the strongest association (red filled circles), exceeding the maximum association at 19 randomly selected SNPs (dashed horizontal lines) by more than 3 fold. SNPs in genes *LY6G6D* and *C5* also have p<0.05 (genes shown in black vertical text). Stricter phenotype definition (severe cholera; hollow red circles) yields 3 additional associated regions with p<0.05, including two containing potassium channel genes (*KCNH7*, *RPS6KB2* and *KCNH5*). Replication results are shown as blue bars. (**B**) The association (red filled circles) in region 2(5) overlaps the signal of selection (hollow black circles, right axis). The peak, bracketed by red dashed lines, encompasses 5 genes. The SNP with the highest CMS score in the genome (pink line) sits in an intron of the gene *SNRNP200*. The region under this peak is well conserved across placental mammals (blue bars).

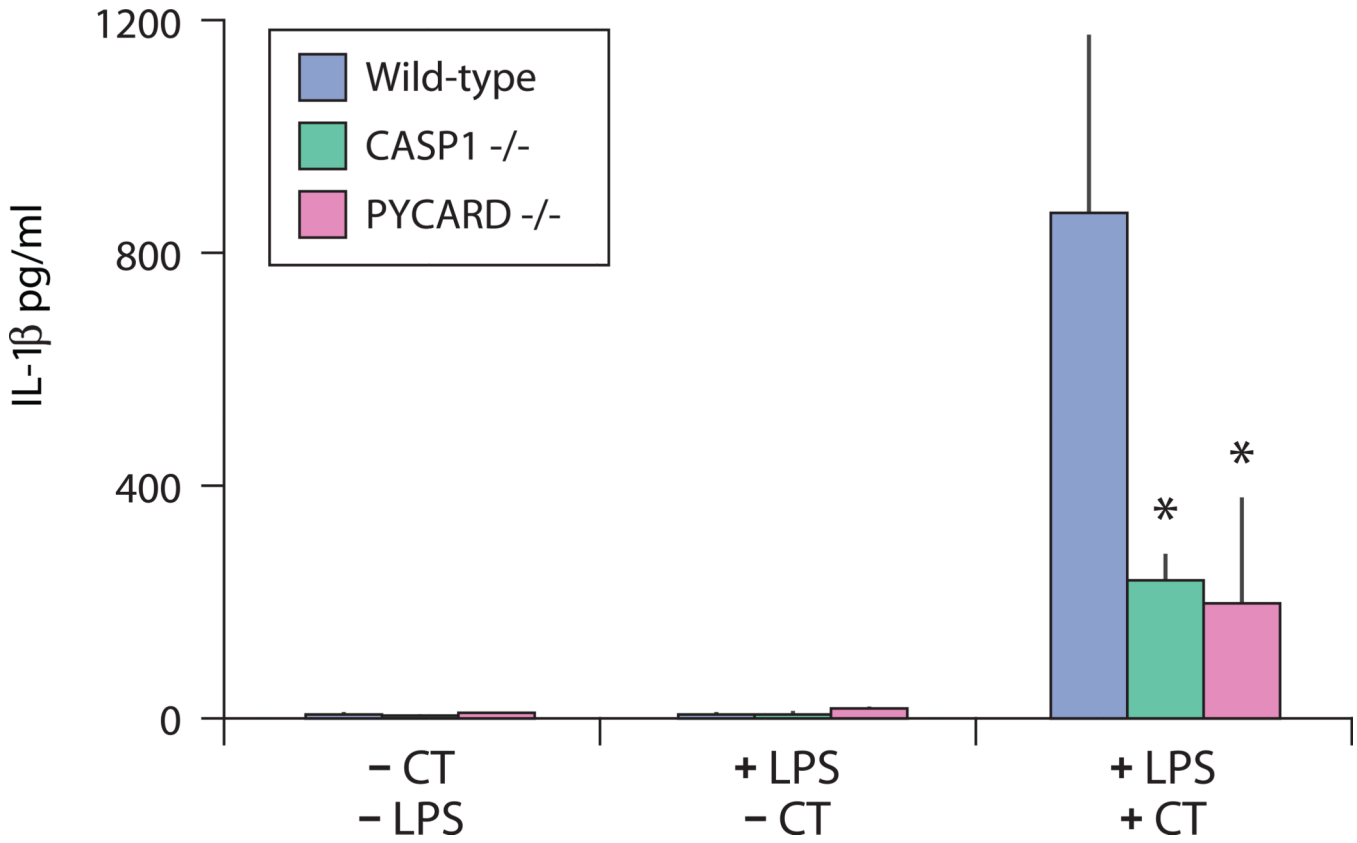

Figure 4. Cholera toxin induces caspase-dependent IL-1β secretion consistent with inflammasome activation

Cholera toxin (CT) induces robust IL-1β secretion in lipopolysaccharide (LPS) primed mouse macrophages but not in macrophages deficient in caspase-1 (green) or PYCARD (purple), key components of inflammasome signaling (starred, p=0.0001 and 0.0005, respectively, when compared to wild-type using Student's t test). The average (bars) and standard error of the mean (vertical black lines) are shown for three experimental replicates, each with three sampling replicates.

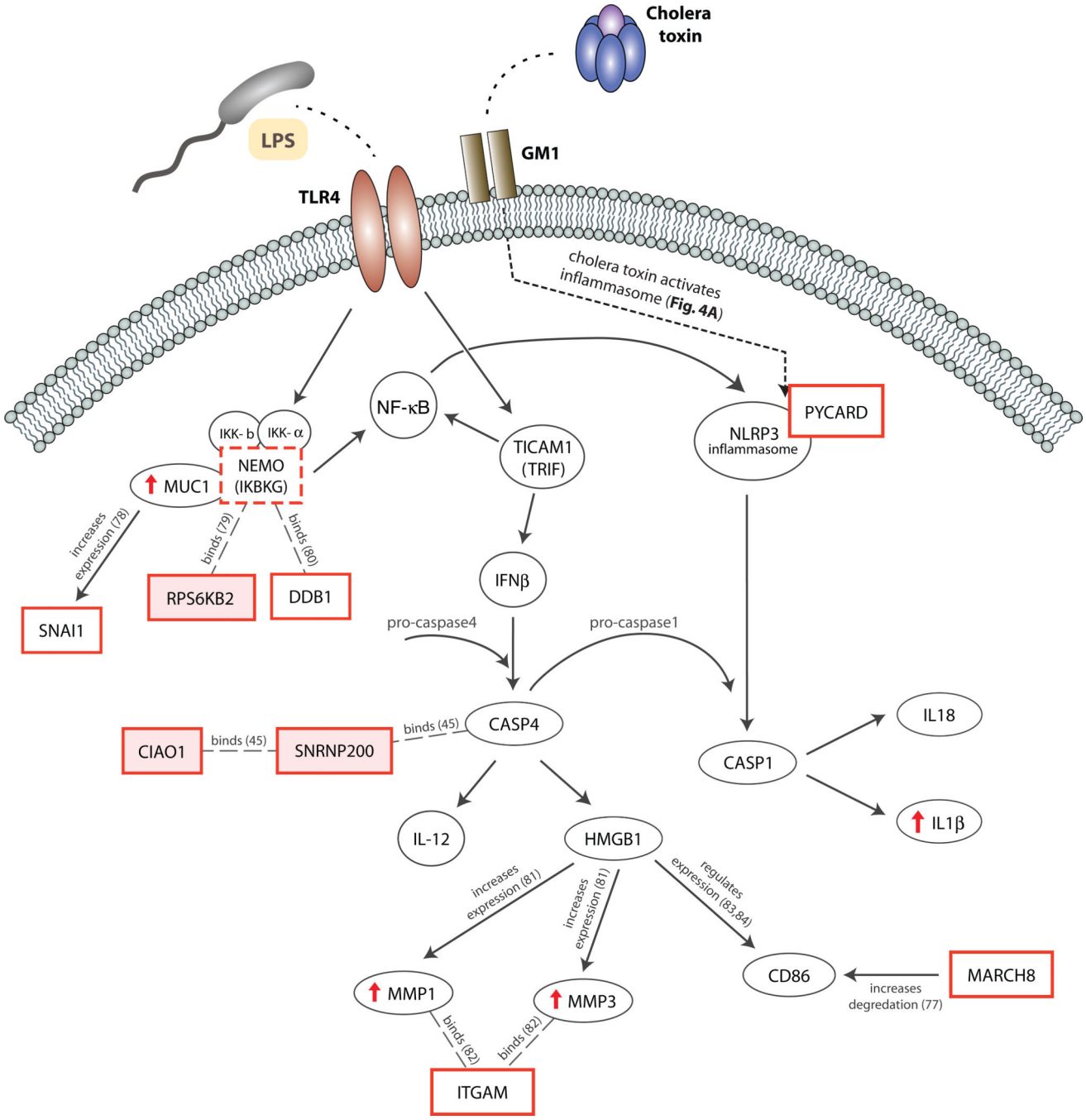

Figure 5. Human innate immune signaling pathways are targeted by selective pressures, including cholera, in the BEB population

We developed a model of the cholera-related immune pathways under selection in the BEB population by mapping all interactions cataloged by Ingenuity's IPA software between strongly selected genes, including *IKBKG*, and genes shown to respond to cholera toxin (see methods). Upon exposure to *V. cholerae*, toll-like receptor 4 (TLR4) and GM1 ganglioside interact with *V. cholerae* LPS and CT, resulting in coordinated activation of the NLRP3 inflammasome and the NF-κB regulatory complex, through multimerization of PYCARD

and assembly of the NEMO complex, respectively. NEMO is encoded by *IKBKG* (red dashed box), the central gene in a gene set that is significantly enriched for selection in the BEB. TLR4 stimulates production of interferon β (IFNβ), which cleaves procaspase-4 into mature caspase-4. The NLRP3/PYCARD complex cleaves pro-caspase-1 into mature caspase-1. Caspase-1 and caspase-4, central mediators of the inflammasome pathway, then facilitate cell death and release of pro-inflammatory cytokines, such as HMGB1, IL-18, and IL-1β. Genes in selected regions (red boxes), including three specifically associated with cholera susceptibility (red shaded boxes) interact with regulatory components of the proposed pathway. Genes that are significantly upregulated during acute cholera are denoted with red arrows. Interactions shown as arrows or dashed lines; those found using Ingenuity Knowledge Base are labeled with references in parentheses(50, 81–88).

Table 1

Gene sets significantly enriched in candidate regions of natural selection in the BEB population.

				North/West					
	BEB population			East Asian		European		Yoruban	
	P_{set}	P_{exp}	Genes	P_{set}	P_{exp}	P_{set}	P_{exp}	P_{set}	P_{exp}
Molecular Signatures Database MSigDB.c4 (851 gene sets with >10 genes)									
MORF_IKBKG (110 genes)	5.2E-05	0.017	PARN, DDB1, DIP2C, ANKRD12, USP19, ADAM15, NUDT3, PTPN9, DRG2, TUBGCP2, MCRS1, PRKCSH, PCGF1, RPS6KB2,PITPNM1, ST14	0.561	1.000	0.436	1.000	0.168	1.000
Potassium channel genes*									
Voltage-gated and Ca2+ activated K+ channels (47 genes)	1.6E-03	0.008	KCNH3, KCNQ5, KCNH5, KCND2, KCNH7			0.160	0.180		
Voltage-gated and Ca2+ activated K+ channels: EAG, ERG, ELK related (8 genes)	5.2E-03	0.030	KCNH3, KCNH5, KCNH7						
Blood systems									
Kell system: KELL and Kell complex genes (8 genes)	4.4E-03	0.007	XKR7, XKR9, KEL	0.025	0.029				

We used INRICH to calculate empirically the significance of the enrichment for each gene set (P_{set}) and the experiment-wide significance corrected for the number of gene sets tested (P_{exp})(35). For the Gene Ontology and potassium channel analyses (marked with *), no significant enrichments were found across all candidate selected regions and the analysis was rerun just on regions with normalized CMS_{GW} scores >5. No P value is shown for sets with 0 selected genes. See Table S7 for all sets with $P_{set}<0.05$.