Supplemental Data

Pathogen-Driven Selection and Worldwide HLA Class I Diversity

Franck Prugnolle, Andrea Manica, Marie Charpentier, Jean-François Guégan, Vanina Guernier, and François Balloux

Supplemental Experimental Procedures

Human Genetic Diversity at HLA Class I Genes (A, B, and C)

We retrieved allelic frequencies for HLA genes from 61 human populations distributed all over the world from the dbMHC database (National Center for Biotechnology Information [NCBI]; http://www.ncbi.nlm.nih.gov/mhc/). More details about those populations are presented in Table S1. All individuals were genotyped with the same standardized sequence-specific oligonucleotide hybridization (SSO) method (see database for more details). We limited the study to class I genes (A, B, and C). From this, we computed the genetic diversity at HLA class I genes within each population, excluding populations with unclear recent ancestry (e.g., African Americans).

Genetic diversity was computed as the expected heterozygosity unbiased estimator

$$[2n/(2n-1)]\left[1-\sum_{i=1}^{k}\hat{p}_i^2\right]$$

(where n is the number of individuals within each population, k is the number of distinct alleles, and $\hat{p}_i$ is the relative allele frequency of allele i in the sample). For this computation, the frequencies of all alleles, even those that were distinct by only one nucleotide substitution, have been considered. One nucleotide substitution can have a strong effect on the efficiency of HLA proteins for pathogen defense, as shown in several studies (see, for example, [S1]), and, hence, two alleles distinct by one nucleotide substitution have been considered equivalent to two alleles distinct by 100 substitutions. Incidentally, this approach also facilitates comparisons between our HLA diversity measure and microsatellite diversity as measured in our previous study relating human genetic diversity and geographic distance from Africa [S2].

Presence/Absence Matrix of Pathogens and Pathogen Richness

A matrix of presence/absence of pathogens for 224 countries was extracted from the GIDEON database (Global Infectious Diseases and Epidemiology Network; http://www.gideononline.com), published by the Tel Aviv Medical Centre. In this database, presence/absence of pathogens in each country is based on the standardized statistics reported by the World Health Organization, National Health Ministries, abstracts of major international meetings, and epidemiology journals. The database is updated weekly.

The matrix of species presence/absence provides distributional information about which species naturally occur in which countries without any reference to their prevalence. Only species that are naturally transmitted in the countries were recorded, meaning that cases of disease transmission linked to immigration or tourism were not taken into account. Pathogen species that were recently eradicated from a country were also considered to be present (e.g., autochthonous malaria in Croatia was officially eradicated in 1962). Obviously, some disease agents are recorded in all countries (e.g., measles and influenza), whereas some others are only present in few, if not only one, countries (endemic species such as the Sabia virus, responsible for the Brazilian hemorrhagic fever). More information can be found at http://www.cyinfo.com/ and in [S3]. Because HLA class I genes (A, B, and C) are mainly involved in the presentation and recognition of intracellular pathogens [S4], we only considered intracellular disease agents (viruses and obligate and facultative intracellular bacteria and protozoa with at least one intracellular stage; see Tables S3–S5). Because no standardized information on pathogen prevalence was available, we simply compiled pathogen richness as the number of distinct intracellular disease agents (number of pathogen species) known from countries from which human populations with HLA information originate.

Human Demography, Neutral Diversity, and Geographic Distance from Africa

Geographic distance of current populations from East Africa along likely ancient colonization routes is strongly correlated to their neutral genetic diversity (H_S^n) ($r^2 = 85\%$) [S2]. This correlation is even stronger ($r^2 = 87\%$) when we use $H_S^{n*} = \text{Log}\,[H_S^n/(1-H_S^n)]$ instead of H_S^n. Geographic distance from Africa can therefore be used to disentangle the effect of past migrations from selection on HLA diversity of human populations. To estimate geographic distance of each population from the postulated East African origin of modern humans, we used a newly developed algorithm based on graph theory. The method is fully described in [S2].

Statistical Analyses

Because both geographic distances from East Africa and pathogen richness are used as two independent potential explanatory variables of the HLA diversity in human populations, we first tested for their potential collinearity. No relationship between pathogens and geographic distances from East Africa was found (e.g., for total pathogen richness, $p = 0.36$), and hence these two predictors were treated as independent.

The PDBS hypothesis was tested with S-PLUS 2000 (Mathsoft) by fitting a weighted Generalized Linear Model with a Gaussian error structure [S5], after transformation of the response variable H_S^{HLA} (HLA genetic diversity) into the synthetic variable $H_S^{HLA*} = \text{Log}\,[H_S^{HLA}/(1-H_S^{HLA})]$. We fitted the model $H_S^{HLA*} = \text{Dist. Africa} + \text{Path. Rich.} + \text{Constant}$, where: H_S^{HLA*} corresponds to the transformed HLA diversity computed in each population and weighted by the inverse of the number of observations within each country to avoid giving too much weight to populations that come from the same country and, hence, share the same predictor variables; Dist. Africa is the geographic distance of HLA-typed populations to the postulated East African origin; and Path. Rich. is total pathogen richness, that is, the total number of intracellular human pathogen species known from within each country. This model allows us to first account for the effect of past human colonization history on HLA diversity (via the variable Dist. Africa) and to test specifically for the effect of pathogen richness on the remaining unexplained variance. The significance of the different variables was assessed with an F-test after a backward stepwise procedure [S5]. When an effect of pathogen richness on HLA diversity was detected, we a posteriori tested for homoscedasticity between samples, with Levene's tests [S5]. No significant deviation from homoscedasticity was detected in any of our models. Cook distances were also used to graphically assess the influence of each observation on the regression. When observations seemed to be overinfluential, they were deleted from the dataset, and the regression analysis was rerun. This procedure never changed the results qualitatively.

Because of the collinearity between virus, bacteria, and protozoa richness, the PDBS hypothesis was tested separately for each aetiological group through independent weighted regression models. For each model, Dist. Africa was again entered as a covariate to account for the effect of human settlement history on HLA diversity.

An alternative, more complete geographic model including sampling latitude was also considered because latitude is a good proxy for climate and could thus provide some information on the environment in which the population evolved. In this model, the absolute latitude of the countries from which populations come was entered as another explanatory variable to test whether geography alone

Table S1. Human HLA-Typed Populations Used in the Study (from the dbMHC Database)

Population	Sampling Site	Dist. Africa (km)	H_s^{HLA}	Pathogen Richness
African Populations				
Kenyan 142	Kenya	950	A (0.945), B (0.961), C (0.899)	94 (47, 10, 31, 6)
Kenyan Highlander	Kenya	950	A (0.929), B (0.948), C (0.889)	94 (47, 10, 31, 6)
Kenyan Lowlander	Kenya	950	A (0.941), B (0.945), C (0.905)	94 (47, 10, 31, 6)
Doggon	Mali	4653	A (0.897), B (0.958), C (0.843)	83 (38, 10, 30, 5)
Mandenka	Senegal	5904	A (0.922), B (0.955)	91 (47, 8, 30, 6)
Zulu	South Africa	4665	A (0.951), B (0.952), C (0.927)	89 (43, 11, 29, 6)
Ugandan	Uganda	1001	A (0.935), B (0.968), C (0.917)	97 (49, 9, 33, 6)
Zambian	Zambia	2781	A (0.899), B (0.952), C (0.904)	79 (37, 8, 29, 5)
Shona	Zimbabwe	3362	A (0.920), B (0.944), C (0.918)	82 (41, 8, 29, 4)
Chaouya	Morocco	5393	A (0.930), B (0.967)	75 (34, 8, 28, 5)
Metalsa	Morocco	5393	A (0.926), B (0.954)	75 (34, 8, 28, 5)
American Populations				
Yupik	Alaska (Canada)	14831	A (0.622), B (0.842), C (0.778)	89 (44, 9, 32, 4)
Brazilian	Brazil	28169	A (0.908), B (0.965)	104 (53, 10, 33, 8)
Guarani-Kaiowa	Brazil	28169	A (0.707), B (0.987), C (0.768)	104 (53, 10, 33, 8)
Guarani-Nandewa	Brazil	28169	A (0.765), B (0.932), C (0.816)	104 (53, 10, 33, 8)
Seri	Mexico	21051	A (0.653), B (0.756)	92 (41, 11, 32, 8)
Mestizos	Mexico	21051	A (0.879), B (0.978), C (0.887)	92 (41, 11, 32, 8)
Amerindian	United States	19311	A (0.872), B (0.946), C (0.915)	114(60, 13, 35, 6)
Bari	Venezuela	25748	A (0.583), B (0.753), C (0.775)	87 (47, 8, 30, 7)
East Asian Populations				
Okinawan	Hawaii (Japan)	12706	A (0.819), B (0.908), C (0.862)	81 (36, 11, 29, 5)
Ryukyan	Japan	12706	A (0.827)	81 (36, 11, 29, 5)
Korean 200	Korea	11148	A (0.874), B (0.946)	78 (34, 11, 30, 3)
Malay	Singapore (Malaysia)	12743	A (0.893), B (0.954), C (0.908)	89 (42, 11, 31, 5)
Chinese	Rep. of China	8820	A (0.855), B (0.934), C (0.881)	94 (42, 12, 34, 6)
Han-Chinese 149	Singapore (Rep. of China)	8820	A (0.862), B (0.956)	94 (42, 12, 34, 6)
Han-Chinese 572	Rep. of China	8820	A (0.853), B (0.991)	94 (42, 12, 34, 6)
Singapore (Chinese)	Singapore (Rep. of China)	8820	A (0.869), B (0.936)	94 (42, 12, 34, 6)
Tuva	Russia	8388	A (0.897), B (0.959), C (0.926)	96 (46, 14, 31, 5)
Buryat	Russia	9732	A (0.910)	96 (46, 14, 31, 5)
Ami 97	Taiwan	11165	A (0.552), B (0.774), C (0.816)	77 (34, 9, 29, 5)
Atayal	Taiwan	11165	A (0.592), B (0.792), C (0.825)	77 (34, 9, 29, 5)
Bunun	Taiwan	11165	A (0.608), B (0.832), C (0.816)	77 (34, 9, 29, 5)
Hakka	Taiwan	11165	A (0.800), B (0.915), C (0.863)	77 (34, 9, 29, 5)
Minnan	Taiwan	11165	A (0.838), B (0.911), C (0.872)	77 (34, 9, 29, 5)
Paiwan 51	Taiwan	11165	A (0.254), B (0.797), C (0.695)	77 (34, 9, 29, 5)
Pazeh	Taiwan	11165	A (0.790), B (0.906), C (0.881)	77 (34, 9, 29, 5)
Puyuma 49	Taiwan	11165	A (0.579), B (0.882), C (0.838)	77 (34, 9, 29, 5)
Rukai	Taiwan	11165	A (0.403), B (0.843), C (0.729)	77 (34, 9, 29, 5)
Saisiat	Taiwan	11165	A (0.638), B (0.603), C (0.525)	77 (34, 9, 29, 5)
Siraya	Taiwan	11165	A (0.737), B (0.917), C (0.889)	77 (34, 9, 29, 5)
Thao	Taiwan	11165	A (0.594), B (0.853), C (0.794)	77 (34, 9, 29, 5)
Toroko	Taiwan	11165	A (0.715), B (0.771), C (0.789)	77 (34, 9, 29, 5)
Tsou	Taiwan	11165	A (0.373), B (0.853), C (0.803)	77 (34, 9, 29, 5)
Yami	Taiwan	11165	A (0.581), B (0.763), C (0.758)	77 (34, 9, 29, 5)
Thai	Singapore (Thailand)	9890	A (0.873), B (0.938), C (0.899)	84 (38, 11, 31, 4)
European Populations				
Bulgarian	Bulgaria	4469	A (0.819), B (0.905), C (0.905)	83 (39, 9, 30, 5)
Croatian	Croatia	4387	A (0.874), B (0.946)	83 (37, 11, 30, 5)
Finn 90	Finland	6764	A (0.799), B (0.968), C (0.901)	76 (37, 8, 28, 3)
Irish	Ireland	7244	A (0.848), B (0.909), C (0.886)	72 (34, 7, 27, 4)
Czech	Czech Republic	5461	A (0.861), B (0.959), C (0.914)	80 (38, 8, 30, 4)
Georgian	Rep. of Georgia	4126	A (0.846), B (0.941), C (0.909)	84 (41, 7, 31, 5)
Middle East Populations				
Arab Druze	Israel	2802	A (0.935), B (0.960), C (0.912)	83 (40, 7, 31, 5)
Omani	Oman	5052	A (0.907), B (0.929)	76 (34, 11, 27, 4)
Kurdish	Rep. of Georgia	4126	A (0.936), B (0.954), C (0.896)	84 (41, 7, 31, 5)

(continued)

Table S1. Continued

Population	Sampling Site	Dist. Africa (km)	H_s^{HLA}	Pathogen Richness
Oceanian Populations				
American Samoan	A. Samoa	22687	A (0.825), B (0.879), C (0.879)	57 (30, 3, 22, 2)
Cape York	Australia	16102	A (0.801), B (0.871), C (0.849)	85 (40, 10, 32, 3)
Groote Eylandt	Australia	16102	A (0.747), B (0.845), C (0.812)	85 (40, 10, 32, 3)
Kimberley	Australia	16102	A (0.515), B (0.755), C (0.742)	85 (40, 10, 32, 3)
Yuendumu	Australia	16102	A (0.700), B (0.837), C (0.862)	85 (40, 10, 32, 3)
Filipino	Philippines	14470	A (0.842), B (0.945), C (0.846)	81 (39, 11, 28, 3)
Ivatan	Philippines	14470	A (0.781), B (0.887), C (0.842)	81 (39, 11, 28, 3)

Sampling Site gives the country where the populations have been sampled from. When the country of origin of the populations differs from the sampling location (e.g., for ethnic groups that have migrated recently), this information is indicated in parentheses. The "country of origin" is the country from which epidemiological data have been used, except for the Yupik population, for which epidemiological data were recorded from Canada because no epidemiological data were available from Alaska. Dist. Africa is the distance of each population from East Africa. H_s^{HLA} gives the HLA class I gene diversities, when data are available. Pathogen Richness refers to the total pathogen richness observed within each country, plus the richness in viruses, in obligate and facultative intracellular bacteria, and in intracellular protozoa respectively (in parentheses).

could explain the observed genetic diversity observed at HLA genes (see Table 2 in the main text for results).

Finally, to ensure that pathogen diversity is not linked to genetic diversity in general, we fitted the same regression models to a set of markers that should not be influenced by the richness of intracellular pathogens: one locus of HLA class II (HLA DRB1, for which allele frequencies were retrieved from the dbMHC database) and ten microsatellite loci randomly chosen from the 377 reported in Rosenberg et al. [S6] and previously used in [S2]. For these loci (see Table 2 for details), we expect only a significant relationship between genetic diversity and geographic distance from East Africa, but no relationship between genetic diversity and pathogen richness.

Table S2. Regressions between HLA A and B Genetic Diversity (H_S^{HLA*}), Geographic Distance from Africa (Dist. Africa), and Intracellular Pathogen Species Richness

		HLA A	HLA B
n		48	48
Model I			
Dist. Africa	r^2	39%***	10%*
Path. Rich.	r^2	6%*	14%**
Model II			
Dist. Africa	r^2	39%***	10%*
Viruses	r^2	8%*	14%**
Model III			
Dist. Africa	r^2	39%***	10%*
Bacteria O	r^2	1.9% ns	6% ns
Model IV			
Dist. Africa	r^2	39%***	17%***
Bacteria F	r^2	$<$1% ns	7%*
Model V			
Dist. Africa	r^2	39%***	10%*
Protozoa	r^2	6.5%*	15%**

In this table, we reanalyze the HLA A and B datasets after subsetting them to include only the populations that are available for HLA C ($n = 48$). Regression models I–V were fitted independently according to the procedure detailed in Statistical Analyses. Path. Rich., Bacteria O, and Bacteria F denote species richness of all intracellular pathogens and obligate and facultative intracellular bacteria, respectively. n represents the number of populations genotyped, and r^2 the proportion of variance explained by each independent variable. For all regressions between pathogen richness and HLA class I diversity, the sign of slopes was always positive. P values for *F* tests: *** $<$ 0.001; ** $<$ 0.01; * $<$ 0.05; ns, non-significant.

Supplemental References

S1. Gao, X.J., Nelson, G.W., Martin, M.P., Phair, J., Kaslow, R., Goedert, J.J., Buchbinder, S., Hoots, K., Vlahov, D., O'Brien, S.J., et al. (2001). Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. N. Engl. J. Med. *344*, 1668–1675.

S2. Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. Curr. Biol. *15*, R159–R160.

S3. Guernier, V., Hochberg, M.E., and Guegan, J.F. (2004). Ecology drives the worldwide distribution of human diseases. PloS Biol. *2*: e141 10.1371/journal.pbio.0020141.

S4. Hughes, A.L., and Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. *32*, 415–435.

S5. Venables, W.N., and Ripley, B.D. (1999). Modern Applied Statistics with S-Plus (Berlin: Springer).

S6. Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. Science *298*, 2381–2385.

Table S3. Family, Genus, Species Common Name, and Name of the Associated Disease for Viruses Reported from at Least One Country Included in the Study

Disease Name	Virus
Adenovirus	Adenoviridae, Adenovirus, Enteric strains classified in genus Mastadenovirus
Argentine hemorrhagic fever	Arenaviridae, Tacaribe complex, Arenavirus: Junin virus
Barmah Forest disease	Togaviridae, Alphavirus: Barmah Forest virus
Bolivian Hemorrhagic fever	Arenaviridae, Tacaribe complex. Arenavirus: Machupo virus
Brazilian hemorrhagic fever	Arenaviridae, Tacaribe complex, Arenavirus: Sabia virus
California encephalitis group	Bunyaviridae, Orthobunyavirus: La Crosse, California encephalitis, Jamestown Canyon
Chikingunya	Togaviridae, Alphavirus: Chikingunya virus.
Colorado tick fever	Reoviridae, Coltivirus: Colorado tick fever virus
Conjunctivitis-viral	Picornavirus, Adenovirus
Cowpox	Poxviridae, Orthopoxvirus: Cowpox virus
Crimean-Congo hemorrhagic fever	Bunyaviridae, Nairovirus: CCHF virus
Cytomegalovirus infection	Herpesviridae, Betaherpesvirinae: Human herpesvirus 5 (Cytomegalovirus)
Dengue	Flaviviridae, Flavivirus: Dengue virus
Eastern equine encephalitis	Togaviridae, Alphavirus: Eastern equine encephalitis virus
Ebola fever	Mononegavirales, Filoviridae, Filovirus: Ebola virus
Enterovirus infection	Picornaviridae: Coxsackievirus, ECHO virus, Enterovirus, Parechovirus
Gastroenteritis—viral	Calicivirus (Norwalk, Hawaii, Sapporo, Snow Mountain, Norovirus); Torovirus or Astrovirus
Hantavirus infections-Old World	Bunyaviridae, Hantavirus: Hantaan, Puumala, Dobrava/Belgrade, Saaremaa & Seoul viruses
Hantavirus pulmonary syndrome	Bunyaviridae, Hantavirus: Sin Nombre, Black Creek Canal, Bayou, New York-1, Andes, etc.
Hendra virus disease	Paramyxoviridae, Megamyxovirus [Henipavirus]: Hendra virus
Hepatitis A	Picornaviridae, Hepatovirus: Hepatitis A virus
Hepatitis B	Hepadnaviridae, Orthohepadnavirus: Hepatitis B virus
Hepatitis C	Flaviviridae, Hepacivirus: Hepatitis C virus
Hepatitis D	Deltavirus: Hepatitis D virus
Hepatitis E	Caliciviridae: Hepatitis E virus
Hepatitis G	Flaviviridae, Hepacivirus: Hepatitis G virus. HGBV-A, B, and C appear to be related.
Herpes simplex encephalitis	Herpesviridae, Alphaherpesvirinae, Simplexvirus: Herpesvirus (usually type 1)
Herpes zoster	Herpesviridae, Alphaherpesvirinae: Varicella-zoster virus
Herpesvirus simiae infection	Herpesvirus simiae infection
HIV infection—initial illness	Retroviridae, Lentivirinae: Human Immunodeficiency Virus
Ilheus	Flaviviridae, Flavivirus
Influenza	Orthomyxoviridae, Orthomyxovirus: Influenza virus
Japanese encephalitis	Flaviviridae, Flavivirus: Japanese encephalitis virus
Karelian fever	Togaviridae, Alphavirus: Sindbis virus
Kyasanur Forest disease	Flaviviridae, Flavivirus: Kyasanur Forest disease virus
Lassa fever	Arenaviridae, Arenavirus: Lassa virus
Louping ill	Flaviviridae, Flavivirus: Louping ill virus
Lymphocytic choriomeningitis	Arenaviridae, Arenavirus: Lymphocytic choriomeningitis virus
Marburg virus disease	Mononegavirales, Filoviridae, Filovirus: Marburg virus
Mayaro	Togaviridae, Alphavirus: Mayaro virus
Measles	Paramyxoviridae, Paramyxovirinae, Morbillivirus: Measles virus
Meningitis—aseptic(viral)	Picornaviridae, enteroviruses
Monkeypox	Poxviridae, Orthopoxvirus: Monkeypox virus
Mononucleosis—infectious	Herpesviridae. Gammaherpesvirinae, Lymphocryptovirus: Human herpesvirus
Mumps	Paramyxoviridae, Paramyxovirinae, Rubulavirus: Mumps virus
Murray Valley encephalitis	Flaviviridae, Flavivirus. Murray Valley encephalitis virus
New World phleboviruses	Bunyaviridae, Orthobunyavirus: Alenquer, Arboledas, Bujaru, Cacao, Candiru, Chagres, Punta Toro
Nipah virus disease	Paramyxoviridae, Megamyxovirus [Henipavirus]: Nipah virus
O'nyong nyong	Togaviridae, Alphavirus: O'nyong nyong virus
Omsk hemorrhagic fever	Flaviviridae, Flavivirus: Omsk hemorrhagic fever virus
Orf	Poxviridae, Parapoxvirus: Orf virus
Oropouche	Bunyaviridae, Orthobunyavirus, Simbu group virus: Oropouche virus
Parainfluenza virus infection	Paramyxoviridae: Respirovirus-Rubulavirus
Parvovirus B19 infection	Parvoviridae, Parvovirinae: Erythrovirus B19
Pleurodynia	Picornaviridae: Coxsackievirus
Pogosta disease	Togaviridae, Alphavirus: Sindbis virus
Poliomyelitis	Picornaviridae, Picornavirus: Polio virus
Powassan	Flaviviridae, Flavivirus: Powassan virus
Pseudocowpox	Poxviridae, Parapoxvirus: Pseudocowpox virus
Rabies	Rhabdoviridae, Mononegavirales, Lyssavirus: Rabies virus
Respiratory syncytial virus infection	Paramyxoviridae, Pneumovirinae: Human respiratory syncytial virus
Rift Valley fever	Bunyaviridae, Phlebovirus: Rift Valley fever virus
Rocio	Flaviviridae, Flavivirus: Rocio virus
Roseola or human herpesvirus 6	Herpesviridae, Betaherpesvirinae, Roseolovirus: Herpesvirus 6 (Herpesvirus 7 as well)
Ross River disease	Togaviridae, Alphavirus: Ross River virus

(continued)

Table S3. Continued

Disease Name	Virus
Rotavirus infection	Reoviridae: Rotavirus
Rubella	Togaviridae: Rubella virus
Sandfly fever	Bunyaviridae, Phlebovirus: Sandfly fever virus (at least three types)
Sindbis	Togaviridae, Alphavirus: Sindbis virus
Smallpox	Poxviridae, Orthopoxvirus: Variola virus
Spondweni	Flaviviridae, Flavivirus: Spondweni virus
St. Louis encephalitis	Flaviviridae, Flavivirus: St. Louis encephalitis virus
Tanapox virus disease	Poxviridae, Yatapoxvirus: Tanapox virus
Thogoto	Orthomyxoviridae, Thogotovirus: Thogoto virus
Tick-borne encephalitis: Central European	Flaviviridae, Flavivirus: Central European encephalitis virus
Tick-borne encephalitis: Russian spring-summer	Flaviviridae, Flavirirus: Russian spring-summer virus
Varicella	Herpesviridae, Alphaherpesvirinae: Human Herpesvirus 3 (Varicella-zoster virus)
Venezuelan equine encephalitis	Togaviridae, Alphavirus: Venezuelan equine encephalitis virus
Venezuelan hemorrhagic fever	Arenaviridae, Tacaribe complex, Arenavirus: Guanarito virus
Vesicular stomatitis	Rhabdoviridae, Vesiculovirus: Vesicular stomatitis virus
Wesselsbron	Flaviridae, Flavivirus: Wesselsbron virus
West Nile fever	Flaviridae, Flavivirus: West Nile virus
Western equine encephalitis	Togaviridae, Alphavirus: Western equine encephalitis virus
Whitewater Arroyo virus infection	Arenaviridae, Arenavirus: Whitewater arroyo virus
Yellow fever	Flaviridae, Flavivirus: Yellow fever virus

Table S4. Genus, Species, and Name of the Associated Disease for Bacteria Reported from at Least One Country Included in the Study

Disease Name	Bacteria
Obligate Intracellular Bacteria	
African tick bite fever	*Rickettsia africae*
Astrakhan fever	*Rickettsia caspii*
Bartonellosis	*Bartonella (Rochalimaea) henselae, quintana, and elizabethiae*
Chlamydia pneumoniae infection	*Chlamydia pneumonae*
Conjunctivitis—inclusion	*Chlamydia trachomatis*
Ehrlichiosis—*E. sennetsu*	*Ehrlichia sennetsu*
Ehrlichiosis—human granulocytic	*Ehrlichia phagocytophila*
Ehrlichiosis—human monocytic	*Ehrlichia chaffeensis*
Flinders Island spotted fever	*Rickettsia honei*
Israeli spotted fever	*Rickettsia israeli*
Japanese spotted fever	*Rickettsia japonica*
Lymphogranuloma venereum	*Chlamydia trachomatis* Type L1-3
Mediterranean spotted fever	*Rickettsia conorii*
North Asian tick typhus	*Rickettsia siberica*
Ornithosis	*Chlamydia psittaci*
Q fever	*Coxiella burnetii*
Queensland tick typhus	*Rickettsia australis*
Rickettsia felis infection	*Rickettsia felis*
Rickettsialpox	*Rickettsia akari*
Rocky Mountain spotted fever	*Rickettsia rickettsii*
South American Bartonellosis	*Bartonella bacilliformis*
Typhus—endemic	*Rickettsia typhi*
Typhus—epidemic	*Rickettsia prowazekii*
Typhus—scrub	*Orientia* [formerly *Rickettsia*] *tsutsugamushi*
Facultative Intracellular Bacteria	
Anthrax	*Bacillus anthracis*
Bacillus cereus	*Bacillus cereus*
Brazilian purpuric fever	*Haemophilus aegyptius*
Brucellosis	*Brucella spp.*
Campylobacteriosis	*Campylobacter jejuni*
Chancroid	*Haemophilus ducreyi*
Chronic meningococcemia	*Neisseria meningitidis*
Clostridial myonecrosis	*Clostridium perfringens*
Endemic syphilis (bejel)	*Treponema pallidum subsp. endemicum*
Erysipeloid	*Erysipelothrix rhusiopathiae*
Glanders	*Burkholderia (Malleomyces) mallei*
Gonorrhea	*Neisseria gonorrhoeae*
Granuloma inguinale	*Calymmatobacterium [Klebsiella] granulomatis*
Legionellosis	*Legionella spp.*
Leprosy	*Mycobacterium leprae*
Leptospirosis	*Leptospira interrogans*
Listeriosis	*Listeria monocytogenes*
Lyme disease	*Borrelia burgdorferi*
Malignant otitis externa	*Pseudomonas aeruginosa*
Melioidosis	*Burkholderia (Pseudomonas) pseudomallei*
Mycobacteriosis—*M. marinum*	*Mycobacterium marinum*
Mycobacteriosis—*M. scrofulaceum*	*Mycobacterium scrofulaceum*
Mycobacteriosis—*M. ulcerans*	*Mycobacterium ulcerans*
Nocardiosis	*Nocardia spp.*
Pertussis	*Bordetella pertussis*
Pharyngitis—bacterial	*Streptococcus pyogenes*
Pinta	*Treponema carateum*
Plague	*Yersinia pestis*
Pyomyositis	*Staphylococcus aureus*
Rhinoscleroma	*Klesbellia spp.*
Rhodococcus equi infection	*Rhodococcus equi*
Shigellosis	*Shigella spp.*
Syphilis	*Treponema pallidum subsp. pallidum*
Tuberculosis	*Mycobacterium tuberculosis*
Tularemia	*Francisella tularensis*
Typhoid and enteric fever	*Salmonella typhi*
Whipple's disease	*Tropheryma whipplei*
Yaws	*Treponema pallidum subsp. pertenue*
Yersiniosis	*Yersinia enterocolitica*

Table S5. Phylum, Genus, Species, and Name of the Associated Disease for Protozoa Reported from at Least One Country Included in the Study

Disease Name	Protozoa
Babesioses	*Sporozoa, Apicomplexa: Babesia microti*, B. CA-1 (U.S.); *or B. divergens,* B. EU1*, and B. bigemina (Europe)*
Leishmaniasis—cutaneous	*Neozoa, Euglenozoa, Kenetoplastea. Flagellate: Leishmania tropica*
Leishmaniasis—mucocutaneous	*Neozoa, Euglenozoa, Kenetoplastea. Flagellate: Leishmania braziliensis*
Leishmaniasis—visceral	*Neozoa, Euglenozoa, Kenetoplastea. Flagellate: Leishmania donovani, L. infantum, L. cruzi*
Malaria	*Sporozoa, Apicomplexa: Plasmodium spp.*
Microsporidiosis	Microspora: Enterocytozoon, Encephalitozoon (Septata), Vittaforma (Nosema), Pleistophora, Trachipleistophora, etc.
Sarcocystosis	*Sporozoa, Apicomplexa: Sarcocystis bovihominis or S. suihominis*
Toxoplasmosis	*Sporozoa, Apicomplexa: Toxoplasma gondii*
Trypanosomiasis—African	*Neozoa, Euglenozoa, Kenetoplastea. Flagellate: Trypanosoma [Trypanozoon] brucei gambiense and T. b. rhodesiense*
Trypanosomiasis—American	*Neozoa, Euglenozoa, Kenetoplastea. Flagellate: Trypanosoma cruzi*