

REVIEW

Human and pathogen genotype-by-genotype interactions in the light of coevolution theory

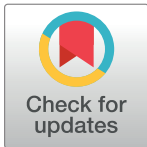
Lars Råberg *

Department of Biology, Lund University, Sweden

* lars.raberg@biol.lu.se

Abstract

Antagonistic coevolution (i.e., reciprocal adaptation and counter-adaptation) between hosts and pathogens has long been considered an important driver of genetic variation. However, direct evidence for this is still scarce, especially in vertebrates. The wealth of data on genetics of susceptibility to infectious disease in humans provides an important resource for understanding host–pathogen coevolution, but studies of humans are rarely framed in coevolutionary theory. Here, I review data from human host–pathogen systems to critically assess the evidence for a key assumption of models of host–pathogen coevolution—the presence of host genotype-by-pathogen genotype interactions (G×G). I also attempt to infer whether observed G×G fit best with “gene-for-gene” or “matching allele” models of coevolution. I find that there are several examples of G×G in humans (involving, e.g., *ABO*, *HBB*, *FUT2*, *SLC11A1*, and HLA genes) that fit assumptions of either gene-for-gene or matching allele models. This means that there is potential for coevolution to drive polymorphism also in humans (and presumably other vertebrates), but further studies are required to investigate how widespread this process is.


 OPEN ACCESS

Citation: Råberg L (2023) Human and pathogen genotype-by-genotype interactions in the light of coevolution theory. PLoS Genet 19(4): e1010685. <https://doi.org/10.1371/journal.pgen.1010685>

Editor: Greg Gibson, Georgia Institute of Technology, UNITED STATES

Published: April 6, 2023

Copyright: © 2023 Råberg. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The author received no specific funding for this work.

Competing interests: The author has declared that no competing interests exist.

Introduction

Population genetic analyses of humans as well as other organisms have shown that immune genes and other genes at the host–pathogen interface are often highly polymorphic. Moreover, many of these polymorphisms are associated with susceptibility to infectious and inflammatory/autoimmune disease and have therefore likely been subject to natural selection [1,2]. Natural selection is normally expected to eliminate genetic variation, so why are immune genes then so variable?

A popular idea is that the high level of polymorphism is a result of host–pathogen coevolution driven by negative frequency-dependent selection (NFDS; [Box 1](#)), often referred to as “Red Queen dynamics” [3,4]. This can occur because infection typically requires pathogen binding to some host molecule to gain access to tissue and/or pathogen evasion of immune recognition to avoid clearance. Regardless of the type of molecular interaction, pathogens should evolve to infect common host genotypes that are then selected against and decline in frequency, followed by pathogen adaptation to alternative host genotypes which are then selected against. Such persistent NFDS as a result of continuous pathogen adaptation to the

currently most frequent host genotype can lead to the maintenance of 2 or more alternative alleles for long time periods at the loci involved.

Box 1. Glossary

Host–pathogen coevolution: a form of antagonistic coevolution, where there is reciprocal selection for adaptation and counter-adaptation in 2 species that affect each other's fitness negatively.

Negative frequency-dependent selection: when the fitness of an allele is negatively correlated with its frequency (direct NFDS) or the frequency of an allele at another locus (indirect NFDS). In case of host–pathogen coevolution, there needs to be indirect NFDS in the sense that the fitness of a host allele depends on the frequency of the pathogen allele with which it interacts [3,10].

The finding that genes at the host–pathogen interface in humans and other organisms often have signatures of balancing selection [5–7] is clearly consistent with the idea of coevolution by NFDS. However, balancing selection on such genes could also be a result of other forms of pathogen-mediated balancing selection, like heterozygote advantage or spatiotemporal heterogeneity in pathogen abundance driven by environmental factors [3]. None of the latter processes involve reciprocal selection for adaptation and counter-adaptation as in coevolution; instead, they represent unidirectional selection by pathogens on the host.

In invertebrates and plants, “time-shift experiments”—where hosts are exposed to pathogens from the past, present, and future—have demonstrated that coevolution by NFDS indeed plays a role in natural populations [8,9]. However, such experiments are difficult to perform on vertebrates, and there is little other evidence that balancing selection in vertebrates is a result of host–pathogen coevolution by NFDS. Moreover, even if NFDS in principle is a very powerful driver of polymorphism, theoretical models have shown that it only occurs in a quite narrow parameter space [10]. Thus, it is relevant to ask: How important is coevolution, with continuous adaptation and counter-adaptation of host and pathogen, as a driver of polymorphism in vertebrates?

A good way to start investigating the role of coevolution by NFDS in vertebrates is to test assumptions that are specific to models of host–pathogen coevolution. The key assumption of classical models of host–pathogen coevolution by NFDS is that infection depends not only on genetic variation in host and pathogen, but also on the combination of host and pathogen genotypes. Thus, in statistical terms, there needs to be a host genotype-by-pathogen genotype interaction (G×G) for susceptibility to infection [3,4,11].

There are 2 basic types of models of host–pathogen coevolution, with different types of G×G; “matching allele” (MA) and “gene-for-gene” (GFG) models [10,12]. Briefly, MA models assume G×G such that different host genotypes are susceptible to different pathogen genotypes, while GFG models assume G×G such that host genotypes differ in the range of pathogen genotypes they are susceptible to. Both scenarios can lead to coevolution by NFDS and the long-term maintenance of polymorphism, but the GFG scenario will only do so if there is a cost of resistance (see Fig 1 for details). Thus, testing for G×G and investigating the nature of G×G provides a key to understanding host–pathogen coevolution.

There are numerous studies demonstrating G×G in plant and invertebrate host–pathogen systems (for examples, see [13,14]), but explicit tests for G×G in vertebrates have been scarce

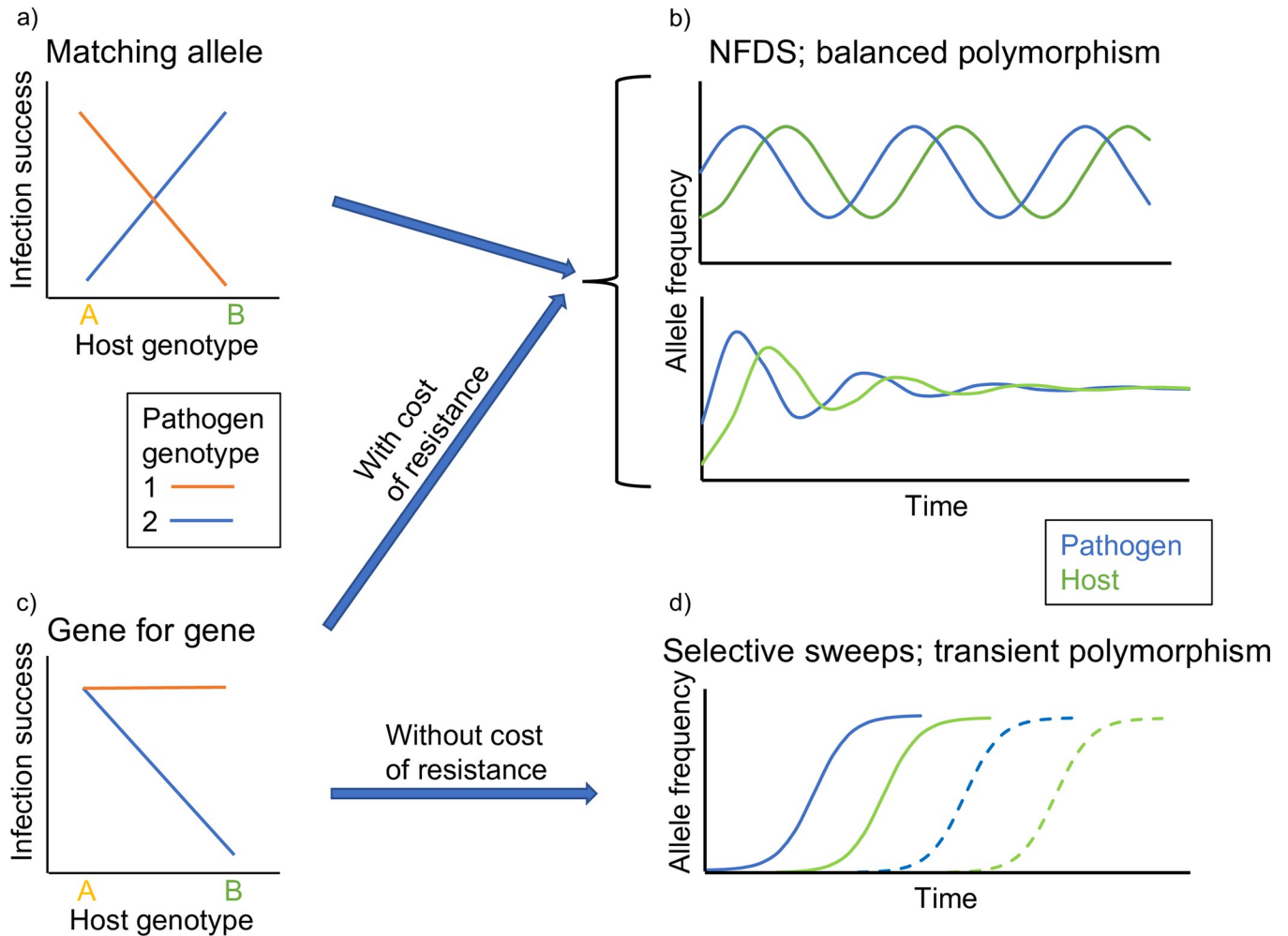


Fig 1. Coevolutionary consequences of different types of G×G. For simplicity, the figure illustrates a scenario where both host and pathogen are haploid and where the G×G involves 1 host locus and 1 pathogen locus (each with 2 different alleles). In MA models, there is a G×G such that different pathogen genotypes infect different host genotypes (a). MA models readily lead to NFDS and the long-term maintenance of polymorphism at interacting loci in both host and pathogen, either in the form of cyclic allele frequencies or a stable polymorphism (b). This occurs because resistance to 1 pathogen genotype comes with a cost in the form of susceptibility to other pathogen genotypes. In other words, under the MA scenario, there is a trade-off between resistance to different pathogen genotypes. In GFG models, there is a G×G such that some pathogen genotypes infect a wider range of genotypes than others (c). In the basic GFG scenario, there is no cost of host resistance or pathogen infectivity. When a host allele that improves resistance without any costs (to the host) occurs in a population, it will be favoured by selection and driven to fixation. Similarly, when a pathogen allele that improves infectivity without costs (to the pathogen) occurs, it will go to fixation. GFG models without costs of resistance or infectivity therefore lead to selective sweeps with only brief, transient polymorphisms, often referred to as arms race coevolution ((d); note that successive sweeps often occur at different sites in the genome, as indicated by different types of lines). However, if there is a cost of host resistance in the currency of another trait related to fitness so that no host genotype has highest fitness under all conditions (and a cost of pathogen infectivity so that no pathogen genotype has highest fitness under all conditions), also GFG models can lead to coevolution by NFDS and the long-term maintenance of polymorphism in the same way as matching allele models (b) [12]. Note that different types of molecular interactions between host and pathogen can result in both MA and GFG type G×G (see [11]). Whether NFDS results in cycles or stable polymorphism (b) depends on the relative importance of 2 different types of NFDS; direct NFDS (where the fitness of an allele is negatively correlated with its frequency) and indirect NFDS (where the fitness of an allele in the host is negatively correlated with the frequency of an allele at the locus involved in G×G in the coevolving pathogen) [10]. Based on figures in [4,11]. GFG, gene-for-gene; G×G, host genotype-by-pathogen genotype interaction; MA, matching allele; NFDS, negative frequency-dependent selection.

<https://doi.org/10.1371/journal.pgen.1010685.g001>

[15]. However, during the last decade, several genome-wide tests for G×G in human host–pathogen systems have been published. Here, I systematically review the evidence for G×G from these studies, as well as candidate gene analyses, and evaluate the implications for our understanding of the importance of coevolution between pathogens and humans (and vertebrates in general) as a cause of balancing selection. I focus on the following questions: For

which human genes is there evidence of G×G? Are these G×G of MA or GFG type? Are there other types of costs (e.g., risk of autoimmune disease) associated with genes involved in G×G (which could help maintain polymorphism in case of GFG type G×G)? Do genes involved in G×G show signatures of balancing selection (as would be expected if they are engaged in coevolution by NFDS)?

G×G in humans

Literature search

Studies of G×G in humans have not used consistent terminology (for example, the term genotype-by-genotype interaction or similar is rarely used in the literature on humans), so it is difficult to perform a focused literature search with narrow search terms. Instead, I identified relevant papers by a combination of broad reading of the literature (particularly review papers of genetics of susceptibility to pathogens in humans) and a literature search with broad search terms ([Box 2](#)).

Box 2. Literature search

I first identified relevant papers by broad reading of the literature, in particular review papers of genetics of susceptibility to pathogens in humans; this yielded a first set of 13 papers showing G×G, involving 8 different pathogens. To find more papers, I performed a literature search in Web of Science Core Collection in Dec 2022. To this end, I extracted key words from the titles and abstracts of the first set of papers and constructed a query with relatively broad search terms, but which still yielded a manageable number of records [Topic = human AND (genetic varia* OR polymorph*) AND (bacteria* OR viral OR virus OR parasite OR pathogen) AND (interact* OR interplay OR “genome to genome”), which yielded approximately 3,900 records]. By scanning titles and abstracts of these records, I identified papers that considered genetic variation of both host and pathogen; these papers (approximately 1% of the records) were examined in detail (both original results and cited references). In the end, this literature search produced 15 additional papers with evidence for G×G. Most of these concerned pathogens and/or human genes already included in the first set of papers; the list of pathogens and host genes involved in G×G should thus be reasonably complete.

I selected studies showing G×G for any infection-related trait; thus, not only analyses of susceptibility to infection (which is the trait that is traditionally the focus of models of coevolution), but also studies using disease severity, pathogen load, immune escape mutations, etc., as outcome. I only considered natural genetic variation, so studies of genetically modified pathogens or human cell lines were excluded.

Study types and prevalence of G×G

I found evidence for G×G in 10 human host–pathogen systems, including protozoan, bacterial, and viral pathogens ([Table 1](#)). Evidence for G×G comes from several different types of studies, from epidemiological analyses to in vitro assays. Moreover, G×G were detected in several different ways. Several studies tested for G×G between 1 or several host candidate genes and pathogen strains. Another approach, employed in some of the most recent studies, is genome-wide testing for G×G in both host and pathogen, that is by performing genome-wide SNP typing of

Table 1. Evidence for G×G in human host–pathogen systems.

Pathogen	Human gene	Pathogen gene/protein	Type of study	G x G test (host–pathogen)	Trait	Effect of host polymorphism for each pathogen genotype	GFG/MA ¹	Signature of balancing selection?	Cost of resistance?
<i>Plasmodium falciparum</i>	<i>HLA-B</i>	<i>csp</i>	Naturally infected (cases)	Candidate gene x candidate gene	Clinical malaria	<i>HLA-B</i> *35 associated with higher frequency of 2 <i>P. falciparum</i> genotypes and lower frequency of 2 genotypes [53].	?	Yes [5,54].	Yes; the haplotype involved in G x G (<i>HLA-B</i> x 35) is associated with a number of diseases, with OR both >1 and <1 [40].
<i>Plasmodium falciparum</i>	<i>HBB</i>	<i>msp-1</i>	Naturally infected (cases)	Candidate gene x candidate gene	<i>P. falciparum</i> infection	Frequencies of <i>msp-1</i> genotypes differ between <i>HBB</i> AS and AA individuals [55].	?	Yes; recent positive/balancing selection [47].	Yes; HbS leads to anemia and associated diseases [56].
<i>Plasmodium falciparum</i>	<i>HBB</i>	3 genes on 2 chromosomes; all in strong LD	Case-control	Candidate genes x genome wide	Severe malaria	HbS protective against severe malaria if infected with parasite having major alleles at all 3 genes (OR≈0.02) but not when infected with parasite having minor alleles (OR≈1) [23].	GFG	Yes; recent positive/balancing selection [47].	Yes; HbS leads to anemia and associated diseases [56].
<i>Streptococcus pneumoniae</i>	<i>SASH1</i> , <i>FARP1</i>	<i>pspC</i>	Case-control	Genome x candidate genes	Pneumococcal meningitis	<i>SASH1</i> : OR for 1 <i>pspC</i> allele >1; OR for 2 other <i>pspC</i> alleles not ≠ 1. <i>FARP1</i> : OR for 1 <i>pspC</i> allele >1; OR for 2 other <i>pspC</i> alleles not ≠ 1 [20].	GFG	No	No known.
<i>Streptococcus pneumoniae</i>	<i>STK32C</i>		Case-control	Genome x pathogen strain ("serotype")	Pneumococcal meningitis	OR for 1 serotype > 1; OR for 6 other serotypes not ≠ 1 [20].	GFG	No	No known.
<i>Mycobacterium tuberculosis</i>	<i>HLA-DRB1</i> , <i>HLA-DQB1</i>		Case-control	Candidate gene x pathogen strain	Tuberculosis	<i>DRB1</i> *09:01 OR and <i>DQB1</i> *03:03 OR for "modern strains" >1; OR for other strains not ≠ 1 [26]	GFG	Yes [5,54]	Yes [40].
<i>Mycobacterium tuberculosis</i>	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i>		Naturally infected (cases)	Candidate gene x pathogen strain	Pulmonary tuberculosis	Specific HLA class I alleles associated with specific <i>M. mycobacterium</i> strains [57].	?	Yes [5,54]	Yes [40].
<i>Mycobacterium tuberculosis</i>	<i>TLR2</i>		Case-control	Candidate gene x pathogen strain ("lineage")	Tuberculous meningitis	OR for Beijing lineage >1; OR for other lineages not ≠ 1 [24].	GFG	No	No known.

(Continued)

Table 1. (Continued)

Pathogen	Human gene	Pathogen gene/protein	Type of study	G x G test (host-pathogen)	Trait	Effect of host polymorphism for each pathogen genotype	GFG/MA ¹	Signature of balancing selection?	Cost of resistance?
<i>Mycobacterium tuberculosis</i>	<i>CD209</i>		Naturally infected (cases)	Candidate gene x pathogen strain	Mortality from pulmonary tuberculosis	Infection with Beijing vs. other strains in patients who died from TB associated with SNP in <i>CD209</i> [58].	?	No	No known.
<i>Mycobacterium tuberculosis</i>	<i>SLC11A1</i>		Naturally infected (cases)	Candidate gene x pathogen strain	Pulmonary tuberculosis	Infection with Beijing vs. other strains associated with 2 SNPs in <i>SLC11A1</i> [59].	?	(Yes) [60].	Yes; see below.
<i>Mycobacterium tuberculosis</i>	<i>SLC11A1</i>		Naturally infected (cases)	Candidate genes x pathogen strain ("lineage")	Tuberculosis severity	Opposite effects of <i>SLC11A1</i> alleles in individuals with lineage "L4.6 Uganda" vs. other lineages [35].	MA	(Yes) [60].	Yes; high expressing allele associated with autoimmune diseases while low expressing allele associated with infectious diseases [44].
<i>Mycobacterium tuberculosis</i>	<i>PPIAP22</i>		Naturally infected (cases)	Genome x pathogen strain ("lineage")	Tuberculosis severity	Opposite effects of <i>PPIAP22</i> alleles on disease severity in individuals with lineage "L4.6 Uganda" vs. other lineages [21].	MA	No	No known.
<i>Helicobacter pylori</i>	<i>ABO</i>	<i>babA</i>	In vitro functional analysis	Candidate gene x pathogen strain ("isolate")	Binding of host receptor	Most strains bind both A and H antigen (generalists), but significant fraction of strains in S America are specialists and bind only H (i.e., blood group O) [29].	Mainly GFG	Yes [46].	Yes; blood group A, B, and AB increase susceptibility to severe malaria [27,61].
<i>Vibrio cholerae</i>	<i>ABO</i>	<i>ctxAB</i>	Case-control	Candidate gene x pathogen strain ("serogroup"/"biotype")	Disease severity	Blood group O confers higher risk of severe disease than blood group A or B, but only when infected with O1 El Tor or O139 strains; in contrast, no association between disease severity and ABO group for classical O1 strains [27].	GFG	Yes [46].	Yes; blood group A, B, and AB increase susceptibility to severe malaria [27,61].
HIV	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i>	48 amino acid residues throughout HIV proteome	Naturally infected (cases)	Genome x genome	Immune escape mutations	48 HIV-1 amino acid variants associated with SNPs in <i>HLA-A</i> , <i>B</i> , or <i>C</i> [17]. See also [30].	?	Yes [5,54].	Yes [40].

(Continued)

Table 1. (Continued)

Pathogen	Human gene	Pathogen gene/protein	Type of study	G x G test (host-pathogen)	Trait	Effect of host polymorphism for each pathogen genotype	GFG/MA ¹	Signature of balancing selection?	Cost of resistance?
HIV	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i>	Protease/RT, Nef, Vpr	Naturally infected (cases)	Candidate genes x candidate genes	Immune escape mutations	Different <i>HLA-A</i> , <i>HLA-B</i> , and <i>HLA-C</i> alleles often select for different amino acid escape mutations at a given HIV residue (at 23/57 residues where polymorphism is associated with specific HLA alleles) [31].	MA for at least 23/57 HIV residues involved in GxG.	Yes [5,54].	Yes [40].
HIV	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i>	31 amino acid residues throughout HIV proteome	Naturally infected (cases)	Candidate genes x genome	Immune escape mutations and viral load	Specific <i>HLA-A</i> , <i>HLA-B</i> , and <i>HLA-C</i> alleles select for specific amino acid escape mutations at 31 HIV residues (but no case where different <i>HLA</i> alleles select for different amino acids at a given residue, as in [31]). In addition, effect of <i>HLA-A/B</i> allele x immune escape mutation on viral load at 3 of the 31 residues, such that viral load is reduced in individuals carrying specific <i>HLA</i> alleles if HIV has not acquired escape mutation [32].	Immune escape?; viral load: GFG.	Yes [5,54].	
HIV	<i>KIR2DL2</i>	Vpu, Env	In vitro infection assay	Candidate gene x candidate gene	Viral replication	NK cells with ≥ 1 copy of <i>KIR2DL2</i> inhibit replication of HIV Vpu-Env(WT/WT) but not Vpu-Env (V/V), but NK cells w/o <i>KIR2DL2</i> inhibit neither HIV genotype [34].	GFG	Yes [62].	Yes; haplotype w <i>KIR2DL2</i> associated w several AID, including T1D and UC [43].
Hepatitis C virus	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i> , <i>DQA1</i> , <i>DRB1</i>	NS3, NS4B	Naturally infected (cases)	Genome x genome	Immune escape mutations	Specific HLA alleles select for specific amino acid escape mutations at certain residues, but no case where different <i>HLA</i> alleles select for different amino acids at a given residue [18]. See also [33].	?	Yes [5,54].	Yes [40].

(Continued)

Table 1. (Continued)

Pathogen	Human gene	Pathogen gene/protein	Type of study	G x G test (host-pathogen)	Trait	Effect of host polymorphism for each pathogen genotype	GFG/MA ¹	Signature of balancing selection?	Cost of resistance?
Hepatitis C virus	<i>IFNL4</i>	NS5B	Naturally infected (cases)	Genome x genome	Immune escape mutations and viral load	Immune escape: <i>IFNL4</i> rs12979860 associated with amino acid variants at 11 HCV residues. Viral load: <i>IFNL4</i> rs12979860 C>T associated with reduced viral load, but only if HCV has serine at site 2414 [18]. See also [63].	Immune escape?; Viral load: GFG	No (but population-specific positive selection; [64]).	No known.
Hepatitis C virus	<i>KIR</i> genes		Naturally infected (cases)	Candidate gene x pathogen strain (HCV genotype)	Risk of hepatocellular carcinoma (HCC)	For HCV genotype 1, 2, and 3, risk of HCC decreases with number of activating <i>KIR</i> genes; for HCV genotype 4, risk is low regardless of number of activating <i>KIR</i> genes [36].	GFG	Yes [62].	Yes, several activating <i>KIR</i> genes associated with autoimmune diseases [43].
Norovirus	<i>FUT2</i>		Challenge studies/case-control/prospective cohort studies	Candidate gene x pathogen strain	Acute gastroenteritis	Challenge studies: Nonfunctional <i>FUT2</i> protects against strain GI.1 and GII.4, but not against GII.2. Case-control studies: Nonfunctional <i>FUT2</i> protects against GII.4, but not against GI.3. Prospective cohort studies: Nonfunctional <i>FUT2</i> protects against GII.3 and GII.4, but not against GI.3, GI.6, GII.1, GII.2, and GII.7 [25].	GFG	Yes [65,66].	Yes; nonfunctional <i>FUT2</i> associated w susceptibility to Crohn's disease [41] and other diseases [42].
Epstein-Barr virus (EBV)	<i>UNC5D</i> , <i>LINC01830</i> , non-coding region on chr 7	BALF5, BBRF1, BRLF1	Naturally infected (cases)	Genome x genome	Immune escape mutations	25 host SNPs in 3 genomic regions (17, 1, and 7 SNPs, respectively) associated with variants in 3 EBV genes [19].	?	Yes, <i>UNC5D</i> [67].	Possibly for <i>UNC5D</i> ; SNP in this gene associated with adolescent idiopathic scoliosis.
Epstein-Barr virus (EBV)	<i>HLA-B</i>	EBNA-1	In vitro functional analysis	Candidate gene x pathogen strain	In vitro immune response	EBV peptide variant HPVG most immunogenic on HLA-B* 35:01, but peptide variant HPVG-D5 most immunogenic on HLA-B* 35:08 [37].	MA	Yes [5,54].	Yes [40].

(Continued)

Table 1. (Continued)

Pathogen	Human gene	Pathogen gene/protein	Type of study	G x G test (host-pathogen)	Trait	Effect of host polymorphism for each pathogen genotype	GFG/MA ¹	Signature of balancing selection?	Cost of resistance?
Human papillomavirus (HPV)	<i>HLA-DR/DQ</i>		Case-control	Candidate gene x pathogen strain	Cervical cancer	Different <i>HLA-DR-DQ</i> haplotypes affect resistance/susceptibility to cancer caused by different HPV types [28].	MA	Yes [5,54].	Yes [40].
Human papillomavirus (HPV)	<i>HLA-DR/DQ</i>	E6	Naturally infected (cases; i.e., cervical cancer patients)	Candidate gene x candidate gene	Infection (or immune escape mutation?)	<i>HLA-DR04-DQ03</i> haplotype associated with HPV E6 L83V variant in cervical cancer patients [68,69].	?	Yes [5,54].	Yes [40].

¹ GFG = gene-for-gene type G×G, MA = matching allele type G×G, ? = type of G×G could not be inferred from published data. For case-control studies of binary traits (presence/absence of infection or disease), GFG was inferred when the OR for ≥1 pathogen genotype was different from 1 while the OR for ≥1 other pathogen genotype was equal to 1, whereas MA was inferred when the OR for different pathogen genotypes were in opposite directions. For analyses of continuous traits (pathogen load, disease severity), GFG was inferred when there was an effect of host genotype on the trait for ≥1 pathogen genotype but no effect of host genotype for other pathogen genotypes, whereas MA was inferred when host genotype had opposite effects on the trait for different pathogen genotypes. For escape mutations, MA was inferred when alternative alleles at a HLA gene were associated with different escape mutations at a given pathogen residue (see main text for more detailed explanations).

<https://doi.org/10.1371/journal.pgen.1010685.t001>

both host and pathogen and then testing for G×G between all pairs of host and pathogen SNPs, referred to as “genome-to-genome” analysis [16]. Other studies used various combinations of candidate gene analysis, pathogen strain identification, and genome-wide analyses.

To gain insight into how common G×G are it is useful to focus on studies based on genome-wide analyses of humans (genome-to-genome studies and genome-wide tests for interactions with pathogen strains), as they should provide a more unbiased estimate of the occurrence of G×G than candidate gene studies. Genome-wide analyses have been performed with viral and bacterial pathogens. All 3 genome-wide tests for G×G with viruses found evidence for G×G [17–19]. Most of these concern immune escape mutations, the only exception being [18], which also found G×G for viral load. Of the genome-wide analyses for G×G with bacterial pathogens, 2 studies found statistically significant G×G [20,21] while 1 did not [22]. In all studies where G×G were found, only 1 or a few host loci were involved. Thus, the currently available data indicate that G×G occur in most host-pathogen pairs, but that at most a few host genes are involved in each pair. It should be noted, though, that the multiple testing burden is considerable in genome-wide tests for G×G [16], so future studies with higher power may reveal that a larger number of host loci are often involved in G×G in each host-pathogen pair.

For which genes is there evidence of G×G?

Human genes with evidence for G×G include some genes that are textbook examples of associations with susceptibility to infectious disease, such as the MHC class I genes (*HLA-A*, *HLA-B*, *HLA-C*; G×G with, e.g., *Plasmodium falciparum* and HIV) and the blood group antigen gene *ABO* (G×G with *Helicobacter pylori* and *Vibrio cholerae*). A recent study also found G×G between *HBB* (encoding the hemoglobin β subunit)—which has well-known effects on susceptibility to malaria—and *P. falciparum* [23]. Specifically, the protective effect of the HbS variant

at *HBB* was found to depend on the genotype at 3 different loci in the *P. falciparum* genome, with all 3 loci in strong linkage disequilibrium such that the minor alleles occur together.

In addition, genes with evidence for G×G include some canonical immune genes (e.g., *TLR2*, *IFNL4*, *KIR2DL2*, *CD209*) and other genes at the host–pathogen interface with well-documented associations with susceptibility to infection (*FUT2*, *SLC11A1*), but also several genes that are not previously recognised in this context (e.g., *FARPI*, *STK32C*, *UNC5D*).

Are these G×G of MA or GFG type?

None of the studies put their results in the context of MA versus GFG. I therefore inferred the type of G×G based on the published data. G×G were detected for several different infection phenotypes, including both binary (e.g., infection status) and continuous traits (e.g., pathogen load or disease severity in infected individuals). For case-control studies of infection status or other binary disease phenotypes, where it is possible to calculate host genotype odds ratios (OR) separately for each pathogen genotype, this information can be used to distinguish MA and GFG. To see this, consider the simplest case where there is G×G between a pair of loci that are bi-allelic in both host and pathogen, as in Fig 1. If there is a trade-off between resistance to different host genotypes as in the MA scenario (Fig 1A), host genotype should be associated with both pathogen genotypes, but in opposite ways. Thus, the OR should be >1 for one pathogen genotype but <1 for the other (note that none of the ORs need to be significantly different from 1, but they should be different from each other). In contrast, if host genotypes differ in the range of pathogen genotypes they are resistant/susceptible to, as in the GFG scenario (Fig 1C), host genotype should only be associated with one of the pathogen genotypes. Thus, the OR should be significantly different from 1 for one pathogen genotype but equal to 1 for the other.

Of the case-control studies with evidence for G×G for infection status or other binary disease phenotypes, 8 present pathogen genotype-specific ORs, with 1 to 3 host genes involved in G×G with each pathogen [20,23–28]. In all but 1 case, the OR is significantly different from 1 for one pathogen genotype but not for others. Thus, in the majority of cases the pattern is most consistent with GFG type G×G. Perhaps the most striking example is *HBB* and resistance to malaria [23]. Here, HbS is protective against severe malaria if an individual is infected with a parasite having the major allele at all 3 loci involved in G×G (OR≈0.02) but not when infected with parasites having the minor allele at all 3 loci (OR≈1) (based on data in Fig 2 of [23]). Similar differences between host genotypes in the range of pathogen strains they are susceptible to (but without any indication of a trade-off between resistance to different strains) are seen with for example *ABO* (with *H. pylori* and *V. cholera*) [27,29] and *FUT2* (with Norovirus) [25]. The only indication of MA type G×G in case-control studies concern HLA class II and risk of cervical cancer caused by human papilloma virus (HPV), where different HLA haplotypes affect susceptibility to different HPV types [28].

There are also several studies that have performed analyses of associations between pathogen and host alleles in chronic viral infections [17–19,30–33]. Such G×G are generally interpreted as being a result of within-host evolution of immune escape, although they could also reflect differences in susceptibility to infection with viruses carrying different alleles at the start of the infection. These studies have primarily found G×G involving HLA genes. It is generally difficult to infer whether these G×G are of MA or GFG type, because specific HLA alleles are often associated with escape mutations at several positions in the viral genome. However, one of the studies of HIV found that different HLA alleles were associated with different amino acid escape mutations at a particular position [31], a pattern clearly indicating a trade-off between resistance to different pathogen genotypes; thus in at least some cases G×G for escape mutations are consistent with the MA scenario.

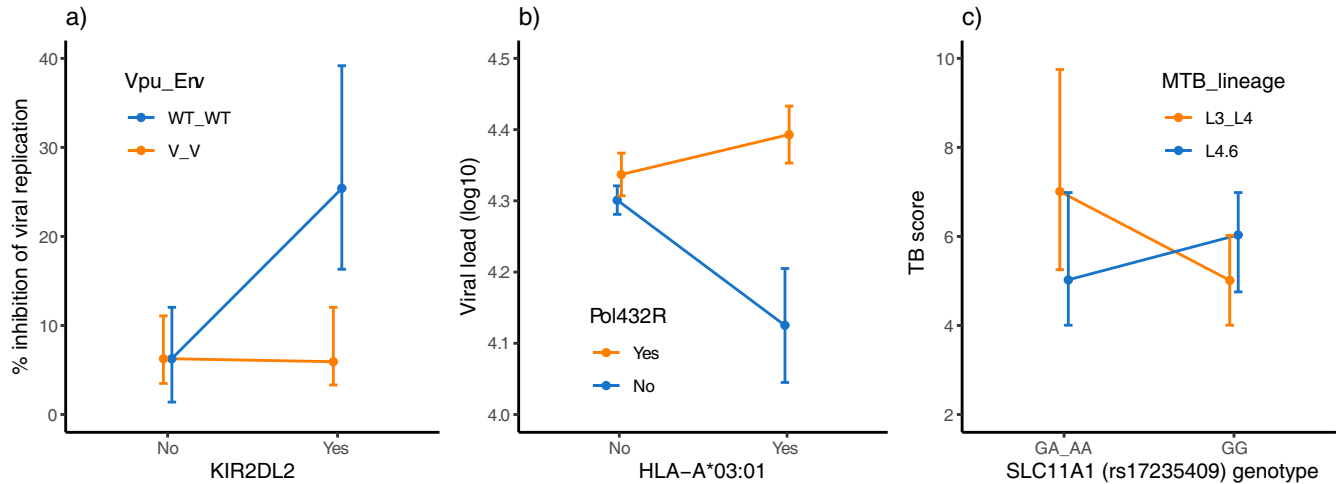


Fig 2. Examples of G×G for continuous traits related to resistance/susceptibility to infectious disease. (a) G×G between NK cell Killer Immunoglobulin Receptor genotype and HIV genotype for inhibition of viral replication (median \pm interquartile range). NK cells with the *KIR2DL2* allele strongly inhibit replication of HIV with wild-type alleles at *vpu* and *env* (WT_WT) but have limited inhibitory effect on HIV with variant alleles (V_V), while NK cells without *KIR2DL2* have limited inhibitory effect on both WT_WT and V_V. Thus, presence/absence of *KIR2DL2* affects the range of HIV genotypes an individual is susceptible to, consistent with the GFG scenario. Data extracted from Fig 1B (day 3) of [34]. (b) G×G between *HLA-A* genotype and HIV genotype for viral load (mean \pm SE). Arginine (R) at residue 432 in the *pol* gene is an immune escape mutation from the *HLA-A* allele 03:01. In individuals carrying A*03:01, viral load is suppressed in infections with virus without the Pol432R escape mutation, while there is no effect of Pol432 genotype on viral load in individuals without A*03:01 (and no G×G for viral load between Pol432 genotype and other HLA alleles). Thus, there is no indication of a trade-off between resistance to Pol432R and other genotypes as would be the case in an MA type G×G; instead the pattern is consistent with a GFG type G×G. Data from [32]. (c) G×G between *SLC11A1* genotype and *M. tuberculosis* lineage for tuberculosis severity (median \pm IQR). A recently evolved *M. tuberculosis* sublineage (L4.6) in combination with homozygosity for an ancestral *SLC11A1* allele (genotype GG) and an original *M. tuberculosis* lineage (L3 or L4) in combination with ≥ 1 derived *SLC11A1* allele (genotype GA or AA) is associated with more severe disease than the other combinations of host and pathogen genotypes (the interaction is highly significant: $P = 0.00022$). Thus, there is a trade-off between resistance to disease by different lineages, consistent with an MA type G×G. Data extracted from Fig 2 in [35]. GFG, gene-for-gene; G×G, host genotype-by-pathogen genotype interaction; MA, matching allele.

<https://doi.org/10.1371/journal.pgen.1010685.g002>

Besides studies based on epidemiological analyses of presence/absence of infectious disease or immune escape mutations, there are also some studies finding G×G for various continuous infection-related traits like pathogen replication and disease severity [18,21,32,34–37]. Given that the trait is associated with both host and pathogen fitness, also G×G affecting such traits could lead to coevolution. A study using an in vitro assay of HIV replication found that NK cells with at least 1 copy of the *KIR2DL2* allele inhibit replication of 1 specific HIV genotype, while NK cells without *KIR2DL2* have limited inhibitory effect regardless of HIV genotype [34], a pattern consistent with a GFG type G×G (Fig 2A). Similarly, a study analysing effects of HIV escape mutations on viral load showed that certain virus genotypes had reduced viral load in individuals carrying a specific MHC allele while there was no effect of virus genotype on viral load in individuals without that allele; also this pattern appear consistent with the GFG scenario (Fig 2B) [32]. In contrast, an analysis of tuberculosis patients showed that *SLC11A1* genotype had opposite effects on disease severity depending on *Mycobacterium tuberculosis* lineage [35], consistent with an MA type G×G (Fig 2C). Overall, 4 of the 7 analyses of continuous traits showed results consistent with GFG type G×G, while 3 are consistent with MA type G×G (Table 1).

Finally, in vitro functional analyses of the ability of different *Helicobacter pylori* isolates to bind host receptors showed that most isolates are generalists and bind both A and H antigen (from individuals with blood group A and O, respectively) while a significant fraction of strains in South America are specialists and bind only H antigen [29], consistent with GFG type G×G.

Are there other types of costs associated with genes involved in G×G?

Since several of the G×G appeared to be of the GFG type, it would be of interest to know if the genes involved are associated with other types of diseases that could lead to the fitness cost necessary to generate NFDS and help maintain polymorphism in case of a GFG type G×G (Fig 1). To identify potential costs of alleles conferring resistance to a particular pathogen genotype, I searched the GWAS catalog [38] (either directly or via LDtrait at LDlink [39] to check if SNPs involved in G×G were in linkage disequilibrium with SNPs associated with other diseases) and PheWAS Resources (*HLA* genes) [40] for disease-associations with genes involved in G×G.

For several of the genes with GFG type G×G, there is indeed strong evidence for costs associated with the allele that confers resistance to a subset of pathogen genotypes. For example, the nonfunctional *FUT2* allele, which protects against some Norovirus strains, is also associated with Crohn's disease and other diseases [41,42]. Similarly, *KIR2DL2*, which inhibits replication of a specific HIV genotype [34], is associated with several autoimmune diseases [43]. Overall, costs are known for about two thirds of the genes with indication of GFG type G×G (Table 1). Interestingly, there is also evidence for costs of resistance in case of *SLC11A1*, which is one of few genes showing clear MA type G×G (where costs are not necessary to generate NFDS; Fig 1). Here, high and low expression alleles are associated with susceptibility to autoimmune and infectious disease, respectively [44].

Do genes involved in G×G show signatures of balancing selection?

If the G×G identified in humans really lead to coevolution by NFDS, one would expect the genes involved to exhibit signatures of balancing selection that can be detected by analyses of population samples of DNA sequence data [45]. For 12 of the 20 genes involved in G×G, there are such signatures of balancing selection, based on genome-wide scans or candidate gene analyses (Table 1). Most show signatures of long-term balancing selection, in some cases—for example, *ABO*—in the form of “trans-species polymorphisms,” meaning that the polymorphism has been maintained by selection in primates for tens of millions of years [46]. An exception to the trend for long-term balancing selection is *HBB* that shows a signature of recent positive or balancing selection (for recent selection the signatures of positive and balancing selection are indistinguishable) [47].

Conclusions

The present review has shown that several human genes are involved in G×G, as assumed by models of host–pathogen coevolution. Most of the G×G seem to fit the GFG rather than MA scenario, particularly for case-control studies of infection status and other binary disease phenotypes, which means a cost of resistance is required for these G×G to lead to maintenance of polymorphism by NFDS. Such costs are known for at least some of the genes with evidence for G×G. Taken together, this shows there is scope for coevolution by NFDS also in vertebrates. These conclusions come with several caveats, though.

First, for G×G to result in coevolution, the phenotypic trait concerned must be associated with both host and pathogen fitness. While most studied traits (Table 1) clearly can affect host fitness, the relevance for pathogen fitness is doubtful in some cases, for example, meningitis in *Streptococcus pneumoniae* infection [20] and risk of cervical cancer caused by HPV [28]. Second, in case of chronic viral infections (HIV, HCV, and EBV), the G×G are thought to be a result of within-host evolution of immune escape, and it is not always clear if these G×G also affect some aspect of host fitness, such as susceptibility to infection or severity of disease, as would be required for coevolution to occur. However, a recent study of HIV found that at least some of the immune escape mutations led to G×G for viral load [32], indicating that G×G

involving immune escape mutations might indeed affect host fitness. Third, inferring if G×G are of GFG or MA type from currently segregating host and pathogen alleles might be misleading. For example, what is actually an MA type G×G might appear to be a GFG type G×G if rare alleles are not sampled [11,48]. Fourth, the preponderance of GFG type G×G in case-control studies of binary disease phenotypes might be an artefact of that these analyses are based on separate analyses of each pathogen strain and only report host polymorphisms where the OR is different from 1 for at least one of the pathogen strains. Thus, these analyses will miss MA type G×G where the OR for 2 pathogen strains are in opposite directions and different from each other, but none is different from 1. Even with these caveats in mind, there are some strong cases for coevolutionarily relevant G×G of both GFG and MA type (GFG: e.g., *HBB*, *ABO*, *FUT2*, and HLA genes; MA: e.g., *SLC11A1* and HLA genes).

The G×G for *HBB* illustrates that different types of pathogen-mediated balancing selection can act on a given gene simultaneously. *HBB* is the textbook example of heterozygote advantage, where individuals with 1 copy of the HbS variant have improved resistance to malaria, whereas HbS homozygosity leads sickle cell disease [49]. The finding that *HBB* is involved in a G×G with *P. falciparum* shows that there might also be NFDS on this gene. It is often expected that several different types of pathogen-mediated balancing selection operate simultaneously on a given gene and *HBB* is perhaps the clearest evidence yet that this is the case.

The G×G for *ABO* and the HLA genes illustrate another aspect of pathogen-mediated balancing selection—that a given gene might be coevolving with more than 1 pathogen simultaneously, so called “diffuse coevolution” [3]. Diffuse coevolution is expected to be common, perhaps the norm, but *ABO* and the HLA genes are as far as I am aware the first cases where specific genes have been shown to be involved in G×G with 2 or more different pathogens, thus demonstrating that there actually is opportunity for diffuse coevolution.

In conclusion, there is some evidence from humans for G×G, a key assumption of models of host–pathogen coevolution by NFDS (but not other types of pathogen-mediated balancing selection). This indicates that balancing selection on genes at the host–pathogen interface in humans (and other vertebrates) could indeed be a result of coevolution, as is commonly assumed. Nevertheless, more studies testing for G×G are clearly desirable, in particular, genome-to-genome studies as they give an unbiased perspective on which genes are involved. Recent development of statistical approaches should facilitate this [50,51]. Still, it is important to recognise that the presence of G×G only shows that there is opportunity for coevolution by NFDS, not that it has occurred. Demonstrating that polymorphism is a result of coevolution would require additional analyses. One way would be to test if there is NFDS. Specifically, balancing selection by antagonistic coevolution requires indirect NFDS, i.e., the fitness of a host allele should be negatively correlated with the frequency of a pathogen allele. This could be tested by following 1 or more populations over time. Advances in the analysis of ancient DNA from both mammals and pathogens should make this possible even for humans and other species with long generations times [52]. Nevertheless, identifying the genes involved in G×G—as described in this review—would be a critical first step.

Acknowledgments

I thank Charlie Cornwallis and Maria Strandh for comments on the manuscript and Rasmus Lykke Marvig for providing data for Fig 2B.

References

1. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet.* 2014; 15:379–393. <https://doi.org/10.1038/nrg3734> PMID: 24776769

2. Quintana-Murci L, Clark AG. Population genetic tools for dissecting innate immunity in humans. *Nat Rev Immunol*. 2013; 13:280–293. <https://doi.org/10.1038/nri3421> PMID: 23470320
3. Ebert D, Fields PD. Host–parasite co-evolution and its genomic signature. *Nat Rev Genet*. 2020; 21:754–768. <https://doi.org/10.1038/s41576-020-0269-1> PMID: 32860017
4. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet*. 2002; 32:569–577. <https://doi.org/10.1038/ng1202-569> PMID: 12457190
5. Bitarello D, De Filippo C, Kleinert P, Meyer D, Andrés AM. Signatures of long-term balancing selection in human genomes. *Genome Biol Evol*. 2018; 10:939–955. <https://doi.org/10.1093/gbe/evy054> PMID: 29608730
6. Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*. 2013; 339:1578–1582. <https://doi.org/10.1126/science.1234070> PMID: 23413192
7. Koenig D, Jörg H, Li R, Bemm F, Slotte T, Neuffer B, et al. Long-term balancing selection drives evolution of immunity genes in Capsella. *Elife*. 2019; 8:e43606. <https://doi.org/10.7554/eLife.43606> PMID: 30806624
8. Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D, et al. Host-parasite “Red Queen” dynamics archived in pond sediment. *Nature*. 2007; 450:870–873. <https://doi.org/10.1038/nature06291> PMID: 18004303
9. Thrall PH, Laine A-L, Ravensdale M, Nemri A, Dodds PN, Barrett LG, et al. Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecol Lett*. 2012; 15:425–435. <https://doi.org/10.1111/j.1461-0248.2012.01749.x> PMID: 22372578
10. Tellier A, Brown JK, Boots M, John S. Theory of Host–Parasite Coevolution: From Ecology to Genomics. *eLS*. 2021; 2:1–10. <https://doi.org/10.1002/9780470015902.a0029063>
11. Dybdahl MF, Jenkins CE, Nuismer SL. Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. *Am Nat*. 2014; 184:1–13. <https://doi.org/10.1086/676591> PMID: 24921596
12. Agrawal A, Lively CM. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evol Ecol Res*. 2002; 4:79–90.
13. Carius HJ, Little TJ, Ebert D. Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution*. 2001; 55:1136–1145. <https://doi.org/10.1111/j.0014-3820.2001.tb00633.x> PMID: 11475049
14. Salvaudon L, Héraudet V, Shykoff JA. Genotype-specific interactions and the trade-off between host and parasite fitness. *BMC Evol Biol*. 2007; 7:189. <https://doi.org/10.1186/1471-2148-7-189> PMID: 17919316
15. Råberg L, Clough D, Hagström Å, Scherman K, Andersson M, Drews A, et al. MHC class II genotype-by-pathogen genotype interaction for infection prevalence in a natural rodent-Borrelia system. *Evolution*. 2022; 76:2067–2075. <https://doi.org/10.1111/evo.14590> PMID: 35909235
16. Fellay J, Pedergrana V. Exploring the interactions between the human and viral genomes. *Hum Genet*. 2020; 139:777–781. <https://doi.org/10.1007/s00439-019-02089-3> PMID: 31729546
17. Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife*. 2013; 2:e01123. <https://doi.org/10.7554/eLife.01123> PMID: 24171102
18. Ansari MA, Pedergrana V, Ip CLC, Magri A, Von Delft A, Bonsall D, et al. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet*. 2017; 49:666–673. <https://doi.org/10.1038/ng.3835> PMID: 28394351
19. Rüeiger S, Hammer C, Loetscher A, McLaren PJ, Lawless D, Naret O, et al. The influence of human genetic variation on Epstein–Barr virus sequence diversity. *Sci Rep*. 2021; 11:4586. <https://doi.org/10.1038/s41598-021-84070-7> PMID: 33633271
20. Lees JA, Ferwerda B, Kremer PHC, Wheeler NE, Serón MV, Croucher NJ, et al. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nat Commun*. 2019; 10:2176. <https://doi.org/10.1038/s41467-019-09976-3> PMID: 31092817
21. McHenry ML, Wampande EM, Joloba ML, Malone LL, Mayanja-Kizza H, Bush WS, et al. Interaction between *M. tuberculosis* lineage and human genetic variants reveals novel pathway associations with severity of TB. *Pathogens*. 2021; 10:1487. <https://doi.org/10.3390/pathogens10111487> PMID: 34832643
22. Nelson CL, Pelak K, Podgoreanu MV, Ahn SH, Scott WK, Allen AS, et al. A genome-wide association study of variants associated with acquisition of *Staphylococcus aureus* bacteremia in a healthcare setting. *BMC Infect Dis*. 2014; 14:83. <https://doi.org/10.1186/1471-2334-14-83> PMID: 24524581

23. Band G, Leffler EM, Jallow M, Sisay-Joof F, Ndila CM, Macharia AW, et al. Malaria protection due to sickle haemoglobin depends on parasite genotype. *Nature*. 2021; 602:106–111. <https://doi.org/10.1038/s41586-021-04288-3> PMID: 34883497
24. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NTN, Thuong NTT, et al. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog*. 2008; 4:e1000034. <https://doi.org/10.1371/journal.ppat.1000034> PMID: 18369480
25. Nordgren J, Svensson L. Genetic Susceptibility to Human Norovirus Infection: An Update. *Viruses*. 2019; 11:226. <https://doi.org/10.3390/v11030226> PMID: 30845670
26. Toyo-oka L, Mahasirimongkol S, Yanai H, Mushiroda T, Wattanapokayakit S, Wichukchinda N, et al. Strain-based HLA association analysis identified HLA-DRB1*09:01 associated with modern strain tuberculosis. *HLA Immune Response Genet*. 2017; 90:149–156. <https://doi.org/10.1111/tan.13070> PMID: 28612994
27. Cooling L. Blood Groups in Infection and Host Susceptibility. *Clin Microbiol Rev*. 2015; 28:801–870. <https://doi.org/10.1128/CMR.00109-14> PMID: 26085552
28. Apple RJ, Erlich HA, Klitz W, Manos MM, Becker TM, Wheeler CM. HLA DR-DQ associations with papillomavirus-type specificity. *Nat Genet*. 1994; 6:157–162.
29. Aspholm-Hurtig M, Dailide G, Lahmann M, Kalia A, Ilver D, Roche N, et al. Functional adaptation of *BabA* the *H. pylori* ABO blood group antigen binding adhesin. *Science*. 2004; 305:519–522. <https://doi.org/10.1126/science.1098801> PMID: 15273394
30. Naret O, Chaturvedi N, Bartha I, Hammer C, Fellay J. Correcting for Population Stratification Reduces False Positive and False Negative Results in Joint Analyses of Host and Pathogen Genomes. *Front Genet*. 2018; 9. <https://doi.org/10.3389/fgene.2018.00266> PMID: 30105048
31. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, et al. Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. *PLoS Pathog*. 2007; 3:e94. <https://doi.org/10.1371/journal.ppat.0030094> PMID: 17616974
32. Gabrielaite M, Bannedbæk M, Zucco AG, Ekenberg C, Murray DD, Kan VL, et al. Human Immunotypes Impose Selection on Viral Genotypes Through Viral Epitope Specificity. *J Infect Dis*. 2021; 224:2053–2063. <https://doi.org/10.1093/infdis/jiab253> PMID: 33974707
33. Timm J, Li B, Daniels MG, Bhattacharya T, Reyrol LL, Allgaier R, et al. Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. *Hepatology*. 2007; 46:339–349. <https://doi.org/10.1002/hep.21702> PMID: 17559151
34. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, Carlson JM, et al. HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature*. 2011; 476:96–100. <https://doi.org/10.1038/nature10237> PMID: 21814282
35. McHenry ML, Bartlett J, Igo RP, Wampande EM, Benchek P, Mayanja-Kizza H, et al. Interaction between host genes and *Mycobacterium tuberculosis* lineage can affect tuberculosis severity: Evidence for coevolution? *PLoS Genet*. 2020; 16:e1008728. <https://doi.org/10.1371/journal.pgen.1008728> PMID: 32352966
36. Littera R, Zamboni F, Tondolo V, Fantola G, Chessa L, Orrù N, et al. Absence of activating killer immunoglobulin-like receptor genes combined with hepatitis C viral genotype is predictive of hepatocellular carcinoma. *Hum Immunol*. 2013; 74:1288–1294. <https://doi.org/10.1016/j.humimm.2013.05.007> PMID: 23756163
37. Liu YC, Chen Z, Neller MA, Miles JJ, Purcell AW, McCluskey J, et al. A Molecular Basis for the Interplay between T Cells, Viral Mutants, and Human Leukocyte Antigen Micropolymorphism. *J Biol Chem*. 2014; 289:16688–16698. <https://doi.org/10.1074/jbc.M114.563502> PMID: 24759101
38. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019; 47:D1005–D1012. <https://doi.org/10.1093/nar/gky1120> PMID: 30445434
39. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015; 31:3555–3557. <https://doi.org/10.1093/bioinformatics/btv402> PMID: 26139635
40. Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med*. 2017; 9:eaai8708. <https://doi.org/10.1126/scitranslmed.aai8708> PMID: 28490672
41. McGovern DPB, Jones MR, Taylor KD, Marcianti K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum Mol Genet*. 2010; 19:3468–3476. <https://doi.org/10.1093/hmg/ddq248> PMID: 20570966

42. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci U S A*. 2011; 108:19030–19035. <https://doi.org/10.1073/pnas.1106408108> PMID: 22068912
43. Moesta AK, Parham P. Diverse functionality among human NK cell receptors for the C1 epitope of HLA-C: KIR2DS2, KIR2DL2, and KIR2DL3. *Front Immunol*. 2012; 3:1–13. <https://doi.org/10.3389/fimmu.2012.00336> PMID: 23189078
44. Blackwell JM, Goswami T, Evans CAW, Sibthorpe D, Papo N, White JK, et al. SLC11A1 (formerly NRAMP1) and disease resistance. *Cell Microbiol*. 2001; 3:773–784. <https://doi.org/10.1046/j.1462-5822.2001.00150.x> PMID: 11736990
45. Fijarczyk A, Babik W. Detecting balancing selection in genomes: Limits and prospects. *Mol Ecol*. 2015; 24:3529–3545. <https://doi.org/10.1111/mec.13226> PMID: 25943689
46. Ségurel L, Thompson EE, Flutre T, Lovstad J, Venkat A, Margulis SW. The ABO blood group is a trans-species polymorphism in primates. *Proc Natl Acad Sci U S A*. 2012; 109:4–11. <https://doi.org/10.1073/pnas.1210603109> PMID: 23091028
47. Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, et al. Screening for recently selected alleles by analysis of human haplotype similarity. *Am J Hum Genet*. 2006; 78:153–159. <https://doi.org/10.1086/499252> PMID: 16385459
48. Frank SA. Specificity versus detectable polymorphism in host-parasite genetics. *Proc R Soc London B*. 1993; 254:191–197. <https://doi.org/10.1098/rspb.1993.0145> PMID: 8108452
49. Aidoo M, Terlouw DJ, Kolczak MS, McElroy PD, Ter Kuile FO, Kariuki S, et al. Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet*. 2002; 359:1311–1312. [https://doi.org/10.1016/S0140-6736\(02\)08273-9](https://doi.org/10.1016/S0140-6736(02)08273-9) PMID: 11965279
50. Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H, et al. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc Natl Acad Sci U S A*. 2018; 115:E5440–E5449. <https://doi.org/10.1073/pnas.1710980115> PMID: 29848634
51. MacPherson A, Otto SP, Nuismer SL. Keeping pace with the Red Queen: Identifying the genetic basis of susceptibility to infectious disease. *Genetics*. 2018; 208:779–789. <https://doi.org/10.1534/genetics.117.300481> PMID: 29223971
52. Kerner G, Patin E, Quintana-Murci L. New insights into human immunity from ancient genomics. *Curr Opin Immunol*. 2021; 72:116–125. <https://doi.org/10.1016/j.coi.2021.04.006> PMID: 33992907
53. Gilbert SC, Plebanski M, Gupta S, Morris J, Cox M, Aidoo M, et al. Association of malaria parasite population structure, HLA, and immunological antagonism. *Science*. 1998; 279:1173–1177. <https://doi.org/10.1126/science.279.5354.1173> PMID: 9469800
54. DeGiorgio M, Lohmueller KE, Nielsen R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet*. 2014; 10:e1004561. <https://doi.org/10.1371/journal.pgen.1004561> PMID: 25144706
55. Ntoumi F, Rogier C, Dieye A, Trape J-F, Millet P, Mercereau-Puijalon O. Imbalanced distribution of *Plasmodium falciparum* MSP-1 genotypes related to sickle-cell trait. *Mol Med*. 1997; 3:581–592. PMID: 9323709
56. Ashley-Koch A, Yang Q, Olney RS. Sickle Hemoglobin (HbS) allele and sickle cell disease: a HuGe review. *Am J Epidemiol*. 2000; 151:839–845. <https://doi.org/10.1093/oxfordjournals.aje.a10288> PMID: 10791557
57. Salie M, Van Der Merwe L, Möller M, Daya M, Van Der Spuy GD, Van Helden PD, et al. Associations between human leukocyte antigen class I variants and the mycobacterium tuberculosis subtypes causing disease. *J Infect Dis*. 2014; 209:216–223. <https://doi.org/10.1093/infdis/jit443> PMID: 23945374
58. Ogarkov O, Mokrousov I, Sinkov V, Zhdanova S, Antipina S, Savilov E. “Lethal” combination of Mycobacterium tuberculosis Beijing genotype and human CD209 -336G allele in Russian male population. *Infect Genet Evol*. 2012; 12:732–736. <https://doi.org/10.1016/j.meegid.2011.10.005> PMID: 22027159
59. van Crevel R, Parwati I, Sahiratmadja E, Marzuki S, Ottenhoff THM, Netea MG, et al. Infection with Mycobacterium tuberculosis Beijing Genotype Strains Is Associated with Polymorphisms in SLC11A1/NRAMP1 in Indonesian Patients with Tuberculosis. *J Infect Dis*. 2009; 200:1671–1674. <https://doi.org/10.1086/648477> PMID: 19863441
60. Barnes I, Duda A, Pybus OG, Thomas MG. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution*. 2011; 65:842–848. <https://doi.org/10.1111/j.1558-5646.2010.01132.x> PMID: 20840594
61. Rowe JA, Handel IG, Thera MA, Deans AM, Lyke KE, Koné A, et al. Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc Natl Acad Sci U S A*. 2007; 104:17471–17476. <https://doi.org/10.1073/pnas.0705390104> PMID: 17959777

62. Gendzekhadze K, Norman PJ, Abi-Rached L, Graef T, Moesta AK, Layrisse Z, et al. Co-evolution of KIR2DL3 with HLA-C in a human population retaining minimal essential diversity of KIR and HLA class I ligands. *Proc Natl Acad Sci U S A*. 2009; 106:18692–18697. <https://doi.org/10.1073/pnas.0906051106> PMID: 19837691
63. Ansari MA, Aranday-Cortes E, Ip CL, Da Silva FA, Lau SH, Bamford C, et al. Interferon lambda 4 impacts the genetic diversity of hepatitis c virus. *Elife*. 2019; 8:1–23. <https://doi.org/10.7554/eLife.42463> PMID: 31478835
64. Key FM, Peter B, Dennis MY, Huerta-Sánchez E, Tang W, Prokunina-Olsson L, et al. Selection on a Variant Associated with Improved Viral Clearance Drives Local, Adaptive Pseudogenization of Interferon Lambda 4 (IFNL4). *PLoS Genet*. 2014; 10:e1004681. <https://doi.org/10.1371/journal.pgen.1004681> PMID: 25329461
65. Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, et al. Targets of balancing selection in the human genome. *Mol Biol Evol*. 2009; 26:2755–2764. <https://doi.org/10.1093/molbev/msp190> PMID: 19713326
66. Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, et al. A natural history of FUT2 polymorphism in humans. *Mol Biol Evol*. 2009; 26:1993–2003. <https://doi.org/10.1093/molbev/msp108> PMID: 19487333
67. Siewert KM, Voight BF. Detecting long-term balancing selection using allele frequency correlation. *Mol Biol Evol*. 2017; 34:2996–3005. <https://doi.org/10.1093/molbev/msx209> PMID: 28981714
68. Zehbe I, Tachezy R, Mytilineos J, Voglino G, Mikyškova I, Delius H, et al. Human papillomavirus 16 E6 polymorphisms in cervical lesions from different European populations and their correlation with human leukocyte antigen class II haplotypes. *Int J Cancer*. 2001; 94:711–716. <https://doi.org/10.1002/ijc.1520> PMID: 11745467
69. Beskow AH, Engelmark MT, Magnusson JJ, Gyllensten UB. Interaction of host and viral risk factors for development of cervical carcinoma in situ. *Int J Cancer*. 2005; 117:690–692. <https://doi.org/10.1002/ijc.21212> PMID: 15929080