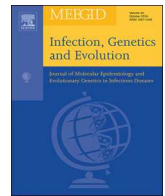




ELSEVIER

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Detecting HLA-infectious disease associations for multi-strain pathogens

Connor F White^{a,c}, Lorenzo Pellis^{a,b,c}, Matt J Keeling^{a,c,d}, Bridget S Penman^{a,d,*}^a Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, CV4 7AL, United Kingdom^b Department of Mathematics, The University of Manchester, Alan Turing Building, Oxford Road, Manchester, M13 9PL, United Kingdom^c Mathematics Institute, University of Warwick, CV4 7AL, United Kingdom^d School of Life Sciences, University of Warwick, CV4 7AL, United Kingdom

ARTICLE INFO

Keywords:

HLA
Epidemiology
Infection
Evolution
Pathogen diversity
Case control study

ABSTRACT

Human Leukocyte Antigen (HLA) molecules play a vital role helping our immune system to detect the presence of pathogens. Previous work to try and ascertain which HLA alleles offer advantages against particular pathogens has generated inconsistent results. We have constructed an epidemiological model to understand why this may occur. The model captures the epidemiology of a multi strain pathogen for which the host's ability to generate immunological memory responses to particular strains depends on that host's HLA genotype. We find that an HLA allele's ability to protect against infection, as measured in a case control study, depends on the population frequency of that HLA allele. Furthermore, our capability to detect associations between HLA alleles and infection with a multi strain pathogen may be affected by the properties of the pathogen itself (i.e. R_0 and length of infectious period). Both host and pathogen genetics must be considered in order to identify true HLA associations. However, in the absence of detailed pathogen genetic information, a negative correlation between the frequency of an HLA type and its apparent protectiveness against disease caused by multi strain pathogen is a strong indication that the HLA type in question is well adapted to a subset of strains of that pathogen.

1. Introduction

An individual's ability to fend off invading pathogens is affected by their genotype. Co-evolution between humans and pathogens has generated extreme diversity in certain human genes, in particular the Human Leukocyte antigens (HLAs) (Spurgin and Richardson, 2010; Jeffery and Bangham, 2000; Hedrick, 2002; Pierini and Lenz, 2018) making HLA loci the most polymorphic in the human genome (Robinson et al., 2014). Understanding which HLA genotypes are best adapted to which infectious diseases is an ongoing challenge. Here we investigate how epidemiological and population genetic factors combine to influence whether individual HLAs appear to protect against infection with multi-strain pathogens.

HLA molecules play an integral role in the human immune system. They are cell surface proteins containing a binding cleft. The binding cleft is loaded with peptides sampled from either inside (class I HLA molecules) or outside (class II HLA molecules) the cell (Horton et al., 2004). T cell receptors bind to the HLA/peptide complex and if the bound peptide is recognised as "non-self" by a T cell, this will trigger an immune response. HLA molecules encoded by different alleles have binding clefts with different properties. The specific peptide fragments which are displayed by HLA molecules determine an individual's T cell

responses, thus HLA genotype acts as a bottleneck, which can shape adaptive immunity.

HLA genotype has been shown to affect the outcome of a wide range of infectious diseases (Tian et al., 2017; Dunstan et al., 2014; McLaren et al., 2015; Sveinbjornsson et al., 2016; Hill et al., 1991; Thursz et al., 1995; Oliveira-Cortez et al., 2016; Carrington and O'Brien, 2003). The first HLA/infectious disease association was demonstrated for malaria: HLA-Bw53 was associated with a reduced chance of developing severe malaria disease in a population in The Gambia (Hill et al., 1991). Also in The Gambia, HLA-DRB1*1302 was associated with a reduction in the probability of developing persistent hepatitis B (HBV) infection (Thursz et al., 1995). Various HLA alleles have been shown to affect the time to Acquired Immune Deficiency Syndrome (AIDS) for Human Immunodeficiency Virus (HIV) infected individuals (Just, 1995); and individual amino acids in the binding clefts both HLA-A and HLA-B molecules can impact HIV setpoint viral load (McLaren et al., 2015). For some common infections including mumps, childhood ear infections and strep throat, a recent Genome Wide Association Study (GWAS) suggests that HLA genotype may affect the probability of experiencing symptomatic infection at all (Tian et al., 2017).

Studies which identify HLA-infectious disease associations typically compare a group of individuals with an infectious disease phenotype

* Corresponding author.

E-mail address: b.penman@warwick.ac.uk (B.S. Penman).<https://doi.org/10.1016/j.meegid.2020.104344>

Received 4 February 2020; Received in revised form 24 April 2020; Accepted 27 April 2020

Available online 05 May 2020

1567-1348/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(cases) to a group of individuals without it (controls) and examine differences in the frequencies of HLA types in the two groups. An over representation of a specific HLA type in the control group could be because it has a protective effect. However, such case control studies do not always give consistent results. A case control study of severe malaria, similar to the previously mentioned study in The Gambia (Hill et al., 1991), was performed in Mali (Lyke et al., 2011). The Mali study did not find HLA-Bw53 to be protective against severe malaria but they did find that HLA-A*30:01 and HLA-A*33:01 increased susceptibility to developing severe malaria. Lyke et al noted that this discrepancy could be due to different strains of malaria parasite circulating in Mali as opposed to The Gambia. If specific HLA alleles are associated with better immune responses to just a subset of pathogen strains, then the effectiveness of HLA alleles will depend on which pathogen strains are circulating in a population.

The interaction between host genotype and pathogen strain in determining disease outcome is highlighted by recent studies of Tuberculosis (TB). A case control study in the Chiang Rai province of Thailand split TB patients into groups infected with modern strains of TB and those infected with ancient strains of TB (Toyo-Oka et al., 2017) (the ancient/modern distinction is based on the presence/absence of a TbD1 deletion in the TB genome). They found HLA DRB1*09:01 to be associated with protection from infection with modern strain tuberculosis. They did not find this association when they grouped patients with modern and ancient tuberculosis strains together. A similar study was performed in Cape Town, South Africa (Salie et al., 2013), distinguishing TB strains by restriction fragment length polymorphism genotyping. Salie et al found that HLA class I types A*01, B*08 and C2¹ were all associated with increased susceptibility to Beijing strain TB; B*27 and C1 were associated with lower susceptibility to the Beijing strain.

The possible impact of HLA type on the epidemiology of infection has been considered in terms of the maintenance of parasite diversity (Gupta and Hill, 1995), and more recently in terms of the effect of HLA type on pathogen R_0 in different populations (Sambaturu et al., 2018). Specifically, Sambaturu et al addressed the impact of HLA type on the spread of H1N1 influenza (Sambaturu et al., 2018), by first classifying HLA types by the range of influenza epitopes they are predicted to present, (Mukherjee and Chandra, 2014), and then making the assumption that the more viral epitopes a host can represent, the lower that host's susceptibility to H1N1 influenza. They show that when a population has a wide range of individual susceptibility to a strain of H1N1 influenza, it can reduce the size of an epidemic from said strain of H1N1 influenza. This previous work highlights the potential significance of HLA diversity for public health. However, no previous model has considered the epidemiological processes underlying why, if an HLA type truly does affect infection, it might not always be detected as having such an effect in a case control study.

MacPherson et al recently analysed the effect of pathogen diversity and host-pathogen coevolution on the ability of genome wide association studies (GWAS) to detect which host genes matter for infection (MacPherson et al., 2018). They elegantly demonstrated that if pathogen diversity is ignored, many important host loci will go undetected by GWAS. In the case of HLA genes, and the further complication of adaptive immunity, the problems highlighted by MacPherson may be compounded.

Here we explore the case of a multi-strain pathogen for which the presence of a specific HLA molecule in the host is necessary to develop an effective memory immune response against a specific pathogen strain. We identify the different epidemiological outcomes which arise as a consequence of this HLA-strain relationship, when population HLA

frequencies vary. We simulate case control studies of infection, and demonstrate the circumstances under which an HLA type offering an advantage against a specific pathogen strain is likely to appear protective or risky against the prevailing local infection.

Neisseria meningitidis, *Streptococcus pneumoniae* and *Plasmodium falciparum* are examples of multi-strain pathogens where humans experience multiple infections; become immune to different strains, and where T cell responses (and hence HLA type) are implicated in the generation of protective immunity (Wiertz et al., 1996; Davenport et al., 2003; Aslam et al., 2010; Aslam et al., 2011; Mordmüller et al., 2017). Our model offers insight into how HLA/strain interactions may impact the epidemiology of such systems. Our model further suggests a technique to detect the existence of HLA/strain specific associations in a system, even if the key properties distinguishing different strains are not yet known.

2. Methods

2.1. The epidemiological model

We consider a pathogen which exists as at least two strains (1 and 2). We assume that the diploid HLA genotype of a host (ij) determines whether or not that host will be able to develop a memory immune response against a particular pathogen strain after infection. To model these possible immune outcomes, we use SIR and SIS models which are commonly used ordinary differential equation (ODE) models in epidemiology (for a full review of such approaches, see (Keeling and Rohani, 2011)). If a host's genotype includes an HLA allele which allows the recognition of strain i , that host will become immune to strain i following infection (SIR dynamics). If a host does not have an HLA allele enabling them to recognise a pathogen strain, that host will not become immune to that pathogen strain on recovery (SIS dynamics). This is a simplifying assumption (although very stark MHC/pathogen strain relationships are observed in nature specifically in chickens, as we detail in the discussion). Our aim here is to ask the question: if certain HLA types make the difference between mounting successful memory immune responses or not (as simulated in the model), will this ever be detectable by our current standard methodology (the case control study)?

The force of infection for pathogen strain i is λ_i . The rate of recovery from pathogen strain i back to being susceptible is σ_i . The rate an infected individual becomes immune to pathogen strain i is μ_i . We simulate a population with a constant size, so births and deaths happen at the same rate d . Pathogen induced mortality is not included in this model. A flow chart of the model is given in Fig. 1.

We use a double letter notation to denote susceptibility (S), infectiousness (I), immunity (R), to pathogen strain 1 (first letter) and to pathogen strain 2 (second letter). For instance host IS is infected with pathogen strain 1 and susceptible to pathogen strain 2; a host with SR is susceptible to pathogen strain 1 and immune to pathogen strain 2. We denote the proportion of the population susceptible to pathogen strain 1 and immune to pathogen strain 2 with genotype ij as N_{ij}^{SR} . The total proportion of a genotype in the population is

$$N_{ij} = N_{ij}^{SS} + N_{ij}^{SI} + N_{ij}^{SR} + N_{ij}^{IS} + N_{ij}^{IR} + N_{ij}^{RS} + N_{ij}^{RI} + N_{ij}^{RI} + N_{ij}^{RR}, \quad (1)$$

where the total size of the population is

$$N = \sum_{ij} N_{ij} = 1. \quad (2)$$

We denote the proportion of hosts who are infected with pathogen strain i as I_i and we denote the proportion of hosts immune to pathogen strain i as R_i .

¹ HLA-C in this study was classified into C1 or C2, a distinction based on the interaction between HLA-C and Killer-cell Immunoglobulin-like Receptor (KIR) molecules.

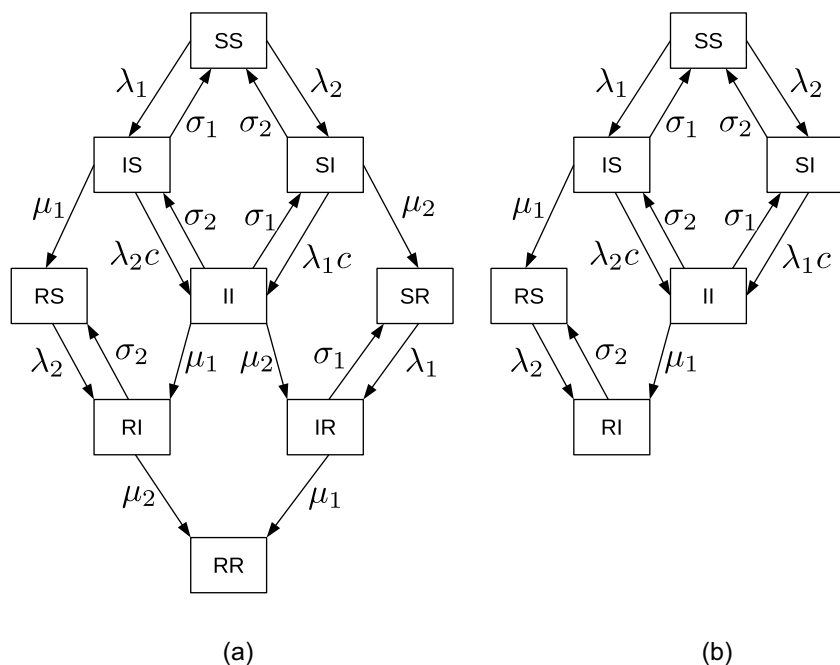


Fig. 1. The possible states of the population and pathways between them. (a) is a flow chart of the possible paths a host of any genotype ij can take from initially being susceptible to all pathogen strains. (b) is a flow chart specifically for a host whose genotype means they can only mount a memory immune response against pathogen strain 1. This flow chart is a case when $\alpha = 0$.

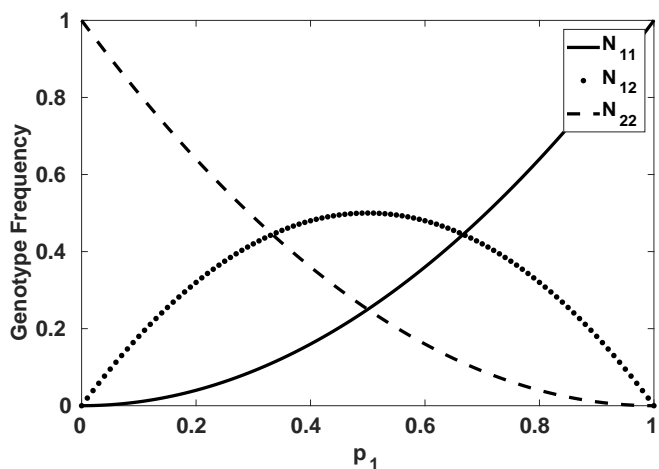


Fig. 2. The frequency of genotypes according to the Hardy-Weinberg principle.

$$\begin{aligned}
 I_1 &= \sum_{(i,j)} (N_{ij}^{IS} + N_{ij}^{IR}) \\
 I_2 &= \sum_{(i,j)} (N_{ij}^{SI} + N_{ij}^{RI}) \\
 R_1 &= \sum_{(i,j)} (N_{ij}^{RS} + N_{ij}^{RI} + N_{ij}^{RR}).
 \end{aligned}
 \tag{3}$$

The transmission rate for pathogen strain i is notated as β_i , therefore the force of infection from pathogen strain i is

$$\lambda_i = \beta_i I_i.
 \tag{4}$$

ODEs for this model can be found in Section 1 in the Appendix.

In order to scale the degree to which a host can be infected with both pathogen strains simultaneously we introduce the parameter c . When $c = 0$ coinfection is impossible. When $c = 1$ infection with one pathogen strain has no effect on the rate a host becomes infected with the other pathogen strain. In order to scale the strength of strain transcending (and host-genotype-independent) immunity, we introduce the parameter α . A proportion α of all recoveries enter the RR state and a proportion $(1 - \alpha)$ recover according to the host's genotype (Fig. 1).

In Sections 3.1, 3.2 and 3.3 of the results, $c = 0$ and $\alpha = 0$. In Section 3.4 we demonstrate the impact of varying these parameters.

Our host population is divided into different possible diploid HLA genotypes (ij) at a single HLA locus. In our main model, we include two pathogen strains (1 and 2) and two HLA alleles (1 and 2). We assume a 1:1 correspondence between HLA alleles and pathogen strains, meaning that the presence of HLA allele 1 is necessary to mount a memory immune response against pathogen strain 1.

We also extend our main model in two ways. We allow for three possible pathogen strains (1, 2 and 3), with 3 corresponding HLA alleles necessary for the recognition of each (see Section 2 of Appendix). We also include a “perfect” or a “useless” HLA allele, recognising both or neither of pathogen strains 1 and 2, alongside strain specific HLA 1 and 2 alleles (see Section 3 of Appendix).

A key parameter of all our models is HLA allele frequency. We notate this as p_i for allele i . The birth rates of different host genotypes are in accordance with the Hardy-Weinberg principle (Relethford, 2012), implying random mating within the population. In a two allele model, the frequency of homozygotes (11 and 22) and heterozygotes (12), are as follows:

$$\begin{aligned}
 N_{11} &= p_1^2 \\
 N_{12} &= 2p_1(1 - p_1) \\
 N_{22} &= (1 - p_1)^2.
 \end{aligned}
 \tag{5}$$

Fig. 2 illustrates how the frequencies of each genotype vary as p_1 varies between 0 and 1.

Starting frequencies of each genotype were assigned in these proportions and the frequencies of each genotype remained unchanged over time. Hardy-Weinberg proportions were similarly used for three allele models, in which 3 homozygotes (11, 22 and 33) and 3 heterozygotes (12, 23, 13) are possible. In three allele models we set $p_2 = p_3 = (1 - p_1)/2$. This means that as p_1 is varied we allocate an equal frequency to alleles 2 and 3. This maximises HLA diversity.

All results described here are obtained by numerically solving the ODEs described above, using solver ode45 in Matlab and calculating quantities from these numerically solved solutions. Matlab code to run the two pathogen and three pathogen ODE models is available in the following repository: github.com/ConnorFrancisWhite/HLA_Infection_Association_Model_White_et_al_2020.

2.2. The odds ratio

The key feature of our model is that possession of a particular HLA type is necessary in order for a host to develop a specific memory immune response against a specific pathogen strain. That HLA type will always be “protective” against infection with the strain it matches. However, when assessing the protectiveness/riskiness of an allele or genotype in a case control study, it is rare that such functional strain definitions are already known. We wish to investigate how the interplay between a diverse host and pathogen makes particular alleles risky or protective against infection in general (i.e. infection with any strain), since this is the property which will most likely be captured by case control studies in practise.

The odds ratio for being infected with any strain given a host has a specific allele i is calculated as follows:

$$OR_i = \left(\frac{P(I|i)}{1 - P(I|i)} \right) \left(\frac{P(I|\hat{i})}{1 - P(I|\hat{i})} \right).$$

Here, $P(I|i)$ is the proportion of hosts with allele i that are infected and $P(I|\hat{i})$ is the proportion of hosts that do not have allele i that are infected. If $OR_i < 1$ allele i is protective against the prevailing local infection. This method of calculating the odds ratio uses proportions of the population at the steady state of the ODE model.

In Section 3.3 we investigate whether the effects we are modelling could be detected in a real world study. We assume this real world study involves 500 cases and 500 controls, and assign different numbers of genotypes to both groups by multiplying relevant steady state proportions from our model by 500. A standard method to calculate the odds ratio for studies involving finite sample sizes of cases and controls, which we denote as OR'_i , is as follows:

$$OR'_i = \left(\frac{n(I|i) + g}{n(\hat{i}|i) + g} \right) \left(\frac{n(I|\hat{i}) + g}{n(\hat{i}|\hat{i}) + g} \right).$$

where $n(I|i)$ is the number of people who are infected and have allele i . $n(\hat{i}|i)$ is the number of people who are not infected and have allele i . $n(I|\hat{i})$ is the number of people who are infected and do not have allele i . $n(\hat{i}|\hat{i})$ is the number of people who are not infected and do not have allele i . When $g > 0$ there is no undefined calculation for the odds ratio even if the number of individuals in one of the categories is 0. A commonly used value for g is $g = 0.5$ (Woolf, 1955; Gart, 1966). (Woolf, 1955) also provides a method for calculating the confidence interval for the odds ratio which we use in Section 3.3. In Figs. 8 and 9 we round our sample sizes to integer values to further simulate a real world study.

3. Results

3.1. The protectiveness of a strain-specific HLA allele against infection is related to the population frequency of that allele

We first consider one pathogen strain in a population that contains two HLA alleles. Only one of these alleles (allele 1) allows a host to generate a memory immune response against the pathogen strain. In Fig. 3a we illustrate how varying HLA frequencies affects OR_1 , the odds ratio for a genotype containing allele 1 being infected with any strain.

If only pathogen strain 1 is present OR_1 is always below 1 for the entire range of p_1 values ($0 < p_1 < 1$) (Fig. 3a, dashed line). The distribution of HLA alleles within the population does not alter the fact that allele 1 is beneficial. This result is intuitive, since the only circulating pathogen strain is one to which hosts with allele 1 can develop memory immunity. No matter how many hosts have allele 1, possessing allele 1 will always make hosts less likely to be infected.

As noted in the introduction, however, many pathogens exist as multiple strains. Let us suppose that among the pathogen variants, some express an immunogenic peptide which can be bound by one HLA type,

and others express a different form of the immunogenic peptide which can be bound by a different HLA type at the same HLA locus. We simulate a two strain, two HLA type system, where HLA allele 1 is necessary to mount a memory immune response against strain 1, and HLA allele 2 is necessary to mount a memory immune response against strain 2. Heterozygotes for alleles 1 and 2 (genotype 12) can generate a memory immune response against both strains. Now as p_1 is increased OR_1 goes from below 1 to above 1 (Fig. 3a, solid line). The frequency of HLA allele 1 in the population is negatively correlated with its ability to protect a host from the prevailing local infection. The same applies to HLA allele 2 (Fig. 3b), noting that $p_2 = 1 - p_1$. To understand why this is happening we need to look at the level of infection from each strain, within the population as p_1 varies.

If allele 1 is more frequent than allele 2, there are more hosts who can become immune to pathogen strain 1 than hosts who can become immune to pathogen strain 2. Pathogen strain 2 will be more successful in such an environment. As p_1 is increased (Fig. 4), the proportion of hosts infected with pathogen strain 1 at equilibrium decreases and the proportion of hosts infected with pathogen strain 2 increases. The distribution of HLA alleles in the population affects the pathogen strain structure in the population. This explains why OR_i is positively correlated with the frequency of allele i .

In a 3 strain, 3 allele extension of the system, OR_1 follows the same trend found with the two pathogen strain model (Fig. 3a, dotted line). As a further extension of the model, in the Supplementary Appendix, we add a different type of third allele to the system where HLA alleles 1 and 2 confer the ability to recognise pathogen strains 1 and 2. This third allele is either a “perfect” allele (conferring the ability to mount an immune response against either strain 1 or strain 2) or a “useless” allele (conferring no ability to mount any immune response).

The presence of the useless allele alongside alleles 1 and 2 and strains 1 and 2 does not substantially alter the pattern shown in Fig. 3a (see Fig. S2) The presence of the perfect allele likewise does not alter the positive correlation between p_1 and OR_1 , although it does reduce the set of circumstances where HLA allele 1 is protective against the prevailing local infection.

We finally tested a system including two pathogen strains, in which HLA allele 1 recognises strain 1, and only the perfect or useless allele is present alongside HLA allele 1. For both of these cases, OR_1 did not cross the value 1 for the entire range of p_1 (Fig. S3). It would seem a system requires at least two pathogen strains and the presence of at least two HLA alleles that differ in their strain specificity in order for OR_1 to go from below 1 to above 1 as p_1 is increased.

3.2. The more rapid the immune response associated with a particular HLA allele against a particular pathogen strain, the less likely that HLA allele is to appear protective

In the results just presented, genotypes containing allele 1 become immune to strain 1 at exactly the same rate that genotypes containing allele 2 become immune to strain 2. However, it is possible that HLA molecules encoded by different HLA alleles differ in their fundamental ability to activate T cells and allow hosts to clear infection. One way to investigate this possibility within our framework is to allow each allele to be associated with a different recovery rate. We retain the strain specificity of the alleles in our main two strain, two allele model, but hosts with allele 1 recover more quickly than hosts with allele 2 after infection with their matched strain.

As shown in Fig. 5, increasing μ_1 relative to μ_2 (enhancing the recovery rate associated with allele 1), changes the behaviour of OR_i (compare dashed lines to solid lines). Lower allele frequencies are still associated with greater protection against infection in general, but OR_1 crosses the value 1 for a much smaller value of p_1 . The faster recovery rate of allele 1 has caused allele 1 to be protective over a smaller region of p_1 , and allele 2 to be protective over a greater range of values of p_1 . This counter intuitive result is explained when we look at the level of

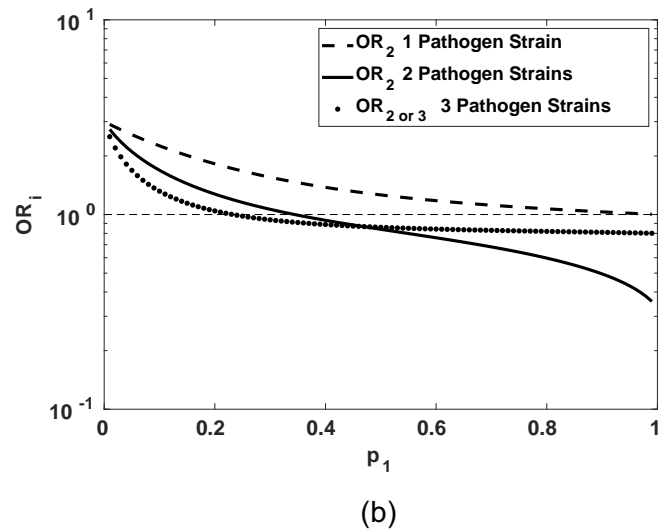
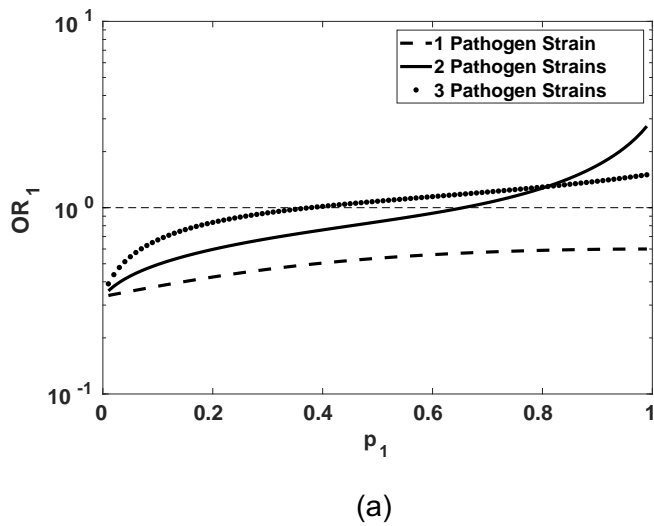


Fig. 3. The relationship between HLA allele frequencies and OR_i . The three different line styles indicate scenarios with 1, 2 and 3 pathogen strains. Panel (a) illustrates how the odds ratio for allele 1 changes as p_1 is increased. Panel (b) illustrates how the odds ratio for every other allele other than allele 1 varies as p_1 is increased. The parameter values are: $d = 0.01$, $\beta_i = 0.06$, $\mu_i = 0.02$ and $\sigma_i = 0.02$ ($i = 1,2,3$).

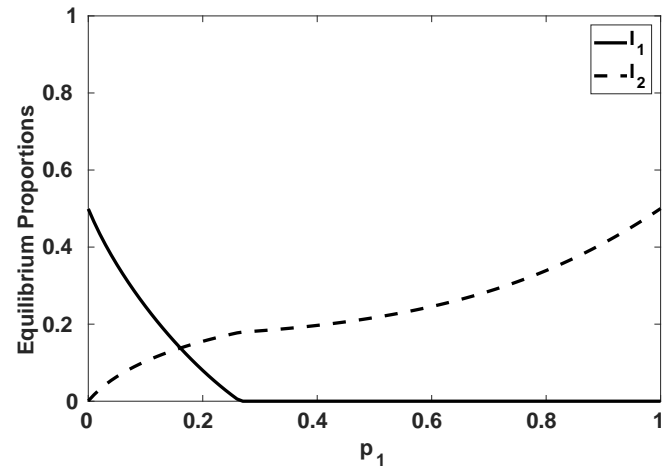
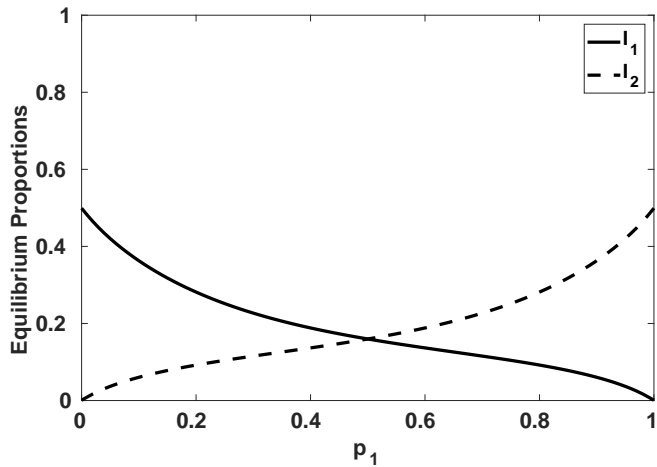


Fig. 4. The relationship between the steady state values of I_1 and I_2 and the frequency of HLA allele 1, in a 2 strain, 2 allele model where possessing allele i is necessary to mount a memory immune response against strain i . The equilibrium values of I_1 and I_2 against p_1 . I_i is the total proportion of the population infected with pathogen strain i . The parameter values are: $d = 0.01$, $\beta_i = 0.06$, $\mu_i = 0.02$ and $\sigma_i = 0.02$ ($i = 1, 2$).

Fig. 6. Pathogen strain dynamics are affected by increasing the recovery rate associated with HLA allele 1 relative to that associated with HLA allele 2. The graph is of I_i against p_1 for the two pathogen strain model. The parameter values are: $d = 0.01$, $\beta_i = 0.06$, $\mu_1 = 0.2$, $\mu_2 = 0.02$ and $\sigma_i = 0.02$ ($i = 1, 2$).

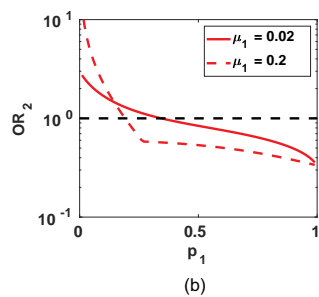
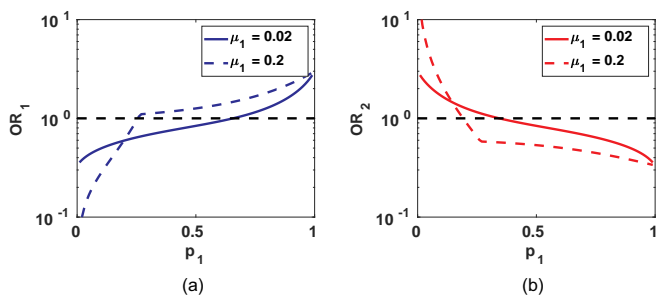


Fig. 5. OR_i and how its relationship with the frequency of allele 1 changes as the rate of recovery associated with HLA allele 1 becomes higher than that of associated with HLA allele 2. OR_1 against p_1 for the two pathogen strain model (a). OR_2 against p_1 for the two pathogen strain model (b). The parameter values are: $d = 0.01$, $\beta_i = 0.06$, $\mu_2 = 0.02$ and $\sigma_i = 0.02$ ($i = 1, 2$).

infection of each pathogen strain at equilibrium (Fig. 6).

When $\mu_1 > \mu_2$, pathogen 1 can only invade the system at low levels of p_1 . The shorter duration of infection with pathogen strain 1 associated with allele 1 makes it harder for pathogen strain 1 to thrive. This in turn causes pathogen strain 2 to be the only circulating pathogen strain for the majority of p_1 values. In this environment only allele 2 will provide any protection.

These results demonstrate that, in a multi-strain pathogen system, the protective effect of an HLA allele may arise from being associated with a relatively slow recovery rate. Counter intuitively, this “poor” HLA allele could end up being protective in a population due to it helping one pathogen strain to out compete another.

3.3. Genotype/infection associations for HLAs which protect against specific pathogen strains are only detectable under limited circumstances

To investigate how strain specificity of HLA/pathogen recognition may impact our ability to detect HLA-infectious disease associations in real world case control studies, we calculated 95% confidence intervals

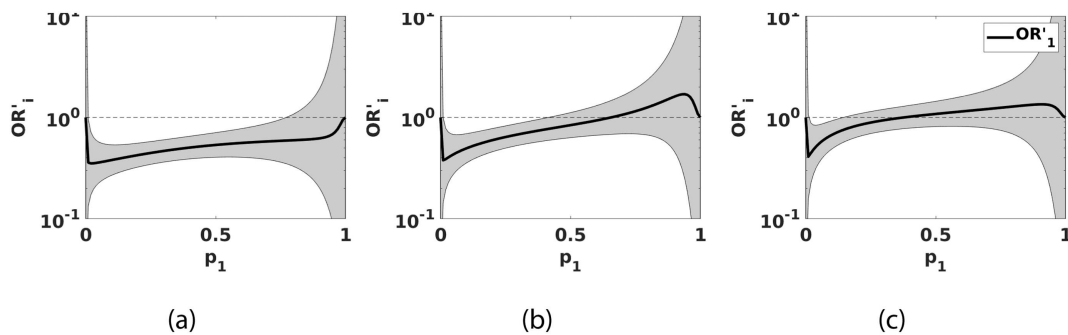


Fig. 7. OR_1' and its relationship with the frequency of HLA allele 1. The highlighted regions are the 95% confidence intervals for OR_1' . (a) illustrates a system with one pathogen strain and two HLA alleles in the population. (b) illustrates a system with two pathogen strains and two strain specific HLA alleles within the population. (c) illustrates a system with three pathogen strains and three strain specific HLA alleles within a population. The confidence intervals were calculated with a sample size of 500 cases and 500 controls (see Methods for further details). The parameter values are: $d = 0.01$, $\beta_i = 0.06$, $\mu_i = 0.02$ and $\sigma_i = 0.02$ ($i = 1, 2, 3$).

for OR_1' , on the assumption of 500 infected cases and 500 disease-free controls (see Methods for further details).

We considered three different systems containing only HLA alleles recognising single pathogen strains. In a two HLA allele system where just pathogen strain 1 is present, for most of the range of p_1 OR_1' has a confidence interval small enough that a case control study would be able to conclude that HLA type has an effect on infection (Fig. 7a). In a two allele system where both strains 1 and 2 are present a case control study would only identify that an HLA type protects against the prevailing local infection for a smaller range of p_1 (Fig. 7b). For the three strain, three allele system, with these parameter values, there is an even smaller range of values of p_1 where a case control study of 500 cases and 500 controls could declare if OR_1' is below 1. This effect arises because the value of p_1 at which the odds ratio moves from below 1 to above 1 (henceforth p_{crit}), shifts to lower values of p_1 as the complexity of the system increases from two to three strains (examined further in Section 4.2 of the Appendix).

Fig. S4 in Section 4.1 of the supplementary material illustrates the behaviour of the system if only 100 cases and 100 controls are used. Now, despite there still being a relationship between HLA frequency and protectiveness (i.e. OR_1 still correlates with p_1), there is no frequency of p_1 at which the 95% confidence interval for OR_1 does not encompass 1. However, if we change the properties of the pathogen by increasing R_0 , we do find values of p_1 at which the 95% confidence interval for OR_1 does not include 1, even in the smaller case control study (Fig. S5).

To further explore how pathogen properties affect the ability of case control studies to detect protective or risky HLAs, we define Ω as the difference between the maximum and minimum values of p_1 at which the odds ratio for HLA type 1 is significantly below (Ω_B) or above (Ω_A) 1 (which we take to be when the 95% confidence interval does not encompass 1). Ω therefore represents the ease with which a HLA protective (Ω_B) or risky (Ω_A) association against local prevailing infection might be detected in our system. If Ω is large, then the detection of the association is less dependent on a specific frequency of p_1 . Figs. 8 and 9 illustrates how Ω_B and Ω_A vary for different recovery rates and different values of the basic reproductive number (R_0) of the pathogen in the 2 strain/2 allele and 3 strain/3 allele scenarios. For these simulations $\mu_i = \sigma_i = \sigma$ thus σ refers to all recovery rates in the system.

Figs. 8 and 9 shows that Ω_B and Ω_A are both 0 when the recovery rate (σ) is low (black region). Detecting any HLA-infection association is difficult if the infectious period is too long. This is unsurprising, because the key to protection against infection in our system is the ability, or not, of a host to mount a memory immune response. A pathogen which has a recovery rate similar to the mortality rate of the host cannot generate a strong HLA-dependent signal within such a system.

Fig. 8 relates to the two strain, two allele system. Ω_B is a measure of our ability to detect whether HLA type 1 is protective against the prevailing local infection. Detecting a protective effect of HLA type 1 in a two strain-two allele system requires that homozygotes without HLA type 1 (e.g. genotype 22) bear the brunt of infections. This can only occur if pathogen strain 1 (to which genotype 22 is especially vulnerable) is circulating in the population. The ability to detect protective associations over a wide range of p_1 (a high value of Ω_B in Fig. 8a) implies that pathogen 1 is able to circulate at a reasonably high frequency in the population even as the proportion of allele 1 becomes relatively high. This requires there to remain a good balance between strain 2, which benefits from the increased frequency of genotype 11, and strain 1. The balance between the strains is affected by both R_0 and σ (explored in Section 4.2 of the Appendix). In most cases, increasing R_0 increases the coexistence of the strains, which is reflected in the broad relationship between R_0 and Ω_B seen in Fig. 8a.

Detecting that HLA type 1 is risky (i.e. increases susceptibility to the prevailing local infection) in the two strain-two allele system (Ω_A) requires genotype 11 to bear the brunt of infections. This requires the presence of a high level of pathogen strain 2 among circulating strains. A high value of Ω_A implies that pathogen 2 is able to dominate the population even at relatively low frequencies of allele 1. This effect, likewise, depends on the relative values of R_0 and σ (see Section 4.2 of Appendix). Broadly, the faster the recovery rate, the easier it is for one pathogen strain to out-compete the other. Ω_A , therefore, tends to increase with increasing σ , as seen in Fig. 8b.

When we move from a two strain, two allele system to a three strain, three allele system (Fig. 9), Ω_B becomes smaller and Ω_A becomes larger. In other words, it has become easier to detect risky associations and harder to detect protective associations. It also appears that so long as σ and R_0 are both above a certain threshold, they have little effect on Ω_A or Ω_B . When we calculate an odds ratio for the effect of HLA-1 on infection in a three allele-three strain system, we are comparing the distribution of “HLA-1 containing genotypes” (11, 12, 13) and “non HLA-1 containing genotypes” (22, 33, 23) among cases and controls. Unlike in the two strain, two allele system, both sets of genotypes now include heterozygotes as well as homozygotes, and both sets of genotypes include a genotype particularly vulnerable to one of the strains (genotype 11 and genotype 33 are both equally vulnerable to strain 2). These factors increase the similarity of the two sets of genotypes being compared by the odds ratio, and mean that a lower value of p_1 is necessary for HLA-1 to be detected as protective against the prevailing local infection. In Section 4.2 of the Appendix we derive an expression for p_{crit} , the frequency at which an HLA type switches from being protective to being risky, and show that p_{crit} is more limited in the three strain, three allele model than in the two strain, two allele model,

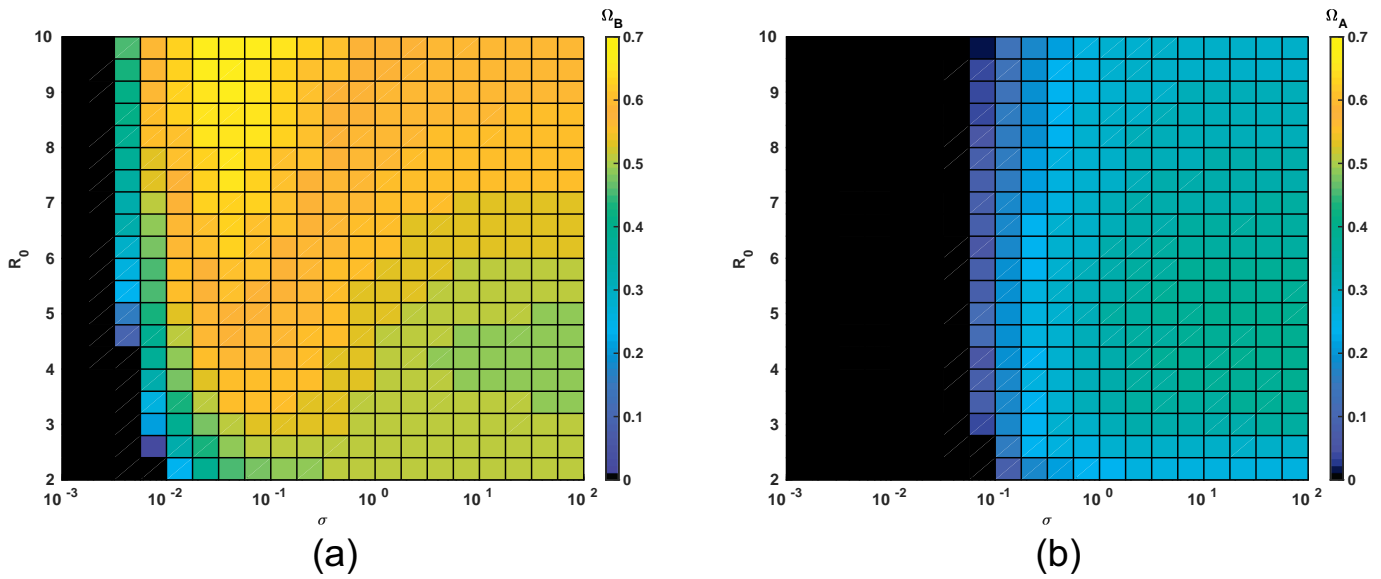


Fig. 8. A measure of how well protective (Ω_B) or risky (Ω_A) associations can be detected in simulated case control studies in the two strain, two allele system. Panel (a) shows Ω_B , panel (b) shows Ω_A . R_0 is increased from 2 to 10 along the y axis of each heat map, σ is increased logarithmically from 10^{-3} to 10^2 along the x axis of each heat map. β_i is calculated $\beta_i = R_0(d + \sigma)$ (for $i = 1, 2$) and $d = 0.01$.

accounting for the plateauing of Ω in Fig. 9.

3.4. Model behaviour is insensitive to co-infection, but breaks down at high levels of strain transcending immunity

As noted in Section 2.2, parameter c controls the effect to which a host can be infected simultaneously with two pathogen strains and α controls the proportion of hosts infected with a pathogen who recover to being immune to both pathogen strains (henceforth strain transcending immunity). So far, we have displayed results where $c = 0$ and $\alpha = 0$.

Our key result is that the protectiveness of a strain-specific HLA allele against the prevailing local infection caused by a multi-strain pathogen is correlated with its population frequency. We define the

breakdown of this trend as the case where neither OR_1 nor OR_2 switch from above/below to below/above 1 over the range of p_1 . Fig. 10 illustrates when this breakdown occurs in the main two strain, two allele model as c and α are varied. We also explore the impact of introducing a discrepancy in fitness between the pathogen strains in the model, achieved by varying the transmission parameter, β .

When there is only a small difference between β_1 and β_2 (Fig. 10a), the relationship between HLA allele frequency and protectiveness against infection exists at all levels of co-infection, and in the presence of some strain transcending immunity, but not when strain transcending immunity is complete ($\alpha = 1$). As the fitness difference between the strains is increased, the relationship between HLA allele frequency and the protectiveness of that allele against the prevailing local infection breaks down at lower levels of strain transcending

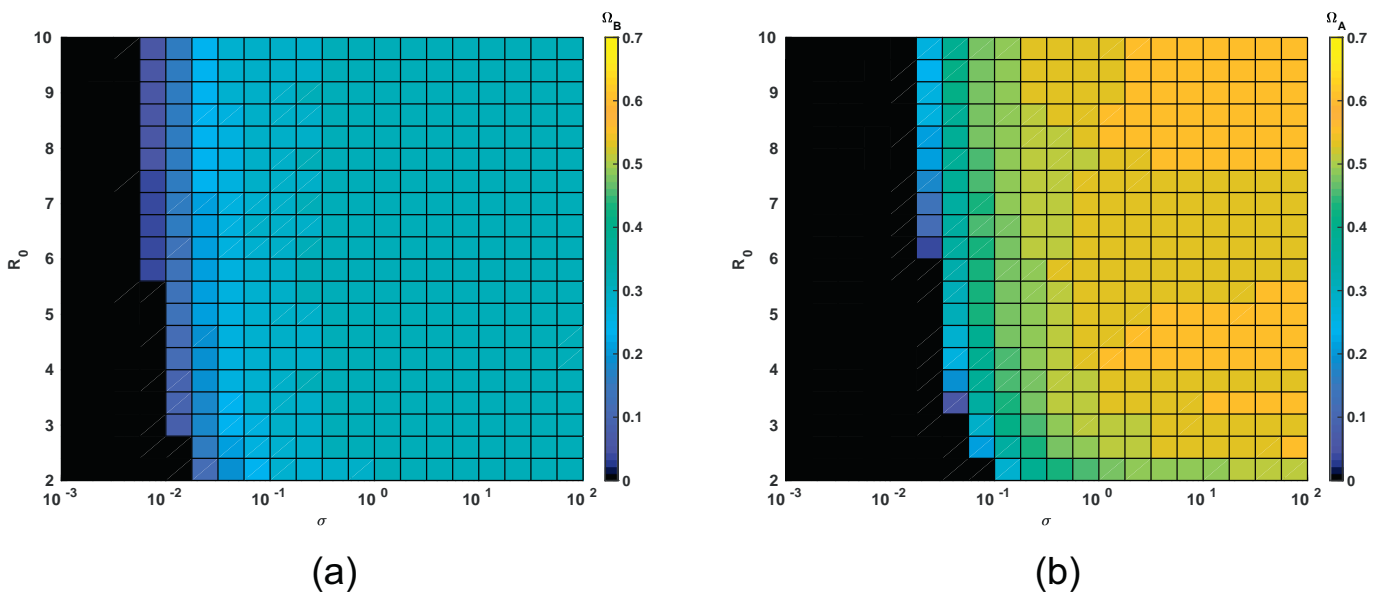


Fig. 9. A measure of how well protective (Ω_B) or risky (Ω_A) associations can be detected in simulated case control studies in the three strain, three allele system. Panel (a) shows Ω_B , panel (b) shows Ω_A . R_0 is increased from 2 to 10 along the y axis of each heat map, σ is increased logarithmically from 10^{-3} to 10^2 along the x axis of each heat map. β_i is calculated $\beta_i = R_0(d + \sigma)$ (for $i = 1,2,3$) and $d = 0.01$.

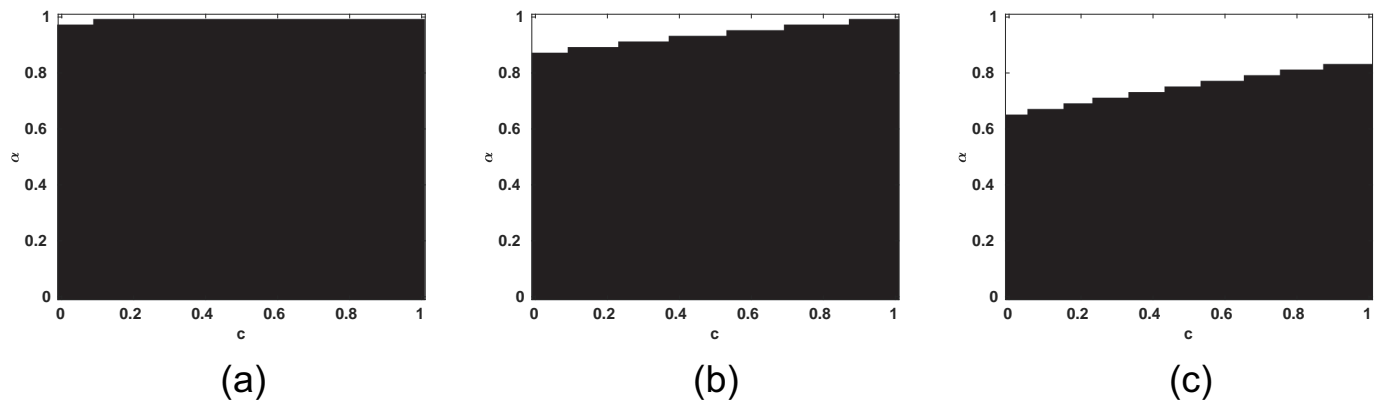


Fig. 10. Showcasing how co-infection and strain transcending immunity affects the outcome of the original model. The three figures are of binary heat maps where a white indicates the trend disappearing, black indicates when the trend is still present. The parameter values are: $d = 0.01$, for (a) $\beta_1 = 0.061$, for (b) $\beta_1 = 0.065$, for (c) $\beta_1 = 0.08$, $\beta_2 = 0.06$, $\mu_i = 0.02$, $\sigma_i = 0.02$ ($i = 1, 2$).

immunity, but this effect can be countered by allowing more co-infection to occur (Fig. 10b and c). Overall, the relationship we have identified requires that both strains of pathogen are able to co-circulate in the population, over at least some values of p_1 . Both strains need to be present in order to drive the relationship between HLA frequency and protectiveness against local prevailing infection. Any process that promotes co-circulation (e.g. co-infection) makes such a relationship more likely, and any process which acts against co-circulation (e.g. fitness differences between strains; strain transcending immunity) makes such a relationship less likely.

4. Discussion

Population level differences in the protectiveness of HLAs against infection with multi-strain pathogens may be a result of HLAs being strain specific in their effects. Here we simulated such HLA-strain interactions in an epidemiological model. We showed that the adaptation of a multi strain pathogen to the HLAs of a host population will generate a negative association between the population frequency of an HLA allele which is beneficial against a specific pathogen strain and the protectiveness of that HLA type against the prevailing local infection. We also showed that if certain HLA types cause hosts to recover from infection with particular pathogen strains more quickly than other HLAs, those HLAs are less likely to protect against infection in general, since the strains against which those HLA types are particularly effective may be outcompeted by other pathogen strains in the population.

All of our simulations included HLA alleles/types which were “protective” in the sense that they conferred the ability to mount an immune response against a specific strain. However, despite these extreme HLA-strain associations, we found that under a great many circumstances no association would be detected between HLA type and infection in general in a typical case control study (Fig. 7 and Fig. S4). This is because the frequencies of the prevailing pathogen strains adapt to the HLAs present in the population. Previous efforts to use case control studies to identify HLA-infection associations for multi strain pathogens may have been hindered by such effects.

We presented results for two HLA allele and three HLA allele systems. Two and three alleles are far fewer than the diversity of HLA alleles present in human populations (e.g. there are 100 HLA-B alleles observed in the German population (Gonzalez-Galarza et al., 2010)). However, classical definitions of HLA alleles, based on the amino acid sequence of the HLA molecule, do not necessarily reflect functionally relevant properties. Class I HLA alleles are grouped into 10 HLA supertypes based on their binding capabilities (Sidney et al., 2008). Associations between HLA supertypes and susceptibility to and severity of tuberculosis have been found (Balamurugan et al., 2004). Other HLA

supertype associations have been found for HIV (Trachtenberg et al., 2003; MacDonald et al., 2000). Functional grouping of HLA alleles has also been performed for specific pathogens such as the H1N1 influenza. Mukherjee et al classified the HLA diversity of human populations into “response types” which are HLA class I genotypes that share similar epitope binding pools to the H1N1 virus (Mukherjee and Chandra, 2014). If we were to classify HLA alleles by whether or not they were able to bind a specific peptide sequence from an immunodominant epitope in a particular pathogen, then there would only be two types of HLA: those that bind the peptide of interest and those that do not. Our two allele and three allele systems (or more properly “two binding type” and “three binding type” systems) can therefore deliver insights into how HLA-strain systems operate, even if our simulated alleles are not directly equivalent to known HLA alleles.

We focused on the impact of strain-specific HLA types which we assumed to be mutually exclusive in their ability to protect against different pathogen strains (i.e. where the ability to display an immunogenic peptide from strain 1 precludes the ability to display an immunogenic peptide from strain 2). This has a precedent, albeit in a non-human system. Experiments in chickens have shown that single MHC types can be associated with completely opposite responses to different pathogen strains (McBride et al., 1981). GB1 line chickens are susceptible to rous sarcoma virus (RSV) subgroup C PR-RSV strain, but resistant to rsv subgroup C B77 strain. GB2 line chickens (of a different MHC type (Briles et al., 1982)) display exactly the opposite pattern (resistant to PR-RSV but susceptible to B77).

MHC/HLA molecules exhibit a range of binding properties. Some are specialist (binding only a narrow range of types of peptides), others are more generalist (able to bind a wider range of peptides). It is possible that the maintenance of both generalist and specialist MHC/HLAs in populations may be because each are best adapted to respond to different types of pathogen (Chappell et al., 2015). Chappell et al (Chappell et al., 2015) highlight Marek’s disease in chickens as an example of an infectious disease where generalist MHCs have been shown to provide the best responses, and HIV in humans as an example where specialist HLAs are associated with the most effective immune responses. Kaufman et al (Kaufman, 2018) gives review of generalist and specialist MHC class I molecules. It might seem that, since generalists can present a wide range of peptides, specialist MHCs/HLAs are redundant and should not be maintained in populations. The continued existence of specialist MHCs/HLAs can be understood if we suppose that sometimes the best immune responses are associated with mounting an immune response against a very particular pathogen peptide. Even if a generalist MHC/HLA can present that peptide, it will not necessarily present it as reliably and consistently as a specialist MHC/HLA with narrower binding properties. As we described

previously, our model applies to systems where it is beneficial for a host HLA to present a particular immunodominant peptide, but where not all HLAs are capable of doing this. This therefore implies the HLA types in our model are relatively narrow in their binding (although we do not mean this to imply they are specialised only to the hypothetical pathogen at hand). As illustrated in Fig. S2, the inclusion of an HLA type capable of presenting peptides from any of the strains in the system (the “perfect allele”, which could also be called a generalist) does not change our overall conclusions provided there are still at least 2 types of more specialist HLA in the system.

Figs. 8 and 9 shows that it is plausible for a real world case control study to detect a significant effect of a strain specific HLA type on infection, driven by the processes modelled here. The only necessary conditions are that the recovery rate of the pathogen is faster than the background mortality rate of the host and the frequency of the HLA type is within a certain range. However, defining “HLA type” is problematic, since the functionally relevant “HLA type” for a particular pathogen might in fact be a set of HLAs encoded by a range of different alleles, which share the property of being able to bind a critically important (and unknown to us) pathogen peptide. Nevertheless, our model suggests a method to identify such systems. If the odds ratio for being infected with a disease caused by a multi-strain pathogen changes for the same HLA type or supertype in different populations, and if there is a positive relationship between that odds ratio and the frequency of that HLA type or supertype, this would be highly suggestive that there are HLA-pathogen strain associations to be found.

In the introduction we noted three pathogens where our model is likely to be particularly relevant: *Neisseria meningitidis*, *Streptococcus pneumoniae* and *Plasmodium falciparum* malaria. Existing work attempting to link the risk of disease caused by these pathogens to host genetics focuses on severe disease outcomes (invasive meningococcal or pneumococcal disease or severe malaria), rather than infection (or “carriage”) per se. Our model does not attempt to simulate the development of severe infection. However, if severe disease is associated with particular strains, and immunity to strains outside of this subset does not prevent severe disease, our model can be interpreted in terms of how HLAs impact the accumulation of immunity to just those strains that are capable of causing severe disease. For *Streptococcus pneumoniae* and *Neisseria meningitidis* it is widely accepted that only a subset of strains cause severe disease (Enright and Spratt, 1998; Peltola, 1983). The situation for *P. falciparum* is complicated by the antigenic variation *P. falciparum* exhibits during infection, but the presence of particular group A var. genes in the genomes of parasites may determine their potential to cause severe disease (Bull et al., 2008). We propose that a correlation between population frequency of an HLA type and the odds ratio for severe disease associated with the presence of that HLA type would be suggestive of an HLA-strain relationship in any of these systems.

Our model fixed HLA frequencies within each simulation. Over the longer term, HLA frequencies themselves must be evolving under selection from pathogens (Jeffery and Bangham, 2000; Prugnolle et al., 2005; Hertz et al., 2011). Extending our model into a co-evolutionary framework (similar to those explored by Penman (Penman et al., 2013) and MacPherson (MacPherson et al., 2018)) would deliver further insights, especially into long term HLA supertype dynamics. We could also increase the complexity of the model further to include for more alleles or multiple linked HLA allele loci, which previous modelling work has shown to enable a range of co-evolutionary outcomes (Penman et al., 2013).

We have explored one way in which epidemiology may affect our ability to detect the importance of HLA type in infectious disease. However, other processes can also cause case control studies to generate conflicting results, including within-host adaptation of chronic viral diseases and epistasis between HLAs and other loci. Within-host adaptation of HIV to escape HLA restricted immune responses has spilled over into population level adaptation that renders previously

protective HLA alleles non protective (Kawashima et al., 2009; Payne et al., 2014). HLAs are also known to interact epistatically with Killer-cell immunoglobulin-like receptors (KIRs) (Khakoo et al., 2004; Martin et al., 2002), and failing to account for KIRs genotype could also lead to HLA-infectious disease associations being overlooked.

Multi-strain pathogens include infections of vast public health significance such as malaria, TB, influenza, and streptococcal and meningococcal bacterial disease. A deeper understanding of the immunogenetics of such infections could pave the way to develop more personalised vaccines or other control measures. However, finding associations between HLA alleles and these infectious diseases has been an ongoing challenge. The model we present here explores the consequences of one potential mechanism of HLA-pathogen interaction and suggests a method to detect the signature of HLA-strain relationships in combined analyses of case control studies in different populations.

Declaration of Competing Interest

None.

Acknowledgements

We gratefully acknowledge the following sources of funding: CFW and MJK are funded by the Engineering and Physical Sciences Research Council UK and the Medical Research Council UK through the MathSys CDT (grant number EP/L015374/1). LP is funded by the Wellcome Trust and the Royal Society UK (grant 202562/Z/16/Z)

Appendix A. Supplementary text and figures

Supplementary text and figures for this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104344>.

References

- Aslam, A., Mason, A., Zemenides, S., Chan, H., Nováková, L., Branny, P., Finn, A., Chapel, H., Ogg, G.S., 2010. Rapid effector function of circulating cd4+ t cells specific for immunodominant regions of the conserved serine/threonine kinase found in *Streptococcus pneumoniae* (stkp) in healthy adults. *FEMS Immunol. Med. Microbiol.* 60 (2), 113–122.
- Aslam, A., Chapel, H., Ogg, G., 2011. Direct ex-vivo evaluation of pneumococcal specific t-cells in healthy adults. *PLoS One* 6 (10), e25367.
- Balamurugan, A., Sharma, S.K., Mehra, N.K., 2004. Human leukocyte antigen class i superotypes influence susceptibility and severity of tuberculosis. *J. Infect. Dis.* 189 (5), 805–811.
- Briles, W., Bumstead, N., Ewert, D., Gilmour, D., Gogusev, J., Hala, K., Koch, C., Longenecker, B., Nordskog, A., Pink, J., et al., 1982. Nomenclature for chicken major histocompatibility (b) complex. *Immunogenetics* 15 (5), 441–447.
- Bull, P.C., Buckee, C.O., Kyes, S., Kortok, M.M., Thathy, V., Guyah, B., Stoute, J.A., Newbold, C.I., Marsh, K., 2008. *Plasmodium falciparum* antigenic variation. mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol. Microbiol.* 68 (6), 1519–1534.
- Carrington, M., O'Brien, S.J., 2003. The influence of hla genotype on aids. *Annu. Rev. Med.* 54 (1), 535–551.
- Chappell, P.E., Meziane, E.K., Harrison, M., Magiera, L., Hermann, C., Mears, L., Wrobel, A.G., Durant, C., Nielsen, L.L., Buus, S., et al., 2015. Expression levels of mhc class i molecules are inversely correlated with promiscuity of peptide binding. *Elife* 4, e05345.
- Davenport, V., Guthrie, T., Findlow, J., Borrow, R., Williams, N.A., Heyderman, R.S., 2003. Evidence for naturally acquired t cell-mediated mucosal immunity to neisseria meningitidis. *J. Immunol.* 171 (8), 4263–4270.
- Dunstan, S.J., Hue, N.T., Han, B., Li, Z., Tram, T.T.B., Sim, K.S., Parry, C.M., Chinh, N.T., Vinh, H., Lan, N.P.H., et al., 2014. Variation at hla-drb1 is associated with resistance to enteric fever. *Nat. Genet.* 46 (12), 1333.
- Enright, M.C., Spratt, B.G., 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 144 (11), 3049–3060.
- Gart, J.J., 1966. Alternative analyses of contingency tables. *J. R. Stat. Soc. Ser. B Methodol.* 164–179.
- Gonzalez-Galarza, F.F., Christmas, S., Middleton, D., Jones, A.R., 2010. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res.* 39 (suppl_1), D913–D919.
- Gupta, S., Hill, A.V., 1995. Dynamic interactions in malaria: host heterogeneity meets parasite polymorphism. *Proc. R. Soc. Lond. B* 261 (1362), 271–277.
- Hedrick, P.W., 2002. Pathogen resistance and genetic variation at mhc loci. *Evolution* 56

- (10), 1902–1908.
- Hertz, T., Nolan, D., James, I., John, M., Gaudieri, S., Phillips, E., Huang, J.C., Riadi, G., Mallal, S., Jovic, N., 2011. Mapping the landscape of host-pathogen coevolution: Hla class I binding and its relationship with evolutionary conservation in human and viral proteins. *J. Virol.* 85 (3), 1310–1321.
- Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J., Greenwood, B.M., 1991. Common west african hla antigens are associated with protection from severe malaria. *Nature* 352 (6336), 595.
- Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot Jr., C.C., Wright, M.W., et al., 2004. Gene map of the extended human mhc. *Nat. Rev. Genet.* 5 (12), 889.
- Jeffery, K.J., Bangham, C.R., 2000. Do infectious diseases drive mhc diversity? *Microbes Infect.* 2 (11), 1335–1341.
- Just, J.J., 1995. Genetic predisposition to hiv-1 infection and acquired immune deficiency virus syndrome: a review of the literature examining associations with hla. *Hum. Immunol.* 44 (3), 156–169.
- Kaufman, J., 2018. Generalists and specialists: a new view of how mhc class I molecules fight infectious pathogens. *Trends Immunol.* 39 (5), 367–379.
- Kawashima, Y., Pfafferoth, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga, H., Fujiwara, M., Hachiya, A., Koizumi, H., et al., 2009. Adaptation of hiv-1 to human leukocyte antigen class I. *Nature* 458 (7238), 641.
- Keeling, M.J., Rohani, P., 2011. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Khakoo, S.I., Thio, C.L., Martin, M.P., Brooks, C.R., Gao, X., Astemborski, J., Cheng, J., Goedert, J.J., Vlahov, D., Hilgartner, M., et al., 2004. Hla and nk cell inhibitory receptor genes in resolving hepatitis c virus infection. *Science* 305 (5685), 872–874.
- Lyke, K.E., Fernández-Viña, M.A., Cao, K., Hollenbach, J., Coulibaly, D., Kone, A.K., Guindo, A., Burdett, L.A., Hartzman, R.J., Wahl, A.R., et al., 2011. Association of hla alleles with *Plasmodium falciparum* severity in malian children. *Tissue Antigens* 77 (6), 562–571.
- MacDonald, K.S., Fowke, K.R., Kimani, J., Dunand, V.A., Nagelkerke, N.J., Blake Ball, T., Oyugi, J., Njagi, E., Gaur, L.K., Brunham, R., et al., 2000. Influence of hla supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J. Infect. Dis.* 181 (5), 1581–1589.
- MacPherson, A., Otto, S.P., Nuismer, S.L., 2018. Keeping pace with the red queen: identifying the genetic basis of susceptibility to infectious disease. *Genetics* 208 (2), 779–789.
- Martin, M.P., Gao, X., Lee, J.-H., Nelson, G.W., Detels, R., Goedert, J.J., Buchbinder, S., Hoots, K., Vlahov, D., Trowsdale, J., et al., 2002. Epistatic interaction between kir3ds1 and hla-b delays the progression to aids. *Nat. Genet.* 31 (4), 429.
- McBride, R., Cutting, J., Schierman, L., Strebel, F., Watanabe, D., 1981. Mhc gene control of growth of avian sarcoma virus-induced tumours in chickens: a study on the role of virus strain. *Int. J. Immunogenet.* 8 (3), 207–214.
- McLaren, P.J., Coulonges, C., Bartha, I., Lenz, T.L., Deutsch, A.J., Bashirova, A., Buchbinder, S., Carrington, M.N., Cossarizza, A., Dalmau, J., et al., 2015. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of hiv-1 virus load. *Proc. Natl. Acad. Sci.* 112 (47), 14658–14663.
- Mordmüller, B., Surat, G., Lagler, H., Chakravarty, S., Ishizuka, A.S., Lalremruata, A., Gmeiner, M., Campo, J.J., Esen, M., Ruben, A.J., et al., 2017. Sterile protection against human malaria by chemoattenuated pfsz vaccine. *Nature* 542 (7642), 445.
- Mukherjee, S., Chandra, N., 2014. Grouping of large populations into few cti immune response-types from influenza h1n1 genome analysis. *Clin. Transl. Immunol.* 3 (8), e24.
- Oliveira-Cortez, A., Melo, A., Chaves, V., Condino-Neto, A., Camargos, P., 2016. Do hla class II genes protect against pulmonary tuberculosis? A systematic review and meta-analysis. *Eur. J. Clin. Microbiol. Infect. Dis.* 35 (10), 1567–1580.
- Payne, R., Muenchhoff, M., Mann, J., Roberts, H.E., Matthews, P., Adland, E., Hempenstall, A., Huang, K.-H., Brockman, M., Brumme, Z., et al., 2014. Impact of hla-driven hiv adaptation on virulence in populations of high hiv seroprevalence. *Proc. Natl. Acad. Sci.* 111 (50), E5393–E5400.
- Peltola, H., 1983. Meningococcal disease: still with us. *Rev. Infect. Dis.* 5 (1), 71–91.
- Penman, B.S., Ashby, B., Buckee, C.O., Gupta, S., 2013. Pathogen selection drives non-overlapping associations between hla loci. *Proc. Natl. Acad. Sci.* 110 (48), 19645–19650.
- Pierini, F., Lenz, T.L., 2018. Divergent allele advantage at human mhc genes: signatures of past and ongoing selection. *Mol. Biol. Evol.* 35 (9), 2145–2158.
- Prugnolle, F., Manica, A., Charpentier, M., Guégan, J.F., Guernier, V., Balloux, F., 2005. Pathogen-driven selection and worldwide hla class I diversity. *Curr. Biol.* 15 (11), 1022–1027.
- Relethford, J.H., 2012. *Human Population Genetics*. 7 John Wiley & Sons.
- Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P., Marsh, S.G., 2014. The ipd and imgt/hla database: allele variant databases. *Nucleic Acids Res.* 43 (D1), D423–D431.
- Salie, M., van der Merwe, L., Möller, M., Daya, M., van der Spuy, G.D., van Helden, P.D., Martin, M.P., Gao, X.-j., Warren, R.M., Carrington, M., et al., 2013. Associations between human leukocyte antigen class I variants and the *Mycobacterium tuberculosis* subtypes causing disease. *J. Infect. Dis.* 209 (2), 216–223.
- Sambaturu, N., Mukherjee, S., López-García, M., Molina-Paris, C., Menon, G.I., Chandra, N., 2018. Role of genetic heterogeneity in determining the epidemiological severity of h1n1 influenza. *PLoS Comput. Biol.* 14 (3), e1006069.
- Sidney, J., Peters, B., Frahm, N., Brander, C., Sette, A., 2008. Hla class I supertypes: a revised and updated classification. *BMC Immunol.* 9 (1), 1.
- Spurgin, L.G., Richardson, D.S., 2010. How pathogen drive genetic diversity: Mhc, mechanisms and misunderstandings. *Proc. R. Soc. B Biol. Sci.* 277 (1684), 979–988.
- Sveinbjornsson, G., Gudbjartsson, D.F., Halldorsson, B.V., Kristinsson, K.G., Gottfredsson, M., Barrett, J.C., Gudmundsson, L.J., Blondal, K., Gylfason, A., Gudjonsson, S.A., et al., 2016. Hla class II sequence variants influence tuberculosis risk in populations of european ancestry. *Nat. Genet.* 48 (3), 318.
- Thursz, M.R., Kwiatkowski, D., Allsopp, C.E., Greenwood, B.M., Thomas, H.C., Hill, A.V., 1995. Association between an mhc class II allele and clearance of hepatitis B virus in the gambia. *N. Engl. J. Med.* 332 (16), 1065–1069.
- Tian, C., Hromatka, B.S., Kiefer, A.K., Eriksson, N., Noble, S.M., Tung, J.Y., Hinds, D.A., 2017. Genome-wide association and hla region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* 8 (1), 599.
- Toyo-Oka, L., Mahasirimongkol, S., Yanai, H., Mushihiro, T., Wattanapokayakit, S., Wichukhinda, N., Yamada, N., Smittipat, N., Juthayothin, T., Palittapongarnpim, P., et al., 2017. Strain-based hla association analysis identified hla-drb1* 09: 01 associated with modern strain tuberculosis. *Hla* 90 (3), 149–156.
- Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., et al., 2003. Advantage of rare hla supertype in hiv disease progression. *Nat. Med.* 9 (7), 928.
- Wiertz, E., Delvig, A., Donders, E., Brugghe, H., Van Unen, L., Timmermans, H., Achtman, M., Hoogerhout, P., Poolman, J., 1996. T-cell responses to outer membrane proteins of neisseria meningitidis: comparative study of the opa, opc, and pora proteins. *Infect. Immun.* 64 (1), 298–304.
- Woolf, B., 1955. On estimating the relation between blood group and disease. *Ann. Hum. Genet.* 19 (4), 251–253.