

Economic Performance, Human Cooperation and the Major Histocompatibility Complex*

José A.M. Borghans

Dep. of Immunology

UMC Utrecht

j.borghans@lab.azu.nl

Lex Borghans

ROA

Maastricht University

l.borghans@roa.unimaas.nl

Bas ter Weel

MERIT

Maastricht University

b.terweel@merit.unimaas.nl

16 June 2005

Abstract

This paper develops a theory and presents empirical evidence of the link between economic achievement and genetic evolution. Important properties for successful analysis of this link are found in the adaptive immune system and particularly in the major histocompatibility complex (MHC). The MHC is crucial for health; its observed level of diversity is related to economic conditions; and, its frequency-dependent distribution ensures very rapid evolution. The theory develops an understanding of the functioning of the immune system when interactions between agents are important by modelling interactions in terms of a linear public-good game. The model exhibits a trade-off revealing that every agent who is better off in having a different adaptive immune system from the other agents in the population is also part of the protecting belt of the other agents in which having similar immune systems is optimal. The theory predicts that low levels of MHC diversity minimize the incidence of infectious diseases in a population and constitute a social optimum, while the evolutionary stable equilibrium with selfish agents induces a maximum level of MHC diversity. The data have been collected from a large number of individual blood samples of people in 63 populations. Consistent with the theory the estimates suggest a negative association between health outcomes and MHC diversity, which suggests that low diversity improves health outcomes from a social perspective. Furthermore, the estimates suggest a negative correlation of MHC diversity with indicators of cooperation, which is consistent with more cooperation inducing the internalization of the health of other agents. Finally, the estimates present a robust negative correlation between MHC diversity and several measures of economic performance, which suggests that societies that are better able to cope with public good dilemmas are economically more successful.

Keywords: Economic outcomes; Cooperation, Genetic diversity; Mutual dependence

JEL Codes: O10; O50; D60; D90

*We would like to thank seminar participants at IZA, MERIT, the Economic Growth Conference in Lucca, the Evolution and Behavior Workshop in Maastricht, and Philippe Aghion, Rob de Boer, Arnaud Dupuy, Oded Galor, Omer Moav, Ido Pen, Gilles Saint-Paul, Adriaan van Zon, Simon Verhulst, Franjo Weissing, and Thomas Ziesemer for helpful discussions and useful comments, and Ella van den Berg-Loonen and Christina Voorter of the Tissue Typing Laboratory at the University Hospital in Maastricht for clarifying measurement issues concerning the major histocompatibility complex. We are grateful to the Netherlands Organisation for Scientific Research (NWO) for financial support.

1 Introduction

This paper develops a theory and presents empirical evidence of the link between economic achievement and genetic evolution. We argue that economic achievement is a crucial determinant of genetic evolution and make plausible that evolutionary forces have favored human characteristics that are well-adapted to economic circumstances. When populations are exposed to distinct economic conditions, they develop in dissimilar ways, so genetic differences can reveal divergent economic circumstances in a population.

To date economic research has only pointed to a theoretical link between economic and evolutionary forces. The economic literature has studied a number of human characteristics important for economic behavior and success, which are potentially subject to evolutionary selection (e.g., altruism and reciprocity). Also differences in preferences (e.g., discount rates and quantity-quality trade-offs for children) have been linked to genetic evolution. However, empirical linkages of these theoretical notions are hampered by a lack of knowledge about the actual genes related to these human characteristics and economic outcomes.¹

We take a different approach and study genes that fulfil a direct and crucial role in human survival and are linked to economic circumstances through the externalities that occur in the evolutionary selection process. The main obstacle to investigate the relationship between genetic evolution and economic outcomes is to select genes for which a significant link with economic circumstances can be obtained. There are three crucial conditions for the potential success of this research: (i) We have to study genes that are of eminent importance for evolutionary success; (ii) the fitness of these genes has to depend on and change with economic circumstances; and (iii) the evolution of these genes has to be relatively fast and keep abreast of changes in economic circumstances.

The genes we investigate in this paper encode the major histocompatibility complex (MHC), in humans called human leukocyte antigen (HLA). We apply information about HLA-A and HLA-B molecules in this analysis because knowledge about these molecules is most advanced in immunology and the classification and measurement of these two types of molecules is less subject to debate than for the other four existing types.² In mammals, but also in other species, the MHC encodes a set of closely linked genetic loci that play a crucial role in the adaptive immune system. The basic role of the MHC is to help control infectious diseases resulting from a pathogen that has managed to infect a host. Basically the MHC presents pathogen parts to the cell membrane, which triggers the growth of T cells to take care of the infection.³ The MHC is the most polymorphic gene known and because of its crucial role in the immune

¹See for example Fehr and Fischbacher [2003] who conclude that “to enhance the study of the evolution of human altruism, there is a great need for empirically testable predictions that are rigorously derived from the evolutionary models.”

²We provide details about this discussion and the advantages of using these data in Section 2 below.

³See Hansen and Sachs [1989] for an excellent overview study of the role of the MHC in immunology.

system it is subject to strong evolutionary pressure. This property of the MHC satisfies the first condition of linking economic outcomes to genetic evolution.

Different molecules do not outperform each other and polymorphism provides direct advantages to individuals with rare molecules. In the long run this would lead to a uniform distribution of MHC molecules within a population. However, the observed distribution of MHC molecules in populations reveals the properties of a frequency dependent distribution in which a small number of MHC molecules are frequently observed. In humans for example there are 19 different HLA-A molecules observed, of which the five most frequently observed cover more than 50 percent of the total frequency, and sometimes up to 90 percent. We argue that this pattern, and especially differences across populations, is a reflection of economic circumstances. There are two forces working in opposite directions. First, since agents are most likely infected through contact with others, embodying a rare type of MHC molecule has an advantage for the individual. The reason for this is that the deviant immune system of the individual agent might recognize pathogens that have not been recognized by the immune system of another agent in his neighborhood. Second, since infectious diseases can be transmitted by several agents, similarity of the adaptive immune system is advantageous from the perspective of a group of agents. The reason for this is that group members with deviant MHC molecules decrease the effectiveness of the protection belt against infections. Hence, the equilibrium with the maximum level of diversity is not necessarily optimal. This trade-off between the individual and society's optimum level of genetic diversity presents an externality that satisfies the second condition of a relationship between genetic evolution and economic circumstances.

Finally, a change in the observed frequency of existing MHC molecules in a population suffices to change the population's genetic diversity. The research advantage of this property of the MHC is that only a fraction of the population with certain MHC molecules needs to change to induce a change in the level of MHC diversity. In addition, the characteristics that need to change are already present in the population and do not require mutation and suppression of previously dominant genes. This leads to very rapid evolutionary processes, requiring only some four to ten generations for a substantial adjustment in the genetic characteristics of a population. Such a short period of substantial genetic changes satisfies the third requirement because these rapid changes are likely to keep pace with changes in economic circumstances.

We develop a theory of genetic evolution and human cooperation in which we assume that populations differ in the degree of cooperation.⁴ We define cooperation as the working together of agents in small groups in which they equally share the resulting output from joint production and model this in terms of

⁴This is consistent with the substantial differences in altruism and reciprocity across populations found in ultimatum games. See Fehr and Fishbacher [2003] for an overview.

a linear public-good game. In the game agents decide upon the fraction of production time they devote to individual and joint production. If an agent does not contribute sufficiently to joint production, he will be punished. When an agent contracts a disease he will be out of the production process for some time to recover. A key feature of the model is that the agent's income depends more on the performance of other agents when he cooperates more. Assuming that the agent's evolutionary success depends positively on his level of income, agents in more cooperative societies have more interest in the health of the agents they cooperate with. Agents not interested in cooperation do not care about the health of others and are better off having an immune system different from the other agents in close vicinity. Thus, when agents are completely selfish and only engage in individual production the evolutionary stable equilibrium will be such that a population's MHC diversity will be maximized leading to a uniform distribution of MHC molecules. Under certain conditions this evolutionary stable equilibrium will be the social optimum, in which the social planner is interested in minimizing the number of infectious diseases in a population. We will show that there will be a difference between this evolutionary stable equilibrium and the social optimum if the probabilities of the transmission and the contracting of different infectious diseases change. In such cases a lower level of MHC diversity is more beneficial for the health outcomes of the population as a whole. Finally, we derive an evolutionary stable equilibrium with cooperation, which equals the social optimum. In such an equilibrium the more agents depend on each other the more they have an interest in minimizing the number of infections in the group they belong to. In this equilibrium the level of MHC diversity is low, improving the health status of the population. As a consequence, economic performance increases and exhibits a negative correlation with MHC diversity.

We collected data from a large number of blood samples of individuals from many populations. We have merged these data to economic indicators at the country level and constructed a sample of 63 populations. We investigate three different relationships. First, a discrepancy between the selfish stable evolutionary equilibrium and the social optimum should be associated with a negative correlation between MHC diversity and health outcomes. The main reason for this is that in societies with higher levels of mutual dependence the MHC diversity is lower and the health status of agents better. Using life expectancy at birth and the fraction of people dying from viral infections as indicators of health we find a robust negative correlation between HLA-A and HLA-B diversity and health outcomes, controlling for several covariates that could interfere with this correlation. Second, cooperation leads to mutual dependence and the health of others will be internalized, so we expect a negative correlation between measures of cooperation and MHC diversity. In addition, cooperation between agents is of a small-scale nature, so we expect to observe particularly strong correlations between indicators capturing voluntary and more local

forms of cooperative behavior. The estimates suggest that measures of trust, civic cooperation and the importance of friends relative to family correlate negatively with MHC diversity, but that more macro-level indicators of cooperation such as democracy, confidence in the federal government and executive constraints are not significantly related to MHC diversity. These relationships turn out to be not driven by measures of fractionalization. Finally, societies that are better able to cope with public-good dilemmas will be economically more successful. More cooperation induces lower optimal levels of MHC diversity, so we expect a negative correlation between MHC diversity and economic performance. Figure I presents a preview of this relationship between the growth of GDP per capita in the period 1960-2000 and a country's MHC diversity. We have plotted the diversity of the populations' HLA-A and HLA-B molecules against GDP per capita growth in Figure Ia (for the diversity of HLA-A) and Ib (for HLA-B). Our estimates generally suggest that a one-standard deviation increase in MHC diversity reduces average annual GDP per capita growth by 0.2 to 0.4 percent. In the regression analysis we control for several covariates and show that the negative correlation between MHC diversity and economic outcomes is robust.

Finally, we discuss a number of possible alternative explanations for our findings. We discuss the general importance of genetic diversity for survival and group-selection arguments. In addition, we discuss the role of epidemics and migration to interfere with our results. The conclusion is that it is highly unlikely that these processes have interfered with our analysis.

The theory and evidence presented in this paper is related to three streams of research on economic growth and human evolution. With regard to the relationship between economics and health outcomes our work relates to Ofek [1999], Acemoglu et al. [2003], Saint-Paul [2003], Acemoglu and Johnson [2004], and Galor and Moav [2004]. They establish a link between health status and life expectancy. Acemoglu et al. [2003] argue that the direct effect of health on income is too small to explain cross-country differences in income. There is however an indirect effect through institutions (human capital, law enforcement etc.) which has positive effects on economic performance. Saint-Paul [2003] and Galor and Moav [2004] explicitly model genetic differences into their theories. Saint-Paul develops a theory of comparative advantage to show that under trade less fit humans are able to survive and to be as productive as more fit humans. Galor and Moav argue that increased pathogen pressure, worsened living conditions and environmental hazards in the movement from hunter-gatherer societies to agricultural societies reduced life expectancy initially, but at the same time selected individuals who were less vulnerable to this movement, which eventually increased life expectancy. However, these contributions do not offer any empirical support for a direct link between genetic evolution and health status.

On the relationship between human cooperation and evolution there is a literature on reciprocity

speculating on the link between genetic differences in human populations and the degree of reciprocity. For example, Henrich et al. [2001] have conducted ultimatum games in fifteen small-scale economies. They present interesting evidence on considerable differences in payoffs to cooperation in these societies. It turns out that the higher the payoffs to cooperation, the greater the level of cooperation in experimental games. What this literature have not established so far is a direct empirical link between cooperation and genetic evolution as we present in this paper. Other related literature in this respect are studies explaining cross-country differences in performance using measures of social capital or trust [e.g., Knack and Keefer 1997]. We show that trust and mutual dependence as mirrored by MHC diversity go together. The relationship with the cross-country growth papers starting with Barro [1991] is that we show that MHC diversity is a very interesting proxy and reliable reflection for cooperation explaining a significant portion of economic development and levels of performance.

The paper proceeds as follows. Section 2 presents the properties of the MHC that are important for economic theory. Section 3 presents the model and establishes the evolutionary equilibrium outcomes. Section 4 presents the empirical results of the relationship between life expectancy and MHC diversity, the correlation between measures of human cooperation and MHC diversity and the relationship between economic outcomes and MHC diversity. In Section 5 we present some alternative explanations for the observed patterns and show that these explanations are invalid. We end with a conclusion and avenues for future research.

2 The Major Histocompatibility Complex

Immunology is concerned with the means by which a host tries to control infectious diseases caused by pathogens.⁵ Protection against pathogens operates at three different levels. First, external defences, such as the skin, hamper the intrusion of pathogens. Second, if pathogens nevertheless manage to enter the host, the innate immune system rapidly tries to eliminate them using a large variety of cells and molecules. Finally, if pathogens are not captured and disposed by the innate immune system, the adaptive immune system, consisting of a huge diversity of lymphocytes, eliminates the pathogen (whenever possible) and retains immunological memory to be protected against future infections with the same pathogen.

In mammals and many other species a set of closely linked genetic loci is of fundamental importance for the adaptive immune system. This set of genes encodes the major histocompatibility complex (MHC).

⁵Pathogens are defined as infectious agents that cause diseases. Microbiologists generally distinguish six classes of pathogens: viruses (e.g., pox, and influenza), bacteria (e.g., syphilis and tuberculosis), fungi (e.g., ringworm and yeast), protozoa (e.g., trypanosome), helminths (e.g., hookworm and tapeworm) and ectoparasites (insects).

MHC molecules play a crucial role in the induction of T lymphocytes of the adaptive immune system.⁶ Its basic role is to present pathogens to the adaptive immune system.⁷

2.1 The Basic Role of MHC in the Immune System

A pathogen that has managed to infect a host, and has escaped the external defence and the innate immune system, may be harmful to the host. In case of a viral infection, infected cells become the producers of new pathogens. The cell decomposes its intracellular proteins, including the viral proteins, into short peptides.⁸ The primary role of MHC molecules is to take small samples from these large collections of peptides and present these peptides on the cell membrane. In response to the presentation of foreign peptides on the cell membrane, the activity and proliferation of specialist T cells is triggered, which ultimately leads to killing of the infected cell. Thus, the presence of a foreign peptide bound to an MHC molecule signals the presence of a foreign pathogen inside the cell. After a while the proliferation of T cells stops and most T cells die. About 5-10 percent of the specific T cells survive and form a population of “memory cells” responsible for future protection.⁹

Figure IIa provides a schematic representation of an MHC molecule. The polymorphic pair on top of the molecule (α_1 and α_2) forms a groove to which peptides are bound. The specific binding properties of different MHC molecules determine which peptides are presented to the T cells and, as a result, which parts of a pathogen are recognized by the immune system. This means that different individuals with different MHC molecules sample different peptides from the same set of proteins to initiate an immunological response to an infection. This can explain differences in disease progression between individuals, as has been shown for HIV and malaria.

Figure IIb illustrates how T cells cooperate with MHC molecules to kill a virally infected cell. Once the cell has been infected, the MHC molecules on the cell surface present peptides, which can be recognized as foreign by specialist T cells. If a T cell recognizes a peptide-MHC complex it is able to respond effectively to the infection and to eventually kill the virally infected cell.

In humans, human leukocyte antigens (HLA) are encoded by genes located on chromosome 6. Six

⁶The adaptive immune system itself consists of B lymphocytes, which produce antibodies, and T lymphocytes. There are two major types of T lymphocytes: cytotoxic CD8 T lymphocytes, which kill pathogens, and helper CD4 T lymphocytes, which play a central role in the induction of B and T lymphocyte responses.

⁷Hansen and Sachs [1989] provide an elaborate overview study of the importance of MHC diversity to the study of immunology. Playfair and Bancroft [2004] review the study of infection and immunity at great length.

⁸For class I MHC molecules short means that these peptides exist of 8-11 amino acids only.

⁹T cells have similarities with B cells, which produce antibodies, but there are some important differences. While B cells can directly recognize extracellular pathogens, T cells only recognize peptides that are being presented by MHC molecules on the surface of cells. T cells can be thought of as hands reaching out to feel the surface of neighboring cells, with some fingers contacting the MHC molecules and others probing for the peptide bound to them. Pathogens that survive inside host cells are thus effectively protected against recognition by B cells, but the MHC molecules betray their presence to T cells.

different types of HLA molecules are distinguished: HLA-A, HLA-B, and HLA-C are class I MHC molecules and are found on the membrane of every body cell, while HLA-DP, HLA-DQ, and HLA-DR are class II MHC molecules and are present only on the membrane of certain cells of the immune system. Roughly speaking, HLA-A, HLA-B, and HLA-C are especially important for immune responses against viruses, while HLA-DP, HLA-DQ, and HLA-DR are mainly deployed to control bacterial infections. MHC class I molecules trigger so-called cytotoxic CD8 T lymphocytes, while MHC class II molecules induce CD4 T helper responses.¹⁰

The MHC is known for its remarkable degree of polymorphism. In mammals MHC genes are the most polymorphic genes present, so that within a population many different alleles per MHC locus exist.¹¹ Since MHC molecules fulfil a crucial role in the host's immune response to pathogens, there is strong evolutionary selection of MHC alleles. If one specific MHC molecule outperforms the others, this molecule is expected to become dominant in the population. For this reason, the large degree of observed polymorphism points to a frequency-dependent selection process of MHC molecules in populations. Since every agent has two sets of chromosomes, maximally two different MHC molecules at each MHC locus are present (one inherited from the mother and one from the father). Due to the high degree of MHC polymorphism in the population, there is a high probability of MHC heterozygosity (i.e., the presence of different alleles at the two chromosomes). Although the evolutionary advantage of MHC heterozygotes could explain the occurrence of polymorphism, a number of studies suggest a direct evolutionary advantage of rare alleles to be the main factor explaining the large degree of polymorphism of the MHC in populations.¹² Such a direct advantage of rare alleles is present, because pathogens tend to adapt to the most common MHC molecules in the host population.

Several empirical studies suggest that MHC polymorphism also directly affects mating processes.¹³ Fish, mice and other animals turn out to prefer sexual partners with dissimilar MHC molecules. In addition, embryos with a large MHC similarity with their mother face a higher risk of preferential abortion. Experiments have shown that humans prefer the body odor of potential partners embodying a dissimilar

¹⁰Cytotoxic T cells are equipped to kill cells harboring intracellular pathogens such as viruses, while helpers cells are designed to switch on the activity of defense cells.

¹¹For this reason it is common among anthropologists to use information about the frequency of different MHC molecules to trace ethnic origins of populations and to study the cultural mechanisms that can help explain why genes and culture may coincide [Durham 1992]. One allele typically dominates within a population, but rare variants are known to survive easily. The existence of a specific allele that is rare within one population but dominant in another reveals ethnic bonds between both populations in the past. Cavalli-Sforza et al. [1994] have contributed to the understanding of the ethnic origins of different groups and populations, also revealing similarities with the diffusion of linguistic elements over the world. Hewlett et al. [2002] have used genes and language as tools for interpreting the genetic diversity among a large number of African ethnic groups. We investigate the diversity of alleles per locus within a population, leaving aside possible ethnic traces revealed by specific or individual MHC alleles.

¹² See e.g., De Boer et al. [2004] and Borghans et al. [2004] who argue that the MHC is polymorphic as a consequence of frequency dependent selection due to host-pathogen co-evolution, and not merely because of MHC heterozygote advantage. In addition, MHC polymorphism allows hosts to draw small and different samples from a pathogen's genome to induce an immune response.

¹³Penn [2002] provides an overview study addressing the relationship between sexual selection and MHC diversity.

MHC genotype, which increases the degree of polymorphism in a population.¹⁴ Both phenomena support the hypothesis that MHC diversity is important for evolutionary success.

Since class I MHC molecules are expressed by virtually all body cells, HLA alleles in humans can be traced by analysis of blood samples. There are two types of tests that are frequently used: serological tests, typing the specific HLA molecules expressed by a cell, and DNA tests, reading the structure of the genes that encode the HLA molecules. DNA typing is more expensive, but has the advantage that no fresh blood is needed, and that its resolution can be higher than that of serological tests. As a result of new DNA techniques the knowledge about subtle differences between HLA molecules is still increasing. At present there are still many questions regarding the specific role of HLA-A versus HLA-B.¹⁵

Originally, using serological tests, both HLA-A and HLA-B alleles have been classified according to a two-digit system, distinguishing 19 different HLA-A alleles varying from A01 to A43, and 42 HLA-B alleles, ranging from B07 to B78. More refined DNA methods enabled researchers to find differences within many two-digit groups. For that reason the classification has changed to a four-digit level. The first two digits indicate the old classification, while the last two digits include information about the particular subtype. At the same time, researchers recognized that many subtypes are actually almost functionally identical in the sense that they share similar peptide binding properties. Based on these binding properties Sette and Sidney [1999] and Lund et al. [2004] propose new groups of HLA “supertypes” of molecules that are functionally similar. For HLA-A these supertypes coincide to a large extent with the original two-digit grouping, but for HLA-B many molecules associated with different two-digit groups turn out to be functionally very similar. The knowledge about HLA-C is more limited, the available frequency tables contain many unknowns, and the number of populations for which HLA-C is available is relatively low. With respect to class II molecules, information for a representative sample of countries is currently not available.

2.2 HLA Data

Appendix A.I provides details about the data and its sources. Here we discuss only a number of the most salient details. We have collected data from several sources about the frequency of HLA molecules in different populations. We have selected all known (two-digit) HLA-A and HLA-B alleles within many different populations, leaving HLA-C and class-II molecules aside.

Our primary data source has been the collection of blood samples following the Eleventh International

¹⁴ See e.g., Wedekind and Fürti [1997] and Milinski and Wedekind [2001]. Reusch et al. [2001] provide interesting evidence on the stickleback mating behavior, which seems to be driven by odor as well.

¹⁵ Kiepiela et al. [2004] provide some evidence that HLA-B is dominant in the co-evolutionary process of the immune system and pathogens.

Histocompatibility Workshop and Conference. At this conference the results of identical HLA typing for a large number of ethnic groups were presented. In this way consistent information about allele frequencies within many populations could be identified. The total number of blood samples used to construct our database from this source equals 10,394. We use several secondary data sources to test the consistency of the data and to add a number of populations to our sample. For some populations HLA typing is obtained from regular tests because of tissue, organ or stem cell donations.¹⁶ To avoid selectivity and non-random sampling, only genetic information from healthy agents has been used, including only one observation when information of several siblings or parents within one family is present in the data.

The data on ethnic groups have been matched to countries to obtain information about macroeconomic indicators, such as GDP per capita. In some cases the link between ethnic group and countries is not trivial. In these cases we matched economic data of a country with information about HLA of the economic dominant ethnic group in that country. For example, for the United States we have used the whites (Caucasian). For other countries such a link is not feasible. Furthermore, for some ethnic groups no economic indicators were present (e.g., the Papuans), whereas for some countries no HLA information was available. This reduced the sample of countries – and consequently the number of observations for the empirical analysis – to 63.

2.3 Frequency-Dependence and Evolution of MHC Diversity

The large degree of MHC polymorphism points to a process of frequency-dependent selection with respect to these molecules. Interestingly, the observed frequency distribution of different MHC molecules within a population does not appear to be a uniform distribution. Figure III shows the average frequency distribution of the 19 different HLA-A molecules and 42 HLA-B molecules in a population. Because different MHC molecules are dominant in different populations, we rank-ordered the frequencies before taking averages over all populations. The graph in the first panel illustrates that in most populations there is one HLA-A molecule with a high frequency, of on average 26.9 percent. The average frequency of the second most-frequent MHC molecule equals 17.0 percent. The relatively high presence of some molecules is more prominent for HLA-A than for HLA-B which is most likely due to the fact that there are 42 B-molecules compared to only 19 A-molecules, and because of the disagreement in the classification of HLA-B molecules, which only recently advanced to the formation of so-called supertypes (e.g., Lund

¹⁶Originally, the name MHC was given to a large genetic region containing genes that determine the success or failure of graft transplantation. If two agents share the same set of MHC molecules, transplantation has the largest chance of being successful. The larger the difference, the higher the probability of rejection of the transplant.

et al. [2004]).¹⁷ Another interesting observation is that the dominant allele differs between populations, which suggests that, along with an individual selective pressure for a large variety of rare MHC alleles, there seems to be a selective pressure at work favoring a number of MHC alleles within populations.

As a measure of diversity we use $1 - \sum_k F_k^2$ where F_k is the frequency of type k in the population. This measure is known in genetics as the degree of heterozygosity (see Cavalli-Sforza et al. [1994]) and as the Herfindahl index in economics, and used here to measure diversity instead of its usual application to measure market concentration.¹⁸ The index equals $1 -$ the probability that two persons in a population share the same molecule.¹⁹ When applied to small samples, a plain calculation of the sample diversity underestimates the diversity within the whole population. For that reason we corrected the estimator by subtracting the expected bias ($F_k(1 - F_k)/n$) from each component F_k^2 . Details about this approach can be found in Appendix A.II. Simulation of the resulting measure of diversity suggests that for the smallest samples in our set ($n = 100$) the bias is reduced from approximately 0.010 (28.6 percent of a standard deviation) to 0.0005 (1.4 percent of one standard deviation) on a diversity scale ranging from 0.700 to 0.900 (with a standard deviation equal to 0.035).

Figure IV shows how the frequency of the five most frequent HLA-A and HLA-B molecules in the 63 populations is related to the measure of diversity. The horizontal axis is defined as the level of diversity within a population and the vertical axis shows the frequency of appearance of particular molecules maximized at 1. It is interesting to observe that there exists a consistent relationship between the overall level of MHC diversity and the specific form of the distribution within populations. Especially the frequency of the most frequent MHC molecule is large when MHC diversity is low. The frequency of the second molecule also increases as MHC diversity decreases, but at a much lower pace. The frequency of the fifth molecule remains almost constant between a level of diversity of 0.754 (0.873) and 0.929 (0.959) for HLA-A (HLA-B) molecules but decreases when diversity falls below this level. Again the observed pattern is more pronounced for HLA-A molecules than it is for HLA-B molecules, which is consistent with the number presented in Figure III.

A theoretical link between the economic and biological domain only materializes when the time horizon of both processes are comparable. This implies that economic circumstances have to remain relatively stable over some period of time for an evolutionary equilibrium to emerge. Evolution of the diversity of MHC alleles within a population can be relatively rapid for two reasons. First, there is a large evolutionary

¹⁷On average the most frequent HLA-B molecule has a frequency of 14.8 percent and the second most frequent B-molecule has a frequency of 11.9 percent.

¹⁸We have considered the robustness of all our results carefully by using different powers in this index: the results are not affected.

¹⁹As a measure of polymorphism the “degree of heterozygosity” seems to be a misnomer. Based on Mendelian expectations, both measures of diversity will be identical however. Since there are reasons why the Mendelian rules do not hold for MHC (e.g., as a result of disassortative mating and preferential abortion), this measure reflects the degree of polymorphism rather than the degree of heterozygosity.

pressure on the MHC. Second, changes in diversity require only changes in MHC frequencies. A change in a distribution, such as presented in Figure III, requires only the characteristics of a fraction of the population to change. Importantly, the characteristics that need to increase in terms of frequency can already be present at high frequency in a population before the evolutionary change starts. We will show that a sufficient change in MHC diversity within a population can occur in 4 to 10 generations. In contrast, for many other evolutionary changes a new and improved gene has to occur through mutation and suppress the previously dominant one, which (if successful at all) takes much longer. For example, Aoki [1991] shows that lactose tolerance among Europeans increased from 5 to 70 percent in a period of about 5,000 years. Similarly, the establishment of the sickle cell gene that protects African farmers against malaria took around 10,000 years.

2.4 Transmission of Pathogens and Immune Responses

Infectious diseases spread in a number of ways between individual hosts. Most skin infections spread by direct skin contact (e.g., pox) or via the water (e.g., hookworm), while most respiratory infections spread by coughing and sneezing, and most intestinal infections by faecal and diarrhoea. Other pathogens spread as a result of sexual contacts (e.g., herpes, HIV and syphilis) and via blood or blood products (e.g., hepatitis B and C and malaria). Finally, insect and animal bites spread pathogens causing for example yellow fever, malaria and rabies.

For some pathogens it is more easy to spread than for others.²⁰ A living agent transmitting a pathogen to another living agent is defined as a vector. When a pathogen spreads through a population some hosts may contract the disease while others do not. These differences between hosts are to a large extent caused by differences in the HLA background. Some hosts manage to present a larger variety of pathogen peptides to their immune system than others, or happen to present peptides that the pathogen cannot mutate. Hosts may even get infected, transmit the virus, but fail to contract the disease itself because of immune control.²¹ The other infected hosts might contract the disease because of less effective immune responses. These cases can occur because the time between infection and the induction of an effective immune response can be used by the pathogen to infect another host.

Transmission of infectious diseases explains why agents with an adaptive immune system that differs from neighboring agents have a selective advantage. If this would be the only factor, the distribution

²⁰In other words, the degree of contact between the host and the pathogen matters. This depends on the numbers of pathogens in the environment and the means by which they typically spread.

²¹An interesting example is the remarkably low rate of HIV disease progression among some HIV-infected prostitutes in Africa who share certain MHC molecules. These MHC molecules seem to induce a much more effective T cell response to HIV than is currently observed [Trachtenberg et al. 2003].

of MHC molecules within populations would be expected to be uniform to ensure a maximum level of MHC diversity between hosts. The non-uniform shape of the frequency-dependent distribution of the MHC described above thus suggests that there are additional forces influencing MHC diversity within populations.

A prominent reason for the observed MHC frequency distribution to be nonuniform is that properties of the immune system of one agent affect the health outcomes of other agents living in the same population. If a host becomes infected but the adaptive immune system deals with the infection (either through existing memory cells or by an effective response of T cells), in the sense that the agent will not act as a vector to infect others, this agent reduces the incidence of the disease in the population. In addition, the prevalence and reproduction rate of this infection in a particular environment will be lower if the proportion of agents with an adequate immune response in the population is higher.²² This is especially true for endemic diseases and for periodically returning (modified) pathogens with relatively high reproduction rates, such as the influenza virus. Of course, if hosts deal ineffectively with infections and act as a vector of the disease, the incidence in the population will be higher.

One specific agent in a group can be infected by all others. The immune responses of the latter therefore act as a protecting belt to him. If every agent has different MHC molecules, which are effective against different pathogens, every pathogen has ample opportunities to get through this protective belt, using the weakest link as host. For this specific agent it is therefore optimal when the others share the same MHC molecules. Conditional on the MHC molecules of the other agents, it is advantageous to have a different MHC. When a new pathogen breaks through this wall, this provides the best opportunities for an effective immune response. The trade-off is that every agent who is better off in being different from the others is also part of the protecting belt of others, in which a lower level of MHC diversity is optimal. The optimal distribution of MHC molecules for the group depends on the relative magnitude of both effects. A selfish agent will however only optimize its own benefits. Being similar is beneficial to the others, while being different is beneficial for himself, so a prisoners dilemma occurs, that will lead to a higher level of MHC diversity than is optimal for the group as a whole.

In economic terms this means that there exists an externality in the immune responses of single agents to build resistance at the level of the population. As a result, there is no reason to assume an equilibrium in which a maximum level of MHC diversity is optimal. In case of a discrepancy between the individual equilibrium and the population's (social) optimum, any change in the degree to which agents internalize this externality affects the frequency distribution of MHC molecules in that population. For populations

²²The reproduction rate refers to the average number of susceptible agents infected by a single host; the R_0 in epidemiology.

in which mutual dependence between agents in terms of production and economic outcomes is more important, the social optimum with lower levels of MHC diversity and a high prevalence of specific types of MHC is more likely to be the observed immune system outcome, compared to populations in which mutual dependence is lower.

3 Model

3.1 Basic Structure

Consider a population with generations of N_g agents. In each generation random and small groups of agents of size $n + 1$ are created. These small groups can cooperate to produce, but agents can also decide to work on their own. Every agent has n randomly selected contemporary agents in close vicinity. Each period, agents have one unit of time available to work. After having contracted a disease a fraction $1 - t_i$ is needed to recover, so a fraction $t_i \leq 1$ of this time can be used to produce. Following the structure of a linear public goods game, each agent can decide to spend a fraction of time ($\bar{\gamma}_i$) to a joint project and $(1 - \bar{\gamma}_i)$ to solo activities. Total output of a cooperative project equals $Q \sum_j \bar{\gamma}_j t_j$, where Q is the return to the public good investments. Output will be shared among the agents, so every agent receives $\frac{Q}{n+1} \sum_j \bar{\gamma}_j t_j$. The marginal per capita return equals $\frac{Q}{n+1}$. The output of a solo project equals $(1 - \bar{\gamma}_i)t_i$.

When $Q < 1$ cooperation is not beneficial and when $Q > n + 1$, agents voluntarily cooperate completely (in a Nash equilibrium): $\bar{\gamma}_i = t_i$. The interesting case is when $1 < Q < n + 1$, where the Nash equilibrium induces solo activities only ($\bar{\gamma}_i = 0$), since $\frac{d(1-\bar{\gamma}_i)}{d\bar{\gamma}_i} > \frac{dQ\bar{\gamma}_i/n+1}{d\bar{\gamma}_i}$. In cases where individual optimization does not lead to social optimality, rules, altruism, social pressure or reciprocal punishment can lead to higher levels of cooperation.

We assume that groups of agents expect a target level of cooperation $\bar{\gamma}$. Agents who put less effort in this joint project are punished, while those who invest more are rewarded. The punishment function is defined as $\rho t_i \left(e^{\frac{\bar{\gamma}-\bar{\gamma}_i}{\bar{\gamma}}} - 1 \right)$, with $\rho \geq 0$. Now, the income of agent i equals

$$Y_i = (1 - \bar{\gamma}_i)t_i + \sum_j \frac{Q}{n+1} \bar{\gamma}_j t_j - \rho t_i \left(e^{\frac{\bar{\gamma}-\bar{\gamma}_i}{\bar{\gamma}}} - 1 \right), \quad (1)$$

where the first term on the left-hand side is solo production, the second term joint production and the final term punishment. The optimal value of $\bar{\gamma}_i$ can be derived from the first order condition and equals

$$\frac{dY_i}{d\bar{\gamma}_i} = t_i \left(-1 + \frac{Q}{n+1} + \frac{\rho}{\bar{\gamma}} t_i \left(e^{\frac{\bar{\gamma}-\bar{\gamma}_i}{\bar{\gamma}}} - 1 \right) \right), \quad (2)$$

which implies that

$$\frac{\bar{\gamma} - \bar{\gamma}_i}{\bar{\gamma}} = \ln \left(\frac{\frac{n+1-Q}{n+1} \bar{\gamma}}{\rho} \right). \quad (3)$$

So every target level of cooperation ($\bar{\gamma}$) is related to actual cooperative behavior ($\bar{\gamma}_i$). We impose that in every population a long-run equilibrium of the social norm will be settled in which the target value of cooperation is equal to the actual degree of cooperation. So the socially expected behavior equals the usual behavior of the agents, i.e.,

$$\bar{\gamma} = \rho \frac{n+1}{n+1-Q}, \quad (4)$$

which reveals that $\bar{\gamma}$ is increasing with the level of punishment and with Q , and decreasing with the size of the group $n+1$. Agent i 's income in this equilibrium then equals

$$Y_i = \left(\frac{(n+1)(1-\bar{\gamma}) + Q\bar{\gamma}}{n+1} \right) t_i + \sum_{j \neq i} \left(\frac{1}{n+1} Q\bar{\gamma} \right) t_j, \quad (5)$$

which exists of an individual and a group component (of course agent i contributes to the group income too). Using the symmetry of the model it can be derived that the expected income per worker equals

$$E(Y_i) = ((1-\bar{\gamma}) + Q\bar{\gamma})E(t_i), \quad (6)$$

which shows that income rises from t_i to Qt_i when $\bar{\gamma}$ shifts from 0 to 1. If the variation between countries in terms of $E(t_i)$ is smaller relative to the variation in Q , differences in cooperation between individuals are the main determinant of income.

Crucial for the model, however, is that differences in the degree of cooperation change evolutionary pressure regarding the immune system. Defining

$$\gamma = \frac{Q\bar{\gamma}}{\frac{(n+1)(1-\bar{\gamma}) + Q\bar{\gamma}}{n+1}} \quad (7)$$

to normalize agent i 's own contribution to one the participation by other agents contributes with weight γ to the performance of i . Consequently, the overall income of i depends more on the health status of others, when cooperation $\bar{\gamma}$ increases. This result shows that cooperation leads to mutual dependency among agents cooperating in small groups.

Assuming that the evolutionary success of i (ES_i) depends positively on his economic performance (as lower output lowers the revenues and subsequently lowers the possibilities to use resources for reproduction and health), evolutionary success is negatively affected by the fraction of time he is recovering from a

disease, but also by the time the other agents are recovering. More formally, evolutionary success of i is a function of individual and group fitness: $ES_i = f(1 - t_i + \gamma \sum_{j \neq i} 1 - t_j)$, with $f' > 0$. When f can be approximated by a linear function, the expected value of the evolutionary pressure equals by approximation

$$E(ES_i) = f(1 - E(t_i) + \gamma n(1 - E(t_j))). \quad (8)$$

An evolutionary stable equilibrium has to satisfy the equilibrium condition and the stability condition.²³ The equilibrium condition implies that $ES_i = C$ ($C > 0$) for all genetic characteristics of i with a positive frequency in the population, while the stability condition implies that the evolutionary success of every characteristic with zero probability is less or equal to this level ($ES_i \leq C$).

We assume m different MHC molecules (indexed k or l) and m types of pathogens (indexed v or w), which can cause an infectious disease. MHC molecules differ in the effectiveness of their immune response to a certain type of pathogen when *contracting* it, and when *transmitting* it. Each pathogen is assumed to have equal probability to be active in a certain (sub)period. When the pathogen arrives each of the n agents have equal probability (α) of direct infection. When a person is infected, there is a probability \bar{p}_{kv} of transmission to all the others in his group. This probability of transmission depends on the combination of the pathogen (v), and the MHC molecules (k) of the infected donor. Combined with the initial probability of infection $p_{kv} = \alpha \bar{p}_{kv}$ describes the probability that a host infects others. Once infected, a host contracts the disease with probability q_{kv} . We distinguish between \bar{p}_{kv} and q_{kv} because hosts can transmit without contracting the disease. There are two interesting specific cases. When $\bar{p}_{kv} = q_{kv}$ everyone who transmits contracts the disease and visa versa. When $q_{kv} = 1$, everyone infected contracts the disease. In case of contracting a disease, production stops for a fixed period of time, until recovery.

For our purposes the most interesting case is when each MHC molecule has an advantage to tackle a specific pathogen, without any other difference in performance. The scheme below gives the values of p_{kv} and q_{kv} for this case. The scheme shows that MHC 1 is good for fighting disease 1 and so on. We use this scheme to illustrate the results of the model and refer to it as the symmetric case. In addition, without loss of generality, we assume that $p_{high} > p_{low}$ and $q_{high} > q_{low}$.

	Transmission			Disease		
	Disease 1	Disease 2	Disease 3	Disease 1	Disease 2	Disease 3
MHC 1	$p_{11} = p_{low}$	$p_{12} = p_{high}$	$p_{13} = p_{high}$	$q_{11} = q_{low}$	$q_{12} = q_{high}$	$q_{13} = q_{high}$
MHC 2	$p_{21} = p_{high}$	$p_{22} = p_{low}$	$p_{23} = p_{high}$	$q_{21} = q_{high}$	$q_{22} = q_{low}$	$q_{23} = q_{high}$
MHC 3	$p_{31} = p_{high}$	$p_{32} = p_{high}$	$p_{33} = p_{low}$	$q_{31} = q_{high}$	$q_{32} = q_{high}$	$q_{33} = q_{low}$

²³See e.g., Maynard Smith and Price [1973] and Van Damme [1991] for definitions of these equilibria.

The probability that someone with MHC type k contracts disease type v equals the probability of transmission from outside the cooperative group (α) of at least one other agent in the group, multiplied by the probability of actually contracting the disease when infected, i.e.

$$P(k, v) = \left(1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}\right) q_{kv}, \quad (9)$$

where π_k is the fraction in the population with MHC type k , with $\sum_k \pi_k = 1$.

A social planner, who aims to minimize the number of infections in a population and who could set the distribution of MHC molecules in a population, would want to maximize income Y , which is equivalent to minimizing the number of diseases (S):

$$S = \sum_v \sum_k P(k, v) = \sum_v \left(\left(1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}\right) N_g \sum_k (\pi_k q_{kv}) \right). \quad (10)$$

3.2 Evolutionary Equilibria

Individual evolutionary selection increases the frequency of MHC molecules that lead to higher evolutionary success. This implies that in an evolutionary stable equilibrium either the evolutionary success of molecule k equals the evolutionary success of molecule \bar{k} , or that the molecule with lower fate has frequency 0.

Assume $\gamma = 0$. For each combination of molecules k and \bar{k} with non-zero frequency, the equilibrium condition for mixed distributions equals

$$\sum_v \left(1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}\right) q_{kv} = \sum_v \left(1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}\right) q_{\bar{k}v}. \quad (11)$$

In the case where $\gamma > 0$, individual success also depends on the health of the other n agents. The expected number of agents who contract the disease upon exposure from virus v when agent i expresses MHC type k , equals

$$\left(1 - (1 - \alpha)(1 - p_{kv})^{(n-1)\pi_k+1} \prod_{l \neq k} (1 - p_{lv})^{(n-1)\pi_l}\right) n \sum_l \pi_l q_{lv}. \quad (12)$$

The condition for an evolutionary stable equilibrium, for each combination of k and \bar{k} with non-zero

frequency, equals

$$\begin{aligned}
& \sum_v (1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}) q_{kv} + \\
& \gamma \left(1 - (1 - \alpha) (1 - p_{kv})^{(n-1)\pi_k+1} \prod_{l \neq k} (1 - p_{lv})^{(n-1)\pi_l} \right) n \sum_l \pi_l q_{lv} \\
& = \\
& \sum_v (1 - (1 - \alpha) \prod_l (1 - p_{lv})^{n\pi_l}) q_{\bar{k}v} + \\
& \gamma \left(1 - (1 - \alpha) (1 - p_{\bar{k}v})^{(n-1)\pi_{\bar{k}}+1} \prod_{l \neq \bar{k}} (1 - p_{lv})^{(n-1)\pi_l} \right) n \sum_l \pi_l q_{lv}.
\end{aligned} \tag{13}$$

In the symmetric case three results can be derived from the social welfare function (10) and the two conditions (11) and (13). We derive and proof these results in three theorems below.

Theorem 1 (Evolutionary Equilibrium without Cooperation): *In the symmetric case when evolutionary selection depends only on individual health (i.e., $\gamma = 0$), and $p_{high} \neq p_{low}$ and $q_{high} \neq q_{low}$, the only evolutionary stable equilibrium equals $\pi_k = 1/m$ for every k*

Proof: Using the symmetric parametrization equation (11) can be written as

$$\begin{aligned}
& \left(1 - (1 - \alpha) (1 - p_{low})^{n\pi_k} (1 - p_{high})^{n\pi_{\bar{k}}} (1 - p_{high})^{n(1-\pi_k-\pi_{\bar{k}})} \right) (q_{low} - q_{high}) + \\
& \left(1 - (1 - \alpha) (1 - p_{high})^{n\pi_k} (1 - p_{low})^{n\pi_{\bar{k}}} (1 - p_{high})^{n(1-\pi_k-\pi_{\bar{k}})} \right) (q_{high} - q_{low}) = 0
\end{aligned} \tag{14}$$

Because $q_{high} \neq q_{low}$ this can be rewritten as

$$\left(\frac{1 - p_{low}}{1 - p_{high}} \right)^{n(\pi_k - \pi_{\bar{k}})} = 1. \tag{15}$$

Now $p_{high} \neq p_{low}$ yields $\pi_k = \pi_{\bar{k}}$, and thus $\pi_k = 1/m$. It is easily verified that the second order condition in this equilibrium holds.

QED

The most interesting result of Theorem 1 is that when evolutionary success depends on individual health only, the distribution of MHC molecules in the population in an evolutionary stable equilibrium is always uniform and thus fully diverse. This observation, however, seems inconsistent with the observed distribution of MHC molecules in populations described in Section 2 and shown in Figures III and IV above.

Theorem 2 (Difference Between Social Optimum and Evolutionary Stable Equilibrium without Cooperation): π_k , for all k , is not a socially optimal distribution of MHC molecules, if $n \log \left(\frac{1-p_{low}}{1-p_{high}} \right) >$

$$2 \frac{q_{high} - q_{low}}{\frac{1}{m} q_{high} + \frac{m-1}{m} q_{low}}.$$

Proof: Consider the situation in which the frequency of MHC molecule k , π_k , can be increased at the expense of molecule \bar{k} , with frequency $\pi_{\bar{k}}$, keeping the frequency of the other MHC molecules constant. Denoting $\bar{\pi} = \sum_{l \neq k, \bar{k}} \pi_l$, implies that $\pi_{\bar{k}} = \bar{\pi} - \pi_k$. The social welfare function (10) then reads

$$\begin{aligned} S = & \left(1 - (1 - \alpha)(1 - p_{low})^{n\pi_k}(1 - p_{high})^{n(\bar{\pi} - \pi_k)} \prod_{l \neq k, \bar{k}} (1 - p_{low})^{n\pi_l}\right) \\ & N_g(\pi_k q_{low} + (1 - \pi_k) q_{high}) \\ & + \left(1 - (1 - \alpha)(1 - p_{high})^{n\pi_k}(1 - p_{low})^{n(\bar{\pi} - \pi_k)} \prod_{l \neq k, \bar{k}} (1 - p_{low})^{n\pi_l}\right) \\ & N_g((1 - \bar{\pi} - \pi_k) q_{low} + (\bar{\pi} + \pi_k) q_{high}) \\ & + \sum_{v \neq k, \bar{k}} \left(1 - (1 - \alpha)(1 - p_{low})^{n(1 - \bar{\pi})} \prod_{l \neq k, \bar{k}} (1 - p_{low})^{n\pi_l}\right) N_g(\pi_k q_{low} + (1 - \pi_k) q_{high}). \end{aligned} \quad (16)$$

It is easily verified that the derivative with respect to π_k of this function, when $\pi_l = 1/m$ for all l equals 0. Therefore, the social welfare function always has an optimum in the distribution $(1/m, 1/m, \dots, 1/m)$ of MHC molecules in the symmetric case.

This optimum is not a minimum when the second derivative of the welfare function is negative in $(1/m, 1/m, \dots, 1/m)$. The second derivative in this point equals:

$$\begin{aligned} \left. \frac{d^2 S}{d\pi_k^2} \right|_{(1/m, 1/m, \dots, 1/m)} &= 2m(q_{high} - q_{low}) + \\ n(q_{low} + (m-1)q_{high}) \log(1 - p_{high}) &- n(q_{low} + (m-1)q_{high}) \log(1 - p_{low}) \end{aligned} \quad (17)$$

This expression is negative when

$$n \log \left(\frac{1 - p_{low}}{1 - p_{high}} \right) > 2 \frac{q_{high} - q_{low}}{\frac{1}{m} q_{high} + \frac{m-1}{m} q_{low}}. \quad (18)$$

QED

Theorem 2 shows that for certain parameter values of the model the social optimum is equal to the evolutionary stable equilibrium, but for other parameters the social optimum is a different one. In the latter case there is a discrepancy between the outcome of the individual evolutionary selection process and the social optimum. This discrepancy will occur when n is sufficiently large, $p_{high} - p_{low}$ is sufficiently large, or $q_{high} - q_{low}$ is sufficiently low. Increasing m also shifts the balance in the advantage of a discrepancy between the social optimum and the individual equilibrium.

Theorem 3 (Evolutionary Stable Equilibrium with Cooperation): *If evolutionary selection depends on individual and co-worker health (i.e., $\gamma > 0$), there is a value of γ for which the evolutionary*

stable equilibrium equals the social optimum. When p_{high} and p_{low} are small, the value of γ for which this equality holds, is close to unity.

Proof: We need to show that for a certain value of γ the condition for an evolutionary stable equilibrium equals the first-order condition for a social optimum. Again, as in (16), consider the situation in which in the social welfare function the frequency of k is changed at the cost of \bar{k} . In an optimum the derivative of the welfare function with respect to π_k should be 0:

$$\begin{aligned}
& \left(1 - (1 - \alpha)(1 - p_{kk})^{n\pi_k}(1 - p_{\bar{k}k})^{n\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{lk})^{n\pi_l}\right)(q_{kk} - q_{\bar{k}k}) \\
& + \left(1 - (1 - \alpha)(1 - p_{k\bar{k}})^{n\pi_k}(1 - p_{\bar{k}\bar{k}})^{n\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{l\bar{k}})^{n\pi_l}\right)(q_{k\bar{k}} - q_{\bar{k}\bar{k}}) \\
& \quad + \gamma(p_{kk} - p_{\bar{k}k}) \\
& \left(1 - (1 - \alpha)(1 - p_{kk})^{(n-1)\pi_k}(1 - p_{\bar{k}k})^{(n-1)\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{lk})^{(n-1)\pi_l}\right)n \sum_l \pi_l q_{lk} \\
& \quad + \gamma(p_{k\bar{k}} - p_{\bar{k}\bar{k}}) \\
& \left(1 - (1 - \alpha)(1 - p_{k\bar{k}})^{(n-1)\pi_k}(1 - p_{\bar{k}\bar{k}})^{(n-1)\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{l\bar{k}})^{(n-1)\pi_l}\right)n \sum_l \pi_l q_{l\bar{k}} = 0.
\end{aligned} \tag{19}$$

The first-order condition for a social optimum equals:

$$\begin{aligned}
& \left(1 - (1 - \alpha)(1 - p_{kk})^{n\pi_k}(1 - p_{\bar{k}k})^{n\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{lk})^{n\pi_l}\right)(q_{kk} - q_{\bar{k}k}) \\
& + \left(1 - (1 - \alpha)(1 - p_{k\bar{k}})^{n\pi_k}(1 - p_{\bar{k}\bar{k}})^{n\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{l\bar{k}})^{n\pi_l}\right)(q_{k\bar{k}} - q_{\bar{k}\bar{k}}) \\
& \quad + \gamma(\log(1 - p_{kk}) - \log(1 - p_{\bar{k}k})) \\
& \left(1 - (1 - \alpha)(1 - p_{kk})^{(n-1)\pi_k}(1 - p_{\bar{k}k})^{(n-1)\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{lk})^{(n-1)\pi_l}\right)n \sum_l \pi_l q_{lk} \\
& \quad + \gamma(\log(1 - p_{k\bar{k}}) - \log(1 - p_{\bar{k}\bar{k}})) \\
& \left(1 - (1 - \alpha)(1 - p_{k\bar{k}})^{(n-1)\pi_k}(1 - p_{\bar{k}\bar{k}})^{(n-1)\pi_{\bar{k}}}\prod_{l \neq k, \bar{k}}(1 - p_{l\bar{k}})^{(n-1)\pi_l}\right)n \sum_l \pi_l q_{l\bar{k}} = 0.
\end{aligned} \tag{20}$$

In the symmetric case these conditions are equal when

$$\gamma = \frac{(1 - p_{low})^{\pi_A k}(1 - p_{high})^{\pi_{\bar{k}}}}{p_{high} - p_{low}}(\log(1 - p_{high}) - \log(1 - p_{low})). \tag{21}$$

Suppose now that $p_{low} = \alpha p_{high}$. Then

$$\lim_{p_{high} \rightarrow 0} \frac{(1 - p_{low})^{\pi_k}(1 - p_{high})^{\pi_{\bar{k}}}}{p_{high} - p_{low}}(\log(1 - p_{high}) - \log(1 - p_{low})) = 1. \tag{22}$$

QED

The core message of Theorems (1-3) is that when the social optimum differs from the frequency distribution of MHC molecules that results from individual selection, the evolutionary stable equilibrium gets closer to the social optimum when γ – the degree to which evolutionary success of one agent depends on the

health of the other n agents – increases. Figure Va provides the frequency distribution in an evolutionary stable equilibrium for different values of γ when $m = 2$. When $\gamma = 0$ the distribution is $1/2, 1/2$. When γ rises, the distribution changes gradually to $0, 1$ (or $1, 0$).

Figure Vb illustrates the results for the model with three types of MHC molecules. The simulation results replicate the typical patterns in the frequency distribution of HLA-A and HLA-B shown in Figure III in Section 2. In equilibrium the frequencies of MHC molecules is not uniform, but each molecule has a different frequency. Furthermore, with decreasing levels of diversity (i.e., when $\gamma \rightarrow 1$) one allele becomes more dominant and the frequency of others decreases, a feature observed in the distribution of molecules in the populations we have collected data for.

Taken together, the model shows that under certain circumstances the health status of a population is better off with low levels of MHC diversity, whereas individual selection processes tend to increase diversity. In more cooperative populations, evolutionary pressure shifts the immune system from the selfish one with high levels of diversity to the more efficient social optimum with lower levels of MHC diversity. Consequently, the level of MHC diversity mirrors a population's level of cooperative behavior. This finding is able to explain the typical non-uniform pattern in HLA-A and HLA-B frequencies observed in Figures III and IV. Moreover, the model puts forward three other predictions. First, if there is indeed a discrepancy between the social optimum and the selfish equilibrium, low levels of MHC diversity in a population should be associated with better health outcomes. Second, when such differences between countries reflect differences in the way people cooperate, MHC diversity should be negatively related to indicators of cooperative behavior. Finally, this mechanism only works when cooperative behavior has a higher output relative to carrying out solo projects. Hence, economic performance should correlate negatively with MHC diversity.

3.3 The Pace of Change of the Evolutionary Process

Apart from equilibrium properties of the model, it is important to get a feeling for some of the most important dynamic properties of the evolutionary process. A crucial requirement for any satisfactory explanation of a relationship between economic circumstances and evolutionary outcomes is that the time scale of the evolutionary adjustment has to be relatively short. Many important evolutionary changes have taken at least several thousands of years, and most important genetic changes took even many millions of years to develop, and it is hard to imagine that the same economic circumstances have ruled during such

a long period of time.²⁴

To analyze the pace of adjustment we simulated the model for a certain value of γ until the frequency distribution converged and then changed γ . Table I shows what happens in a case with four different MHC molecules, for which we change the value of γ such that the equilibrium diversity shifts from 0.700 to 0.600. Column (1) and (2) illustrate the equilibrium frequency distributions for both situations. In an equilibrium every MHC molecule has the same fitness, i.e. two agents who differ in their MHC molecule must have the same expected number of children. When changing γ in a situation in which an equilibrium confers, the fitness of each molecule changes. Column (4) provides the fitness of each MHC molecule immediately after the shift in γ . Since γ increases, the fitness of relatively frequent (rare) molecules increases (decreases). At present there is no evidence about the actual size of fitness differentials and we calibrated the simulation to obtain a 1.050 fitness parameter for the most successful allele. This seems a conservative estimate since it implies that people with MHC-1 have a five percent higher effective fertility rate than the average level of reproduction in this population. This is a very small difference in terms of actual offspring, which would most likely not be observed in reality for a reasonably long period of time. Under these assumptions it takes seven generations to bridge the half-life of the difference in diversity between the old and new situation. When people get offspring at age 20, this equals a time-span of 140 years.

The numbers in Table I illustrate why the pace of this evolutionary change is relatively high. In Column (3) we provide information about the frequency differences between both equilibria. Adding all increases or all decreases shows that only 17.62 percent of the population has to “change genes” to obtain the new equilibrium. Furthermore, the genes that need to increase in frequency are already present in the population and need not be developed. An increase of five percent in population size, increases the fraction of the population expressing MHC-1 with $0.050 * 35.980 = 1.799$ percentage points in the next generation.

To address the robustness of this process we carried out several simulations, including different parameter values. Generally, we find that four to ten generations suffice to reduce the MHC diversity in a population by 50 percent, which makes us confident that our measure of human evolution changes with changes in economic circumstances.

²⁴The currently most developed nations are the OECD member states, but only some 1,000 years ago the situation was fundamentally different. The Chinese society was flourishing while Europe and most of the other OECD member states were stuck at very low levels of economic development. At the same time the level of development in what are now the oil producing countries in Asia was far higher than in Europe.

4 Results

4.1 Health Outcomes

The first implication of our model is that a discrepancy between the individual stable evolutionary equilibrium and the social optimum, should be associated with a correlation between the MHC diversity in a population and its health outcomes. Regarding the central role of class I MHC molecules in the control of viral infections, there should especially be a lower incidence of viral infections in populations with a low MHC diversity.

To test the relationship between health outcomes and MHC diversity we use information about life expectancy to evaluate the effect of MHC diversity on health outcomes in general. The number of deaths caused by viral or bacterial infections as a fraction of the total population is used as a more detailed proxy for health outcomes. Life expectancy at birth is taken from the 2002 World Development Indicators, and refers to the situation in 1990.²⁵ Information on mortality is taken from the 2000 Mortality Database of the World Health Organization from which we have taken the data on infectious diseases and split this information into viral, bacterial and other infections. We apply HLA-A diversity as our measure of MHC diversity for reasons detailed in Section 2 above. The estimates using HLA-B diversity yield qualitatively similar results.

Figure VIa presents a first glance of the relationship between health outcomes and MHC diversity by showing a negative and statistically significant correlation between life expectancy at birth in 1990 and HLA-A diversity for the sample of 63 countries. The horizontal axis measures HLA-A diversity and the vertical axis life expectancy at birth in 1990. Table II provides the main results for using life expectancy at birth as the dependent variable. Column (1) shows the OLS relationship between MHC diversity and life expectancy at birth in 1990 by estimating the following model:

$$LE_{c1990} = C + \alpha_1 MHC_c + \epsilon_c, \quad (23)$$

where LE_{c1990} is life expectancy at birth in country c in 1990. The main variable of interest is MHC_c , the level of HLA-A diversity in country c . The parameter α_1 measures whether HLA-A diversity has an effect on life expectancy. ϵ_c is an error term with the usual properties, capturing all omitted factors. Column (1) shows a statistically significant correlation between the level of HLA-A diversity and life expectancy at birth. The estimate (standard error) of α_1 is $-57.797(26.943)$, which is significant at the five percent level. This coefficient suggests that a one-standard deviation (i.e., 0.035) increase in MHC diversity reduces

²⁵For 1990 the number of populations for which life expectancy is available is the largest. Using life expectancy at birth in 1995 or 2000 yields similar estimates but lacks information about a number of developing countries.

equilibrium life expectancy by 0.265 years. These seem to be estimates of reasonable magnitude relative to the mean (standard deviation) of life expectancy at birth in 1990 in the sample, which is 69.201 (7.691) years.

To check the robustness of these results we conducted analyses including covariates that could explain part of the observed correlation. Since lower MHC diversity mirrors mutual dependence and will be associated with higher levels of economic performance, this estimate might include both the direct effect of a better immune response and the indirect effect of MHC diversity on life expectancy working through income. Effective treatment and preventive measures are likely to lower the prevalence of infectious diseases in a population because they reduce the number of susceptible agents in a population by, for example, vaccination programs yielding immunization, lower the duration of the infectious state of the host, and the distance over which spread and transmission of the disease occurs. Furthermore, a more advanced health system could lower the optimal level of MHC diversity because the threat of pathogens will diminish. We therefore control for differences in health systems. To account for differences in the treatment of diseases and the incidence of preventive measures we use the log of health expenditures per capita in constant 1990 US\$ in 1995. In addition, we average two measures of child immunization to capture vaccination coverage of children under one year of age: (i) child immunization against diphtheria, pertussis (or whooping cough), and tetanus (DPT) after receiving three doses of vaccine; and (ii) child immunization against measles after receiving one dose of vaccine. All information is taken from the 2002 World Development Indicators. Finally, we include an OECD dummy variable to capture any possible effect of differences between the more developed and developing countries.

Pathogen pressure is known to be different in various parts of the world, and to be relatively high close to the equator because the incidence of infections as well as outbreaks of new infections in single areas is higher. Since a higher level of pathogen pressure in an environment will lead to more diseases, it could induce a higher equilibrium level of MHC diversity in a population. Since such a relationship between pathogen pressure and MHC diversity could bias our estimates we use several proxies as a control for pathogen pressure. From Hall and Jones [1999] we use absolute latitude to capture the effect of populations residing close to the equator having a higher prevalence and incidence of infectious diseases. From Gallup et al. [2001] we add the following information into one variable: (i) the fraction of tropical area of a country, (ii) the fraction of the population living in tropical areas, and (iii) the tropical climate zone which is an indicator measuring the extent to which a country is regarded tropical. The variable used in the regression equation measures the standardized averages of the three indicators.

The regression equation now becomes

$$LE_{c1990} = C + \alpha_1 MHC_c + \alpha_2 PP_c + \alpha_3 LA_c + \alpha_4 HE_c + \alpha_5 IMM_c + \alpha_6 OECD_c + \epsilon_c, \quad (24)$$

where PP_c measures the degree of tropical features in country c and LA_c is absolute latitude, which together should capture the effect of different circumstances with regard to pathogen pressure. HE_c is the log of health expenditures per capita in 1995 and IMM_c is country c 's average immunization rate against DPT and measles, which together form the set of variables capturing possible differences in the country's level of health advancement.

Columns (2) and (3) in Table II present estimates of different variants of equation (24). OLS estimates in Column (2), where only the covariates measuring possible differences in pathogen pressure and the OECD dummy variable are added to equation (23), show a coefficient (standard error) for α_1 equal to $-56.505(18.344)$, which is significant at the one percent level and comparable to the estimate reported in Column (1). Adding these covariates on differences in pathogen pressure yields a predicted effect of a one-standard deviation increase in MHC diversity on equilibrium life expectancy of 0.259 years, which is equivalent to a 10 percentage point increase in MHC diversity reducing life expectancy at birth by 0.735 years.

Adding covariates aimed to capture the population's health status, HE_c and IMM_c , yields an OLS estimate for α_1 equal to $-40.045(14.942)$, which is also significant at the one percent level but somewhat lower compared to the estimate obtained from the previous analyzes (Column(3)). The predicted effect of a one-standard deviation increase in MHC diversity on equilibrium life expectancy is equal to -0.184 , which yields a reduction of life expectancy at birth of 0.521 years when MHC diversity is increased by 10 percentage points.

The measure of MHC diversity includes most likely some measurement error, which could bias its estimates towards zero. To deal with measurement error in MHC diversity we instrument HLA-A diversity with HLA-B diversity (the correlation coefficient between HLA-A and HLA-B diversity equals 0.508 and is significant at the one percent level). The estimates of this 2SLS strategy are reported in Columns (4)-(6) of Table II. The estimates suggest that taking care of the measurement error increases the estimated coefficients for α_1 considerably. The coefficients (standard errors) are now $-82.146(44.129)$; $-75.687(36.833)$; and $-79.069(29.787)$ for the three models estimated using 2SLS instead of OLS. These 2SLS estimates suggest that a one-standard deviation increase in MHC diversity reduces life expectancy at birth by 0.376, 0.347 and 0.362 years, respectively.

Columns (7)-(10) document the robustness of these findings to alternative samples. We report both the OLS and 2SLS regression results. A first concern is whether African countries, where the persistence of

a relatively low level of life expectancy (these countries are typically situated in the bottom half of Figure VIa) might have different causes, are driving these results. The estimates in Columns (7) and (8) exclude the ten African countries in our sample, but show no statistically significant effect on the results. Both the OLS and 2SLS estimates of α_1 are quantitatively lower to those in previous columns but still suggest a statistically significant reduction in life expectancy at birth when MHC diversity is higher.

Finally, we add the log of GDP per capita in 1960 from the 2002 World Development Indicators (GDP_{c1960}). The argument for including this covariate is that higher levels of income might have a direct effect on life expectancy at birth because people are able to acquire, for example, better food and housing. Adding this covariate changes the regression model into

$$LE_{c1990} = C + \alpha_1 MHC_c + \alpha_2 PP_c + \alpha_3 LA_c + \alpha_4 HE_c + \alpha_5 IMM_c + \alpha_6 GDP_{c1960} + \alpha_7 OECD_c + \epsilon_c. \quad (25)$$

The estimates for α_1 reported in columns (9) and (10) do not look different from estimating comparable equations reported in columns (3) and (6) of Table II; actually, the estimates on α_1 are even slightly higher. The coefficient on GDP per capita in 1960 is statistically significant, which suggests some direct causal effect of income on life expectancy in our sample of countries.²⁶ Hence, the effects of income on life expectancy seem to go through more economic fortunes and higher levels of mutual dependence, captured by (the inverse of) MHC diversity.

Because of the specific role of MHC type I molecules, our model predicts that MHC diversity is especially related to differences in health status related to viral infections. To take a closer look at the specific diseases in a population we use mortality measures from the WHO Mortality Database measuring the incidence of death in a population as a result of infectious diseases in general, and viral infections, and bacterial infections in particular.²⁷ Bacterial and viral infections together comprise a large number of infectious diseases. We do not include HIV infections because a number of countries have extreme high AIDS dead rates while other countries do not report these numbers for political and socioeconomic reasons.²⁸ In addition, differences in the availability of medication and the way in which governments deal with the HIV epidemic capture effects we do not want to take into account in the empirical analysis.

Figure VIb and VIc show the relationship between the fraction of the population that has died from infectious diseases multiplied by 1,000 and HLA-A diversity. Figure VIb does so for viral infections and

²⁶Note that Cuba does not report information on income variables, so the sample is reduced to 62 countries.

²⁷The data are taken from the 2000 WHO Mortality Database or the most recent year if 2000 is unavailable. The Appendix provides more details about the definition of infectious diseases from this data source. Gallup et al. [2001] provide information for the incidence of single (infectious) diseases. We do not use this information since we would aim to capture the effects of overall pathogen pressure and not the incidence of individual diseases. Information about infectious diseases is available for 47 populations only. We do not have information for the following countries: Botswana, Cape Verde, China, the Democratic Republic of Congo, India, Indonesia, Kenya, the Federal Republic of Macedonia, Morocco, Oman, Pakistan, Senegal, Turkey, Uganda, Vietnam, and Zambia. In addition, we again lose Cuba in the regressions including GDP_{c1960} .

²⁸For some countries there appears to be an unclassified group of infectious diseases, which we do not take into account when defining lethal bacterial and viral infections.

illustrates a strong positive relationship. Figure VIc illustrates the relationship for bacterial infections and HLA-A diversity. There does not appear to be a statistically significant relationship in the latter case. These two pictures are consistent with the role of class I MHC molecules in the immune response to different infections and add confidence to our measure of MHC diversity. Table III provides the main results for using VI_c , the fraction of the population that has died from viral infections (multiplied by 1,000), as the dependent variable. Column (1) shows the OLS relationship between MHC diversity and viral infections by estimating the following model:

$$VI_c = C + \alpha_1 MHC_c + \epsilon_c, \quad (26)$$

For bacterial infections we estimate the same equation, replacing VI_c by BA_c . The estimate for α_1 recovers the positive relationship between HLA-A diversity and the fraction of the population that has died as a result of viral infections and suggests that a one-standard deviation increase in HLA-A diversity increases mortality by 4.63 percent. Column (2) reports the estimate for α_1 using BA_c as the dependent variable. Consistent with the illustration in Figure VIc, MHC class I diversity and bacterial infections are not related.

Similar to the analysis above, additional columns report results of estimating the following equation:

$$VI_c = C + \alpha_1 MHC_c + \alpha_2 PP_c + \alpha_3 LA_c + \alpha_4 HE_c + \alpha_5 IMM_c + \alpha_6 GDP_{c1960} + \alpha_7 OECD_c + \epsilon_c. \quad (27)$$

Columns (3) and (4) report the OLS estimates and Columns (5) and (6) the 2SLS estimates where we have again applied HLA-B diversity as an instrument for HLA-A diversity. The estimates suggest a statistically significant relationship between MHC_c and VI_c .

Exclusion of the three African countries in the sample of 47 countries and adding GDP_{c1960} does not change the conclusion. It is interesting to note that GDP_{c1960} is not significant in the estimates reported in the final two columns of Table III. This estimate suggests no direct causal effect of income on VI_c in our sample of countries. Hence, the effects of becoming richer on health outcomes seem to go more through higher levels of mutual dependence, captured by (the inverse of) MHC diversity, than through income directly. For the estimates using life expectancy as a measure of health outcomes we also found some direct effect of economic performance.

Taken together these results suggest that higher levels of MHC class I diversity are indeed related to lower life expectancy and a higher incidence of people dying from viral infections. Our proxies for health outcomes are related to mortality, while in the model the fact that diseases prevent agents from working, already generates the results. Ideally we would want to measure the average days of absence from work due to diseases for our sample of countries, but such data are not available for a large set of countries. It

is very plausible to assume however that the causes of death reflect the kind of diseases from which many more individuals than only those who die, suffer in a population. Finally, from the perspective of our theory, the estimates imply that there is a discrepancy between the individual evolutionary equilibrium, that leads to high levels of MHC diversity, and the social optimum. These conclusions remain when we control for observed differences in pathogen pressure and differences in advancement of health systems and economic fortunes between the populations in the sample.

4.2 Human Cooperation

The second empirical prediction of the model is a relationship between measures of cooperation and MHC diversity. Mutual dependence is the result of the successful cooperative behavior in the linear public good game, making free-riding inefficient. The higher the level at which agents cooperate with each other, to achieve a higher level of individual output, the higher the level of mutual dependence. Higher levels of mutual dependence result in agents being more dependent on the health of the agents they cooperate with. Consequently, the health of other agents will be “internalized” in the evolution of the immune system of the individual agent and MHC diversity will be lower in an evolutionary stable equilibrium.

Empirically we expect a negative correlation between measures of cooperation and MHC diversity. In addition, the agent’s cooperative behavior with the group is voluntary and of a small-scale nature (only agents within close vicinity cooperate and benefit from mutual dependence), so we should observe particularly strong correlations between voluntary and more local forms of cooperation compared to relatively enforced and population-wide measures of cooperation. To establish the empirical predictions of the former we apply measures of the World Values Surveys such as measures of trust, civic cooperation, and the importance of friends relative to the importance of family, which directly relate to small-scale bonds (other than direct genetic bonds) between individual agents. Appendix A.III provides details on the construction and definitions of these measures of cooperation. In contrast, the more macroeconomic and nationally enforced measures of institutions, such as indicators of a country’s level of democracy, people’s confidence in federal government bodies, a nation’s executive constraints, rule of law and control of corruption, should not necessarily reveal a statistically significant relationship with MHC diversity because these indicators do not establish mutual dependence in local communities of agents working together.²⁹

Table IV presents the estimation results of the following regression model:

$$HC_c = C + \alpha_1 MHC_c + \alpha_2 FR_c + \alpha_3 OECD_c + \epsilon_c, \quad (28)$$

²⁹Additionally, some have argued that higher levels of economic performance do not increase macro-level cooperation, such as democracy. See Acemoglu et al. [2005].

where HC_c is a measure of human cooperation in country c and FR_c is a measure the degree of fractionalization in country c . Higher levels of fractionalization are likely to reduce the level of human cooperation in a population and at the same time have an impact on mutual dependence, measured through the MHC diversity of a population. We incorporate the average of three measures of fractionalization in all regressions, suggested in Alesina et al. [2003]: ethnic, language and religious fractionalization. Neither the correlation coefficients between the individual measures of fractionalization and MHC diversity nor the correlation between the composite measure and MHC diversity are statistically significant. This makes us confident that we do not pick up ethnic diversity or other measures of fractionalization in the MHC measure.³⁰

The first panel of Table IV presents the OLS estimates of equation (28) for the local measures of HC_c . The coefficient of interest is α_1 , which takes on statistically significant and negative values for all three measures we applied for HC_c . Trust and civic cooperation are defined in accordance with the study of Knack and Keefer [1997]. Trust measures the general level of how much agents trust each other, and civic cooperation is the average of three behavioral variables concerning benefiting from fraud, tax evasion and paying for public facilities, such as public transport, when making use of them. The predicted effects suggest that a one-standard deviation increase in MHC diversity reduces the level of trust and civic cooperation by 33.7 and 29.3 percent, respectively.³¹

The relative importance of friends is measured as the difference between the population's average importance of having close friends and having close family ties. This indicator signals the presence of reliable partners in human cooperation. Since everyday cooperation in production is often outside the direct family ties this indicator is of interest for the outcomes of the public good game. The coefficient for α_1 suggests that a one-standard deviation increase in MHC diversity reduces the relative importance of friends by 24.1 percent.

The second panel of Table IV reports the OLS estimates of applying more macroeconomic measures of HC_c to establish a relationship with MHC diversity. We use measures of confidence in the federal government from the World Values Survey,³² a measure of average democracy and executive constraints in the

³⁰The correlation (standard error) between MHC diversity and ethnic fractionalization is equal to 0.167(0.195), for language fractionalization the coefficient equals 0.065(0.618) and for religious fractionalization the correlation coefficient is negative and equal to $-0.103(0.412)$. For the composite measure the correlation coefficient is equal to 0.055(0.674) for a sample of 60 countries. There is no information available for all three measures for Cape Verde, Cuba and Puerto Rico. Potentially, fractionalization might have an impact on cooperation and mutual dependence. See for example Easterly and Levine [1997] who relate fractionalization in African societies to their poor economic performance and Caselli and Coleman [2004] who build a theory of ethnic conflict based on fractionalization.

³¹We do not have information about these measures of micro-level human cooperation for a considerable number of countries, which reduces the sample to a minimum of 51 countries in the regression using civic cooperation as the dependent variable. In particular, we lose Botswana, Cape Verde, the Democratic Republic of Congo, Cuba, Israel, Kenya, Mongolia, Oman, Senegal, Turkey and Zambia.

³²The measure is a composite indicator defined as the country's average score on confidence in the government, civil service, parliament, armed forces, legal system, and police.

period 1960-2000 from Jagers and Marshall [2000], and finally two measures of government effectiveness from Kaufmann et al. [2003]: rule of law and control of corruption.³³ These last two measures could also have some effect on local communities, since they relate to the features of a public good game in the sense that stricter norms and values with respect to the control of corruption and a stronger rule of law makes people more willing and more confident to take part in cooperation games.

The estimates of equation (28) using these measures as dependent variables reveal that none of these indicators of human cooperation exhibit a statistically significant relationship with MHC diversity, which suggests that mutual dependence is consistent with the model of a local (i.e., small-scale) public good game and does not work for societies as a whole. The measures of rule of law and control of corruption are significant at the ten percent level, which is consistent with these measures having some relation with small-scale human cooperation.

4.3 Economic Performance

The final empirical relationship we want to establish is one between economic outcomes and MHC diversity. Ultimately, the basic idea of the public goods game is that it increases output and mutual dependency. As a consequence, societies that are better able to cope with public good dilemmas by means of altruism, social norms or punishment, will be more successful in economic terms and at the same time produce a higher degree of mutual dependency. This level of mutual dependency is mirrored by the level of MHC diversity in a population.

Column (1) in the first panel of Table V reports OLS estimates of the following model:

$$\Delta GDP_{c_{1960-2000}} = C + \alpha_1 GDP_{c_{1960}} + \alpha_2 MHC_c + \alpha_3 OECD_c + \epsilon_c, \quad (29)$$

where $\Delta GDP_{c_{1960-2000}}$ is defined as the log difference between GDP per capita in 2000 and 1960, and $GDP_{c_{1960}}$ as the log of GDP per capita in 1960. The coefficient of interest is the one on α_2 . The estimate of α_2 illustrates a strong and negative correlation between MHC diversity and economic growth over this forty year period, which is significant at the one percent level. The estimate suggests that a one-standard deviation increase in MHC diversity reduces economic growth over this period by 0.435 percent. Given the average level (standard deviation) of growth of 0.923(0.612) for the sample as a whole, this is a relatively large effect.³⁴

³³Recently, Glaeser et al. [2004] have criticized some of these country-level indicators of institutions to be unsuitable for analyzing economic growth. They argue that most of these measures do not reflect constraints on the government and often not durable for long periods of time. We believe that, taken as one body of indicators, the measures are suitable for carrying out the simple analysis we propose here.

³⁴Similar estimates for α_2 are obtained when using HLA-B diversity instead of HLA-A diversity.

In subsequent columns we replace $\Delta GDP_{c1960-2000}$ by more recent changes in GDP per capita. Column (2) reports the estimates for the period 1970-2000 and Column (3) for the period 1980-2000 for HLA-A diversity as measure of mutual dependence. Finally, Columns (4)-(6) present the 2SLS estimates, where we use HLA-B as an instrument for HLA-A diversity. All estimates are statistically significant and the predicted effects are relatively large.

There might be some concern about regional correlation, since the Asian countries seem to have grown fast and reveal a relatively low level of HLA-A diversity. In addition, the African countries are generally poor and have been behind in terms of growth as well. At first site, they seem to reveal relatively high levels of MHC diversity. To control for these possible continental differences, we present two sets of estimates in Columns (7)-(10). One set of OLS and 2SLS estimates for the 1960-2000 period excluding Asia and another set excluding the African countries. The estimates for α_2 remain statistically significant in all these four specifications. In addition, estimating the model using continent dummy variables instead of $OECD_c$, or excluding certain regions from the regression analysis, gives similar results.

Column (1) of Table VI reports the OLS estimation results of estimating the following model:

$$\begin{aligned} \Delta GDP_{c1960-2000} = & C + \alpha_1 GDP_{c1960} + \alpha_2 MHC_c + \alpha_3 PP_c + \alpha_4 LA_c \\ & + \alpha_5 HE_c + \alpha_6 IMM_I + \alpha_7 FR_c + \alpha_8 OECD_c \epsilon_c. \end{aligned} \quad (30)$$

We added the covariats HE_c , IMM_c , PP_c and LA_c because differences in health systems and pathogen pressure could potentially have an impact on α_2 , the coefficient of interest. In addition, fractionalization (FR_c) could potentially have an impact as well. The estimate (standard error) for α_2 is equal to $-5.192(1.749)$ and significant at the one percent level. The estimate implies that a one-standard deviation increase in MHC diversity reduces economic growth by 0.291 percent. Compared to the previous table, the effect of MHC diversity on economic growth is reduced by about one-third, but still sizeable and economically significant.

Column (2) reports the 2SLS estimate of the same equation where we again instrumented HLA-A diversity with HLA-B diversity. The estimate for α_2 equals $-8.270(3.633)$ and is significant at the five percent level. Compared to the estimate reported in Column (4) of Table V, the size of the effect of MHC diversity on growth is here too reduced by about one-third from -0.613 to -0.463 .

Finally, we have applied a different measures of economic performance to estimate equation (30) to illustrate the relationship with MHC diversity. Using the current (2000) log level of GDP per capita yields a similar negative relationship with MHC diversity, as is illustrated in Column (3) of Table VI.³⁵ The estimate suggests that a one-standard deviation increase in MHC diversity reduces the level of GDP per

³⁵Similar results are obtained for HLA-B diversity.

capita by 10.8 percent. Estimating the model using continental dummies instead of an OECD dummy yields similar results.³⁶

Overall these estimates suggest a strong and statistically significant effect of MHC diversity on economic performance over time. In addition, the level of mutual dependence is able to explain the level of GDP per capita, which makes us confident that we are capturing relatively long-term trends in income next to patterns of medium-run growth.

4.4 Further Estimates

4.4.1 Overall Picture

The analyses in the previous section have shown that there is a robust relationship between MHC diversity and health outcomes $LE_{c_{1990}}$, human cooperation HC_c , and economic performance $\Delta GDP_{c_{1960-2000}}$, as predicted by the model. If health outcomes, human cooperation, and economic performance are mutually related, a link between MHC diversity and one of these three variables would already be sufficient to get a correlation with all three outcome variables. To further explore the robustness of our findings we investigate whether the significant relationships between MHC diversity and the three groups of output variables remain intact when we control for the other two output variables. If all the theoretical relationships would hold without any additional disturbances, the inclusion of more than one of the four core variables would lead to multicollinearity. In practice each of these dimensions will be subject to many other processes too. Therefore, this exercise could yield evidence about the existence of a direct link between MHC diversity and each of the three outcome variables.

More specifically in Columns (1)-(3) of Table VII we present estimates for the following three models:

$$LE_{c_{1990}} = C + \alpha_1 MHC_c + \alpha_2 \Delta GDP_{c_{1960-2000}} + \alpha_3 HC_c + \alpha_4 OECD_c + \epsilon_c, \quad (31)$$

$$HC_c = C + \alpha_1 MHC_c + \alpha_2 \Delta GDP_{c_{1960-2000}} + \alpha_3 LE_{c_{1990}} + \alpha_4 OECD_c + \epsilon_c, \quad (32)$$

and

$$\begin{aligned} \Delta GDP_{c_{1960-2000}} = C + \alpha_0 GDP_{c_{1960}} + \alpha_1 MHC_c + \alpha_2 HC_c + \\ \alpha_3 LE_{c_{1990}} + \alpha_4 OECD_c + \epsilon_c. \end{aligned} \quad (33)$$

For HC_c we only report the results of using trust as a proxy for human cooperation, because this variable has shown the strongest link with MHC diversity in the estimates reported in Section 4.2. The parameter

³⁶Reducing the sample by excluding negatively growing countries, such as Zambia, Congo, Venezuela and Senegal, excluding the high growth countries, such as Singapore, Korea, and Oman, or by excluding the Asian or African countries does not change the main conclusion drawn from Table VI.

of main interest is α_1 . The estimate for α_1 is statistically significant and negative in all three regressions, which suggests a direct effect of MHC diversity on all three dependent variables. Note that the estimates for α_2 and α_3 are insignificant in all three specifications.

The estimates for α_1 do not change substantially when we include other covariates into the regression equations. Columns (4)-(6) in Table VII report estimates from estimating the following three models:

$$LE_{c1990} = C + \alpha_1 MHC_c + \alpha_2 \Delta GDP_{c1960-2000} + \alpha_3 HC_c + \alpha_4 OECD_c + \alpha_5 HE_c + \alpha_6 IMM_c + \alpha_7 PP_c + \alpha_8 LA_c + \alpha_9 FR_c + \epsilon_c, \quad (34)$$

$$HC_c = C + \alpha_1 MHC_c + \alpha_2 \Delta GDP_{c1960-2000} + \alpha_3 LE_{c1990} + \alpha_4 OECD_c + \alpha_5 HE_c + \alpha_6 IMM_c + \alpha_7 PP_c + \alpha_8 LA_c + \alpha_9 FR_c + \epsilon_c, \quad (35)$$

and

$$\Delta GDP_{c1960-2000} = C + \alpha_0 GDP_{c1960} + \alpha_1 MHC_c + \alpha_2 HC_c + \alpha_3 LE_{c1990} + \alpha_4 OECD_c + \alpha_5 HE_c + \alpha_6 IMM_c + \alpha_7 PP_c + \alpha_8 LA_c + \alpha_9 FR_c + \epsilon_c. \quad (36)$$

Although some of the added covariates have a significant effect on the dependent variables, the estimates for α_1 remain statistically significant suggesting a direct effect of MHC diversity on the three measures of interest.

4.4.2 Inclusion of Other Regressors

The body of literature concerned with explaining cross-country differences in economic outcomes has suggested a number of additional variables explaining economic growth. This section presents estimates of including the most often used covariates in a growth regression. In particular we estimate the following equation

$$\Delta GDP_{c1960-2000} = C + \alpha_0 GDP_{c1960} + \alpha_1 MHC_c + \alpha_2 HC_c + \alpha_3 LE_{c1990} + \alpha_4 OECD_c + \alpha_5 HE_c + \alpha_6 IMM_c + \alpha_7 PP_c + \alpha_8 LA_c + \alpha_9 FR_c + \alpha_{10} OTH_c + \epsilon_c, \quad (37)$$

where OTH_c is either country c 's trade with others, M_2 /GDP, growth of the labor force, fertility, the investment rate, or gross enrolment in primary and secondary education. Trade is defined as the log difference of imports and exports in 1995 and 1960; M_2 /GDP is a financial variable suggested by King and Levine [1993] to capture primary financial development measured in 1995; growth of the labor force is defined as the log difference between the population aged 16-65 in 2000 and 1960; fertility is the fertility rate in 1990; investment rate is the price level of investment goods in 1960; and education is measured

using Barro’s measure of the proportion of eligible students enrolled in secondary and primary education in 1960 [Barro 1991].

The results are shown in Table VIII where each of the columns reports the inclusion of one of the covariates captured by OTH_c . The estimate for α_1 turns out to be relatively insensitive to the inclusion of these covariates and remains statistically significant in all specifications reported.

5 Alternative Explanations

5.1 Benefits of Diversity and Group Selection

The theoretical prediction and empirical observation that – from a social perspective – a lower level of MHC diversity is beneficial may strike the reader as a counter-intuitive result. The most likely reason is a general intuitive notion that diversity in general, and especially genetic diversity, is advantageous. It is important to notice that our finding about the social planner’s desire to establish a lower level of MHC diversity is related to the specific properties of the immune system and the role of the MHC in the adaptive part of the immune system. In economic life there are many examples in which either diversity or similarity is the preferred equilibrium outcome. The gains from the division of labor reveal (comparative) advantages of different abilities to perform specific tasks, but a population speaking the same language has economic gains as well.

The equilibrium in our model and estimates putting forward the societal benefits of lower MHC diversity is consistent with the observation that in immunization programs the individual risk of being hurt by these programs is larger than the individual benefits of being protected. At the same time, the social benefits of vaccination are much larger than the individual risk. What is crucial in our model is that mutual dependence shifts the evolutionary trade-off. The fact that the measure we deploy is a diversity measure is just a technical feature of the functioning of the immune system.

Another aspect of this discussion is that genetic diversity is important for the preservation of the ecological system. It is true that species with more genetic diversity have higher probabilities of survival during a very dramatic change of the natural environment. These are not the kinds of episodes we want to emphasize in our model for two major reasons. First, if genetic diversity in humans is valuable to prevent us from being wiped out during a dramatic event, it would require diversity in all relevant genes, not only diversity at the MHC. In practice, humans do show some degree of genetic diversity, but the degree of MHC diversity goes far beyond this “natural” level of diversity. Clearly, at the level of the MHC a different force is at work to preserve the species at stake. Second, even in the case where genetic diversity

is beneficial and crucial for survival, it is difficult to imagine that selection forces would significantly alter the level of diversity over a reasonable time-span (reasonable being here reasonable in economic terms). In our model, evolutionary pressure is defined at the individual level. The evolutionary success of one agent only depends on the agents in his direct environment. For the selection of societies with sufficient levels of genetic diversity, a process of group selection would be required. Such a process of group selection implies that selection depends on survival and extinction of complete societies. It remains an open question whether group selection processes are important in explaining (human) evolution, but if they are important these processes will require extremely long periods of time to show an effect. While in general – with human generations of about 20 years – a fast evolving gene takes at least 10,000 years to become important, group selection processes are likely to evolve with lower levels of evolutionary pressure and will require “generations” (i.e., the time span of the development of a small group of people into a mature society) of hundreds of years, making these processes unobservable for sensible economic analyzes. A reasonable time span for group selection processes would be of the order of magnitude of a couple of million of years. It is therefore more likely that group selection has played a role in the development of primitive species than in the development of differences between contemporary human populations.

5.2 Epidemics

An issue that is often put forward is whether a negative relationship between MHC diversity and economic performance can be explained by large epidemics. The argument used is that these large epidemics have led to a strong selection of people who have characteristics that enable survival. After the disease has exterminated all susceptible humans the degree of diversity will be lower. For example, it has been argued that the low degree of MHC diversity of Native Americans can be related to the diseases brought to America by white European immigrants. If epidemic infectious diseases have a direct negative impact on the economic performance of a population, the expected effect would be a positive correlation between MHC diversity and economic performance. If epidemic diseases result in selection, economic performance might improve since the fraction of strong people will increase. However, there is no empirical evidence that populations that suffered large losses in the past due to epidemics are the ones with a higher level of economic performance. Furthermore, if evolutionary success is related to specific alleles, it is not clear why the evolutionary process has not increased the presence of these particular alleles over time.

5.3 Migration

Different populations tend to consist of people with dissimilar forefathers. When a country faces an inflow of many immigrants from countries with different alleles, the degree of MHC diversity in this country will rise. In that case MHC diversity is a reflection of group heterogeneity. A negative relationship between MHC diversity and economic performance would be found if economic performance benefits from homogeneity. The most interesting country to investigate from this perspective is the United States because it has experienced the most dramatic inflow of people from different European countries. The most obvious comparison is between the white U.S. population and people from Western European countries, such as England, France and Italy. Since the ancestors of U.S. whites are from these countries, the MHC frequencies in the United States should reflect these diverse origins, and MHC diversity in the United States should substantially exceed diversity in these European populations. However, the reverse turns out to be the case: HLA-A diversity among the white U.S. population is equal to 0.855, whereas it is equal to 0.893 in France, 0.874 in Italy, and 0.875 in the United Kingdom, which are significant differences given the 0.035 standard deviation of HLA-A diversity. A similar argument applies to HLA-A diversity in Australia, which equals 0.838. This observation about diversity in the United States and Australia is consistent with our simulation results that the frequency dependent nature of MHC diversity makes the diversity change relatively fast.

6 Conclusions and Directions for Future Research

In this paper we have shown that there exists an empirical link between economic circumstances and genetic evolution. We have argued that three properties are required for a successful empirical analysis: (i) We have to study genes that are of eminent importance for evolutionary success; (ii) the fitness of these genes has to depend on, and change with, economic circumstances; and, (iii) the evolution of these genes has to be relatively fast and keep abreast of changes in economic circumstances. The MHC turns out to be a good candidate for these analyses. Based on an evolutionary theory of the main properties of the adaptive immune system, we predict that low levels of MHC diversity minimize the incidence of infectious diseases in a population, while the evolutionary equilibrium with selfish agents leads to maximum levels of diversity. Assuming that populations differ in their degree of cooperation, three important predications have been made. First, if there is indeed a discrepancy between the social optimum and the selfish equilibrium, low levels of MHC diversity in a population should be associated with better health outcomes. Second, when differences in MHC diversity between countries reflect differences in the way people cooperate, MHC

diversity should be negatively related to indicators of cooperative behavior. Finally, this mechanism only works when cooperative behavior has a higher output relative to carrying out solo projects. Hence, economic performance should correlate negatively with MHC diversity. The model is able to replicate the crucial properties of the observed MHC distribution. Furthermore, based on data of MHC frequencies in 63 populations, we find a significantly negative correlation between health outcomes and MHC diversity, and between MHC diversity and indicators of cooperation. Finally, we find a robust negative correlation between MHC diversity and several measures of economic performance.

The findings presented in this paper are of general interest for several academic disciplines. In biology the role of HLA has been analyzed at the individual human level in relation to sexual selection and the progression of diseases. The main findings suggest that people with rare MHC alleles are better protected against diseases because rare MHC molecules complement the protection already provided by the MHC molecules of others. Furthermore, rare MHC molecules seem to be an individually optimal answer to the danger of host pathogen co-evolution. We have shown that taking into account economic externalities implies that the equilibrium level of MHC diversity runs counter to polymorphism and promotes homogeneity. In anthropology, researchers have focused on HLA polymorphism to explain linguistic and cultural similarities and disparities between populations. We have added to this that community homogeneity, measured by trust, is likely to lead to relatively rapid changes in the frequency distribution of the HLA and its degree of population diversity. From the viewpoint of behavioral ecology (which has focused on the issue of human cooperation by means of communication, monitoring and enforcement of norms) we add a genetic link involving human cooperation to explain why humans might act cooperatively. Finally, we add to the economics discipline that (long-term) economic performance can be explained in part by human cooperation through HLA homogeneity, thereby not repudiating the role of other, more traditional factors.

There are several avenues for future research. First, investigating the human immune system from the perspective of differences between individual and social benefits might shed light on the biological question why the level of diversity of HLA-A is generally lower than the level of diversity of HLA-B and why HLA diversity between populations is associated with the existence of one dominant type which differs between different populations. One possible reason could be that the types of diseases HLA-A is protecting against are related to larger externalities than the ones HLA-B is taking care of. Relating the existence of infectious diseases, e.g. the several mutants of HIV, to HLA diversity would be an obvious avenue to proceed. A first attempt by Trachtenberg et al. [2003] – analyzing the association of HLA super-types with HIV disease progression rates in a population of HIV-infected men – has suggested that HIV adapts to the

most frequent alleles in a population, providing a selective advantage for those who express rare alleles. Second, we have used genetic information for the dominant groups in a country. Comparison of levels of HLA diversity within and between different populations living in the same country or region could reveal information about the level of human cooperation within and between different groups in a society. Third, MHC diversity in animals is related to sexual selection and disassortive mating. From economic studies concerning sexual selection we know that economic circumstances such as income inequality affect the sexual selection process significantly. An interesting question to explore is whether or not these external influences also affect MHC polymorphism and a population's long-term health status.

Appendix

A.I. MHC Data Sources

In 1991, the organizers of the Eleventh International Histocompatibility Workshop and Conference asked participating laboratories from countries all over the world to collect blood samples from representative samples of populations. Each laboratory received a set of sera to type the various HLA antigens.³⁷ Juji et al. [1992] provide detailed information about the sera used, the distribution of the sera to the laboratories, and the procedures for testing blood samples. In this way unique and consistent information about allele frequencies of many ethnic groups has been brought together and carefully documented. We have selected individuals from 72 populations that could potentially be matched to countries based on blood samples of 10,394 persons. The major advantage of this undertaking is that we are able to use data about HLA types of a representative group of (healthy) people in each population, collected using identical typing procedures. To the best of our knowledge there are three alternative sources of HLA frequencies for different human populations. First, Cavalli-Sforza et al. [1994] provide tables for HLA-A and HLA-B frequencies. The main limitation of these data is that only frequencies of alleles found consistently in all populations in the world are reported, because these have been found most relevant for anthropologic purposes. To construct a correct measure of HLA diversity, information about rare alleles is also needed. Second, the National Center for Biotechnology Information (NCBI) in Bethesda MD provides an online database of HLA typing studies classified by population.³⁸ Third, a comparable database is provided by Derek Middleton at <http://www.allelefrequencies.net/>. The main disadvantage of these latter two databases is that methods of typing differ between the several studies documented on these web pages. In addition, many of the studies included in these databases refer to specific and small sub-populations within countries.

For the empirical analysis, we have used all sources but applied the data from the Eleventh International Histocompatibility Workshop and Conference as the primary data source. As a secondary source we have made use of the NCBI data and lastly we added information from <http://www.allelefrequencies.net/> for countries not included in the other two data sources. Information from the Netherlands is obtained from a study by Voorter, Drent and Van den Berg-Loonen [2004]. We did not use the information collected by Cavalli-Sforza et al. [1994] because it is impossible to construct a reliable diversity index using these data.

³⁷ Because reliability of the sera was regarded crucial for this project, a couple of WHO approved antigens could not be identified by this procedure.

³⁸ See <http://www.ncbi.nlm.nih.gov/IEB/Research/GVWG/IHWG/ihwg.cgi?ID=9&cmd=PRJOV>.

Apart from sampling errors, especially the match between ethnic groups and countries and the method applied for typing the molecules is likely to lead to differences in the measurement of HLA diversity. Data from the Eleventh International Histocompatibility Workshop and Conference are likely to be of a relatively high quality compared to the other sources, since all laboratories have used exactly the same sera for typing HLA and are based on representative samples of the ethnic groups considered for analysis. To investigate the robustness of our findings with respect to potential measurement problems in HLA diversity, we carried out all regressions for countries represented only in the data of the Eleventh International Histocompatibility Workshop and Conference as well. We did not find any significant differences with the empirical results from the complete sample.

The HLA has been measured to represent genetic diversity within one ethnic group rather than diversity within a population. For this reason we matched countries to the HLA diversity of the dominant ethnic group, e.g. for the United States we make use of the HLA diversity among U.S. whites, and to characterize HLA diversity in China we use the Northern Han. This procedure excludes information for several ethnic groups because they refer to a group within a country that is not dominant or to ethnic groups spread over a number of countries. We end up with information about HLA diversity for 62 countries: 43 from the Proceedings of the Eleventh International Histocompatibility Workshop and Conference, 11 from the database of the NCBI and eight from Derek Middleton’s database. We add information from the Netherlands from Voorter, Drent and Van den Berg-Loonen [2004] to complete our sample of 63 countries.

A.II. Estimation of HLA diversity

Diversity is defined as $1 - \sum_k F_k^2$ where F_k is the frequency of k in the population. Using the distribution of MHC molecules in the sample (\hat{F}_k), to calculate diversity $\sum_k F_k^2$ gives an upward biased estimate, and thus an underestimation of diversity:

$$E\left(\sum_k \hat{F}_k^2\right) = E\left(\sum_k (F_k + \epsilon_k)^2\right) = \sum_k F_k^2 - \sum_k E(\epsilon_k^2). \quad (38)$$

We approximate

$$E(\epsilon_k^2) = \frac{F_k(1 - F_k)}{n} \quad (39)$$

with

$$\frac{\hat{F}_k(1 - \hat{F}_k)}{n} \quad (40)$$

and subtract this expected bias from the diversity in the sample to obtain an estimate of the diversity in the population. Simulation for distributions of HLA-A and HLA-B molecules similar to those observed in the data reveals that this procedure reduces the bias almost completely.

A.III. Data Definitions and Sources

Table AI presents an overview of the variables, the sources and definitions used in the paper. Measures of MHC diversity are described and defined above.

References

- Acemoglu, Daron, and Simon Johnson, "Disease and Development: The Effect of Life Expectancy on Economic Growth," Working Paper, MIT, November 2004.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson, "Disease and Development in Historical Perspective," *Journal of the European Economic Association*, I (2003), 397-405.
- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared, "Income and Democracy," NBER Working Paper No. 11205, March 2005.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, VIII (2003), 155-194.
- Aoki, Kenichi, "Time Required for Gene Frequency Change in a Deterministic Model of Gene Culture Coevolution, With Special Reference to the Lactose Absorption Problem," *Theoretical Population Biology*, XL (1991), 354-368.
- Barro, Robert J., "Economic Growth in a Cross-Section of Nations," *Quarterly Journal of Economics*, CVI (1991), 407-443.
- Borghans, José A. M., Joost B. Beltman, and Rob J. de Boer, "MHC Polymorphism Under Host-Pathogen Coevolution," *Immunogenetics*, LV (2004), 732-739.
- Caselli, Francesco, and Wilbur John Coleman II, "On the Theory of Ethnic Conflict," Working Paper, Harvard University, 2004.
- Cavalli-Sforza, Luigi Luca, Paolo Menozzi, and Alberto Piazza, *The History and Geography of Human Genes* (Princeton University Press, Princeton NJ, 1994).
- De Boer, Rob J., José A.M. Borghans, Michiel van Boven, Can Kesmir, and Franz J. Weissing, "Heterozygote Advantage Fails to Explain the High Degree of Polymorphism of the MHC," *Immunogenetics*, LV (2004), 725-731.
- Durham, William H., *Coevolution: Genes, Culture, and Human Diversity* (Stanford University Press, San Francisco CA, 1992).
- Easterly, William, and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, CXII (1997), 1203-1250.
- Fehr, Ernst, and Urs Fischbacher, "The Nature of Human Altruism," *Nature*, CDXXV (2003), 785-791.
- Gallup, John L., Mellinger, Andrew, and Sachs, Jeffrey D., "Geography Datasets," Center for International Development, Harvard University, 2001.
- Galor, Oded, and Omer Moav, "Natural Selection and the Origin of Economic Growth," *Quarterly Journal of Economics*, CXVII (2002), 1133-1192.
- Galor, Oded, and Omer Moav, "Natural Selection and the Evolution of Life Expectancy," Working Paper Brown/Hebrew University, August 2004.
- Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, "Do Institutions Cause Growth?" *Journal of Economic Growth*, IX (2004), 271-303.
- Hall, Robert E., and Charles I. Jones, "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, CXIV (1999), 83-116.
- Hansen, Ted H., and David H. Sachs, "The Major Histocompatibility Complex," in: William E. Paul (ed.), *Fundamental Immunology, Second Edition* (Raven Press Ltd., New York NY, 1989), 445-487.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath, "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, XCI (2001), 73-78.

- Hewlett, Barry S., Annalisa De Silvestri, and C. Rosalba Guglielmino, "Semes and Genes in Africa," *Current Anthropology*, XLIII (2002), 313-321.
- Jagers, Keith, and Monty G. Marshall, "Polity IV Project," Center for International Development and Conflict Management, University of Maryland, 2000.
- Juji, Takeo, Tatsuya Akaza, Katsushi Tokunaga, Hiromi Miyoshi, and Kouichi Kashiwase, "The Serology Studies of the Eleventh International Histocompatibility Workshop: An Overview," in: Kimiyoshi Tsuji, Miki Aizawa, and Takehiko Sasazuki (eds.), *Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, (Oxford: Oxford University Press, 1992), 83-108.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, "Governance Matters III: Governance Indicators for 1996-2002," World Bank Policy Research Department Working Paper, 2003.
- Kiepiela, Photini, Alasdair J. Leslie, Isobella Honeyborne, Danni Ramduth, Christina Thobakgale, Senica Chetty, Prinisha Rathnavalu, Corey Moore, Katja J. Pfafferott, Louise Hilton, Peter Zimbwa, Sarah Moore, Todd Allen, Christian Brander, Marylyn M. Addo, Marcus Altfeld, Ian James, Simon Mallal, Michael Bunce, Linda D. Barber, James Szinger, Cheryl Day, Paul Klenerman, James Mullins, Bette Korber, Hoosen M. Coovadia, Bruce D. Walker, and Philip J.R. Goulder (2004), "Dominant Influence of HLA-B in Mediating the Potential Co-Evolution of HIV and HLA," *Nature*, CDXXXII (2004), 769-774.
- King, Robert G., and Ross Levine, "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics*, CVIII (1993), 717-737.
- Knack, Stephen, and Philip Keefer, "Does Social Capital Have a Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics*, CXII (1997), 1251-1288.
- Lund, Ole, Morten Nielsen, Can Kesmir, Anders Gorm Petersen, Claus Lundegaard, Peter Worning, Christina Sylverster-Hvid, Kasper Lamberth, Gustav Rder, Sune Justesen, Sren Buus, and Sren Brunak, "Definition of Supertypes for HLA Molecules Using Clustering of Specificity Matrices," *Immunogenetics*, LV (2004), 797 - 810.
- Maynard Smith, John and George R. Price, The Logic of Animal Conflict, *Nature*, CCXLVI (1973), 15-18.
- Milinski, Manfred, and Claus Wedekind, "Evidence for MHC-Correlated Perfume Preferences in Humans," *Behavioral Ecology*, XII (2001), 140-149.
- Ofek, Haim, *Second Nature: Economic Origins of Human Evolution* (Cambridge University Press, Cambridge, 1999).
- Penn, Dustin J., "The Scent of Genetic Compatibility: Sexual Selection and the Major Histocompatibility Complex," *Ethology*, CVIII (2002), 1-21.
- Playfair, John H.L., and Gregory J. Bancroft, *Infection and Immunity* (Oxford University Press, Oxford, 2004).
- Reusch, Thorsten B.H., Michael A. Haberli, Peter B. Aeschlimann, and Manfred Milinski, "Female Sticklebacks Count Alleles in a Strategy of Selection Explaining MHC Polymorphism," *Nature*, CDXIV (2001), 300-302.
- Robinson, Mary Ann, and Thomas J. Kindt, "Major Histocompatibility Complex and Genes," in: William E. Paul (ed.), *Fundamental Immunology, Second Edition* (Raven Press Ltd., New York NY, 1989), 489-539.
- Saint-Paul, Gilles, "On Market Forces and Human Evolution," Working Paper, Toulouse, July 2003.
- Sette, Alessandro, and John Sidney, "Nine Major HLA Class I Supertypes Account for the Vast Preponderance of HLA-A and -B Polymorphism," *Immunogenetics*, L (1999), 201-212.
- Trachtenberg, Elizabeth, Bette Korber, Cristina Sollars, Thomas B. Kepler, Peter T. Hraber, Elizabeth Hayes, Robert Funkhouser, Michael Fugate, James Theiler, Yen S. Hsu, Kevin Kunstman, Samuel Wu,

John Phair, Henry Erlich and Steven Wolinsky, "Advantage of Rare HLA Supertype in HIV Disease Progression," *Nature Medicine*, IX (2003), 928-935.

Van Damme, Eric, *Stability and Perfection of Nash Equilibria* (Springer Verlag, Berlin, 1991).

Voorter, Cristina E.M., Marjolein Drent, and Ella M. van den Berg-Loonen, "Severe Pulmonary Sarcoidosis is Strongly Associated with HLA-DQB1*0602," Working Paper, Maastricht University, 2004.

Wedekind, Claus, and S. Furi, "Body Odor Preferences in Men and Women: Do They Aim for Specific MHC Combinations or Simple Heterogeneity?" *Proceedings of the Royal Society of London, Series B*, CCLXIV (1997), 1471-1479.

Table I
Simulation Results of the Pace of Change of the Evolutionary Process

	(1)	(2)	(3)	(4)
	Before	After	Difference	Fitness immediately after change
Diversity	0.700	0.600		
MHC-1	35.98	49.00	+13.02	1.0500
MHC-2	34.24	38.84	+4.60	1.0428
MHC-3	21.61	8.95	-12.66	0.9797
MCH-4	8.17	3.21	-4.96	0.9275
Percentage changing alleles			17.62	

Table II
Health Outcomes and MHC Diversity
(Dependent Variable: Life Expectancy at Birth in 1990)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	2SLS	2SLS	2SLS	OLS	2SLS	OLS	2SLS
<i>MHC_i</i>	-57.797 (26.943)	-56.505 (18.344)	-40.045 (14.942)	-82.146 (44.129)	-75.687 (36.833)	-79.069 (29.787)	-25.607 (11.969)	-44.795 (23.562)	-47.145 (14.434)	-95.562 (30.527)
<i>PP_i</i>		-5.074 (1.295)	-5.393 (1.033)		-5.203 (1.324)	-5.629 (1.107)	-2.484 (0.934)	-2.627 (0.972)	-5.085 (1.029)	-5.274 (1.138)
<i>LA_i</i>		-0.035 (0.069)	-0.153 (0.058)		-0.042 (0.070)	-0.163 (0.062)	-0.108 (0.045)	-0.109 (0.046)	-0.132 (0.058)	-0.137 (0.064)
<i>HE_i</i>			3.082 (0.668)			2.842 (0.723)	1.962 (0.561)	1.847 (0.594)	1.755 (0.900)	1.263 (1.022)
<i>IMM_i</i>			0.111 (0.046)			0.114 (0.053)	0.074 (0.043)	0.079 (0.050)	0.105 (0.046)	0.105 (0.055)
<i>GDP_{i1960}</i>									1.873 (0.818)	2.184 (0.915)
<i>OECD_i</i>	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R ²	0.055	0.567	0.726	0.348	0.546	0.701	0.602	0.582	0.755	0.717
<i>n</i>	63	63	63	63	63	63	53	53	62	62
Pred. effect of a one-standard deviation increase in MHC diversity	-0.265	-0.259	-0.184	-0.376	-0.347	-0.362	-0.117	-0.205	-0.216	-0.438
Pred. effect of 10 point increase in MHC diversity	-0.751	-0.735	-0.521	-1.068	-0.984	-1.028	-0.333	-0.582	-0.613	-1.242

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table III
 Infectious Diseases and MHC Diversity
 (Dependent Variable: Fraction of the Population that Has Died From Viral Infections in 2000 (Multiplied by 1,000))

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	2SLS	2SLS	OLS	2SLS	OLS	2SLS
MHC_i	0.306 (0.092)	0.001 (0.001)	0.348 (0.084)	0.326 (0.083)	0.431 (0.131)	0.460 (0.130)	0.249 (0.087)	0.386 (0.150)	0.330 (0.081)	0.438 (0.125)
PP_i			0.011 (0.006)	0.013 (0.006)	0.013 (0.007)	0.015 (0.007)	0.012 (0.007)	0.013 (0.007)	0.010 (0.006)	0.011 (0.007)
LA_i			0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
HE_i				-0.007 (0.004)		-0.005 (0.004)	-0.006 (0.004)	-0.005 (0.004)	-0.011 (0.006)	-0.012 (0.006)
IMM_i				0.000 (0.001)		0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
GDP_{i1960}									0.007 (0.005)	0.007 (0.005)
$OECD_i$	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.180	0.006	0.359	0.391	0.293	0.357	0.307	0.276	0.443	0.402
n	47	47	47	47	47	47	44	44	46	46
Pred. effect of a one-standard deviation increase in MHC diversity	0.463	0.000	0.527	0.493	0.652	0.696	0.377	0.584	0.500	0.663
Pred. effect of 10 point increase in MHC diversity	1.314	0.001	1.495	1.400	1.851	1.976	1.069	1.658	1.417	1.881

Note: Variables are defined in the Appendix. Standard errors are reported in brackets. Column (2) reports an estimate for MHC_i using bacterial infections as the dependent variable.

Table IV
Human Cooperation and MHC Diversity
(Dependent Variables: Eight Measures Capturing Human Cooperation)

	MHC_i	FR_i	Adj. R^2	n	Pred. effect of a one-st.dev increase in MHC diversity
<i>Micro-level or local measures of human cooperation</i>					
Trust	-1.414 (0.492)	-0.019 (0.100)	0.286	53	-0.337
Civic Cooperation	-4.576 (2.178)	0.040 (0.446)	0.028	51	-0.293
Relative Importance of Friends	-1.457 (0.648)	0.098 (0.132)	0.413	52	-0.241
<i>Macro-level measures of human cooperation</i>					
Confidence in the Federal Government	-1.701 (1.333)	0.306 (0.255)	0.064	47	-0.189
Average Democracy 1960-2000	3.187 (10.834)	-0.928 (2.515)	0.096	41	0.044
Average Executive Constraints 1960-2000	-0.786 (5.233)	1.730 (1.215)	0.361	41	-0.019
Rule of Law	-4.706 (2.383)	-0.704 (0.442)	0.622	58	-0.161
Control of Corruption	-4.512 (2.495)	-0.472 (0.463)	0.639	58	-0.144

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table V
Economic Performance and MHC Diversity
(Dependent Variables: GDP per Capita Growth for the Indicated Periods)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	1960-2000	1970-2000	1980-2000	1960-2000	1970-2000	1980-2000	1960-2000	1960-2000	1960-2000	1960-2000
	OLS	OLS	OLS	2SLS	2SLS	2SLS	OLS	2SLS	OLS	2SLS
MHC_i	-7.589 (2.011)	-6.412 (1.779)	-4.609 (1.329)	-10.705 (4.133)	-6.180 (3.535)	-5.925 (2.712)	-7.939 (1.808)	-13.030 (4.303)	-0.499 (2.449)	-6.287 (3.818)
Initial GDP_i	-0.093 (0.080)	-0.108 (0.077)	-0.117 (0.053)	-0.095 (0.081)	-0.108 (0.077)	-0.118 (0.053)	-0.118 (0.071)	-0.105 (0.077)	-0.115 (0.084)	-0.113 (0.085)
Adj. R^2	0.191	0.209	0.212	0.099	0.069	0.128	0.285	0.159	0.151	0.129
n	62	53	61	62	53	61	52	52	53	53
Predicted effect of a one-standard deviation increase in MHC diversity	-0.435	-0.444	-0.398	-0.613	-0.428	-0.511	-0.321	-0.857	-0.262	-0.329

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table VI
Economic Performance and MHC Diversity Including Covariates
(Dependent Variables: GDP per Capita (Growth) for the Indicated Periods)

	(1)	(2)	(3)
	1960-2000	1960-2000	2000
	OLS	2SLS	OLS
<i>MHC_i</i>	-5.192 (1.749)	-8.270 (3.633)	-4.971 (2.323)
Initial <i>GDP_i</i>	-0.425 (0.108)	-0.390 (0.116)	
<i>PP_i</i>	-0.147 (0.143)	-0.174 (0.149)	-0.079 (0.176)
<i>LA_i</i>	-0.017 (0.008)	-0.017 (0.008)	-0.013 (0.010)
<i>HE_i</i>	0.410 (0.111)	0.376 (0.119)	0.786 (0.106)
<i>IMM_i</i>	0.011 (0.005)	0.009 (0.006)	0.005 (0.007)
<i>FR_i</i>	-0.937 (0.348)	-0.934 (0.359)	-0.901 (0.341)
<i>OECD_i</i>	Yes	Yes	Yes
Adj. R ²	0.489	0.457	0.852
<i>n</i>	60	60	60
Predicted effect of a one- standard deviation increase in MHC diversity	-0.291	-0.463	-0.108

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table VII
Overall Estimates
(Dependent Variables: Life Expectancy, Human Cooperation and GDP per Capita Growth)

	(1)	(2)	(3)	(4)	(5)	(6)
	LE_i	HC_i	$GDP_{i1960-2000}$	LE_i	HC_i	$GDP_{i1960-2000}$
	OLS	OLS	OLS	OLS	OLS	OLS
MHC_i	-68.609 (23.023)	-1.537 (0.632)	7.454 (1.942)	-67.176 (16.902)	-1.195 (0.587)	-6.192 (1.721)
Initial GDP_i			-0.170 (0.078)			-0.470 (0.098)
LE_i		-0.002 (0.004)	0.007 (0.015)		0.001 (0.006)	-0.013 (0.015)
HC_i	-2.321 (5.331)		-0.290 (0.480)	0.439 (4.183)		0.363 (0.418)
$GDP_{i1960-2000}$	-1.489 (1.540)	-0.010 (0.042)		-2.595 (1.183)	0.014 (0.045)	
HE_i				3.318 (0.712)	-0.062 (0.030)	0.570 (0.112)
IMM_i				0.050 (0.056)	0.000 (0.002)	-0.003 (0.006)
PP_i				-3.555 (1.150)	0.002 (0.046)	-0.137 (0.127)
LA_i				-0.125 (0.063)	0.002 (0.002)	-0.015 (0.006)
FR_i				-8.485 (2.846)	0.047 (0.114)	-0.768 (0.296)
$OECD_i$	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.428	0.274	0.309	0.716	0.331	0.568
n	54	54	54	53	53	53

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table VIII
Robustness to the Inclusion of Other Covariates
(Dependent Variables: GDP per Capita Growth 1960-2000)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
MHC_i	-4.195 (1.430)	-3.888 (1.808)	-6.321 (1.737)	-6.067 (1.768)	-6.308 (1.669)	-5.493 (1.824)
Initial GDP_i	-0.465 (0.078)	-0.474 (0.093)	-0.505 (0.108)	-0.455 (0.106)	-0.432 (0.097)	-0.470 (0.099)
LE_i	0.014 (0.018)	0.003 (0.015)	-0.011 (0.016)	-0.018 (0.021)	-0.012 (0.015)	-0.009 (0.017)
HC_i	-0.439 (0.347)	-0.099 (0.476)	0.316 (0.424)	0.395 (0.430)	0.347 (0.405)	0.317 (0.431)
HE_i	0.240 (0.095)	0.456 (0.107)	0.602 (0.120)	0.550 (0.124)	0.601 (0.110)	0.587 (0.118)
IMM_i	0.001 (0.005)	-0.005 (0.006)	-0.003 (0.006)	-0.003 (0.006)	-0.002 (0.006)	-0.002 (0.006)
PP_i	-0.020 (0.095)	-0.002 (0.123)	-0.128 (0.128)	-0.148 (0.131)	-0.107 (0.124)	-0.125 (0.130)
LA_i	0.001 (0.007)	-0.009 (0.006)	-0.012 (0.008)	-0.016 (0.007)	-0.013 (0.006)	-0.014 (0.007)
FR_i	-0.451 (0.251)	-0.436 (0.302)	-0.777 (0.297)	-0.759 (0.300)	-0.839 (0.289)	-0.729 (0.308)
	Trade	M_2/GDP	Lab. For.	Fertility	Investment	Education
OTH_i	-0.137 (0.149)	0.008 (0.002)	0.170 (0.214)	-0.043 (0.109)	-0.208 (0.107)	0.164 (0.082)
$OECD_i$	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.755	0.704	0.565	0.559	0.595	0.561
n	29	39	53	53	53	53

Note: Variables are defined in the Appendix. Standard errors are reported in brackets.

Table AI
Definitions of Variables Used

Variable	Definition	Source
Per capita income measures (GDP)	GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 1995 U.S. \$. Dollar figures for GDP are converted from domestic currencies using 1995 official exchange rates. From the GDP per capita data we have computed the growth rates by taking the log differences for the relevant period.	World Bank Development Indicators [2002]
Gross enrolment in education	The total number of pupils enrolled at primary and secondary level in public and private schools.	Barro [1991]
Life expectancy	The number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	World Bank Development Indicators [2002]
Fertility	The number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with prevailing age-specific fertility rates.	World Bank Development Indicators [2002]
Fractionalization	The fractionalization data are measuring the percentage of foreign population (ethnic), of non-domestic languages and religion in a country in a particular available year (mostly in the 1980s and 1990s). See http://www.stanford.edu/~wacziarg/downloads/ for the data.	Alesina et al. [2003]
Trust	To assess the level of trust in a population we have used the answers to the following question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?". Our trust indicator is the weighted average of the two possible answers.	World Values Surveys
Civic Cooperation	This measure is a composite measure of the attitude towards illegal behavior in which people take advantage at the cost of society. The question used reads: "Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something in between," using a card with figures ranging from 10="Never Justifiable," to 1="Always Justifiable." For the empirical analysis we have used "cheating on taxes if you have a chance," "avoiding a fare on public transport," "accepting bribes," "drunk driving," and "throwing away litter".	World Values Surveys
Health expenditures	The sum of public and private health expenditures as a percentage of GDP. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health but does not include provision of water and sanitation.	World Bank Development Indicators [2002]
Immunization	Child immunization measures the rate of vaccination coverage of children under one year of age.	World Bank Development Indicators [2002]
Exports and imports	Exports (Imports) of goods and services represent the value of all goods and other market services exported (received) to (from) the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude labor and property income (formerly called factor services) as well as transfer payments.	World Bank Development Indicators [2002]
M ₂	The sum of currency outside banks, demand deposits other than those of the central government, and the time, savings, and foreign currency deposits of resident sectors other than the central government. It corresponds to lines 34 and 35 in the International Monetary Fund's	World Bank Development Indicators [2002]

	(IMF) International Financial Statistics (IFS).	
Labor force	People who meet the International Labour Organization (ILO) definition of the economically active population: all people who supply labor for the production of goods and services during a specified period. It includes both the employed and the unemployed.	World Bank Development Indicators [2002]
Investment	Price level of investment goods in 1960. Data are in constant 1995 U.S. \$.	World Bank Development Indicators [2002]
Democracy	The variable ranges from 0 to 10 where higher values equal a higher degree of democracy. Like Glaeser et al. [2004] we use the average from 1960 through 2000.	Jagers and Marshall [2000]
Executive constraints	The variable ranges from 0 to 7 where higher values equal a greater extent of constraints on the power of chief executives. Like Glaeser et al. [2004] we use the average from 1960 through 2000.	Jagers and Marshall [2000]
Confidence in federal government	The extent of the citizens' confidence in the legal system, government, parliament, civil service, armed forces and police. The question reads: "I'm going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?"	World Values Surveys
Rule of law	The governance indicators reflect the statistical compilation of responses on the quality of governance given by a large number of enterprise, citizen and expert survey respondents in industrial and developing countries, as reported by a number of survey institutes, think tanks, non-governmental organizations, and international organizations. The aggregate indicators in no way reflect the official position of the World Bank, its Executive Directors, or the countries they represent. As discussed in detail in the accompanying papers, countries' relative positions on these indicators are subject to margins of error that are clearly indicated. Consequently, precise country rankings should not be inferred from this data.	Kaufmann, Kraay and Mastruzzi [2003]
Control of corruption	The governance indicators presented here reflect the statistical compilation of responses on the quality of governance given by a large number of enterprise, citizen and expert survey respondents in industrial and developing countries, as reported by a number of survey institutes, think tanks, non-governmental organizations, and international organizations. The aggregate indicators in no way reflect the official position of the World Bank, its Executive Directors, or the countries they represent. As discussed in detail in the accompanying papers, countries' relative positions on these indicators are subject to margins of error that are clearly indicated. Consequently, precise country rankings should not be inferred from this data.	Kaufmann, Kraay and Mastruzzi [2003]
Infectious diseases	The disease variables are defined as the number of death through a certain disease as a percentage of the population in a country. See http://www3.who.int/whosis/menu.cfm?path=whosis.mort&language=english .	World Health Organization; Mortality Database
Tropical area information	Proportion of country's land area within geographical tropics. Proportion of country's population living in geographical tropics. Fraction tropical climate zone.	Gallup, Mellinger and Sachs [2001]
Importance of friends and family	The average importance of family and friends in a country on a scale from 1 (not important) to 5 (very important).	World Values Surveys

Figure Ia
HLA-A Diversity and GDP per Capita Growth 1960-2000

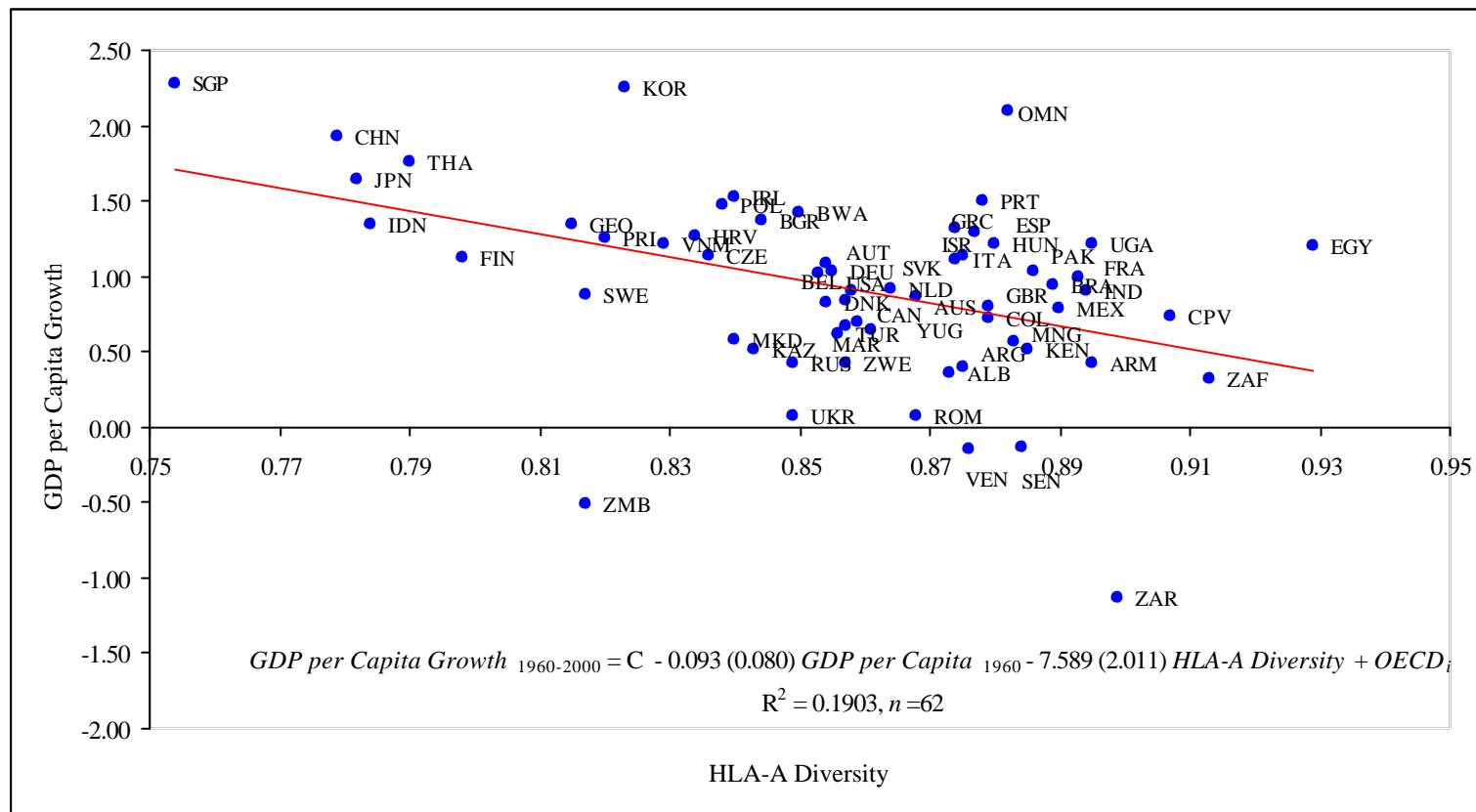


Figure Ib
HLA-B Diversity and GDP per Capita Growth 1960-2000

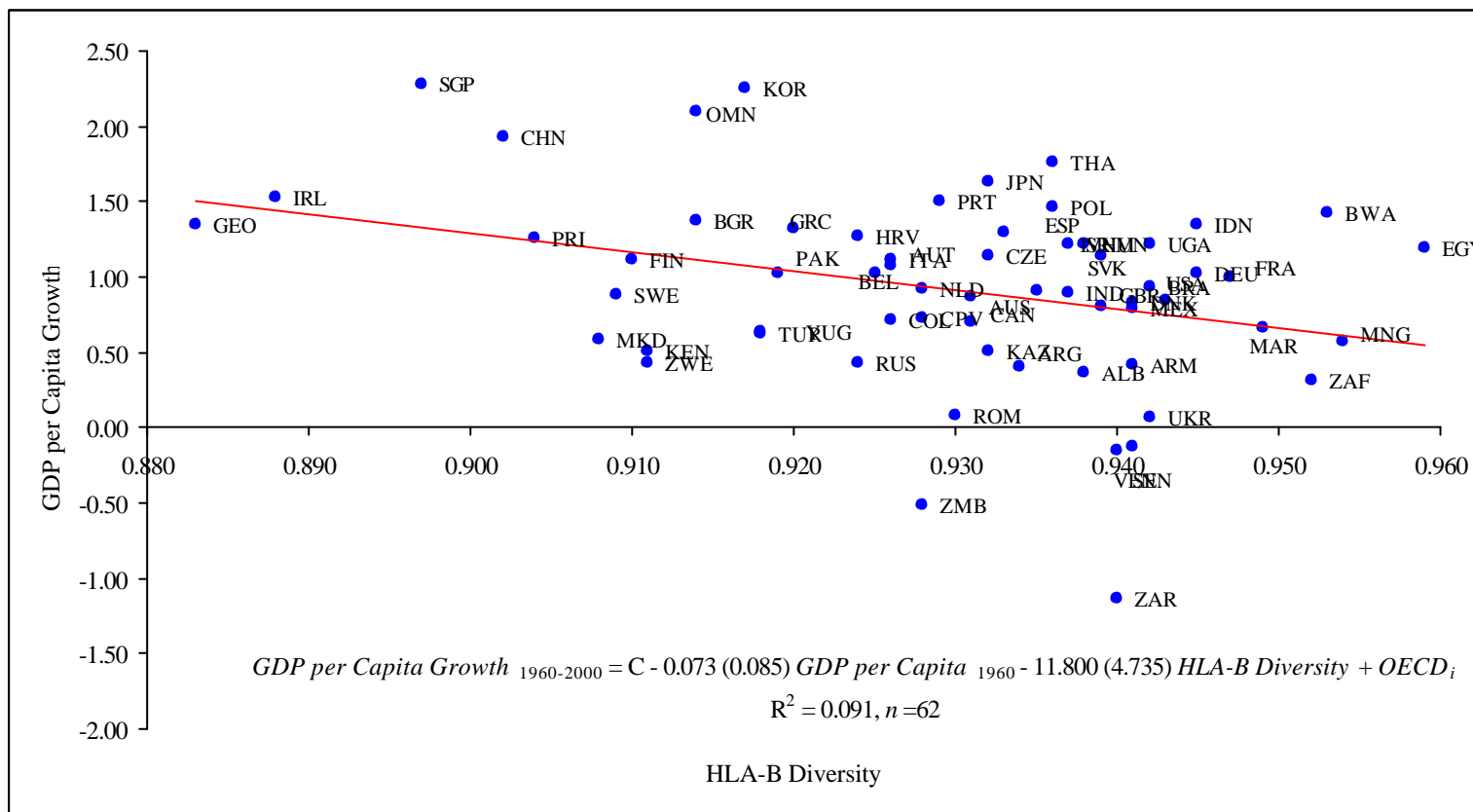
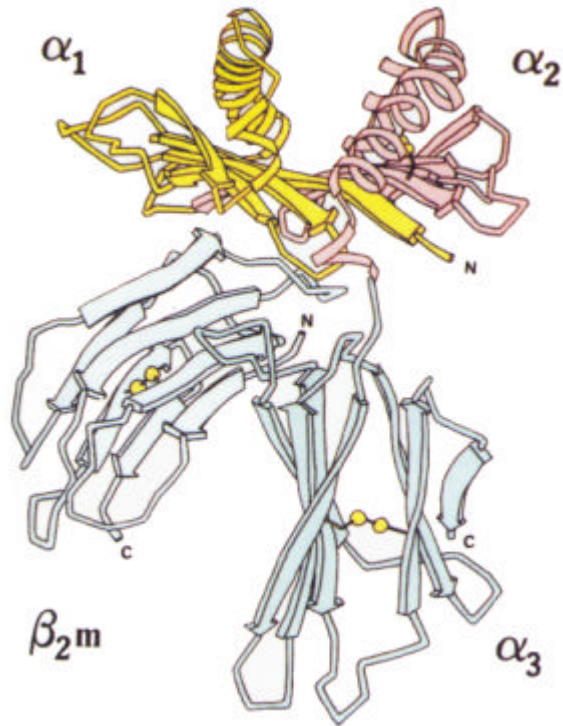
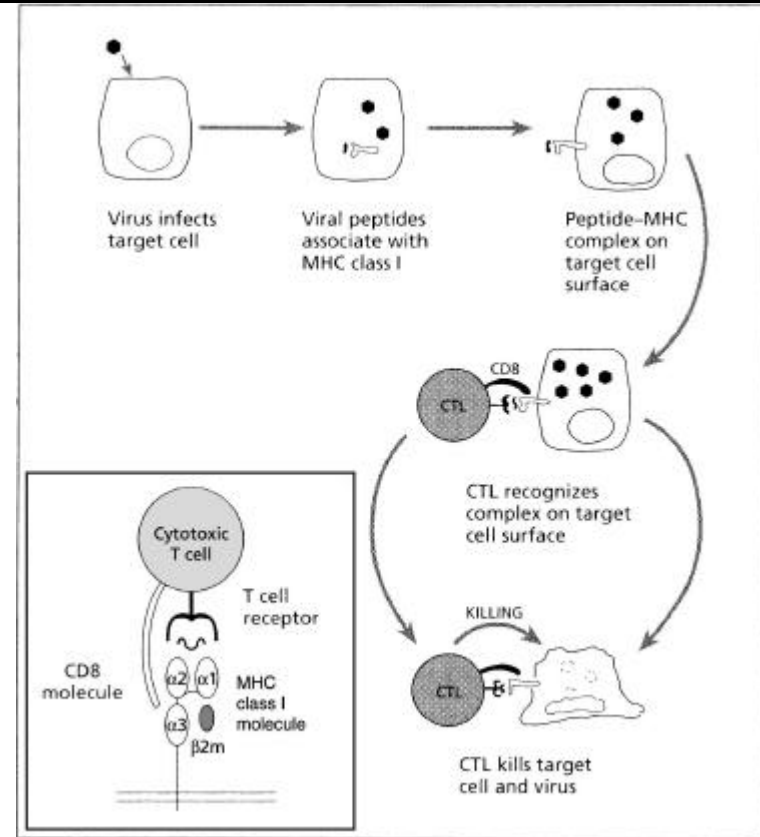


Figure II
The Workings of the Immune System and the Role of the MHC

(A)
Schematic Representation of HLA -A02 Showing the Four Domains with the Polymorphic α_1 and α_2 at the Top.



(B)
T Cell Response to Viral Infection and the Role of MHC Molecules in Presenting Peptides

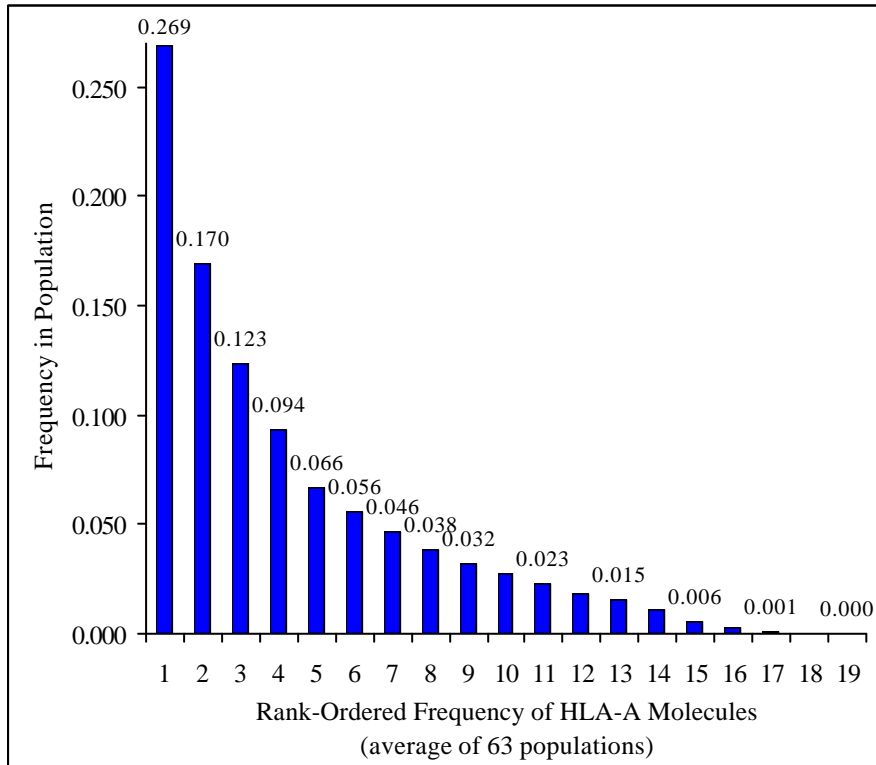


Source: Robinson and Kindt [1989], pp. 512-513.

Source: Adapted from Playfair and Bancroft [2004]

Figure III
Rank-Ordered Frequency Distribution of HLA-A and HLA-B in 63 Populations

(A)
HLA-A



(B)
HLA-B

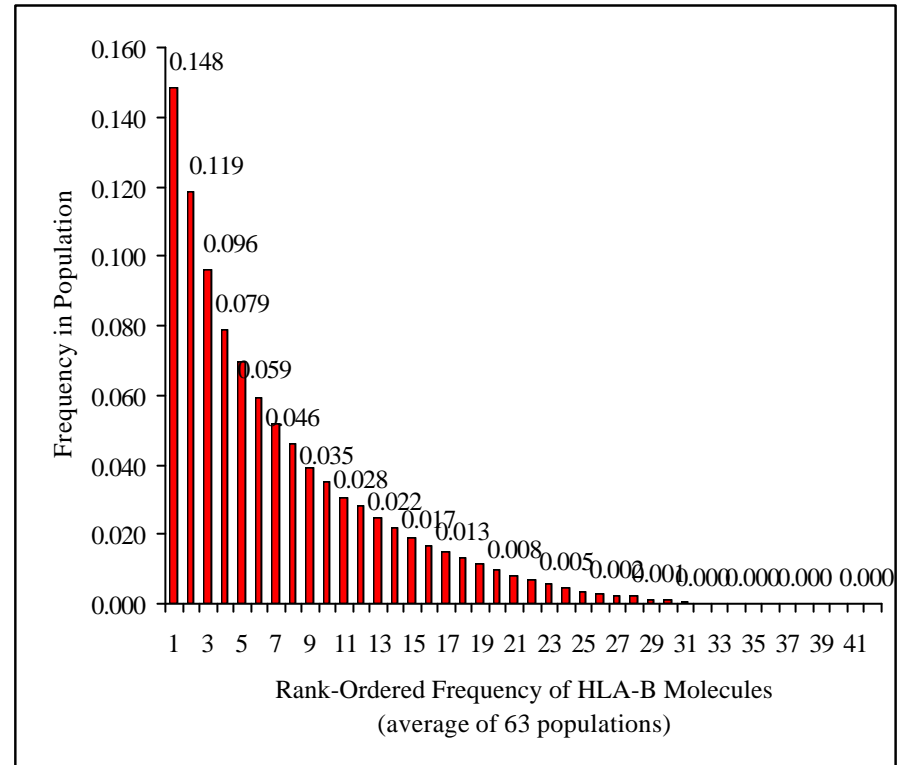
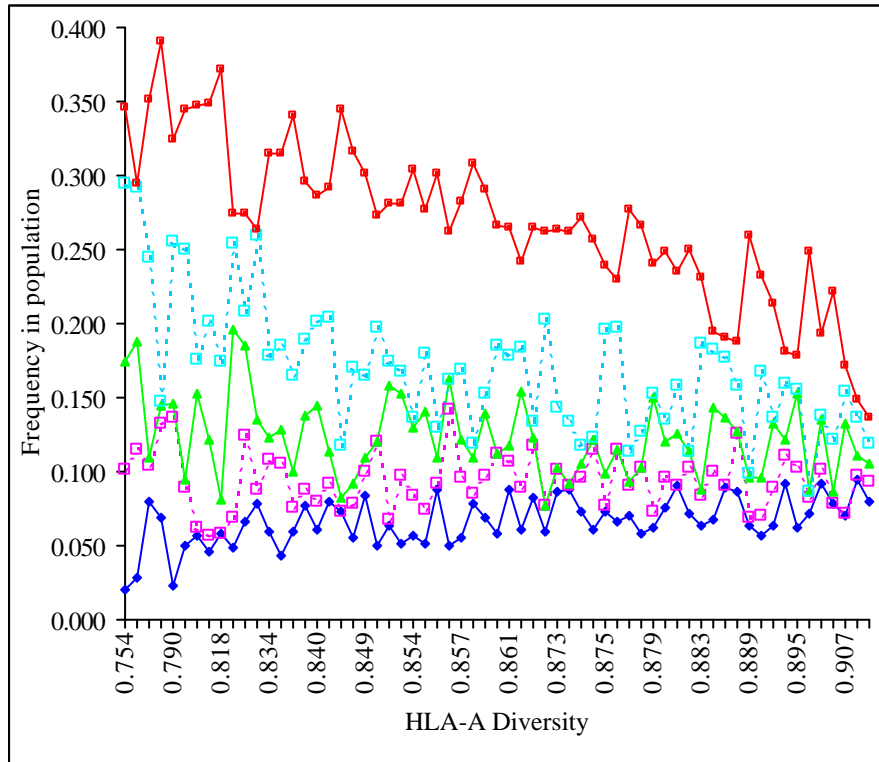


Figure IV
Frequency of HLA-A and HLA-B Alleles and Levels of Diversity

(A)
Appearance of the Five Most Frequent HLA -A Molecules and Diversity



(B)
Appearance of the Five Most Frequent HLA -B Molecules and Diversity

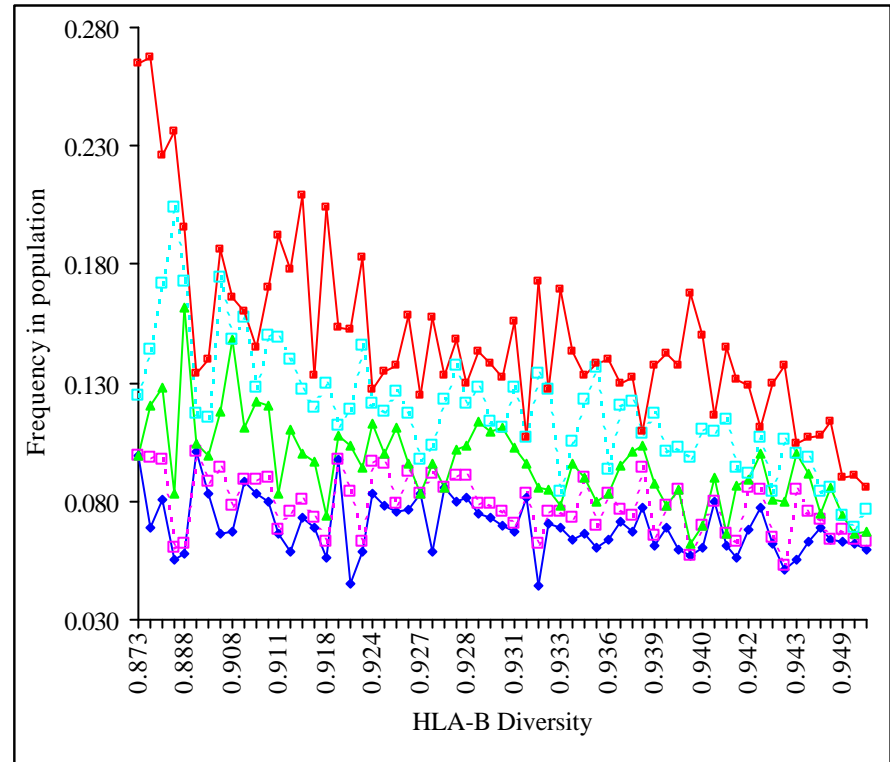
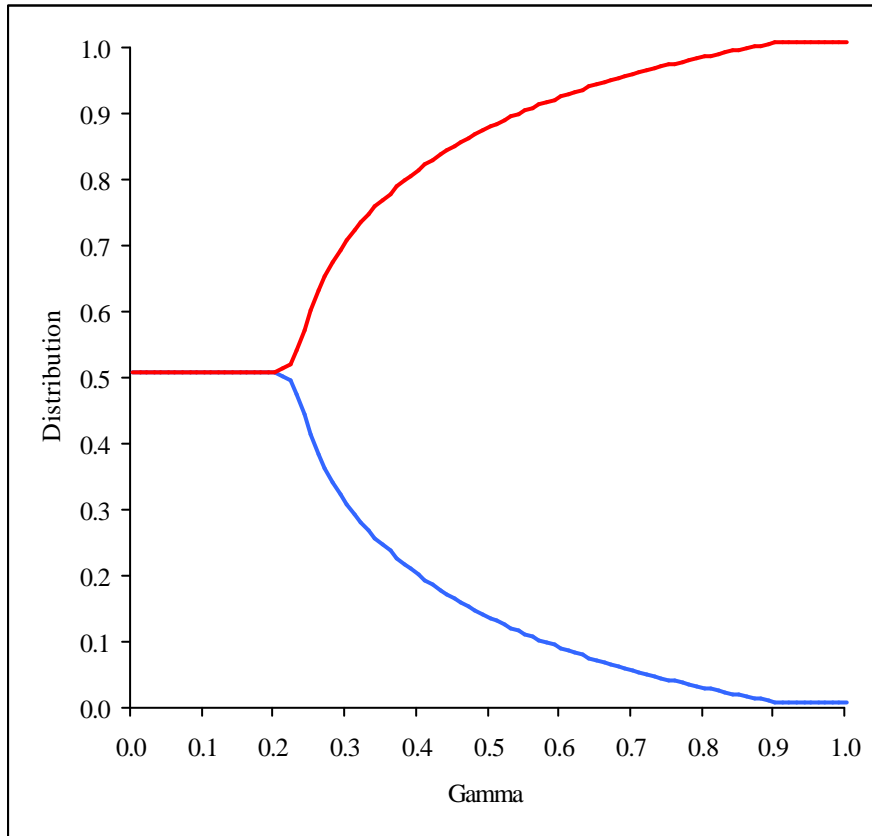


Figure V
Simulation Results of Human Cooperation and MHC Diversity

(A)
Results for Two MHC Molecules



(B)
Results for Three MHC Molecules

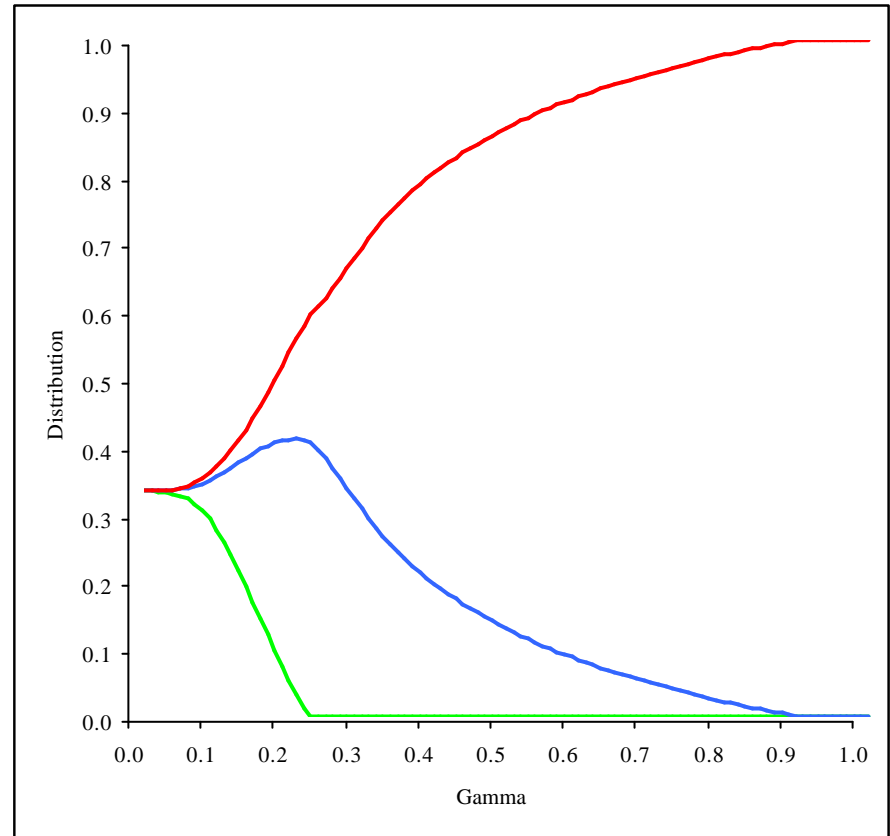
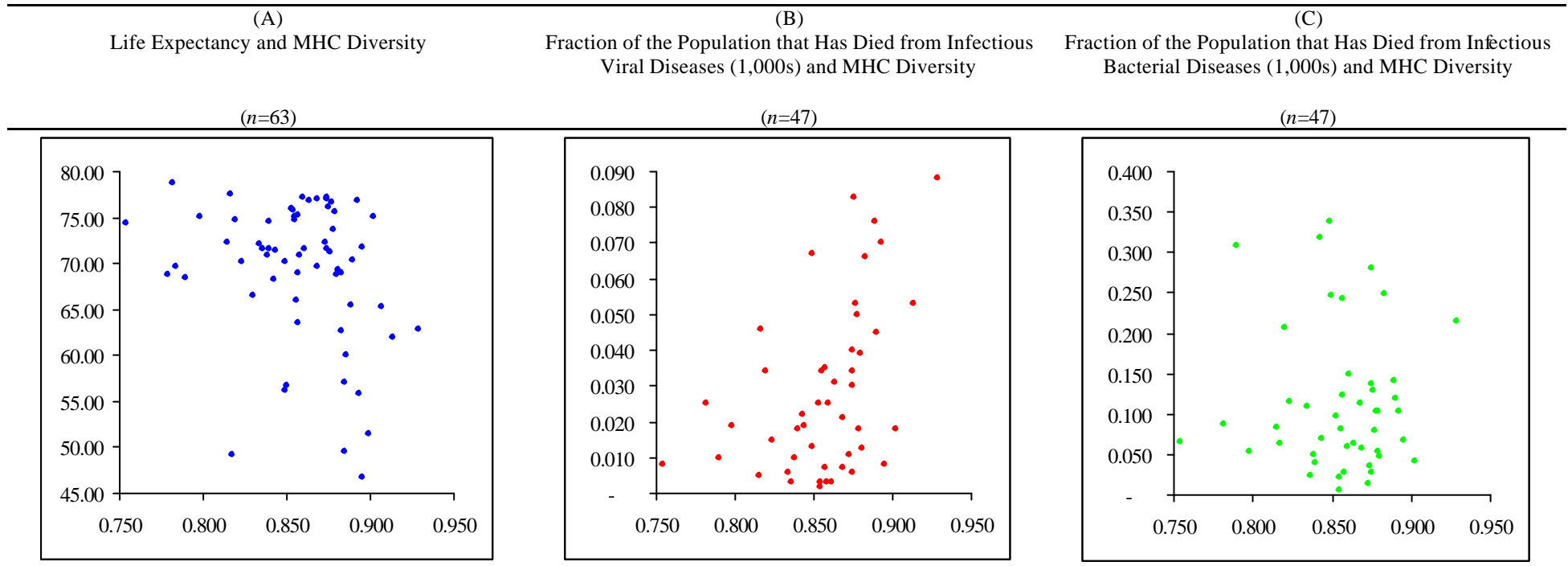


Figure VI
Health Outcomes and MHC Diversity



Note: Variables are defined in the Appendix. The horizontal axis measures MHC diversity (HLA -A diversity in this case) and the vertical axis illustrates the health outcomes.