The maintenance of strain structure in populations of recombining infectious agents

SUNETRA GUPTA[1], MARTIN C.J. MAIDEN[2], IAN M. FEAVERS[2], SEAN NEE[1], ROBERT M. MAY[1] & ROY M. ANDERSON[1]

[1]*The Wellcome Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[2]*Division of Bacteriology, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Herts, EN6 3QG, UK*

Using mathematical models that combine population genetic and epidemiological processes, we resolve the paradox that many important pathogens appear to persist as discrete strains despite the constant exchange of genetic material. We show that dominant polymorphic determinants (that is, those that elicit the most effective immune responses) will be organized into nonoverlapping combinations as a result of selection by the host immune system, thereby defining a set of discrete independently transmitted strains. By analysing 222 isolates of *Neisseria meningitidis*, we show that two highly polymorphic epitopes of the outer membrane protein PorA exist in nonoverlapping combinations as predicted by this general framework. The model indicates that dominant polymorphic determinants will be in linkage disequilibrium, despite frequent genetic exchange, even though they may be encoded by several unlinked genes. This suggests that the detection of nonrandom associations between epitope regions can be employed as a novel strategem for identifying dominant polymorphic antigens.

For many important pathogens of humans, such as *Plasmodium falciparum*[1], HIV-1 (ref. 2) and many infectious bacteria[3], epidemiological studies indicate that distinct strains persist in host populations over long periods of time. However, many of these pathogens have the opportunity for the exchange of genetic material because of, for example, sexual processes in the case of *P. falciparum*[4], coinfection of the same host cell for HIV-1 (ref. 5), or transformation in bacteria[6]. These two observations appear paradoxical, because it would seem that strain structure should be lost rapidly as a result of genetic recombination.

With the rapid development of molecular methods for typing microorganisms, and for studying the extent of genetic exchange, it is becoming increasingly clear that sex in many organisms such as *Escherichia coli*[7] does not appear to disrupt the "clonal" structure of the population. It has been proposed that extrinsic elements such as periodic selection, in combination with low rates of genetic exchange, may limit the genetically effective population size of organisms[8]. In this paper, however, we consider the more general situation where there are no constraints on genetic exchange, and selection is generated endogenously by host immune responses to the circulating strains.

In organisms that lack genetic exchange, a set of discrete strains may be clonally maintained, even in the absence of selection, because of nonrandom associations (or linkage disequilibrium) between multiple loci. However, it can easily be demonstrated[9] that even the smallest degree of outcrossing will eventually result in a random association (or linkage equilibrium) of alleles. Thus, even in populations of infectious agents that recombine infrequently, unless a "strain" is defined by a single polymorphic determinant, we would expect a rapid loss of strain structure due to the disintegration of associations between the relevant loci. Instead, many microorganisms that exchange genetic material appear to be transmitted in a manner that is consistent with a discrete rather than continuous population structure. In this paper, we explore this paradox using a simple mathematical model, and propose a solution as to how these conflicting observations may be reconciled.

A simple model for genetic exchange in infectious agents

A "strain" may be defined in epidemiological terms simply by the loci that affect its transmission rather than by its entire genetic constitution. As it is the host immune response that has the most profound effect on transmission (in the absence of drugs and other interventions), we assume that the relevant loci will encode a set of dominant immunogens (that is, those that elicit the most effective responses). Thus, for the case where there are two immunologically dominant loci, each with two alleles or variants, the four possible types of strains are *ay*, *ax*, *bx* and *by*, where *a* and *b* are alleles at one locus, and *x* and *y* are alleles at the second locus.

To assess the impact of genetic exchange on the population structure of infectious disease agents, we have developed a mathematical model (see Methods) for the dynamics of a set of recombining parasite strains within a homogeneous host population. To reflect the process of genetic exchange, we assume that the progeny of parasites within hosts that are infectious for two or more strains will consist of defined fractions of the various combinations of the different strains. This is indicated in Fig. 1 for the two-locus two-allele example presented above. The diagram shows that individuals infectious for both *ay* and *bx* are able to generate infectious elements of all four possible genotypes, while those singly infected can only release infectious progeny of the same genotype.

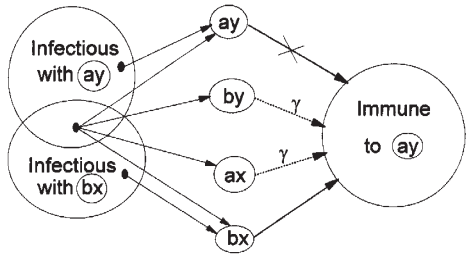

Fig. 1 Two principal assumptions of the model in the case where strains are defined by two loci each with two alleles. The infectious stages are thus given as *ay*, *by*, *ax* and *bx*. Those infectious with only one strain (for example, *ay* or *bx*, as shown in the diagram) can only generate infectious stages of the same strain. However, those infected with both *ay* and *bx* (shown by the overlap between the two compartments) can generate infectious progeny of all four types. An individual immune to a given strain, for example, *ay* as shown in the diagram, cannot be infected by *ay*, but may be infected by the totally unrelated strain *bx*. The transmission of related strains (in this case *by* and *ax*) within an individual immune to *ay* is modified by a factor γ.

We consider the simplest case where infection with a given strain confers lifelong strain-specific immunity. We assume that once an individual is rendered immune to a given strain, say *ay*, other strains sharing any of the relevant determinants (for example, *ax* and *by*) will be at a disadvantage within this host. Specifically, we assume that the capacity of such a host to transmit these strains will be reduced by a factor γ, which we will term the "degree of cross-protection." The factor γ is effectively a measure of the cost to a strain of infecting a host who is capable of mounting a protective immune response against some, but not all, of its specific variants. A strain that does not share any alleles with *ay* (that is, *bx*) will not be at any disadvantage within a host immune to *ay*. Any population may thus be grouped into set of "discordant" strains that do not share any alleles with each other, and therefore do not interfere with each other's transmission.

Analytical studies of the behavior of this model show that, in the limit where "cross-protection" between strains sharing alleles is complete (γ = 1), one set of discordant strains will always strongly dominate, in terms of abundance, over all other strains. Figure 2*a* illustrates this for the two-locus two-allele example. We have assumed that all four strains have the same basic reproductive number or R_0 (the average number of secondary cases of infection generated by one primary case in a susceptible population[10]). Thus, although the strains are identical in every respect, the system still settles to an "asymmetric" equilibrium. In the other limit, where cross-protection is absent (γ = 0), all strains independently equilibrate at the same level $(1 - 1/R_0)$, creating a "symmetric" equilibrium as shown in Fig. 2*b*. The transition from symmetric equilibrium to asymmetric equilibrium occurs when the degree of cross- protection, γ, exceeds a critical value, $\hat{\gamma}$, which may be given, in the two-locus two-allele case, as:

$$\hat{\gamma} = [1 - 1/R_0^i] / [1 - 1/(R_0^i R_0^{i'})]$$

Here R_0^i and $R_0^{i'}$ are the basic reproductive rates of the dominant pair of strains and R_0^i is the lower value of the subordinate pair. It can be shown by standard techniques of stability analysis that when γ exceeds this critical value, the symmetric equilibrium becomes unstable and bifurcates to give upper and lower asymmetric branches, as illustrated in Fig. 3. The switch from the symmetric equilibrium to the asymmetric equilibrium will manifest itself as a transition from linkage equilibrium to a high level of linkage disequilibrium. We conjecture that these broad patterns pertain more generally in *m* locus/*n* allele (n^m) systems.

The stability of the system against invasion by new recombinants is illustrated in Fig. 4, in which an antigenically distinct strain (say *by*) is introduced into a population within which a given strain (say *ax*) is already endemic. It can be seen that even recombinants (that is, *ay* and *bx*) of very much higher R_0 are unable to displace the original strains, provided the average duration of infectiousness is short. This property suggests that the persistence of a distinct subset of strains over long periods of time is likely to be more commonly observed for infectious agents with short infection periods in the host (for example, many common directly transmitted viral and bacterial infections) than for pathogens with longer infectious periods (for example, persistent viral infections such as hepatitis B, and many protozoan infections).

The important and general principle revealed by this analysis is that populations of infectious agents will be structured into independently transmitted strains by a dominant immune response against a polymorphic determinant, when there is a moderate to strong cross-protection between genotypes that share alleles at the loci that define a strain. When this is the case,

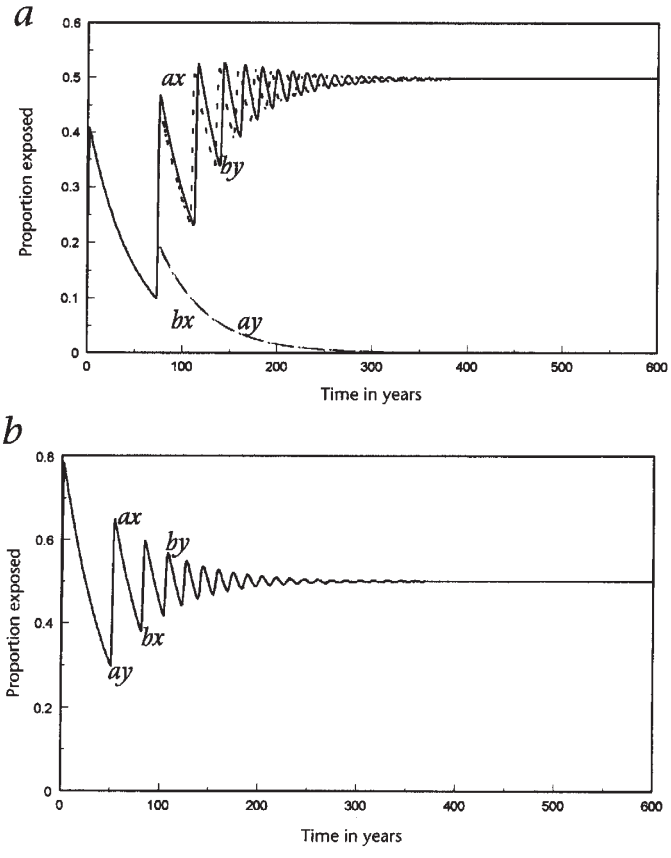

Fig. 2 The temporal changes in the proportions of host population exposed (z_i) to four strains: *ax* (heavy solid line), *by* (light dashes), *ay* (light dashes and dots) and *bx* (heavy dashed line), where the degree of cross-immunity between strains sharing alleles is variously (*a*) γ = 1, (*b*) γ = 0. The strains have the same R_0 (= 2), but the proportion initially infected varies slightly, that is, $y_{ax} = y_{bx} = y_{ay} = 0.0001$; $y_{by} = 0.0001001$.

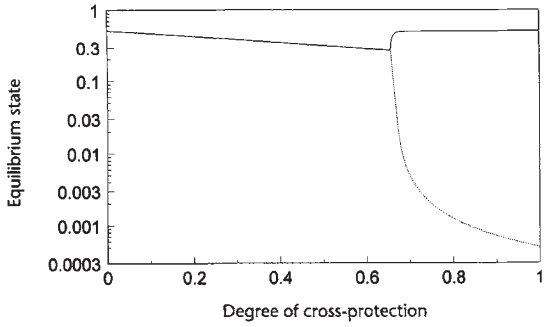

Fig. 3 The stable equilibrium state as a function of the degree of cross-immunity. Model parameters are identical to those used in Fig. 1. The switch from the symmetric equilibrium to the asymmetric equilibrium may be interpreted as a transition from linkage equilibrium to a high level of linkage disequilibrium.

one set of discordant strains (with nonoverlapping combinations of alleles) will dominate over all other strains. The subordinate groups will continue to be generated via recombination events, but they will persist only at very low prevalences, because of strong selection against them by hosts who are immune to the dominant strains. Despite the strains being defined by combinations of alleles at unlinked loci within this model structure, the results equally apply to the generation of novel combinations of epitopes through intragenic recombination and to the generation of novel sequences of antigenic variants through recombination in multigene families such as variant surface glycoproteins (VSG) of trypanosomes and variant surface antigens (VSA) of *Plasmodium falciparum*.

Neisseria meningitidis: A test of the model

We tested the model by examining the association between the two principal antigenic regions VR1 and VR2 of the outer membrane protein PorA of *Neisseria meningitidis* in a dataset[11] containing 222 isolates of serogroups B and C, collected in England and Wales in the period 1989–1991. The meningococcus is a transformable and antigenically diverse commensal bacterium that can cause severe infections, particularly in infants. Ninety-five percent of cases of disease are caused by organisms belonging to three (A, B, and C) of the nine known serogroups, which are defined by their capsular polysaccharide. Meningococci of serogroup A have a clonal population structure[12]; this is consistent with the epidemic life history of these bacteria, which ensures that they will have little opportunity for recombination with other meningococci[13,14]. In contrast to serogroup A, serogroup B and C meningococci do not have a strongly clonal population structure[15], although "clusters" and "complexes" of genetically related bacteria, which contain antigenically diverse strains generated by recombination, do occur in these serogroups[16]. The UK population experiences a high carriage rate of serogroup B and C meningococci of up to 10% of the population at any one time[17]. During carriage, many meningococci are acapsulate[18], thus subcapsular antigens are potentially exposed to immune selection. We have focused on the subcapsular antigen PorA, which has two principal antigenic regions VR1 and VR2 (ref. 19). It has been established that the human immune system is capable of mounting a bactericidal immune response to VR1 and VR2 epitopes[20], and that VR1 and VR2 families can be reassorted by horizontal genetic transfer[6,13]. The data in Fig. 5 show the distribution of VR1 and VR2 combinations in 222 endemic serogroup B and C meningococci isolated from diverse geographical locations in England and Wales during the period 1989–1991. The dataset shown in Fig. 5*a* records the association of 8 epitope families of VR1 with 14 epitope families of VR2 in the 222 isolates. Figure 5*b* shows, for comparison, the expected distribution under the null hypothesis of random association between the epitopes. As is clear from inspection, not only do the data depart significantly from the null

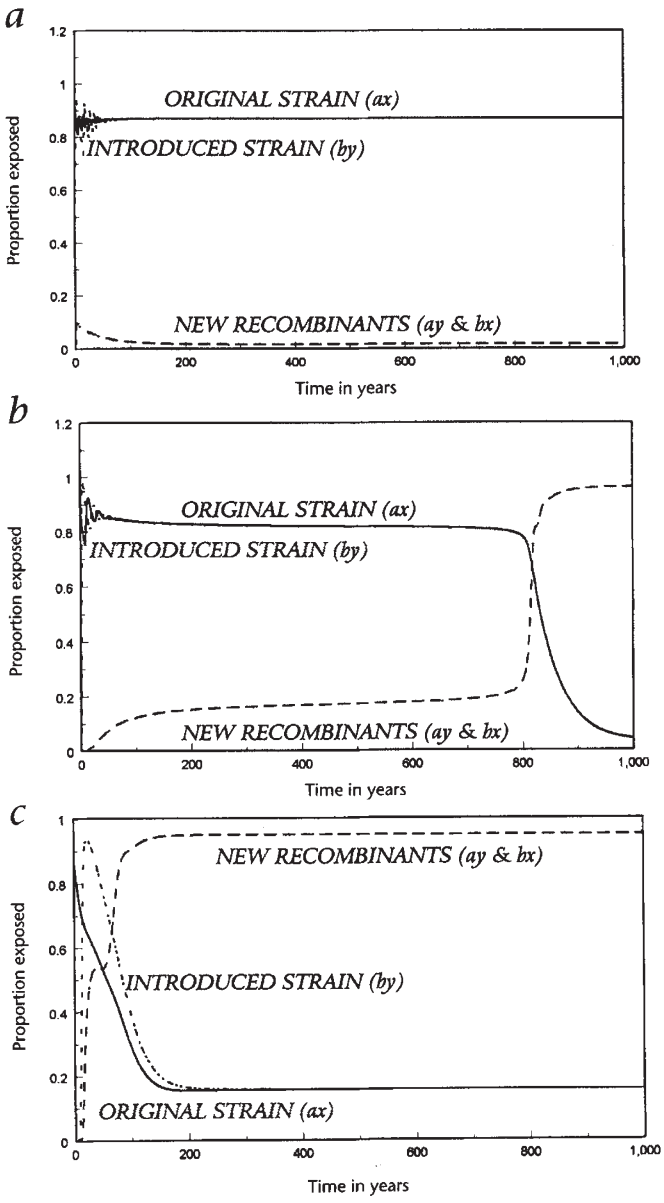

Fig. 4 The temporal changes in the proportions exposed (z_i) to four strains, in the situation where where a new strain *by* (represented by dots and dashes) is introduced within a host population that already harbors a single antigenically different strain *ax* (solid line). It is assumed, in this simulation, that the new recombinant types, *ay* and *bx* (dashed lines), have an R_0 of 28.5, in contrast to the R_0 values of the original (*ax*) and introduced (*by*) strains, which are assumed to be 8. Fig. 3 records the outcome of this genetic interaction for different values of the duration of infectiousness ($1/\sigma$), namely (*a*) 0.1 years (*b*) 1 year and (*c*) 10 years.

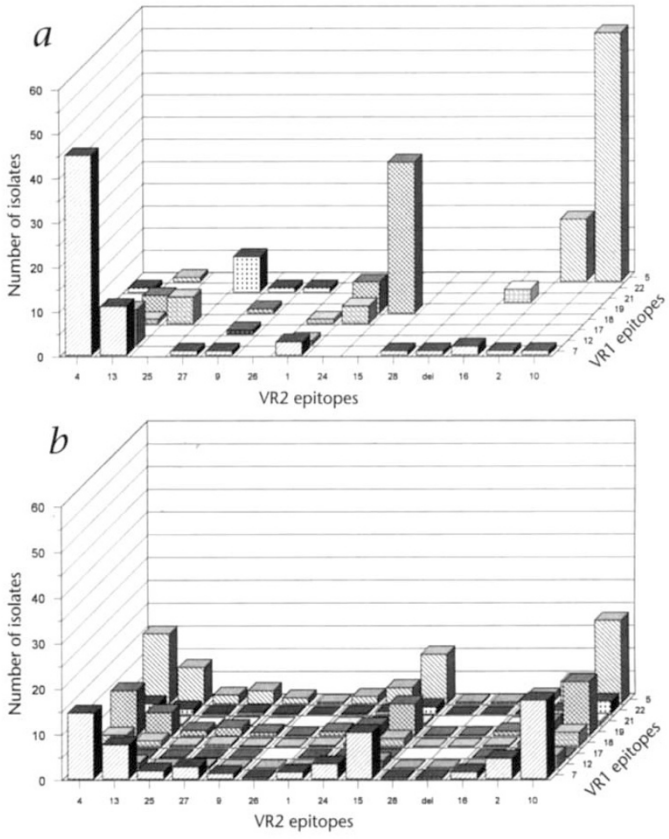

Fig. 5 *a*, The association between the two principal antigenic regions VR1 and VR2 of the outer membrane protein PorA of *Neisseria meningitidis* in a dataset[17] containing 222 isolates of serogroups B and C, collected in England and Wales in the period 1989–1991. The epitope families have been arranged on the axes to make clear the diagonal association. *b*, For comparison, the expected distribution under the null hypothesis of random association between the epitopes.

hypothesis ($\chi^2_{15} = 419$, $P << 0.01$), but they also have the diagonal structure, that is, nonoverlapping epitope combinations, predicted by the model. The data assume particular relevance when we recognize that, with more than two alleles at each locus, a nonrandom association can manifest itself in many ways other than the diagonal signature of strain structure.

The DNA sequences that encode the different VR families in both VR1 and VR2 are highly diverse and novel combinations are likely to arise only by recombination. There is much evidence for recombination events reassorting VR families[6] but only a few, nonoverlapping combinations persist at high frequency[11], as predicted by the model. Although it is conceivable that some combinations of the external surface loops containing these epitopes are not permitted because of protein structural requirements, it is unlikely that such arguments would account for the diagonal structure of the data. These data thus provide a clear illustration of the principles articulated by the model.

A new strategy for identifying dominant antigens

Nonrandom associations between immunogenic loci or epitopes will only occur when the degree of cross-protection, γ, is moderate to high (Fig. 3). This suggests that such nonrandom associations or "linkage disequilibrium" will only be observed among genes encoding the dominant polymorphic antigens, because the cost of sharing alleles at a given locus γ must correlate strongly with the efficacy of the associated immune response. Epitopes located in VRs of PorA proteins have been shown to elicit a T cell-dependent, anamnestic immune response[21]; that such an immune response may structure the parasite population into nonoverlapping repertoires therefore suggests that it has a significant effect on transmission.

This method may also be employed to discriminate between different polymorphic antigens in terms of their impact on parasite transmission, as our model predicts that subdominant polymorphic antigens will exist in a greater variety of mosaic forms than dominant antigens. This may explain why merozoite surface antigens (MSP-1 and MSP-2) of *Plasmodium falciparum* exist in forms that appear to have been generated by recombination[4,22], whereas epidemiological and experimental data suggest that there is limited overlap between the VSA repertoires of different strains[1,23]. Variant surface antigens of *P. falciparum* have been implicated by immunoepidemiological and molecular studies as the dominant polymorphic antigen that structures the parasite into independently transmitted strains[1]. As this molecule is known to undergo antigenic variation within the host[24], it has further been suggested that each *P. falciparum* strain is composed of a set of variants with a limited degree of overlap with the variant repertoires of other strains[23]. It has been argued that, in the absence of clonality, nonoverlapping variant repertoires can only persist if the dominant antigen is defined by a single or a tightly linked set of genes[25]. To the contrary, our analyses show that strain structure can survive under conditions of random mating even if the dominant polymorphic antigen is encoded by several unlinked genes. The recent discovery of the genes that encode the VSAs (ref. 26–28) should provide an opportunity to determine whether the latter do in fact elicit a significantly protective dominant immune response that structures *P. falciparum* into independently transmitted strains.

Discussion

We have shown how immune selection by the host can cause populations of infectious pathogens to self-organize into a stable collection of independently transmitted strains with nonoverlapping repertoires of dominant polymorphic determinants, despite the effects of recombination. Similar results are likely to be obtained by substituting mutation for recombination. We have assumed, in our analyses, that immunity to a given strain does not interfere with the generation of that strain through recombination events. This is certainly true for *P. falciparum*, where recombination takes place within the mosquito vector. However, in the case where the infectious progeny are also under selection by the host immune system, the cost of sharing alleles will increase, thus increasing the probability of the emergence of strain structure.

We have shown that two variable epitope regions within the PorA protein of *N. meningitidis* exist as nonoverlapping combinations as predicted by the model. A recent study of the kinetics of the *omp1* gene of *Chlamydia trachomatis* in two Gambian villages shows a similar pattern of nonoverlapping combinations of specific epitopes between genovars, despite evidence of genetic exchange[29].

The dynamics of a parasite population that is structured into strains by a dominant immune response may be influenced at other levels by weaker responses to both conserved and other polymorphic determinants. For instance, the typical age-distribution of *P. falciparum* prevalence suggests that although the VSA categorizes the population into strains, some degree of infection-

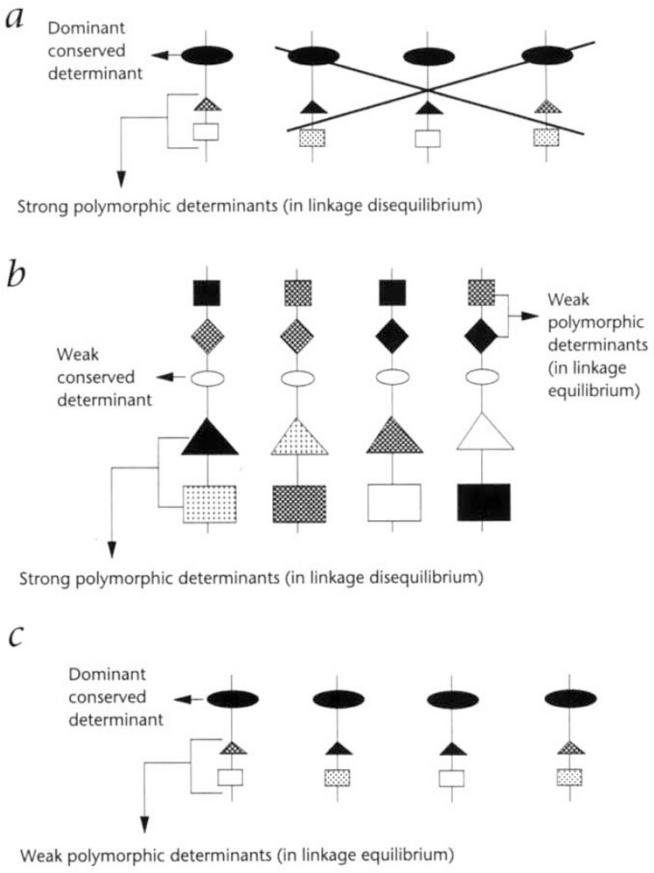

Fig. 6 The different types of pathogen population structures precipitated by (*a*) a strong conserved determinant acting in tandem with strong polymorphic determinants, (*b*) strong polymorphic determinants dominating over weak conserved and polymorphic determinants and (*c*) a strong conserved response dominating over weak polymorphic determinants.

blocking immunity to all strains is eventually established after repeated exposure to a conserved antigen[30]. A conserved response such as antibodies to the carboxy terminus of MSP-1 may reach densities at which merozoite invasion is blocked, as suggested by immunization experiments[31]. Similarly, immunity to sporozoite antigens[32,33] or cytotoxic T lymphocytes against liver stages[34] may, after many exposures, achieve densities that render such responses effective against infection by all strains. It is because these conserved responses are relatively weak that strong immune responses against the VSAs may be able to structure the parasite population into strains. A strong immune response against a conserved determinant would effectively reduce the system, through competitive exclusion, to a single transmission entity with the highest basic reproductive rate[35]. An example of this may be *Neisseria meningitidis* of serogroup A which has a strongly immunogenic capsular polysaccharide[36] in contrast to serogroup B (ref. 37) (which is why immunization with the former is effective, and with the latter is not). Long-lasting VR1 and VR2 associations have been shown in epidemic meningococci of serogroup A (ref. 12). This may be a consequence of the competitive exclusion of other strains (VR1-VR2 combinations) of serogroup A because of strong immunity against the shared capsular polysaccharide. For serogroup B meningococci, immunity against the capsular polysaccharide is too weak to result in the selection of a single, most transmissible, strain. Instead, immune selection occurs at the subcapsular level, resulting in nonoverlapping combinations of VR1-VR2 as discussed above. These two paradigms are illustrated in Fig. 6. Figure 6*a* schematically illustrates how competitive exclusion by a dominant conserved response can reduce the parasite population to a single strain. In Fig. 6*b*, the conserved determinant is weak in comparison with one set of polymorphic determinants, which are therefore able to structure the population into independently transmitted strains. These strongly immunogenic polymorphic determinants are thus in linkage disequilibrium, in contrast with other weak polymorphic determinants, which will be in linkage equilibrium. We have suggested that this property may be exploited as a strategy for identifying dominant antigens, provided that caution is exercised to eliminate other explanations for the occurrence of linkage disequilibrium such as structural constraints.

Finally, weak polymorphic determinants may exist alongside strong conserved determinants. A possible example is the hepatitis B virus (HBV), where a strong response against the *a* determinant is the primary structuring force in the pathogen population[38]. In contrast to the VR1 and VR2 epitope families of *N. meningitidis*, the polymorphic subdeterminants *d/y* and *w/r* of HBV do not exhibit linkage disequilibrium[39]. This suggests that the influence of these subdeterminants on the transmission of the virus is relatively low, such that the level of cross-protection (γ) between the subtypes *adw, adr, ayw* and *ayr* is not sufficiently large to counteract the diversifying force of recombination. That recombination does indeed occur in HBV has recently been demonstrated by Bollyky *et al.*[40]. Thus *d/y* and *w/r* of HBV form an example of weak polymorphic determinants in linkage equilibrium existing alongside a dominant conserved determinant *a*. Figure 6*c* schematically illustrates this scenario where all combinations of polymorphic determinants (in linkage equilibrium) function as a single transmissible entity or "strain" because of a dominant conserved determinant.

To date, analyses of linkage disequilibrium patterns within infectious pathogens have, for the most part, addressed the question of clonality[3,41]. Our theoretical framework suggests that such studies may also be employed to identify dominant, strain-structuring immune responses in the large number of pathogens that, despite recombination, appear to be transmitted as discrete strains.

Methods

The model. For simplicity (but without significant loss of generality), we consider the specific case of a sexually reproducing pathogen population where the developmental stage present in the human host, and subject to selection by immunological attack, is haploid. Each combination of alleles (for example, four possible combinations in the case of two loci, each with two alleles) represents a particular antigenic type. The model describes changes in the proportions of the host population who are, respectively, immune (z_i) and infectious (y_i) with respect to strain *i*. We assume, for simplicity, that an individual enters the "immune" and "infectious" categories simultaneously upon infection. We have assumed, since we are defining the strains by the strongest polymorphic determinants, that the "strain-specific" immunity generated by these determinants is lifelong, so that losses from the immune categories occur at the same rate as mortality:

$$dz_i/dt = \lambda_i (1 - z_i) - \mu z_i \quad (1)$$

$$dy_i/dt = \lambda_i (1 - z_i) [1 - \gamma(1 - \phi_i)] - \sigma y_i \quad (2)$$

Here the parameters $1/\mu$ and $1/\sigma$ record, respectively, the life expectancy of the host and the average duration of infectiousness; $\phi_i = \Pi(1 - z_j)$ with j denoting any strain that shares alleles one or more alleles with i; λ_i is the per capita rate of infection for a given strain; and the parameter γ measures the degree to which infection with a given strain limits the transmission of strains that share any of its alleles. Following conventional lines of analysis[8], the basic reproductive number of strain i, R_0^i, is simply β_i/σ, where β_i is a combination of parameters affecting the transmission of strain i ($\lambda_i \sim \beta_i$).

The force of infection, λ_i, can be explicitly formulated as a linear combination of (the probability of a set of strains coinfecting a host) $\times$ (the fraction of the resultant progeny that is likely to be of genotype i). For instance, in the two-locus/two-allele case, when $\gamma = 1$, $\lambda = \beta_i [y_i (1 - y_{i'}) + (\alpha/4)(y_i y_{i'} + y_j y_{j'})]$, where i' is the discordant of i (that is, shares no alleles), whereas j and j' are the other discordant pair (each sharing one allele with i). The parameter α indicates the degree to which the transmission capacity of a host is affected by infection with two strains as compared with a singly infected host (if the transmission capacity is unaffected, $\alpha = 1$; if it is doubled, $\alpha = 2$; and so on). When $\gamma = 1$, coinfection is only possible with i and i' or with j and j'. When $\gamma < 1$, all possible combinations of i, i', j and j' (whose probabilities are in turn conditional on γ) must be considered; the resulting expression for λ_i is straightforward but lengthy. In the limit $\mu/\sigma << 1$, however, the latter simplifies, yielding the following solutions:

(1) All $y_i \sim O\mu$, whence $\lambda_i = \beta_i y_i (1 + O\mu)$, $\nexists i$. As illustrated by Fig. 1 and 2, such a solution becomes unstable at a critical value of γ, $\hat{\gamma}$, determined by the relative R_0 values.

(2) For one discordant pair, $y_{i,i'} \sim O\mu$, whereas for the other pair, $y_{j,j'} \sim O\mu^2$. In this case $\lambda_i = \beta_i y_i (1 + O\mu)$ with a similar expression for $\lambda_{i'}$. For λ_j and $\lambda_{j'}$, however, we have $\lambda_j = \beta_j (y_j + (\alpha/4) y_i y_{i'}) (1 + O\mu)$. These solutions can exist only if $\gamma > \hat{\gamma}$.

(3) Solutions with all but one $y_i \sim O\mu$ and one with $y_k \sim O\mu^2$ can be obtained but they are not stable.

These results can be extended in a straightforward way to m loci, each with two alleles. (Note: The symbol O indicates "of the order".)

Acknowledgments

We are very grateful to Brian Spratt for his contribution to this work. We would like to thank Eddie Holmes and Adrian Hill for constructive comments. S.G. and R.M.A. thank the Wellcome Trust, R.M.M. thanks the Royal Society, S.N. thanks the B.B.S.R.C., and M.C.J.M. and I.M.F. thank the British Society for the Study of Infection for financial support.

RECEIVED 23 OCTOBER 1995; ACCEPTED 20 FEBRUARY 1996

1. Gupta, S., Trenholme, K., Anderson, R.M. & Day, K.P. Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science* **263**, 961–963 (1994).
2. Louwagie, J. *et al.* Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7**, 769–780 (1993).
3. Maynard Smith, J., Dowson, C.G. & Spratt, B.G. Localized sex in bacteria. *Nature* **349**, 29–31 (1991)
4. Walliker, D. Genetic recombination in malaria parasites. *Exp. Parasitol.* **69**, 303–309 (1989).
5. Robertson, D.L., Sharp, P.M., McCutchan, F.E. & Hahn, B.H. Recombination in HIV-1. *Nature* **374**, 124–126 (1995).
6. Feavers, I.M., Heath, A.B., Bygraves, J.A. & Maiden, M.C.J. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Mol. Microbiol.* **6**, 489–495 (1992)
7. Desjardins, P., Picard, B., Kaltenbock, B., Elion, J. & Denamur, E. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: Evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J. Mol. Evol.* **41**, 440–448 (1995).
8. Levin, B. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* **99**, 1–23 (1981).
9. Hartl, D. & Clark, A.G. *Principles of Population Genetics* (Sinauer Associates, Sunderland, Massachusetts, 1989).
10. Anderson, R.M. & May, R.M. *Infectious Diseases of Humans: Dynamics & Control* (Oxford Univ. Press, Oxford, 1991).
11. Feavers, I.M.*et al.* Antigenic diversity of meningococcal outer membrane protein PorA has implications for epidemiological analysis and vaccine design *Clin. Diagn. Lab. Immunol.* (in the press).
12. Suker, J. *et al.* The porA gene in serogroup A meningococci: Evolutionary stability and mechanism of genetic variation. *Mol. Microbiol.* **12**, 253–265 (1994).
13. Maiden, M.C.J. Population genetics of a transformable bacterium: The influence of horizontal genetic exchange on the biology of Neisseria meningitidis. *FEMS Microbiol. Lett.* **112**, 243–250 (1993).
14. Maiden, M.C.J. & Feavers, I.M. Population genetics and global epidemiology of the human pathogen *Neisseria meningitidis*. in *Population Genetics of Bacteria* (eds. Baumberg, S., Young, J.P.W., Saunders, J.R. & Wellington, E.M.H.) 269–293 (Cambridge Univ. Press, Cambridge, 1995).
15. Maynard Smith, J.M., Smith, N.H., O'Rourke, M. & Spratt, B.G. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**, 4384–4388 (1993).
16. Spratt, B.G., Smith N.H., Zhou J., O'Rourke M. & Feil, E. The population genetics of the pathogenic *Neisseria*. in *Population Genetics of Bacteria* (eds. Baumberg, S., Young, J.P.W., Saunders, J.R. & Wellington, E.M.H.) 143 (Cambridge Univ. Press, Cambridge, 1995).
17. Cartwright, K.A.V., Stuart, J.M., Jones, D.M. & Noah, N.D. The Stonehouse survey: Nasopharyngeal carriage of meningococci and *Neisseria lactamica*. *Epidemiol. Infect.* **99**, 591–601 (1987).
18. Craven, D.E. *et al.* Adherence of isolates of *Neisseria meningitidis* from patients and carriers to human buccal epithelial cells. *J. Infect. Dis.* **142**, 556–568 (1980).
19. Maiden, M.C.J., Suker, J., McKenna, A.J., Bygraves, J.A. & Feavers, I.M. Comparison of the class 1 outer membrane proteins of eight serological reference strains of *Neisseria meningitidis*. *Mol. Microbiol.* **5**, 727–736 (1991).
20. Mandrell, R.E. & Zollinger W.D. Human immune response to meningococcal outer membrane protein epitopes after natural infection or vaccination. *Infect. Immun.* **57**, 1590–1598 (1989).
21. Wiertz, E.J. *et al.* T cell recognition of *Neisseria meningitidis* class 1 outer membrane proteins. Identification of T cell epitopes with selected synthetic peptides and determination of HLA restriction elements. *J. Immunol.* **147**, 2012–2018 (1991).
22. Hughes, A.L. Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. *Mol. Biol. Evol.* **9**, 381–393 (1992).
23. Gupta, S. & Day, K.P. A strain theory of malarial transmission. *Parasitol. Today* **10**, 476–481 (1994).
24. Roberts D.J. *et al.* Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357**, 689–692 (1992).
25. Tibayrenc, M. Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*: The clonality/sexuality debate revisited. *Parasitol. Today* **10**, 456–457 (1994).
26. Baruch, D.I. *et al.* Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).
27. Su, X-z. *et al.* The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100 (1995).
28. Smith, J.D. *et al.* Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
29. Hayes, L.J. *et al.* Extent and kinetics of genetic change in the *omp1* gene of *Chlamydia trachomatis* in two villages with endemic trachoma. *J. Infect. Dis.* **172**, 268–272 (1995).
30. Gupta, S. & Day, K.P. A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasite Immunol.* **16**, 361–370 (1994).
31. Daly, T.M. & Long, C.A. A recombinant 15-kilodalton carboxyl-terminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein 1 induces a protective immune response in mice. *Infect. Immun.* **61**, 2462–2467 (1993) .
32. Hoffman, S. *et al.* Naturally acquired antibodies to sporozoites do not prevent malaria: Vaccine development implications. *Science* **237**, 639–642 (1987).
33. Jones, T.R., Ballou, W.R. & Hoffman, S.L. Antibodies to the circumsporozoite protein and protective immunity to malaria sporozoites. *Prog. Clin. Parasitol.* **3**, 103–117 (1993).
34. Hill, A.V.S. *et al.* Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* **360**, 434–439 (1992).
35. Gupta, S., Swinton, J. & Anderson, R.M. Theoretical studies of the effects of genetic heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. Series B* **256**, 231–238 (1994).
36. Gotschlich, E.C., Goldschneider, I. & Artenstein, M.S. Human immunity to the meningococcus. IV. Immunogenicity of group A and group C meningococcal polysaccharides in human volunteers. *J. Exp. Med.* **129**, 1367–1384 (1969).
37. Wyle, F.A. *et al.* Immunologic response of man to group B meningococcal polysaccharide vaccines. *J. Infect. Dis.* **126**, 514–521 (1972).
38. Carman, W.F., Thomas, H.C., Zuckerman, A.J. & Harrison, T. Hepatitis B virus: Molecular variants. in *Viral Hepatitis* (eds. Zuckerman, A.J. & Thomas, H.C.) 115–136 (Churchill Livingstone, Edinburgh, 1993)
39. Norder, H. *et al.* Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. *J. Gen. Virol.* **74**, 1341–1348 (1993).
40. Bollyky, P.L., Rambaut A., Harvey, P.H. & Holmes, E.C. Recombination between sequences of Hepatitis B Virus from different genotypes. *J. Mol. Evol.* (in the press).
41. Tibayrenc, M., Kjellberg, F. & Ayala, F.J. A clonal theory of parasitic protozoa: The population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences *Proc. Natl. Acad. Sci. USA* **87**, 2414-2418 (1990).