The Interface Between Epidemiology and Population Genetics

S. Paterson and M.E. Viney

Modern biology increasingly integrates disparate disciplines. Here, Steve Paterson and Mark Viney examine the interface between epidemiology and population genetics. They argue that infection and inheritance can be considered as analogous processes, and that epidemiology and population genetics share many common features. They consider the potential for existing population genetic theory to dissect epidemiological patterns in field studies and they consider other relationships between genetics and epidemiology that provide a research challenge for the future.

Epidemiology is the study of disease dynamics within a population. Current models of infection dynamics attempt to describe the process of infection from one host to another and the consequence of this process on host and parasite populations. Population genetics is concerned with the inheritance of genes at the population level. There is, therefore, a clear analogy between inheritance and infection. The transmission of genes from one generation to the next and infection – the transmission of parasites from one host to another – and the population-level consequences of each of these processes. Inheritance and infection both occur within populations and, in this respect, both population genetics and epidemiology are concerned with problems of scale; specifically, to extend a basic biological process (inheritance or infection) that occurs at the individual level to the population-level consequence of that process.

The interface of inheritance and infection

Modeling drug resistance in parasite populations is one area where population genetics and epidemiology come together. The task is to construct a model that combines both the spread of drug-resistance genes through a parasite population and the spread of drug-resistant parasites through a host population (Box 1). Recent work by Smith et al. provides a good example of how this can be achieved. Here, the basic unit of the model is the parasite itself, rather than the infected host. Parasites of different anthelmintic genotypes (eg RR, Rr or rr, where R is the anthelmintic-resistance allele) were modeled deterministically within a host population that was subjected to various anthelmintic dosing regimens. In each parasite generation, adult parasites mate and produce progeny whose genotype frequencies are determined by Hardy–Weinberg processes from the allele frequencies of the parents. This mating function allowed the frequency of the anthelmintic-resistant and -sensitive alleles to be followed within the parasite population. One important prediction from this model was that there was little difference in the rate of spread of anthelmintic-resistant parasites under chemoprophylactic or chemotherapeutic dosing regimens. Therefore, this model directly uses population genetic and epidemiological modeling to understand the movement of anthelmintic parasites through a host population subject to different dosing regimens and, thus, is powerfully predictive.

Steve Paterson and Mark Viney are at the School of Biological Sciences, University of Bristol, Woodland Road, Bristol, UK. BS8 1UG. Tel: +44 117 928 7470, Fax: +44 117 925 7374, e-mail: S.Paterson@bristol.ac.uk
Transmission and gene flow

Parasite transmission at the individual scale describes the infection of one host by another. Although this is the scale at which it is often considered, this process also extends to the population level and, at the largest scale, can describe the spread of a parasite species, strain or genotype across a continent. The concept of gene flow in population genetics describes the movement of genes through or between populations.

In this respect, it is analogous to parasite transmission at the population level. Just as parasite transmission at this scale arises from a combination of individual infections, gene flow arises from a combination of the mendelian inheritance of genes at the individual scale. Thus, population genetics can help build theoretical models of infectious disease and can also be used empirically to understand patterns of parasite transmission in the field.

The measurement of transmission rates in the field is difficult. The ability to make such measurements can be used to predict the spread of an epidemic or the effectiveness of a vaccine or other therapy at the level of a community, country or continent. The central problem is determining the origin of a new infection because a naive individual might be exposed to many carriers. An additional problem is that hosts can be superinfected, and it can be difficult to know when a host acquires a new infection, let alone the origin of that infection.

A common form of population genetic analysis is to partition genetic variation within a hierarchical structure, i.e. to assay genetic variation at the level of the individual – the ‘subpopulation’ and the total population. This approach has been used successfully in several field studies using genetic markers. It not only provides a description of the structure of genetic variation found within a parasite population, but also gives estimates of the rates of gene flow in the population and, hence, transmission through the host population. A good example of this style of analysis is the work by Blouin et al. Here, genetic variation of trichostrongylid nematodes within populations of white-tailed deer and domestic cattle across the USA was compared. Greater genetic structuring was found within the parasite population of deer compared with that of cattle. It was concluded that this resulted from the higher rates of gene flow in the parasites of cattle which was, in turn, caused by the movement of cattle by humans. Conversely, the lower rates of gene flow in parasite populations of deer was thought likely to reflect the fact that populations of wild deer showed less geographical movement. Another example is a study by Anderson et al., which determined that there was little gene flow between the Ascaris populations of pigs and humans in Guatemala, even though infected pigs and humans could both be found in the same community. The authors concluded that two populations of Ascaris exist, which are separated by host preference.

As parasites are also compartmentalized within hosts, could partitioning genetic variation be used to measure between-individual transmission dynamics between hosts? Unfortunately this is not the case: genetic differentiation and, hence the power of this analysis, is quickly lost except under very low rates of gene flow. Thus, this style of analysis appears best suited to assaying genetic structure over large geographical areas or between parasite populations, as in the examples above. The assay used to analyse genetic variation must be appropriate to the scale of the study. A successful approach to dissecting microepidemiological patterns has been used to determine the transmission of HIV. As HIV replicates, mutations accumulate in its genome, and this genetic variation can be identified by sequencing. From sequences of HIV isolates taken from infected individuals in the UK and Ireland, this genetic variation can be used to construct a phylogenetic tree that mirrors the spread of the virus through the population. This work identifies distinct lineages of HIV grouped within haemophiliacs and within intravenous drug users.
of an underlying process at the individual level to the population-level consequences of that process? A particular challenge common to both epidemiology and population genetics is the introduction of heterogeneity into the underlying process – infection or inheritance – and its consequence upon population-level dynamics.

The use of genetic markers in field studies has highlighted the importance of population structure within host populations. This structure leads to heterogeneity in between-individual transmission rates. The simplest models of micro- and macroparasitic infection assume that individuals of the host population mix freely, and that infection of an individual might be the result of a contact with any other individual in that population. In reality, a host population might be distributed spatially and temporally and might be split into demographic groups. Infection is, therefore, most likely to occur between spatial neighbours or between individuals in the same demographic group. A current focus of epidemiology is to model spatial and demographic heterogeneity, with the aim of understanding the fundamental processes underlying infection dynamics and of providing a framework for evaluating potential control strategies for infectious disease. The difficulty is how to model the population-level dynamics of heterogeneous populations accurately without retaining the details of every member of the population.

To do this, it is crucial that theoretical models are directed by observations from the field. A good example of modeling demographic heterogeneity comes from measles. This is an excellent test-bed for many models of epidemics because, as a notifiable disease, accurate records exist of every case of measles since World War II. By considering the biology of the system, the importance of social and geographical structure in the contact network has become apparent, i.e. heterogeneity in transmission rates within families, between children at the same school and between communities. The estimation of small-scale transmission patterns from field data is an essential first step to understanding the influence of other contact networks on the movement of infection through a population.

Transmission heterogeneity

Many of the theoretical challenges faced by epidemiologists and population geneticists are problems of scale. How does one go from a mechanistic explanation whereas homosexuals show a more diverse origin of HIV. At the finest scale, this approach has identified the transmission of HIV from one individual to another – in one notable case showing infection of five individuals by a dentist. The estimation of small-scale transmission patterns from field data is an essential first step to understanding the influence of other contact networks on the movement of infection through a population.

Transmission heterogeneity

Many of the theoretical challenges faced by epidemiologists and population geneticists are problems of scale. How does one go from a mechanistic explanation...
the population can ultimately be linked to every other via a contact network. Several possible contact networks are shown in Box 2. An extremely regular system of contacts (Box 2 Fig. 1a) is biologically unrealistic. Empirical evidence from studies of HIV transmission highlights the importance of structure within contact networks, with the number of contacts between individuals varying through space and time. Box 2 Fig. 1b–d gives alternative schemes that attempt to capture features of spatial heterogeneity in contact networks16,19–21.

Modeling spatial heterogeneity is a topic where epidemiologists have sought inspiration from outside the traditional ecological literature. In this regard, the physical sciences have provided some interesting approaches. One problem from physics concerns predicting the large-scale properties of individual interacting elements, such as grains of sand in a sandpile or the alignment of magnetic moments in ‘spin glasses’ (an otherwise unappalling group of metal alloys)22. A successful approach has been to set out a lattice of cellular automata, each of which acts according to a set of simple probabilistic rules depending upon the state of its neighbours. Rhodes and Anderson used this method to describe measles epidemics23. They constructed a lattice consisting of sites that were either unoccupied, occupied by uninfected individuals or occupied by infected individuals (Box 2 Fig. 1c). Infected individuals could infect their neighbours with a certain probability. They were able to show that the size of epidemics, in both simulated studies and using data from the Faroe Islands, followed a characteristic power law distribution; i.e., a distribution that shows an inverse straight-line relationship between the log of the frequency of an event and the log of the size of that event, in this case epidemic size. Such power laws have been found in a wide range of natural phenomena, such as earthquakes and extinctions in the fossil record, and are highly indicative of complex, dynamic systems poised between stasis and chaos24.

Another interesting example of spatial modeling of disease epidemics comes from the ‘small-world’ networks of Watts and Strogatz, which attempt to mimic social interactions20. An example of such a network is shown in Box 2 Fig. 1d. Interactions are most likely between nearest neighbours, but occasional long-range interactions also occur. These long-range interactions reduce the average distance between any two individuals within the network, while the fact that most interactions are still with near neighbours maintains the spatial clustering of the network. They showed that an epidemic can move through such a network more easily than through a regular lattice. Heterogeneity in infection patterns is a biological reality and must be incorporated into epidemiological models. As outlined above, there are several approaches to modeling heterogeneity in disease transmission, and the approach used will depend upon both the disease and the questions one wants to ask. However, the crucial test of any model is how well it explains observed patterns of data. In the future, these models will not only have to explain the incidence of disease (as for the measles examples above), but also the observed patterns of genetic variation exhibited by genetic markers in field studies. As these epidemiological models succeed in explaining the population genetics of parasites, their power as predictive tools of infectious disease will increase enormously.

**Emergent properties and epistasis**

Spatial modeling in epidemiology has highlighted the importance of interactions between individuals for infection dynamics. Interactions between genes might play just as an important role in population genetics. Genetically complex traits such as weight, height, etc. are governed not by a single locus, but by a large number of loci. How do these loci act to produce a complex trait? In the simplest case, the effect of each locus is additive (Fig. 1a) – the sum of the effects of all the loci gives the expected value of the trait25. In this case, the population-level dynamics, i.e., the response of the trait and of the allele frequencies to selection, are well understood. The situation is considerably more complicated if the loci governing a complex trait interact with each other (Fig. 1b). This is called epistasis. Here, the value of the trait is a complicated function of all loci, reflecting the number and strength of interactions among all of the loci24. The population-level consequences of epistatic interactions at just two or three loci are still poorly understood, but it is a pressing concern given current genome research, which offers the potential to investigate the genetic architecture underlying complex traits23,25,26.

The root of the problem of understanding epistatic interactions is somewhat similar to the problem of modeling spatial dynamics in epidemiology. In the case of epistasis, genes act as part of a network of interacting elements. In the case of epidemiology, individuals infect each other within a contact network. The problem is one of scale. The evolution of an entire genome or the temporal dynamics of an epidemic is likely to be an emergent property of the system that is not readily apparent from the small-scale process of inheritance or infection22,27,28. Understanding how such emergent properties can arise is, therefore, a challenge common to both epidemiology and population genetics.

**Infection and disease**

Epidemiology studies the population dynamics of disease; however, infection *per se* is not equivalent to
reviews

disease. Infection is the presence of infectious agents in a host; disease is the deviation from normal health. Thus, disease is a possible, but not a necessary, outcome of infection. Consider hookworm infection. To understand hookworm disease, it is not enough to know whether a patient is infected. Rather, one also needs to know how many hookworms are present within a patient. An infection of 50 worms can cause mild symptoms, whereas an infection of 1000 worms will result in severe symptoms. The state of an individual’s immunity is also an important factor in the progression of a disease. With malaria infection, individuals with clinical immunity can have a malaria infection but have no disease or pathology. In contrast, those without such immunity will have disease and pathology for similar levels of infection.

Therefore, an epidemiological model should follow not only the progress of an infection, but also the possible consequences of infection, namely disease. To some extent, the impact of infection is considered in epidemiological models that monitor the immune status of a host or that consider several different parasite strains. In practice, the disease associated with the infection is considered in epidemiological models that monitor the immune status of a host or that consider several different parasite strains. In practice, the disease associated with the infection is considered in epidemiological models that monitor the immune status of a host or that consider several different parasite strains. In practice, the disease associated with the infection is considered in epidemiological models that monitor the immune status of a host or that consider several different parasite strains.

Working at the interface

What benefits will result from work at the interface of population genetics and epidemiology? First, population genetic and epidemiological models are now becoming integrated, especially in models of anthelmintic resistance. This is a trend to be welcomed because it helps to produce predictive models for use in parasite control programmes. Second, field studies using molecular markers and population genetic theory should encourage the development of epidemiological models that provide a deeper understanding of the impact of heterogeneity on the processes of transmission. Third, genetics, epidemiology and, indeed, disease all show or have the potential for emergent properties. These properties are not yet understood and are an important area of current and future research.

References

12 Paterson, S. et al. (2000) Inferring infection processes of a parasitic nematode using population genetics. Parasitology 120, 185–194

Call for Papers


Abstracts are due 15 December 2000

Parasitology Today, vol. 16, no. 12, 2000