
http://researchrepository.murdoch.edu.au/7069/
The molecular epidemiology of parasite infections: tools and applications

AJ Lymbery\textsuperscript{a,b*}, RCA Thompson\textsuperscript{b}

\textsuperscript{a}Fish Health Unit and \textsuperscript{b}WHO Collaborating Centre for the Molecular Epidemiology of Parasitic Infections, School of Veterinary and Biomedical Sciences, Murdoch University, Murdoch WA 6150, Australia

*Corresponding author. Tel +64 8 93602729; Fax +64 8 9360 7512.
E-mail address: a.lymbery@murdoch.edu.au

ABSTRACT

Molecular epidemiology, broadly defined, is the application of molecular genetic techniques to the dynamics of disease in a population. In this review, we briefly describe molecular and analytical tools available for molecular epidemiological studies and then provide an overview of how they can be applied to better understand parasitic disease. A range of new molecular tools have been developed in recent years, allowing for the direct examination of parasites from clinical or environmental samples, and providing access to relatively cheap, rapid, high throughput molecular assays. At the same time, new analytical approaches, in particular those derived from coalescent theory, have been developed to provide more robust estimates of evolutionary processes and demographic parameters from multilocus, genotypic data. To date, the primary application of molecular epidemiology has been to provide specific and sensitive identification of parasites and to resolve taxonomic issues, particularly at the species level and below. Population genetic studies have also been used to determine the extent of genetic diversity among populations of parasites and the degree to which this diversity is associated with different host cycles or epidemiologically important phenotypes. Many of these studies have also shed new light on transmission cycles of parasites, particularly the extent to which zoonotic transmission occurs, and on the prevalence and importance of mixed infections with different parasite species or intraspecific variants (polyparasitism). A major challenge, and one which is now being addressed by an increasing number of studies, is to find and utilise genetic markers for complex traits of epidemiological significance, such as drug resistance, zoonotic potential and virulence.

Keywords: molecular epidemiology; parasite identification; species delimitation; transmission; genetic markers
1. What is molecular epidemiology?

Epidemiology is the study of the causation and dynamics of disease in a population. For parasitic diseases, this is determined by the transmission of the parasite between hosts, and how this transmission affects the dispersal of the parasite within and among host populations [1]. To control parasitic disease, therefore, we need to understand parasite ecology, particularly transmission dynamics, how life cycles may interact and the nature of interactions within the host. This requires an input from both population and evolutionary biology, to determine, for example, the genetic structure and evolution of infectious agents, their population biology, and the evolutionary consequences of medical and public health interventions [2].

Traditionally, both epidemiology and parasite ecology have concentrated on an empirical approach. Epidemiological studies typically begin with a description of the frequency and distribution of disease and then attempt to associate these patterns with the frequency and distribution of independent variables or risk factors. Identifying risk factors is important because it allows for targeted control programs, but the efficacy of such control programs hinges upon knowing how the risk factors interact with the parasite’s life cycle to increase exposure. Ecological studies of parasite life cycles usually start with a description of parasite prevalence, and sometimes also intensity of infection, within different host species. These data, often accompanied by in vitro and experimental infection studies, can then be used to infer the major pathways of parasite transmission. In the last 20 years, these empirical studies of the ecology of parasite life cycles and the epidemiology of parasitic disease have been complemented by a more theoretical approach, which uses mathematical models of parasite and host population sizes to guide epidemiological interpretation [e.g.3,4,5].

The application of molecular and analytical tools, derived largely from the fields of population genetics and systematics, can contribute enormously to both empirical and theoretical studies of the epidemiology of parasitic disease. Molecular epidemiological approaches enable the reconstruction of evolutionary relationships between parasites over a wide range of temporal and spatial scales, improving our ability to identify parasites, track their movements, relate their spread to environmental factors and understand the role they play in disease causation [6].

In this review, we briefly describe the range of molecular and analytical tools available for molecular epidemiological studies and then provide an overview of how they can be applied to better understand the causation and dynamics of parasitic diseases. Tibayrenc [7,8] suggested that molecular epidemiological studies of parasitic diseases could be classified into two different types, depending on whether they were concerned purely with identification of the causative agents of disease, or whether they considered the impact of genetic variation on “downstream functions”, such as transmission, infectivity, virulence or drug resistance. To date, the main applications of molecular techniques have been in parasite identification rather than to study patterns of disease progression or transmission. This has been due partly to the inability of molecular tools to distinguish
genetic variation at the appropriate level of resolution for addressing downstream function and partly to the inadequacy of analytical methods to interpret genetic variation in an ecologically meaningful fashion. Rapid advances in both these areas mean that that an increasing number of molecular epidemiological studies are addressing questions of function, although parasite identification remains a critical issue. From a practical perspective, genotyping the agents should not become a dominant aim of molecular epidemiological investigations, since the existence of different genotypes does not imply they necessarily have some phenotypic importance. The scope and potential of molecular epidemiology is much greater, and in this regard the search for genetic markers for ‘medically’ important traits such as infectivity, drug resistance and virulence present important challenges for molecular epidemiological investigations.

2. Molecular tools

Our ability, using molecular techniques, to detect and characterise the genetic variability of infectious agents, particularly at the intraspecific level, can be seen as the foundation for most molecular epidemiological studies [9]. The application of appropriate molecular tools will aid in the identification and surveillance of infectious agents and in determining sources of infection. The availability of such tools, particularly those based on the polymerase chain reaction (PCR), which allow direct examination of clinical or environmental isolates, has had an enormous impact on the genetic characterisation, diagnosis and taxonomy of parasites. They also obviate the need for laboratory amplification of parasite isolates, which was a major limiting factor in characterising parasites refractory to in vitro culture, and may lead to biased sampling of natural diversity by the selective amplification of those parasites amenable to culture.

Using PCR, defined gene sequences of infectious agents can be detected from small quantities of material and the resultant data can be used not only for diagnosis, but also to assess the effect of interventions on the population structure of infectious agents, assessment of intraspecies diversity, and transmission studies. The value of such tools is greatest if they can be applied directly to faecal or tissue specimens, as well as environmental samples, and if there is the potential to automate such procedures. Table 1 summarises the available molecular tools and their application. Emphasis will be given in the future to establishing high throughput molecular assays such as pyrosequencing, as well as their field applicability. Pyrosequencing techniques have the added advantage of allowing the simultaneous detection of multiple species/genotypes in a single sample [10,11,12]. Multiplex PCR (mPCR) also enables the amplification of more than one target of interest in a PCR by using multiple primer pairs and producing amplicons of different size [13]. Loop-mediated isothermal DNA amplification (LAMP) is a newly developed, rapid, quantitative, highly sensitive and specific nucleic acid-based, non-PCR diagnostic tool [14,15,16], applicable to ‘low-cost’ laboratory settings. This simple molecular test can be carried out on a bench with a heating block instead of a thermal cycler and may prove to be an invaluable ‘field friendly’ tool for screening and quantifying infections in host populations while providing important genotypic information [17].
Choosing an appropriate marker for molecular epidemiological studies requires consideration of the required level of resolution for the study, the precision of the genetic data collected and the historical information content of the data.

Genetic markers, although identified in individuals, are influenced by processes which are more readily measured at the level of populations, such as mode of reproduction, breeding system, mutation, migration and selection. Population level processes eventually influence speciation and thus all cladogenetic events in the history of a lineage. Therefore, by using genetic markers with appropriate rates of change, we should be able to examine evolutionary patterns and processes at all levels throughout the hierarchy of life, from individuals to kingdoms. In this context, emphasis has been given to the importance of appropriate analysis and the value of characterising the genetic diversity of infectious agents at different levels of specificity [18,19,20]. The latter requires choosing molecular tools which are capable of discriminating genetic variants at different hierarchical levels and the region of DNA examined must be appropriate to the level of questions being asked [13,20,21,22,23]: e.g. taxonomy, diagnosis, population genetics, evolutionary relationships, isolate tracking etc. (Table 1). This is primarily a question of choosing a genomic region with an appropriate signal to noise ratio; too little variation will provide a signal which is too weak to discriminate among groups, whereas too much variation will swamp the signal with uninformative noise.

The choice of genetic marker typically involves a trade off between technical convenience and precision. Markers such as RAPDs and AFLPs do not require specific sequence information from the target genome, and hence can be utilized more readily for less well studied species. The variation they detect, however, may be non-heritable and even when heritable is dominant rather than co-dominant. This means that alternative alleles at a locus cannot be distinguished, greatly reducing the range and power of analytical techniques which can be applied to the resultant data. It also means that the data cannot be compared effectively over different studies, and are therefore limited in usefulness to the particular time and place where they were collected.

Traditional genetic markers, such as allozymes and RFLPs have little historical information content. That is, we do not usually know the phylogenetic relationship between alternative alleles or haplotypes and the data are therefore analysed as allele or haplotype frequency differences among groups. Sequence data, however, do provide historical information because the phylogeny of sequences can usually be inferred. This enables sequence data to be analysed in ways which are not possible for allele frequency data.

3. Analytical tools

Concomitant with the development of new genetic markers and the ability to rapidly genotype large numbers of genetic loci has been the development of new analytical tools to interpret these multilocus genotypes, and a blurring of the boundaries between
population genetic and systematic analyses. Traditionally, these fields have been quite distinct in their analytical approaches. Population genetics aims to describe and understand the processes underlying the distribution of genetic variation within and among populations of the same species, while systematics aims to describe and organize the pattern of evolutionary relationships among species and higher taxa. For sexually reproducing organisms, evolutionary relationships above and below the species level are quite different in nature [24]. Below the species level, relationships between genes sampled from different individuals are not hierarchical because homologous genes from the two parents combine in their offspring. Above the species level, however, relationships between genes sampled from different taxa are hierarchical because they are a consequence of speciation followed by long periods of reproductive isolation. Methods developed for inferring phylogenetic relationships above the species level rely on assumptions that are often violated by the reticulate relationships between individuals below the species level. Population genetic analyses, therefore, have not traditionally considered genealogical relationships among genes or among individuals, and systematic analyses have traditionally ignored the possibility of reticulate relationships. This, however, is now changing, driven in part by the availability of DNA sequences and other molecular markers with historical information content and in part by increased computing power, which makes more feasible the application of statistical techniques such as maximum likelihood and Bayesian methods [25].

3.1. Systematic analyses

In recent years, there has been an increasing interest in representing phylogenetic relationships above the species level as networks rather than as strictly bifurcating trees [26,27]. This is partly because of the desire to present character conflict or uncertainty in the reconstructed phylogeny, even when the true evolutionary relationship is believed to be hierarchical, and partly because of the recognition that evolutionary events such as hybridisation, horizontal gene transfer and symbiosis may create true non-hierarchical relationships. Character conflict is often represented by a split network, where parallel edges connect the nodes, while reticulate networks, where some nodes have more than two parents, are often used as an explicit representation of complex evolutionary events [27]. A number of different network construction methods are available, including median networks [28], neighbour-nets [29] and reticulograms [30].

3.2. Population genetic analyses

The traditional approach to population genetic analysis is based on allele frequencies, without regard to historical relationships between these alleles. To infer the action of evolutionary processes, such genetic drift, migration or selection, allele frequencies are compared with equilibrium expectations, derived from particular models of population structure. More recent studies have incorporated explicit tests of the effects of landscape heterogeneity on evolutionary processes, an approach known as landscape genetics [31,32]. Although this approach has proved very powerful and yielded important insights into the ecology of parasites and the epidemiology of parasitic diseases, it also has
important limitations. First, when applied to genetic markers such as DNA sequences and microsatellites, it does not utilize their historical information content. Even worse, if the markers are extremely variable, allele frequencies become meaningless because every sequence is different. Second, the equilibrium assumptions upon which many analyses depend are not always valid. This is a particular problem for parasites of people and domestic animals which may be subjected to rapid, long distance dispersal as a consequence of host movements. Third, traditional population genetic analyses require an a priori demarcation of breeding populations or demes. This is not always straightforward, especially for parasites which must leave the host to complete their life cycle.

Assignment methods, which use information from genetic markers to ascertain population membership of individuals or groups of individuals, have the potential to overcome some of the limitations of traditional allele frequency approaches [33]. In particular, assignment methods do not require that a stable equilibrium has been achieved between opposing evolutionary forces (although they do usually assume that the population is in Hardy-Weinberg and linkage equilibrium), so they are often more appropriate for parasites with recent history of invasions or range expansions. Despite their advantages, there are limitations to the use of many assignment methods with parasite species. In particular, assumptions of Hardy-Weinberg and linkage equilibria limit their usefulness for species which do not reproduce sexually or have high rates of inbreeding. They also may not distinguish population subdivision from other processes such as small population size, inbreeding and genetic bottlenecks, which may cause departures from Hardy-Weinberg and linkage equilibria.

Phylogenetic methods make use of intraspecific gene genealogies rather than allele frequencies to infer evolutionary processes, usually in a geographical context (phylogeography). The great strength of phylogenetic methods of analysis is that they add a temporal dimension which can be related to spatial organization among alleles [25]. Important information can be obtained from intraspecific gene phylogenies even in the absence of a population genetic model [e.g. 34], but an explicit population genetic model provides extra power to test specific hypotheses about evolutionary processes [35].

The two main approaches to inferring evolutionary processes from a reconstructed intraspecific phylogeny are nested clade analysis and coalescent-based methods, particularly approximate Bayesian computation. Nested clade analysis is a method of inferring the role of contemporary evolutionary processes, such as gene flow, and historical events, such as population fragmentation, range expansion or colonization, from the geographic structure of intraspecific gene clades [36,37]. Briefly, a series of hierarchically nested clades are defined from the phylogeny. The geographic distributions between clades at different hierarchical levels are statistically compared and the pattern of the comparison is used to test hypotheses about evolutionary processes which have been developed from simulation models. Nested clade analysis, while it has been applied to infer evolutionary processes responsible for genetic structure in parasites [e.g. 38], has
been widely criticized for the subjective nature of the inference process [39,40; although see 41].

Coalescent approaches to population genetic analysis are based on coalescent theory, which is a mathematical description of the genealogical history of a sample of neutral alleles from a population. A pivotal result from coalescent theory is that coalescence time, the time at which two alleles share a most recent common ancestor, is a function of the demographic history of the population [42]. In theory, this enables a likelihood function to be calculated, which considers both the probability of obtaining the observed data given an intraspecific phylogeny, and the probability of the phylogeny given certain genetic or demographic parameters [43]. Calculating the likelihood is not usually computationally feasible and instead it is approximated by a variety of techniques. The most widely used technique is approximate Bayesian computation, in which the data are compressed into summary statistics and calculation of the likelihood is replaced by a comparison of observed and simulated data [44]. These methods have not been utilized widely for parasites, probably because models relating gene phylogenies to evolutionary processes are complicated by the need to consider the epidemiology of parasitic infection. For microparasites, where successful reproduction requires both replication within hosts and transmission between hosts, gene phylogenies in a component population (all the members of a parasite species in all the individuals of a particular host species) composed of many infrapopulations (all the members of a parasite species within a single host individual), should be reasonably approximated by standard metapopulation models [35]. Different patterns of immune response in the host, however, can produce quite different gene phylogenies in the parasite [45] and will have to be explicitly incorporated in coalescent models to infer genetic parameters. For macroparasites, the metapopulation model is complicated by the extent to which parasite infrapopulations have a stable recurrence of generations, which is determined by the extent of correlated transmission of offspring from one host to the next ([46]; see also section 4.1.2).

4. **Identification and classification of the causative agents**

Parasite control depends upon the rapid, accurate detection and identification of the aetiological agents, so that cycles of transmission can be inferred and the potential for interaction between cycles determined. Effective control also requires the ability to characterise parasites from different stages in their life cycles in tissues, blood, faeces or the environment, on the basis of epidemiologically useful features. These include host specificity, public health significance in terms of zoonotic potential, virulence and drug sensitivity. Traditional diagnostic techniques involving microscopy have thus been complemented by a variety of molecular tools that provide additional information about the causative agents.

Molecular identification is particularly important when discriminating different parasites with morphologically identical life cycle stages, such as eggs or cysts, from faecal samples, or when attempting to match different life cycle stages of the same parasite from intermediate and definitive hosts [47]. For example, it is now emerging that in some
endemic areas humans may be infected with more than one species of hookworm, and the eggs expelled in the faeces are morphologically identical [48]. Fortunately, PCR-based procedures have been developed which can differentiate between all the relevant genera and species of hookworm of public health and veterinary significance [49]. It is important to be able to distinguish between the two main genera of human hookworm, *Ancylostoma* and *Necator*, because of their different pathogenic potential, but within the genus *Ancylostoma*, there are two species, *A. duodenale* and *A. ceylanicum*, of which the latter is zoonotic. The emergence of *A. ceylanicum* in South East Asia is a major impediment to control where mass chemotherapy is used, because dogs are the zoonotic reservoir of *A. ceylanicum* and are not targeted in mass chemotherapy programs [48]. Similarly, humans may be infected with more than one species of taeniid cestode in some endemic areas where there are a variety of susceptible intermediate hosts, particularly pigs and cattle [50]. In such situations, the epidemiology and control of human taeniasis and cysticercosis is dependent upon determining cycles of transmission and sources of infection. As with hookworm, morphological discrimination of taeniid species is not possible on the basis of the parasite stages passed in human faeces but is readily achieved with PCR-based procedures [48].

In addition to providing rapid and sensitive identification of established parasite taxa, the application of molecular tools has also helped to resolve taxonomic issues that may have resulted in controversy in the past, when new species or ‘strains’ were described on the basis of host occurrence, phenotypic characteristics and/or epidemiological observations. The resolution of taxonomic issues using molecular tools often occurs in two distinct stages. First, different genetic groups are found within what is ostensibly a single, morphologically defined species, and then these groups are defined as taxonomic categories, either at the intraspecific level or as different species or higher taxa.

4.1. Factors which promote genetic structure

The genetic structure (i.e. the extent to which genetic variation is distributed among, rather than within populations) of a species is determined by the interplay of different evolutionary forces, principally genetic drift, selection and migration. These evolutionary forces are themselves influenced by a range of biological and ecological factors such as mode of reproduction, breeding system, effective population size and dispersal ability. The extent to which we recognize intraspecific groups of parasites will be a function of the extent to which intraspecific variation is structured among different hosts or among different geographic areas. This, in turn, will be determined primarily by the mode of reproduction of the parasite and the fragmentation of parasite populations among host individuals.

4.1.1. Mode of reproduction

In asexual reproduction, new, genetically identical individuals are produced by a single parent without genetic recombination. Although viruses and bacteria reproduce predominantly asexually, they may sometimes exchange genetic material. Most parasitic
protozoa reproduce asexually, although they may also have an obligate or facultative sexual phase in their life cycle. As a result, three different types of population structure have been proposed for viral, bacterial and protozoan parasites: panmictic, as a result of frequent genetic exchange; clonal, resulting from little or no genetic exchange; and epidemic, where a basic panmictic structure is masked by occasional clonal expansion of certain genotypes [51,52].

The extent of clonality in a species can be inferred through the pattern of single locus and multilocus genetic diversity. At the single locus level, clonality will lead to an excess of heterozygotes in diploid organisms, while at the multilocus level, clonality will produce widespread, identical genotypes, non-random associations between alleles at different loci (linkage disequilibrium) and congruence in different intraspecific gene phylogenies [53,54,55]. A clonal population structure does not imply that genetic exchange is absent in the species, only that it is too rare to erode the basic genetic patterns of clonality. Using these criteria, Tibayrenc et al. [53,56] identified a number of species of parasitic protozoa, including Entamoeba histolytica, Giardia duodenalis, Leishmania tropica, L. major, Trypanosoma brucei, T. cruzi and T. vivax, as having an essentially clonal population structure. Subsequent studies have shown that the situation is rather more complex and different populations of the same species often show different degrees of clonality [57]. MacLeod et al. [52,58], for example, found that Trypanosoma brucei isolated from livestock in Botswana showed an epidemic population structure, while the same species isolated from humans in the same locality, had a clonal population structure. In Europe and North America, Toxoplasma gondii is considered to have a predominantly clonal population structure, with three main clonal lineages, referred to as Types I, II and III, accounting for >85% of strains isolated from humans and domestic animals [59]. However, recent studies of isolates of T. gondii in wildlife from North and South America, as well as Australia have uncovered more biological and genetic diversity [60,61,62]. Although some strains infecting wildlife appear to be recombinant genotypes derived from crosses between the archetypal clonal lineages, others are atypical strains which possess completely novel alleles. This diversity appears to be driven by regular cycles of sexual reproduction, with occasional expansion of clonal lineages by carnivory or self-mating [63,64]. A panmictic population structure is therefore thought to exist in South America and parts of North America, with an epidemic expansion of successful clones through most of North America and Europe, where wildlife apparently plays a less significant role in transmission of the parasite [60,65].

4.1.2. Fragmentation among hosts

Some parasitic protozoa and most helminths have an obligate phase of sexual reproduction, involving genetic exchange, during their life cycle. For these parasites, the factor of overwhelming importance in determining genetic structure is the fragmentation of populations among hosts. The infrapopulation in a single definitive host represents the breeding group. Eggs or larvae are passed into the external environment and/or one or more intermediate hosts, so the progeny from different infrapopulations are mixed each
The extent of mixing at different spatial scales determines the extent of genetic differentiation among infrapopulations, among component populations in different host species and among suprapopulations (all the individuals of a parasite species within all hosts and in the environment) in different geographic areas. Progeny mixing is influenced by a wide variety of intrinsic and extrinsic factors, including transmission dynamics, asexual amplification of larval stages, inbreeding rate in the definitive host and host migration rate.

At the level of different definitive hosts of the same species, if offspring are transmitted vertically or as a clump from one definitive host to the next over several generations, then infrapopulations will effectively function as demes. This is likely to promote inbreeding, leading to the reduction of within-host genetic variation and an increase in among-host genetic variation through genetic drift [46] (Fig. 1a). Evidence for a recurrence of generations within individual infrapopulations has been found in a number of studies. For example, infrapopulations of lice, *Geomydoecus actuosi*, infecting pocket gophers (*Thomomys bottae*) have heterozygote deficiencies (indicating inbreeding) and are strongly structured, with 9.2% of genetic variance distributed among hosts [66]. Lice are transmitted exclusively by inter host contact and in pocket gophers this principally occurs during mating encounters and the rearing of young. Anderson *et al*. [67] found that *Ascaris* worms bearing identical mtDNA haplotypes were found within the same human or pig host more frequently than expected by chance. They suggested that this resulted from the spatial clumping of genetically related eggs in the environment.

For many parasites, however, there is little evidence of genetic structuring among infrapopulations, indicating that clumped transmission is rare. In populations of *Teleodorsagia* from sheep, for example, 98% of genetic variation occurs within infrapopulations [68], a result consistent with other studies of trichostrongyloid nematodes [69,70]. Even in parasites which have a life cycle predisposed to self-reinfection, such as *Strongyloides ratti*, less than 5% of genetic variation is distributed among infrapopulations in definitive hosts [71].

The recurrence of generations within individual definitive hosts may be enhanced by asexual reproduction in intermediate hosts, as occurs in many protozoans, digenean trematodes and cestodes (Fig.1b). Transmission of clones from the intermediate host to the definitive host appears to lead to enhanced structuring of infrapopulations of *Plasmodium falciparum* in people [72] and the cestode, *Fascioloides magna*, in white-tailed deer (*Odocoileus virginianus*) [73]. Asexual reproduction in intermediate hosts does not necessarily lead, however, to a stable recurrence of generations within, and enhanced genetic diversity among definitive hosts, because it will be countered by factors such as reduced variance in reproductive success between clones and enhanced mobility of both definitive and intermediate hosts [74]. Theron *et al*. [75] for example, found that infrapopulations of *Schistosoma mansoni* in rats (*Rattus rattus*) from Guadaloupe, contained a mean of 34 different multilocus genotypes per host despite the fact that snail intermediate hosts (*Biomphalaria glabrata*) contained only 1.1 genotypes per host, on average. The transmission of multiple genotypes to the definitive host is likely due to the
mobility of rats, their weak immune response, allowing multiple infections, and to spatial aggregation of infected snails around limited water resources.

For parasites that are able to utilize more than one species of definitive or intermediate host, the likelihood of genetic structuring between different component populations depends on the extent to which the different host species utilize different resources, and will be enhanced by the same processes that lead to structuring among infrapopulations, that is, clump transmission and asexual multiplication. For example, Wang et al. [76] identified two major genetic clusters of *Schistosoma japonicum* infecting different definitive host species in Anhui province, China; one in cattle, water buffalo and humans, and the other in goats, pigs, dogs and cats. The authors suggest that this differentiation is due to spatial resource sharing by cattle, water buffalo and humans.

At the level of suprapopulations of parasites in different geographic areas, the major determinant of genetic structure is host mobility, through its effect on parasite gene flow. For example, trichostrongylid nematode parasites of livestock (*Ostertagia ostertagi*, *Teleodorsagia circumcincta*, *Haemonchus placei* and *H. contortus*), that are regularly transported by people between distant locations, have less genetic structure than a related parasite of wild deer (*Mazamastrongylus odocoilei*) [77]. Similarly, ticks (*Ixodes uriae*) on Atlantic puffins (*Fratercula arctica*) are much less genetically structured than the same tick species on black legged kitiwakes (*Rissa tridactyla*); presumably because puffins move between local colonies much more frequently than kitiwakes [78]. For parasites with indirect life cycles, intermediate host mobility as well as definitive host mobility may influence genetic structure. Criscione and Blouin [79] found that three species of digenean trematodes (*Deropegus aspina A*, *Deropegus aspina B* and *Plagioporus shawi*) of salmonids (*Oncorhynchus* spp.) which cycle exclusively in aquatic hosts, are more strongly structured than a fourth species (*Nanophyetus salmincola*) whose life cycle includes highly mobile terrestrial hosts.

Host mobility can prevent geographic differentiation of parasite suprapopulations even when parasite infrapopulations are highly structured. In the digenean trematode, *Fascioloides magna*, for example, where asexual multiplication in intermediate hosts leads to strong genetic differentiation among infrapopulations in white tailed deer (*Odocoileus virginianus*), there is little differentiation among flukes in different geographic areas, presumably because of long distance dispersal by deer [73]. In the cestode *Echinococcus granulosus* in Australia, there is no significant genetic variation between populations from intermediate hosts in different geographic areas separated by more than 4,000 km, despite evidence of high effective selfing rates in definitive hosts, due to clumped transmission of clones [80]. This lack of geographic differentiation is presumably due to high mobility of both intermediate hosts (sheep and kangaroos) and definitive hosts (dogs and dingoes) of *Echinococcus granulosus* in Australia.

4.2. Delimiting strains and species
When a substantial part of the genetic variation within a species of parasite is associated with distinguishable biological or ecological characteristics, such as morphology, host associations, development rate, infectivity, pathogenicity or drug resistance, we usually wish to recognize the different variants with some formal or informal taxonomic designation. Nomenclature is essential for effective communication and provides the stability that underpins epidemiological investigations [81]. The lack of morphological differences between many inter- and intraspecific variants has, in the past, compounded an often confusing taxonomic picture, which in many cases has taken decades to resolve. Such was the situation with Trichinella and Echinococcus, but as a result of the application of molecular tools many taxonomic issues have been resolved and as a consequence, communication has been markedly enhanced. In itself, giving something a taxonomic designation with the support of molecular data is not a ‘molecular epidemiological’ study unless it can be put into an epidemiological context. For example, being able to discriminate between E. histolytica and E. dispar was only possible with the development of molecular tools which gave confidence to the species names proposed and a terminology that underpins epidemiological investigations [81,82].

The only formal taxonomic category below the species level is the subspecies, traditionally defined as a geographically localized intraspecific group that differs genetically (and taxonomically) from other such groups [83]. There are, however, a plethora of other terms such as isolate, stock, line, strain and discrete typing unit, which have been used informally to describe intraspecific variation (Table 2). These terms have often been defined in different ways and in a biological sense their value is questionable because they may tell us nothing about the evolutionary history or evolutionary potential of the groups concerned. Their application to parasitic organisms however, has been of great practical significance, because they are often related to important features of parasitic disease. In Trypanosoma brucei, for example, three subspecies have historically been defined on the basis of geographic and host distribution, and the clinical course of disease. T. b. gambiense is a human parasite distributed through western and central Africa, causing chronic disease. T. b. rhodesiense is a human parasite distributed through eastern and southern Africa causing acute disease and T. b. brucei infects domestic and game animals, but not humans and is widely distributed throughout sub-Saharan Africa. A genetic basis to human infectivity appears to have been established, at least for some isolates, in the expressivity of the serum-resistance-associated (SRA) gene product [84,85]. Population genetic studies using a range of genetic markers, however, suggest that neither T. b. gambiense or T. b. rhodesiense form monophyletic groups. It appears that T. brucei has acquired the ability to infect humans on four separate occasions, twice within the subspecies T. b. gambiense and twice within the subspecies T. b. rhodesiense [52]. Therefore, while it seems inappropriate to retain the three subspecific designations, the three groups could still be referred to as different strains, because of differences in geographic distribution and human infectivity.

Debate over what constitutes a species has been an enduring source of confusion in biology, with a multitude of different species concepts proposed [86,87]. Among parasitologists, this has frequently led to disillusionment with the prospect of identifying
a single species concept that includes all groups of parasites and the use of a purely phenotypic definition of species [88,89]. We believe, however, that the problem posed by this multiplicity of different species concepts can be overcome by recognizing a fundamental distinction between conceptual views of what constitutes a species and operational criteria for delimiting different species [90]. Many existing species concepts, such as the biological species concept, the phylogenetic species concept and the cohesion species concept, differ only in their operational criteria for species delimitation; conceptually they agree that species represent the contemporary tips of an evolutionary lineage [86]. They can therefore be equated with the evolutionary species concept; which states that that a species is a single lineage of organisms with a common evolutionary trajectory, distinguishable from other such lineages [91]. The evolutionary species concept is applicable to most eukaryotic organisms, regardless of their mode of reproduction or breeding system, although it may be difficult to apply if horizontal gene transfer is common between distant lineages [89]. Delimiting species under an evolutionary species concept requires a determination of when lineages have a common evolutionary trajectory (indicating that they are the same species) or when they have different evolutionary trajectories (indicating that they are separate species).

One approach to delimiting species is to utilise the pattern of evolutionary relationships among lineages, such as genetic distance, monophyly or exclusivity, as a guide to their evolutionary trajectory. Genetic distance between lineages, usually inferred with mtDNA or ribosomal ITS markers, has frequently been used as an indicator of specific status. For example, Mcnish et al. [92] suggested that isolates of *Hymenolepis nana* in Australia actually exist as two cryptic or sibling (morphologically identical, but genetically different) species, based on a sequence divergence of 5% in the mitochondrial cytochrome *c* oxidase 1 gene. Such a genetic yardstick approach is a useful prospecting tool for suggesting the possibility of different cryptic species within a morphologically similar group, but it does not provide an infallible guide [93]. More reliable indicators of species status can be gained from a phylogenetic analysis of the putative species. Organisms following the same evolutionary trajectory should be monophyletic (derived from the same ancestral taxon) and exclusive (more closely related to each other than they are to any individuals outside the group) [94]. The morphologically defined species *Echinococcus granulosus*, for example, has now been split into a number of different species, because phylogenies based on mtDNA sequence data indicate that strains of *E. granulosus* are not monophyletic [95,96].

Another approach to delimiting species, complementary to utilising the pattern of evolutionary relationships among lineages, is to examine the processes, such as gene flow and ecological constraints, which are responsible for maintaining a cohesive evolutionary trajectory. Studies utilising this approach have typically focused on gene flow, although that is, of course, only applicable to organisms which regularly exchange genes in their life cycle. For parasites which can be cultured under controlled conditions, experimental crosses can be used to determine the ability of two populations or lineages to hybridise. Le Jambre [97], for example, established mixed populations of *Haemonchus contortus*
and H. placet in recipient sheep and found that hybrid offspring had markedly reduced fertility.

Crossing experiments, however, are not possible or practical for most parasite species and the extent of gene flow is usually monitored in the field. Fixed genetic differences between populations in sympatry provide strong evidence that they are on different evolutionary trajectories and therefore represent different species. Many studies have used genetic markers to identify such non-recombining lineages of parasites existing in the same geographic area and often in the same host individual. The implication from these studies is that we may have underestimated the number of independently evolving species in almost all groups of parasites. For example, historically only two major zoonotic species of anisakid nematodes have been recognised; the herring worm or whale worm Anisakis simplex, and the cod worm or seal worm, Pseudoterranova decipiens, both with an apparently cosmopolitan distribution. Recent molecular genetic studies, however, have shown that both of these morphospecies actually comprise a number of genetically differentiated sibling species, often with distinct geographic and/or host ranges [98,99]. Similarly, morphological studies have identified approximately 175 species of avian blood parasites of the genera Haemoproteus and Plasmodium, but mtDNA sequencing indicates that the real number of species may be around 10,000, almost two orders of magnitude greater [100]. Even in very well studied groups of parasites, such as the trichostrongylid nematodes infecting livestock, new cryptic species are being discovered. Grillo et al. [101] found that worms from a goat farm in France, morphologically identified as Teleodorsagia circumcincta, were in fact two separate species, with little gene flow between them.

5. Determining transmission cycles

Theoretical models of the population dynamics of micro- and macroparasites have provided important insights into the key features of parasite transmission, particularly with respect to the parameters that determine the basic reproduction ratio (R₀) of the parasite. Such models, however, usually make simplifying assumptions that ignore parasite, host and environmental heterogeneity, and obtaining empirical data on how these heterogeneities affect parasite transmission is an important requirement for developing more realistic population dynamic models. A key empirical demand, and one that can be addressed with molecular epidemiological tools, is to estimate networks of parasite transmission, both within and among species of hosts. Among species of host, most interest has centred on the application of molecular tools to infer the frequency of zoonotic transmission in a range of parasite species.

5.1. Echinococcus, Giardia and Cryptosporidium

Species of these three parasite genera clearly share little in terms of their biology and phylogenetic relationships. However, they all have three characteristics in common: a wide host range and questions of host specificity; an early taxonomy poorly supported by limited and/or questionable morphological discrimination and based largely on host
occurrence; and uncertainty about their zoonotic and public health significance [102]. With *Echinococcus*, *Giardia* and to some extent *Cryptosporidium*, molecular tools have helped to resolve taxonomic issues and have supported the proposals of early taxonomists [22,103,104]. Evidence of morphological differences between isolates of *Echinococcus* from different hosts can now be supported by extensive molecular evidence of genetic variation and as such can be used as reliable and cost effective diagnostic markers in field studies, particularly in developing countries where costs and lack of equipment is an issue. This is not the case with *Giardia*, where initial taxonomic descriptions based on host occurrence could not be supported by morphological differences. Molecular tools therefore are the only method for identification and are proving of value epidemiologically. With *Cryptosporidium*, robust molecular epidemiological tools are available but they have principally been utilised for taxonomic purposes and their full epidemiological potential has still to be realised.

Species of *Echinococcus* have a two-host life cycle involving an herbivorous or omnivorous intermediate host and a carnivorous definitive host. The parasites demonstrate high definitive host specificity but low intermediate host specificity, which has raised questions about the rigidity of cycles of transmission and the zoonotic potential of populations maintained in different host assemblages [105]. As with *Giardia* and *Cryptosporidium*, the epidemiology of infections with *Echinococcus* spp., particularly in humans (cystic or alveolar echinococcosis), was based on a species taxonomy established largely on host occurrence. This was questioned on taxonomic grounds and in the absence of evidence of genetic distinctness between the parasites from different intermediate hosts, there was uncertainty for many years whether cycles involving sheep, cattle, pigs, camels, kangaroos etc. could interact. These questions have subsequently been resolved with the advent of reliable, robust and reproducible molecular tools which have not only supported the early taxonomy, but also demonstrated the distinctness of transmission cycles and the potential for interaction, particularly with respect to zoonotic transmission [95,103,106]. Importantly, these molecular epidemiological studies have given confidence to the morphological characters used for species discrimination which now offer a simple, cost effective means of parasite identification in endemic foci where the application of molecular tools may not be practical or cost-effective [107,108].

Host specificity and zoonotic potential have been the key drivers of epidemiological investigations on species of *Giardia*, ubiquitous enteric protozoan parasites of mammals. Frequent reports of infection in companion animals, livestock and aquatic mammals have led to much discussion and controversy over their role as zoonotic reservoirs of infection (104,109). The lack of any significant morphological variability, but considerable evidence of phenotypic differences between isolates, fuelled this debate over many years. The subsequent application of PCR-based molecular tools has resolved questions of host specificity, taxonomy and zoonotic potential, but not the frequency of zoonotic transmission [104,109]. The application of multilocus genotyping to isolates of *Giardia* from human and other mammalian hosts in different parts of the world has clearly demonstrated the occurrence of zoonotic species in the same geographic areas, supporting the potential for zoonotic transmission [110,111]. However, finding infection with
zoonotic genotypes only demonstrates the potential for transmission – not the occurrence of actual transmission. This is not a reflection of the tools available but of the lack of focus in the study design, and from an epidemiological point of view, genotyping disparate collections of isolates may not be informative.

A number of studies in defined endemic foci have provided convincing evidence of zoonotic transmission involving dogs and humans [112,113,114]. Although the emphasis of these studies has been on the dog as a reservoir of human infection, some recent reports investigating the molecular epidemiology of infections with species of *Giardia* in wild primates have provided further evidence of zoonotic transmission in localised foci, and have also demonstrated that ‘reverse zoonotic transmission’ (zooanthroponotic) is an important factor that must be considered in understanding the epidemiology of infections with these parasites [115,116,117].

There are two zoonotic species/assemblages of *Giardia* which are geographically widespread and as more isolates are genotyped, contrasting patterns are emerging of their distribution. For example, studies in Europe had suggested that *G. enterica*/assemblage B has a predominantly human distribution [118], but a recent study of dogs in the USA found a higher frequency of infections with *G. enterica*/assemblage B than with *G. duodenalis*/assemblage A [119]. Thus in North America at least, we cannot assume that *G. duodenalis*/assemblage A is the most common of the zoonotic assemblages found in non-human hosts. Indeed in wildlife, *G. enterica*/assemblage B often predominates [117] whereas in cattle, *G. duodenalis*/assemblage A is most often reported [119]. However, there is extensive genetic sub-structuring within *G. enterica*/assemblage B, and it is possible that some subgroups are more commonly associated with zoonotic infections than others. From an epidemiological perspective, there is increasing evidence of differences in virulence between the zoonotic species of *Giardia* and how this manifests clinically in different host species and in different circumstances associated with nutritional deficiencies and/or polyparasitism requires much more research [120].

A lack of morphological differences between isolates of *Cryptosporidium* from different host species, waterborne outbreaks and circumstantial evidence that livestock could have been the source of water contamination were the main drivers for the development of molecular epidemiological tools for these parasites. These have proved to be very useful in determining sources of infection and risk factors of public health significance [121] but have not resulted in significant epidemiological information since then. Although the tools are available, overall, molecular epidemiological studies on *Cryptosporidium* infections are considered to be still in their infancy [122]. Molecular tools have, however, improved our understanding of species-level taxonomy. This has largely been of value in understanding the extent of non-human reservoirs of zoonotic infection, but with limited application to population genetic studies. The epidemiological potential of such studies has been demonstrated by Mallon et al. [123] and Peng et al. [124], who have provided evidence that the population structure of *C. parvum (=C. pestis [125]) and C. hominis is more complex than previously suggested.
Sub-genotyping has continued to reveal genetic sub-structuring within \textit{C. parvum}, but whether this is reflected in variation in host specificity and zoonotic potential remains unclear. Using multilocus genotyping, Grinberg et al. \cite{126} provided evidence supporting the suggestion of Hunter and Thompson \cite{121} of distinct anthroponootic \textit{C. parvum} cycles that do not involve cattle. The recent demonstration of zoonotic transmission to humans in the UK of a newly described genotype of \textit{Cryptosporidium} from rabbits \cite{127} has raised concerns that humans may be at risk of infection from rabbits in other geographic areas, as recently proposed in Australia where the rabbit genotype has been identified \cite{128}. Although given a new species name, \textit{C. cuniculi}, by Robinson et al. \cite{129}, it is genetically very close to \textit{C. parvum} and \textit{C. hominis} and thus it may be prudent to reconsider the taxonomic status of all three species in the future. Recent advances in nucleic-based approaches for the diagnosis and analysis of genetic diversity in species of \textit{Cryptosporidium} \cite{22} represent a significant step towards an improved understanding of epidemiology and population structure \cite{122}.

5.2. Wildlife and zoonoses

The role of wildlife as reservoirs of infections that may be transmissible to humans is a controversial issue. Human factors clearly play a role in increasing the risk of any ‘spillover’ of infections from wildlife through encroachment, land clearing, hunting etc. However, there is increasing evidence of ‘reverse zoonotic’ transmission from humans to wildlife, a factor that must be embraced if the so-called ‘one health’ concept considers diseases of wildlife as important as those affecting humans and domestic animals. It is important to understand parasite biodiversity in wildlife in terms of conservation \cite{130,131}. This requires the surveillance of fauna that may often be endangered or threatened. In such circumstances, non-invasive sampling and the application of molecular tools can provide data which, in the past, were only available following opportunistic necropsy. For example, from an anthropocentric viewpoint, an understanding of the species of \textit{Plasmodium} affecting primates in the wild led to the identification of a new zoonosis, with \textit{P. knowlesi} found in humans using molecular tools \cite{132,133}. In contrast, the application of molecular tools has identified a novel, genetically distinct form of \textit{Leishmania} in macropod marsupials in Australia, as well as a new non-sandfly vector \cite{134-136}. It is likely that the ecology of both the novel form of \textit{Leishmania} and its vector have been associated with wildlife well before human settlement of Australia, but the discovery has raised concerns about the transmission to wildlife of introduced, pathogenic forms of \textit{Leishmania} from humans or dogs \cite{137}.

Native rats (\textit{Rattus macleari}) on Christmas Island became extinct following the introduction of flea-infested \textit{R. rattus} in the early 1900’s. It was proposed that this could have been due to infection of naïve native rats with \textit{T. lewisi} transmitted by fleas from \textit{R. rattus} \cite{138,139}. There was no way of proving this until ancient DNA techniques were applied to museum specimens and demonstrated the presence of \textit{T. lewisi} in native rats after colonisation of Christmas Island by \textit{R. rattus} but not before \cite{139}. In Western Australia, a comprehensive program of non-invasive parasite surveillance in native wildlife has revealed a diversity of novel \textit{Trypanosoma} genotypes \cite{140}. All are
stercorarian trypanosomes, with some closely related to the causative agent of Chagas
disease, *T. cruzi*. Chagas disease is no longer confined to South America because of
increasing human migration to non-endemic regions [141,142]. The occurrence of
trypanosomes related to *T. cruzi* in Australian native wildlife raises question about the
vectorial potential of *T. cruzi* in Australia, which cannot as yet be answered [137]. As
with *Leishmania* sp. in Australia, the application of molecular tools will prove invaluable
in addressing these questions.

6. Polyparasitism

The term polyparasitism refers to concurrent/concomitant/co-/mixed infections of either
different species and/or intraspecific variants of parasites, the latter a more recently
recognised phenomenon with the advent of molecular tools. It is not a newly discovered
phenomenon as Stoll [143] demonstrated when estimating the huge global burden of
human nematode infections. This was greater than the global population at the time,
reflecting the large number of mixed infections in developing countries. The seminal
studies of Buck et al. [144-147] highlighted the problems of assessing morbidity due to
multiple parasite infections in highly endemic foci. Keusch and Migasena [148] also
emphasised that polyparasitism was the rule rather than the exception, and that the
possibility of either synergistic or antagonistic effects must be considered in planning
public health intervention programs. Yet this phenomenon has been overshadowed for
many years by the widespread use of the DALY measuring system to determine the
impact on health of parasitic disease, which does not take into consideration the “co-
morbidities of polyparasitism” [149]. Payne and colleagues [149] suggest a new approach
where co-infections with more than one infectious agent are defined as a specific disease,
for example malaria, hookworm, malaria + hookworm, malaria + schistosomes, etc. This
would appear logical and should be readily achievable with diseases such as malaria and
hookworm, where the pathogenic mechanisms of the individual aetiological agents are
reasonably well understood.

Awareness of the significance of polyparasitism in terms of malaria, schistosomes and
more recently gastrointestinal helminths, is now well established [150-156]. Molecular
epidemiological tools will contribute enormously to a better understanding of
polyparasitism by providing the means to identify parasites in situations where
morphological characterisation is not possible. Probably the earliest example of this was
the demonstration of mixed infections with genetic variants of *Plasmodium falciparum* in
humans [157]. The potential clinical impact of this had been demonstrated previously in
rodent models of *P. chabaudi* [158]. The far reaching implications of such mixed
infections on vaccination and drug treatment strategies in malaria, with respect to
interfering with competitive interactions and promoting increased virulence, are now well
known and continue to be the subject of much debate [159].

In contrast, much less is known about the occurrence of genetically mixed infections of
trypanosomes [160]. Molecular epidemiological studies have demonstrated extensive
genetic diversity in *Trypanosoma cruzi* in endemic regions of South America. Studies
have also revealed the wide host range of *T. cruzi* in terms of sylvatic cycles and vectors, as well as the occurrence of mixed infections in reservoir hosts, which is perhaps not unexpected given the high levels of genetic diversity which characterise *T. cruzi* [161]. Experimental studies in mice have also demonstrated that in mixed infections, strains exhibit predilection for different tissues (histiotropism). It is therefore surprising that the appropriate molecular tools have only recently been applied to human infections. These have demonstrated the occurrence of mixed infections with different strains of *T. cruzi* [162-164] and raised questions about the impact that differences in virulence, drug sensitivity and histiotropism among strains of *T. cruzi* will have on the management of Chagas disease [163]. Some years ago it was suggested that *T. cruzi* populations in a patient’s bloodstream could be dissimilar to the parasite population that causes tissue damage [165]. Using molecular tools, this has now been shown to be the case in mixed infections where the most prevalent genotype in the bloodstream is different to that in cardiac tissue [164,166]. This can complicate treatment involving heart transplantation, where chemotherapy is used to counter reactivation of the parasite following post-surgical immunosuppressive therapy, since in mixed infections parasite reactivation can occur at different times [166].

The epidemiology of mixed infections with different genera of protozoa demonstrates another situation where molecular tools provide the means to identify the causative agents. For example, in South America, overlapping zones of transmission of *T. cruzi* and *Leishmania* spp. occur frequently, resulting in mixed infections in humans[167,168]. Molecular tools provide a valuable alternative to traditional diagnostic approaches, which suffer from cross-reactivity, in areas where successful treatment depends upon accurate and speedy diagnosis. In wildlife, molecular tools have contributed to a better understanding of the pathogenesis of infections with *Toxoplasma gondii* in sea otters. A recent study has shown that in 42% of cases, sea otters infected with *T. gondii* were coinfectected with *Sarcocystis neurona* and that such mixed infections were more virulent than single infections [169]. Similarly, experimental studies in mice coinfectected with *T. gondii* and *Trypanosoma lewisi* demonstrated increased virulence in mixed infections [170-172]. Molecular epidemiological studies of threatened native wildlife in Western Australia have revealed the occurrence of mixed infections with *T. gondii* and *Trypanosoma* spp., which may be associated with the decline of *Bettongia penicillata* [173]. Mixed infections with different genotypes of *T. gondii* are common in Australian wildlife, but it is not yet known how this influences the virulence of infections [62].

The prevalence of mixed infections with enteric protozoa in developing countries show that they are the rule rather than the exception, with children most at risk harbouring at least two species of protozoan [174-178]. The most commonly represented protozoa in concurrent infections are *Giardia* spp., *Blastocystis* spp., and *Entamoeba coli*, which are usually also found with the cestode *Hymenolepis nana* [174-176,179,180]. Depending on the endemic area, *E. histolytica* and *E. dispar* may also occur in mixed infections. It is not possible to identify all these species using morphology alone and thus molecular tools now provide the means to undertake epidemiological studies.
Multiple low-intensity infections of gastrointestinal helminths and schistosomes have been shown to confer an increased risk of anaemia, emphasising that this common pattern of infection is not clinically benign [181]. Chronic helminth infections could also have a significant influence over the immune response and hence susceptibility to other pathogens including microparasites, both systemic and enteric, as well as influencing the outcome of vaccine trials [182-186]. In order to determine the impact of polyparasitism in these situations, accurate diagnosis is again important, for example in determining the species of schistosome or hookworm involved [187]. Molecular tools will therefore provide the basis to better understand the epidemiology of mixed helminth and protozoan infections. For example, the inability to identify hookworm species in the study by Sorensen et al. [188] of gastrointestinal infections in Guatemalan children, leaves open questions about the potential of zoonotic transmission, if dogs are a reservoir of *A. ceylanicum* as in South East Asia (see above).

7. Markers for traits of epidemiological importance

Studies of genetic diversity may be of great practical significance not only in parasite identification, dissecting interactions within the host and tracking transmission among hosts, but also in inferring the distribution of genetic variation in traits of epidemiological importance, such as drug resistance, zoonotic potential and virulence, and in predicting the evolutionary response in such traits to selection pressures imposed by nature or by human intervention. A number of studies using neutral allozyme or mtDNA markers, for example, have found relatively high levels of genetic diversity within populations and little population genetic structure in helminth macroparasites [47]. An obvious implication from such studies is that populations of parasites might respond rapidly to selection imposed by drug treatment and that gene flow would rapidly spread resistance alleles to other populations.

There is however, a limit to the usefulness of non-neutral genetic markers in epidemiological studies. Although there are a number of examples of resistance to drugs which appear to be largely under the control of a single gene (e.g. resistance to chloroquine in *Plasmodium falciparum* [189] and to benzimidazole in *Haemonchus contortus* [190]), most of the parasite traits which interest us from an epidemiological viewpoint are likely to be polygenic, quantitative traits [19]. A number of studies over a wide range of taxa have shown that there is a poor correlation between the among-group genetic variance in neutral, single gene markers (typically measured by $F_{ST}$) and the among-group genetic variance in complex, quantitative traits (as measured by $Q_{ST}$) [191-193].

7.1. Finding genetic markers for quantitative traits

An obvious solution to this problem of non-correspondence between $F_{ST}$ and $Q_{ST}$ would appear to be to utilize more non-neutral genetic markers, particularly quantitative trait loci (QTLs), in molecular epidemiological studies. There are two common approaches to the identification of QTLs for polygenic traits; linkage studies, which test for correlated
segregation of marker alleles and phenotypic values in families or the offspring of experimental crosses, and association studies, which test for the correlated occurrence of marker alleles and phenotypic values in natural populations [194]. Either approach may use neutral markers, which are distributed randomly throughout the genome, or candidate markers, which are chosen because of some a priori evidence of their effect upon the trait of interest.

The dominant design for the last 20 years has been linkage analysis with neutral markers to find regions of the genome in which putative QTLs may be located, followed by a search for candidate loci within the mapped region. This approach has been used successfully to map QTLs for traits of epidemiological significance in a number of parasite species; for example drug resistance in \textit{Plasmodium falciparum} [195], virulence in \textit{Toxoplasma gondii} [196] and resistance to malaria parasites in \textit{Anopheles gambiae} [197]. Linkage analyses have been so successful because, as closely related individuals have large portions of their genome in common, a relatively small number of polymorphic neutral markers (usually 500 or less) are usually sufficient to detect linked regions [198]. They do, however, require sophisticated genetic crosses, which are often not feasible for parasites with complex life cycles and, while they are more powerful than association studies for detecting QTLs with large effects on a trait, they are less powerful for detecting QTLs with small effects, because patterns of allele sharing in such cases will be less striking between relatives than between unrelated individuals [198,199].

Association analyses have been limited in the past by the need for very dense neutral marker coverage of the genome (typically tens or hundreds of thousands of markers) and a paucity of suitable candidate genes. These limitations are rapidly being overcome. In recent years, the growth of genome databases for an increasing number of parasite species has vastly increased the number and genome coverage of SNPs and polymorphic microsatellite loci, making genome-wide association studies much more feasible. At the same time, the development of new genomic/proteomic tools, such as cDNA microarrays, have greatly expanded the potential for identifying suitable candidate loci, even for species where genome databases are not available (195,200). While these new techniques hold much promise for the detection of QTLs for epidemiologically important traits, there are still a number of challenges that need to be faced.

7.2. The importance of understanding population structure

Unless candidate gene markers are being used, genome-wide association analyses rely on linkage disequilibrium (non-random association of alleles) between neutral markers and QTLs. Linkage disequilibrium can arise, not only because of physical linkage between loci, but also because of recent mutations, epistatic selection, genetic drift (especially in founding populations) or the admixture of genetically differentiated populations [201]. Asexual reproduction and inbreeding, which limit recombination between loci, enhance the maintenance of linkage disequilibrium and QTLs are therefore more likely to be detected in parasite species with a clonal population structure [194].
For species which have regular cycles of sexual reproduction, careful selection of populations for association studies can increase the chances of QTL detection; isolated populations which have been recently derived from a small number of founders, for example, should have enhanced linkage disequilibrium and would provide good targets for an initial genome-wide QTL scan. On the other hand, an inadequate knowledge of population genetic structure can lead to false-positive associations between markers and QTLs. This may arise if unobserved subgroups within the population, which will differ in allele frequencies throughout the genome, also differ in mean values for the epidemiological trait of interest [199]. This is of particular concern for parasites, because subgroups are often indistinguishable morphologically and can only be detected by a thorough population genetic study.

7.3. The importance of understanding genetic architecture

QTLs, once identified, can potentially be used to directly examine the genetic structure of epidemiologically important traits, which will have major advantages for identifying risk factors for parasitic disease and predicting the outcomes of treatment and control regimes. Caution will be needed, however, in the interpretation of such data. Theoretical analyses suggest that, unless the QTL explains a large proportion of genetic variance in the trait, then among-group QTL variance, just like among-group neutral marker variance, may be uncorrelated with genetic variance in the trait itself [202]. The extent of trait variance explained by a QTL is essentially a question of genetic architecture, i.e. of the number, effect size and allelic interactions of all the loci affecting the trait [203].

A number of mapping studies have identified QTLs explaining a substantial portion of the variance in quantitative epidemiological traits. Ferdig et al. [204], for example, found that 65% of the variance in quinine resistance in Plasmodium falciparum was attributable to allelic variation at two loci, while Behnke et al. [196] were able to explain 90% of the variance in virulence between two clonal lineages of Toxoplasma gondii by allelic variation at a single QTL. Results such as these suggest a skewed distribution of effects; with a few QTLs with relatively large effects (major genes) influencing most of the genetic variance, while the remainder of the genetic variance is influenced by a large number of QTLs with much smaller effects (minor genes). This is an encouraging finding, for it might indicate that we can concentrate on major genes and effectively ignore the contribution from minor genes when investigating the genetic structure of epidemiologically important traits. Unfortunately, the reality is likely to be much more complex. First, QTL mapping studies invariably tend to underestimate the number of QTLs affecting a trait and overestimate their effects, especially when small numbers of progeny are analysed [205,206]. Second, mapping studies do not necessarily provide a reliable guide to the contribution made by a QTL to the genetic variance of a quantitative trait in natural populations, because the extent of this contribution depends critically upon allele frequencies, which will differ among populations [207]. Finally, there is increasing evidence that epistatic effects (interactions among loci) explain a substantial portion the genetic variance in many quantitative traits [203,206,208]; this means that the effect of any particular QTL will vary depending on the genetic background.
There is, therefore, an urgent need to investigate the genetic architecture of epidemiologically important traits in parasites. This work has hardly begun. Indeed, even basic studies on the heritability of quantitative traits in parasites are rare, principally because traditional quantitative genetic analyses rely on determining the covariance structure for phenotypic resemblances between organisms with known degrees of relatedness. Anderson et al. [209] show how this can be achieved without using expensive and complicated breeding designs, by inferring relatedness from allele sharing at microsatellite loci in natural populations of Plasmodium falciparum.

7.4. The importance of understanding the phenotype

Finding and utilising QTLs for traits of epidemiological importance will be enhanced, not only by well designed mapping studies using the rapidly developing array of new genomic and proteomic tools, but also by well defined and accurately measured phenotypes [195]. This is a particular problem for complex, emergent traits, such as virulence, which depend on both parasite and host for their expression. Virulence is usually defined as the parasite-induced increase in host mortality or reduction in host fitness. While it is often regarded as a parasite trait, it should be more properly thought of as a trait which emerges from an interaction between parasite and host [210,211].

For example, differences in virulence between the three main clonal lineages of Toxoplasma gondii have been mapped to a small number of major genes or “intrinsic” virulence QTLs, which encode kinases or pseudokinases found in apical secretory organelles of the parasite [196,212,213]. The measure of virulence in all these studies was the mortality rate of mice challenged with a standard intra-peritoneal dose of parasites. Virulence, however, is not a static property of the genome of T. gondii and strains of the parasite that are highly virulent in one species of host may be completely avirulent in another host species [62]. From the point of view of the parasite, host species is an environmental factor that may markedly influence the phenotypic expression of virulence.

For emergent traits such as virulence, there are advantages in more carefully defining the parasite components of the phenotype, for example traits that affect invasion of the host, nutrient acquisition, modulation of the host cell cycle and evasion of the host immune response. Such “intermediate” traits [214] may be able to be measured more precisely and will be influenced by a smaller number of genetic and environmental factors than the emergent trait. QTLs influencing intermediate traits should therefore be easier to identify in association studies, they should explain a greater proportion of genetic variance in the trait and their influence should be less affected by different hosts or other environments.

8. Conclusions

Molecular epidemiological studies have had a major impact on the identification and classification of parasites, which has in turn improved our ability to unravel transmission
patterns and identify risk factors for parasite infection. The development of new molecular tools allowing relatively cheap, high throughput molecular assays and of new analytical approaches for population genetic studies will continue to drive a deeper understanding of parasite ecology and the epidemiology of parasitic disease. A major challenge for the immediate future will be to correlate genetic and phenotypic variation to identify markers for complex epidemiological traits, such as drug resistance, zoonotic potential and virulence.

References


[64] Wendte JM, Gibson AK, Grigg ME. Population genetics of *Toxoplasma gondii*: new perspectives from parasite genotypes in wildlife Vet Parasitol 2011; (in press).


[125] Slapeta J. Naming of Cryptosporidium pestis is in accordance with the ICZN Code and the name is available for this taxon previously recognised as C parvum `bovine genotype’. Vet Parasitol 2011; (in press).


Table 1.
The characterisation of genetic diversity in parasites (modified from [20]). In some cases, there may be overlap between the tools (regions of DNA) used and function. This will depend on the group of parasites being studied and the level of variation detectable by a particular approach. Abbreviations: AFLP – amplified fragment length polymorphism; ITS – internal transcribed spacer; LAMP – Loop-mediated isothermal DNA amplification; mPCR – multiplex PCR; PCR – polymerase chain reaction; PCR-RFLP – PCR-coupled restriction fragment length.

<table>
<thead>
<tr>
<th>Function</th>
<th>Purpose</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination above species level</td>
<td>Systematics</td>
<td>Highly conserved coding regions e.g. SSU rDNA, certain mitochondrial genes</td>
</tr>
<tr>
<td>Discrimination between species</td>
<td>Systematics / diagnosis / epidemiology</td>
<td>Moderately conserved regions e.g. coding mitochondrial genes, ITS rDNA, and other loci (e.g. house-keeping genes such as GDH, TPI, HSP, Actin, etc.); mPCR, LAMP</td>
</tr>
<tr>
<td>Discrimination between intraspecific variants</td>
<td>Population genetics / breeding systems (e.g. cross vs self fertilisation) / host specificity / molecular epidemiology / conservation (e.g. predicting susceptibility to pathogens) / biosecurity (exotic and emerging pathogens)</td>
<td>Variable regions e.g. allozymes, RAPD, AFLP, PFGE, PCR-RFLP, pyrosequencing, mPCR, LAMP, qPCR</td>
</tr>
<tr>
<td>Discrimination between individual isolates / clonal lineages / subgenotypes / ecological interactions within host</td>
<td>‘Fingerprinting’ / Molecular epidemiology – tracking transmission of subgenotypes / sources of infection and risk factors / competitive interactions and course of infection</td>
<td>Fingerprinting techniques e.g. Mini / microsatellites, SSCP, qPCR</td>
</tr>
<tr>
<td>Genetic markers / linking phenotype and genotype</td>
<td>Identifying phenotypic traits of clinical and epidemiological significance, e.g. virulence, infectivity, drug sensitivity</td>
<td>Genotype linked to phenotype via i) genetic map; ii) RDA; iii) sequencing and / or RT PCR of genes thought to be linked to phenotypic traits</td>
</tr>
</tbody>
</table>
Table 2
Definitions of the main intraspecific taxonomic terms used for parasites, adapted from [7,8,19].

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolate</td>
<td>An intraspecific group of (typically asexually) reproducing microparasites (viruses, bacteria, protozoa) or larval trematodes or cestodes, that have been obtained from a particular host individual at a particular time. Does not necessarily imply a clonal group.</td>
</tr>
<tr>
<td>Stock</td>
<td>An isolate that has been cultured in the laboratory for some time.</td>
</tr>
<tr>
<td>Line</td>
<td>A subgroup of a reproducing isolate which is genetically and phenotypically distinct, and therefore represents a single, clonal lineage.</td>
</tr>
<tr>
<td>Strain</td>
<td>An intraspecific group of parasites that differs genetically from other such groups in one or more traits of relevance to the treatment or control of parasitic disease</td>
</tr>
<tr>
<td>Discrete typing unit (DTU)</td>
<td>Similar to strain, but generally reserved for a monophyletic group (clade)</td>
</tr>
</tbody>
</table>
Fig 1. Factors which promote recurrence of generations and therefore genetic structuring among infrapopulations of parasites.
(a) Clumped transmission of infective stages released into the environment. (b) Asexual multiplication of larvae in intermediate host. Open triangles represent definitive hosts, open rectangles represent intermediate hosts. Closed circles and diamonds are parasites from different infrapopulations. Adapted from [46].