

Population Genetics and Molecular Epidemiology of Eukaryotes*

RONALD E. BLANTON¹

¹Center for Global Health & Diseases, Case Western Reserve University, Cleveland, OH 44106

ABSTRACT Molecular epidemiology uses the distribution and organization of a pathogen's DNA to understand the distribution and determinants of disease. Since the biology of DNA for eukaryotic pathogens differs substantially from that of bacteria, the analytic approach to their molecular epidemiology can also differ. While many of the genotyping techniques presented earlier in this series, "Advances in Molecular Epidemiology of Infectious Diseases," can be applied to eukaryotes, the output must be interpreted in the light of how DNA is distributed from one generation to the next. In some cases, parasite populations can be evaluated in ways reminiscent of bacteria. They differ, however, when analyzed as sexually reproducing organisms, where all individuals are unique but the genetic composition of the population does not change unless a limited set of events occurs. It is these events (migration, mutation, nonrandom mating, selection, and genetic drift) that are of interest. At a given time, not all of them are likely to be equally important, so the list can easily be narrowed down to understand the driving forces behind the population as it is now and even what it will look like in the future. The main population characteristics measured to assess these events are differentiation and diversity, interpreted in the light of what is known about the population from observation. The population genetics of eukaryotes is important for planning and evaluation of control measures, surveillance, outbreak investigation, and monitoring of the development and spread of drug resistance. *This article is part of a curated collection.

INTRODUCTION

The study of molecular epidemiology of parasitic infections and their vectors is meant to answer the same types of questions as for bacterial or viral infections. As with bacteria and viruses, the molecular epidemiology of eukaryotic infections follows the distribution and dynamics of microbial DNA. The key difference, however,

is precisely this biology, which at times demands a distinct approach to molecular epidemiologic investigation of infections caused by eukaryotic organisms. In bacterial reproduction, each individual passes down an essentially identical copy of all the DNA to the next generation. Some eukaryotic pathogens also reproduce in this way, while many others reproduce sexually for at least part of their life cycle, and for a few it is not entirely clear (1). An asexually reproducing individual generates a clone of itself but can also exhibit promiscuous horizontal gene transfer, which is a major source of variation and adaptation (2). This is not sex, however. Sex is the biologically necessary programmed recombination (**crossing over**) and random shuffling (**reassortment**) of chromosomal DNA in the process of reproduction. (A glossary of abbreviations and terms in bold type is provided at the end of this article.) This results in an enormous reservoir of variation. For example, for schistosomes with 8 pairs of chromosomes, and considering

Received: 14 March 2018, **Accepted:** 6 August 2018,
Published: 2 November 2018

Editor: Michael Sadowsky, BioTechnology Institute, University of Minnesota, St. Paul, MN; Lee W. Riley, Divisions of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA

Citation: Blanton RE. 2018. Population genetics and molecular epidemiology of eukaryotes. *Microbiol Spectrum* 6(6):AME-0002-2018. doi:10.1128/microbiolspec.AME-0002-2018.

Correspondence: Ronald E. Blanton, reb6@case.edu

© 2018 American Society for Microbiology. All rights reserved.

Curated Collection: [Advances in Molecular Epidemiology of Infectious Diseases](#).

there are some differences on each chromosome of the pair, reassortment alone offers >65,000 possible combinations of alleles with each mating. This of course pales beside the human genome, with 46 chromosomes and 7×10^{13} potential combinations. This is in addition to variation from exchange of genetic segments with crossing over. In nature, bacteria are heterogeneous conglomerates or communities (2, 3), but when they cause disease, especially in epidemics, it is generally a clone (see [Laboratory Methods in Molecular Epidemiology: Bacterial Infections for a discussion of the meaning of “clone.”](#)) that is responsible and that will be tracked. Sexual reproduction in some protozoa, many parasitic worms, and most vectors never results in a clone (with the exception of identical twins). There is genetic conservation, however, within a group of organisms that tends to breed together. In genetics, this is the working definition of a population. For sexually reproducing organisms, the population is the epidemiologic unit to track. Within the group, allele frequencies and thus traits are conserved under well-defined conditions. The unique power of the genetics of populations is that it reflects not only present individuals but also the population's past and the potential for subsequent generations (4). Many parasites exhibit both sexual and asexual modes of reproduction, but these life stages are often distributed in different hosts. Treatment of their molecular epidemiology is doubly complex, but it can be simplified for many questions by considering their biology just in the human host. As with most genetics, the output from population genetics has to be interpreted within the context of the question being asked and the biology and epidemiology of organisms being studied. The whole field of population genetics is perhaps the most complex area of genetics, but it rests on simple precepts. This review outlines the basic models used in population genetics that are directly applicable to problems of public health epidemiology.

KEY POINTS

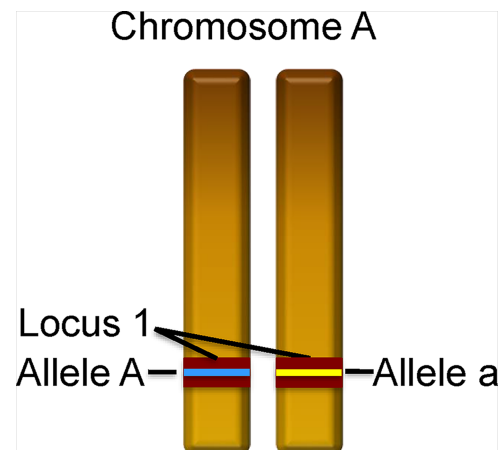
- Asexual reproduction usually produces a clone; sexual reproduction never does.
- A population is a group of organisms that tends to breed together.
- Allele and genotype frequencies describe populations.
- Allelic/genotypic frequencies and traits tend to be maintained within groups of interbreeding organisms (derived from the Hardy-Weinberg equilibrium).
- Gross changes in allelic/genotypic frequencies are due to a limited set of events.
- Allele and genotype frequencies can be used to infer population histories.
- Indices and statistics can be used to compare and assess population history and to project population dynamics.

DEFINING GENOTYPE IN EUKARYOTIC ORGANISMS

Some expressions may not be familiar to some readers, so it is important to define these early and in simple terms. A glossary is provided at the end of this review. One of the dividing lines between the DNA of bacteria and sexually reproducing parasites and vectors of human disease is their physical structure and organization. Sexually reproducing organisms pass some portion of their life cycle in which their chromosomes exist as nearly identical pairs (**diploid**). Some organisms, the parasite causing malaria, for example, also have only one copy of each chromosome (**haploid**) during their asexual stage, and this is the stage that infects humans. A similar location on each of the chromosomes of a diploid individual or those of different haploid individuals is a **locus**, and differences between loci (nucleotides, insertions, deletions, copies) are **alleles** (Fig. 1).

The geometry of DNA also strongly differentiates eukaryotes from bacteria (Fig. 2). There are a number

FIGURE 1 There are 2 nearly identical copies of all chromosomes for diploid organisms, except those responsible for determining sex. The corresponding location on each chromosome is a locus. Specific differences or similarities within a locus are alleles. The identity of both alleles at a locus is the genotype. In this case the genotype is Aa, i.e., heterozygous. Were it AA, the genotype would be homozygous.



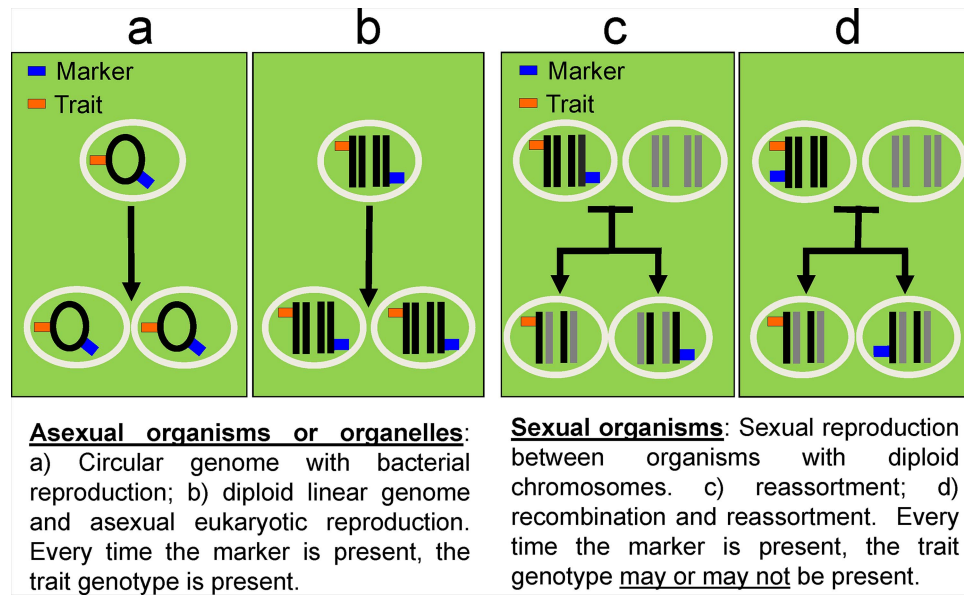


FIGURE 2 (a and b) Asexual organisms or organelles. **(a)** Circular genome with bacterial reproduction; **(b)** diploid linear genome and asexual eukaryotic reproduction. Every time the marker is present, the trait genotype is present. **(c and d)** Sexual organisms. Shown are diagrams for sexual reproduction between organisms with diploid chromosomes. **(c)** Reassortment; **(d)** recombination and reassortment. Every time the marker is present, the trait genotype may or may not be present.

of exceptions, but only a framework for understanding essential differences is presented here. Most prokaryotes have a single circular chromosome (*Leptospira* spp., *Vibrio cholerae*, and some others are exceptions with 2 circular chromosomes; several species also possess a linear chromosome), whereas even the simplest eukaryotes, such as yeasts, have multiple linear chromosomes. A specific marker on a bacterial chromosome is transmitted at reproduction together with any other marker or trait in the genome. The same also occurs with an asexually reproducing eukaryote despite its having multiple linear chromosomes. A marker on the genome of a sexually reproducing eukaryote, in contrast, will have a 50% chance of being transmitted away from any marker it is not very close to. Because of crossing over, markers near each other can be linked, i.e., occur together more frequently than expected by random chance in subsequent generations. This is physical **linkage** of the loci. At the population level, some loci that are not near each other can also be associated nonrandomly due to demography or population history. The nonrandom association between loci is termed **linkage disequilibrium** (LD). This is a statistical result and not necessarily recognition of a physical relationship. The labeling of each allele present at the same locus on each chromosome constitutes the genotype. A locus with the same poly-

morphism at the same site on each of the chromosome is **homozygous**, and one with a different polymorphism is **heterozygous**.

OPTIONS FOR MOLECULAR EPIDEMIOLOGY OF EUKARYOTES

- Study asexual parasites and clonal groups.
- Use markers in organelles that behave like bacteria (e.g., mitochondria).
- Use a marker close to the trait of interest (if known).
- Use many markers throughout the genome or perform sequencing.
- Study the whole group of organisms in which the trait is present (population genetics).

HARDY-WEINBERG EQUILIBRIUM: THE POPULATION NULL HYPOTHESIS

A population has a mathematical definition based on allele frequencies, which ultimately contributes to the development of tools for key measures of differentiation and diversity. Allele frequencies can differentiate populations, and genotypic frequencies can do so with

even greater resolution. The Hardy-Weinberg equilibrium (HWE) describes a relationship between allelic frequency and genotypic frequency. It is a simple mathematical relationship that is the definition of a population. If we use the letters “A” and “a” to represent different alleles at a single **diallelic** locus and “ p ” and “ q ” to represent their respective frequencies, a population with a p of 0.8 and q of 0.2 is clearly different from a population in which p is 0.2 and q is 0.8, especially where this kind of result is found at multiple loci. Allele frequencies are not always the most sensitive measure of differentiation, however. The same allele frequency may still be found in what are clearly distinct populations if assessed for genotypic frequencies. Alleles combine to form genotypes, so allelic frequencies are closely related to genotypic frequencies. For a diallelic locus where we know the frequency of each allele, the sum of these frequencies is 1 (or $p + q = 1$). For sexually reproducing organisms, the next generation arises from the combination of alleles from a pool of males with alleles from a pool of females. If we imagine that individuals from these pools will pair at random, the subsequent distribution of alleles in genotypes is equivalent to flipping 2 coins. For independent, random events, the probability of 2 events occurring simultaneously is the product of their frequencies [$(p + q)_{\text{female}} \times (p + q)_{\text{male}} = 1$]. The genotypic frequencies of the offspring for such a population should be $p^2 + 2pq + q^2$, if all assumptions are met, where p^2 and q^2 are the frequencies for the homozygotes and $2pq$ the heterozygotes. This expression is the **HWE**. This simple quadratic equation is the basis for all population genetics even when it is not measured directly. It represents the expected genotypic frequencies for a given set of allelic frequencies. It is one of the most stable mathematical relationships in nature. It is so much the expectation that when not observed in sequencing projects, it can suggest sequencing errors. It is the null hypothesis and mathematical definition of a stable population at equilibrium; nothing is happening.

The relationship HWE describes is true under a set of 5 to 10 assumptions that represent the most important factors that influence population genetic structure. The 5 most common assumptions are that there is

1. Random mating (**panmixia**, assortative mating)
2. No selection
3. No migration
4. An infinite population
5. No mutation

It is rare to have any of these assumptions met in nature, but the proportions are so resilient that the assumptions have to be severely violated to disturb this relationship. Even if they are violated, the proportions will be reestablished within 1 or 2 generations once the population is stabilized. As with most models, the underlying assumptions are the most important aspects. They are the basis for most conversations in population genetics.

MARKERS

Microsatellites, single nucleotide polymorphisms (SNPs), and sequencing are currently the marker types most employed in population genetics. Microsatellites are short tandem repeats of 2 to 8 nucleotides (reviewed in reference 5). Microsatellites have fallen out of favor in studies of statistical genetics or gene finding, since SNPs and sequencing provide better resolution at the level of individuals. Microsatellites, however, remain important in population genetics since they are mostly neutral for selection and have higher allelic richness and information content than single SNPs, and individuals are not the focus of population genetics. The microsatellite's rapid mutation rate (10^{-2} to 10^{-5} per generation) and stepwise mode of mutation can limit their application to questions that extend over short time scales and to certain statistical approaches. SNPs have lower rates of mutation (10^{-8}) in eukaryotes, often are diallelic, are 10 times more abundant (6, 7), and have high processivity and scorability. Sequencing essentially provides a very dense panel of SNPs and identifies rare variants as well as structural polymorphisms (copy number or insertions/deletions).

Mitochondrial and ribosomal DNA markers are much less abundant, less polymorphic, and thus less informative than microsatellites or SNPs. Some are under selection, and in the case of mitochondrial DNA, the genome is haploid (only 1 copy of chromosomal DNA) and may or may not have sex-specific inheritance depending on the species (8). They are useful for phylogeny studies, may be more economical to use in laboratories with limited capabilities, and are sometimes combined with other markers. Some aspects of malaria, such as multiplicity of infection or populations on a continental scale, are studied as asexually reproducing organisms, which they are in the human host. The ideal markers for this turn out to be from organelle DNA, since these structures behave like bacteria. The analysis then is based on markers usually scored simply as present or absent as for DNA barcodes (9).

MEASURES OF DIFFERENTIATION AND DIVERSITY

Evolution and conservation are the areas most often addressed by using population genetics. These two areas deal with essentially the same phenomenon but at different time scales; thus, the questions, the approaches, and the interpretation will differ depending on the nature of the problem. The relevant public health questions in population genetics focus on identity and dynamics of the group rather than individuals, extend over short time scales, and are directed at the control or extinction of a parasite or vector. Who is sleeping with whom, modes of reproduction, evolution, and the last common ancestor are all important in different contexts. These issues may be useful to help explain anomalies and can influence interpretation, but they are rarely answers to issues of control or intervention. Understanding how diverse a population is or the degree of difference between populations, combined with good study design, will contribute directly to determining the impact of control measures, host or parasite demographics, resistance, risk, and resilience or fragility of the population.

The field of population genetics depends heavily on mathematical analyses, some simple and some very sophisticated, to answer these questions. Mathematical treatments of all of the indices and statistics of differentiation and diversity can be easily obtained online or from textbooks or other publications and are not included here. Fortunately for the mathematically challenged, many open-source, individual computer programs and modules are available in R. The risk that goes with all readymade programs is a failure to understand what is being asked or the assumptions and limitations of the approach being taken. A list of some frequently used programs is provided at the end of this review ([Table 1](#)).

POPULATION DIFFERENTIATION INDICES

In addition to the HWE, populations can be further differentiated by other statistical tests. These are based on comparing population heterozygosity (F_{st} , G_{st} , G'_{st}) or the effective number of alleles (Jost's D).

F_{st} , G_{st} , and G'_{st}

In the 1950s, Sewall Wright and Gustave Malécot developed a family of indices (the fixation indices), one of which (the F_{st}) describes the likelihood of homozygosity (fixation) at a single diallelic locus based on **heterozygosity** of a subpopulation (s) compared to the total population (t). Theoretically, values should range from

0 (identity—every individual is genotypically the same) to 1 (no similarity—every individual is genotypically different). Nei extended the F_{st} to handle the case of more than two alleles and developed the G_{st} ([10](#)). Although the term F_{st} is often used in the literature, formally most studies today perform the G_{st} . When highly polymorphic loci, such as microsatellites, are genotyped, the G_{st} can severely underestimate differentiation and will not range from 0 to 1. Hedrick adjusted the range of values for the G_{st} by dividing it by its maximum possible value given the markers used and forcing the range back to 0 to 1 ([11](#)). This is the G'_{st} . The G'_{st} makes possible the full numerical range of differentiation. F_{st} and G_{st} relate to inherent properties of populations, contain evolutionary information, and approximate how close populations are to fixation ([12](#)). F_{st} -like measures have been and continue to be widely used to describe population structure, and their characteristics and behavior are well known. There are additional related indices (e.g., ϕ_{st} [[13](#)], analysis of molecular variance (AMOVA), R_{st} [[14](#)], and θ [[15](#)]) that address other aspects of differing genetic models, unequal sample sizes, accounting for haploid genomes, mechanisms of mutation, and selection.

Jost's D

Jost's D is sometimes referred to as simply D , but since there are numerous other D 's related to genetics and statistics **Jost's D** (after its creator, Lou Jost [[16](#)]) is used here. There can be logical inconsistencies for estimating differentiation based directly on heterozygosity. Ratios of pooled subpopulations to total population diversities tend toward zero when the subpopulation diversity is high ([16](#)). Unlike those based directly on heterozygosity, Jost's D is based on the effective allele number (see below). It has the property of yielding a linear response to changes in allele frequencies and is independent of subpopulation diversity. Unlike F_{st} , G_{st} , and similar indices, Jost's D does not carry information relevant to the evolutionary processes responsible for the present composition of a subpopulation. It is described by supporters and detractors as only a measure of differentiation ([17](#)). It was never meant to do more.

Whitlock provides one of the best comparisons of these 2 approaches ([17](#)):

This (Jost's) D differs from F_{st} in a fundamental definitional way: F_{st} measures deviations from panmixia, while D measures deviations from total differentiation. As a result, their denominators differ, and thus, the two indices can behave quite differently. D indicates the proportion of allelic diversity that lies among populations, while F_{st} is proportional to the variance of

TABLE 1 Selected population genetics programs^a

Program	Functions	Platform	Update	Authors	URL
Admixture	Clustering	Windows	2015	Alexander et al.	https://www.genetics.ucla.edu/software/admixture/
Arlequin†‡	HWE, Fst, AMOVA	Windows Mac Linux	2015	Excoffier and Lischer	http://cmpg.unibe.ch/software/arlequin35/
Bottleneck†	Expected heterozygosity	Windows	1999	Cornuet et al.	http://www1.montpellier.inra.fr/CBGP/software/Bottleneck/bottleneck.html
diveRsity	HWE, Fst, Ar	R package	2017	Keenan	http://diversityinlife.weebly.com/
DnaSP‡	Fst, Gst	Windows Mac Linux	2018	Rozas et al.	http://www.ub.edu/dnaspl/
estMOI	Haplotype generation	Windows	2014	Assefa et al.	http://pathogenseq.lshtm.ac.uk
Fluctuate†	Ne	Windows Mac Linux	2002	Kuhner	http://evolution.genetics.washington.edu/lamarck.html
FSTAT	HWE, Fst, Rst	Windows	2005	Goudet	http://www2.unil.ch/popgen/softwares/fstat.htm
Genealex†	HWE, Ae, Fst, Rst, Gst, G'st, Jost's D, Nm and AMOVA	Windows Mac	2016	Peakall and Smouse	http://biology-assets.anu.edu.au/GenAlEx/
Genepop	HWE, Nm, Fst*	Online Windows Mac Linux Unix	2017	Raymond and Rousset	http://genepop.curtin.edu.au/
Genetix	HWE, Fst, Gst	Windows	2004	Belkhir et al.	http://kimura.univ-montp2.fr/genetix/
MLNE†	Ne, m, Nm	Windows	2003	Wang	https://www.zsl.org/science/software/mlne
mmod	D_{est} , G'st, and ϕ_{ST}	R package	2017	Winter et al.	https://github.com/dwinter/mmod
Pegas	HWE, Fst, AMOVA	R package	2017	Paradis	https://cran.r-project.org/web/packages/pegas/index.html
Popgen	HWE, Ar	R package	2016	Kamvar	http://popgen.nescent.org/
PopGen Report	HWE, summary statistics, D_{est} , G'st and ϕ_{ST}	R package	2015	Adamack and Gruber	http://www.popgenreport.org/
Poptree2	Gst, standardized Gst, and Jost's D	Online, Windows, DOS, UNIX	2014	Takezaki et al.	http://www.med.kagawa-u.ac.jp/~genomelb/takezaki/poptree2/index.html
SpadeR†	Jost's D	Online R package	2015	Chao	http://chao.stat.nthu.edu.tw/wordpress/
SPAGeDi	Fst, Gst, Rst	Windows Mac Linux	2015	Hardy and Vekemans	http://ebe.ulb.ac.be/ebe/SPAGeDi.html
Structure	Clustering	Windows Mac	2012	Pritchard et al.	http://web.stanford.edu/group/pritchardlab/structure.html
Tfpga	HWE, Fst, genetic distance	Windows	2000	Miller	http://www.marksgeneticssoftware.net/tfpga.htm

^aMost programs perform additional analyses, but the emphasis here is for measures of differentiation and diversity discussed in this part. All programs accept discrete genotypes, except DnaSP. Some (†) accept allele frequency data and others (‡) raw DNA sequences. Programs producing estimates of Fst using Weir and Cockerham (15) are indicated by an asterisk. Additional lists of programs can be found in the work of Excoffier and Heckel (128) or at the following websites: Population genetics in R (http://grunwaldlab.github.io/Population_Genetics_in_R/), University of Washington (<https://courses.washington.edu/popgen/Software.htm>), and Softlinks (<http://softlinks.amnh.org/popgen.html>).

allele frequency among populations. D is more related to the genetic distance between populations than to the variance in allele frequencies; it may be preferable to call D a genetic distance measure.

There has been controversy about the use of these different types of indices. There should not be. They address different but related questions and resolve different analytic problems. It should be recognized that the G'st and Jost's D yield similar results when the number of populations is small and the markers have a small number of alleles. The G'st and Jost's D have given similar results in our own studies using microsatellites (18) and in simulation (17), with G'st giving values slightly higher than those for Jost's D. Some authors recommend calculating both Gst and Jost's D, in part to satisfy everyone and in part to obtain the useful infor-

mation about population diversity their departure may provide.

In relation to public health, most questions about parasites and vectors deal with near-term events of 2 to 20 generations, rather than events occurring on an evolutionary scale. The result we often seek is simply to know how different the allelic composition of one group is compared to another. In addition, the population or subpopulation structures may be such that there cannot be strong assumptions of HWE. The use of other markers, such as SNPs and sequencing, may require other approaches, such as principal-component or factor analysis (19). Apart from differentiation, individuals can be clustered into populations by model-based, structured association methods used in programs such as Structure (20) and Admixture (21). These programs use

approximation to HWE between populations as part of their clustering rule. They do provide a value for F_{st} to quantify the differences between the clusters formed.

DIVERSITY

Diversity, like differentiation, has myriad formulations and interpretations. The simplest expression is mean **heterozygosity** (H). For microsatellite data, this is usually high due to the intrinsically high mutation rate of these markers, and markers with higher variability are usually selected. Allelic richness (A_r) is simply a count of the number of alleles at each locus. Differences in sample size and marker choice will necessarily result in differences in allele number. This can be adjusted for by statistical methods such as rarefaction (22) to standardize sample sizes between comparison groups. The effective allele number (A_e) is also a measure of diversity and needs no adjustment for sample size. The maximum heterozygosity under HWE occurs when the frequencies of each of the alleles at a locus are equal. The A_e represents the number of alleles with equal frequencies in an ideal population that will produce the same heterozygosity as that of the actual population.

The most informative measure of diversity is the effective population size (N_e), a concept also introduced by Sewell Wright. It addresses the essential reason that diversity is important; namely, it reflects the strength of **genetic drift**. Genetic drift is the effect of random transmission of alleles during reproduction to succeeding generations. When numbers of reproducing individuals are small, the genetic composition of the population of offspring can differ, by chance, from what is expected given the composition of the parents. For examples, if two coins are flipped, it would not be that unusual for both to come up heads. If a thousand coins are flipped, the ratio of heads to tails will always be very near the expected 50:50 ratio. Thus, genetic drift is like what happens when few coins are flipped. Chance variation from the expected distribution is stronger when populations are small or reduced.

N_e can represent the number of actively breeding individuals in the population (inbreeding N_e) or the number of individuals in an ideal population needed to reconstitute the diversity in an actual population (variance N_e). It is almost always less than the census population. It is a key value in conservation genetics and population genetics in general, since it reflects the history and potential of a population. Increasing drift (decreasing N_e) tends to neutralize the force of directional selection, permits retention of deleterious mutations and

hampers the ability of populations to adapt to stresses (23). Like differentiation, there are several approaches to N_e that can produce different values and are designed to measure different aspects of diversity.

Despite its importance, N_e can be difficult to estimate in wild populations due to uncertainties of the demographic, genetic, and biological contexts (24, 25). It can be affected by sample size, overlapping generations, sampling interval, sex ratios, **gene flow**, age structure, variation in family size, fluctuating population size, or selection. Increasing the numbers of markers is less important than large samples for accurate estimates of N_e . Sampling the number of individuals that represent as much as 10% of the value of N_e has been recommended (24). Its interpretation can also be uncertain. Estimated N_e has been used as an aid in predicting extinction using the concept of a minimum viable population size. Some have suggested that a population at an N_e of 50 to 500 will experience extinction in the short or long run, respectively (26). Others have argued that this might occur at an N_e of 5,000 (27). While it is clear that lack of diversity has an impact on extinction (28), it is also clear that there cannot be a single number or range for the minimum viable population size for all populations (29, 30). In any case, theory suggests that there is a number defined by the amount of genetic drift below which populations are likely to go extinct on their own. Determination of the range for this number will be context specific and will require multiple species-specific studies under multiple conditions. This kind of analysis might contribute to developing a stopping rule as control measures approach elimination.

APPLICATION OF MOLECULAR TECHNIQUES TO THE EPIDEMIOLOGY OF EUKARYOTIC PATHOGENS

As we emphasize throughout reviews in this series, the molecular method employed to assess the epidemiology of a pathogen or vector depends on the question being asked and the biology of the organism's DNA. Many protozoa employ asexual reproduction, and the tools applied to their study resemble the pattern recognition approaches used for bacteria: DNA fingerprinting, repetitive element palindromic PCR, enterobacterial repetitive intergenic consensus, spoligotyping etc. Although this part has emphasized population genetics, the study of malaria and other protozoa with similar life histories may present a special case relative to helminths. While malaria is a sexually reproducing organism in the mosquito, in humans it reproduces asexually, and

its multiplicity of infection (MOI) is rarely more than 3 even when based on genomic sequence (31) and usually less than 2. Despite infection with multiple genotypes, these infections remain simple, so several simple marker systems and analyses suffice for some questions. Helminths, by contrast, tend not to reproduce asexually in humans, and infections are acquired one or a few parasites at a time over years (32). This produces a rich mix of multilocus genotypes, a situation not easily addressed by analysis of banding patterns on a gel. To differentiate species, or individuals of the same species, or to identify resistance when the gene(s) responsible is known, requires few markers and a readout for presence or absence. Depending on the question, however, larger numbers of polymorphic markers and standard population genetics approaches offer better differentiation, an understanding of the reasons for diversity and population histories, as much for malaria as for helminths.

A major role for molecular techniques in infection control programs is detection. In many cases, PCR is more sensitive (33) than microscopy and permits differentiation of closely related species, such as the filarial worms (34), hookworm species (35), or *Leishmania* (36, 37). Beyond this, simple prevalence or incidence of infection alone may not reveal local differences in the stability of the parasite population or its response to control measures. An analysis of population genetics with adequate sampling better reflects the whole population, since the genetics represents sampling from a much wider range of individuals and over several generations. The pattern of transmission of DNA over generations in itself provides information about the relationship of multiple unseen members of the species. For example, reduction in population size may be accompanied by increasing LD where inbreeding and genetic drift result in reduced diversity (38). An increase in differentiation from one period to the next may indicate a new introduction or an epidemic burst from a resident subpopulation. Similarity between geographic isolates (low F_{st}) also suggests a high degree of movement between regions, i.e., high gene flow. Interpretation still depends on the context: presence of control measure, change in morbidity or prevalence, a change in vectors, new hosts, etc. Each of these conditions may require a different public health response. Some examples of the application of molecular methods to address problems of parasite epidemiology are given in this section.

Molecular methods become essential for the assessment of population status when transmission is declining in order to improve detection, to focus attention on “hot spots,” to evaluate the impact of interventions, and to

scan for evidence of resistance. This section is a partial review of studies that apply molecular techniques to questions of public health and parasites. This section also has a heavy emphasis on malaria and some helminths. *Schistosoma mansoni* and *Plasmodium vivax* are addressed in greater detail in later reviews, but a few examples are included here. Given the importance of *Plasmodium falciparum* for global health, a large amount of attention is focused on this pathogen in this work.

Distinguishing Recrudescence from Reinfection and Determining MOI

For studies of treatment efficacy, new infections must be distinguished from persistence of the pretreatment infection. For most *P. falciparum* studies, the highly polymorphic loci merozoite surface proteins 1 and 2 (*msp1* and *msp2*) and the glutamate-rich protein (*glurp*) have become standard tools for infection control (39, 40). They are specifically recommended by the WHO (41) to accompany trials of drug efficacy, and there is extensive literature on their application for this purpose (39, 42–53).

In addition to distinguishing persistence from reinfection, *msp1*, *msp2*, and *glurp* are commonly used to determine MOI. Prior to molecular analyses, human infections with malaria were often considered completely clonal. We now know that infection with multiple genotypes is common, albeit with a small number of individuals, so the MOI ranges from 1 to 3. MOI is used as an indicator of the intensity of transmission and diversity (54). While this set of markers clearly has utility and often performs the purpose for which it was designed, it has some significant technical and analytic limitations. To reduce cost, these markers are often analyzed on agarose gels. The joint amplification of these markers in multiplex assays, and the resolution available on agarose gels, likely underestimates diversity compared to their detection by capillary electrophoresis (39). There is also inherent variability in agarose gel electrophoresis that needs to be taken into account when establishing criteria for band sizes (49, 51). Even under ideal conditions, the markers tend to underestimate MOI compared to sequencing approaches (31). As control programs achieve lower prevalence, the loss of diversity can further decrease the ability to differentiate relapse from new infections locally (50), while genetic differentiation between foci of infection increases (55). The small number of markers also does not permit a higher-order analysis of population dynamics such as effective population size or differentiation.

Distinguishing Epidemic from Endemic Disease Occurrence

Reduced variability is a key characteristic of founder, bottlenecked, and outbreak populations (56). With bacteria and some protozoa (57), an outbreak often means the presence of a clonal population. In an outbreak of *P. falciparum* in Cape Verde reported by Arez et al. (57), the population was genetically homogeneous based on the *msp1*, *msp2*, and *glurp* loci. Since the area was previously one of very low transmission, this suggested that the increased incidence was due to a single introduction and single initial transmission, given that the parasite needs to complete a sexual cycle for continued transmission. More commonly, for outbreaks with sexually reproducing organisms, diversity is low but not clonal and LD is high. These characteristics have previously identified a spike in *P. vivax* infections in the Amazon basin as an epidemic circulating on top of endemic transmission (58).

Diversity can also be used in other ways to interpret population structure. Where there is a low level of endemic circulation, there will be low diversity, fewer multiple infections, and high LD due to larger amounts of inbreeding. This was the expectation in a survey of 130 samples from 7 districts in Bangladesh, where malaria is considered hypoendemic based on the number of symptomatic cases. Instead, a relatively high MOI and high heterozygosity, more typical of regions of hyperendemicity in Africa, were observed, and this is consistent with a much larger population than expected. This indicated to the authors that the current national strategy should now include asymptomatic cases, since they likely contributed to the overall size of the population and persistence (54).

Distinguishing Pathovars versus Commensal Flora or Saprophytes

Distinguishing pathogenic varieties (**pathovars**) of parasites from nonpathogenic ones is primarily a question of distinguishing species. This task, therefore, employs marker systems that provide simple “present” or “absent” responses. A classic problem in the diagnosis of amebiasis (a complex of conditions including severe dysentery and liver abscess) is how to distinguish *Entamoeba histolytica* (the principal cause) from *Entamoeba moshkovskii* (an uncommon cause) and *Entamoeba dispar* (a commensal). They are all morphologically indistinguishable. Serologic tests have been developed but are less sensitive than PCR for 16S rDNA (33, 59). The diagnosis of large tapeworms is similarly difficult primarily due to a newly identified species,

Taenia asiatica. For many decades, dogma held that there were 2 species of tapeworms in the genus *Taenia* infecting humans. Differentiation was important, since the species using the pig as an intermediate host (*Taenia solium*) produced an invasive intermediate form causing serious neurologic disease in humans (cysticercosis), while the species developing in cattle (*Taenia saginata*) did not. These species could be distinguished morphologically. In the early 1990s, however, a parasite with the morphological features of *T. saginata* and infecting pigs was found to be invasive for humans (60). This new species was designated *T. asiatica*. Sequences within the valine tRNA and NADH genes were used to design a species-specific PCR capable of distinguishing these 3 important species. The complete sequence of the mitochondrial genome (61, 62) provided a level of resolution that allowed in-depth analysis of the population status and dynamics of this new pathogen. These studies also showed the close phylogenetic similarity of *T. asiatica* to *T. saginata* (63) as well as evidence of hybridization, which, unhappily, may indicate a path for rescuing characteristics of the pathogenic species from extinction (64).

Ascarid nematode worms have the ability to infect a variety of vertebrates, but it is uncertain whether worms from different hosts represent different species. When the pig pathogen *Ascaris suum* infects humans, the larvae can be invasive and may migrate aberrantly to the viscera, including to the central nervous system. The morphologically identical *Ascaris lumbricoides* is not invasive. The two can be differentiated by PCR amplification of the ribosomal internal transcribed spacer and by using a commercial kit based on NADH dehydrogenase subunit 5, yet there is still controversy (65–67). Molecular studies have not been sufficient to resolve this question, since it has not been clear what is meant by “species” in this case and to what degree it matters.

In addition to distinguishing pathogens from commensals, molecular techniques have also been used to differentiate more pathogenic strains from lesser pathogens within the same species. Infection with *Onchocerca volvulus* is a cause of blindness and a debilitating dermatitis, but blindness and other severe complications are more common with the savanna form than with strains found in the forest. In transition zones and in a landscape being shaped by human activity, these geographic descriptions do not always correspond to the parasite strain being transmitted. Several investigators (34, 68–70) helped develop a system for differentiating species and strains of *Onchocerca* based on PCR amplification of a highly repeated 150-bp element followed

by hybridization with a strain-specific probe (O-150). This marker system has been widely adopted by the WHO's Onchocerciasis Control Program to estimate the potential for severe disease in multiple populations of the parasite (71). The use of a single marker does not carry implications for status of the population in and of itself, but combined with demographic data, it can indicate migration events (72) and inform the planning of control/elimination measures (73–75).

Identifying New Hosts and Modes of Transmission

Many helminthic parasites distribute their development among different hosts, and the epidemiology of the infection is greatly influenced by the biology of these hosts. Further, the host range of a given parasite is not always clear. In several recent cases, the zoonotic contribution to human disease could be appreciated only by molecular techniques. Morphologically similar parasites may infect human and/or animal hosts without belonging to the same species (76). In other cases, there may be new or unappreciated hosts, and these must be verified using molecular techniques. The elimination of Guinea worm suffered a recent setback when it became apparent that dogs were previously unappreciated as or had become major hosts of the parasite (77, 78). Whole-genome sequencing verified the identity of the parasite in both humans and dogs as the same Guinea worm species, *Dracunculus medinensis*. This necessitated a reanalysis of the elimination strategy. Similarly, sequencing of *Hymenolepis nana* from humans and rats demonstrated that the same species was being transmitted and confirmed rodent involvement in the life cycle (79), with attendant implications for control measures.

In Denmark, human infection with ascarid worms is uncommon (1%), but the infection is commonly found in Danish pigs (25%). By PCR-restriction fragment length polymorphism analysis, parasites from infected Danes clustered with the local pig parasites rather than human infections elsewhere in the world (80). Thus, the few observed human infections in Denmark were native *A. suum* and not imported cases of *A. lumbricoides*. A similar association was observed in the United Kingdom (81). The implication for public health is that in these countries, whether or not there are separate species, human infection results from a zoonosis, and control in the human population should target control in pigs.

Finally, *Schistosoma japonicum*, a trematode, unlike the other major human schistosomes, is primarily zoonotic. The risk to humans tracks with the biology of the other infected species. Rudge et al. (129) observed local

geographic clustering of *S. japonicum* based on microsatellite genotypes and *F_{st}*. They also clustered with parasites infecting the predominant local nonhuman host species. This suggested that rodents might represent a more important reservoir than initially thought in some areas. The highest gene flow was between humans and domestic ruminants, as suggested by low *F_{st}* values and Bayesian clustering with the program Structure (20). Identification of the principal local hosts by a molecular approach contributes to planning local control measures.

Tracking Geographic and Temporal Distribution of Infectious Disease Agents

Panels of 23 or 24 SNP markers for *P. falciparum* (82, 83) have been derived from bacterium-like organelles, such as mitochondria and apicoplasts. These cellular structures have the advantage of not being subject to sexual recombination (84), and markers can be analyzed simply as present or absent, analogous to supermarket barcode labels. DNA barcoding was originally developed as a taxonomic tool. The 23-SNP barcode panel was able to classify most isolates by continent of isolation with 92% predictive accuracy. It was able to determine that a mystery isolate that infected a Dutch resident living near Amsterdam's international airport originated in the continent of Africa but was unable to determine if the source was from East or West Africa. A limitation of such markers is their resolution, since the questions are often over a continental scale and resolution deteriorates with increasing diversity. A subsequent panel of 42 nuclear SNPs forms a barcode that may have better resolution and is able to distinguish between populations from Brazil/French Guiana, Ethiopia, and Sri Lanka (85).

Genomic sequencing, while currently the most costly approach, has the advantage of providing higher resolution. For *P. vivax*, most studies have used sequences from the mitochondrial genome (86–90). Paradoxically, while differentiating regional populations, mitochondrial sequences were less useful for differentiating continental origins (90). Mitochondrial genome sequences were critical for identifying the origin for the 1993 re-emergence of vivax malaria in South Korea as coming from a regional source (89) rather than a more distant importation.

In and of itself, geographic origin may not be a key epidemiologic factor, unless associated with a pathogen phenotype, a new invasion, or as a factor in evaluation or planning for a control program. The study of longitudinal microsatellite diversity of *P. vivax* populations

has been able to resolve many questions about the nature of transmission (91). In the Amazon basin, for example, increased incidence during a 3-year study was associated with diminished genetic diversity signaling a localized outbreak by a limited number of strains. As the incidence fell, diversity returned to preoutbreak levels (58). In contrast, during a resurgence of vivax malaria in Greece (92) and Sri Lanka (93), high parasite diversity indicated contribution from multiple strains, consistent with rapid expansion of the local population. In the context of declining prevalence, this suggested multiple introductions rather than clonal expansion. Options for control and prevention differ for these 2 situations.

Understanding the Distribution of Resistance

Molecular tools for the study of the epidemiology of drug resistance rely primarily on identification of the gene or genes involved. The paradigm of “one drug, one resistance gene,” however, clearly does not always reflect reality for parasites any more than it does for bacteria. Antimalarials constitute the category of most concern at present. For chloroquine and the combination sulfadoxine and pyrimethamine (Fansidar), there is a well-defined and limited number of genes responsible. In contrast, artemisinin resistance is the result of a summation of several mutations in multiple genes, with key variants tipping the balance toward resistance (94, 95). These patterns correspond to what have been termed, respectively, hard selective sweeps, where resistance arises on one or a few genes, and soft selective sweeps, which are the result of the summation and concentration of multiple preexisting characteristics to result in a resistant phenotype. Even for what appear to be hard selective sweeps, multiple mutations are required in 2 genes for resistance to chloroquine (*pfcr* and *pfmdr1*) (96) and sulfadoxine-pyrimethamine (*pfdhps* and *pfdftr*) (reviewed in reference 97). The epidemiology of antimalarials, then, depends on the genetic mechanism for the development of resistance. Mutations within the kelch motif protein 13 gene serve as markers for artemisinin resistance, at least in Asia (94). An approach for detection of resistance without identification of a responsible gene is discussed in a later review.

Since 1996, the WHO has recommended national surveillance for antimalarial resistance every 2 years. Initially the advocated methodology was based purely on monitoring response to treatment (98), but in 2009 standard PCR was added for a more sensitive measure of infection and to distinguish persistence from reinfection. There is as yet no standard recommendation for surveillance based purely on identification of resistance

genes. A surveillance program for malaria and resistance genes revealed that subclinical infections were more frequent than symptomatic cases (OR 3.7) and served as a likely reservoir for chloroquine resistance genes *pfcr* and *pfmdr1* (99). The existence and importance of this reservoir could be uncovered only by a molecular method. Molecular surveillance of specific polymorphisms was also used to demonstrate high degrees of susceptibility (100–103), sustained resistance (104), or high risk for resistance (105) in specific countries. In each case, ongoing molecular surveillance is suggested as an early warning mechanism for the arrival or development of resistance. An excellent resource for the distribution of artemisinin, chloroquine, and sulfadoxine-pyrimethamine resistance mutants is the Worldwide Antimalarial Resistance Network website (<http://www.wwarn.org>). Detailed and updated interactive maps and tables are provided. Similar approaches are taken using equivalent *P. vivax* genes (106).

A limited number of markers is sufficient to track resistance within a country and across the globe when the gene(s) is known. Some of these studies allow assessment for resistance arising within a population, such as the assessment of mutations contributing to artemisinin resistance in migrants in Myanmar (105) or geographic clustering of sulfadoxine-pyrimethamine resistance in the Republic of the Congo (107). Another dimension to the potential for the spread of resistance is the degree of gene flow between populations. Since resistance usually arises in one or a few areas, the mobility of parasite genotypes in general will predict the rapidity and extent of spread of a resistant phenotype. Based on using multiple microsatellite markers, low differentiation indices ($F_{st} \sim 0.02$) predict the ready dissemination of any new resistance characteristic that arrives in areas of Mali (108) or the Sudan (109). In a similar way, measures of gene flow are useful for planning the release of modified mosquitoes in vector control programs (110).

Anderson et al. used SNPs to provide a new much higher estimate of N_e to study the development of artemisinin resistance in Southeast Asia (111). They were able to estimate that mutations in the most important resistance locus (*kelch13*) arose hundreds of times, implying that the drugs used in combination with artemisinin were likely to be only partially effective. This lent support to the use of 3 effective drugs for treatment.

Few resistance genes against anthelmintics have been identified, but the number of drugs in common use is small to begin with. Recently, a gene for resistance to the drug oxamniquine was found (112), but this drug is

no longer commonly used. The target of albendazole, a key drug for nearly all nematode infections, is β -tubulin. In veterinary practice, resistance is becoming common, since it relies on point mutations within this gene. Surveillance for some of these mutations in human parasites showed their frequencies to be high following treatment (113). Decreased efficacy, however, was observed only in *Trichuris trichiura* and not in hookworm or *Ascaris lumbricoides*, for which the relevant mutants were few.

Planning Control Strategies

The very structure of populations is a useful guide for planning control measures. The ability of parasites to migrate into a control area can be gauged by looking at differentiation as an indication of gene flow. In the Comoros archipelago, gene flow for *P. falciparum* was high between all 4 islands ($F_{st} < 0.05$), except the most southern of them (>0.10). This was still not considered a sufficient degree of isolation to prevent rapid repopulation. The authors felt that the best strategy would be control measures initiated simultaneously on all islands to avoid rapid repopulation (114). Similar degrees of gene flow with similar recommendations were observed on islands in Lake Victoria (115). A survey across West Africa likewise demonstrated extensive gene flow (most F_{st} values were <0.03), and for that reason, the authors were pessimistic about the development of a fragmented landscape that would facilitate elimination without a very extensive fall in prevalence (116). The distribution of *P. falciparum* in Malaysia stands in contrast to that in West Africa. Here endemicity has been declining for some time, and there are high levels of differentiation between sites and elevated LD, suggesting elevated inbreeding (55, 117). While the fragmentation suggests that an island-by-island approach may be successful for malaria control, the increased LD and even clonal expansion in some areas may make monitoring difficult, since it will become more difficult to distinguish recrudescence from reinfection where individuals are highly related.

Surveillance and Monitoring Response to Intervention

Molecular techniques contribute to the measurement of the success of control programs by increasing the sensitivity of detection and assessing the impact on target populations. In addition to morbidity reduction, the aim of many programs today is to eliminate transmission, which requires increasingly more sensitive methods of detection. Lamberton et al. (118) assessed the impact of ivermectin treatment on transmission of onchocerciasis

in Ghana, where treatment and control programs have been ongoing for 3 to 24 years, depending on locale. Using species-specific ribosomal markers and the O-150 marker on captured black flies, they observed a negative correlation between the prevalence of *O. volvulus* in the vector and duration of the treatment program, yet there was still vector positivity after 24 years of control efforts. This meant that treatment with ivermectin alone was unlikely lead to elimination. This emphasized the need for new approaches with triple-combination therapy, which is showing remarkable ability to maintain long-term elimination in humans (119) and is now recommended by the WHO.

Senegal intensified its malaria control efforts starting in 2006 through their National Malaria Control Program (NMCP), with apparent marked reduction in transmission. The long-term effect on the population could be tracked through analysis of 24 microsatellite markers (82, 83) as well as complete sequencing of a sample of parasites. SNPs developed from the sequences were optimized into sets of barcodes, and the extent of similarity and identity were used to model rates of reproduction. The pattern of inbreeding and extensive regions of LD mirrored the significant reduction from 2006 to 2010. Reexpansion of unique barcodes in 2013-2014 were consistent with the rebound observed in clinical incidence data (120).

In a similar fashion, Ingasia et al. in Kenya surveyed *P. falciparum* populations from the 2 well-defined ecological zones of lowland and highland transmission as well as Eastern versus Western Kenya (121). Of particular importance, this group wanted to assess the effect on the population of the introduction of artemether-lumefantrine. The study employed 12 microsatellite loci genotyped in >200 isolates from 5 sites. In Eastern Kenya, where the infection has been in decline due to intensive treatment and use of bed nets, parasite diversity was low. Otherwise, there was high diversity elsewhere in the country, suggesting high effective population size and little effect. In the highlands there was high within-population LD, but overall differentiation between sites was moderate ($F_{st} 0.072$). In Kenya, where gene flow is high, the prevalences of antimalarial resistance genes were comparable across geographic and ecological regions as predicted from population structure (122).

Not all efforts at malaria control have been associated with evidence of a genetic impact on the parasite population. Another study on falciparum malaria in Kenya employed 8 microsatellite loci to assess parasite genetic structure over a 10-year period following the introduc-

tion of insecticide-treated bed nets (123). The expectation was for a decrease in diversity but no significant changes in MOI, heterozygosity, or LD. Differentiation between the populations over this period remained low, at an average F_{st} of 0.04. The high diversity and low differentiation suggested that high gene flow helped maintain the population in this area. Haiti was another country where stepped-up efforts to control and eliminate malaria occurred, especially following the 2010 earthquake. Microsatellite analysis with 12 neutral loci indicated high gene flow, high diversity, and little evidence for a genetic bottleneck (124). The approach to control needed to be modified.

Finally, Malawi instituted a multipronged control effort directed at vector control with indoor residual spraying, widespread use of bed nets, rapid diagnosis, and treatment. Parasites were sampled 2 years before and twice during a 5-year period following implementation of these measures. Many of the methods mentioned in this review were applied to evaluate the impact of these efforts. No changes were noted for MOI, heterozygosity, genotypic richness, multilocus LD, or N_e . There was some modest reduction in heterozygosity for multiple-genotype infections (125), but the general conclusion was that these efforts were insufficient to effect significant change in the overall population and that other approaches should be incorporated. It was also possible that the sample size gave insufficient power to detect changes in all of the parameters measured. The report of this study further acknowledged 2 other studies of malaria (93, 126) with similar poor responses when the starting prevalence was very high.

CONCLUDING REMARKS

Some problems in the study of molecular epidemiology of eukaryotic pathogens can be addressed with a few markers and categorical interpretations for their presence or absence (e.g., MOI, savannah versus forest forms, and pathovar versus nonpathovar). Markers for antimicrobial resistance sometimes fall into this category, but as with bacteria, we are finding that resistance can be a complex and multigenic trait in parasites. The application of population genetics to the study of parasite epidemiology is more complex but promises enormous rewards in bringing to light processes at work within the pathogen community. Every control program whose aim is elimination should apply this kind of analysis to assess success and understand failure. Otherwise, these programs operate blindly without fully appreciating how the outcome, good or bad, was achieved.

A population genetic analysis is not a perfect crystal ball but contributes important clues about the continued risk of transmission, the quality of the intervention (coverage), with what resources the parasites as a group responded (drug resistance, immigration, reservoirs, force of infection, and persistence), and, consequently, how best to approach control or elimination.

TERMS AND ABBREVIATIONS

Ae. Effective allele number. Ae is a measure of diversity and is related to heterozygosity: $A_e = 1/(1 - H_{exp})$.

Allele. Point of variation at homologous loci.

Ar. Allelic richness, i.e., the total number of different alleles in a population. Ar is a measure of diversity.

Crossing over. The exchange of homologous segments of different chromosomes during gametogenesis in sexually reproducing organisms.

Diallelic. Having 2 alleles (types of nucleotide variation) at a given locus.

Diploid. The status of organisms or cells that have 2 homologous chromosomes.

Gene flow. The geographic movement of genes due to migration and reproduction of organisms.

Genetic drift. The effect of random transmission of alleles during reproduction to succeeding generations. Gamete and zygote formation in sexual reproduction is probabilistic, so expected combinations appear at expected frequencies when the numbers are large. When the numbers are small, unusual combinations occur. Populations with small numbers develop unexpected allelic and genotypic frequencies over time. For this reason, infinite population size is one of the criteria for the HWE.

Haploid. The status of organisms or cells with 1 copy of each chromosome.

Heterozygosity. The percent of a population that is heterozygous at a given locus or the average percent over all loci. This is often taken to be one measure of diversity, but it is dependent on the kind of marker used. Microsatellites often have high heterozygosity (50 to 80%), while SNP heterozygosity is generally less than 50%.

Homozygote/heterozygote. For diploid organisms, homozygotes have no variation between DNA sequence or structure at homologous loci, while heterozygotes differ.

HWE. Hardy-Weinberg equilibrium. HWE is a description of the relationship of allele frequency to genotypic frequency in an ideal population. It is the null hypothesis in population genetics.

LD. Linkage disequilibrium. LD is the observation of the joint occurrence of alleles at 2 or more loci more often than expected given their frequency. The concept refers to families and populations and involves demographic and evolutionary processes more than the physical geography of loci in the genome (reviewed in reference 38). The expectation under HWE is that there will be no association between alleles on different loci.

Linkage. The degree to which loci within the genome are inherited randomly due to their proximity. Loci that are close to each other tend to be transmitted together to the next generation.

Locus. An identified region of a chromosome.

MOL. Multiplicity of infection.

Ne. Effective population size. Ne is the sensitivity of a population to genetic drift, the number of breeding pairs in a population, or the minimum number of individuals in an ideal population required to reconstitute the diversity in a real population. It is a measure of diversity.

Panmixia. The state where all individuals have an equal chance of reproducing with all others.

Pathovars. Pathogenic varieties.

Reassortment. The mixing of genetic material in the process of reproduction. In sexually reproducing organisms, this includes the random assignment of chromosomes to the haploid sex cells (gametes) as well as the process of exchanging homologous sections of chromosomes (see crossing over). In viruses, such as influenza virus, segments from one viral strain or species may be mixed and segregated with those of a different strain.

CURATED COLLECTION

[Click here to read other articles in this collection.](#)

REFERENCES

1. Rougeron V, De Meeüs T, Kako Ouraga S, Hide M, Bañuls AL. 2010. "Everything you always wanted to know about sex (but were afraid to ask)" in *Leishmania* after two decades of laboratory and field analyses. *PLoS Pathog* 6:e1001004. <http://dx.doi.org/10.1371/journal.ppat.1001004>.
2. Rosen MJ, Davison M, Bhaya D, Fisher DS. 2015. Microbial diversity. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science* 348:1019–1023. <http://dx.doi.org/10.1126/science.aaa4456>.
3. Krause DJ, Whitaker RJ. 2015. Inferring speciation processes from patterns of natural variation in microbial genomes. *Syst Biol* 64:926–935. <http://dx.doi.org/10.1093/sysbio/syv050>.
4. Falconer D. 1960. *Introduction to Quantitative Genetics*. Ronald Press Company, New York, NY.
5. Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5:435–445. <http://dx.doi.org/10.1038/nrg1348>.
6. Hartl DL. 2014. *Essential Genetics*, 6th ed. Jones & Bartlett Learning, Burlington, MA.
7. Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17:6463–6471. <http://dx.doi.org/10.1093/nar/17.16.6463>.
8. Hoeh WR, Blakley KH, Brown WM. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251:1488–1490. <http://dx.doi.org/10.1126/science.1672472>.
9. Schindel DE, Miller SE. 2005. DNA barcoding a useful tool for taxonomists. *Nature* 435:17. <http://dx.doi.org/10.1038/435017b>.
10. Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* 70:3321–3323. <http://dx.doi.org/10.1073/pnas.70.12.3321>.
11. Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633–1638. <http://dx.doi.org/10.1111/j.0014-3820.2005.tb01814.x>.
12. Jost L, Archer F, Flanagan S, Gaggiotti O, Hoban S, Latch E. 2 January 2018. Differentiation measures for conservation genetics. *Evol Appl* <https://doi.org/10.1111/eva.12590>.
13. Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
14. Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.
15. Weir BS, Cockerham CC. 1984. Estimating *F*-statistics for analysis of population structure. *Evolution* 38:1358–1370.
16. Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Mol Ecol* 17:4015–4026. <http://dx.doi.org/10.1111/j.1365-294X.2008.03887.x>.
17. Whitlock MC. 2011. G_{ST} and D do not replace F_{ST} . *Mol Ecol* 20:1083–1091. <http://dx.doi.org/10.1111/j.1365-294X.2010.04996.x>.
18. Blank WA, Test MR, Liu SF, Lewis FA, Blanton RE. 2010. Long-term genetic stability and population dynamics of laboratory strains of *Schistosoma mansoni*. *J Parasitol* 96:900–907. <http://dx.doi.org/10.1645/JGE-2463.1>.
19. Fumagalli M, Vieira FG, Korneliussen TS, Linderöth T, Huerta-Sánchez E, Albrechtsen A, Nielsen R. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* 195:979–992. <http://dx.doi.org/10.1534/genetics.113.154740>.
20. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
21. Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664. <http://dx.doi.org/10.1101/gr.094052.109>.
22. Leberg PL. 2002. Estimating allelic richness: effects of sample size and bottlenecks. *Mol Ecol* 11:2445–2449. <http://dx.doi.org/10.1046/j.1365-294X.2002.01612.x>.
23. Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol* 25:438–449. <http://dx.doi.org/10.1111/j.1523-1739.2010.01637.x>.
24. Palstra FP, Ruzzante DE. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* 17:3428–3447. <http://dx.doi.org/10.1111/j.1365-294X.2008.03842.x>.
25. Wang J, Caballero A. 1999. Developments in predicting the effective size of subdivided populations. *Heredity* 82:212–226. <http://dx.doi.org/10.1038/sj.hdy.6884670>.

26. Franklin I, Frankham R. 1998. How large must populations be to retain evolutionary potential? *Anim Conserv* 1:69–73. <http://dx.doi.org/10.1111/j.1469-1795.1998.tb00228.x>.
27. Lande R. 1995. Mutation and conservation. *Conserv Biol* 9:782–791. <http://dx.doi.org/10.1046/j.1523-1739.1995.09040782.x>.
28. Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. *Proc Natl Acad Sci U S A* 101:15261–15264. <http://dx.doi.org/10.1073/pnas.0403809101>.
29. Flather CH, Hayward GD, Beissinger SR, Stephens PA. 2011. Minimum viable populations: is there a ‘magic number’ for conservation practitioners? *Trends Ecol Evol* 26:307–316. <http://dx.doi.org/10.1016/j.tree.2011.03.001>.
30. Traill LW, Bradshaw CJ, Brook BW. 2007. Minimum viable population size: a meta-analysis of 30 years of published estimates. *Biol Conserv* 139:159–166. <http://dx.doi.org/10.1016/j.biocon.2007.06.011>.
31. Assefa SA, Preston MD, Campino S, Ocholla H, Sutherland CJ, Clark TG. 2014. estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* 30:1292–1294. <http://dx.doi.org/10.1093/bioinformatics/btu005>.
32. Stothard JR, Sousa-Figueiredo JC, Betson M, Adiriko M, Arinaitwe M, Rowell C, Besiyege F, Kabaterine NB. 2011. *Schistosoma mansoni* infections in young children: when are schistosome antigens in urine, eggs in stool and antibodies to eggs first detectable? *PLoS Negl Trop Dis* 5:e938. <http://dx.doi.org/10.1371/journal.pntd.0000938>.
33. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements AC, Gomes SJ, Traub R, McCarthy JS. 2016. Application of a multiplex quantitative PCR to assess prevalence and intensity of intestinal parasite infections in a controlled clinical trial. *PLoS Negl Trop Dis* 10:e0004380. <http://dx.doi.org/10.1371/journal.pntd.0004380>.
34. Meredith SE, Lando G, Gbakima AA, Zimmerman PA, Unnasch TR. 1991. *Onchocerca volvulus*: application of the polymerase chain reaction to identification and strain differentiation of the parasite. *Exp Parasitol* 73:335–344. [http://dx.doi.org/10.1016/0014-4894\(91\)90105-6](http://dx.doi.org/10.1016/0014-4894(91)90105-6).
35. Zhan B, Li T, Xiao S, Zheng F, Hawdon JM. 2001. Species-specific identification of human hookworms by PCR of the mitochondrial cytochrome oxidase I gene. *J Parasitol* 87:1227–1229. [http://dx.doi.org/10.1645/0022-3395\(2001\)087\[1227:SSIOHH\]2.0.CO;2](http://dx.doi.org/10.1645/0022-3395(2001)087[1227:SSIOHH]2.0.CO;2).
36. Bensoussan E, Nasereddin A, Jonas F, Schnur LF, Jaffe CL. 2006. Comparison of PCR assays for diagnosis of cutaneous leishmaniasis. *J Clin Microbiol* 44:1435–1439. <http://dx.doi.org/10.1128/JCM.44.4.1435-1439.2006>.
37. Jara M, Adaui V, Valencia BM, Martinez D, Alba M, Castrillon C, Cruz M, Cruz I, Van der Auwera G, Llanos-Cuentas A, Dujardin JC, Arevalo J. 2013. Real-time PCR assay for detection and quantification of *Leishmania (Viannia)* organisms in skin and mucosal lesions: exploratory study of parasite load and clinical parameters. *J Clin Microbiol* 51:1826–1833. <http://dx.doi.org/10.1128/JCM.00208-13>.
38. Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9:477–485. <http://dx.doi.org/10.1038/nrg2361>.
39. Messerli C, Hofmann NE, Beck HP, Felger I. 2016. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. *Antimicrob Agents Chemother* 61:61.
40. Snounou G, Beck HP. 1998. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitol Today* 14:462–467. [http://dx.doi.org/10.1016/S0169-4758\(98\)01340-4](http://dx.doi.org/10.1016/S0169-4758(98)01340-4).
41. Felger I, Snounou G. 2008. *Recommended Genotyping Procedures (RGPs) To Identify Parasite Populations*. World Health Organization and Medicines for Malaria Venture, Geneva, Switzerland.
42. Adjuk M, Agnamey P, Babiker A, Borrmann S, Brasseur P, Cisse M, Cobelens F, Diallo S, Faucher JF, Garner P, Gikunda S, Kremsner PG, Krishna S, Lell B, Loolpapit M, Matsiegui PB, Missinou MA, Mwanza J, Ntoumi F, Olliaro P, Osimbo P, Rezbach P, Some E, Taylor WR. 2002. Amodiaquine-artesunate versus amodiaquine for uncomplicated *Plasmodium falciparum* malaria in African children: a randomised, multicentre trial. *Lancet* 359:1365–1372. [http://dx.doi.org/10.1016/S0140-6736\(02\)08348-4](http://dx.doi.org/10.1016/S0140-6736(02)08348-4).
43. Adjuk M, Babiker A, Garner P, Olliaro P, Taylor W, White N, International Artemisinin Study Group. 2004. Artesunate combinations for treatment of malaria: meta-analysis. *Lancet* 363:9–17. [http://dx.doi.org/10.1016/S0140-6736\(03\)15162-8](http://dx.doi.org/10.1016/S0140-6736(03)15162-8).
44. Cattamanchi A, Kyabayinze D, Hubbard A, Rosenthal PJ, Dorsey G. 2003. Distinguishing recrudescence from reinfection in a longitudinal antimalarial drug efficacy study: comparison of results based on genotyping of msp-1, msp-2, and glurp. *Am J Trop Med Hyg* 68:133–139.
45. Happi CT, Gbotosho GO, Sowunmi A, Falade CO, Akinboye DO, Gerena L, Kyle DE, Milhous W, Wirth DF, Oduola AM. 2004. Molecular analysis of *Plasmodium falciparum* recrudescence malaria infections in children treated with chloroquine in Nigeria. *Am J Trop Med Hyg* 70:20–26.
46. Imwong M, Jindakhad T, Kunasol C, Sutawong K, Vejakama P, Dondorp AM. 2015. An outbreak of artemisinin resistant falciparum malaria in Eastern Thailand. *Sci Rep* 5:17412. <http://dx.doi.org/10.1038/srep17412>.
47. Mugittu K, Adjuk M, Snounou G, Ntoumi F, Taylor W, Mshinda H, Olliaro P, Beck HP. 2006. Molecular genotyping to distinguish between recrudescence and new infections in treatment trials of *Plasmodium falciparum* malaria conducted in Sub-Saharan Africa: adjustment of parasitological outcomes and assessment of genotyping effectiveness. *Trop Med Int Health* 11:1350–1359. <http://dx.doi.org/10.1111/j.1365-3156.2006.01688.x>.
48. Mutabingwa TK, Anthony D, Heller A, Hallett R, Ahmed J, Drakeley C, Greenwood BM, Whitty CJ. 2005. Amodiaquine alone, amodiaquine +sulfadoxine-pyrimethamine, amodiaquine+artesunate, and artemether-lumefantrine for outpatient treatment of malaria in Tanzanian children: a four-arm randomised effectiveness trial. *Lancet* 365:1474–1480. [http://dx.doi.org/10.1016/S0140-6736\(05\)66417-3](http://dx.doi.org/10.1016/S0140-6736(05)66417-3).
49. Mwingira F, Nkwengulila G, Schoepflin S, Sumari D, Beck HP, Snounou G, Felger I, Olliaro P, Mugittu K. 2011. *Plasmodium falciparum* msp1, msp2 and glurp allele frequency and diversity in sub-Saharan Africa. *Malar J* 10:79. <http://dx.doi.org/10.1186/1475-2875-10-79>.
50. Oyebola MK, Idowu ET, Olukosi YA, Iwalokun BA, Agomo CO, Ajibaye OO, Tola M, Otubanjo AO. 2014. Genetic diversity and complexity of *Plasmodium falciparum* infections in Lagos, Nigeria. *Asian Pac J Trop Biomed* 4(Suppl 1):S87–S91. <http://dx.doi.org/10.12980/APJTB.4.2014C1301>.
51. Rouse P, Mkulama MA, Thuma PE, Mharakurwa S. 2008. Distinction of *Plasmodium falciparum* recrudescence and re-infection by MSP2 genotyping: a caution about unstandardized classification criteria. *Malar J* 7:185. <http://dx.doi.org/10.1186/1475-2875-7-185>.
52. Sirima SB, Tiono AB, Gansané A, Diarra A, Ouédraogo A, Konaté AT, Kiechel JR, Morgan CC, Olliaro PL, Taylor WR. 2009. The efficacy and safety of a new fixed-dose combination of amodiaquine and artesunate in young African children with acute uncomplicated *Plasmodium falciparum*. *Malar J* 8:48. <http://dx.doi.org/10.1186/1475-2875-8-48>.
53. Van Ha N, Dyk Dao L, Rabinovich SA. 2002. Use of nested PCR for differential diagnosis of falciparum malaria reinfection and relapse in drug-resistant patients. *Bull Exp Biol Med* 134:379–381. <http://dx.doi.org/10.1023/A:1021968517124>.
54. Alam MS, Elahi R, Mohon AN, Al-Amin HM, Kibria MG, Khan WA, Khanum H, Haque R. 2016. *Plasmodium falciparum* genetic diversity in Bangladesh does not suggest a hypoendemic population structure. *Am J Trop Med Hyg* 94:1245–1250. <http://dx.doi.org/10.4269/ajtmh.15-0446>.
55. Mohd Abd Razak MR, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, Jelip J, Rundt C, Imwong M, Mudin RN, Abdullah NR. 2016. Genetic diversity of *Plasmodium falciparum* populations in malaria declining areas of Sabah, East Malaysia. *PLoS One* 11:e0152415. <http://dx.doi.org/10.1371/journal.pone.0152415>.

56. Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *Proc Natl Acad Sci U S A* 90:4384–4388. <http://dx.doi.org/10.1073/pnas.90.10.4384>.
57. Arez AP, Snounou G, Pinto J, Sousa CA, Modiano D, Ribeiro H, Franco AS, Alves J, do Rosario VE. 1999. A clonal *Plasmodium falciparum* population in an isolated outbreak of malaria in the Republic of Cabo Verde. *Parasitology* 118:347–355. <http://dx.doi.org/10.1017/S0031182099003972>.
58. Batista CL, Barbosa S, Da Silva Bastos M, Viana SA, Ferreira MU. 2015. Genetic diversity of *Plasmodium vivax* over time and space: a community-based study in rural Amazonia. *Parasitology* 142:374–384. <http://dx.doi.org/10.1017/S0031182014001176>.
59. Parija SC, Mandal J, Ponnambath DK. 2014. Laboratory methods of identification of *Entamoeba histolytica* and its differentiation from look-alike *Entamoeba* spp. *Trop Parasitol* 4:90–95. <http://dx.doi.org/10.4103/2229-5070.138535>.
60. Eom KS, Rim HJ. 1993. Morphologic descriptions of *Taenia asiatica* sp. n. *Korean J Parasitol* 31:1–6. <http://dx.doi.org/10.3347/kjp.1993.31.1.1>.
61. Jeon HK, Kim KH, Eom KS. 2007. Complete sequence of the mitochondrial genome of *Taenia saginata*: comparison with *T. solium* and *T. asiatica*. *Parasitol Int* 56:243–246. <http://dx.doi.org/10.1016/j.parint.2007.04.001>.
62. Jeon HK, Lee KH, Kim KH, Hwang UW, Eom KS. 2005. Complete sequence and structure of the mitochondrial genome of the human tapeworm, *Taenia asiatica* (Platyhelminthes; Cestoda). *Parasitology* 130:717–726. <http://dx.doi.org/10.1017/S0031182004007164>.
63. Bowles J, McManus DP. 1994. Genetic characterization of the Asian *Taenia*, a newly described taeniid cestode of humans. *Am J Trop Med Hyg* 50:33–44. <http://dx.doi.org/10.4269/ajtmh.1994.50.1.TM0500010033>.
64. Ale A, Victor B, Praet N, Gabriël S, Speybroeck N, Dorny P, Devleeschauwer B. 2014. Epidemiology and genetic diversity of *Taenia asiatica*: a systematic review. *Parasit Vectors* 7:45. <http://dx.doi.org/10.1186/1756-3305-7-45>.
65. da Silva Alves EB, Conceição MJ, Leles D. 2016. *Ascaris lumbricoides*, *Ascaris suum*, or “*Ascaris lumbricoides*”? *J Infect Dis* 213:1355.
66. Leles D, Gardner SL, Reinhard K, Iñiguez A, Araujo A. 2012. Are *Ascaris lumbricoides* and *Ascaris suum* a single species? *Parasit Vectors* 5:42. <http://dx.doi.org/10.1186/1756-3305-5-42>.
67. Peng W, Criscione CD. 2012. Ascariasis in people and pigs: new inferences from DNA analysis of worm populations. *Infect Genet Evol* 12:227–235. <http://dx.doi.org/10.1016/j.meegid.2012.01.012>.
68. Erttmann KD, Meredith SE, Greene BM, Unnasch TR. 1990. Isolation and characterization of form specific DNA sequences of *O. volvulus*. *Acta Leiden* 59:253–260.
69. Erttmann KD, Unnasch TR, Greene BM, Albiez EJ, Boateng J, Denke AM, Ferraroni JJ, Karam M, Schulz-Key H, Williams PN. 1987. A DNA sequence specific for forest form *Onchocerca volvulus*. *Nature* 327:415–417. <http://dx.doi.org/10.1038/327415a0>.
70. Zimmerman PA, Dadzie KY, De Sole G, Remme J, Alley ES, Unnasch TR. 1992. *Onchocerca volvulus* DNA probe classification correlates with epidemiologic patterns of blindness. *J Infect Dis* 165:964–968. <http://dx.doi.org/10.1093/infdis/165.5.964>.
71. World Health Organization. 1995. *Onchocerciasis and Its Control*. WHO, Geneva, Switzerland.
72. Ogunrinade A, Boaky D, Merriweather A, Unnasch TR. 1999. Distribution of the blinding and nonblinding strains of *Onchocerca volvulus* in Nigeria. *J Infect Dis* 179:1577–1579. <http://dx.doi.org/10.1086/314784>.
73. Gonzalez RJ, Cruz-Ortiz N, Rizzo N, Richards J, Zea-Flores G, Domínguez A, Sauerbrey M, Catú E, Oliva O, Richards FO, Lindblade KA. 2009. Successful interruption of transmission of *Onchocerca volvulus* in the Escuintla-Guatemala focus, Guatemala. *PLoS Negl Trop Dis* 3:e404. <http://dx.doi.org/10.1371/journal.pntd.0000404>.
74. Lloyd MM, Gilbert R, Taha NT, Weil GJ, Meite A, Kouakou IM, Fischer PU. 2015. Conventional parasitology and DNA-based diagnostic methods for onchocerciasis elimination programmes. *Acta Trop* 146:114–118. <http://dx.doi.org/10.1016/j.actatropica.2015.03.019>.
75. Vlamincx J, Fischer PU, Weil GJ. 2015. Diagnostic tools for onchocerciasis elimination programs. *Trends Parasitol* 31:571–582. <http://dx.doi.org/10.1016/j.pt.2015.06.007>.
76. Bimi L, Freeman AR, Eberhard ML, Ruiz-Tiben E, Pieniazek NJ. 2005. Differentiating *Dracunculus medinensis* from *D. insignis*, by the sequence analysis of the 18S rRNA gene. *Ann Trop Med Parasitol* 99:511–517. <http://dx.doi.org/10.1179/136485905X51355>.
77. Callaway E. 2016. Dogs thwart effort to eradicate Guinea worm. *Nature* 529:10–11. <http://dx.doi.org/10.1038/529010a>.
78. Eberhard ML, Ruiz-Tiben E, Hopkins DR, Farrell C, Toe F, Weiss A, Withers PC, Jr, Jenks MH, Thiele EA, Cotton JA, Hance Z, Holroyd N, Cama VA, Tahir MA, Mounda T. 2014. The peculiar epidemiology of dracunculiasis in Chad. *Am J Trop Med Hyg* 90:61–70. <http://dx.doi.org/10.4269/ajtmh.13-0554>.
79. Macnish MG, Morgan-Ryan UM, Monis PT, Behnke JM, Thompson RC. 2002. A molecular phylogeny of nuclear and mitochondrial sequences in *Hymenolepis nana* (Cestoda) supports the existence of a cryptic species. *Parasitology* 125:567–575. <http://dx.doi.org/10.1017/S0031182002002366>.
80. Nejsum P, Parker ED, Jr, Frydenberg J, Roepstorff A, Boes J, Haque R, Astrup I, Prag J, Skov Sørensen UB. 2005. Ascariasis is a zoonosis in Denmark. *J Clin Microbiol* 43:1142–1148. <http://dx.doi.org/10.1128/JCM.43.3.1142-1148.2005>.
81. Bendall RP, Barlow M, Betson M, Stothard JR, Nejsum P. 2011. Zoonotic ascariasis, United Kingdom. *Emerg Infect Dis* 17:1964–1966. <http://dx.doi.org/10.3201/eid1710.101826>.
82. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, Rosen D, Angelino E, Sabeti PC, Wirth DF, Wiegand RC. 2008. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar J* 7:223. <http://dx.doi.org/10.1186/1475-2875-7-223>.
83. Preston MD, Campino S, Assefa SA, Echeverry DF, Ocholla H, Amambua-Ngwa A, Stewart LB, Conway DJ, Borrmann S, Michon P, Zongo I, Ouedraogo JB, Djimde AA, Doumbo OK, Nosten F, Pain A, Bousema T, Drakeley CJ, Fairhurst RM, Sutherland CJ, Roper C, Clark TG. 2014. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. *Nat Commun* 5:4052. <http://dx.doi.org/10.1038/ncomms5052>.
84. Su XZ. 2014. Tracing the geographic origins of *Plasmodium falciparum* malaria parasites. *Pathog Glob Health* 108:261–262. <http://dx.doi.org/10.1179/2047772414Z.000000000225>.
85. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, Hamilton E, Ferreira MU, Karunaweera ND, Serre D, Zimmerman PA, Sá JM, Welles TE, Musset L, Legrand E, Melnikov A, Neafsey DE, Volkman SK, Wirth DF, Sabeti PC. 2015. Development of a single nucleotide polymorphism barcode to genotype *Plasmodium vivax* infections. *PLoS Negl Trop Dis* 9:e0003539. <http://dx.doi.org/10.1371/journal.pntd.0003539>.
86. Culleton R, Coban C, Zeyrek FY, Cravo P, Kaneko A, Randrianarivelojosa M, Andrianaranjaka V, Kano S, Farnert A, Arez AP, Sharp PM, Carter R, Tanabe K. 2011. The origins of African *Plasmodium vivax*: insights from mitochondrial genome sequencing. *PLoS One* 6:e29137. <http://dx.doi.org/10.1371/journal.pone.0029137>.
87. Gelabert P, Sandoval-Velasco M, Olalde I, Fregel R, Rieux A, Escosa R, Aranda C, Paaijmans K, Mueller I, Gilbert MT, Lalueza-Fox C. 2016. Mitochondrial DNA from the eradicated European *Plasmodium vivax* and *P. falciparum* from 70-year-old slides from the Ebro Delta in Spain. *Proc Natl Acad Sci U S A* 113:11495–11500. <http://dx.doi.org/10.1073/pnas.1611017113>.
88. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, Vallejo AF, Herrera S, Arévalo-Herrera M, Fan Q, Wang Y, Cui L, Lucas CM, Durand S, Sanchez JF, Baldeviano GC, Lescano AG, Laman M,

- Barnadas C, Barry A, Mueller I, Kazura JW, Eapen A, Kanagaraj D, Valecha N, Ferreira MU, Roobsoong W, Nguitragool W, Sattabonkot J, Gamboa D, Kosek M, Vinetz JM, González-Cerón L, Birren BW, Neafsey DE, Carlton JM. 2016. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat Genet* 48:953–958. <http://dx.doi.org/10.1038/ng.3588>.
89. Iwagami M, Hwang SY, Fukumoto M, Hayakawa T, Tanabe K, Kim SH, Kho WG, Kano S. 2010. Geographical origin of *Plasmodium vivax* in the Republic of Korea: haplotype network analysis based on the parasite's mitochondrial genome. *Malar J* 9:184. <http://dx.doi.org/10.1186/1475-2875-9-184>.
90. Rodrigues PT, Alves JM, Santamaria AM, Calzada JE, Xayavong M, Parise M, da Silva AJ, Ferreira MU. 2014. Using mitochondrial genome sequences to track the origin of imported *Plasmodium vivax* infections diagnosed in the United States. *Am J Trop Med Hyg* 90:1102–1108. <http://dx.doi.org/10.4269/ajtmh.13-0588>.
91. Ferreira MU, Rodrigues PT. 2014. Tracking malaria parasites in the eradication era. *Trends Parasitol* 30:465–466. <http://dx.doi.org/10.1016/j.pt.2014.08.003>.
92. Spanakos G, Alifrangis M, Schousboe ML, Patsoula E, Tegos N, Hansson HH, Bygbjerg IC, Vakalis NC, Tseroni M, Kremastinou J, Hadjichristodoulou C. 2013. Genotyping *Plasmodium vivax* isolates from the 2011 outbreak in Greece. *Malar J* 12:463. <http://dx.doi.org/10.1186/1475-2875-12-463>.
93. Gunawardena S, Ferreira MU, Kapilananda GM, Wirth DF, Karunaweera ND. 2014. The Sri Lankan paradox: high genetic diversity in *Plasmodium vivax* populations despite decreasing levels of malaria transmission. *Parasitology* 141:880–890. <http://dx.doi.org/10.1017/S0031182013002278>.
94. Malaria GEN, MalariaGEN *Plasmodium falciparum* Community Project. 2016. Genomic epidemiology of artemisinin resistant malaria. *Life* 5:5.
95. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amarantunga C, Lim P, Mead D, Oyola SO, Dhorda M, Imwong M, Woodrow C, Manske M, Stalker J, Drury E, Campino S, Amenga-Etego L, Thanh TN, Tran HT, Ringwald P, Bethell D, Nosten F, Phyto AP, Pukrittayakamee S, Chotivanich K, Chuor CM, Nguon C, Suon S, Sreng S, Newton PN, Mayxay M, Khanhavong M, Hongvanthong B, Htut Y, Han KT, Kyaw MP, Faiz MA, Fanello CI, Onyamboko M, Mokuolu OA, Jacob CG, Takala-Harrison S, Plowe CV, Day NP, Dondorp AM, Spencer CC, McVean G, Fairhurst RM, White NJ, Kwiatkowski DP. 2015. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 47:226–234. <http://dx.doi.org/10.1038/ng.3189>.
96. Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su XZ. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 418:320–323. <http://dx.doi.org/10.1038/nature00813>.
97. Wongsrichanalai C, Pickard AL, Wernsdorfer WH, Meshnick SR. 2002. Epidemiology of drug-resistant malaria. *Lancet Infect Dis* 2:209–218. [http://dx.doi.org/10.1016/S1473-3099\(02\)00239-6](http://dx.doi.org/10.1016/S1473-3099(02)00239-6).
98. World Health Organization. 2009. *Methods for Surveillance of Antimalarial Drug Efficacy*. WHO, Geneva, Switzerland.
99. Brown T, Smith LS, Oo EK, Shawng K, Lee TJ, Sullivan D, Beyrer C, Richards AK. 2012. Molecular surveillance for drug-resistant *Plasmodium falciparum* in clinical and subclinical populations from three border regions of Burma/Myanmar: cross-sectional data and a systematic review of resistance studies. *Malar J* 11:333. <http://dx.doi.org/10.1186/1475-2875-11-333>.
100. Koukouikila-Koussounda F, Jeyaraj S, Nguetse CN, Nkonganyi CN, Kokou KC, Etoka-Beka MK, Ntouni F, Velavan TP. 2017. Molecular surveillance of *Plasmodium falciparum* drug resistance in the Republic of Congo: four and nine years after the introduction of artemisinin-based combination therapy. *Malar J* 16:155. <http://dx.doi.org/10.1186/s12936-017-1816-x>.
101. Lucchi NW, Komino F, Okoth SA, Goldman I, Onyona P, Wiegand RE, Juma E, Shi YP, Barnwell JW, Udhayakumar V, Kariuki S. 2015. In vitro and molecular surveillance for antimalarial drug resistance in *Plasmodium falciparum* parasites in Western Kenya reveals sustained artemisinin sensitivity and increased chloroquine sensitivity. *Antimicrob Agents Chemother* 59:7540–7547. <http://dx.doi.org/10.1128/AAC.01894-15>.
102. Menegon M, Nurahmed AM, Talha AA, Nour BY, Severini C. 2016. Molecular surveillance of antimalarial drug resistance related genes in *Plasmodium falciparum* isolates from Eritrea. *Acta Trop* 157:158–161. <http://dx.doi.org/10.1016/j.actatropica.2016.02.007>.
103. Norahmad NA, Mohd Abd Razak MR, Abdullah NR, Sastu UR, Imwong M, Muniandy PK, Saat MN, Muhammad A, Jelip J, Tikuson M, Yusof N, Rundi C, Mudin RN, Syed Mohamed AF. 2016. Prevalence of *Plasmodium falciparum* molecular markers of antimalarial drug resistance in a residual malaria focus area in Sabah, Malaysia. *PLoS One* 11:e0165515. <http://dx.doi.org/10.1371/journal.pone.0165515>.
104. Kateera F, Nsohya SL, Tukwasibwe S, Hakizimana E, Mutesa L, Mens PF, Grobusch MP, van Vugt M, Kumar N. 2016. Molecular surveillance of *Plasmodium falciparum* drug resistance markers reveals partial recovery of chloroquine susceptibility but sustained sulfadoxine-pyrimethamine resistance at two sites of different malaria transmission intensities in Rwanda. *Acta Trop* 164:329–336. <http://dx.doi.org/10.1016/j.actatropica.2016.09.008>.
105. Nyunt MH, Wang B, Aye KM, Aye KH, Han JH, Lee SK, Han KT, Htut Y, Han ET. 2017. Molecular surveillance of artemisinin resistance *falciparum* malaria among migrant goldmine workers in Myanmar. *Malar J* 16:97. <http://dx.doi.org/10.1186/s12936-017-1753-8>.
106. Mint Lekweiry K, Ould Mohamed Salem Boukhary A, Gaillard T, Wurtz N, Bogreau H, Hafid JE, Trape JF, Bouchiba H, Ould Ahmedou Salem MS, Pradines B, Rogier C, Basco LK, Briolant S. 2012. Molecular surveillance of drug-resistant *Plasmodium vivax* using pvdhfr, pvdhps and pvmr1 markers in Nouakchott, Mauritania. *J Antimicrob Chemother* 67:367–374. <http://dx.doi.org/10.1093/jac/dkr464>.
107. Taylor SM, Antonia AL, Parobek CM, Juliano JJ, Janko M, Emch M, Alam MT, Udhayakumar V, Tshefu AK, Meshnick SR. 2013. *Plasmodium falciparum* sulfadoxine resistance is geographically and genetically clustered within the DR Congo. *Sci Rep* 3:1165. <http://dx.doi.org/10.1038/srep01165>.
108. Nsimba B, Jafari-Guemouri S, Malonga DA, Mouata AM, Kiori J, Louya F, Yocka D, Malanda M, Durand R, Le Bras J. 2005. Epidemiology of drug-resistant malaria in Republic of Congo: using molecular evidence for monitoring antimalarial drug resistance combined with assessment of antimalarial drug use. *Trop Med Int Health* 10:1030–1037. <http://dx.doi.org/10.1111/j.1365-3156.2005.01490.x>.
109. Abdel-Muhsin AA, Mackinnon MJ, Awadalla P, Ali E, Suleiman S, Ahmed S, Walliker D, Babiker HA. 2003. Local differentiation in *Plasmodium falciparum* drug resistance genes in Sudan. *Parasitology* 126:391–400. <http://dx.doi.org/10.1017/S0031182003003020>.
110. Lanzaro G, Tripet F. 2003. Gene flow among populations of *Anopheles gambiae*: a critical review, p 109–132. In Takken W, Scott TW (ed), *Ecological Aspects for the Application of Genetically Modified Mosquitoes*. Frontis Press, Wageningen, the Netherlands.
111. Anderson TJ, Nair S, McDew-White M, Cheeseman IH, Nkhoma S, Bilgic F, McGready R, Ashley E, Pyae Phyto A, White NJ, Nosten F. 2017. Population parameters underlying an ongoing soft sweep in Southeast Asian malaria parasites. *Mol Biol Evol* 34:131–144. <http://dx.doi.org/10.1093/molbev/msw228>.
112. Valentim CL, Cioli D, Chevalier FD, Cao X, Taylor AB, Holloway SP, Pica-Mattoccia L, Guidi A, Basso A, Tsai JJ, Berriman M, Carvalho-Queiroz C, Almeida M, Aguilar H, Frantz DE, Hart PJ, LoVerde PT, Anderson TJ. 2013. Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. *Science* 342:1385–1389. <http://dx.doi.org/10.1126/science.1243106>.

113. Diawara A, Halpenny CM, Churcher TS, Mwandawiro C, Kihara J, Kaplan RM, Streit TG, Idaghdour Y, Scott ME, Basáñez MG, Prichard RK. 2013. Association between response to albendazole treatment and β -tubulin genotype frequencies in soil-transmitted helminths. *PLoS Negl Trop Dis* 7:e2247. <http://dx.doi.org/10.1371/journal.pntd.0002247>.
114. Rebaudet S, Bogreau H, Silāi R, Lepere JF, Bertaux L, Pradines B, Delmont J, Gautret P, Parola P, Rogier C. 2010. Genetic structure of *Plasmodium falciparum* and elimination of malaria, Comoros archipelago. *Emerg Infect Dis* 16:1686–1694. <http://dx.doi.org/10.3201/eid1611.100694>.
115. Mulenge FM, Hunja CW, Magiri E, Culleton R, Kaneko A, Aman RA. 2016. Genetic diversity and population structure of *Plasmodium falciparum* in Lake Victoria Islands, a region of intense transmission. *Am J Trop Med Hyg* 95:1077–1085. <http://dx.doi.org/10.4269/ajtmh.16-0383>.
116. Mobegi VA, Loua KM, Ahouidi AD, Satoguina J, Nwakanma DC, Amambua-Ngwa A, Conway DJ. 2012. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar J* 11:223. <http://dx.doi.org/10.1186/1475-2875-11-223>.
117. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, Shamsul S, Singh B. 2005. Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *J Infect Dis* 191:1558–1564. <http://dx.doi.org/10.1086/429338>.
118. Lamberton PH, Cheke RA, Winskill P, Tirados I, Walker M, Osei-Atweneboana MY, Biritwum NK, Tetteh-Kumah A, Boakye DA, Wilson MD, Post RJ, Basáñez MG. 2015. Onchocerciasis transmission in Ghana: persistence under different control strategies and the role of the simuliid vectors. *PLoS Negl Trop Dis* 9:e0003688. <http://dx.doi.org/10.1371/journal.pntd.0003688>.
119. Fischer PU, King CL, Jacobson JA, Weil GJ. 2017. Potential value of triple drug therapy with ivermectin, diethylcarbamazine, and albendazole (IDA) to accelerate elimination of lymphatic filariasis and onchocerciasis in Africa. *PLoS Negl Trop Dis* 11:e0005163. <http://dx.doi.org/10.1371/journal.pntd.0005163>.
120. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang HH, Wong W, Baro N, Ndiaye D, Fall FB, Ndiop M, Ba M, Milner DA, Jr, Taylor TE, Neafsey DE, Volkman SK, Eckhoff PA, Hartl DL, Wirth DF. 2015. Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci U S A* 112:7067–7072. <http://dx.doi.org/10.1073/pnas.1505691112>.
121. Ingasia LA, Cheruiyot J, Okoth SA, Andagalu B, Kamau E. 2016. Genetic variability and population structure of *Plasmodium falciparum* parasite populations from different malaria ecological regions of Kenya. *Infect Genet Evol* 39:372–380. <http://dx.doi.org/10.1016/j.meegid.2015.10.013>.
122. Bonizzoni M, Afrane Y, Baliraine FN, Amenya DA, Githeko AK, Yan G. 2009. Genetic structure of *Plasmodium falciparum* populations between lowland and highland sites and antimalarial drug resistance in Western Kenya. *Infect Genet Evol* 9:806–812. <http://dx.doi.org/10.1016/j.meegid.2009.04.015>.
123. Gatei W, Gimnig JE, Hawley W, Ter Kuile F, Otero C, Iriemenam NC, Shah MP, Howard PP, Omosun YO, Terlouw DJ, Nahlen B, Slutsker L, Hamel MJ, Kariuki S, Walker E, Shi YP. 2015. Genetic diversity of *Plasmodium falciparum* parasite by microsatellite markers after scale-up of insecticide-treated bed nets in western Kenya. *Malar J* 13(Suppl 1):495. <http://dx.doi.org/10.1186/s12936-015-1003-x>.
124. Carter TE, Malloy H, Existe A, Memnon G, St Victor Y, Okech BA, Mulligan CJ. 2015. Genetic diversity of *Plasmodium falciparum* in Haiti: insights from microsatellite markers. *PLoS One* 10:e0140416. <http://dx.doi.org/10.1371/journal.pone.0140416>.
125. Sisya TJ, Kamn'gona RM, Varetta JA, Fulakeza JM, Mukaka MF, Seydel KB, Laufer MK, Taylor TE, Nkhoma SC. 2015. Subtle changes in *Plasmodium falciparum* infection complexity following enhanced intervention in Malawi. *Acta Trop* 142:108–114. <http://dx.doi.org/10.1016/j.actatropica.2014.11.008>.
126. Nkhoma SC, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, Anderson TJ. 2012. Close kinship within multiple-genotype malaria parasite infections. *Proc Biol Sci* 279:2589–2598. <http://dx.doi.org/10.1098/rspb.2012.0113>.
127. Takezaki N, Nei M, Tamura K. 2014. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* 27:747–752.
128. Excoffier L, Heckel G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 7:745–758.
129. Rudge JW, Carabin H, Balolong E, Tallo V, Shrivastava J, Lu DB, Basanez MG, Olveda R, McGarvey ST, Webster JP. 2008. Population genetics of *Schistosoma japonicum* within the Philippines suggest high levels of transmission between humans and dogs. *PLoS Negl Trop Dis* 2:e340.