

## Glossary

Note that the terminology used in this paper sometimes differs from common usage.

**Allele.** An instance of the parasite genome, e.g.  $n$  haploid individuals correspond to  $n$  alleles.

**Coalescence.** If two lineages are traced back in time, coalescence occurs when they meet in the same ancestral allele.

**Cotransmission.** Occurs when a host or a vector is carrying a mixture of parasite alleles with different ancestral histories, subsequent to an episode of superinfection.

**Crossing rate of transmission chains ( $\chi$ ).** The proportion of hosts that acquire parasites from more than one transmission chain, i.e. from more than one host in the previous generation. This is equivalent to proportion of hosts that are superinfected.

**Effective number of hosts ( $N_H$ ).** The number of hosts that effectively transmit parasites in each generation of the transmission graph. This is a form of population bottleneck.

**Effective recombination.** Recombination between genetically distinct alleles that acts to change the DNA sequence of a haplotype locus.

**Effective recombination parameter  $\phi_t$ .** The probability that, if recombination occurs at a locus at time  $t$ , this will change the DNA sequence of the locus.

**Haplotype locus.** A locus that extends over multiple nucleotide positions. A haplotype is a specific DNA sequence observed at a haplotype locus.

**Heterozygosity ( $H$ ).** The probability that two alleles sampled randomly from some population are heterozygous, i.e. that they have different DNA sequences.

**Homozygosity ( $G$ ).** The probability that two alleles sampled randomly from some population are homozygous, i.e. that they have the same DNA sequence.

**Lineage.** A path that traces the ancestry of an allele at a point locus, going backwards in time through the transmission graph. A point locus is not affected by recombination, so a lineage can be traced back over many generations despite frequent recombination events.

**Locus.** A specific location in the genome. This can be either a single nucleotide position (a point locus) or a sequence extending over multiple nucleotide positions (a haplotype locus).

**Nucleotide diversity ( $\pi$ ).** The probability that two alleles are heterozygous at a random nucleotide position in the genome.

**Point locus.** A specific single nucleotide position in the genome.

**Quantum of transmission ( $Q$ ).** The number of parasite alleles transmitted from one host to the next via a vector.  $Q$  summarises a complex series of bottlenecks in host-vector and vector-host transmission occurring during one generation of the parasite life-cycle.

**Serial interval of transmission ( $\tau$ ).** The length of time corresponding to one generation of the transmission graph, i.e the mean interval between parasites entering one host and the next on a transmission chain.

**Superinfection.** Infection of a host with parasites from more than one source host in the previous generation. This is equivalent to crossing of transmission chains.

**Transmission bottleneck.** A population bottleneck that affects the number of alleles passed from one generation of the genomic transmission graph to the next. The quantum of transmission  $Q$  and the effective number of hosts  $N_H$  can both be considered as transmission bottlenecks.

**Transmission chain.** A sequence of host-to-host transmission events, defined by selecting a node in the transmission graph and tracing a path forwards in time along the edges to another node.