Host-parasite coevolution and its genomic signature

Dieter Ebert[1,2,*] *& Peter D. Fields*[1]

[1] University of Basel, Department of Environmental Sciences, Zoology, Vesalgasse 1, 4051 Basel, Switzerland

and

[2] Wissenschaftskolleg zu Berlin, Wallotstrasse 19, 14193 Berlin, Germany

[*]e-mail: dieter.ebert@unibas.ch

ORCID IDs:
0000-0003-2653-3772 (D.E.)
0000-0003-2959-2524 (P.D.F.)

Manuscript accepted for publication in:
Nature Review Genetics, Vol. 21, December 2020
https://doi.org/10.1038/ s41576-020-0269-1

Abstract | Studies in diverse biological systems have indicated that host–parasite coevolution is responsible for the extraordinary genetic diversity seen in some genomic regions, such as major histocompatibility (MHC) genes in jawed vertebrates and resistance genes in plants. This diversity is believed to evolve under balancing selection on hosts by parasites. However, the mechanisms that link the genomic signatures in these regions to the underlying coevolutionary process are only slowly emerging. We still lack a clear picture of the coevolutionary concepts and of the genetic basis of the coevolving phenotypic traits in the interacting antagonists. Emerging genomic tools that provide new options for identifying underlying genes will contribute to a fuller understanding of the coevolutionary process.

Introduction

Host–parasite coevolution occurs when selection by parasites [G] triggers diverse host adaptations that reduce the costs of infection, which in turn prompts parasites to adapt anew to their hosts. This process may be among the most important generators of biological diversity over the past 3.5 billion years [1,2], including the generation and maintenance of genetic diversity within populations and species, and the sharing of certain variants across species boundaries [3-9]. As such, the genomes of organisms would be expected to show signatures of this host–parasite coevolution. Recognizing and characterizing these genomic signatures [G] would be expected to lead to a better understanding of how diversity has evolved, and is still evolving, across the tree of life. Furthermore, identifying genomic signatures of coevolution can help to narrow down the loci under selection and provide mechanistic insights into host–parasite interactions.

Antagonistic coevolution has been defined as reciprocal selection between two closely interacting species [10,11]. This definition focuses on the phenotypic traits of the coevolving antagonists that negatively influence each other. It specifies that it is the traits responsible for the interaction and their underlying genes that coevolve—not the species [12,13]. It also sets antagonistic coevolution apart from scenarios of host–parasite co-association and the resulting long-term patterns of co-speciation or co-cladogenesis, which do not require adaptive processes [14].

The process of antagonistic coevolution can be described by two types of model: models of specific coevolution, in which one host and one parasite species interact (sometimes called pairwise coevolution); and models of unspecific coevolution, better known as diffuse coevolution [10], in which multiple hosts and/or parasite species contribute to the process. Models of specific coevolution can be further grouped into selective sweep [G] models, in which novel variants are selected for and rise to high frequencies, and balancing selection models, in which alternative variants at specific loci fluctuate in frequency over time [11,15]. The different modes of action underlying these models create different signatures in the genomes of the antagonists. Thus, observation of a particular signature in a genome allows (within limits) conclusions to be drawn about the evolutionary mechanism producing them (Table 1).

A genomic perspective of coevolution considers the entire region around the coevolving genes, because making use of the additional information present in the flanking sites strongly increases the power of the analysis to detect genomic signatures. Despite being functionally independent, sites physically tightly-linked to the coevolving genes are influenced by linked selection [G], which causes genetic hitchhiking [G], and, unlike unlinked sites, their fate is determined by the dynamics of the selected genes and the rate of recombination among them [16,17]. Thus, genomic signatures of coevolution can be detected in populations by comparing patterns of

genetic variation in these regions with the patterns in the genomic background that presumably evolved by neutral evolutionary processes. The genomic signatures of models of specific coevolution have received much attention, resulting in a good picture of the expected patterns in population samples of genomes. By contrast, the genomic signatures expected under diffuse coevolution have not yet been determined; nevertheless, some heuristics exist.

In this Review, we address conceptual issues concerning host–parasite coevolution and how they manifest in the genomes of the antagonists. We highlight the differences between specific and diffuse coevolution and point out the role of spatial population structure in shaping genomic signatures. Using these concepts, we discuss the evolution of trans-species polymorphisms. Finally, we address new cogenomic [G] methods that allow coevolving loci in hosts and parasites to be directly pinpointed and what we can learn from these recent developments.

Models of specific host–parasite coevolution

The two models of specific coevolution, selective sweep selection and balancing selection, have different effects on genetic variation. Selective sweep coevolution is arguably one of the most important generators of macroevolutionary patterns, and explains differences in immune systems among lineages, drives speciation and is implicated in some of the major transitions in evolution [2,18-20]. However, it fails to explain the extraordinary genetic diversity observed at some host resistance loci. By contrast, coevolution by balancing selection [G] is best known for its potential to maintain high levels of genetic diversity within populations and species. The theory states that hosts and parasites undergo continuous antagonistic coevolution, often referred to as Red Queen dynamics or trench warfare, which results in a balance of different variants at loci related to host defence and parasite offence [21-23]. Recent studies have revealed that there are more genomic regions in plants and animals that undergo balancing selection than previously thought [8,24-29].

The study of coevolution by analysing patterns of genetic variation has a long tradition, but has a number of limitations that make it hard to reach strong conclusions about underlying evolutionary processes. For example, genomic studies on coevolution are mostly performed separately in host and parasite genomes (but see ‘Joint analysis of host and parasite genomes’). Thus, a given finding may be difficult to attribute to host–parasite coevolution or to other processes, because the bioinformatic and population genetic approaches that are used to identify genomic regions with characteristics of selective sweeps or balancing selection do not identify the cause of selection — antagonistic interactions are only one of many possibilities. Additional efforts are needed to differentiate between these options. This issue is particularly pronounced for balancing selection, as different mechanisms (such as overdominance [G], local adaptation [G], direct negative frequency dependent

selection [G] (NFDS) and indirect negative frequency dependent selection [G]) can all produce genomic signatures of balancing selection. However, only indirect NFDS is associated with antagonistic coevolution, and so distinguishing among these mechanisms is crucial.

[H2] Selective sweep coevolution. A selective sweep describes the rapid increase in frequency of a beneficial variant, such as a novel mutation. In the extreme case, the variant will reach fixation, replacing alternative variants. Adaptive evolution in this form is considered to be a dominant driver of protein adaptation and species divergence [30,31]. During host–parasite coevolution, both antagonists may experience sweeps at loci that play a functional role in their interaction (FIG. 1) [11,15,32-35]. For coevolution to occur, sweeps do not need to alternate between hosts and parasites. Variants at multiple sites may even spread at the same time, as long as they are decoupled by recombination, as otherwise selective interference [G] takes place [36 33,34].

A selective sweep leaves a characteristic valley of locally reduced genetic variation in population samples of genomes. These diversity valleys are formed because variants that are in proximity to the sweeping variant hitchhike along and replace alternative variants, which results in a corresponding pattern of increased linkage disequilibrium [G] [37]. Such valleys allow us to detect the approximate position of the beneficial mutant, but the genomic signatures are rarely clear and lose their distinct features over time as recombination breaks down associations and as new mutations arise. Thus, ancient selective sweeps are no longer detectable, although they leave their traces in signatures of positive selection that result in patterns of divergence among lineages [20]. Likewise, sweeping mutants in the initial phase of a sweep are hard to detect, as the reduction in local genetic diversity is not yet pronounced (FIG. 2). Additionally, each sweep event has a different history: each one is associated with a different selection coefficient [G] and may occur at a different location in the genome and possibly experience different local recombination rates [38]. Finally, signatures of sweeps strongly depend on the initial variation on which selection acted [37].

Variants spreading to fixation in coevolving hosts and parasites may be functionally independent, other than that they each provide a benefit to their carrier and are disadvantageous to the antagonist [23]. However, when particular host or parasite proteins interact, selective sweep coevolution can occur through alternating changes to functionally important segments of these proteins. For example, the restriction factor *TRIM5*, an antiviral protein in Old World primates seems to have coevolved in tight interaction with capsid proteins in lentiviruses [39].

Due to their high rate of evolution, parasites are among the fastest changing selecting agents host organisms experience. Therefore, the speed with which a host

population can respond to a new variant of a parasite is important. However, selective sweep evolution from a novel mutation to fixation (that is, a hard sweep) is a rather slow process. In particular, the initial increase in allele frequencies is slow: in a large population, a novel mutation with a 10 % fitness advantage takes about 200 generations to reach a frequency of 10 %, but only about 15 generations more to reach 50 % [40,41]. Thus, sweeps starting from standing genetic variation (known as soft sweeps) can occur faster than hard sweeps, as the initial allele frequencies are higher. Indeed, evidence is accumulating that many sweeps observed in natural populations are soft sweeps [41-43]. Interestingly, standing genetic variation is often high for loci related to host–parasite interactions and thus can drive faster sweeps. Although it is not fully known why this is the case, the two-speed genome model [44-46] attempts to explain this paradox by suggesting that specific regions in host and parasite genomes have strongly elevated mutation rates, constantly re-fuelling standing genetic variation, whereas housekeeping genes have lower mutation rates and evolve by purifying selection.

Coevolution by balancing selection. The Red Queen model of balancing selection was introduced by Clarke [47] to explain unexpectedly high genetic diversity within populations. It is based on the assumption of strong host–genotype by parasite–genotype interactions [48-50] (FIG. 3a), which has been finding increasing empirical support [51-53]. As opposed to the selective sweep model, in which mutant alleles have a general advantage over the wild type alleles, genetic variants under balancing selection provide an advantage only in specific relation to a corresponding genetic variant in the antagonist. These host and parasite genes are functionally coupled, forming the driving force for reciprocal selection [22,54-56]. According to the Red Queen model, a parasite allele will increase in frequency when the host allele that allows it to infect is common; as hosts carrying this allele succumb to increased parasitism, their numbers will decline, and—with a time lag—so too do parasite genotypes that depend on these particular host genotypes, that is, the loci in the host and the parasite are under NFDS. As parasites with particular infectivity alleles track corresponding host alleles, cyclical dynamics of host and parasite alleles might arise, with parasite cycles lagging behind the corresponding host allele cycles [57,58]. NFDS reduces the likelihood that alleles go extinct by chance because rare alleles gain an advantage; thus, genetic polymorphisms in hosts and parasites can be maintained, and the frequencies of the functional variants at the involved loci are balanced over time (FIG. 3b). In this form of balancing selection, an allele's selection coefficient does not depend directly on its own frequency, but rather depends on the frequencies of specific alleles in the antagonist. Therefore, it is called indirect NFDS [21].

Relative to the rest of the genome, genomic regions containing genes under balancing selection are expected to show high genetic diversity and a more even

frequency spectrum of variants, resulting in positive Tajima's D [G] values [16,26,59,60] (Table 2). Scans of host genomes have revealed ample evidence of regions under balancing selection, with the most well-known being the MHC class 1 and 2 genes of jawed vertebrates and the AB0 blood group system of higher primates, but also other regions of vertebrate genomes related to immune function [25-28,61-64]. Scans in plants and invertebrates have also revealed balancing selection in diverse genomic regions, again often regions associated with immune function [8,65-71]. However, current bioinformatic and population genetic methods may still miss many regions of interest. To produce a recognizable genomic signature, selection should be strong and should have acted for a sufficiently long time period (Fig. 2). Regions around sites under selection usually show strong local linkage disequilibrium, thus local recombination rate will also affect the ability to detect signatures. Ideally, the recombination rate around the selected loci should be low to allow linked SNPs to hitchhike; high recombination rates would reduce the region to the very site under balancing selection itself [16,21]. However, some level of recombination is helpful for the detection of the sites under selection, as otherwise the entire region will be inherited as one linkage block.

Less attention has been given to balancing selection in parasites than in hosts. Mapping parasite loci is more difficult because parasites often cannot be cultured and phenotyped outside the host. Furthermore, they often have no or infrequent genetic recombination, have extreme population structures and demography, carry extra chromosomal genetic elements (such as plasmids), and have substantial divergence in gene content as a result of horizontal gene transfer: each of these features may produce genomic signatures that make it more difficult to identify the signature of balancing selection [72,73]. New bioinformatic tools for genome analysis are now being developed that address some of these challenges (see Table 3 in ref [73]), and theoretical analyses suggest that balancing selection might show stronger genomic signatures in parasites than hosts [21]. Examples of loci likely under balancing selection have been described for a range of parasites [74-79]. For West Nile Virus in mosquitos and Deformed Wing Virus in honey bees it has been suggested that genetic diversity is maintained by selection from RNA interference (RNAi) in the host [78,80,81]. Interestingly, the genes for RNAi in *Drosophila* were suggested to evolve by very high rates of positive selection with evidence for recent selective sweeps [82]. Two closely-related human pathogenic bacteria, *Staphylococcus aureus* and *S. epidermidis* [74,75], show evidence for balanced polymorphisms, but for different genes. The hrpA gene of the plant pathogen *Pseudomonas syringae*, which encodes part of the type III secretion system, was found to be under balancing selection [76]. In addition, a rare example of a balanced polymorphism involving structural variation, which determines the presence or absence of alternative pathogenicity islands, was described in *Pseudomonas viridiflava*, a plant pathogen often found on *Arabidopsis* [83]. Interestingly, several of

these bacteria are known for their wide host range, hinting that diffuse coevolution may contribute to the observed patterns (see 'Diffuse coevolution' section). Sites with high genetic diversity and polymorphisms have also been described in human malaria parasites [84,85], although the signature of balancing selection here may be confounded with processes resulting from acquired immunity. Finally, in the planktonic crustacean *Daphnia magna* and its highly specific bacterial parasite *Pasteuria ramosa*, evidence for balancing selection comes from both the infectivity loci in the bacteria [79] and from the resistance loci in the host [53].

The genomic signatures of coevolution in hosts and parasites cannot be expected to resemble each other, as the two antagonists experience selection in very different ways. For parasites, the fitness difference between infecting and failing to infect a host is very large, whereas the lifetime fitness difference between hosts that resist a parasite attack or not is likely much smaller [86]. In addition, some hosts may not encounter the parasite during their lifetime, reducing selection for resistance even further. Hosts and parasites also may have different generation times, recombination rates, effective population sizes, ploidy-levels, mutation rates, population structures and demography, to name just a few dissimilarities. Therefore, the footprint of coevolution and the genomic signatures of the two coevolving antagonists may look very different and may sometimes only be detectable in one of the antagonists [21].

Initially, population genetic modelling of coevolution by NFDS was performed for infinite population sizes, which strongly reduces the chance of extinction of host and parasite variants. Real populations, however, are of finite size and subject to stochastic effects, so functional variants may be lost due to genetic drift [G], bringing–in the most extreme case–coevolution to a halt [21,87-89]. These effects can substantially blur the genomic signatures of coevolution in natural populations, making their interpretation more difficult [21]. However, spatial structures of host populations can strongly reduce these stochastic events, as long as sub-populations are not evolving in synchrony [90-92].

Specific coevolution and spatial structure Simple models of host–parasite coevolution generally focus on evolutionary dynamics in a single, large, panmictic [G] population. However, populations are almost always spatially distributed, with an extended geographic structure and gene flow among populations. This spatial structuring profoundly influences coevolution and divergence [58,90,91,93-97] and accounting for it can strongly refine population genomic analyses [33,73,98].

Research on subdivided populations has examined how coevolution influences the spread of variants related to host–parasite interactions, thereby contributing to patterns of genetic variation on a species-level scale (rather than a population-level scale) [96,99,100]. In a selective sweep scenario, strong gene flow allows globally beneficial mutations to spread quickly from population to population, whereas weak gene flow

leads to population divergence [37,101]. More complex patterns of divergence may arise, for example when gene flow is rare and unbalanced among subpopulations, when it differs for the two antagonists, and when populations adapt locally to other environmental factors [102]. Genomic patterns of such processes are hard to study, especially if the evolving genes are not known beforehand [37,101].

By contrast, under balancing selection, gene flow enables alleles to persist in the overall gene pool much longer, as it can bring locally extinct alleles back into the population (FIG. 3b) [60,93,100,103,104]. In addition, as immigrating alleles are likely to be rare upon their arrival, they are expected to have an advantage and experience a lower rate of extinction and a higher likelihood of spreading locally [104,105], thereby increasing their effective migration rate compared to neutral genetic variants. The number of alleles maintained on a species level thus rises accordingly, and the loci under NFDS are expected to show less differentiation among populations than neutral loci that do not benefit from NFDS [79,106]. This difference is seen when comparing isolation-by-distance (IBD) patterns for these two groups of loci: whereas neutral loci show increasing differentiation (visible, for example, as an increased fixation index [G] (F_{ST})) with increasing distance among populations, loci involved in NFDS show no or a much weaker pattern of IBD [107], which is the opposite of what local positive selection would produce [104,108,109]. However, while the immigrant advantage reduces differentiation at the loci under NFDS among populations over the long term, it may drive the divergence of neighbouring populations in the short term, for example if strong selection occurs at a given gene but for different variants. Such patterns have been observed in some population genetic studies of resistance genes [100,106,110]. Thus, populations may show more variable pairwise F_{ST} at sites under NFDS than at neutral sites.

Although limited gene flow causes populations to show patterns of IBD [111,112], the perspective described above does not include common ecological and biological circumstances that may influence gene flow, such as spatially divergent selection, population size and genetic drift, metapopulation and source-sink population structure, and historic events. Parasites may differ in local abundance and may not be present in every host population. Thus, the propensity for host–parasite coevolution may vary in space and time. Host genes involved in host–parasite interactions may become neutral in the absence of the parasite, or even detrimental if there are costs of resistance. The combined dynamics of these and other evolutionary and ecological processes that influence spatial and temporal variation have been described as the geographic mosaic of coevolution, with hot spots showing strong and rapid coevolutionary dynamics, and cold-spots marked by slow or no coevolution [113,114]. For any given system, the greater the proportion of coevolutionary cold spots, the weaker the overall signature of balancing selection. However, even at cold spots, one may find elevated genetic diversity at loci that are under balancing selection in hot spots,

because gene flow from hot spots may prevent extinction of variants in cold spots. Species-wide, the overall genomic signature of balancing selection may, therefore, be relatively insensitive to the geographic mosaic of coevolution. However, this will certainly depend on the interplay of migration, genetic drift and selection.

Trans-species-polymorphism

Because alleles under balancing selection are less likely to be lost from a gene-pool, they segregate for longer time periods than genes undergoing neutral evolution or directional selection [G] [16]. This theory can be tested using coalescence approaches that compare regions under balancing selection with the genomic background [38]. In extreme cases, ancient identical-by-descent genetic variants at polymorphic sites may have pre-existed even before the last speciation event, so that these variants are shared across closely related species (FIG. 4a). This trans-species-polymorphism (TSP) is indicated when haplotypes from the two species cluster by allele and not by species (FIG. 4b) [16,115], as expected for neutrally evolving alleles. In the context of host-parasite interactions, the signal of TSP is very clear in MHC genes, in the AB0 blood group system of higher primates and in several other vertebrate loci [25,29,64,115-117]. It has also been observed in plants [8,118,119], but rarely in invertebrates [68]. Genomic regions with evidence of TSP are typically enriched with immune function genes [6,8,26,97,105,119,120], suggesting that host–parasite coevolution might be the driving force behind a genomic signature of TSP [20,120,121]. Experimental data further support a link between TSP and parasitism [105,122].

Given the large time scale involved in TSP, the impact of the coevolving genes on the immediate genomic neighbourhood of the selected sites will be reduced, because recombination decouples selected sites from linked neutral sites [16,123]. Therefore, long-term polymorphisms of single SNPs are difficult to detect by analysing genomic signatures alone. This can be seen with the SNP in the human haemoglobin gene responsible for the balanced sickle-cell anaemia–malaria resistance polymorphism in large parts of Africa. Long-term balancing selection can best be detected when the balanced haplotypes do not recombine (for example, if they are located in inverted regions of the genome or are part of a supergene [G]), or if multiple polymorphisms are together under selection, so that they maintain linkage disequilibrium across polymorphic sites [25, 53, 123].

Diffuse coevolution

The two models of specific coevolution described above assume that the relevant interactions are between one host and one parasite species. However, interactions may include multiple parasites and/or multiple hosts. This scenario is known as diffuse coevolution [10]. Diffuse coevolution refers particularly to the interactions of functional guilds [G], such as diverse species of herbivores and their plant hosts or parasites and

hosts. In its simplest form, coevolution is diffuse when a host trait and the underlying genes in the genome evolve in response to at least two parasites, or a parasite trait evolves in response to selection caused by more than one host. In such cases, coevolution is influenced by a combination of frequency dependence (according to the genetic composition of the parasite populations) and density dependence (according to the abundance of the parasite species). Furthermore, the parasite composition may differ in different sub-populations of the host and may change over time (owing to diversifying selection and/or spatio-temporal dynamics)[124-126]. The more hosts or parasites that participate, the more complex the evolution of the participants will be and the harder it will be to predict the genomic signature of the diffusely evolving regions [13].

The need for consideration of diffuse coevolution is underlined by examples of host loci that interact with multiple parasites. Human variants of CCR5, TRIM5α and APOBECG3 interact with HIV-1, but were suggested to have been under positive selection by other viruses with which they interacted in the past [127-129]. The tomato R-gene [G] *cf2* confers resistance to the parasitic nematode *Globodera rostochiensis* and to the fungus *Cladosporium fulvum* [130]. Parasites from three kingdoms interact in part with the same proteins in *Arabidopsis* [131]. The MHC region of diverse vertebrates are well known for their interactions with many parasites [125,126,132,133].

The converse, a parasite interacting with more than one host, seems also to be widespread. Sympatric species of sticklebacks are infected by the same parasites, which interact with the host's MHC [116]. The pathogenic bacteria *Staphylococcus aureus*, *Pseudomonas viridiflava* and *Pseudomonas syringae* are known to infect various host species with overlapping genetic mechanisms [74,76,134]. Finally, hosts may even exchange immune genes to fight parasites: for example, related *Arabidopsis* species have been shown to exchange resistance genes through hybridization and introgression [135]. Likewise, parasites may exchange coevolving genes via horizontal gene transfer to overcome host defenses [136].

Coevolution may become even more diffuse when the interactions of various hosts and parasite species show spatio-temporal dynamics. Coevolution in such multi-species scenarios is based on variable interactions between changing communities of hosts and parasites. Thus, the signatures observed in the host and parasite genomes cannot be attributed to one ecological setting, but result from a history with diverse settings with different interactions in host and parasite communities. Each setting may result in a different evolutionary trajectory and may well include temporal phases of more narrowly defined coevolution by NFDS or selective sweeps between pairs of antagonists [137]. The signatures of these phases of specific coevolution will not be recognizable in the genome, unless the ecological setting in which they occur are sufficiently stable for a given period of time.

Models of single hosts coevolving with multiple parasites support the idea that long term maintenance of genetic polymorphisms is possible [7,48,59,124,138,139]. Empirical data from systems likely under diffuse coevolution are consistent with this idea, leading to the widely held belief that the complexity of multi-species interactions maintains genetic diversity, including TSP [92,116,125-127,140-142]. Thus, diffuse coevolution offers an explanation for the "missing antagonist problem" that typifies TSP studies, which typically lack knowledge of the parasite that coevolved with the host in the past [120]. If TSP is caused by diffuse coevolution there is no single coevolving parasite, but an association with a changing pool of parasite species over time and space. Indeed, most of the associations of the human MHC with parasites (such as HIV, West Nile virus, dengue, hepatitis B, hepatitis C, tuberculosis and leprosy [133]) are believed to be rather young—much younger than 5 million years, the approximate date of the last common human–chimpanzee ancestor for which TSP at the MHC was observed [25,29,116]. Thus, the parasites we see interacting with humans now may not be the same as those that interacted with us some million years ago. Malaria, which is a parasite specific to humans, may be an exception. Recent work suggests that human–malaria coevolution could be as old as the human split from our closest living relatives [143-146]. However, the hypothesis that malaria is a missing antagonist, explaining TSP in humans, requires further investigation.

To further our understanding of TSP it will be necessary to distinguish between cases of specific long-term coevolution and diffuse coevolution. Although demonstrating that TSP is the consequence of ancient specific coevolution is challenging, we can make testable predictions. In the strictest case of specific coevolution by balancing selection, the aim would be to demonstrate that two closely related host species with evidence of TSP are parasitized by two closely related species of coevolving parasites (FIG. 4a) and that the genes in question play a functional role in the interaction of the two pairs of antagonists (FIG. 3a). In such a scenario, an overlay of long-term Red-Queen dynamics with the co-speciation of the host and the parasite might be observed[14](FIG. 4b). Such a strict set of conditions may not be very likely, but the requirements may be relaxed to accommodate ecological, historical, epidemiological and biogeographical features [116,138].

Joint analysis of host and parasite genomes

Unlike the approaches described above, emerging cogenomic [G] methods jointly analyse the genomes of hosts and those of their parasites. These methods focus on the genes responsible for the phenotypic interactions of the antagonists and allow subsequent analysis of their genomic signatures [72,147-149].

[H2] Identifying the genes involved in host–parasite interactions. A cogenomics approach to identify the genes that directly interact with each other in hosts and

parasites is best illustrated with a host–parasite matrix that shows the functional specificity of variants of the antagonists that interact with each other to produce phenotypes—disease (compatible) and resistance (incompatible) (FIG. 3a). Although many different matrices have been proposed [150-155], we still do not know much about them in natural host-parasite systems. Studies of interacting host–parasite genes have a long tradition in plant–pathogen systems, starting with H.H. Flor's gene-for-gene system for flax and one of its rust pathogens [156] (for reviews see: [22,91]), but only a few examples have so far been confirmed for animal systems [52,53]. It is widely believed that specific interaction matrices like these underlie coevolution by NFDS and thus are responsible for maintaining genetic diversity. Finding the genes underlying such matrices is still cumbersome and time consuming, but this may change with the development of cogenomics approaches that detect interspecies (or intergenomic) linkage-disequilibrium (iLD). If the only polymorphism in a population that explains variation in infection success is the interaction between the host's A-locus and the parasite's B-locus, then infected individuals could only ever carry the host–parasite allele combinations A/B and a/b. Other combinations (A/b and a/B) would not produce infections (FIG. 3a). This non-random association between host and parasite alleles produces iLD because the phenotype (infection) depends on the combination of genetic variants in different species. The statistical signal of iLD is only seen at the interacting loci, not at sites in the genetic background, as long as recombination efficiently decouples selected loci from non-selected loci (FIG. 5a). Using iLD is a powerful tool to detect interacting genes for a number of reasons. First, its presence indicates that host and parasite individuals carry alleles related to each other in a fitness context, that is, for a trait (infection) likely under selection. Second, it allows the interacting loci of both antagonists to be pinpointed. Third, it can be used even when only samples of infected hosts (such as infected patients) are available. Fourth, it is free from assumptions about the shape of the interaction matrix, which is important, as we currently have little understanding of how the matrix is structured. Finally, the method is not limited to model species, as it does not require previous knowledge about the system.

However, statistical limitations are apparent. The method works best when the alleles in question have intermediate frequencies and when the number of tests to be conducted (every host polymorphism is tested in relation to every parasite polymorphism) is not too large relative to the sample size: as the number of tests, and therefore false positives, increases with the product of the numbers of polymorphisms in both antagonists, larger sample sizes are required for antagonists with large genomes. In addition to these limitations, other factors can reduce the power to detect significant iLD. For example, if genes involved in iLD have epistatic (that is, non-additive) interactions with other loci in the same genome, the statistical power to detect them is reduced. Furthermore, simulations of different coevolutionary models

showed that the inferential power of iLD will vary over time due to the underlying dynamics of a host-parasite interaction matrix, and relevant loci will be more identifiable if coevolution occurs by NFDS than if by selective sweep coevolution [157]. Finally, multiple infections (that is, more than one parasite genotype being involved in the infection of individual hosts) may cause problems in the analysis. Multiple infections of humans and animals are believed to be very common [158].

A number of pioneering cogenomics studies have been performed with human patients infected with different parasites. In genome-to-genome analyses, iLD was used to test for associations between SNPs in humans with SNPs in HIV-1 [147], hepatitis C [159] and *Streptococcus pneumoniae* [160]. These studies were able to identify previously unknown interacting genes in the host and parasite, with the virus studies being more powerful because they have smaller genomes than *Streptococcus* and require fewer tests. This method is currently being further developed to increase its sensitivity and to take population stratification into account [161,162]. A modified version of this method was used to test for associations of SNPs in human candidate genes with phylogenetic lineages of *Mycobacterium tuberculosis*, revealing a host SNP – parasite lineage association [163].

Another cogenomics method, Analysis with a Two-Organism Mixed Model (ATOMM), also aims to find interacting loci in hosts and parasites (FIG. 5b)[164]. This method requires experimental data for phenotypes, such as infection and disease symptoms, from all possible combinations of host and parasite genotypes. It allows researchers to map these phenotypes to host and parasite genomes, while accounting for confounding factors such as phylogenetic or spatial structure.

Although not yet routinely used, cogenomic methods offer exciting possibilities for future studies of host–parasite interactions. The new pairs of loci identified with these methods can then be analysed for their genomic signatures, allowing the picture of the signature of coevolution in host and parasite genomes to be fine-tuned. The use of the genome-to-genome method with non-model organisms will allow the study of systems previously out of the reach of coevolutionary research.

[H2] Using cogenomics to infer signatures of the coevolutionary process. As of yet, we do not have access to analytical derivations that determine exactly how different coevolutionary processes affect allele frequencies at loci across host–parasite genomes. However, we can simulate much of the biological complexity that arises as a result of the coevolutionary process and explore how genomic signatures might differ under particular scenarios [157] (BOX 1). For example, one approach simulates coevolutionary dynamics to generate allele frequency expectations, which can then be used to look back in time to determine the coalescent properties that might generate these expected allele frequencies at loci in host and parasite genomes [21]. This general idea has been adapted into an inferential framework using Approximate Bayesian

519 Computation [G] (ABC) [149] for population genetic simulations of different
520 coevolutionary scenarios. The results of these simulations can be summarized by both
521 traditional population genetic summaries (Table 2) as well as novel summaries of the
522 host and parasite allele frequency spectra concurrently, which are then statistically
523 compared with polymorphism data obtained from host and parasite genomes to
524 determine their similarity. This approach enables the identification of loci that underlie
525 the coevolutionary process in both host and parasite genomes simultaneously, as well
526 as identifying important eco-evolutionary parameters such as the cost of infectivity.
527 The method can be applied to genome wide polymorphism data gathered from
528 controlled laboratory experiments or from natural populations. It may be used either
529 in concert with the iLD approach or entirely independently, as the false positive rate
530 for detecting genomic regions undergoing coevolution is rather low.
531

532 ## Conclusions and future perspectives

533 In a world of rapidly increasing transportation and migration, accelerating climate
534 change, de- and reforestation, altered agricultural and food handling practises, and
535 increasing human and livestock densities, host–parasite contact rates are already very
536 high and are only expected to increase, resulting in more intense reciprocal selection.
537 These changes affect many naturally coevolving systems in which humans have a
538 vested interest, such as malaria–mosquito, virus (including dengue, zika, West-Nile)–
539 mosquito, *Borrelia*–ticks, and *Mycoplasma*–house finch, as well as parasites believed
540 to coevolve with humans [163,165-167]. Although we still have a limited understanding of
541 how coevolution works in natural populations, the knowledge derived from the state-
542 of-the-art population genomic methodology described here can guide and inform
543 coevolutionary research and experimentation. We are now in a position to approach
544 new questions and gain a new perspective on old ones.
545 With few exceptions, we know little about the role of structural variation, such
546 as copy number variation and inversions, in the coevolutionary process. Copy number
547 variation, which is well known in the MHC, can be maintained by balancing selection
548 (reviewed in [125]). Supergenes have also been shown to play an important part in
549 balanced polymorphisms in various traits [168], and have recently been suggested to
550 have a decisive role in host–parasite coevolution [53]. Although it is still cumbersome to
551 identify structural polymorphisms in large samples, improved methods will certainly
552 make this an important aspect of coevolutionary research.
553 A broad survey of the literature on coevolution would suggest a taxon-specific
554 division between plant systems, centred on gene-for-gene infection models, and
555 animal systems, focused on matching-allele-models without invoking costs. It is
556 currently not clear if this division reflects a bias in our research efforts or has a
557 biological basis. Now, cogenomic methods will enable us to move beyond a few model
558 systems and rapidly collect more data on the type of infection matrix that

559 predominates in natural populations. This information is particularly required for
560 animal systems and non-agricultural plant systems, for which there are only a few
561 examples of interaction matrices. Understanding the costs of resistance will need
562 fitness assays in the presence and absence of the parasite.
563 Another challenge is to understand how genes within the same genome
564 interact to modify the dynamic with genes in the antagonist. Epistasis has long been
565 thought to be important in host-parasite interactions [48], and epistasis between host
566 resistance loci has been proposed for a number of systems [22,52,53,133,169]. However, the
567 generality and importance of these observations for host–parasite coevolution is not
568 yet clear. Epistasis is central to the theory of coevolution and of the evolution and
569 maintenance of genetic recombination. Genetic recombination works to speed up
570 evolutionary responses under antagonistic coevolution only if the recombining loci
571 show epistatic interactions [48,170,171] because epistasis may create negative linkage
572 disequilibrium among resistance alleles.
573 Knowing which genes are functionally linked and what epistatic relationships
574 exist would allow us to predict how reciprocal selection acts, whether specific or
575 diffuse coevolution occurs, and what traces we might expect at the genome level.
576 After decades of coevolution research, we still do not have examples of temporal
577 dynamics allele frequencies and their associated phenotypes in either hosts or
578 parasites, or an understanding of how the corresponding genomic sites in the
579 antagonists are functionally linked. However, recent progress in the field promises
580 change in the near future. Finally, our understanding of the genomic signatures of
581 coevolution is still largely correlational. Rigorous experimental work, including
582 experimental evolution studies, can help us scrutinize the evidence for coevolution,
583 link it to genomic signatures, and test specific model predictions [56,172-174].
584
585
586

References

1 Majerus, M., Amos, W. & Hurst, G. *Evolution: The Four Billion Year War* (Addison, Wesley Longman, 1996).

2 Jack, R. & Du Pasquier, L. *Evolutionary Concepts in Immunology*. (Springer Nature Switzerland, 2019).

3 Fumagalli, M. *et al.* Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLoS Genet.* **7**, doi:10.1371/journal.pgen.1002355 (2011).

4 Karasov, T. L., Horton, M. W. & Bergelson, J. Genomic variability as a driver of plant-pathogen coevolution? *Current Opinion in Plant Biology* **18**, 24-30, doi:10.1016/j.pbi.2013.12.003 (2014).

5 Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. & Potts, W. K. The Nature of Selection on the Major Histocompatibility Complex. *Crit Rev Immunol* **37**, 75-120, doi:10.1615/CritRevImmunol.v37.i2-6.10 (2017).

6 Lenz, T. L., Hafer, N., Samonte, I. E., Yeates, S. E. & Milinski, M. Cryptic haplotype-specific gamete selection yields offspring with optimal MHC immune genes. *Evolution* **72**, 2478-2490, doi:10.1111/evo.13591 (2018).

7 Penman, B. S. & Gupta, S. Detecting signatures of past pathogen selection on human HLA loci: are there needles in the haystack? *Parasitology* **145**, 731-739, doi:10.1017/S0031182017001159 (2018).

8 Koenig, D. *et al.* Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife* **8**, doi:10.7554/eLife.43606 (2019).

9 Guoy, A. & Excoffier, L. Polygenic Patterns of Adaptive Introgression in Modern Humans Are Mainly Shaped by Response to Pathogens. *Mol. Biol. Evol.* **37**, 1420-1433, doi:10.1093/molbev/msz306 (2020).

10 Janzen, D. H. When Is It Coevolution. *Evolution* **34**, 611-612, doi:Doi 10.2307/2408229 (1980).

11 Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32**, 569-577 (2002).

12 Kiester, A. R., Lande, R. & Schemske, D. W. Models of coevolution and speciation in plants and their pollinators. *Am. Nat.* **124**, 220-243 (1984).

13 Wade, M. J. The co-evolutionary genetics of ecological communities. *Nat. Rev. Genet.* **8**, 185-195 (2007).

14 de Vienne, D. M. *et al.* Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist* **198**, 347-385, doi:10.1111/nph.12150 (2013).

15 Ebert, D. Host-parasite coevolution: Insights from the Daphnia-parasite model system. *Curr. Opin. Mircobiol.* **11**, 290-301 (2008).

16 Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2**, 379-384, doi:10.1371/journal.pgen.0020064 (2006).
This is an authorative review on the population genetics of balancing selection.

17 Zivkovic, D., John, S., Verin, M., Stephan, W. & Tellier, A. Neutral genomic signatures of host-parasite coevolution. *BMC Evol Biol* **19**, doi:10.1186/s12862-019-1556-3 (2019).

18 Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution*. (Oxford University Press, 1995).

19 Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539-548, doi:10.1038/nrg2812 (2010).

20 Sironi, M., Cagliani, R., Forni, D. & Clerici, M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat. Rev. Genet.* **16**, 224-236, doi:10.1038/nrg3905 (2015).

21 Tellier, A., Moreno-Gamez, S. & Stephan, W. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution* **68**, 2211-2224, doi:10.1111/evo.12427 (2014).

22 Thrall, P. H., Barrett, L. G., Dodds, P. N. & Burdon, J. J. Epidemiological and Evolutionary Outcomes in Gene-for-Gene and Matching Allele Models. *Frontiers in Plant Science* **6**, doi:10.3389/fpls.2015.01084 (2016).

23 Ebert, D. Open questions: what are the genes underlying antagonistic coevolution? *Bmc Biology* **16**, doi:10.1186/s12915-018-0583-7 (2018).

24 Fischer, M. C., Foll, M., Heckel, G. & Excoffier, L. Continental-Scale Footprint of Balancing and Positive Selection in a Small Rodent (*Microtus arvalis*). *Plos One* **9**, doi:10.1371/journal.pone.0112332 (2014).

25 Leffler, E. M. *et al.* Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science* **339**, 1578-1582, doi:10.1126/science.1234070 (2013).
This study reports on many sites with TSPs and ancient balancing selection in genomes of humans and chimpanzee.

26 Key, F. M., Teixeira, J. C., de Filippo, C. & Andres, A. M. Advantageous diversity maintained by balancing selection in humans. *Curr Opin Genet Dev* **29**, 45-51, doi:10.1016/j.gde.2014.08.001 (2014).

27 Schweizer, R. M. *et al.* Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. *Mol. Biol. Evol.* **35**, 1190-1209, doi:10.1093/molbev/msy031 (2018).

28 Bitarello, B. D. *et al.* Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biology and Evolution* **10**, 939-955, doi:10.1093/gbe/evy054 (2018).

29 Cheng, X. H. & DeGiorgio, M. Detection of Shared Balancing Selection in the Absence of Trans-Species Polymorphism. *Mol. Biol. Evol.* **36**, 177-199, doi:10.1093/molbev/msy202 (2019).

30 Enard, D., Cai, L., Gwennap, C. & Petrov, D. A. Viruses are a dominant driver of protein adaptation in mammals. *Elife* **5**, doi:10.7554/eLife.12469 (2016).

31 Schirrmann, M. K. *et al.* Genomewide signatures of selection in Epichloe reveal candidate genes for host specialization. *Mol. Ecol.* **27**, 3070-3086, doi:10.1111/mec.14585 (2018).

32 Persoons, A. *et al.* The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Mol. Ecol.* **26**, 1902-1918, doi:10.1111/mec.13980 (2017).

33 Mohd-Assaad, N., McDonald, B. A. & Croll, D. Genome-Wide Detection of Genes Under Positive Selection in Worldwide Populations of the Barley Scald Pathogen. *Genome Biology and Evolution* **10**, 1315-1332, doi:10.1093/gbe/evy087 (2018).

34 Badouin, H. *et al.* Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol. Ecol.* **26**, 2041-2062, doi:10.1111/mec.13976 (2017).

35 Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. The evolution of RNAi as a defence against viruses and transposable elements. *Phil. Trans. R. Soc. B* **364**, 99-115, doi:10.1098/Rstb.2008.0168 (2009).

36 Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827-839, doi:10.1038/nrg3564 (2013).

37 Hermisson, J. & Pennings, P. S. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol Evol* **8**, 700-716, doi:10.1111/2041-210x.12808 (2017).

38 Hahn, M. W. *Molecular Population Genetics*. (Sinauer Associates, 2018).

39 McCarthy, K. R., Kirmaier, A., Autissier, P. & Johnson, W. E. Evolutionary and Functional Analysis of Old World Primate TRIM5 Reveals the Ancient Emergence of Primate Lentiviruses and Convergent Evolution Targeting a Conserved Capsid Interface. *PLoS Pathog.* **11**, doi:10.1371/journal.ppat.1005085 (2015).

40 Elena, S. F., Cooper, V. S. & Lenski, R. E. Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**, 1802-1804 (1996).

41 Anderson, T. J. C. *et al.* Population Parameters Underlying an Ongoing Soft Sweep in Southeast Asian Malaria Parasites. *Mol. Biol. Evol.* **34**, 131-144, doi:10.1093/molbev/msw228 (2017).

42 Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLoS Genet.* **11**, e1005004, doi:10.1371/journal.pgen.1005004 (2015).

43 Sanchez-Vallet, A. *et al.* The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *Ann. Rev. Phytopathol.* **56**, 21-40, doi:10.1146/annurev-phyto-080516-035303 (2018).

44 Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24-33 (1994).

45 Raffaele, S. *et al.* Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. *Science* **330**, 1540-1543, doi:10.1126/science.1193070 (2010).

46 Croll, D. & McDonald, B. A. The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *PLoS Pathog.* **8**, doi:10.1371/journal.ppat.1002608 (2012).

47 Clarke, B. C. The ecological relationship of host-parasite relationships. in *Genetic aspects of host-parasite relationships* (eds A.E.R. Taylor & R.M. Muller) 87-104 (Blackwell, Oxford, 1976).

48 Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites. *Proc. Nat. Acad. Sci. USA* **87**, 3566-3573 (1990).

49 Fenton, A., Antonovics, J. & Brockhurst, Michael A. Inverse-Gene-for-Gene Infection Genetics and Coevolutionary Dynamics. *Am. Nat.* **174**, E230-E242, doi:10.1086/645087 (2009).

50 Schmid-Hempel, P. *Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics*. (Oxford University Press, 2011).
This is a comprehensive text book summarizing the field of host-parasite evolution and coevolution.

51 Ben Khalifa, M., Simon, V., Fakhfakh, H. & Moury, B. Tunisian Potato virus Y isolates with unnecessary pathogenicity towards pepper: support for the matching allele model in eIF4E resistance-potyvirus interactions. *Plant Pathol.* **61**, 441-447, doi:10.1111/j.1365-3059.2011.02540.x (2012).

52 Luijckx, P., Fienberg, H., Duneau, D. & Ebert, D. A Matching-Allele Model Explains Host Resistance to Parasites. *Curr. Biol.* **23**, 1085-1088, doi:10.1016/J.Cub.2013.04.064 (2013).
This study provides an early example of a well worked out matching-allele host-parasite interaction matrix.

53 Bento, G. *et al.* The genetic basis of resistance and matching-allele interactions of a host-parasite system: The Daphnia magna-Pasteuria ramosa model. *PLoS Genet.* **13**, doi:10.1371/journal.pgen.1006596 (2017).

54 King, K. C., Jokela, J. & Lively, C. M. Parasites, Sex, and Clonal Diversity in Natural Snail Populations. *Evolution* **65**, 1474-1481, doi:Doi 10.1111/J.1558-5646.2010.01215.X (2011).

55 Ashby, B. & Boots, M. Multi-mode fluctuating selection in host-parasite coevolution. *Ecol. Lett.* **20**, 357-365, doi:10.1111/ele.12734 (2017).

56 Papkou, A. *et al.* The genomic basis of Red Queen dynamics during rapid reciprocal host-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **116**, 923-928, doi:10.1073/pnas.1810402116 (2019).
This study of experimental evolution with nematodes and a bacterial pathogen demonstrate the complexity of coevolutionary interactions emerging in seemingly simple systems.

57 Koskella, B. & Lively, C. M. Evidence for Negative Frequency-Dependent Selection during Experimental Coevolution of a Freshwater Snail and a Sterilizing Trematode. *Evolution* **63**, 2213-2221, doi:10.1111/J.1558-5646.2009.00711.X (2009).

58 Lively, C. M. Habitat Heterogeneity, Host Population Structure, and Parasite Local Adaptation. *J. Hered.* **109**, 29-37, doi:10.1093/jhered/esx100 (2018).

59 Ejsmond, M. J., Babik, W. & Radwan, J. MHC allele frequency distributions under parasite-driven selection: A simulation model. *BMC Evol Biol* **10**, doi:10.1186/1471-2148-10-332 (2010).

60 Fijarczyk, A. & Babik, W. Detecting balancing selection in genomes: limits and prospects. *Mol. Ecol.* **24**, 3529-3545, doi:10.1111/mec.13226 (2015).

61 Bubb, K. L. *et al.* Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* **173**, 2165-2177, doi:10.1534/genetics.106.055715 (2006).

62 Cagliani, R. *et al.* The signature of long-standing balancing selection at the human defensin beta-1 promoter. *Genome Biol* **9**, R143, doi:10.1186/gb-2008-9-9-r143 (2008).

63 Fumagalli, M. *et al.* Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Research* **19**, 199-212, doi:10.1101/Gr.082768.108 (2009).

64 Segurel, L. *et al.* The ABO blood group is a trans-species polymorphism in primates. *Proc. Natl. Acad. Sci. USA* **109**, 18493-18498, doi:10.1073/pnas.1210603109 (2012).

65 Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. C. Evolutionary dynamics of plant R-genes. *Science* **292**, 2281-2285, doi:10.1126/science.1061337 (2001).

66 Hoerger, A. C. *et al.* Balancing Selection at the Tomato RCR3 Guardee Gene Family Maintains Variation in Strength of Pathogen Defense. *PLoS Genet.* **8**, doi:10.1371/journal.pgen.1002813 (2012).

67 Llaurens, V., Whibley, A. & Joron, M. Genetic architecture and balancing selection: the life and death of differentiated variants. *Mol. Ecol.* **26**, 2430-2448, doi:10.1111/mec.14051 (2017).

68 Croze, M. *et al.* A genome-wide scan for genes under balancing selection in Drosophila melanogaster. *BMC Evol Biol* **17**, doi:10.1186/s12862-016-0857-z (2017).

69 Buckley, J., Holub, E. B., Koch, M. A., Vergeer, P. & Mable, B. K. Restriction associated DNA-genotyping at multiple spatial scales in *Arabidopsis lyrata* reveals signatures of pathogen-mediated selection. *Bmc Genomics* **19**, doi:10.1186/s12864-018-4806-7 (2018).

70 Wu, Q. *et al.* Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. *Genome Biology* **18**, doi:10.1186/s13059-017-1342-8 (2017).

71 Unckless, R. L., Howick, V. M. & Lazzaro, B. P. Convergent Balancing Selection on an Antimicrobial Peptide in Drosophila. *Curr. Biol.* **26**, 257-262, doi:10.1016/j.cub.2015.11.063 (2016).

72 Bartoli, C. & Roux, F. Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. *Frontiers in Plant Science* **8**, doi:10.3389/fpls.2017.00763 (2017).

73 Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* **18**, 41-50, doi:10.1038/nrg.2016.132 (2017).

74 Thomas, J. C., Godfrey, P. A., Feldgarden, M. & Robinson, A. Candidate Targets of Balancing Selection in the Genome of Staphylococcus aureus. *Mol. Biol. Evol.* **29**, 1175-1186, doi:10.1093/molbev/msr286 (2012).

75 Zhang, L. F., Thomas, J. C., Didelot, X. & Robinson, D. A. Molecular Signatures Identify a Candidate Target of Balancing Selection in an arcD-Like Gene of Staphylococcus epidermidis. *J. Mol. Evol.* **75**, 43-54, doi:10.1007/s00239-012-9520-5 (2012).

76 Guttman, D. S., Gropp, S. J., Morgan, R. L. & Wang, P. W. Diversifying selection drives the evolution of the type III secretion system pilus of Pseudomonas syringae. *Mol. Biol. Evol.* **23**, 2342-2354, doi:10.1093/molbev/msl103 (2006).

77 Castillo, J. A. & Agathos, S. N. A genome-wide scan for genes under balancing selection in the plant pathogen Ralstonia solanacearum. *BMC Evol Biol* **19**, doi:10.1186/s12862-019-1456-6 (2019).

78 Ryabov, E. V. *et al.* Dynamic evolution in the key honey bee pathogen deformed wing virus: Novel insights into virulence and competition using reverse genetics. *PLoS Biol.* **17**, doi:10.1371/journal.pbio.3000502 (2019).

79 Andras, J. P., Fields, P. D., Du Pasquier, L., Fredericksen, M. & Ebert, D. Genome-wide association analysis identifies a genetic basis of infectivity in a model bacterial pathogen. . *Mol. Biol. Evol.* (in press), doi:10.1093/molbev/msaa173 (2020).

80 Brackney, D. E., Beane, J. E. & Ebel, G. D. RNAi Targeting of West Nile Virus in Mosquito Midguts Promotes Virus Diversification. *PLoS Pathog.* **5**, doi:10.1371/journal.ppat.1000502 (2009).

81 Brackney, D. E., Schirtzinger, E. E., Harrison, T. D., Ebel, G. D. & Hanley, K. A. Modulation of Flavivirus Population Diversity by RNA Interference. *J. Virol.* **89**, 4035-4039, doi:10.1128/Jvi.02612-14 (2015).

82 Obbard, D. J., Jiggins, F. M., Halligan, D. L. & Little, T. J. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* **16**, 580-585, doi:Doi 10.1016/J.Cub.2006.01.065 (2006).

83 Araki, H. *et al.* Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. *Proc. Natl. Acad. Sci. USA* **103**, 5887-5892, doi:10.1073/pnas.0601431103 (2006).

84 Nygaard, S. *et al.* Long- and Short-Term Selective Forces on Malaria Parasite Genomes. *PLoS Genet.* **6**, doi:10.1371/journal.pgen.1001099 (2010).

85 Ochola, L. I. *et al.* Allele Frequency-Based and Polymorphism-Versus-Divergence Indices of Balancing Selection in a New Filtered Set of Polymorphic Genes in Plasmodium falciparum. *Mol. Biol. Evol.* **27**, 2344-2351, doi:10.1093/molbev/msq119 (2010).

86 Salathe, M., Kouyos, R. D., Regoes, R. R. & Bonhoeffer, S. Rapid parasite adaptation drives selection for high recombination rates. *Evolution* **62**, 295-300 (2008).

87 Gokhale, C. S., Papkou, A., Traulsen, A. & Schulenburg, H. Lotka-Volterra dynamics kills the Red Queen: population size fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC Evol Biol* **13**, doi:10.1186/1471-2148-13-254 (2013).

88 Schenk, H., Schulenburg, H. & Traulsen, A. How long do Red Queen dynamics survive under genetic drift? A comparative analysis of evolutionary and eco-evolutionary models. *BMC Evol Biol* **20**, doi:10.1186/s12862-019-1562-5 (2020).

89 MacPherson, A., Keeling, M. J. & Otto, S. P. Coevolution does not slow the rate of loss of heterozygosity in a stochastic host-parasite model with constant population size. *bioRxiv*, doi:10.1101/2020.04.07.024661 (2020).

90 Thrall, P. H. & Burdon, J. J. Effect of resistance variation in a natural plant host-pathogen metapopulation on disease dynamics. *Plant Pathol.* **49**, 767-773, doi:DOI 10.1046/j.1365-3059.2000.00523.x (2000).

91 Brown, J. K. M. & Tellier, A. Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. *Annu Rev Phytopathol* **49**, 345-367, doi:10.1146/Annurev-Phyto-072910-09530l (2011).

92 Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. & Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genetics* **36**, 298-311, doi:10.1016/j.tig.2020.01.008 (2020).

93 Charlesworth, B., Nordborg, M. & Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res. Camb.* **70**, 155-174 (1997).

94 Eizaguirre, C., Lenz, T. L., Kalbe, M. & Milinski, M. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecol. Lett.* **15**, 723-731, doi:10.1111/j.1461-0248.2012.01791.x (2012).

95 Rico, Y. *et al.* Spatial patterns of immunogenetic and neutral variation underscore the conservation value of small, isolated American badger populations. *Evolutionary Applications* **9**, 1271-1284, doi:10.1111/eva.12410 (2016).

96 Jousimo, J. *et al.* Ecological and evolutionary effects of fragmentation on infectious disease dynamics. *Science* **344**, 1289-1293, doi:10.1126/science.1253621 (2014).

97 Crispo, E. *et al.* The evolution of the major histocompatibility complex in upstream versus downstream river populations of the longnose dace. *Ecol Evol* **7**, 3297-3311, doi:10.1002/ece3.2839 (2017).

98 Keller, M. F. *et al.* Trans-ethnic meta-analysis of white blood cell phenotypes. *Human Molecular Genetics* **23**, 6944-6960, doi:10.1093/hmg/ddu401 (2014).

99 Morgan, A. D., Gandon, S. & Buckling, A. The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* **437**, 253-256 (2005).

100 Thrall, P. H. *et al.* Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecol. Lett.* **15**, 425-435, doi:10.1111/J.1461-0248.2012.01749.X (2012).

101 Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225-1241 (2004).

102 Croll, D. & McDonald, B. A. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol. Ecol.* **26**, 2027-2040, doi:10.1111/mec.13870 (2017).

103 Laine, A. L., Burdon, J. J., Dodds, P. N. & Thrall, P. H. Spatial variation in disease resistance: from molecules to metapopulations. *J. Ecol.* **99**, 96-112, doi:10.1111/J.1365-2745.2010.01738.X (2011).

104 Bolnick, D. I. & Stutz, W. E. Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature* **546**, 285-288, doi:10.1038/nature22351 (2017).

This experimental study with fish shows that rare immigrants have an advantage over resident genotypes. The study demonstrates elegantly that resistance genes have higher effective migration rates.

105 Phillips, K. P. *et al.* Immunogenetic novelty confers a selective advantage in host-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **115**, 1552-1557, doi:10.1073/pnas.1708597115 (2018).

106 Rico, Y., Morris-Pocock, J., Zigouris, J., Nocera, J. J. & Kyle, C. J. Lack of Spatial Immunogenetic Structure among Wolverine (*Gulo gulo*) Populations Suggestive of Broad Scale Balancing Selection. *Plos One* **10**, doi:10.1371/journal.pone.0140170 (2015).

107 Leducq, J. B. *et al.* Effect of balancing selection on spatial genetic structure within populations: theoretical investigations on the self-incompatibility locus and empirical studies in Arabidopsis halleri. *Heredity* **106**, 319-329, doi:10.1038/hdy.2010.68 (2011).

108 Castric, V., Bechsgaard, J., Schierup, M. H. & Vekemans, X. Repeated Adaptive Introgression at a Gene under Multiallelic Balancing Selection. *PLoS Genet.* **4**, doi:10.1371/journal.pgen.1000168 (2008).

109 Hoban, S. *et al.* Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *Am. Nat.* **188**, 379-397, doi:10.1086/688018 (2016).

110 Borg, A. A., Pedersen, S. A., Jensen, H. & Westerdahl, H. Variation in MHC genotypes in two populations of house sparrow (Passer domesticus) with different population histories. *Ecol Evol* **1**, 145-159, doi:10.1002/ece3.13 (2011).

111 Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98-U95, doi:10.1038/nature07331 (2008).

112 Fields, P. D., Reisser, C., Dukic, M., Haag, C. R. & Ebert, D. Genes mirror geography in *Daphnia magna*. *Mol. Ecol.* **24**, 4521-4536, doi:10.1111/mec.13324 (2015).

113 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University of Chicago Press, 2005).

114 Laine, A. L., Barres, B., Numminen, E. & Siren, J. P. Variable opportunities for outcrossing result in hotspots of novel genetic variation in a pathogen metapopulation. *eLife* **8**, doi:10.7554/eLife.47091 (2019).

115 Klein, J. *Immunology*. (Blackwell, 1990).

116 Lenz, T. L., Eizaguirre, C., Kalbe, M. & Milinski, M. Evaluating Patterns of Convergent Evolution and Trans-Species Polymorphism at Mhc Immunogenes in Two Sympatric Stickleback Species. *Evolution* **67**, 2400-2412, doi:10.1111/evo.12124 (2013).
This study demonstrates TSP in two sympatric stickleback fish sharing the same parasites. The authors were able to rule out convergent evolution as an alternative explanation for TSP.

117 Tesicky, M. & Vinkler, M. Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon? *J Immunol Res*, doi:10.1155/2015/838035 (2015).

118 Mboup, M., Fischer, I., Lainer, H. & Stephan, W. Trans-Species Polymorphism and Allele-Specific Expression in the CBF Gene Family of Wild Tomatoes. *Mol. Biol. Evol.* **29**, 3641-3652, doi:10.1093/molbev/mss176 (2012).

119 Novikova, P. Y. *et al.* Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics* **48**, 1077–1082, doi:10.1038/ng.3617 (2016).

120 Azevedo, L., Serrano, C., Amorim, A. & Cooper, D. N. Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Hum Genomics* **9**, doi:10.1186/s40246-015-0043-1 (2015).

121 Lenz, T. L. Computational Prediction of Mhc Ii-Antigen Binding Supports Divergent Allele Advantage and Explains Trans-Species Polymorphism. *Evolution* **65**, 2380-2390, doi:10.1111/j.1558-5646.2011.01288.x (2011).

122 Eizaguirre, C., Lenz, T. L., Kalbe, M. & Milinski, M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications* **3**, doi:10.1038/ncomms1632 (2012).

123 Gao, Z. Y., Przeworski, M. & Sella, G. Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. *Evolution* **69**, 431-446, doi:10.1111/evo.12567 (2015).

124 Hedrick, P. W. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902-1908 (2002).

125 Eizaguirre, C. & Lenz, T. L. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *J. Fish Biol.* **77**, 2023-2047, doi:10.1111/j.1095-8649.2010.02819.x (2010).

126 Osborne, M. J., Pilger, T. J., Lusk, J. D. & Turner, T. F. Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class II in the endangered Rio Grande silvery minnow. *Mol. Ecol.* **26**, 471-489, doi:10.1111/mec.13936 (2017).

127 Daugherty, M. D. & Malik, H. S. Rules of Engagement: Molecular Insights from Host-Virus Arms Races. *Annu Rev Genet* **46**, 677-700, doi:10.1146/annurev-genet-110711-155522 (2012).

128 Cagliani, R. *et al.* A Positively Selected Apobec3h Haplotype Is Associated with Natural Resistance to Hiv-1 Infection. *Evolution* **65**, 3311-3322, doi:10.1111/j.1558-5646.2011.01368.x (2011).

129 Davis, Z. H. *et al.* Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Molecular Cell* **57**, 349-360, doi:10.1016/j.molcel.2014.11.026 (2015).

130 Lozano-Torres, J. L. *et al.* Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proc. Natl. Acad. Sci. USA* **109**, 10119-10124, doi:10.1073/pnas.1202867109 (2012).

131 Wessling, R. *et al.* Convergent Targeting of a Common Host Protein-Network by Pathogen Effectors from Three Kingdoms of Life. *Cell Host & Microbe* **16**, 364-375, doi:10.1016/j.chom.2014.08.004 (2014).

132 Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H. & Milinski, M. Parasite selection for immunogenetic optimality. *Science* **301**, 1343-1343 (2003).

133 Matzaraki, V., Kumar, V., Wijmenga, C. & Zhernakova, A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology* **18**, doi:10.1186/s13059-017-1207-1 (2017).

134 Karasov, T. L., Barrett, L., Hershberg, R. & Bergelson, J. Similar levels of gene content variation observed for Pseudomonas syringae populations extracted from single and multiple host species. *Plos One* **12**, doi:10.1371/journal.pone.0184195 (2017).

135 Bechsgaard, J., Jorgensen, T. H. & Schierup, M. H. Evidence for Adaptive Introgression of Disease Resistance Genes Among Closely Related Arabidopsis Species. *G3-Genes Genomes Genetics* **7**, 2677-2683, doi:10.1534/g3.117.043984 (2017).

136 Gluck-Thaler, E. & Slot, J. C. Dimensions of Horizontal Gene Transfer in Eukaryotic Microbial Pathogens. *PLoS Pathog.* **11**, doi:ARTN e1005156 10.1371/journal.ppat.1005156 (2015).

137 Campbell, M. C., Ashong, B., Teng, S. L., Harvey, J. & Cross, C. N. Multiple selective sweeps of ancient polymorphisms in and around LT alpha located in the MHC class III region on chromosome 6. *BMC Evol Biol* **19**, 218, doi:10.1186/s12862-019-1557-2 (2019).

138 Karasov, T. L. *et al.* The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* **512**, 436-U472, doi:10.1038/nature13439 (2014).

139 Rabajante, J. F. *et al.* Red Queen dynamics in multi-host and multi-parasite interaction system. *Scientific Reports* **5**, doi:10.1038/srep10004 (2015).

140 Kamath, P. L., Turner, W. C., Kusters, M. & Getz, W. M. Parasite-mediated selection drives an immunogenetic trade-off in plains zebras (Equus quagga). *Proc. R. Soc. B* **281**, doi:10.1098/rspb.2014.0077 (2014).

141 Nadeem, A. & Wahl, L. M. Prophage as a genetic reservoir: Promoting diversity and driving innovation in the host community. *Evolution* **71**, 2080-2089, doi:10.1111/evo.13287 (2017).

142 Fortuna, M. A. *et al.* Coevolutionary dynamics shape the structure of bacteria-phage infection networks. *Evolution* **73**, 1001-1011, doi:10.1111/evo.13731 (2019).

143 Silva, J. C. *et al.* Genome sequences reveal divergence times of malaria parasite lineages. *Parasitology* **138**, 1737-1749, doi:10.1017/S0031182010001575 (2011).

144 Galen, S. C. *et al.* The polyphyly of Plasmodium: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *Roy Soc Open Sci* **5**, doi:10.1098/rsos.171780 (2018).

145 Otto, T. D. *et al.* Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* **3**, 687-697, doi:10.1038/s41564-018-0162-2 (2018).

146 Pacheco, M. A. *et al.* Mode and Rate of Evolution of Haemosporidian Mitochondrial Genomes: Timing the Radiation of Avian Parasites. *Mol. Biol. Evol.* **35**, 383-403, doi:10.1093/molbev/msx285 (2018).

147 Bartha, I. *et al.* A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* **2**, doi:10.7554/eLife.01123 (2013).
This study describes a prioneering method in cogenomics, applied to interacting genomic sites in hosts and parasites.

148 Lees, J. A., Tonkin-Hill, G. & Bentley, S. D. GENOME WATCH Stronger together. *Nat. Rev. Microbiol.* **15**, 516-516, doi:10.1038/nrmicro.2017.95 (2017).

149 Märkle, H. & Tellier, A. Inference of coevolutionary dynamics and parameters from host and parasite polymorphism data of repeated experiments. *PLoS Computational Biology* **16**, e1007668, doi:10.1371/journal.pcbi.1007668 (2020).

150 Otto, S. P. & Nuismer, S. L. Species interactions and the evolution of sex. *Science* **304**, 1018-1020 (2004).

151 Tellier, A. & Brown, J. K. M. Polymorphism in multilocus host-parasite coevolutionary interactions. *Genetics* **177**, 1777-1790, doi:10.1534/Genetics.107.074393 (2007).

152 Engelstadter, J. & Bonhoeffer, S. Red Queen Dynamics with Non-Standard Fitness Interactions. *PLoS Computational Biology* **5**, doi:10.1371/Journal.Pcbi.1000469 (2009).

153 Best, A. *et al.* The Evolution of Host-Parasite Range. *Am. Nat.* **176**, 63-71, doi:10.1086/653002 (2010).

154 Fenton, A., Antonovics, J. & Brockhurst, M. A. Two-Step Infection Processes Can Lead to Coevolution between Functionally Independent Infection and Resistance Pathways. *Evolution* **66**, 2030-2041, doi:10.1111/J.1558-5646.2012.01578.X (2012).

155 Kwiatkowski, M., Engelstadter, J. & Vorburger, C. On Genetic Specificity in Symbiont-Mediated Host-Parasite Coevolution. *Plos Computational Biology* **8**, doi:10.1371/journal.pcbi.1002633 (2012).

156 Flor, H. H. Host-Parasite Interaction in Flax Rust - Its Genetics and Other Implications. *Phytopathol.* **45**, 680-685 (1955).

157 Märkle, H., Tellier, A. & John, S. Cross-Species association statistics for genome-wide studies of host and parasite polymorphism data. *bioRxiv*, 726166, doi:10.1101/726166 (2019).

158 Balmer, O. & Tanner, M. Prevalence and implications of multiple-strain infections. *Lancet Infect Dis* **11**, 868-878, doi:10.1016/S1473-3099(11)70241-9 (2011).

159 Ansari, M. A. *et al.* Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nature Genetics* **49**, 666–673, doi:10.1038/ng.3835 (2017).
This study describes a strong example of the application of cogenomics to find interacting loci in humans infected with hepatitis C-virus.

160 Lees, J. A. *et al.* Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nature Communications* **10**, doi:10.1038/s41467-019-09976-3 (2019).

161 Naret, O. *et al.* Correcting for Population Stratification Reduces False Positive and False Negative Results in Joint Analyses of Host and Pathogen Genomes. *Frontiers in Genetics* **9**, doi:10.3389/fgene.2018.00266 (2018).

162 Ansari, M. A. *et al.* Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *eLife* **8**, doi:10.7554/eLife.42463 (2019).

163 McHenry, M. L. *et al.* Interaction between host genes and Mycobacterium tuberculosis lineage can affect tuberculosis severity: evidence for coevolution? *PLoS Genet* **16**, e1008728, doi:10.1371/journal.pgen.1008728 (2020).

164 Wang, M. Y. *et al.* Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc. Natl. Acad. Sci. USA* **115**, E5440-E5449, doi:10.1073/pnas.1710980115 (2018).
This study describes the development of a powerful cogenomics method that utilizes data from an interaction matrix of all combinations of host and parasite genotypes to find the genomic sites that underlay the interaction.

165 Hill, A. V. S., Jepson, A., Plebanski, M. & Gilbert, S. C. Genetic analysis of host-parasite coevolution in human malaria. *Phil. Trans. R. Soc. B* **352**, 1317-1325 (1997).

166 Heeney, J. L., Dalgleish, A. G. & Weiss, R. A. Origins of HIV and the evolution of resistance to AIDS. *Science* **313**, 462-466, doi:10.1126/science.1123016 (2006).

167 Hertz, T. *et al.* Mapping the Landscape of Host-Pathogen Coevolution: HLA Class I Binding and Its Relationship with Evolutionary Conservation in Human and Viral Proteins. *J. Virol.* **85**, 1310-1321, doi:10.1128/Jvi.01966-10 (2011).

168 Schwander, T., Libbrecht, R. & Keller, L. Supergenes and Complex Phenotypes. *Curr. Biol.* **24**, R288-R294, doi:10.1016/j.cub.2014.01.056 (2014).

169 Lenz, T. L. *et al.* Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nature Genetics* **47**, 1085–1090, doi:10.1038/ng.3379 (2015).

170 Salathe, M., Kouyos, R. D. & Bonhoeffer, S. The state of affairs in the kingdom of the Red Queen. *Trends Ecol. Evol.* **23**, 439-445 (2008).

171 da Silva, J. & Galbraith, J. D. Hill-Robertson interference maintained by Red Queen dynamics favours the evolution of sex. *J. evol. Biol.* **30**, 994-1010, doi:10.1111/jeb.13068 (2017).

172 Kubinak, J. L. *et al.* Experimental viral evolution reveals major histocompatibility complex polymorphisms as the primary host factors controlling pathogen adaptation and virulence. *Genes Immun* **14**, 365-372, doi:10.1038/gene.2013.27 (2013).

173 Brockhurst, M. A. & Koskella, B. Experimental coevolution of species interactions. *Trends Ecol. Evol.* **28**, 367-375, doi:10.1016/j.tree.2013.02.009 (2013).

174 Retel, C. *et al.* The feedback between selection and demography shapes genomic diversity during coevolution. *Sci Adv* **5**, doi:10.1126/sciadv.aax0530 (2019).

175 Figueroa, F., Gúnther, E. & Klein, J. MHC polymorphism pre-dating speciation. *Nature* **335**, 265-267, doi:10.1038/335265a0 (1988).

176 Mcdonald, J. H. & Kreitman, M. Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature* **351**, 652-654, doi:10.1038/351652a0 (1991).

177 Eyre-Walker, A. & Keightley, P. D. Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Mol. Biol. Evol.* **26**, 2097-2108, doi:10.1093/molbev/msp119 (2009).

178 Cheng, X. & DeGiorgio, M. Detection of Shared Balancing Selection in the Absence of Trans-Species Polymorphism. *Mol. Biol. Evol.* **36**, 177-199, doi:10.1093/molbev/msy202 (2018).

179 Nielsen, R. Molecular Signatures of Natural Selection. *Ann. Rev. Genet.* **39**, 197-218, doi:10.1146/annurev.genet.39.073003.112420 (2005).

180 Siewert, K. M. & Voight, B. F. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Mol. Biol. Evol.* **34**, 2996-3005, doi:10.1093/molbev/msx209 (2017).

181 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol.* **4**, e72, doi:10.1371/journal.pbio.0040072 (2006).

182 Messer, P. W. & Petrov, D. A. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* **28**, 659-669, doi:10.1016/j.tree.2013.08.003 (2013).

183 DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A Model-Based Approach for Identifying Signatures of Ancient Balancing Selection in Genetic Data. *PLoS Genet.* **10**, e1004561, doi:10.1371/journal.pgen.1004561 (2014).

184 Kim, Y. & Stephan, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**, 765-777 (2002).

185 Cheng, X. & DeGiorgio, M. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *bioRxiv*, 645887, doi:10.1101/645887 (2019).

186 Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410-418, doi:10.1016/j.tree.2010.04.001 (2010).

187 Schrider, D. R. & Kern, A. D. Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends Genetics* **34**, 301-312, doi:10.1016/j.tig.2017.12.005 (2018).
This paper offers an accessible description of both present applications and possible future developments of supervised machine learning for understanding signatures of selection in genomic scale data.

188 Raynal, L. *et al.* ABC random forests for Bayesian parameter inference. *Bioinformatics* **35**, 1720-1728, doi:10.1093/bioinformatics/bty867 (2018).

189 Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS Genet.* **10**, e1004342, doi:10.1371/journal.pgen.1004342 (2014).

190 Kelleher, J., Etheridge, A. M. & McVean, G. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology* **12**, e1004842, doi:10.1371/journal.pcbi.1004842 (2016).

191 Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W. & Ralph, P. L. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Mol Ecol Resour* **19**, 552-566, doi:10.1111/1755-0998.12968 (2019).
This paper describes the implementation of tree-sequence recording into the already multifaceted and powerful SLiM simulation framework, and provides one of the most important schemes needed to model neutral and non-neutral dynamics on genome-scale data.

192 Hejase, H. A., Dukler, N. & Siepel, A. From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection. *Trends Genetics*, doi:10.1016/j.tig.2019.12.008 (2020).
This is an exceptionally comprehensive review of both historical and present approaches for detecting forms of positive selection. While the focus is on positive selection, many of the focal methodologies would, with some modification, be applicable for detecting the many signatures of host-parasite coevolution.

Acknowledgements

We thank the Ebert lab for fruitful discussion and Suzanne Zweizig for comments on the language of the manuscript. This work is supported by a grant from the Swiss National Science Foundation.

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

The authors declare no competing interests.

Table 1. Comparison of features of specific coevolution by selective sweep and balancing selection.

Feature	Selective sweep coevolution	Balancing selection coevolution
Form of selection	Positive selection drives sweeps; selection is directional.	Negative frequency-dependent selection gives common alleles a disadvantage; selection results in a balance of the frequencies of genetic variants.
Functional polymorphisms	Visible only during selective sweeps.	Maintained constantly and potentially for very long time periods.
Underlying genetic system	Beneficial mutation in host and parasite at any locus in the nuclear or cytoplasmic genome may sweep.	Frequencies of alternative alleles at a few selected loci are balanced.
Role of mutations	Mutations define the onset of new selective sweeps (hard sweeps).	Mutations are not necessary but do create rare variants, which may be selected and contribute to balancing selection or even replace a previous variant.
Temporal continuity	Process can be highly stochastic and does not need to be continuous; long periods without sweeps are possible.	Process must operate continuously because genetic variants may otherwise be lost. In a spatial setting, previously lost alleles may be reintroduced from other populations.
Time scale of phenotypic change	Relatively slow because new mutations take a long time to reach a high enough frequency to be recognized. Sweeps starting from standing genetic variation progress more quickly.	Fast because genetic variants are always at intermediate frequencies where selection results in fast changes.
Population divergence	Sweeps drive population and species divergence.	Population divergence is prevented in the long term, though it may occur in the short term.
Evolutionary outcome	Creates macroevolutionary patterns (lineage divergence).	Explains high levels of genetic diversity within populations and species.
Introgression among species	May introduce beneficial new alleles that can sweep.	May introduce new functional variants that can contribute to balancing selection, but may create a fake picture of trans-species polymorphism.

Table 2. Genomic summaries used to determine the evolutionary process underlying a genomic signature.

Signature type refers to the two models of specific coevolution that may be supported with this summary. Time-scale refers to the approximate period during which the summary is able to detect a signal for the genomic signature relative to the genomic background (compare Fig. 2). CLR, Composite Likelihood Ratio Test; HKA, Hudson, Kreitman and Aguadé test; iHS, integrated haplotype score; LD, linkage disequilibrium; MK, McDonald-Kreitman; NCD, Non-central deviation; SFS, site frequency spectrum; TSP, trans-species polymorphism.

Name of summary	Type of summary	Description	Signature type (relative to genomic background)	Time-scale	Local genomic scale of analysis	Other comments
Allele vs. species-tree discordance [175]	Summary statistic	Compares phylogenetic relationships of genes to those of species.	Discordant trees indicate balancing selection and/or TSP	Deep	Distinct genes and/or windows across the genome	_
Elevated ratio of adaptive to non-adaptive divergence [176,177]	Summary statistic and likelihood method	Measures the ratio of adaptive to synonymous divergence	High values indicate positive selection	Deep-to-intermediate	Gene classes can be compared	Explicit tests arise from the MK test or its derivatives
HKA_{trans} [178]	Summary statistic	Adaptation of the standard HKA test to better accommodate genomic data from multiple species	Positive values of chi-squared test statistic are suggestive of balancing selection	Deep-to-intermediate	Windows across genome	_
NCD_{trans} [28,178]	Summary statistic	Extension of the NCD summary statistic to accommodate multi-species data	Increasing values are suggestive of balancing selection	Deep-to-intermediate	Windows across genome	See also [178]. Describes the NCD1 summary which includes an outgroup and fixed differences as part of the test
Tajima's D and HKA [179]	Summary statistic	Excess of low frequency class polymorphisms	Decreased values indicate positive selection/ Selective sweep	Intermediate-to-recent	Windows across genome	Summary can be confounded with demography due to population expansion causing similar measures of the summary statistic
		Increased values indicate balancing selection	Increased values indicate balancing selection	Deep-to-intermediate	Windows across genome	Summary can be confounded with demography due to population bottlenecks

						causing similar measures of the summary statistic
Perturbations from equilibrium SFS [179]	Summary statistic and likelihood method	Intermediate-skewed allele frequency classes	Excess derived allele frequency classes as compared to neutral expectation is indicative of balancing selection	Deep-to-intermediate	Windows across genome	Summary can be confounded with demography due to population bottlenecks causing similar measures of the summary statistic
		Excess minor and/or derived allele frequency classes	Excess minor or derived allele frequency classes as compared to neutral expectation is indicative of positive selection and seletive sweeps	Intermediate-to-recent	Windows across genome	Summary can be confounded with demography due to population expansion causing similar measures of the summary statistic
β [180]	Summary statistic	Measure of allele frequency correlation and overall mutation rate.	Values greater than 0 are indicative of balancing selection	Intermediate	Windows across genome	
iHS [181]	Summary statistic	Lengths of haplotype homozygosity or the decelerated decay of LD	Increased lengths of homozygosity and the slower decay of LD compared to neutral expectations is indicative of a selective sweep	Recent	Windows across genome	
LD [179]	Summary statistic	Distinct increase in LD amongst a subset of adjacent loci	Elevated values above genome wide background indicate a selective sweep	Recent	Windows across genome	
H_1, H_{12},H_2 [42,182]	Summary statistic	Frequencies of ranked haplotypes, that is haplotype spectra	Increased frequencies of distinct, ranked haplotypes (and their relative frequencies)	Deep-to-intermediate	Windows (SNPs) across genome	

			is indicative of a sweep (hard versus soft).			
T_{trans} [178]	Likelihood method	Adaptation of the T statistic of [183] to specifically detect TSP	Fit of the observed data to a model of long-term balancing selection or TSP is greater than that of a model of neutral evolution ($CLR_{TSP} > CLR_{Neutral}$)	Deep-to-intermediate	Windows across genome	
T statistic [183]	Likelihood method	Aggregate of summary statistics of genomic diversity parameterize a maximum likelihood-based comparison of the fit of a neutral vs. balanced polymorphism model	Fit of the observed data to a model of balancing selection is greater than that of a model of neutral evolution ($CLR_{Balancing\ selection} > CLR_{Neutral}$)	Intermediate-to-recent	Windows across genome	
CLR [184]	Likelihood method	Aggregate of summary statistics of genomic diversity parameterize a maximum likelihood-based comparison of the fit of a neutral vs.selective sweep model	Fit of the observed data to a model of balancing selection is greater than that of a model of neutral evolution ($CLR_{Sweep} > CLR_{Neutral}$)	Recent	Windows across genome	Refers specifically to the model described in ref. [184] and its derivatives; while CLR is used sometimes to refer to this specific summary, composite likelihood ratio tests are a general statistical procedure for model comparison.

Figure legends

Figure 1 | **Schematic representation of selective sweep coevolution in the gene pools of a host and parasite population.** For each antagonist, timelines of alleles at different loci are shown in different colours. Over the shown time period, each of the alleles is replaced at least once by a mutant, indicating selective sweeps. Genetic variation at loci, that is the presence of multiple alleles in the gene pool at a given moment in time, is only visible during the sweep. Host and parasite sweeps are not linked. Loci can be present anywhere in the genome. Other loci may undergo sweeps at the same time for reasons unrelated to the coevolutionary interactions.

Figure 2 | **The temporal dynamics of genomic signatures.**
Approximate time scales during which genomic signatures for balancing selection (blue), positive selection and/or selective sweep (red), and interspecies linkage disequilibrium (purple) are detectable with common (though not exhaustive) population genetic tests (Table 2). While signatures of balancing selection (especially those that are associated with trans-species polymorphism (TSP) or approximations thereof; see [185]) and positive selection may extend quite deep into evolutionary time, patterns of very rapid evolutionary change associated with positive selection or selective sweeps will show a much more limited time depth. Concomitantly, the number of generations a given selective regime has to be in place before inference of its signature is possible may be different between balancing selection and positive selection and, on the shortest times scales and with a sample from a single population, the signatures of balancing selection and positive selection may be indistinguishable. As a selective sweep grows older, its signature will disappear and become more similar to the signature of positive selection, which is visible only as elevated ratios of adaptive to synonymous divergence. Patterns of interspecies linkage disequilibrium provide no historical record of the coevolutionary process and so additional methods (such as those described in this figure) are required to determine what (if any) selective regime has been at play at the statistically associated loci. It should be noted, the suggested time frame of inference for individual tests may differ substantially for individual datasets as a result of experimental factors, such as sample size, and as the result of species-specific factors, such as demography. Allele frequency classes may be a component of summary statistic (such as Tajima's D) or may be the specific metric of interest (for example, site-frequency spectra, or SFS). T_{TRANS} = likelihood ratio test statistics for detecting TSP; HKA_{TRANS} = modified version of the HKA test to accommodate genomic data from multiple species in order to detect TSP; NCD = non-central deviation summary to determine deviation from neutral allele frequency expectations; H_1,H_2, and H_{12} = Frequencies of the most frequent haplotype, the frequency of the second most frequent haplotype, and the frequency of the first and second most frequent haplotype, respectively; NCD_{TRANS} = version of the NCD summary statistic extended to multispecies datasets in order to detect TSP [28,178]; HKA = Hudson,

Kreitman and Aguadé test; SFS = site frequency spectrum; iHS = integrated haplotype score [181]; LD = linkage disequilibrium [179]; CLR = composite likelihood ratio test [184]; β = measure of allele frequency correlation [180] (see Table 2).

Figure 3 | **Host resistotype and parasite infectotype interaction matrix and balancing selection.** A simple matching allele model of two host resistotypes and two parasite infectotypes is sufficient to create long term balancing selection within populations. **a** | Haploid hosts with the H-allele have resistotype C (C for compatible interactions, that is, the host is susceptible) for parasites with the P-allele and I for incompatible (the host is resistant) for parasites with the p-allele; hosts with the h-allele have resistotype I for parasites with the P-allele and C for parasites with the p-allele. The haploid parasite with the P-allele has the infectotype C for hosts with the H allele and I for hosts with the h-allele, whereas parasites with the p-allele have infectotype I for hosts with the H-allele and C for hosts with the h-allele. A mutant of an allele may produce a new allele (H', h', P' and p'), but if the function of the allele is not affected in regards to its interaction with the parasite, the resistotypes and infectotypes will stay the same. **b** | A schematic representation of balancing selection is shown in a host population that diverges into two populations. Allele colours correspond to functional types of the host. Initially, the H-allele (blue) gives rise to the h-allele (red), which has a different resistotype. Afterwards, the two allele types are maintained by balancing selection. Their infection profile might resemble the matching allele model shown in part a. Further mutations in the H- and h-alleles change the genotype of these alleles, but not their resistotype. The H-allele persists until it is replaced by the H' allele (light blue) in population 1. The h-allele gives rise to the h' allele (pink) in population 2. H and H', as well as h and h', change in relative frequency independent of selection by the parasite. Gene flow between populations can introduce new alleles (h' into population 1) and reintroduce extinct alleles (h' into population 2), which can replace the resident allele.

Figure 4 | **An idealized scenario for host–parasite coevolution by long-term balancing selection leading to trans-species polymorphisms.**
a | When balancing selection maintains allelic variants over long time periods, trans-species polymorphism (TSP) may be visible, with the polymorphisms existing prior to the split into two species. In contrast to the scenario in FIG. 3b, here, speciation completely blocks gene flow and hybridization. If TSP results from strict long-term coevolution, similar evolutionary histories are expected for the functionally linked variants in hosts and parasites (FIG. 3a). **b** | For the scenario outlined in part a, the genes undergoing long-term balancing selection are expected to cluster by function in

a gene tree (red and blue colour), not by species. Host and parasite genealogies should, however, be congruent (that is, have the same topology), indicating that speciation events occurred in parallel.

Figure 5 | **Cogenomic approaches to find genes involved in host parasite interactions.** A cogenomic analysis tests for associations between polymorphisms in host and parasite genomes. **a** | The genome-to-genome method analyses samples of hosts naturally infected with a parasite to find host variants in strong interspecies linkage disequilibrium (iLD) with variants in the parasites. Phenotypic data are not necessary. The figure shows two pairs of sites in the host and parasite genomes that significantly associate with each other, that is, they show strong iLD. **b** | The analysis with a Two-Organism Mixed Model (ATOMM) includes phenotypic data from a host-parasite infection matrix, which are analysed together with polymorphism data for both the host and parasite isolates [164]. The method allows interacting loci in hosts and parasites to be detected. Host and parasite genomes are represented as a series of squares, where each square indicates the position of a polymorphism in the sampled genomes. Shades of blue and red indicate different host and parasite genomes, respectively. Unequal numbers of host and parasite isolates can be used.

Box 1. Emerging cogenomics methodologies

Likelihood-free inference

The biological complexity inherent to the coevolutionary process and the distinct limitations of population genetic summary statistics, including the lack of a suitable and/or tractable likelihood function [G], for dealing with certain aspects of this complexity have driven researchers to seek new solutions [149]. A very successful approach for by-passing some limitations of individual summary statistics and likelihood-based inference has been Approximate Bayesian Computation (ABC). In ABC, the requirement of a likelihood function is bypassed by the use of simulated datasets. Summarization of these simulations via summary statistics followed by a statistical evaluation, or rejection procedure, identifies which simulations most closely approximate the set of summary statistic values obtained from an observed dataset, thereby allowing inference of the posterior probabilities of model parameters of interest [186]. Another exciting approach that allows the aggregate use of simulation and summary statistics to evaluate complex biological dynamics is supervised machine learning [G]. It works on the principle of using training sets to generate a predictive model, for example a simulated dataset of the relevant evolutionary scenarios to a given problem of interest. This predictive model is then used to determine how a set of input variables predicts a given response [187]. However, these different approaches should not be considered mutually exclusive[188].

Ancestral recombination graphs

To identify a signature of selective sweep or balancing selection using the methods described above, a clear understanding of how a neutral coalescent differs from a non-neutral one is crucial. Given that so much useful information can be gleaned from the flanking genomic regions as well as the locus experiencing selection, traditional summaries of the coalescent that rely on treating loci or genomic windows as (semi-) independent will throw away important, if not crucial, information needed to infer the dynamics of the coevolutionary process. The ancestral recombination graph [G] (ARG) provides a representation of the relationships among genomic segments, mediated through recombination, as a network [189]. Historically, it has been computationally prohibitive to reconstruct these ARGs in even small genomic data sets. However, recent advances in the data structures required to encode the information of an ARG [190], and in the simulation of complicated, non-neutral perturbations of the coalescent in the context of whole genomes [191], have begun to allow the inference of both selective sweeps and balancing selection from reconstructed ARGs [189,192]. It is perhaps not inconceivable that future advances may well allow for the direct reconstruction of the pairwise relationships of genomic segments in host and parasite, which is mediated though the reproductive processes of host and parasite, and the infection process that connects them.

Glossary

PARASITES
Organisms, including pathogens, that take advantage of other organisms (hosts), thereby instigating a process of selection by the host to defend against the parasite.

GENOMIC SIGNATURES
Characteristic patterns of genetic variation, observed at a genomic region in a sample of genomes.

SELECTIVE SWEEP
The spread of a beneficial mutant and the hitchhiking of genetic variants close to it in the genome. Beneficial mutants may have arisen de novo or were segregating in the population before the sweep and become beneficial after a change in conditions.

LINKED SELECTION
The evolution of (nearly) neutral SNPs influenced by selection on loci physically linked to them.

GENETIC HITCHHIKING
The process by which a genetic variant changes in frequency because it is physically linked to another variant that is changing in frequency due to selection.

COGENOMIC
The simultaneous analysis of the function, structure and evolution of pairs of associated genomes in closely-interacting organisms, such as host and parasites.

BALANCING SELECTION
Balancing selection occurs because the alleles involved have, on average, a selective advantage that correlates negatively with their frequency within a population or species. This term does not include loci that are at a balance between gaining and losing variants, such as mutation-drift and mutation-selection balance.

OVERDOMINANCE
Describes the scenario in which heterozygotes have a more extreme phenotypic trait value than all homozygotes. Overdominance for fitness results in balancing selection for the alleles causing the advantage for the heterozygote genotypes.

LOCAL ADAPTATION
If the effect of an allele is habitat specific, such that it is beneficial in one habitat and detrimental in another and vice versa for the alternative allele, local adaptation may evolve with directional selection within each population. Local adaption is a powerful mechanism to maintain genetic diversity within species.

DIRECT NEGATIVE FREQUENCY-DEPENDENT SELECTION
Occurs when the selective benefit of an allele depends directly and negatively on its frequency, for example at sex-determining loci and plant self-incompatibility loci.

INDIRECT NEGATIVE FREQUENCY-DEPENDENT SELECTION
Occurs when the selective benefit of an allele depends on the frequency of an allele in a coevolving species.

SELECTIVE INTERFERENCE
In clonal, but not in sexual, populations beneficial mutations interfere with each other, such that at a given moment the fittest mutation will outcompete weaker beneficial mutations. Interference can also affect the spread of mutants in genomic regions with low recombination rates.

LINKAGE DISEQUILIBRIUM (LD)
A statistical measure of the distribution of combinations of alleles at different loci, which is zero if this distribution follows the expectation based only on allele frequencies. Non-zero values of LD can arise due to hitchhiking, selection on allele combinations and stochastic processes and may occur among loci without physical linkage.

SELECTION COEFFICIENT
A measure of fitness of genotypes or alleles relative to a reference, such as the ancestral form.

TAJIMA'S *D*
A population genetic summary statistic describing the frequency distribution of polymorphisms in a population, with *D* being zero under neutral evolution and positive under balancing selection.

GENETIC DRIFT
A neutral evolutionary process that influences allele frequencies based on the random sampling of genetic variants during reproduction.

PANMICTIC
Random mating within a population.

FIXATION INDEX (F_{ST})
A measure of genetic differentiation of spatially structured populations, usually estimated from single-nucleotide polymorphism (SNP) or microsatellite data.

DIRECTIONAL SELECTION
A mode of natural selection by which a genetic variant is predicted to spread to fixation (also known as positive selection).

SUPERGENE
A group of tightly linked genes on a chromosome that are inherited together as a haplotype and often have reduced recombination.

FUNCTIONAL GUILDS
Groups of organisms with similar life-style characteristics that perform the same ecological function, such as gut-parasites, pollinators and filter-feeders.

R-GENE
Resistance genes of plants that convey resistance against diseases by producing R proteins.

APPROXIMATE BAYESIAN COMPUTATION (ABC)
Bayesian statistical approach wherein parameter inference and model selection is conducted in the absence of likelihood functions. Instead, ABC relies on summary statistics and simulations to infer posterior distributions of parameters and/or models of interest.

LIKELIHOOD FUNCTION
Analytical formulation of a set of parameters, which can be used to assess the fit of a given observed dataset to a predetermined model.

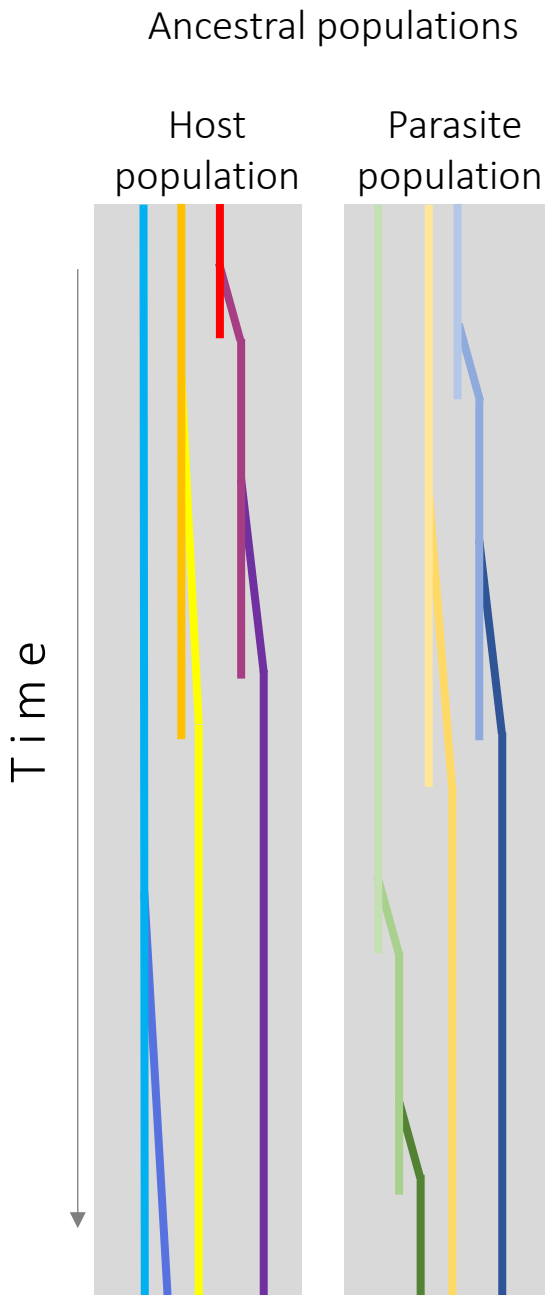
SUPERVISED MACHINE LEARNING
Machine learning is a statistical methodology that uses artificial intelligence to automate inferential processes with minimal explicit instruction. Supervised machine learning is a type of machine learning that uses (labelled) training sets to generate a target function when the correspondence between the function of interest and response variable are known. This target function can then be applied to unclassified (unlabelled) data to make statistical inferences.

ANCESTRAL RECOMBINATION GRAPH (ARG)
A genealogical or phylogenetic representation of the network of coalescence and recombination events in a collection of orthologous DNA sequences.

ToC summary

Host-parasite co-evolution is expected to leave signatures of selection in the genome of both antagonists. Ebert and Fields discuss what is known about these signatures, how they relate to co-evolutionary processes, and how they can help identify the genes underlying the coevolving phenotypes.

Figure 1.



A beneficial host mutant (violet) arises and replaces the ancestral allele (red).

A beneficial parasite mutant (blue) arises and replaces the ancestral allele (light blue).

A further beneficial host mutant (dark violet) arises and replaces the allele (violet) that earlier swept through the population.

More sweeps occur in both the host and the parasite population.

In relatively short succession, a beneficial mutant arises and replaces the ancestral parasite allele (light green), and is then itself replaced by a further beneficial mutant (dark green).

A beneficial mutant (dark blue) sweeps very slowly in the host population, causing a long time period of a polymorphism.

Figure 2.

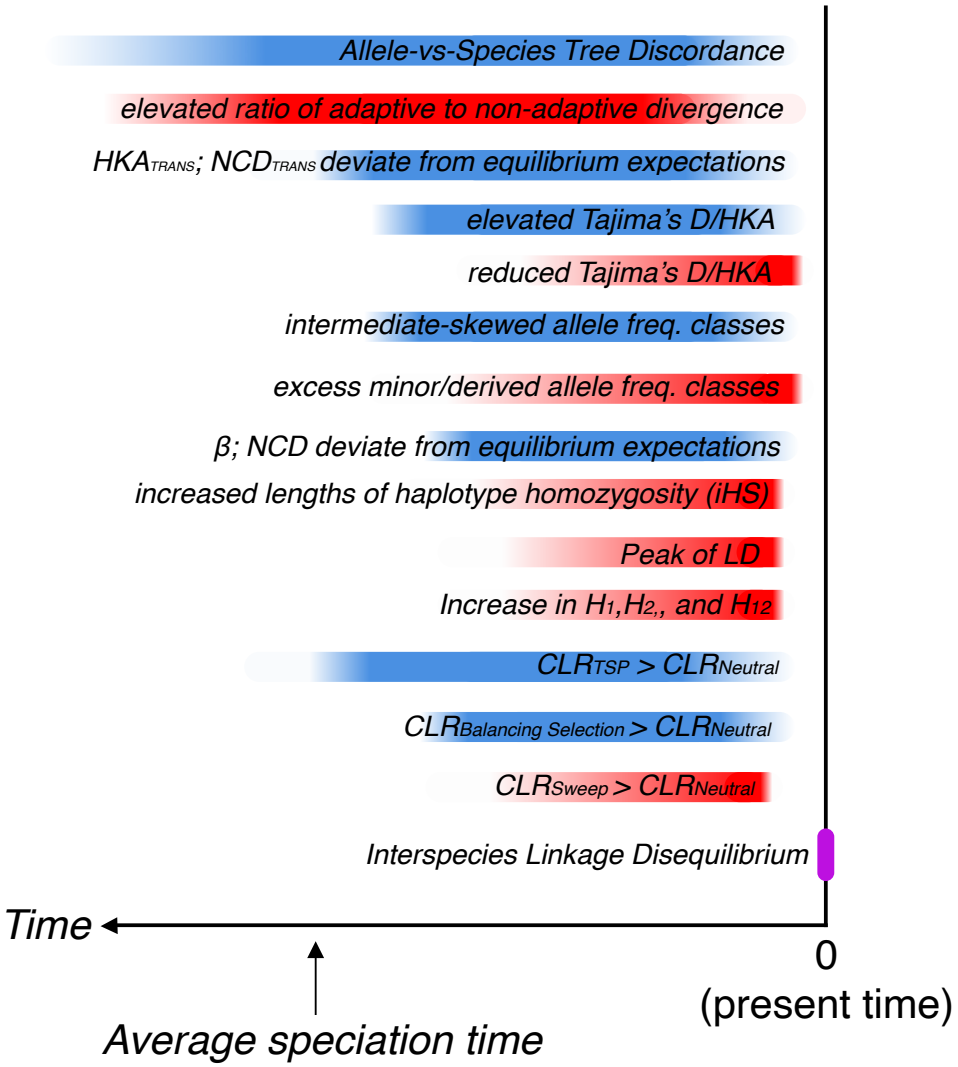
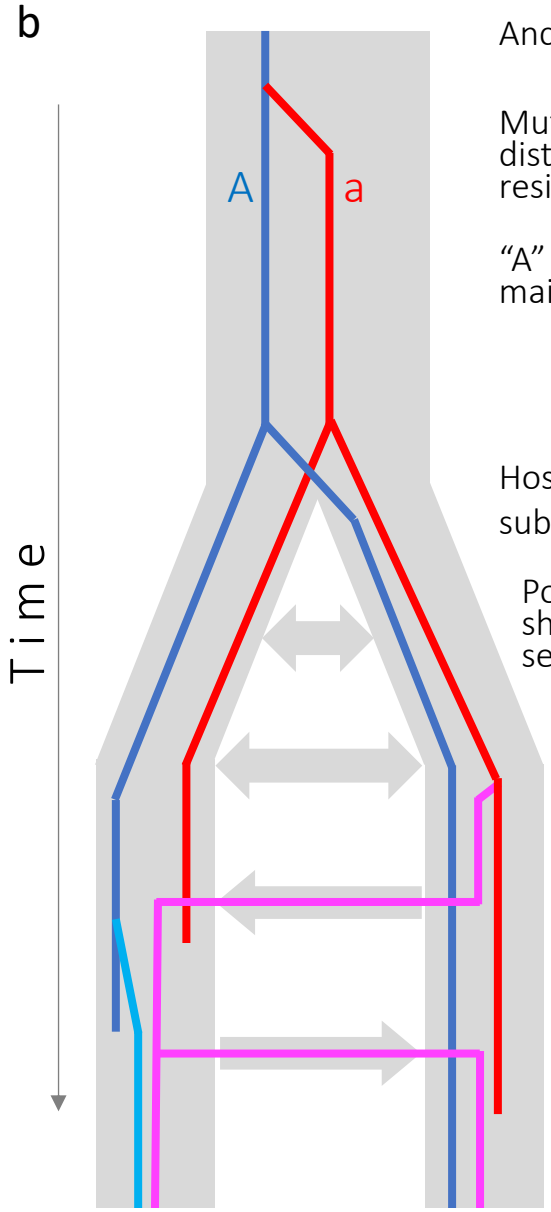

Allele-vs-Species Tree Discordance
elevated ratio of adaptive to non-adaptive divergence
HKA$_{TRANS}$; NCD$_{TRANS}$ deviate from equilibrium expectations
elevated Tajima's D/HKA
reduced Tajima's D/HKA
intermediate-skewed allele freq. classes
excess minor/derived allele freq. classes
β; NCD deviate from equilibrium expectations
increased lengths of haplotype homozygosity (iHS)
Peak of LD
Increase in H_1,H_2,, and H_{12}
$CLR_{TSP} > CLR_{Neutral}$
$CLR_{Balancing\ Selection} > CLR_{Neutral}$
$CLR_{Sweep} > CLR_{Neutral}$
Interspecies Linkage Disequilibrium
Time
0
(present time)
Average speciation time

Fig. 3.

a

		Parasite alleles	
		B, B'	**b, b'**
Host alleles	**A, A'**	**compatible, C**	**incompatible, I**
	a, a'	**incompatible, I**	**compatible, C**

b



Ancestral, monomorphic host population

Mutant arises in the host. “a” is functionally distinguished from “A”, i.e. it has a different resistotype (see Fig. 2).

“A” / “a” resistotype polymorphism is maintained by balancing selection

Host population diverges into two subpopulations

Populations diverge at neutral sites (not shown), but not at sites under balancing selection.

Geneflow among populations reduces further divergence

A new mutant (pink) arises

Geneflow introduces the pink allele in Population 1. Pink replaces red allele.

Light blue allele replaces ancestral dark blue allele

Geneflow brings extinct pink allele back, which then replaces the red allele.

Fig. 4.

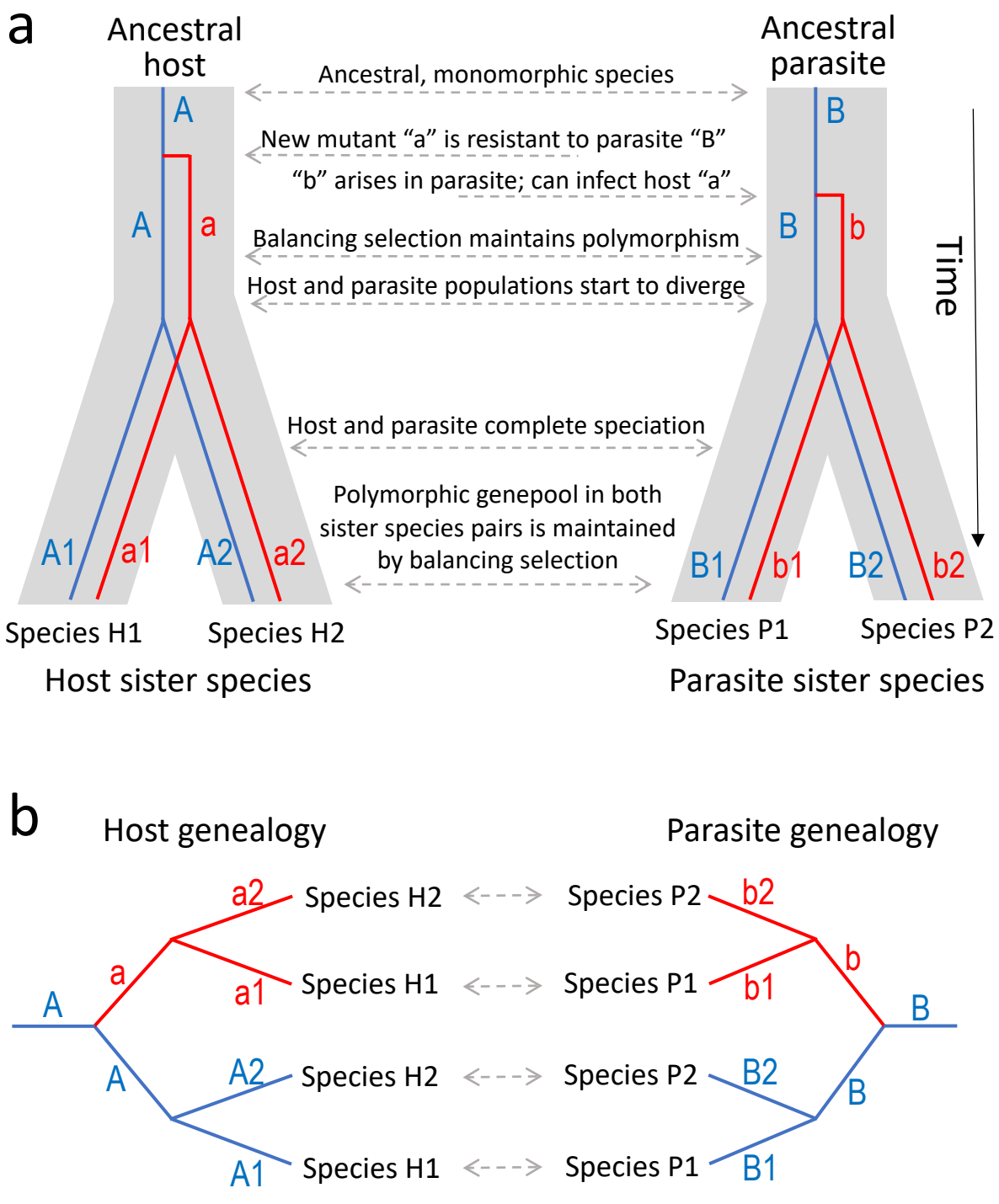
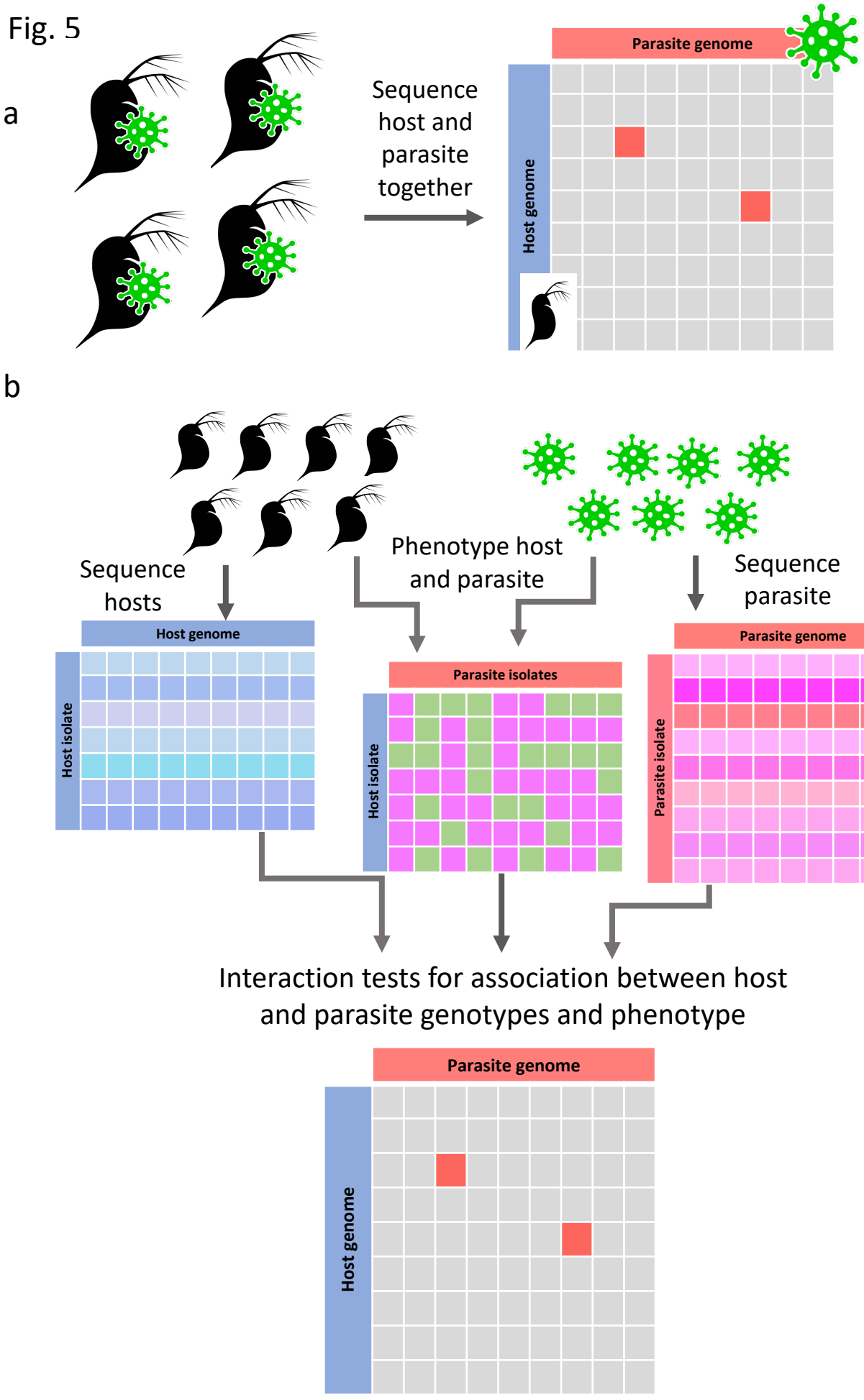

a
Ancestral host
Ancestral parasite
Ancestral, monomorphic species
New mutant “a” is resistant to parasite “B”
“b” arises in parasite; can infect host “a”
Balancing selection maintains polymorphism
Host and parasite populations start to diverge
Host and parasite complete speciation
Polymorphic genepool in both sister species pairs is maintained by balancing selection
Time
A
a
A1
a1
A2
a2
B
b
B1
b1
B2
b2
Species H1
Species H2
Species P1
Species P2
Host sister species
Parasite sister species
b
Host genealogy
Parasite genealogy
Species H2
Species P2
Species H1
Species P1
Species H2
Species P2
Species H1
Species P1

Fig. 5


a
Sequence host and parasite together
Parasite genome
Host genome
b
Sequence hosts
Phenotype host and parasite
Sequence parasite
Host genome
Host isolate
Parasite isolates
Host isolate
Parasite genome
Parasite isolate
Interaction tests for association between host and parasite genotypes and phenotype
Parasite genome
Host genome