Received: 20 January 2021 | Revised: 9 May 2021 | Accepted: 20 May 2021

DOI: 10.1111/mec.16001

INVITED REVIEW AND SYNTHESES

MOLECULAR ECOLOGY WILEY

Novel genomic approaches to study antagonistic coevolution between hosts and parasites

Hanna Märkle[1,2] | Sona John[1] | Amandine Cornille[3] | Peter D. Fields[4] | Aurélien Tellier[1]

[1]Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Freising, Germany

[2]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

[3]INRAE, CNRS, AgroParisTech, GQE - Le Moulon, Université Paris-Saclay, Gif-sur-Yvette, France

[4]Department of Environmental Sciences, University of Basel, Zoology, Basel, Switzerland

Correspondence
Aurélien Tellier, Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Freising, Germany.
Email: tellier@wzw.tum.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: TE809/3 - Project 274542535; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Special Topic Network from the European Society for Evolutionary Biology (STN-ESEB); Technische Universitat Munchen

Abstract

Host-parasite coevolution is ubiquitous, shaping genetic and phenotypic diversity and the evolutionary trajectory of interacting species. With the advances of high throughput sequencing technologies applicable to model and non-model organisms alike, it is now feasible to study in greater detail (a) the genetic underpinnings of coevolution, (b) the speed and type of dynamics at coevolving loci, and (c) the genomic consequences of coevolution. This review focuses on three recently developed approaches that leverage information from host and parasite full genome data simultaneously to pinpoint coevolving loci and draw inference on the coevolutionary history. First, co-genome-wide association study (co-GWAS) methods allow pinpointing the loci underlying host-parasite interactions. These methods focus on detecting associations between genetic variants and the outcome of experimental infection tests or on correlations between genomes of naturally infected hosts and their infecting parasites. Second, extensions to population genomics methods can detect genes under coevolution and infer the coevolutionary history, such as fitness costs. Third, correlations between host and parasite population size in time are indicative of coevolution, and polymorphism levels across independent spatially distributed populations of hosts and parasites can reveal coevolutionary loci and infer coevolutionary history. We describe the principles of these three approaches and discuss their advantages and limitations based on coevolutionary theory. We present recommendations for their application to various host (prokaryotes, fungi, plants, and animals) and parasite (viruses, bacteria, fungi, and macroparasites) species. We conclude by pointing out methodological and theoretical gaps to be filled to extract maximum information from full genome data and thereby to shed light on the molecular underpinnings of coevolution.

KEYWORDS
balancing selection, genetic drift, genomics, host-parasite coevolution, inference, positive selection

1 | INTRODUCTION

Species interactions are ubiquitous in natural populations. They are often characterized by interspecific genotype-by-genotype (GxG) interactions, which can drive evolutionary change in the interacting species. Coevolution is a well-known example of interspecies interactions that can result in reciprocal evolutionary change. Depending on the fitness of interacting species, coevolutionary interactions fall in a continuum between mutualistic (positive/positive), antagonistic (positive/negative), and competitive (negative/negative) interactions. Antagonistic coevolutionary interactions characterized by one species, the parasite, increasing its fitness at the expense of its host's fitness, are of particular interest in medicine and agriculture (Brown, 2015; Fitzpatrick et al., 2020; Pflughoeft & Versalovic, 2012). Thus, field collections, experimental work, and recent advances in sequencing technology are continuously contributing to understanding the extent to which hosts (and particularly multicellular plants or animals) interact with micro-organisms (Fitzpatrick et al., 2020; Pflughoeft & Versalovic, 2012).

With sequencing advances, a so-far unprecedented amount of host and parasite genomic data is becoming available. These data include samples from single and multiple natural populations (including humans), at one or several time points or from experimental coevolution set-ups (e.g., Andras et al., 2020; Ansari et al., 2017; Frickel et al., 2018; Retel, Kowallik, et al., 2019). This increasing availability of genomic data (e.g., Andras et al., 2020; Bartha, 2013; Frickel et al., 2018; Lees, 2019; Papkou et al., 2019; Retel, Kowallik, et al., 2019) offers a valuable source of information to address long-standing questions of host-parasite coevolution and to enhance our understanding of antagonistic interactions. These questions include uncovering the extent to which two species coevolve and the genetic underpinnings of coevolution, assessing the speed and duration of coevolution and its genomic consequences, and understanding links between ecological aspects of coevolution, spatial structure, and genomic (co)evolution. Note that although we mainly refer to host-pathogen examples, the outlined principles apply to any type of micro- or macroparasite or pathogen, and we thereafter use the term parasite generically.

In this review, we first aim to highlight the coevolutionary and noncoevolutionary processes driving the genome evolution of hosts and parasites from a theoretical population genetics perspective. This allows us to (a) describe the information available in host and parasite genomic data and (b) explicit the rationale for the analysis of these data. In doing so, we restrict ourselves to a population genomics perspective, and we do not address genome evolution under cospeciation (measured by cophylogeny), which results from long-term macroevolutionary processes (e.g., de Vienne et al., 2013). Second, we describe how different population genomic methods can foster our understanding of coevolution. We mainly focus on recently developed methods that allow for analysing host and parasite genomic data in a joint framework. Although, we focus on antagonist coevolutionary interactions, the general reasoning and outlined principles are also applicable to other interspecific interactions characterized by GxG interactions (e.g., driven by frequency-dependent selection) such as coevolutionary systems (mutualistic symbioses, plant-pollinators, or prey-predator interactions) or to host-microbial populations with conflicts between cooperator and cheater strains.

A substantial advantage of joint genomic analyses over the conventional single-species methods is to explicitly account for the reciprocal nature of coevolutionary interactions. Since the development of these methods is quite recent, we illustrate their potential applicability on two study systems on which single-species methods have been previously successfully applied to demonstrate coevolution: (a) the invertebrate host *Daphnia magna* and a variety of parasites, and (b) hosts of the plant genus *Silene* and several *Microbotryum* pathogens. These two host-parasite study systems are well understood with respect to their ecology, are amenable to laboratory experiments, and exhibit a wealth of genomic resources. We suggest that these two model systems are prime candidates for the application of the presented new joint host-parasite genomic methods and can be combined with empirical approaches such as polymorphism analysis, gene expression, and functional validation of candidate genes. We also discuss the current pitfalls and shortcomings of these new methods and some guidance for experimental design. We finally conclude by discussing the future developments needed to improve the analysis of host and parasite genomic data.

1.1 | Dynamics and characteristics of antagonistic coevolutionary interactions

Antagonistic coevolution results in reciprocal changes in the distribution of traits involved in the interaction (for example, resistance in host and infectivity in parasite) and the allele frequencies at the genes underlying these traits (Dawkins & Krebs, 1979; Janzen, 1980). These changes in trait/allele frequencies span a continuum ranging from so-called arms-race to trench-warfare dynamics. Under the arms-race dynamics (Bergelson et al., 2001; Dawkins & Krebs, 1979; Holub, 2001; Woolhouse et al., 2002) beneficial traits/alleles reach fixation, while under trench-warfare dynamics (Stahl et al., 1999; Woolhouse et al., 2002) several traits/alleles persist at intermediate frequencies over extended periods. The trench warfare dynamics are either characterized by persistent fluctuations of phenotype/allele frequencies (also called fluctuating selection dynamics) or their convergence to stable equilibrium values over time.

Deterministic mathematical models of coevolution have been used to generate predictions on the expected type of dynamics for hosts and parasites with various life-history traits (e.gAshby & Boots, 2017; Fenton et al., 2009; Haldane & Jayakar, 1963; Leonard, 1993; Tellier & Brown, 2007). However, the simultaneous action of stochastic processes within host and parasite populations may cause deviations from the predicted deterministic outcome (Gokhale et al., 2013; MacPherson et al., 2021; Schenk et al., 2020; Tellier et al., 2014). Stochastic fluctuations, such as genetic drift, can drive alleles maintained at intermediate frequencies to fixation by chance. Therefore, theoretical results suggest arms-race dynamics to occur

under a broader set of conditions than trench-warfare dynamics in natural populations (Gokhale et al., 2013; MacPherson et al., 2021; Schenk et al., 2020; Tellier et al., 2014).

In addition to reciprocal changes in trait distributions and allele frequencies, coevolution is also likely to cause population size changes in both species due to eco-evolutionary feedback resulting from epidemiological dynamics (Ashby et al., 2019; Gokhale et al., 2013; May & Anderson, 1983). The underlying principle is that allele frequency change (evolutionary change) affects host and parasite fitness and results in changes in population sizes (ecological change). These changes, in turn, impact the allelic fitness and frequencies and feeds back onto the evolutionary change (Ashby et al., 2019, Figure 1). Host-parasite eco-evolutionary feedbacks can result in correlated changes of host and parasite population size over time. These correlated population size changes are termed as the codemographic history (Živković et al., 2019). Furthermore, many host-parasite coevolutionary interactions extend beyond single host and parasite populations. Rather, single host and parasite populations are embedded in landscapes with varying extents of homogeneity and are connected to other populations by varying levels of gene flow (Thompson, 2005). The amount of gene flow determines the extent of synchrony of coevolutionary dynamics between populations (Sasaki et al., 2002; Tellier & Brown, 2011). In summary, host-parasite coevolutionary dynamics result in local changes in phenotypes/allele frequencies and population sizes (Figure 1) and have a temporal and a spatial component. Note that the coevolutionary dynamics and the codemographic history (change in population size) generally apply to other interspecies interactions with GxG interactions (e.g., example, driven by frequency or density-dependent selection). Still, the specific conditions for stability or fixation of trait/allele frequency depend on the system under consideration (prey-predator, plant-pollinator, etc).

1.2 | How do coevolutionary processes link to host and parasite genomic data?

From a coevolutionary perspective, host and parasite genomes can be conceptually partitioned into two loci categories, namely neutral (regarding coevolution) and coevolving loci (Figure 1). As opposed to neutral loci, coevolving loci contribute to the phenotypic traits which are under coevolutionary selection. Hence, coevolutionary selection shapes polymorphism patterns at the latter loci (see Figure 1). In hosts, most genes supposedly do not affect the outcome

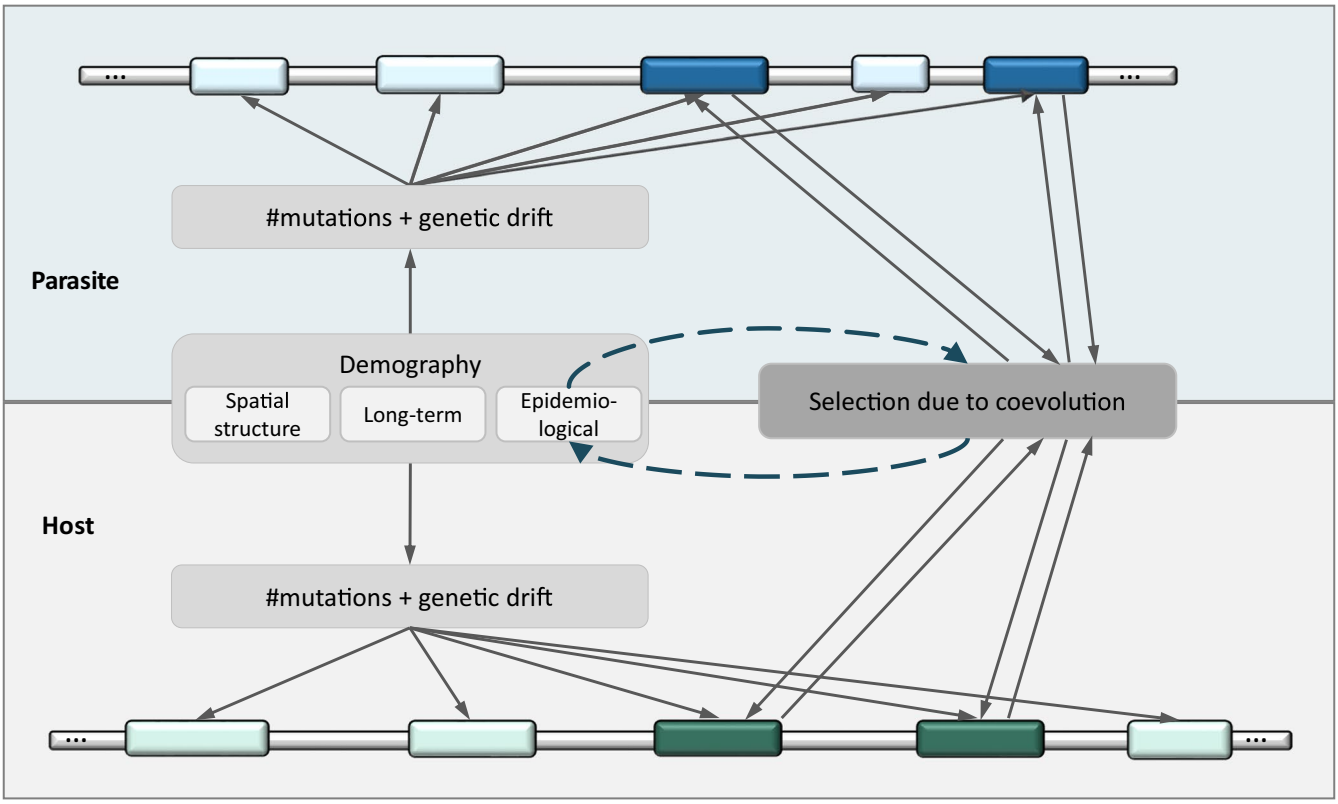

FIGURE 1 Schematic illustration of coevolution shaping genomic signatures in the host and the parasite. A part of the parasite (host) genome shown in grey on the top (bottom). Coloured boxes symbolize single loci/genes. Genes involved in the coevolutionary interaction (such as resistance or effector genes) are coloured in dark blue (parasite) and dark green (host). Neutral loci for the coevolutionary interaction are in light colours. Coevolution causes allele frequency changes at the interacting loci and shapes polymorphism patterns at and around the coevolving genes. Coevolution can also involve eco-evolutionary feedbacks due to epidemiological dynamics and hence, host and parasite population size changes. These population size changes, in turn, will affect levels of genetic drift and the population mutation rate, both affecting all loci in the genome simultaneously.

of coevolution, and, thus, their effect can be considered neutral regarding coevolution. In contrast, the number of parasite genes that are not involved in coevolutionary interactions can be small. This pattern should especially apply to obligate parasites which require an infected host to complete their lifecycle (Möller & Stukenbrock, 2017). Coevolving loci can be further classified based on their genetic contribution (major vs. minor) to a coevolving phenotype and whether they form the genetic basis of qualitative or quantitative traits. Coevolving loci in the host, include (a) major resistance genes defining host-parasite qualitative specificity and resulting in a strong defence response (e.g., nucleotide binding leucine-rich repeat receptors (NLRs) in plants; Stam et al., 2019; or the major histocompatibility complex (MHC) in animals Ansari, 2017), and (b) minor resistance genes (Poland et al., 2009) defining the quantitative strength of host resistance response (Figure 1). Coevolving loci in the parasite include (a) major genes defining parasite specificity and infectivity, such as effectors (Toruño et al., 2016), and (b) minor genes defining the quantitative effect of the infection on host fitness (so-called virulence in animal/human epidemiology or disease severity/aggressiveness in plant pathology) or disease transmission (Figure 1). For example, while being a crucial parasite quantitative trait, parasite transmission has its effect superseded by major genes such as effectors with respect to defining the outcome of the (GxG) interaction with host genotypes. It is noteworthy that additional abiotic and biotic factors can further shape signatures at coevolving loci. This relationship is especially true for crops in which resistance to parasites can be only an indirect outcome of breeder selection schemes (Brown, 2015).

Genome-wide stochastic processes such as genetic drift, mutation, recombination, and gene flow between populations (Figure 1) shape polymorphism patterns at both types (coevolving and noncoevolving) of loci. In particular, the amount of genetic drift depends on population size changes during the coevolutionary interaction. These population size changes can either directly result from coevolution itself (the codemographic history) or processes independent of coevolution, such as changes in the abiotic environment or range expansion. The type and speed of coevolutionary dynamics (how fast allele frequencies cycle or reach fixation) determines genetic polymorphism at the coevolving loci (Figure 1). It is expected that arms-race dynamics generates signatures of (recurrent) selective sweeps, whereas trench-warfare dynamics generates balancing selection signatures (Bergelson et al., 2001; Ebert & Fields, 2020; Holub, 2001; Woolhouse et al., 2002). Signatures of selective sweeps include locally decreased nucleotide diversity level, complex patterns of linkage disequilibrium around the site under selection, and an excess of low and high-frequency SNP variants relative to the genomic background (Stephan, 2019). Conversely, balancing selection signatures are characterized by elevated polymorphism levels, elevated linkage disequilibrium, and an excess of intermediate frequency variants (Charlesworth, 2006). However, Tellier et al., (2014) have shown that these predictions on resulting coevolutionary signatures are often too simplistic. First, genetic drift can broadly impact the signatures at the coevolving loci and render them indistinguishable from noncoevolving loci. Second, under trench-warfare dynamics, the genetic signatures at the involved loci depend on the allele frequencies at the expected theoretical equilibrium state and their proximity to frequency zero (loss) or frequency one (fixation) (Märkle & Tellier, 2020). Overall, trench-warfare dynamics might be relatively uncommon in natural populations, but if they occur and are long-lived, they should result in detectable polymorphism signatures of balancing selection. Moreover, polymorphism signatures of arms-race and trench-warfare are more likely to be detected in parasite than in host polymorphism data (Tellier et al., 2014). Selective sweeps and balancing selection at genes involved in coevolution have been extensively searched for in several host species (Ebert & Fields, 2020; references therein) and both signatures are found in the animal host species *Daphnia magna* (see Box 1), while only selective sweeps were found in plant-pathogen species of the genus *Microbotryum* (see Box 2).

In summary, the polymorphism patterns at different genes/loci (neutral vs. coevolving) in a genome are differentially affected by various selective and random processes acting during coevolution (Retel et al., 2019). This pattern suggests that ideally, the analysis of polymorphism patterns across different genes/loci should allow for assigning individual loci to either category (neutral or coevolving). While performing this assignation, we are specifically interested in answering three sets of coevolutionary questions: (a) What are the genetic determinants of coevolution? (b) What are the selective processes underlying coevolution? (c) Do eco-evo feedbacks occur, and what is the role of spatial structure in coevolution? So far, these questions have been mainly addressed by employing methods originally designed to analyse single species (only one of the antagonists). In the following, we describe recently developed methods that jointly analyse host and parasite polymorphism data, paving the way to answer these questions at finer resolution by explicitly taking the reciprocal nature of host-parasite interactions into account.

2 | RECENTLY DEVELOPED METHODS TO ANSWER COEVOLUTIONARY QUESTIONS

2.1 | Revealing the genetic underpinnings of coevolution

The first set of questions relates to uncovering the genetic underpinnings of coevolution, which includes unravelling the molecular basis of coevolution and counting the number of genes involved. Due to the relevance of these questions for medicine and plant and animal breeding in agriculture, the molecular mapping and study of the genes underlying host-parasite interactions have several decades of history and started with classic genetics (mutant screen) and quantitative genetic approaches (trait mapping). With the advance of next-generation sequencing methods, it has become possible to pinpoint the genes underlying resistance or infectivity phenotypes using

BOX 1 Evidence for coevolution in *Daphnia* and parasite species

The small crustaceans of the genus *Daphnia* and its parasites are a handy empirical model for understanding coevolutionary dynamics (Ebert, 2008). The species *D. magna* is host to a broad taxonomic range of pathogen systems, including microsporidians (*Ordospora colligata* and *Hamiltosporidium tvaerminnensis*; Haag et al., 2019), bacteria (*Pasteuria ramosa*; Bourgeois et al., 2017), and viruses (Daphnia Iridovirus-1; Toenshoff et al., 2018). Interestingly the spectrum of coevolutionary signatures, from trench warfare/balancing selection (*P. ramosa*) to arms race/selective sweeps (*H. tvaerminnensis*), have been discerned in the host's genome.

Routtu and Ebert (2015) used an F2 mapping panel to identify distinct quantitative trait loci (QTL) underlying resistance of *D. magna* to the microsporidian *H. tvaerminnensis*. Cabalzar et al. (2019) showed a distinct quantitative genetic signature of selection in the host arising from the coevolutionary process with *H. tvaerminnensis*. The long term coevolutionary process of microsporidia and host species shows a general pattern of genome reduction in the parasite (Wadi & Reinke, 2020; but see Haag et al., 2019), though published research has not yet localized a distinct genomic region associated with infectivity or virulence in the genome of these parasites.

In contrast to microsporidian parasites of *D. magna*, the sterilizing bacterial pathogen *P. ramosa* has yielded distinct signatures (trench warfare and arms race) of the coevolutionary process in both host and parasite genomes. Analysis of F2 mapping panels by Luijckx et al., (2013) and Bento et al., (2017) suggested a matching-allele model (MA) that determines the resistance of *D. magna* to *P. ramosa* infection. Routtu and Ebert (2015) identified a distinct QTL responsible for resistance to some strains of *P. ramosa*, and Bento et al., (2017) revealed that the genetic basis of host resistance is determined by a supergene region which varies in size from about 60 to 120 kbp, including annotation enrichment for fucosyltransferases and uncharacterized protein. more recently, Andras et al., (2020) applied a GWAS approach developed for bacterial genomes (Collins & Didelot, 2018), to identify a distinct *Pasteuria* collagen-like gene (PCL) responsible for infection of specific host genotypes. A GWAS by Ameline et al., (2021) and a QTL study by Bento et al., (2020) identified two novel additional, chromosomally distinct, genomic regions of the *D. magna* genome which determine resistance to other *P. ramosa* strains. Furthermore, while within chromosome epistatic interactions were already suggested in order to understand host resistance to *P. ramosa*, Ameline et al., (2021) and Bento et al., (2020) showed that among chromosomal interactions are involved in determining the resistance of *D. magna* to distinct strains of *P. ramosa*.

BOX 2 Coevolution at quantitative traits in the spatially structured *Silene* – *Microbotryum* system

Another well-documented host-parasite system is the species complex ofare anther-smut fungi, *Microbotryum* spp., castrating their *Silene* sp. hosts. Anther-smut fungi are highly specialized on their *Silene* host. There is accumulating evidence for a complex genomic basis of the *Microbotryum-Silene* coevolution involving many minor loci and noticeably the absence of major gene-for-gene loci. Genome scans of selective sweeps within populations of *M. lychnidis-dioicae* and *M. silenes-dioicae* detected a recent signature of positive selection (selective sweeps) in effector genes (Badouin et al., 2017). The number, localization, and presence/absence polymorphism of the detected genes differed among the two studied *Microbotryum* species, suggesting different coevolutionary dynamics intensities (Badouin et al., 2017). Gene expression analyses of *Microbotryum* strains *in planta* versus in vitro further revealed that those candidate genes are differentially expressed and probably involved in the specialization and coevolution. However, functional validation of these candidate genes is still missing. Genomic data at the population level for the *Silene* host are also still lacking, because of the large size of the *Silene* genome (2.8 Gb, Krasovec et al., 2018). However, extensive population genetics analyses using microsatellite markers of *S. latifolia*, *S. nutans* and *S. dioca* species provide a first glimpse into the coevolutionary processes in the genus (Feurtey et al., 2016; Hartmann et al., 2020; Martin et al., 2017). The recent combination of microsatellite data from population samples of both the host and *Microbotryum* species shows congruence of the population spatial genetic structure between the *Silene* and *Microbotryum*, even when correcting for isolation by distance (Feurtey et al., 2016).

Additionally, it is also of interest to understand the link between coevolution and specialization/divergence by comparing different plant hosts (*Silene* sp. or *Dianthus* sp.)-*Microbotryum* system. Whereas strict host specialization is often the rule on *Silene* species (Hartmann et al., 2020), *Microbotryum* shows broader and overlapping specialties on *Dianthus* hosts (Petit et al., 2017). Comparing patterns of coevolution in the two host genera and taking into account the host and the pathogen's demographic history, remain challenging but exciting avenues to unravel the genomic bases of coevolution in this multispecies system.

genome-wide association studies (GWAS). GWAS methods quantify the statistical correlation between genetic markers (biallelic single nucleotide polymorphism, SNPs) and a trait of interest (e.g., level of infection) in a sample of individuals. Several GWAS have revealed the genetic variation underlying resistance and disease susceptibility

in hosts (Bourgeois et al., 2017 in *Daphnia magna*; Nemri et al., 2010 in *Arabidopsis thaliana*) or infectivity in parasites (Saur et al., 2019 based on transcriptome data). Note that the system of *Daphnia magna* and its parasites has been particularly well studied (Box 1). However, MacPherson et al., (2018) showed that coevolutionary dynamics, namely, temporal changes in allele frequencies, alter the measured allelic effect sizes in single-species GWAS, thus decreasing their statistical power.

Two new types of GWAS take advantage of the increasing simultaneous availability of host and parasite genomic data. The overall idea is to perform an association study incorporating host and parasite single nucleotide polymorphism (SNP) data (obtained via genomic sequencing or transcriptomics data). The statistical regression is built either on measures of a phenotypic trait in experimental infections (we term this as experimental co-GWAS) (Figure 2a) or on observing parasite strains associated with a given host genotype (we term this as natural co-GWAS) (Figure 2b).

2.1.1 | New methods using joint genome analyses: Experimental co-GWAS

To overcome the shortcomings of single-species GWAS, MacPherson et al., (2018) laid down the foundations for a new type of GWAS that simultaneously incorporates host and parasite genetic information. The principle of the co-GWAS is to separate the variance in infection as the combination of parasite genotypes, host genotypes, and the GxG interactions. Along with the challenge of obtaining genetic information of both species, co-GWAS is computationally demanding as it involves a large number of pairwise comparison of host-parasite genome data. Using a rather simplified two-species co-GWAS scheme, MacPherson et al., (2018) could estimate allelic effect sizes that are robust to reciprocal changes in allele frequencies and thus, boost the amount of phenotypic variance explained compared to single-species GWAS. Wang et al., (2018) introduced a full co-GWAS method (analysis with a two-organism mixed model [ATOMM]),

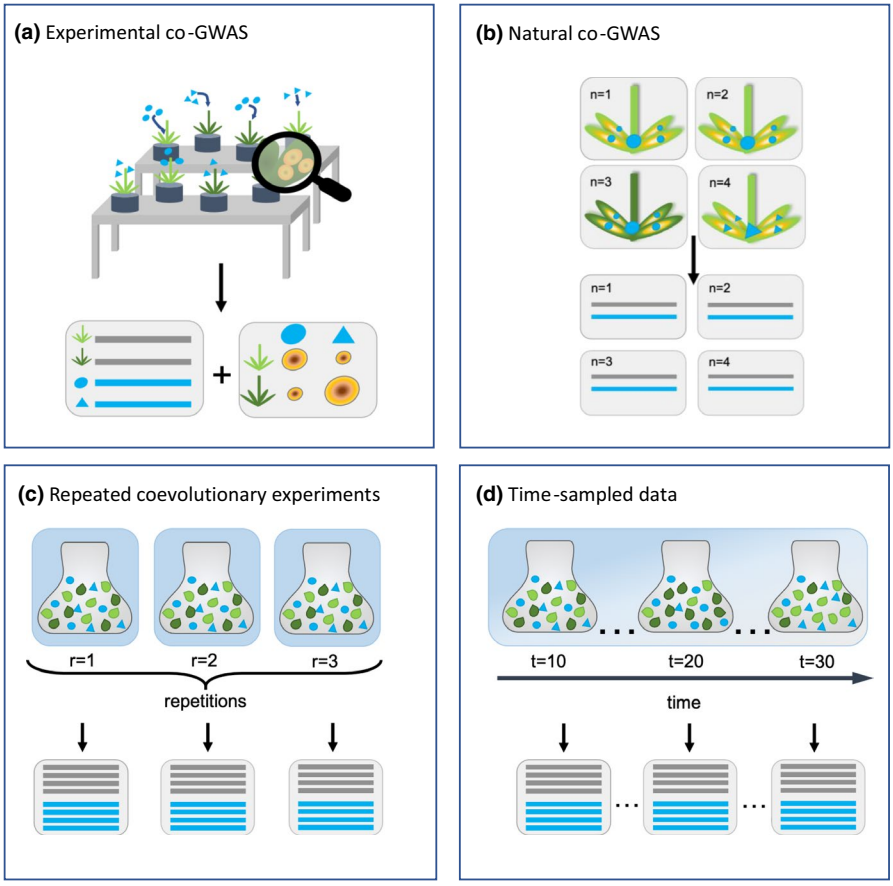

FIGURE 2 Different types of host-parasite coevolutionary experiments or natural study generating host and parasite genomic data. Single host genetic/genomic sequences are shown in grey, and parasite sequences in blue. Illustrations are for two different host phenotypes/genotypes (light green and dark green) and two different parasite phenotypes/genotypes (triangles and circles). (a) In an experimental co-GWAS several host individuals are experimentally infected with distinct parasite individuals in controlled conditions. For each infection, there is a measure of the resulting infection phenotype (such as the size of lesions, strength of hypersensitive response), and host and parasite individuals are sequenced. (b) In natural co-GWAS, host sequences and parasite sequences are obtained for infected host individuals and analysed in a pairwise manner (all pairs of host and parasite sequences). (c) Repeated laboratory coevolutionary experiments start from the same initial conditions, and several samples of hosts and parasite sequences are obtained from each repetition at a particular time point. (d) In a time-series coevolutionary experiment, coevolutionary interaction conducted in controlled conditions is followed over time, and host and parasite samples taken at a series of time points.

which takes the reciprocal nature of host-parasite interactions into account. The ATOMM method requires phenotypic data measured for each host-parasite pair in a controlled infection experiment and the corresponding genome-wide host and parasite genotype data (can also be used with a high-density SNP array). The co-GWAS uses a two-way mixed-effects model that incorporates the effects of host and parasite genetic variants and their interaction as fixed effects. To address population structure, three different genetic relatedness matrices are added as random effects to the model. These matrices include one for the host, one for the parasite, and one capturing the additive-by-additive polygenic interaction effects between the host and the parasite. The method, developed with bacterial pathogens in mind, can deal with different genetic variants, namely SNPs or presence/absence polymorphism. The method is also generalized for multi-allelic variants (Table 1). The ATOMM method is applicable to phenotypic traits that are multivariate normal or binomial-like traits (the trait can fall into one of several discrete categories). It is possible to either marginally map the trait to the host genome, the parasite genome, or perform an interaction test between host and pathogen variants. The authors tested their method by performing an experimental co-GWAS in the *Arabidopsis thaliana-Xanthomonas arbicola* plant-pathosystem. A total of 130 inbred *A. thaliana* accessions from the 1001 genomes project were infected with 22 *X. arbicola* strains and quantitative disease resistance was measured as a response trait. Wang et al., (2018) showed the existence of host-strain-specific quantitative resistance and a suggestive lack of broad-spectrum quantitative disease resistance against *X. arbicola*.

2.1.2 | New methods using joint genome analyses: Natural co-GWAS

Natural co-GWAS have been proposed to study samples from natural populations where it is not possible to perform genotype by genotype (GxG) infection experiments, but in which the sequence data of infected host individuals along with the infecting parasite strain can be collected simultaneously (such as data of infected human patients) (Bartha, 2013; for an illustration see Figure 2 in Bartoli & Roux, 2017). The underlying idea is that the experimental infection has been performed by nature, so to say. Computation of the statistical association for all pairs of biallelic host SNPs and biallelic parasite SNPs found in the samples follows. This association is analogous to an "interspecies linkage disequilibrium" (a concept first proposed in Fenton et al., 2009; and lately in Ebert & Fields, 2020). Note that we prefer the term "interspecies association" or "cross-species association" to avoid confusion with the intraspecies genomic linkage disequilibrium. Specifically, the co-GWAS methods compute the statistical association of each pair of host and parasite alleles, assuming the parasite allelic state as the logistic regression variable (Ansari, 2017). A graphical representation highlighting the significant SNP associations of such genome-to-genome comparisons can be found in Ansari (2017).

To our knowledge, three studies have been conducted so far on human hosts and the genomes of successfully infecting (a) HIV-1 (Bartha, 2013), (b) hepatitis C (Ansari, 2017), and (c) pneumococcal meningitis (Lees, 2019) strains. The pioneering study by Bartha (2013) combined sequences of 1,071 human genomes and the corresponding HIV-1 strains, aiming to map the host genetic pressure on the HIV-1 genome. As expected, the strongest association signal is obtained between human SNPs tagging HLA class 1 alleles and viral mutations in their corresponding CTL (cytotoxic T lymphocytes which kill infected cells) epitopes, thus validating the power of such a pairwise comparison analysis. Different regions previously unknown to have any impact on the interaction were also identified. Similarly, the association between 601 HCV-infected patients and the hepatitis C virus showed that the viral polymorphism is strongly associated with the immune system's human gene components (Ansari, 2017). Note, however, that viruses (such as HIV and HCV) have comparatively small genomes, and it becomes challenging to perform natural co-GWAS on bacterial genomes (such as *Streptococcus pneumoniae*, the causal agent of pneumococcal meningitis) because of a large number of comparisons which hinders the detection of significant associations after correcting for multiple testing (such as a Bonferroni correction). To overcome this challenge, Lees (2019) defined pneumococcal lineages and tested the association between pathogen lineages (transformed as a bi-state genotype) and given host genotype.

These studies of human diseases based on clinical samples demonstrate the power of natural co-GWAS to understand better the genomic factors that control infectious diseases and successfully locate previously unknown genomic regions associated with the disease outcome (as expected by Bartoli & Roux, 2017). These studies also highlight the statistical advantage of using parasite genomic variation as a response variable (in the GWAS statistical framework) to identify human genes of interest rather than conventional clinical outcome measures (resistance to infection, clinical presentation, disease progression, or death). The latter are complex phenotypic outcomes resulting from multiple interactions between multiple host genes and parasite genes influenced by other physiological processes. However, note that natural co-GWAS do not detect coevolution in the strict sense, but as all GWAS, detect statistical associations at the polymorphism level. Therefore, natural co-GWAS based on present-day data most probably reveal adaptation of viruses and bacteria to the current human population diversity as these parasites have a much shorter generation time than their human host. As a result, it remains unclear whether the identified human genes in these studies have evolved in response to the studied diseases (HIV, HCV, meningitis) or are polymorphic due to neutrality, diffuse coevolution, or other selective factors.

2.1.3 | New methods using joint genome analyses: Association indices

As an addition to natural co-GWAS, two cross-species association indices have been developed based on pairwise comparison of host and parasite SNPs, using randomly chosen infected (along with their

TABLE 1 Existing methods, the applicable data type, the sample size required (for host and parasite), advantages and weaknesses as well as open questions to date.

Method	Applicable to	Sample size needed	Advantages	Weaknesses	Remaining questions
Natural co-GWAs (Bartha et al., Ansari et al., Lees et al.)	Sample of infected host along with the associated parasite from natural populations. Possible to incorporate non-infected host data to the association statistic	Several hundreds of host and associated parasite samples.	Ideal for studying natural systems non-amenable to laboratory experiments	Not designed to study coevolution, but can indicate possible candidate loci. Identifiability of host and parasite individuals is required. Not easily applicable to pathogens with large genome size because of the huge number of SNP pair comparisons.	How to handle multi-strain (co) or (super) infections. How to account for population structure with data from several populations?
Experimental co-GWAs (Wang et al.)	Experimental infection of all possible host and parasite pairs in the lab.	On the order of 100 to few hundreds of host and parasite samples.	Ideal when host lines and parasite strains are defined and infection performed in the lab.	Requires large set of controlled infections. The phenotypic outcome should be accurately measurable.	How to design the best sampling scheme in spatially structured populations? What laboratory conditions should be chosen to perform the infection?
ABC inference (Märkle and Tellier)	Data from repeated coevolutionary experiments	>50 host and >50 parasite sequences at candidate genes per repetition.	Full genome data are not necessary (sequence capture of candidate genes are enough). Does not require any infection data.	Not easily applicable to hosts with long generation times. A priori knowledge on the infection matrix in necessary as an input in the ABC.	How to use data from several populations or time-series data? How to incorporate multi-locus interaction and polygenic traits?
Spatial correlations (Nuismer et al.)	Host-parasite pairs of loci across several populations	At least 30 populations and >100 host and >100 parasite samples per population.	Full genome data are not necessary (sequence capture of candidate genes are enough). Does not require any infection data.	False discovery rate increases with increased number of genes, populations need to be independent.	How to apply this method to species with non-recombining genomes
Long-term demographic history (Hecht et al.)	Host and parasite full genome data with a correct and contiguous assembly.	Few (2–10) full genomes	Very little *a priori* information on the past demography is required. Can highlight clear one-sided demographic patterns.	Several factors can confound the interpretation of the results. Only applicable to sexually reproducing species.	How to handle population structure? How to deal with complex demographic scenarios?

parasites) and noninfected hosts (Märkle et al., 2021) (here randomly implies an unbiased sample of the coevolving populations). These indices measure the degree of association between allele frequencies in the host (infected, noninfected and both types) and allele frequencies in the parasite samples, mirroring measures of linkage disequilibrium in population genetics. Therefore, the authors can derive the theoretical expectations for the distributions of these association indices based on the theoretically expected allele/site frequency spectrum from population genetics (while the expected statistical power is not easily predictable for co-GWAS). This study indicates that the power of association indices, and by extension of natural and experimental co-GWAS, to reveal genes under coevolution varies in time, depends on the type of coevolutionary dynamics, and is maximum when the alleles are at intermediate frequencies. Thus, coevolving loci are more likely to be detected under (long-term) trench-warfare than under the arms-race dynamics. Furthermore, obtaining time samples improves the statistical power of these indices in particular, and most likely of all co-GWAS methods, as they capture the temporal fluctuations in allele frequencies at the coevolutionary loci (Märkle et al., 2021). The association indices do not replace co-GWAS to pinpoint loci under coevolution, but allow to link the results of co-GWAS with the theoretical models of coevolution and thus could be used for statistical inference of coevolutionary parameters at significant pairs of associated SNPs found in co-GWAS (Märkle et al., 2021).

These results have implications for the experimental design of co-GWAS. As the genes underpinning coevolution revealed by co-GWAS are more likely to be those under trench warfare dynamics, it is desirable to study host-parasite systems where the effect of genetic drift is relatively weak (Tellier et al., 2014). We, thus, advise the construction of a panel of samples from the whole species range rather than small local populations (e.g., the 130 *A. thaliana* sample panel, Wang et al., 2018). The rationale is that if coevolution is pervasive in space, trench warfare dynamics are more likely to occur and be long-term in a large spatially structured population with a large effective population size. Conversely, revealing genes under arms-race dynamics requires one to obtain samples from several independent populations that are asynchronous in their coevolutionary process (e.g., Haag et al., 2019). Further, it is crucial to keep in mind that co-GWAS can only pinpoint genes under coevolution if they are polymorphic, a state which is only transiently observable under arms race dynamics. Finally, we also point out that for parasites with large genomes (bacteria, *Plasmodium*, fungi), there is a need to reduce the dimensionality of the data set by defining lineages based on relatedness to avoid high false-negative rates (due to corrections for multiple testing). However, defining parasite lineages becomes probably problematic for sexually reproducing parasites (*Plasmodium*, some fungi) due to intragenomic recombination. In the latter case, we suggest regrouping parasite strains based on their genetic relatedness as measured by population genomic clustering methods (e.g., PCA, Structure, Admixture, DAPC). Finally, we note that the sample sizes in the natural co-GWAS performed on human-parasite systems (Ansari, 2017; Bartha, 2013; Lees et al., 2019) are large (several thousand) and obtained at a single point in time. However, Märkle et al., (2021) predict that few hundreds of samples at one-time point (on the order of sample sizes in Wang et al., 2018) might be enough to obtain adequate statistical power. It is also predicted that the statistical power of co-GWAS would be increased if samples are available at several time points. This expectation implies that natural and experimental co-GWAS approaches are not mutually exclusive, and for ad hoc systems (e.g., *Daphnia magna*), it would be highly interesting to perform both simultaneously to assess the influence of environmental variation on the outcome of coevolution. However, as it may be difficult for many natural systems to obtain many samples, an alternative would be to obtain hundreds of samples widespread in space and time (over several years). We note that such extensive sampling is short to be readily available for the host-parasite systems presented in Boxes 1 and 2. Using an experimental co-GWAS approach, we suggest studying crop coevolution with pathogens by using comparisons between multiple crop varieties and multiple parasite strains obtained across different locations and across different years.

2.1.4 | Inferring the selection pressure underlying coevolution

The second group of questions is related to the speed, timing, and attributes of coevolution. Here, the aim is to understand (a) the type of coevolutionary dynamics at the different genes (trench-warfare, arms-race), (b) which species is ahead in the coevolutionary dynamics, (c) if coevolution between a pair of species is strict (i.e., one host and one parasite) or diffuse (i.e., including several species), and (d) since when the two species have been coevolving. Only a few methods exist to answer these questions as they require statistical inference of the coevolutionary selective pressures at the coevolutionary loci.

Population genomics studies aim to (a) scan the genomes for genes under positive or balancing selection, and (b) draw a statistical inference of the past reciprocal selective history. Genome scan methods, applied to either host or parasite data (Ebert & Fields, 2020 and references therein), can be used to detect genes with statistically significant signatures of selection compared to the genomic average while accounting for the population's past demographic history (Stephan, 2019). These methods usually make use of any of the following properties: the distribution of SNPs (frequency-wise and spatially along the sequence), patterns of linkage disequilibrium, the number of substitutions compared to an outgroup species, and the ratio of nonsynonymous to synonymous substitutions. However, variation in recombination rates along the genome, confounding effects of past demography (e.g., severe bottlenecks, admixture, population expansion, etc), or strong genetic drift can limit the detection of selection signatures (e.g., Stephan, 2019). These factors may explain the fact that, despite genome scans for positive or balancing selection being conducted chiefly on wild host or wild parasite species (Boxes 1 and 2), on crops and their pathogens or humans and

their parasites, there are still few documented and demonstrated host-parasite pairs of truly coevolving genes (with complementary gene expression and functional studies of the candidate genes, see reviews in Ebert & Fields, 2020; Möller & Stukenbrock, 2017; Petit-Houdenot & Fudal, 2017).

2.1.5 | New methods using joint genome analyses: Inference of coevolution at coevolving loci

To go beyond mere genome scans for selection and to make use of methods designed to infer the amount of selective pressure, Märkle and Tellier (2020) assessed the amount of information on the parameters underlying the coevolutionary interaction that can be retrieved from host and parasite polymorphism data by means of approximate Bayesian computation (ABC). ABC is a computational method using so-called summary statistics of the data set (observed and simulated) to estimate posterior probabilities of models or parameters of interest when likelihood calculations are intractable (Csilléry et al., 2010; Sunnåker et al., 2013). Therefore, an extensive number of simulations are run with the parameter values of the underlying model sampled from prior distributions, which summarize the current (prior) knowledge. A rejection step retains all simulations for which the summary statistics fall within a certain distance to the observed data's summary statistics. The retained simulations are used to estimate a posterior probability for each competing model (model choice) or generate a posterior distribution for the parameters of interest (Csilléry et al., 2010; Sunnåker et al., 2013).

The ABC in Märkle and Tellier (2020) uses polymorphism data of repeated host-parasite coevolutionary experiments (Figure 2c) to account for the effect of genetic drift on the resulting genetic signatures. It assumes that sequences or SNP data of a sample of *n* hosts and *n* parasites are obtained for each of *r*-repetitions of a coevolutionary experiment (resulting in *r* x *n* samples for each species). Population genetics statistics (e.g., based on the site frequency spectrum (SFS), etc) are calculated separately for each repetition of the host and parasite sequences and then averaged over *r*-repetitions. These averaged population genetics statistics constitute the set of summary statistics used in the ABC.

The ABC model choice is used to distinguish pairs of coevolving host and parasite loci from pairs of neutral (noncoevolving) loci. Some scenarios have high model choice accuracy, which decreases when either the equilibrium frequencies under trench-warfare dynamics are too close to the boundaries (i.e., allele frequency of zero or one) or when alleles get rapidly fixed under arms-race dynamics. In a second step, inference of parameters of the past coevolutionary history (fitness costs) takes place. The best inference results when jointly using host and parasite summary statistics rather than only using either of those. This result is consistent with host allele frequencies and hence, the resulting sequence data, being informative on the parasite parameters and vice-versa (Märkle & Tellier, 2020; Tellier & Brown, 2007; Tellier et al., 2014). In general, the parameter estimations are more accurate when coevolution follows trench-warfare dynamics and when data from more repetitions are available (a minimum of 10). This dependency highlights the need to account for genetic drift when developing new methods to analyse host-parasite coevolution. Sample sizes of at least $n = 50$ for each species are advisable. Finally, identifiability issues arise when distinct parameter combinations result in similar coevolutionary dynamics, and hence, the inference accuracy decreases.

In summary, the study by Märkle and Tellier (2020) demonstrates the potential to infer information on the past coevolutionary history (fitness costs) by jointly using host and parasite polymorphism data at the coevolving loci in an ABC framework. This method could be applied to sets of host and parasite candidate genes obtained by sequence capture (e.g., Stam et al., 2019 for NLRs in tomato) and opens the door for further methodological developments. Yet, the method is so far only tested for two theoretical models of coevolution at major genes under the simplifying assumptions of constant host and parasite population sizes, assuming data from repeated coevolution experiments (e.g., microcosm experiments) and that the same loci drive the interaction across all repetitions. Due to these theoretical shortcomings, the approach may not yet be applicable to many host-parasite systems. The method is currently further developed to infer coevolution across multiple (more or less synchronized) populations instead of using multiple repetitions. It may be possible in the future to apply an extended version of this joint inference method to full genome data of *D. magna* and its parasites and on the *Microbotryum* – *Silene* systems for which the spatial structure is known, and several populations have been sampled, and their genomes are available (Boxes 1 and 2). Finally, current inference approaches can be improved and made more widely applicable by recent progress in population genomics, such as more efficient and flexible forward-in-time simulators (e.g., Haller & Messer, 2019) and machine learning methods (e.g., Sanchez et al., 2020).

2.1.6 | Further applications: Towards joint genome analyses in experimental coevolution studies

Genome-wide data of hosts and parasites can also be obtained from controlled laboratory coevolutionary experiments of systems with very short generation times (bacteria and phage, *C. elegans* and virus, algae, and microparasite) (Frickel et al., 2018; Hall et al., 2011; Papkou et al., 2019; Retel, Kowallik, et al., 2019). Sequencing of host individuals and parasite strains occurs at the beginning, during, and at the end of the coevolutionary experiment, which consists of repeatedly following the evolution of hosts exposed to the parasite and hosts evolving without parasite pressure (Figure 2c). These studies aim to decipher the genetic bases of coevolution. The power of these studies depends on the length of the experiments (how many generations) but also and chiefly on the (a) amount of genetic drift, (b) mutation rate, (c) recombination rate, and (d) host and parasite genetic diversity present at the onset of the experiment. Similarly to a GWAS, genes potentially underlying coevolution are those with aberrant SNP/allele frequency changes when coevolutionary

replicates are compared to host-only replicates (Frickel et al., 2018; Papkou et al., 2019; Retel, Kowallik, et al., 2019). Large host and parasite population sizes, which can arise if recurrent bottlenecks due to the in vitro multiplication rounds and eco-evo feedbacks are not too severe, can favour selection over genetic drift and the occurrence of mutations upon which selection can occur (Frickel et al., 2018; Hall et al., 2011; Retel, Kowallik, et al., 2019). On the other hand, for species with limited population sizes and mutation rates, it is more efficient to start the experiment with large genetic diversity (Papkou et al., 2019). Thus far, analysis of host and parasite genome data are usually performed independently, so there is no specific statistical inference of the coevolutionary process. In addition to replicates, these controlled coevolutionary experiments also provide time samples of the coevolutionary process (Figure 2c and 2d). Hence, natural co-GWAS and association indices can, in principle, be applied to experimental coevolution studies on individuals at the end of the experiment. Nonetheless, it is often difficult in such fast-evolving and microscopic organisms to clearly define and isolate infected or noninfected single host individuals and the corresponding parasite strain (Hall et al., 2011; Retel, Kowallik, et al., 2019). We, therefore, speculate that natural and experimental co-GWAS would more likely be possible in the *C. elegans* system (Papkou et al., 2019), where it is possible to identify infected individuals, though the host sample size may be too small.

2.2 | Eco-evo feedbacks and spatial structure: Insights from genomics

All previous methods and theories are built on the simplifying assumptions of constant population sizes and a host-parasite pair coevolving in a single population. The third group of questions links ecological aspects of coevolution and spatial structure of host and parasite populations with their genomic evolution. We would like to (a) know if eco-evo feedbacks and codemographic history are observable in the genome data, (b) reveal if coevolution is spatially confined to single populations or occurs over the whole species ranges, (c) assess if a species is locally adapted/maladapted to its antagonist, and (d) understand the influence of spatial heterogeneity on coevolution. Answering these questions requires intensive sampling of whole-genome data in space and time (time samples, see Figure 2d).

2.2.1 | New methods using joint genome analyses: Inference of codemographic history

An interesting characteristic of many strict host-parasite interactions is the parasite's dependency on the host for its reproduction, constraining the parasite population size. Therefore, host and parasite population size changes in time may not be fully independent. Thus, inferring historical changes in population size may contain information on the coevolutionary process. We define this correlation in population sizes as the codemographic history (Živković et al., 2019). Codemographic changes can occur on two different time scales (long-term vs. short-term) and be due to two distinct mechanisms (one-sided vs. reciprocal codemographic history) (Figure 3a and Figure 3b). First, long-term changes (thousands of generations) in host population size possibly drive changes in the parasite population size (Hecht et al., 2018) (Figure 3a). Long-term correlations in population sizes can be, for example, due to the expansion of the host population, for example, when colonizing new habitats. This correlation can result in a subsequent expansion of the parasite (as speculated in Hecht et al., 2018). Thus, this type of change can be defined as one-sided as changes in host population size initiate them. Second, eco-evolutionary feedback due to epidemiological dynamics potentially generates short-term changes (few hundreds of generations) in host and parasite population sizes, termed as the reciprocal codemographic history in Živković

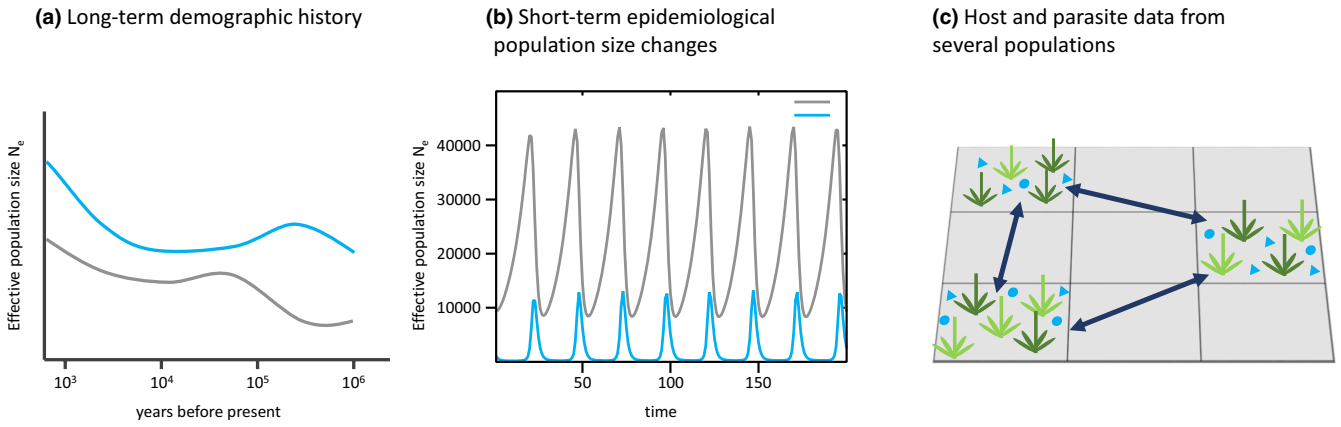

FIGURE 3 Codemographic history and spatial structure as fundamental mechanisms resulting from or shaping the coevolutionary history. (a) The correlation between the long-term demographic history of the host (grey) and parasite (blue) can be indicative of a tight interaction between the host and the parasite (one-sided codemography). (b) Short-term population size change due to eco-evolutionary feedback can result in detectable signatures in the genome wide SFS (reciprocal codemography). (c) Data from several populations connected by gene flow can elucidate host-parasite coevolution's spatial context.

et al., (2019) (Figure 3b). Recent studies show that correlated host and parasite population size changes on both timescales can be inferred based on full-genome SNP data (Hecht et al., 2018; Živković et al., 2019) (Figure 1).

Hecht et al., (2018) developed an add-on to inference methods based on the sequentially Markovian coalescent (SMC) to detect correlations between host and parasite population sizes due to the long-term one-side codemographic history. SMC methods are used to reconstruct genealogies along the genome and infer the past demographic history (Li & Durbin, 2011; Sellinger et al., 2020). The genealogies and transitions between genealogies along the genome due to recombination allow for estimating the hidden states of the model, which are the piecewise constant population sizes in time. Hecht et al., (2018) tested for correlations between these size estimates independently obtained for the host and the parasite populations (Figure 3a). However, one can argue that this approach is presumably too simplistic. Even if hosts increase their range, the habitat may not be suitable for pathogens, so pathogen expansion is uncertain. One further key issue is that the estimate of effective population size in time-based on SMC methods depends on many factors such as population structure, changes in the reproduction mode, and habitat variation. Effective host population size is further affected by changes in host selfing-rate or seed banking (Sellinger et al., 2020) which do not necessarily change host availability for the parasite. In this regard, the method of Hecht et al., (2018) is incomplete and should be further investigated and tested for more complex scenarios of one-sided codemographic history.

Population size changes resulting from the reciprocal codemographic history (due to eco-evolutionary feedback) are too fast to be captured by a coalescent approach. However, Živković et al., (2019) have shown that changes in the genome-wide distribution of allele (SNP) frequencies between time samples during coevolutionary dynamics can be informative on host and parasite population size changes. They also demonstrated that the changes in the SFS are only observable if the population sizes are small enough at the start of the coevolutionary dynamics. In most eco-evo host-parasite coevolutionary models (Ashby & Boots, 2017; Ashby et al., 2019; Gokhale et al., 2013; Živković et al., 2019), the parasite population exhibits more severe size variations over time (recurrent bottlenecks) than the host population. Such drastic bottlenecks can be detected by changes in the SFS over time. The resolution and statistical power depend on the number of parasite generations per host and the number of parasites per host individual (Živković et al., 2019). The genome data across the different time samples in the experimental coevolution study of Retel, Kowallik, et al., (2019) seem to match these predictions. Currently, the required host and parasite time sample data are only available from experimental coevolution systems. However, sequencing hosts and their parasites over time in short-lived plants and invertebrates should also be feasible and help to document the occurrence of eco-evo feedbacks in these systems. Especially, the *Daphnia* system could be used for such a study as recovery and sequencing of time samples from sediments for host and parasites is possible (Decaestecker et al., 2007) (Box 1).

An apparent limitation of these demographic inference approaches (Figure 1) is their dependence on recombination in the analysed host and parasite genomes, generating multiple coalescent genealogies along the sequence. Thus, it may most likely not be possible to obtain independent estimates of the co-demographic history and the underlying selection pressures for asexual viruses, bacteria, or fungi. In asexual organisms, we conjecture that current methods based on phylogeny (e.g., estimation of the reproductive ratio or expansion of the parasite population within an epidemic, Stadler et al., 2012) may only have limited power to infer coevolutionary histories.

2.2.2 | Spatial correlation of host and parasite allele frequencies

Spatial structure is important in host-parasite interactions (Figure 3c). A large body of theoretical models has been built to predict which antagonist is ahead in the coevolutionary interaction (i.e., local adapted) depending on local population sizes and gene flow levels. Both host and parasite spatial structure can be inferred from full genome data (the neutral loci) and compared to one another (Feurtey et al., 2016, Box 2). Moreover, given sufficiently low levels of gene flow, data from single coevolving populations can be considered independent coevolutionary replicates for inference.

Nuimser et al., (2017) proposed a method to identify coevolving loci by testing for significant spatial covariation between host and parasite genetic marker frequencies. The method has the premise that if host-parasite local adaptation is detectable in reciprocal cross-infection, it should be due to spatially covarying allele frequencies at the functional genetic loci. The method starts by calculating allele frequencies for a set of host and parasite genetic markers for each population within a metapopulation. Based on the marker frequencies, the spatial covariance is calculated for each pair of host-parasite markers (e.g., biallelic SNPs). The resulting covariance matrix is converted into a correlation matrix which is screened for significant correlations using a Student's *t* test. To test their method, the authors simulated data under different types of matching-alleles models (1–3 loci, with epistatic or additive effects), varying degrees of local adaptation, and varying the number of populations sampled. Therefore, a coevolving locus was randomly placed in a set of neutral markers. Overall, the method exhibits a very high statistical power if local adaptation is strong and allele frequency estimates are available for more than 30 populations (while they consider only 100 loci and large sample sizes of >100 per population). Note that if full genome data (many more than 100 loci) are to be used, the effect of multiple-testing and the resulting FDR should be assessed for the given sample sizes. Furthermore, the method's strength substantially drops when local adaptation is weak, the number of populations is small (<20), and more than one locus is involved in the interaction. We thus recommend using this method if strong local

adaptation is demonstrated a priori in a reciprocal cross-infection experiment, a large number of independent populations can be sampled (>30 populations), and coevolution depends on very few loci with a large phenotypic effect. Note as well that the method relies on three crucial assumptions (as also discussed by Nuismer et al., 2017). First, populations must be independent of one another, which implies that gene flow levels among populations must be reasonably small. Practically, the rate of gene flow can be measured using full genome data and classic population genetics methods. Second, the marker distribution along the genome must be sufficiently dense (e.g., full genome data or high-density SNP array) to pinpoint the loci under coevolution. Third, markers must be independent of one another, which requires a sufficiently high recombination rate. Further, the method can result in false positives if the covariation of host and parasite marker frequencies is due to the adaptation of both species to the shared environment (Nuismer et al., 2017). By sequencing full genomes from several individuals across several populations, the *Daphnia-parasites* and *Silene-Microbotryum* systems could be used to test and further develop these methods.

3 | CONCLUSION AND OUTLOOK

We have summarized recently developed methods to jointly analyse host and parasite genomic data and, thus, to account for the reciprocity of these interactions explicitly. These methods provide promising avenues to extend our understanding of multiple aspects of host-parasite coevolution. The development of these methods is often based on a specific type of data and specific (and sometimes quite simplistic) assumptions regarding (a) the nature of the coevolutionary interaction, (b) the life-history of the interacting species, and (c) the underlying genomic architecture. Yet, every host-parasite system is unique in terms of life-history traits, the genomic features (such as genome size, recombination rates, number of loci involved, etc), the extent of host and parasite population structure, its experimental tractability, and the feasibility of obtaining particular types of genomic data (cost-wise, time samples, spatial samples, samples from replicated experiments). Table 1 summarizes current limitations, sample size requirements, and potential extensions of the described methods. In their current state, some of these methods might be only applicable to a limited number of host-parasite systems, and there is a need for further theoretical and statistical developments, as well as to optimize the combination of existing methods (gene expression, functional studies, experimental coevolution, etc). We advocate using several empirical and experimental approaches simultaneously to study coevolution, as the observed genomic patterns can result from different coevolutionary and noncoevolutionary processes. We end this review by (a) suggesting some theoretical and methodological developments for the near future and (b) highlighting three main challenges and unanswered questions regarding coevolution that require longer-term theoretical and empirical work.

3.1 | Further directions for theoretical/methodological developments

Theoretical and empirical analyses (see Box 1 and 2) have focused on revealing genes with a major contribution to the phenotypic traits underlying host-parasite interactions. To expand our understanding of coevolution due to minor quantitative genes or combinations of major genes with epistatic interactions, it is required to develop more sophisticated methods to leverage additional information hidden in population genomic data (i.e., at neutral and non-neutral loci, Figure 1).

First, from a theoretical point of view, it is desirable to improve our understanding of the effect of eco-evolutionary feedback on genomic signatures at both coevolving and neutral loci. The study by Živković et al., (2019) shows the potential of indirectly observing short-term coevolutionary dynamics by observing changes in the genome-wide neutral host and parasite site frequency spectra resulting from corresponding host and parasite population size changes. Yet, the effect of eco-evo feedbacks on genetic signatures at the coevolving loci is, to our knowledge, still poorly understood. Furthermore, we still lack a systematic investigation of the effect of different types of host-parasite coevolutionary interactions and dynamics on polymorphism data and how they relate to optimal sampling schemes (replicates, spatial and temporal number of time points, and the number of samples). It is also required to assess how host and parasite specificities affect the power and accuracy of the joint analysis methods, and more specifically, to quantify to which extent GxG specificity (and possibly epistasis) is inferable from host and parasite genomic data. The two well-understood systems of host-parasite coevolution we highlight in Boxes 1 and 2 make them the primary test candidates for these new developments.

Second, the inference of the past coevolutionary history from coevolving loci should be developed further. Märkle and Tellier (2020) have only scratched the surface of the realm of possibilities. The assumption of only a few major loci controlling coevolution could be very strong and possibly invalid for many host-parasite pairs (see Box 2). We suggest developing inference methods specifically designed to consider more realistic scenarios of coevolution with (a) quantitative trait loci, (b) clusters of resistance genes, (c) networks of interacting genes in the host and the parasite, and (d) diffuse coevolution between host and parasite communities. These developments may greatly benefit from recent advances in the field of host and parasite pangenomics (e.g., van der Weyer, 2019), integration of gene network structure into the analysis framework, advances in population genomic machine learning methods (e.g., Sanchez et al., 2020), more efficient and flexible forward-in-time simulators (e.g., Haller & Messer, 2019), and the integration of metagenomic samples in time and space (e.g., Toju et al., 2017). Furthermore, additional prior information such as the presence/absence and function of parasite effectors could provide additional power. Ultimately, these methods should be made accessible to a broader community by integrating them into user-friendly interfaces.

Third, from a theoretical and methodological point of view, developments are needed to disentangle the confounding effects of different phenomena, resulting in similar genomic signatures (identifiability issue). For example, signatures similar to those of coevolution can be the result of seasonal or fluctuating selection. Distinguishing these two scenarios requires developing methods that can correct for the impact of known demography and life-history traits (Sellinger et al., 2020). Another major challenge is identifying neutral and coevolving loci in nonrecombining hosts and parasites (viruses, bacteria, fungi, and asexual aphids). This development is crucial in understanding the pathogen evolution as most viral or bacterial parasites have a small genome with strong linkage. For the *Daphnia* system (Box 1), the availability of asexual and sexual parasite sequenced genomes can guide the development of new methods tailored for lack of recombination.

3.2 | Remaining unanswered questions

One of the main challenges remains to test whether two species are genuinely coevolving. Several studies have pointed out that correlations between host and parasite traits are not necessarily due to coevolution, and coevolution does not necessarily result in detectable correlations (Janzen, 1980; Nuismer et al., 2010). It is desirable to use joint analysis of genome data from hosts and parasites to distinguish strict/diffuse coevolution from unilateral evolution (Nuismer & Week, 2019), comparing different populations under various coevolutionary pressures. A second far-reaching question is to what extent epigenetics, especially genome methylation, affect coevolution or if methylation can be the underlying coevolution mechanism. Models of methylation and epigenetics developed in the context of species adaptation and the corresponding population genetics results and methods are readily available (Vidalis et al., 2016) and could be potentially adapted to host-parasite interactions. Finally, both anatagonists are usually part of a broader community, and thus, most interactions involve multiple hosts and parasites. Sequencing the composition and genetic diversity of microbes (the microbiome) or host (plant, animal) communities is becoming feasible via metagenomics. Consequently, it would be desirable to (a) improve our understanding of host-parasite coevolution in a community context and generate predictions on their genomic effects, and (b) correspondingly extend existing inference methods to accommodate for multi-species interactions.

ACKNOWLEDGEMENTS

Hanna Märkle and Sona John were funded by the Deutsche Forschungsgemeinschaft grant TE809/3 (project 274542535) to Aurélien Tellier within the SPP 1819 "Rapid evolutionary adaptation: Potential and constraints". Hanna Märkle is funded by a Victor Dropkin postdoctoral fellowship. Amandine Cornille was funded by the ATIP-Avenir program. Peter D. Fields is funded by a grant from the Swiss National Science Foundation. Amandine Cornille and Aurélien Tellier are members of the STN "Ecological genomics of coevolutionary interactions" funded by the European Society for Evolutionary Biology (ESEB). Open access funding enabled and organized by Projekt DEAL.

ORCID

Hanna Märkle https://orcid.org/0000-0002-4686-4854
Sona John https://orcid.org/0000-0002-2827-0008
Amandine Cornille https://orcid.org/0000-0002-5348-7081
Peter D. Fields https://orcid.org/0000-0003-2959-2524
Aurélien Tellier https://orcid.org/0000-0002-8895-0785

REFERENCES

Ameline, C., Bourgeois, Y., Vögtli, F., Savola, E., Andras, J., Engelstädter, J., & Ebert, D. (2021). A two-locus system with strong epistasis underlies rapid parasite-mediated evolution of host resistance. *Molecular Biology and Evolution*, *M38*(4), 1512–1538. https://doi.org/10.1093/molbev/msaa311

Andras, J. P., Fields, P. D., Du Pasquier, L., Fredericksen, M., & Ebert, D. (2020). Genome-wide association analysis identifies a genetic basis of infectivity in a model bacterial pathogen. *Molecular Biology and Evolution*, *37*, 3439–3452. https://doi.org/10.1093/molbev/msaa173

Ansari, M. A., Pedergnana, V., L C Ip, C., Magri, A., Von Delft, A., Bonsall, D., Chaturvedi, N., Bartha, I., Smith, D., Nicholson, G., McVean, G., Trebes, A., Piazza, P., Fellay, J., Cooke, G., Foster, G. R., Hudson, E., McLauchlan, J., Simmonds, P., … Spencer, C. C. A. (2017). 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. https://doi.org/10.1038/ng.3835

Ashby, B., & Boots, M. (2017). Multi-mode fluctuating selection in host–parasite coevolution. *Ecology Letters*, *20*, 357–365. https://doi.org/10.1111/ele.12734

Ashby, B., Iritani, R., Best, A., White, A., & Boots, M. (2019). Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution. *Journal of Theoretical Biology*, *464*, 115–125. https://doi.org/10.1016/j.jtbi.2018.12.031

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., Le Prieur, S., Jeziorski, C., Branca, A., & Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, *26*, 2041–2062.

Bartha, I., Carlson, J. M., Brumme, C. J., McLaren, P. J., Brumme, Z. L., John, M., Haas, D. W., Martinez-Picado, J., Dalmau, J., López-Galíndez, C., Casado, C., Rauch, A., Günthard, H. F., Bernasconi, E., Vernazza, P., Klimkait, T., Yerly, S., O'Brien, S. J., & Listgarten, J. … Fellay, J. (2013). A genome-to-genome Analysis of Associations between Human Genetic Variation, HIV-1 sequence diversity, and viral control, *eLife*, *2*, e01123.

Bartoli, C., & Roux, F. (2017). Genome-wide association studies in plant pathosystems: Toward an ecological genomics approach. *Frontiers in Plant Science*, *8*, 763. https://doi.org/10.3389/fpls.2017.00763

Bento, G., Fields, P. D., Duneau, D., & Ebert, D. (2020). An alternative route of bacterial infection associated with a novel resistance locus in the *Daphnia-Pasteuria* host-parasite system. *Heredity*, *125*, 173–183. https://doi.org/10.1038/s41437-020-0332-x

Bento, G., Routtu, J., Fields, P. D., Bourgeois, Y., Du Pasquier, L., & Ebert, D. (2017). The genetic basis of resistance and matching-allele interactions of a host-parasite system: The *Daphnia magna-Pasteuria ramosa* model. *PLoS Genetics*, *13*, e1006596. https://doi.org/10.1371/journal.pgen.1006596

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

Bourgeois, Y., Roulin, A. C., Müller, K., & Ebert, D. (2017). Parasitism drives host genome evolution: Insights from the *Pasteuria ramosa* - *Daphnia magna* system: Brief Communication. *Evolution*, *71*, 1106–1113.

Brown, J. K. M. (2015). Durable resistance of crops to disease: A Darwinian perspective. *Annual Review of Phytopathology*, *53*, 513–539.

Cabalzar, A. P., Fields, P. D., Kato, Y., Watanabe, H., & Ebert, D. (2019). Parasite-mediated selection in a natural metapopulation of Daphnia magna. *Molecular Ecology*, *28*, 4770–4785. https://doi.org/10.1111/mec.15260

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*, e64. https://doi.org/10.1371/journal.pgen.0020064

Collins, C., & Didelot, X. (2018). A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Computational Biology*, *14*, e1005958. https://doi.org/10.1371/journal.pcbi.1005958

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*, 410–418. https://doi.org/10.1016/j.tree.2010.04.001

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society B*, *205*, 489–511.

De Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, *198*(2), 347–385. https://doi.org/10.1111/nph.12150

Decaestecker, E., Gaba, S., Raeymaekers, J. A. M., Stoks, R., Van Kerckhoven, L., Ebert, D., & De Meester, L. (2007). Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, *450*, 870–873. https://doi.org/10.1038/nature06291

Ebert, D. (2008). Host-parasite coevolution: Insights from the Daphnia-parasite model system. *Current Opinion in Microbiology*, *11*, 290–301. https://doi.org/10.1016/j.mib.2008.05.012

Ebert, D., & Fields, P. D. (2020). Host-parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, *21*, 754–768. https://doi.org/10.1038/s41576-020-0269-1

Fenton, A., Antonovics, J., & Brockhurst, M. A. (2009). Inverse-gene-for-gene infection genetics and coevolutionary dynamics. *American Naturalist*, *174*, E230–E242. https://doi.org/10.1086/645087

Feurtey, A., Gladieux, P., Hood, M. E., Snirc, A., Cornille, A., Rosenthal, L., & Giraud, T. (2016). Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. *New Phytologist*, *212*, 668–679. https://doi.org/10.1111/nph.14125

Fitzpatrick, C. R., Salas-González, I., Conway, J. M., Finkel, O. M., Gilbert, S., Russ, D., Teixeira, P. J. P. L., & Dangl, J. L. (2020). The Plant Microbiome: From ecology to reductionism and beyond. *Annual Review of Microbiology*, *74*, 81–100. https://doi.org/10.1146/annurev-micro-022620-014327

Frickel, J., Feulner, P. G. D., Karakoc, E., & Becks, L. (2018). Population size changes and selection drive patterns of parallel evolution in a host-virus system. *Nature Communications*, *9*, 1706. https://doi.org/10.1038/s41467-018-03990-7

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

Haag, K. L., Pombert, J.-F., Sun, Y., de Albuquerque, N. R. M., Batliner, B., Fields, P., Lopes, T. F., & Ebert, D. (2019). Microsporidia with vertical transmission were likely shaped by nonadaptive processes. *Genome Biology and Evolution*, *12*, 3599–3614. https://doi.org/10.1093/gbe/evz270

Haldane, J. B. S., & Jayakar, S. D. (1963). Polymorphism due to selection depending on the composition of a population. *Journal of Genetics*, *58*, 318–323. https://doi.org/10.1007/BF02986301

Hall, A. R., Scanlan, P. D., Morgan, A. D., & Buckling, A. (2011). Host-parasite coevolutionary arms races give way to fluctuating selection: Bacteria-phage coevolutionary dynamics. *Ecology Letters*, *14*, 635–642. https://doi.org/10.1111/j.1461-0248.2011.01624.x

Haller, B. C., & Messer, P. W. (2019). SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Molecular Biology and Evolution*, *36*, 632–637. https://doi.org/10.1093/molbev/msy228

Hartmann, F. E., Snirc, A., Cornille, A., Godé, C., Touzet, P., Van Rossum, F., Fournier, E., Le Prieur, S., Shykoff, J., & Giraud, T. (2020). Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants *Silene italica* and the *Silene nutans* species complex. *Molecular Ecology*, *29*, 1154–1172.

Hecht, L. B. B., Thompson, P. C., & Rosenthal, B. M. (2018). Comparative demography elucidates the longevity of parasitic and symbiotic relationships. *Proceedings of the Royal Society B*, *285*, 20181032. https://doi.org/10.1098/rspb.2018.1032

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Reviews Genetics*, *2*, 516–527. https://doi.org/10.1038/35080508

Janzen, D. H. (1980). When is it coevolution? *Evolution*, *34*, 611–612. https://doi.org/10.1111/j.1558-5646.1980.tb04849.x

Krasovec, M., Chester, M., Ridout, K., & Filatov, D. A. (2018). The mutation rate and the age of the sex chromosomes in *Silene latifolia*. *Current Biology*, *28*, 1832–1838.e4. https://doi.org/10.1016/j.cub.2018.04.069

Lees, J. A., Ferwerda, B., Kremer, P. H. C., Wheeler, N. E., Serón, M. V., Croucher, N. J., Gladstone, R. A., Bootsma, H. J., Rots, N. Y., Wijmega-Monsuur, A. J., Sanders, E. A. M., Trzciński, K., Wyllie, A. L., Zwinderman, A. H., van den Berg, L. H., van Rheenen, W., Veldink, J. H., Harboe, Z. B., Lundbo, L. F., … van de Beek, D. (2019). Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nature Communications*, *10*, 2176. https://doi.org/10.1038/s41467-019-09976-3

Leonard, K. (1993). Stability of equilibria in a gene-for-gene coevolution model of host-parasite interactions. *Phytopathology*, *84*, 70–77. https://doi.org/10.1094/Phyto-84-70

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*, 493–496. https://doi.org/10.1038/nature10231

Luijckx, P., Fienberg, H., Duneau, D., & Ebert, D. (2013). A matching-allele model explains host resistance to parasites. *Current Biology*, *23*, 1085–1088. https://doi.org/10.1016/j.cub.2013.04.064

MacPherson, A., Keeling, M. J., & Otto, S. P. (2021). Coevolution fails to maintain genetic variation in a host–parasite model with constant finite population size. *Theoretical Population Biology*, *137*, 10–21. https://doi.org/10.1016/j.tpb.2020.12.001

MacPherson, A., Otto, S. P., & Nuismer, S. L. (2018). Keeping pace with the red queen: Identifying the Genetic Basis of Susceptibility to Infectious Disease. *Genetics*, *208*, 779–789. https://doi.org/10.1534/genetics.117.300481

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

Märkle, H., Tellier, A., & John, S. (2021). Cross-Species association statistics for genome-wide studies of host and parasite polymorphism data. bioRxiv 726166. https://doi.org/10.1101/726166

Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., Delph, L. F., & Van Rossum, F. (2017). Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution*, *71*, 1519–1531.

May, R. M., & Anderson, R. M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society B*, *219*, 281–313.

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, *15*, 756–771. https://doi.org/10.1038/nrmicro.2017.76

Nemri, A., Atwell, S., Tarone, A. M., Huang, Y. S., Zhao, K., Studholme, D. J., Nordborg, M., & Jones, J. D. G. (2010). Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. *Proceedings of the National Academy of Sciences*, *107*, 10302–10307. https://doi.org/10.1073/pnas.0913160107

Nuismer, S. L., Gomulkiewicz, R., & Ridenhour, B. J. (2010). When is correlation coevolution? *American Naturalist*, *175*, 525–537. https://doi.org/10.1086/651591

Nuismer, S. L., Jenkins, C. E., & Dybdahl, M. F. (2017). Identifying coevolving loci using interspecific genetic correlations. *Ecology and Evolution*, *7*, 6894–6903. https://doi.org/10.1002/ece3.3107

Nuismer, S. L., & Week, B. (2019). Approximate Bayesian estimation of coevolutionary arms races. *PLoS Computational Biology*, *15*, 1–21. https://doi.org/10.1371/journal.pcbi.1006988

Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., Barg, M.-C., Rosenstiel, P. C., Teotónio, H., & Schulenburg, H. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences USA*, *116*, 923–928. https://doi.org/10.1073/pnas.1810402116

Petit, E., Silver, C., Cornille, A., Gladieux, P., Rosenthal, L., Bruns, E., Yee, S., Antonovics, J., Giraud, T., & Hood, M. E. (2017). Co-occurrence and hybridization of anther-smut pathogens specialized on *Dianthus* hosts. *Molecular Ecology*, *26*, 1877–1890.

Petit-Houdenot, Y., & Fudal, I. (2017). Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. *Frontiers in Plant Science*, *8*, 1072. https://doi.org/10.3389/fpls.2017.01072

Pflughoeft, K. J., & Versalovic, J. (2012). Human microbiome in health and disease. *Annual Review of Pathology: Mechanisms of Disease*, *7*, 99–122. https://doi.org/10.1146/annurev-pathol-011811-132421

Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C., & Nelson, R. J. (2009). Shades of gray: The world of quantitative disease resistance. *Trends in Plant Science*, *14*, 21–29. https://doi.org/10.1016/j.tplants.2008.10.006

Retel, C., Kowallik, V., Huang, W., Werner, B., Künzel, S., Becks, L., & Feulner, P. G. D. (2019). The feedback between selection and demography shapes genomic diversity during coevolution. *Science Advances*, *10*, eaax0530.

Retel, C., Märkle, H., Becks, L., & Feulner, P. G. D. (2019). Ecological and evolutionary processes shaping viral genetic diversity. *Viruses*, *11*, 220. https://doi.org/10.3390/v11030220

Routtu, J., & Ebert, D. (2015). Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasites. *Heredity*, *114*, 241–248. https://doi.org/10.1038/hdy.2014.97

Sanchez, T., Cury, J., Charpiat, G., & Jay, F. (2020). Deep learning for population size history inference: Design, comparison and combination with approximate Bayesian computation. *Molecular Ecology Resources*, 1–16. https://doi.org/10.1111/1755-0998.13224

Sasaki, A., Hamilton, W. D., & Ubeda, F. (2002). Clone mixtures and a pacemaker: New facets of Red-Queen theory and ecology. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 761–772. https://doi.org/10.1098/rspb.2001.1837

Saur, I. M., Bauer, S., Kracher, B., Lu, X., Franzeskakis, L., Müller, M. C., Sabelleck, B., Kümmel, F., Panstruga, R., Maekawa, T., & Schulze-Lefert, P. (2019). Multiple pairs of allelic MLA immune receptor-powdery mildew AVRA effectors argue for a direct recognition mechanism. *Elife*, *8*, e44471.

Schenk, H., Schulenburg, H., & Traulsen, A. (2020). How long do Red Queen dynamics survive under genetic drift? A comparative analysis of evolutionary and eco-evolutionary models. *BMC Evolutionary Biology*, *20*, 8. https://doi.org/10.1186/s12862-019-1562-5

Sellinger, T. P. P., Abu Awad, D., Moest, M., & Tellier, A. (2020). Inference of past demography, dormancy and self-fertilization rates from whole genome sequence data. *PLoS Genetics*, *16*, e1008698. https://doi.org/10.1371/journal.pgen.1008698

Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H. F., Drummond, A. J., & Bonhoeffer, S. (2012). Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution*, *29*, 347–357. https://doi.org/10.1093/molbev/msr217

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M., & Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature*, *400*, 667–671. https://doi.org/10.1038/23260

Stam, R., Silva-Arias, G. A., & Tellier, A. (2019). Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats. *New Phytologist*, *224*, 367–379. https://doi.org/10.1111/nph.16017

Stephan, W. (2019). Selective sweeps. *Genetics*, *211*, 5–13. https://doi.org/10.1534/genetics.118.301319

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology*, *9*, e1002803. https://doi.org/10.1371/journal.pcbi.1002803

Tellier, A., & Brown, J. K. M. (2007). Stability of genetic polymorphism in host-parasite interactions. *Proceedings of the Royal Society B*, *274*, 809–817. https://doi.org/10.1098/rspb.2006.0281

Tellier, A., & Brown, J. K. M. (2011). Spatial heterogeneity, frequency-dependent selection and polymorphism in host-parasite interactions. *BMC Evolutionary Biology*, *11*, 319. https://doi.org/10.1186/1471-2148-11-319

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

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

Toenshoff, E. R., Fields, P. D., Bourgeois, Y. X., & Ebert, D. (2018). The End of a 60-year Riddle: Identification and genomic characterization of an iridovirus, the causative agent of white fat cell disease in Zooplankton. *G3 Genes|genomes|genetics*, *8*, 1259–1272. https://doi.org/10.1534/g3.117.300429

Toju, H., Yamamichi, M., Guimarães, P. R., Olesen, J. M., Mougi, A., Yoshida, T., & Thompson, J. N. (2017). Species-rich networks and eco-evolutionary synthesis at the metacommunity level. *Nature Ecology & Evolution*, *1*(2), 1–11. https://doi.org/10.1038/s41559-016-0024

Toruño, T. Y., Stergiopoulos, I., & Coaker, G. (2016). Plant-pathogen effectors: Cellular probes interfering with plant defenses in spatial and temporal manners. *Annual Review of Phytopathology*, *54*, 419–441. https://doi.org/10.1146/annurev-phyto-080615-100204

Van de Weyer, A.-L., Monteiro, F., Furzer, O. J., Nishimura, M. T., Cevik, V., Witek, K., Jones, J. D. G., Dangl, J. L., Weigel, D., & Bemm, F. (2019). A species-wide inventory of NLR genes and alleles in Arabidopsis thaliana. *Cell*, *178*(5), 1260–1272. https://doi.org/10.1016/j.cell.2019.07.038

Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., & Johannes, F. (2016). Methylome evolution in plants. *Genome Biology*, *17*, 264. https://doi.org/10.1186/s13059-016-1127-5

Wadi, L., & Reinke, A. W. (2020). Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLoS Path*, *16*, e1008276. https://doi.org/10.1371/journal.ppat.1008276

Wang, M., Roux, F., Bartoli, C., Huard-Chauveau, C., Meyer, C., Lee, H., Roby, D., McPeek, M. S., & Bergelson, J. (2018). Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proceedings of the National Academy of Sciences USA*, *115*, E5440–E5449. https://doi.org/10.1073/pnas.1710980115

Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B., & Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics*, *32*, 569–577. https://doi.org/10.1038/ng1202-569

Živković, D., John, S., Verin, M., Stephan, W., & Tellier, A. (2019). Neutral genomic signatures of host-parasite coevolution. *BMC Evolutionary Biology*, *19*, 230. https://doi.org/10.1186/s12862-019-1556-3

How to cite this article: Märkle, H., John S., Cornille A., Fields P. D., & Tellier A. (2021). Novel genomic approaches to study antagonistic coevolution between hosts and parasites. *Molecular Ecology*, 30, 3660–3676. https://doi.org/10.1111/mec.16001