



HAL
open science

Novel genomic approaches to study antagonistic coevolution between hosts and parasites

Hanna Märkle, Sona John, Amandine Cornille, Peter D Fields, Aurélien Tellier

► To cite this version:

Hanna Märkle, Sona John, Amandine Cornille, Peter D Fields, Aurélien Tellier. Novel genomic approaches to study antagonistic coevolution between hosts and parasites. *Molecular Ecology*, In press, 10.1111/mec.16001 . hal-03298987

HAL Id: hal-03298987

<https://hal.science/hal-03298987v1>

Submitted on 25 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MOLECULAR ECOLOGY

Novel genomic approaches to study antagonistic coevolution between hosts and parasites

Journal:	<i>Molecular Ecology</i>
Manuscript ID	MEC-21-0080.R1
Manuscript Type:	Invited Reviews and Syntheses
Date Submitted by the Author:	n/a
Complete List of Authors:	Märkle, Hanna; Technical University of Munich, Professorship for Population Genetics, Department of Life Science Systems; University of Chicago, Department of Ecology and Evolution John, Sona; Technical University of Munich, Professorship for Population genetics, Department of Life Science Systems Cornille, Amandine; Univeristé de Paris Sud, Ecologie Systématique et Evolution Fields, Peter; University of Basel, Zoological Institute Tellier, Aurélien; Technical University of Munich, Section of Population Genetics, Department of Life Science Systems
Keywords:	Coevolution, Ecological Genetics, Experimental Evolution, Host Parasite Interactions, Natural Selection and Contemporary Evolution, Species Interactions

1 **Novel genomic approaches to study antagonistic coevolution between hosts**
2 **and parasites**

3 Hanna Märkle (1,2), Sona John (1), Amandine Cornille (3), Peter D. Fields (4), Aurélien Tellier (1)

4 1 Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences,
5 Technical University of Munich, Liesel-Beckmann Strasse 2, 85354 Freising, Germany

6 2 Department of Ecology and Evolution, University of Chicago, 1101 E 57th St, Chicago, IL 60637, USA

7 3 Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette,
8 France

9 4 University of Basel, Department of Environmental Sciences, Zoology, Vesalgasse 1, 4051 Basel,
10 Switzerland

11 **Corresponding authors:** tellier@wzw.tum.de

12 **Keywords:** host-parasite coevolution, positive selection, balancing selection, genetic drift, inference,
13 genomics

14

15 Abstract

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

35

36 1 Introduction

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

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

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

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

85 1.1 Dynamics and characteristics of **antagonistic** coevolutionary interactions

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

95 **Deterministic mathematical models of coevolution have been used to generate predictions on the**
96 **expected type of dynamics for hosts and parasites with various life-history traits** (e.g., Ashby and Boots,
97 2017; Fenton et al., 2009; Haldane and Jayakar, 1963; Leonard, 1993; Tellier and Brown, 2007). However,
98 the simultaneous action of stochastic processes within host and parasite populations may cause deviations
99 from the predicted deterministic outcome (Gokhale et al., 2013; MacPherson et al., 2020; Schenk et al.,
100 2020; Tellier et al., 2014). Stochastic fluctuations, such as genetic drift, can drive alleles maintained at
101 intermediate frequencies to fixation by chance. Therefore, theoretical results suggest arms-race dynamics to
102 occur under a broader set of conditions than trench-warfare dynamics in natural populations (Gokhale et al.,
103 2013; MacPherson et al., 2020; Schenk et al., 2020; Tellier et al., 2014).

104 Besides reciprocal changes in trait distributions and allele frequencies, coevolution is also likely to
105 cause population size changes in both species due to eco-evolutionary feedback resulting from
106 epidemiological dynamics (Ashby et al., 2019; Gokhale et al., 2013; May and Anderson, 1983). **The**
107 **underlying principle is that allele frequency change (evolutionary change) affects host and parasite fitness**
108 **and results in changes in population sizes (ecological change). These changes, in turn, impact the allelic**

109 fitness and frequencies and feeds back onto the evolutionary change (Ashby et al., 2019, Figure 1). Host-
110 parasite eco-evolutionary feedbacks can result in correlated changes of host and parasite population size
111 over time. These correlated population size changes are termed as the co-demographic history (Živković et
112 al., 2019). Furthermore, many host-parasite coevolutionary interactions extend beyond single host and
113 parasite populations. Rather, single host and parasite populations are embedded in landscapes with varying
114 extents of homogeneity and are connected to other populations by varying levels of gene flow (Thompson,
115 2005). The amount of gene flow determines the extent of synchrony of coevolutionary dynamics between
116 populations (Sasaki et al., 2002; Tellier and Brown, 2011). In summary, host-parasite coevolutionary
117 dynamics result in local changes in phenotypes/allele frequencies and population sizes (Figure 1) and have
118 a temporal and a spatial component. Note that the coevolutionary dynamics and the co-demographic history
119 (change in population size) generally apply to other inter-species interactions with GxG interactions (for
120 example, driven by frequency or density-dependent selection). Still, the specific conditions for stability or
121 fixation of trait/allele frequency depend on the system under consideration (prey-predator, plant-pollinator,
122 etc.).

123 **1.2 How do coevolutionary processes link to host and parasite genomic data?**

124 From a coevolutionary perspective, host and parasite genomes can be conceptually partitioned into
125 two loci categories, namely neutral (regarding coevolution) and coevolving loci (Figure 1). As opposed to
126 neutral loci, coevolving loci contribute to the phenotypic traits which are under coevolutionary selection.
127 Hence, coevolutionary selection shapes polymorphism patterns at the latter loci (see Figure 1). In hosts,
128 most genes supposedly do not affect the outcome of coevolution, and, thus, their effect can be considered
129 neutral regarding coevolution. In contrast, the number of parasite genes that are not involved in
130 coevolutionary interactions can be small. This pattern should especially apply to obligate parasites which
131 require an infected host to complete their lifecycle (Möller and Stukenbrock, 2017). Coevolving loci can be
132 further classified based on their genetic contribution (major vs. minor) to a coevolving phenotype and

133 whether they form the genetic basis of qualitative or quantitative traits. Coevolving loci in the host, include
134 (i) major resistance genes defining host-parasite qualitative specificity and resulting in a strong defence
135 response (e.g., Nucleotide binding Leucine-rich repeat Receptors (NLRs) in plants; Stam et al., 2019; or the
136 Major Histocompatibility Complex (MHC) in animals Ansari et al., 2017), and (ii) minor resistance genes
137 (Poland et al., 2009) defining the quantitative strength of host resistance response (Figure 1). Coevolving
138 loci in the parasite include (i) major genes defining parasite specificity and infectivity, such as effectors
139 (Toruño et al., 2016), and (ii) minor genes defining the quantitative effect of the infection on host fitness
140 (so-called virulence in animal/human epidemiology or disease severity/aggressiveness in plant pathology)
141 or disease transmission (Figure 1). For example, while being a crucial parasite quantitative trait, parasite
142 transmission has its effect superseded by major genes such as effectors with respect to defining the outcome
143 of the (GxG) interaction with host genotypes. It is noteworthy that additional abiotic and biotic factors can
144 further shape signatures at coevolving loci. This relationship is especially true for crops in which resistance
145 to parasites can be only an indirect outcome of breeder selection schemes (Brown 2015).

146 Genome-wide stochastic processes such as genetic drift, mutation, recombination, and gene flow
147 between populations (Figure 1) shape polymorphism patterns at both types (coevolving and non-
148 coevolving) of loci. In particular, the amount of genetic drift depends on population size changes during the
149 coevolutionary interaction. These population size changes can either directly result from coevolution itself
150 (the co-demographic history) or processes independent of coevolution, such as changes in the abiotic
151 environment or range expansion. The type and speed of coevolutionary dynamics (how fast allele
152 frequencies cycle or reach fixation) determines genetic polymorphism at the coevolving loci (Figure 1). It
153 is expected that arms-race dynamics generates signatures of (recurrent) selective sweeps, whereas trench-
154 warfare dynamics generates balancing selection signatures (Bergelson et al., 2001; Ebert and Fields, 2020;
155 Holub, 2001; Woolhouse et al., 2002). Signatures of selective sweeps include locally decreased nucleotide
156 diversity level, complex patterns of linkage disequilibrium around the site under selection, and an excess of
157 low and high-frequency SNP variants relative to the genomic background (Stephan, 2019). Conversely,

158 balancing selection signatures are characterized by elevated polymorphism levels, elevated linkage
159 disequilibrium, and an excess of intermediate frequency variants (Charlesworth, 2006). However, Tellier *et*
160 *al.* (2014) have shown that these predictions on resulting coevolutionary signatures are often too simplistic.
161 First, genetic drift can broadly impact the signatures at the coevolving loci and render them indistinguishable
162 from non-coevolving loci. Second, under trench-warfare dynamics, the genetic signatures at the involved
163 loci depend on the allele frequencies at the expected theoretical equilibrium state and their proximity to
164 frequency zero (loss) or frequency one (fixation) (Märkle and Tellier, 2020). Overall, trench-warfare
165 dynamics might be relatively uncommon in natural populations, but if they occur and are long-lived, they
166 should result in detectable polymorphism signatures of balancing selection. Moreover, polymorphism
167 signatures of arms-race and trench-warfare are more likely to be detected in parasite than in host
168 polymorphism data (Tellier *et al.*, 2014). Selective sweeps and balancing selection at genes involved in
169 coevolution have been extensively searched for in several host species (Ebert and Fields, 2020 and
170 references therein) and both signatures are found in the animal host species *Daphnia magna* (see Box 1),
171 while only selective sweeps were found in plant-pathogen species of the genus *Microbotryum* (see Box 2).

172 In summary, the polymorphism patterns at different genes/loci (neutral vs. coevolving) in a genome
173 are differentially affected by various selective and random processes acting during coevolution (Retel *et al.*,
174 2019b). This pattern suggests that ideally, the analysis of polymorphism patterns across different genes/loci
175 should allow for assigning individual loci to either category (neutral or coevolving). While performing this
176 assignation, we are specifically interested in answering three sets of coevolutionary questions: (i) What are
177 the genetic determinants of coevolution? (ii) What are the selective processes underlying coevolution? (iii)
178 Do eco-evo feedbacks occur, and what is the role of spatial structure in coevolution? So far, these questions
179 have been mainly addressed by employing methods originally designed to analyse single species (only one
180 of the antagonists). In the following, we describe recently developed methods that jointly analyse host and
181 parasite polymorphism data, paving the way to answer these questions at finer resolution by explicitly taking
182 the reciprocal nature of host-parasite interactions into account.

183 2 Recently developed methods to answer coevolutionary questions

184 2.1 Revealing the genetic underpinnings of coevolution

185 The first set of questions relates to uncovering the genetic underpinnings of coevolution, which
186 includes unravelling the molecular basis of coevolution and counting the number of genes involved. Due to
187 the relevance of these questions for medicine and plant and animal breeding in agriculture, the molecular
188 mapping and study of the genes underlying host-parasite interactions have several decades of history and
189 started with classic genetics (mutant screen) and quantitative genetics approaches (trait mapping). With the
190 advance of next-generation sequencing methods, it has become possible to pinpoint the genes underlying
191 resistance or infectivity phenotypes using genome-wide association studies (GWAs). GWAs methods
192 quantify the statistical correlation between genetic markers (bi-allelic single nucleotide polymorphism,
193 SNPs) and a trait of interest (*e.g.*, level of infection) in a sample of individuals. Several GWAs have revealed
194 the genetic variation underlying resistance and disease susceptibility in hosts (Bourgeois et al., 2017 in
195 *Daphnia magna*; Nemri et al., 2010 in *Arabidopsis thaliana*) or infectivity in parasites (Saur et al. 2019
196 based on transcriptome data). Note that the system of *Daphnia magna* and its parasites has been particularly
197 well studied (Box 1). However, MacPherson et al. (2018) showed that coevolutionary dynamics, *i.e.*,
198 temporal changes in allele frequencies, alter the measured allelic effect sizes in single-species GWAs, thus
199 decreasing their statistical power.

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

206 2.1.1 *New methods using joint genome analyses: Experimental co-GWAs*

207 To overcome the shortcomings of single-species GWAs, MacPherson et al. (2018) laid down the
208 foundations for a new type of GWAs that simultaneously incorporates host and parasite genetic information.
209 The principle of the co-GWAs is to separate the variance in infection as the combination of parasite
210 genotypes, host genotypes, and the GxG interactions. Along with the challenge of obtaining genetic
211 information of both species, co-GWAs is computationally demanding as it involves a large number of
212 pairwise comparison of host-parasite genome data. Using a rather simplified two-species co-GWAs scheme,
213 MacPherson et al. (2018) could estimate allelic effect sizes that are robust to reciprocal changes in allele
214 frequencies and thus, boost the amount of phenotypic variance explained compared to single-species GWAs.
215 Wang et al. (2018) introduced a full co-GWAs method (Analysis with a Two-Organism Mixed Model, in
216 short ATOMM), which takes the reciprocal nature of host-parasite interactions into account. The ATOMM
217 method requires phenotypic data measured for each host-parasite pair in a controlled infection experiment
218 and the corresponding genome-wide host and parasite genotype data (can also be used with a high-density
219 SNP array). The co-GWAs uses a two-way mixed-effects model that incorporates the effects of host and
220 parasite genetic variants and their interaction as fixed effects. To address population structure, three different
221 genetic relatedness matrices are added as random effects to the model. These matrices include one for the
222 host, one for the parasite, and one capturing the additive-by-additive polygenic interaction effects between
223 the host and the parasite. The method, developed with bacterial pathogens in mind, can deal with different
224 genetic variants, namely SNPs or presence/absence polymorphism. The method is also generalized for
225 multi-allelic variants (Table 1). The ATOMM method is applicable to phenotypic traits that are multivariate
226 normal or binomial-like traits (the trait can fall into one of several discrete categories). It is possible to either
227 marginally map the trait to the host genome, the parasite genome, or perform an interaction test between
228 host and pathogen variants. The authors tested their method by performing an experimental co-GWAs in
229 the *Arabidopsis thaliana*-*Xanthomonas arboricola* plant-pathosystem. A total of 130 inbred *A. thaliana*
230 accessions from the 1001 genomes project were infected with 22 *X. arboricola* strains and quantitative disease

231 resistance was measured as a response trait. Wang et al. (2018) showed the existence of host-strain-specific
232 quantitative resistance and a suggestive lack of broad-spectrum quantitative disease resistance against *X.*
233 *arbicola*.

234 2.1.2 *New methods using joint genome analyses: Natural co-GWAs*

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

248 **To** our knowledge, three studies have been conducted so far on human hosts and the genomes of
249 successfully infecting (i) HIV-1 (Bartha et al., 2013), (ii) hepatitis C (Ansari et al., 2017), and (iii)
250 pneumococcal meningitis (Lees et al., 2019) strains. The pioneering study by Bartha et al. (2013) combined
251 sequences of 1,071 human genomes and the corresponding HIV-1 strains, aiming to map the host genetic
252 pressure on the HIV-1 genome. As expected, the strongest association signal is obtained between human
253 SNPs tagging HLA class 1 alleles and viral mutations in their corresponding CTL (cytotoxic T lymphocytes
254 which kill infected cells) epitopes, thus validating the power of such a pairwise comparison analysis.

255 Different regions previously unknown to have any impact on the interaction were also identified. Similarly,
256 the association between 601 HCV-infected patients and the hepatitis C virus showed that the viral
257 polymorphism is strongly associated with the immune system's human gene components (Ansari et al.,
258 2017). Note, however, that viruses (such as HIV and HCV) have comparatively small genomes, and it
259 becomes challenging to perform natural co-GWAs on bacterial genomes (such as *Streptococcus*
260 *pneumoniae*, the causal agent of pneumococcal meningitis) because of a large number of comparisons which
261 hinders the detection of significant associations after correcting for multiple testing (such as a Bonferroni
262 correction). To overcome this challenge, Lees et al. (2019) defined pneumococcal lineages and tested the
263 association between pathogen lineages (transformed as a bi-state genotype) and given host genotype.

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

278 2.1.3 *New methods using joint genome analyses: association indices*

279 As an addition to natural co-GWAs studies, two cross-species association indices have been
280 developed based on pairwise comparison of host and parasite SNPs, using randomly chosen infected (along
281 with their parasites) and non-infected hosts (Märkle et al., 2021) (here randomly implies an unbiased sample
282 of the coevolving populations). **These indices measure the degree of association between allele frequencies
283 in the host (infected, non-infected and both types) and allele frequencies in the parasite samples, mirroring
284 measures of linkage disequilibrium in population genetics. Therefore, the authors can derive the theoretical
285 expectations for the distributions of these association indices based on the theoretically expected allele/site
286 frequency spectrum from population genetics (while the expected statistical power is not easily predictable
287 for co-GWAs).** This study indicates that the power of association indices, and by extension of natural and
288 experimental co-GWAs, to reveal genes under coevolution varies in time, depends on the type of
289 coevolutionary dynamics, and is maximum when the alleles are at intermediate frequencies. Thus,
290 coevolving loci are more likely to be detected under (long-term) trench-warfare than under the arms-race
291 dynamics. Furthermore, obtaining time samples improves the statistical power of these indices in particular,
292 and most likely of all co-GWAs methods, as they capture the temporal fluctuations in allele frequencies at
293 the coevolutionary loci (Märkle *et al.*, 2021). **The association indices do not replace co-GWAs to pinpoint
294 loci under coevolution, but allow to link the results of co-GWAs with the theoretical models of coevolution
295 and thus could be used for statistical inference of coevolutionary parameters at significant pairs of associated
296 SNPs found in co-GWAs (Märkle *et al.*, 2021).**

297 These results have implications for the experimental design of co-GWAs studies. As the genes
298 underpinning coevolution revealed by co-GWAs are more likely to be those under trench warfare dynamics,
299 it is desirable to study host-parasite systems where the effect of genetic drift is relatively weak (Tellier et
300 al., 2014). We, thus, advise the construction of a panel of samples from the whole species range rather than
301 small local populations (*e.g.*, the 130 *A. thaliana* sample panel, Wang et al., 2018). The rationale is that if
302 coevolution is pervasive in space, trench warfare dynamics are more likely to occur and be long-term in a

303 large spatially structured population with a large effective population size. Conversely, revealing genes
304 under arms-race dynamics requires one to obtain samples from several independent populations that are
305 asynchronous in their coevolutionary process (*e.g.*, Haag et al., 2019). Further, it is crucial to keep in mind
306 that co-GWAs can only pinpoint genes under coevolution if they are polymorphic, a state which is only
307 transiently observable under arms race dynamics. Finally, we also point out that for parasites with large
308 genomes (bacteria, *Plasmodium*, fungi), there is a need to reduce the dataset's dimensionality by defining
309 lineages based on relatedness to avoid high false-negative rates (due to corrections for multiple testing).
310 However, defining parasite lineages becomes likely problematic for sexually reproducing parasites
311 (*Plasmodium*, some fungi) due to intra-genomic recombination. In the latter case, we suggest regrouping
312 parasite strains based on their genetic relatedness as measured by population genomic clustering methods
313 (*e.g.*, PCA, Structure, Admixture, DAPC). Finally, we note that the sample sizes in the natural co-GWAs
314 performed on human-parasite systems (Ansari et al., 2017; Bartha et al., 2013; Lees et al., 2019) are large
315 (several thousand) and obtained at a single point in time. However, Märkle et al. (2021) predict that few
316 hundreds of samples at one-time point (on the order of sample sizes in Wang et al., 2018) might be enough
317 to obtain adequate statistical power. It is also predicted that the statistical power of co-GWAs would be
318 increased if samples are available at several time points. This expectation implies that natural and
319 experimental co-GWAs approaches are not mutually exclusive, and for ad hoc systems (*e.g.*, *Daphnia*
320 *magna*), it would be highly interesting to perform both simultaneously to assess the influence of
321 environmental variation on the outcome of coevolution. However, as it may be difficult for many natural
322 systems to obtain many samples, an alternative would be to obtain hundreds of samples widespread in space
323 and time (over several years). We note that such extensive sampling is short to be readily available for the
324 host-parasite systems presented in Boxes 1 and 2. Using an experimental co-GWAs approach, we suggest
325 studying crop coevolution with pathogens by using comparisons between multiple crop varieties and
326 multiple parasite strains obtained across different locations and across different years.

327 2.2 Inferring the selection pressure underlying coevolution

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

334 Population genomics studies aim to 1) scan the genomes for genes under positive or balancing
335 selection and 2) draw a statistical inference of the past reciprocal selective history. Genome scan methods,
336 applied to either host or parasite data (Ebert and Fields, 2020 and references therein), can be used to detect
337 genes with statistically significant signatures of selection compared to the genomic average while
338 accounting for the population's past demographic history (Stephan, 2019). These methods usually make use
339 of any of the following properties: the distribution of SNPs (frequency-wise and spatially along the
340 sequence), patterns of linkage disequilibrium, the number of substitutions compared to an outgroup species,
341 and the ratio of non-synonymous to synonymous substitutions. However, variation in recombination rates
342 along the genome, confounding effects of past demography (*e.g.*, severe bottlenecks, admixture, population
343 expansion, etc.), or strong genetic drift can limit the detection of selection signatures (*e.g.*, Stephan 2019).
344 These factors may explain the fact that, despite genome scans for positive or balancing selection being
345 conducted chiefly on wild host or wild parasite species (Boxes 1 and 2), on crops and their pathogens or
346 humans and their parasites, there are still few documented and demonstrated host-parasite pairs of truly
347 coevolving genes (with complementary gene expression and functional studies of the candidate genes, see
348 reviews in Ebert and Fields, 2020; Möller and Stukenbrock, 2017; Petit-Houdenot and Fudal, 2017).

349 2.2.1 *New methods using joint genome analyses: Inference of coevolution at coevolving loci*

350 To go beyond mere genome scans for selection and to make use of methods designed to infer the
351 amount of selective pressure, Märkle and Tellier (2020) assessed the amount of information on the
352 parameters underlying the coevolutionary interaction that can be retrieved from host and parasite
353 polymorphism data by means of approximate Bayesian computation (ABC). ABC is a computational
354 method using so-called summary statistics of the dataset (observed and simulated) to estimate posterior
355 probabilities of models or parameters of interest when likelihood calculations are intractable (Csilléry et al.,
356 2010; Sunnåker et al., 2013). Therefore, an extensive number of simulations are run with the parameter
357 values of the underlying model sampled from prior distributions, which summarize the current (prior)
358 knowledge. A rejection step retains all simulations for which the summary statistics fall within a certain
359 distance to the observed data's summary statistics. The retained simulations are used to estimate a posterior
360 probability for each competing model (model choice) or generate a posterior distribution for the parameters
361 of interest (Csilléry et al., 2010; Sunnåker et al., 2013).

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

369 The ABC model choice is used to distinguish pairs of coevolving host and parasite loci from pairs
370 of neutral (non-coevolving) loci. Some scenarios have high model choice accuracy, which decreases when
371 either the equilibrium frequencies under trench-warfare dynamics are too close to the boundaries (*i.e.*, allele
372 frequency of zero or one) or when alleles get rapidly fixed under arms-race dynamics. In a second step,
373 inference of parameters of the past coevolutionary history (fitness costs) takes place. The best inference

374 results when jointly using host and parasite summary statistics rather than only using either of those. This
375 result is consistent with host allele frequencies and hence, the resulting sequence data, being informative on
376 the parasite parameters and vice-versa (Märkle and Tellier, 2020; Tellier et al., 2014; Tellier and Brown,
377 2007). In general, the parameter estimations are more accurate when coevolution follows trench-warfare
378 dynamics and when data from more repetitions are available (a minimum of 10). This dependency highlights
379 the need to account for genetic drift when developing new methods to analyse host-parasite coevolution.
380 Sample sizes of at least $n=50$ for each species are advisable. Finally, identifiability issues arise when distinct
381 parameter combinations result in similar coevolutionary dynamics, and hence, the inference accuracy
382 decreases.

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

399 2.2.2 *Further applications: towards joint genome analyses in experimental coevolution studies*

400 Genome-wide data of hosts and parasites can also be obtained from controlled laboratory
401 coevolutionary experiments of systems with very short generation times (bacteria and phage, *C. elegans* and
402 virus, algae, and microparasite) (Frickel et al., 2018; Hall et al., 2011; Papkou et al., 2019; Retel et al.,
403 2019a). Sequencing of host individuals and parasite strains occurs at the beginning, during, and at the end
404 of the coevolutionary experiment, which consists of repeatedly following the evolution of hosts exposed to
405 the parasite and hosts evolving without parasite pressure (Figure 2c). These studies aim to decipher the
406 genetic bases of coevolution. **The power of these studies** depends on the length of the experiments (how
407 many generations) but also and chiefly on 1) the amount of genetic drift, 2) the mutation rate, 3) the
408 recombination rate, and 4) the host and parasite genetic diversity present at the onset of the experiment.
409 Similarly to a GWAs, genes potentially underlying coevolution are those with aberrant SNP/allele frequency
410 changes when coevolutionary replicates are compared to host-only replicates (Frickel et al., 2018; Papkou
411 et al., 2019; Retel et al., 2019a). Large host and parasite population sizes, which can arise if recurrent
412 bottlenecks due to the *in vitro* multiplication rounds and eco-evo feedbacks are not too severe, can favour
413 selection over genetic drift and the occurrence of mutations upon which selection can occur (Frickel et al.,
414 2018; Hall et al., 2011; Retel et al., 2019a). On the other hand, for species with limited population sizes and
415 mutation rates, it is more efficient to start the experiment with large genetic diversity (Papkou et al., 2019).
416 Thus far, analysis of host and parasite genome data are usually performed independently, so there is no
417 specific statistical inference of the coevolutionary process. In addition to replicates, these controlled
418 coevolutionary experiments also provide time samples of the coevolutionary process (Figures 2c and 2d).
419 Hence, natural co-GWAs and association indices can, in principle, be applied to experimental coevolution
420 studies on individuals at the end of the experiment. Nonetheless, it is often difficult in such fast-evolving
421 and microscopic organisms to clearly define and isolate infected or non-infected single host individuals and
422 the corresponding parasite strain (Hall et al., 2011; Retel et al., 2019a). We, therefore, speculate that natural

423 and experimental co-GWAs would more likely be possible in the *C. elegans* system (Papkou et al., 2019),
424 where it is possible to identify infected individuals, though the host sample size may be too small.

425 **2.3 Eco-evo feedbacks and spatial structure: insights from genomics**

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

434 *2.3.1 New methods using joint genome analyses: Inference of co-demographic history.*

435 An interesting characteristic of many strict host-parasite interactions is the parasite's dependency
436 on the host for its reproduction, constraining the parasite population size. **Therefore, host and parasite**
437 **population size changes in time may not be fully independent. Thus, inferring historical changes in**
438 **population size may** contain information on the coevolutionary process. We define this correlation in
439 population sizes as the co-demographic history (Živković et al., 2019). Co-demographic changes can occur
440 on two different time scales (long-term vs. short-term) and be due to two distinct mechanisms (one-sided
441 vs. reciprocal co-demographic history) (Figure 3a and Figure 3b). First, long-term changes (thousands of
442 generations) in host population size possibly drive changes in the parasite population size (Hecht et al.,
443 2018) (Figure 3a). Long-term correlations in population sizes can be, for example, due to the expansion of
444 the host population, *e.g.*, when colonizing new habitats. This correlation can result in a subsequent
445 expansion of the parasite (as speculated in Hecht et al., 2018). Thus, this type of change can be defined as
446 one-sided as changes in host population size initiate them. Second, eco-evolutionary feedback due to

447 epidemiological dynamics potentially generates short-term changes (few hundreds of generations) in host
448 and parasite population sizes, termed as the reciprocal co-demographic history in Živković et al. (2019)
449 (Figure 3b). **Recent studies show that correlated host and parasite population size changes on both timescales**
450 **can be inferred based on full-genome SNP data** (Hecht et al., 2018; Živković et al., 2019) (Figure 1).

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

466 Population size changes resulting from the reciprocal co-demographic history (due to eco-
467 evolutionary feedback) are too fast to be captured by a coalescent approach. However, Živković et al. (2019)
468 have shown that changes in the genome-wide distribution of allele (SNP) frequencies between time samples
469 during coevolutionary dynamics can be informative on host and parasite population size changes. They also
470 demonstrated that the changes in the SFS are only observable if the population sizes are small enough at the
471 start of the coevolutionary dynamics. In most eco-evo host-parasite coevolutionary models (Ashby et al.,

472 2019; Ashby and Boots, 2017; Gokhale et al., 2013; Živković et al., 2019), the parasite population exhibits
473 more severe size variations over time (recurrent bottlenecks) than the host population. Such drastic
474 bottlenecks can be detected by changes in the SFS over time. The resolution and statistical power depend
475 on the number of parasite generations per host and the number of parasites per host individual (Živković et
476 al., 2019). The genome data across the different time samples in the experimental coevolution study of Retel
477 et al. (2019a) seem to match these predictions. **Currently, the required host and parasite time sample data
478 are only available from experimental coevolution systems. However, sequencing hosts and their parasites
479 over time in short-lived plants and invertebrates should also be feasible and help to document the occurrence
480 of eco-evo feedbacks in these systems. Especially, the *Daphnia* system could be used for such a study as
481 recovery and sequencing of time samples from sediments for host and parasites is possible (Decaestecker et
482 al., 2007) (Box 1).**

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

490 2.3.2 *Spatial correlation of host and parasite allele frequencies:*

491 Spatial structure is important in host-parasite interactions (Figure 3c). A large body of theoretical
492 models has been built to predict which antagonist is ahead in the coevolutionary interaction (that is local
493 adapted) depending on local population sizes and gene flow levels. Both host and parasite spatial structure
494 can be inferred from full genome data (the neutral loci) and compared to one another (Feurtey et al., 2016,

495 Box 2). Moreover, given sufficiently low levels of gene flow, data from single coevolving populations can
496 be considered independent coevolutionary replicates for inference.

497 Nuimser et al. (2017) proposed a method to identify coevolving loci by testing for significant spatial
498 covariation between host and parasite genetic marker frequencies. The method has the premise that if host-
499 parasite local adaptation is detectable in reciprocal cross-infection, it should be due to spatially co-varying
500 allele frequencies at the functional genetic loci. The method starts by calculating allele frequencies for a set
501 of host and parasite genetic markers for each population within a meta-population. Based on the marker
502 frequencies, the spatial covariance is calculated for each pair of host-parasite markers (e.g., bi-allelic SNPs).
503 The resulting covariance matrix is converted into a correlation matrix which is screened for significant
504 correlations using a Student's t-test. To test their method, the authors simulated data under different types
505 of matching-alleles models (1-3 loci, with epistatic or additive effects), varying degrees of local adaptation,
506 and varying the number of populations sampled. Therefore, a coevolving locus was randomly placed in a
507 set of neutral markers. Overall, the method exhibits a very high statistical power if local adaptation is strong
508 and allele frequency estimates are available for more than 30 populations (while they consider only 100 loci
509 and large sample sizes of >100 per population). Note that if full genome data (many more than 100 loci) are
510 to be used, the effect of multiple-testing and the resulting FDR should be assessed for the given sample
511 sizes. Furthermore, the method's strength substantially drops when local adaptation is weak, the number of
512 populations is small (<20), and more than one locus is involved in the interaction. We thus recommend
513 using this method if strong local adaptation is demonstrated *a priori* in a reciprocal cross-infection
514 experiment, a large number of independent populations can be sampled (>30 populations), and coevolution
515 depends on very few loci with a large phenotypic effect. Note as well that the method relies on three crucial
516 assumptions (as also discussed by Nuimser et al., 2017). First, populations must be independent of one
517 another, which implies that gene flow levels among populations must be reasonably small. Practically, the
518 rate of gene flow can be measured using full genome data and classic population genetics methods. Second,
519 the marker distribution along the genome must be sufficiently dense (e.g., full genome data or high-density

520 SNP array) to pinpoint the loci under coevolution. Third, markers must be independent of one another,
521 which requires a sufficiently high recombination rate. Further, the method can result in false positives if the
522 covariation of host and parasite marker frequencies is due to the adaptation of both species to the shared
523 environment. (Nuismer et al., 2017). *By sequencing full genomes from several individuals across several*
524 *populations, the *Daphnia-parasites* and *Silene-Microbotryum* systems could be used to test and further*
525 *develop these methods.*

526 **3 Conclusion and outlook**

527 *Above, we have summarized recently developed methods to jointly analyze host and parasite*
528 *genomic data and, thus, to account for the reciprocity of these interactions explicitly. These methods provide*
529 *promising avenues to extend our understanding of multiple aspects of host-parasite coevolution. The*
530 *development of these methods is often based on a specific type of data and specific (and sometimes quite*
531 *simplistic) assumptions regarding (i) the nature of the coevolutionary interaction, (ii) the life-history of the*
532 *interacting species, and (iii) the underlying genomic architecture. Yet, every host-parasite system is unique*
533 *in terms of life-history traits, the genomic features (such as genome size, recombination rates, number of*
534 *loci involved, etc.), the extent of host and parasite population structure, its experimental tractability, and the*
535 *feasibility of obtaining particular types of genomic data (cost-wise, time samples, spatial samples, samples*
536 *from replicated experiments). Table 1 summarizes current limitations, sample size requirements, and*
537 *potential extensions of the described methods. In their current state, some of these methods might be only*
538 *applicable to a limited number of host-parasite systems, and there is a need for further theoretical and*
539 *statistical developments, as well as to optimize the combination of existing methods (gene expression,*
540 *functional studies, experimental coevolution, etc.). We advocate using several empirical and experimental*
541 *approaches simultaneously to study coevolution, as the observed genomic patterns can result from different*
542 *coevolutionary and non-coevolutionary processes. We end this review by (i) suggesting some theoretical*

543 and methodological developments for the near future and highlighting (ii) three main challenges and
544 unanswered questions regarding coevolution that require longer-term theoretical and empirical work.

545 **3.1 Further directions for theoretical/methodological developments**

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

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

564 Second, the inference of the past coevolutionary history from coevolving loci should be developed
565 further. Märkle and Tellier (2020) have only scratched the surface of the realm of possibilities. The
566 assumption of only a few major loci controlling coevolution could be very strong and possibly invalid for

567 many host-parasite pairs (see Box 2). We suggest developing inference methods specifically designed to
568 consider more realistic scenarios of coevolution with (i) quantitative trait loci, (ii) clusters of resistance
569 genes, (iii) networks of interacting genes in the host and the parasite, and (iv) diffuse coevolution between
570 host and parasite communities. These developments may greatly benefit from recent advances in the field
571 of host and parasite pangenomics (e.g. van der Weyer *et al.* 2019), integration of gene network structure
572 into the analysis framework, advances in population genomic machine learning methods (e.g., Sanchez *et*
573 *al.* 2020), more efficient and flexible forward-in-time simulators (e.g., Haller and Messer, 2019), and the
574 integration of metagenomic samples in time and space (e.g., Toju *et al.* 2017). Furthermore, additional prior
575 information such as the presence/absence and function of parasite effectors could provide additional power.
576 Ultimately, these methods should be made accessible to a broader community by integrating them into user-
577 friendly interfaces.

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

587 **3.2 Remaining unanswered questions**

588 One of the main challenges remains to test whether two species are genuinely coevolving. Several
589 studies have pointed out that correlations between host and parasite traits are not necessarily due to
590 coevolution, and coevolution does not necessarily result in detectable correlations (Janzen, 1980; Nuismer

591 et al., 2010). It is desirable to use joint analysis of genome data from hosts and parasites to distinguish
592 strict/diffuse coevolution from unilateral evolution (Nuismer and Week, 2019), comparing different
593 populations under various coevolutionary pressures. A second far-reaching question is to what extent
594 epigenetics, especially genome methylation, affect coevolution or if methylation can be the underlying
595 coevolution mechanism. Models of methylation and epigenetics developed in the context of species
596 adaptation and the corresponding population genetics results and methods are readily available (Vidalis et
597 al., 2016) and could be potentially adapted to host-parasite interactions. Finally, both antagonists are
598 usually part of a broader community, and thus, most interactions involve multiple hosts and parasites.
599 Sequencing the composition and genetic diversity of microbes (the microbiome) or host (plant, animal)
600 communities is becoming feasible **via metagenomics**. Consequently, it would be desirable to (i) improve
601 our understanding of host-parasite coevolution in a community context and generate predictions on their
602 genomic effects, and (ii) correspondingly extend existing inference methods to accommodate for multi-
603 species interactions.

604

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 and 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 and Didelot, 2018), to identify a distinct *Pasteuria* collagen-like gene (PCL) responsible for infection of specific host genotypes. A GWAs study by Ameline et al. (2021) and a QTL

630 study by Bento et al. (2020) identified two novel additional, chromosomally distinct, genomic regions of
631 the *D. magna* genome which determine resistance to other *P. ramosa* strains. Furthermore, while within
632 chromosome epistatic interactions were already suggested in order to understand host resistance to *P.*
633 *ramosa*, Ameline et al. (2021) and Bento et al. (2020) showed that among chromosomal interactions are
634 involved in determining the resistance of *D. magna* to distinct strains of *P. ramosa*.

635

For Review Only

636 **Box 2 Coevolution at quantitative traits in the spatially structured *Silene* – *Microbotryum* system.**

637 Another well-documented host-parasite system is the species complex of anther-smut fungi,
638 *Microbotryum* spp., castrating their *Silene* sp. hosts. Anther-smut fungi are highly specialized on their *Silene*
639 host. There is accumulating evidence for a complex genomic basis of the *Microbotryum*-*Silene* coevolution
640 involving many minor loci and noticeably the absence of major gene-for-gene loci. Genome scans of
641 selective sweeps within populations of *M. lychnidis-dioicae* and *M. silenes-dioicae* detected a recent
642 signature of positive selection (selective sweeps) in effector genes (Badouin et al., 2017). The number,
643 localization, and presence/absence polymorphism of the detected genes differed among the two studied
644 *Microbotryum* species, suggesting different coevolutionary dynamics intensities (Badouin et al., 2017).
645 Gene expression analyses of *Microbotryum* strains *in planta* vs. *in vitro* further revealed that those candidate
646 genes are differentially expressed and likely involved in the specialization and coevolution. However,
647 functional validation of these candidate genes is still missing. Genomic data at the population level for the
648 *Silene* host are also still lacking, because of the large size of the *Silene* genome (2.8 Gb, Krasovec et al.,
649 2018). However, extensive population genetics analyses using microsatellite markers of *S. latifolia*, *S.*
650 *nutans* and *S. dioica* species provide a first glimpse into the coevolutionary processes in the genus (Feurtey
651 et al., 2016; Hartmann et al., 2020; Martin et al., 2017). The recent combination of microsatellite data from
652 population samples of both the host and *Microbotryum* species shows congruence of the population spatial
653 genetic structure between the *Silene* and *Microbotryum*, even when correcting for isolation by distance
654 (Feurtey et al., 2016).

655 Additionally, it is also of interest to understand the link between coevolution and specialization/divergence
656 by comparing different plant hosts (*Silene* sp. or *Dianthus* sp.)- *Microbotryum* system. Whereas strict host
657 specialization is often the rule on *Silene* species (Hartmann et al., 2020), *Microbotryum* shows broader and
658 overlapping specialties on *Dianthus* hosts (Petit et al., 2017). Comparing patterns of coevolution in the two

659 host genera and taking into account the host and the pathogen's demographic history, remain challenging
660 but exciting avenues to unravel the genomic bases of coevolution in this multi-species system.

661
662
663
664

Acknowledgements

665 HM and SJ were funded by the Deutsche Forschungsgemeinschaft grant TE809/3 (project 274542535) to
666 AT within the SPP 1819 “Rapid evolutionary adaptation: Potential and constraints”. HM is funded by
667 a Victor Dropkin postdoctoral fellowship. AC was funded by the ATIP-Avenir program. PF is funded by a
668 grant from the Swiss National Science Foundation.

669
670
671

Bibliography

- 672 Ameline, C., Bourgeois, Y., Vöggtli, F., Savola, E., Andras, J., Engelstädter, J., Ebert, D., 2021. A two-
673 locus system with strong epistasis underlies rapid parasite-mediated evolution of host resistance.
674 Mol. Biol. Evol. M38(4), 1512-1538.
- 675 Andras, J.P., Fields, P.D., Du Pasquier, L., Fredericksen, M., Ebert, D., 2020. Genome-wide association
676 analysis identifies a genetic basis of infectivity in a model bacterial pathogen. Mol. Biol. Evol. 37,
677 3439–3452.
- 678 Ansari, M.A., et al., 2017. Genome-to-genome analysis highlights the effect of the human innate and
679 adaptive immune systems on the hepatitis C virus. Nat. Genet. 49, 666–673.
- 680 Ashby, B., Boots, M., 2017. Multi-mode fluctuating selection in host–parasite coevolution. Ecol. Lett. 20,
681 357–365.
- 682 Ashby, B., Iritani, R., Best, A., White, A., Boots, M., 2019. Understanding the role of eco-evolutionary
683 feedbacks in host-parasite coevolution. J. Theor. Biol. 464, 115–125.
- 684 Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguilera, G., Snirc, A., Le Prieur, S., Jeziorski, C.,
685 Branca, A., Giraud, T., 2017. Widespread selective sweeps throughout the genome of model plant
686 pathogenic fungi and identification of effector candidates. Mol. Ecol. 26, 2041–2062.
- 687 Bartha, I., et al., 2013. A genome-to-genome analysis of associations between human genetic variation,
688 HIV-1 sequence diversity, and viral control. eLife 2, e01123.
- 689 Bartoli, C., Roux, F., 2017. Genome-Wide Association Studies In Plant Pathosystems: Toward an
690 Ecological Genomics Approach. Front. Plant Sci. 8, 763.
- 691 Bento, G., Fields, P.D., Duneau, D., Ebert, D., 2020. An alternative route of bacterial infection associated
692 with a novel resistance locus in the *Daphnia-Pasteuria* host-parasite system. Heredity 125, 173–
693 183.
- 694 Bento, G., Routtu, J., Fields, P.D., Bourgeois, Y., Du Pasquier, L., Ebert, D., 2017. The genetic basis of
695 resistance and matching-allele interactions of a host-parasite system: The *Daphnia magna*-
696 *Pasteuria ramosa* model. PLoS Genet. 13, e1006596.
- 697 Bergelson, J., Kreitman, M., Stahl, E.A., Tian, D., 2001. Evolutionary dynamics of plant R-genes. Science
698 292, 2281–5.

- 699 Bourgeois, Y., Roulin, A.C., Müller, K., Ebert, D., 2017. Parasitism drives host genome evolution:
700 Insights from the *Pasteuria ramosa* - *Daphnia magna* system: Brief Communication. *Evolution*
701 71, 1106–1113.
- 702 Brown, J.K.M., 2015. Durable resistance of crops to disease: a Darwinian perspective. *Ann Rev*
703 *Phytopathol*, 53, 513–539.
- 704 Cabalzar, A.P., Fields, P.D., Kato, Y., Watanabe, H., Ebert, D., 2019.
705 Parasite-mediated selection in a natural metapopulation of *Daphnia magna*. *Mol. Ecol.* 28, 4770–
706 4785.
- 707 Charlesworth, D., 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS*
708 *Genet.* 2, e64.
- 709 Collins, C., Didelot, X., 2018. A phylogenetic method to perform genome-wide association studies in
710 microbes that accounts for population structure and recombination. *PLoS Comput. Biol.* 14,
711 e1005958.
- 712 Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian Computation
713 (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418.
- 714 Dawkins, R., Krebs, J.R., 1979. Arms races between and within species. *Proc. R. Soc. B.* 205, 489–511.
- 715 Decaestecker, E., Gaba, S., Raeymaekers, J.A.M., Stoks, R., Van Kerckhoven, L., Ebert, D., De Meester,
716 L., 2007. Host-parasite ‘Red Queen’ dynamics archived in pond sediment. *Nature* 450, 870–873.
- 717 De Vienne, D.M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M.E. and Giraud, T., 2013.
718 Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and
719 relation to coevolution. *New Phytol.*, 198(2), pp.347-385.
- 720 Ebert, D., 2008. Host-parasite coevolution: Insights from the *Daphnia*-parasite model system. *Curr. Opin.*
721 *Microbiol.* 11, 290–301.
- 722 Ebert, D., Fields, P.D., 2020. Host-parasite co-evolution and its genomic signature. *Nat. Rev. Genet.* 21,
723 754–768.
- 724 Fenton, A., Antonovics, J., Brockhurst, M.A., 2009. Inverse-gene-for-gene infection genetics and
725 coevolutionary dynamics. *Am. Nat.* 174, E230-42.
- 726 Feurtey, A., Gladieux, P., Hood, M.E., Snirc, A., Cornille, A., Rosenthal, L., Giraud, T., 2016. Strong
727 phylogeographic co-structure between the anther-smut fungus and its white campion host. *New*
728 *Phytol.* 212, 668–679.
- 729 Fitzpatrick, C.R., Salas-González, I., Conway, J.M., Finkel, O.M., Gilbert, S., Russ, D., Teixeira, P.J.P.L.,
730 Dangl, J.L., 2020. The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annu.*
731 *Rev. Microbiol.* 74, 81–100.
- 732 Frickel, J., Feulner, P.G.D., Karakoc, E., Becks, L., 2018. Population size changes and selection drive
733 patterns of parallel evolution in a host-virus system. *Nat. Comm.* 9, 1706.
- 734 Gokhale, C.S., Papkou, A., Traulsen, A., Schulenburg, H., 2013. Lotka-Volterra dynamics kills the Red
735 Queen: population size fluctuations and associated stochasticity dramatically change host-parasite
736 coevolution. *BMC Evol. Biol.* 13, 254.
- 737 Haag, K.L., Pombert, J.-F., Sun, Y., de Albuquerque, N.R.M., Batliner, B., Fields, P., Lopes, T.F., Ebert,
738 D., 2019. Microsporidia with vertical transmission were likely shaped by nonadaptive processes.
739 *Genome Biol. Evol.* 12, 3599–3614.
- 740 Haldane, J.B.S., Jayakar, S.D., 1963. Polymorphism due to selection depending on the composition of a
741 population. *J. Genet.* 58, 318–323.

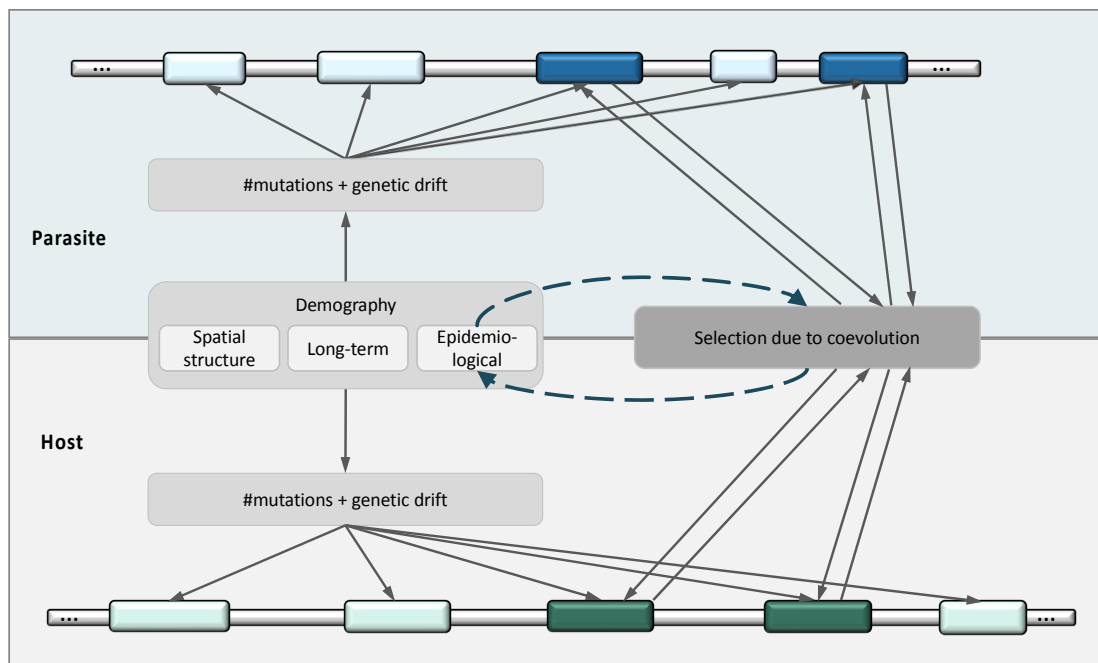
- 741 Hall, A.R., Scanlan, P.D., Morgan, A.D., Buckling, A., 2011. Host-parasite coevolutionary arms races
742 give way to fluctuating selection: Bacteria-phage coevolutionary dynamics. *Ecol. Lett.* 14, 635–
743 642.
- 744 Haller, B.C. and Messer, P.W., 2019. SLiM 3: forward genetic simulations beyond the Wright–Fisher
745 model. *Mol Biol Evol*, 36, 632–637.
- 746 Hartmann, F.E., Snirc, A., Cornille, A., Godé, C., Touzet, P., Van Rossum, F., Fournier, E., Le Prieur, S.,
747 Shykoff, J., Giraud, T., 2020. Congruent population genetic structures and divergence histories in
748 anther-smut fungi and their host plants *Silene italica* and the *Silene nutans* species complex. *Mol.*
749 *Ecol.* 29, 1154–1172.
- 750 Hecht, L.B.B., Thompson, P.C., Rosenthal, B.M., 2018. Comparative demography elucidates the
751 longevity of parasitic and symbiotic relationships. *Proc. R. Soc. B.* 285, 20181032.
- 752 Holub, E.B., 2001. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2,
753 516–27.
- 754 Janzen, D.H., 1980. When is it coevolution? *Evolution* 34, 611–612.
- 755 Krasovec, M., Chester, M., Ridout, K., Filatov, D.A., 2018. The mutation rate and the age of the sex
756 chromosomes in *Silene latifolia*. *Cur. Biol.* 28, 1832-1838.e4.
- 757 Lees, J.A., et al., 2019. Joint sequencing of human and pathogen genomes reveals the genetics of
758 pneumococcal meningitis. *Nat. Comm.* 10, 2176.
- 759 Leonard, K., 1993. Stability of Equilibria in a Gene-for-Gene Coevolution Model of Host-Parasite
760 Interactions. *Phytopathology* 84, 70–77.
- 761 Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences.
762 *Nature* 475, 493–496.
- 763 Luijckx, P., Fienberg, H., Duneau, D., Ebert, D., 2013. A matching-allele model explains host resistance
764 to parasites. *Cur. Biol.* 23, 1085–1088.
- 765 MacPherson, A., Keeling, M.J., Otto, S.P., 2021. Coevolution fails to maintain genetic variation in a host–
766 parasite model with constant finite population size. *Theor. Popul. Biol.* 137, 10–21.
- 767 MacPherson, A., Otto, S.P., Nuismer, S.L., 2018. Keeping Pace with the Red Queen: Identifying the
768 Genetic Basis of Susceptibility to Infectious Disease. *Genetics* 208, 779–789.
- 769 Märkle, H., Tellier, A., 2020. Inference of coevolutionary dynamics and parameters from host and parasite
770 polymorphism data of repeated experiments. *PLoS Comput. Biol.* 16, e1007668.
- 771 Märkle, H., Tellier, A., John, S., 2021. Cross-Species association statistics for genome-wide studies of
772 host and parasite polymorphism data. *bioRxiv* 726166. <https://doi.org/10.1101/726166>
- 773 Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., Delph, L.F., Van Rossum, F., 2017.
774 Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation
775 in allopatry. *Evolution* 71, 1519–1531.
- 776 May, R.M., Anderson, R.M., 1983. Epidemiology and genetics in the coevolution of parasites and hosts.
777 *Proc. R. Soc. B.* 219, 281–313.
- 778 Möller, M., Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens. *Nat.*
779 *Rev. Microbiol.* 15, 756–771.
- 780 Nemri, A., Atwell, S., Tarone, A.M., Huang, Y.S., Zhao, K., Studholme, D.J., Nordborg, M., Jones,
781 J.D.G., 2010. Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance
782 using combined association and linkage mapping. *PNAS* 107, 10302–10307.

- 783 Nuismer, S.L., Gomulkiewicz, R., Ridenhour, B.J., 2010. When is correlation coevolution? *Am Nat* 175,
784 525–37.
- 785 Nuismer, S.L., Jenkins, C.E., Dybdahl, M.F., 2017. Identifying coevolving loci using interspecific genetic
786 correlations. *Ecol. Evol.* 7, 6894–6903.
- 787 Nuismer, S.L., Week, B., 2019. Approximate Bayesian estimation of coevolutionary arms races. *PLoS*
788 *Comput. Biol.* 15, 1–21.
- 789 Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., Barg, M.-C., Rosenstiel, P.C.,
790 Teotónio, H., Schulenburg, H., 2019. The genomic basis of Red Queen dynamics during rapid
791 reciprocal host–pathogen coevolution. *PNAS* 116, 923–928.
- 792 Petit, E., Silver, C., Cornille, A., Gladieux, P., Rosenthal, L., Bruns, E., Yee, S., Antonovics, J., Giraud,
793 T., Hood, M.E., 2017. Co-occurrence and hybridization of anther-smut pathogens specialized on
794 *Dianthus* hosts. *Mol. Ecol.* 26, 1877–1890.
- 795 Petit-Houdenet, Y., Fudal, I., 2017. Complex Interactions between Fungal Avirulence Genes and Their
796 Corresponding Plant Resistance Genes and Consequences for Disease Resistance Management.
797 *Front. Plant. Sci.* 8, 1072.
- 798 Pflughoeft, K.J., Versalovic, J., 2012. Human Microbiome in Health and Disease. *Annu. Rev. Pathol.* 7,
799 99–122.
- 800 Poland, J.A., Balint-Kurti, P.J., Wissler, R.J., Pratt, R.C., Nelson, R.J., 2009. Shades of gray: the world of
801 quantitative disease resistance. *Trend Plant Sci.* 14, 21–29.
- 802 Retel, C., Kowallik, V., Huang, W., Werner, B., Künzel, S., Becks, L., Feulner, P.G.D., 2019a. The
803 feedback between selection and demography shapes genomic diversity during coevolution.
804 *Science Adv.* 10:eaax0530.
- 805 Retel, C., Märkle, H., Becks, L., Feulner, P.G.D., 2019b. Ecological and Evolutionary Processes Shaping
806 Viral Genetic Diversity. *Viruses* 11, 220.
- 807 Routtu, J., Ebert, D., 2015. Genetic architecture of resistance in *Daphnia* hosts against two species of host-
808 specific parasites. *Heredity* 114, 241–8.
- 809 Sanchez, T., Cury, J., Charpiat, G. and Jay, F., 2020. Deep learning for population size history inference:
810 Design, comparison and combination with approximate Bayesian computation. *Mol Ecol Res.* 00,
811 01-16.
- 812 Sasaki, A., Hamilton, W.D., Ubeda, F., 2002. Clone mixtures and a pacemaker: new facets of Red-Queen
813 theory and ecology. *Proc. R. Soc. B.* 269, 761–72. Sellinger, T.P.P., Abu Awad, D., Moest, M.,
814 Tellier, A., 2020. Inference of past demography, dormancy and self-fertilization rates from whole
815 genome sequence data. *PLoS Genet.* 16, e1008698.
- 816 Saur, I.M., Bauer, S., Kracher, B., Lu, X., Franzesakis, L., Müller, M.C., Sabelleck, B., Kümmler, F.,
817 Panstruga, R., Maekawa, T. and Schulze-Lefert, P., 2019. Multiple pairs of allelic MLA immune
818 receptor-powdery mildew AVR effectors argue for a direct recognition mechanism. *Elife*, 8,
819 p.e44471.
- 820 Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P.,
821 Xie, D., Günthard, H.F., Drummond, A.J., Bonhoeffer, S., 2012. Estimating the basic reproductive
822 number from viral sequence data. *Mol. Biol. Evol.* 29, 347–57.
- 823 Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M., Bergelson, J., 1999. Dynamics of disease resistance
824 polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature* 400, 667–71.

- 825 Stam, R., Silva-Arias, G.A., Tellier, A., 2019. Subsets of NLR genes show differential signatures of
826 adaptation during colonization of new habitats. *New Phytol.* 224, 367–379.
- 827 Stephan, W., 2019. Selective Sweeps. *Genetics* 211, 5–13.
- 828 Sunnåker, M., Busetto, A.G., Numminen, E., Corander, J., Foll, M., Dessimoz, C., 2013. Approximate
829 Bayesian computation. *PLoS Comput. Biol.* 9, e1002803.
- 830 Tellier, A., Brown, J.K.M., 2011. Spatial heterogeneity, frequency-dependent selection and polymorphism
831 in host-parasite interactions. *BMC Evol. Biol.* 11, 319.
- 832 Tellier, A., Brown, J.K.M., 2007. Stability of genetic polymorphism in host-parasite interactions. *Proc. R.
833 Soc. B.* 274, 809–17.
- 834 Tellier, A., Moreno-Gómez, S., Stephan, W., 2014. Speed of adaptation and genomic footprints of host-
835 parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68, 2211–24.
- 836 Thompson, J.N., 2005. *The Geographic Mosaic of Coevolution*. The University of Chicago Press.
- 837 Toenshoff, E.R., Fields, P.D., Bourgeois, Y.X., Ebert, D., 2018. The End of a 60-year Riddle:
838 Identification and Genomic Characterization of an Iridovirus, the Causative Agent of White Fat
839 Cell Disease in Zooplankton. *G3* 8, 1259–1272.
- 840 Toju, H., Yamamichi, M., Guimarães, P. R., Olesen, J. M., Mougi, A., Yoshida, T., Thompson, J. N. ,
841 2017. Species-rich networks and eco-evolutionary synthesis at the metacommunity level. *Nature
842 Ecology & Evolution*, 1(2), 1-11.
- 843 Toruño, T.Y., Stergiopoulos, I., Coaker, G., 2016. Plant-Pathogen Effectors: Cellular Probes Interfering
844 with Plant Defenses in Spatial and Temporal Manners. *Annu. Rev. Phytopathol.* 54, 419–441.
- 845 Van de Weyer, A. L., et al., 2019. A species-wide inventory of NLR genes and alleles in *Arabidopsis
846 thaliana*. *Cell*, 178(5), 1260-1272.
- 847 Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., Johannes, F., 2016. Methylome
848 evolution in plants. *Genome Biol.* 17, 264.
- 849 Wadi, L., Reinke, A.W., 2020. Evolution of microsporidia: An extremely successful group of eukaryotic
850 intracellular parasites. *PLoS Pathog.* 16, e1008276.
- 851 Wang, M., Roux, F., Bartoli, C., Huard-Chauveau, C., Meyer, C., Lee, H., Roby, D., McPeck, M.S.,
852 Bergelson, J., 2018. Two-way mixed-effects methods for joint association analysis using both host
853 and pathogen genomes. *PNAS* 115, E5440–E5449.
- 854 Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., Levin, B.R., 2002. Biological and
855 biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* 32, 569–77.
- 856 Živković, D., John, S., Verin, M., Stephan, W., Tellier, A., 2019. Neutral genomic signatures of host-
857 parasite coevolution. *BMC Evol. Biol.* 19, 230.

858
859

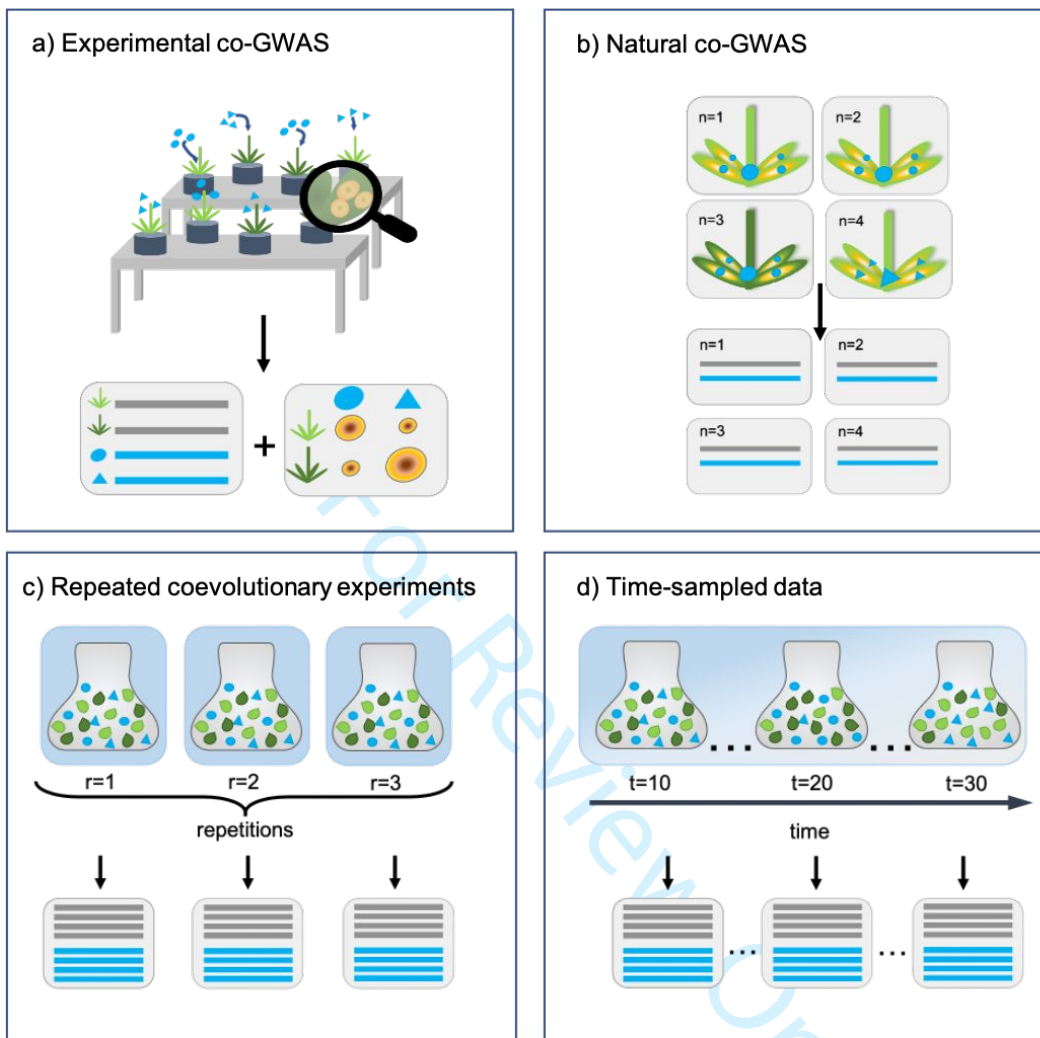
FIGURES



860

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

870



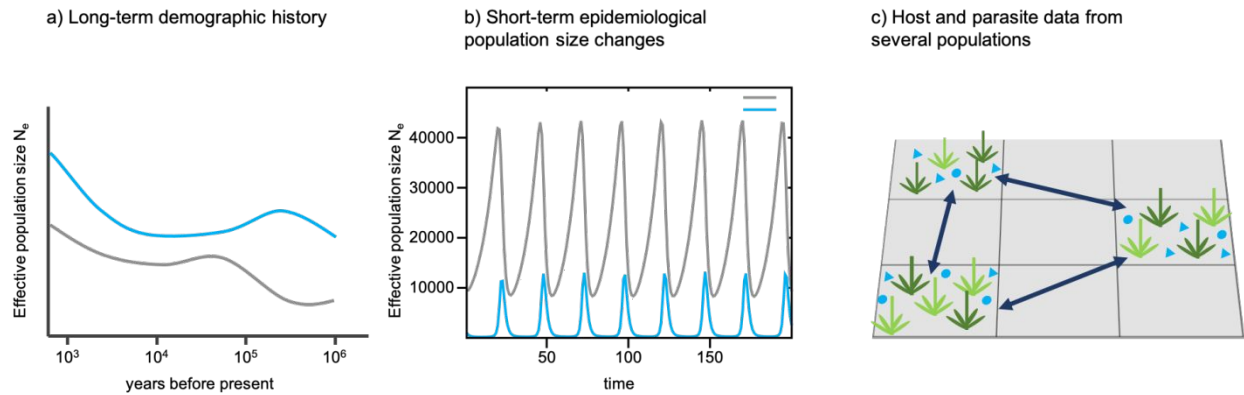
871

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

883 interaction conducted in controlled conditions is followed over time, and host and parasite samples taken at
884 a series of time points.

885

For Review Only



886

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

893

894

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

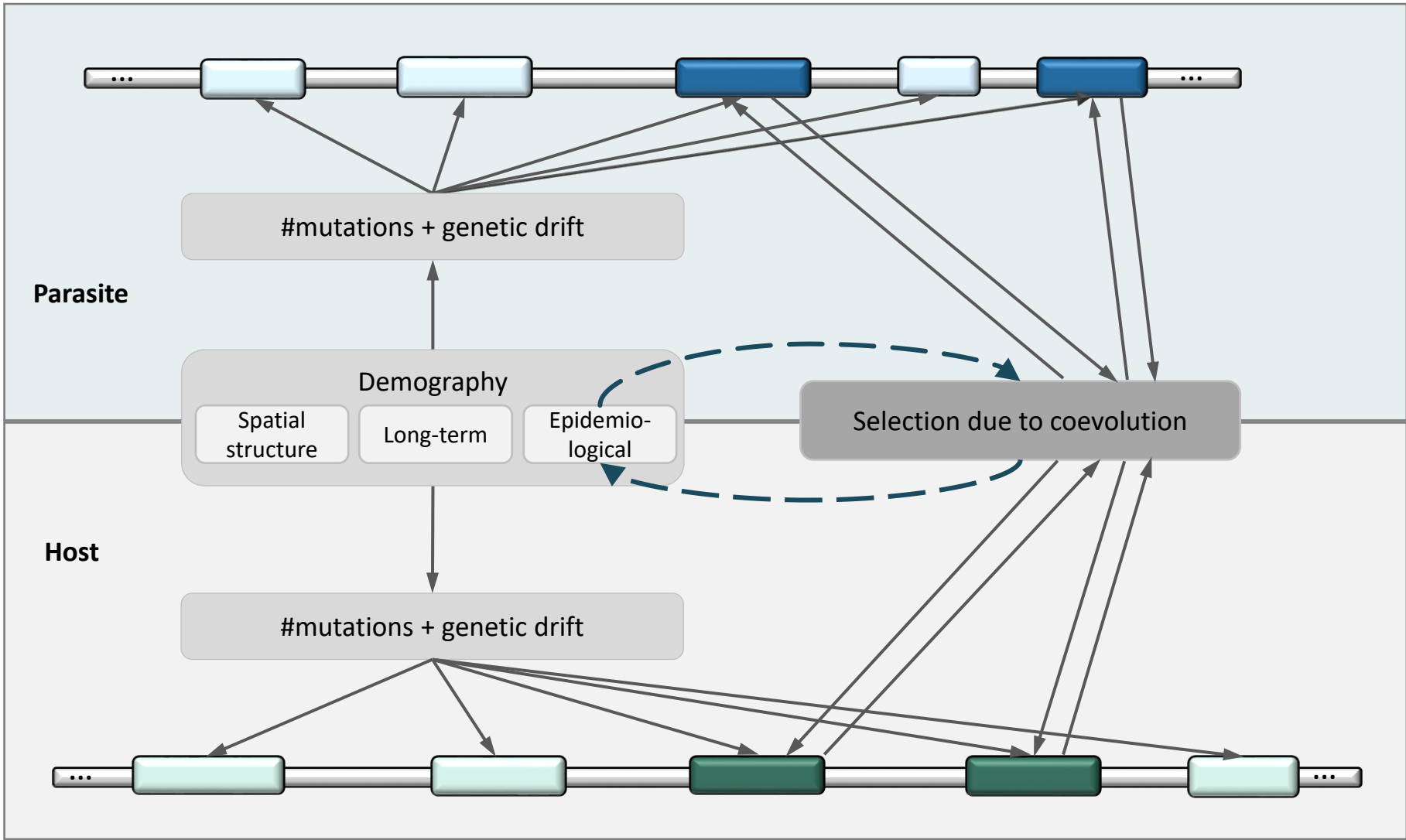
897

898

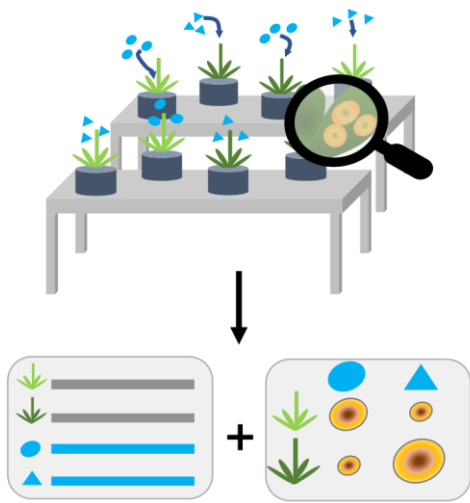
For Review Only

899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941

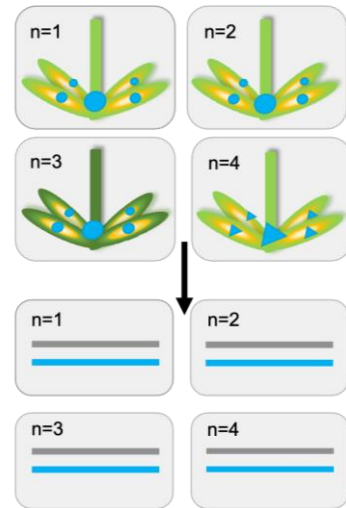
Method	Applicable to	Sample size needed	Advantages	Weaknesses	Remaining questions
Natural co-GWAs (Bartha et al., Ansari et al., Lees et al.)	Sample of infected host along with the associated parasite from natural populations. Possible to incorporate non-infected host data to the association statistic	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 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 <i>a priori</i> information on the past demography is required. Can highlight clear one-sided demography 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?



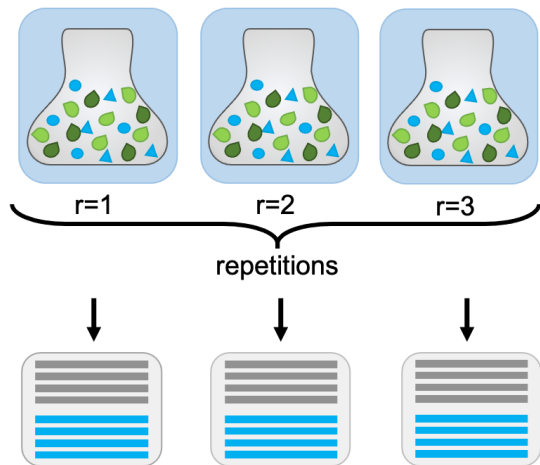
a) Experimental co-GWAS



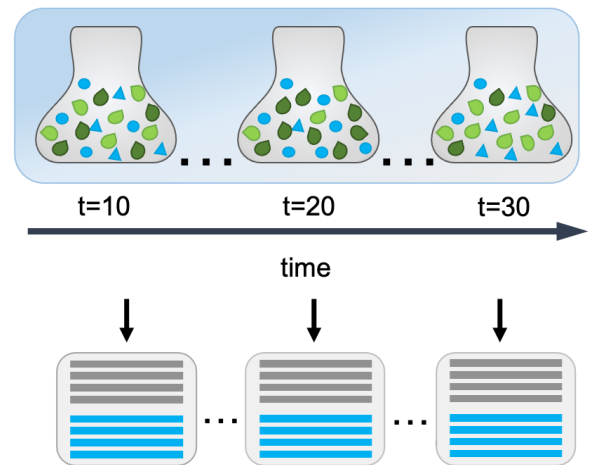
b) Natural co-GWAS



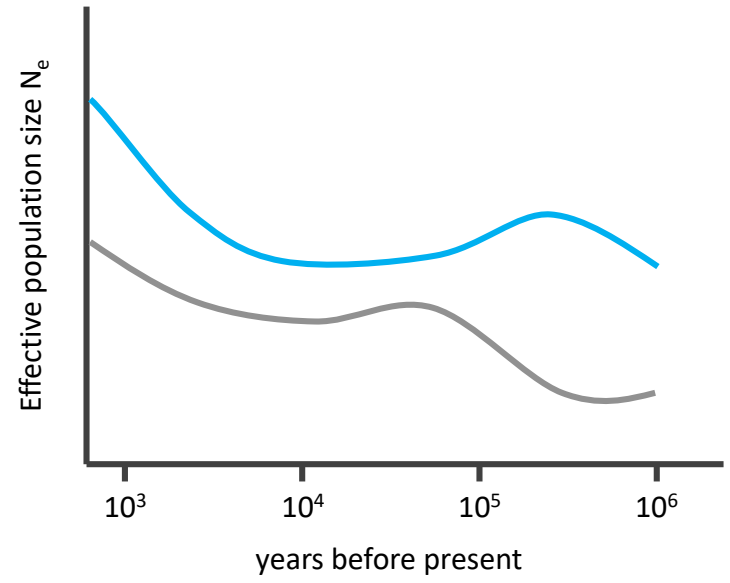
c) Repeated coevolutionary experiments



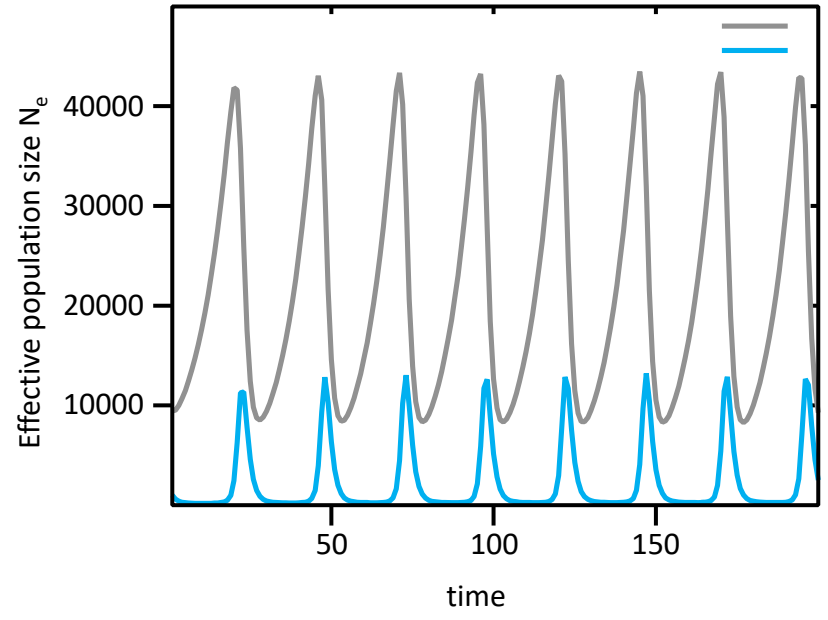
d) Time-sampled data



a) Long-term demographic history



b) Short-term epidemiological population size changes



c) Host and parasite data from several populations

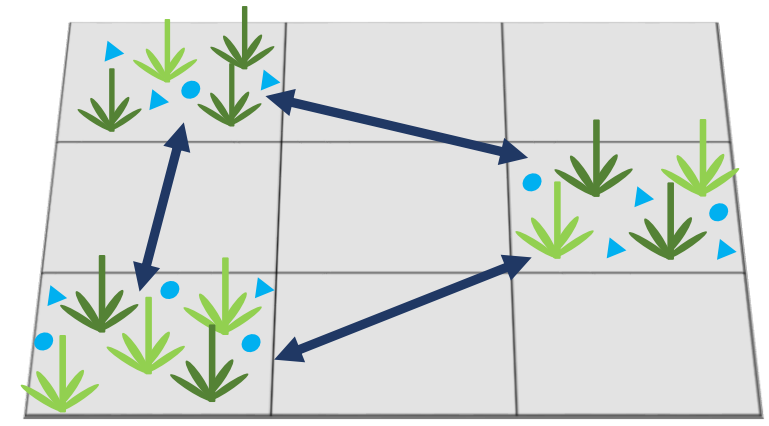


Table 1 Existing methods, the applicable data type, the sample size required (for host and parasite), advantages and weaknesses as well as open questions to date.

For Review Only

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 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 <i>a priori</i> 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?