Trends in Genetics

CellPress

Opinion

Unraveling coevolutionary dynamics using ecological genomics

Cornille Amandine [1,*,@], Dieter Ebert [2], Eva Stukenbrock [3], Ricardo C. Rodríguez de la Vega, [4] Peter Tiffin [5], Daniel Croll [6,*,@] and Aurélien Tellier [7,*]

Coevolutionary interactions, from the delicate co-dependency in mutualistic interactions to the antagonistic relationship of hosts and parasites, are a ubiquitous driver of adaptation. Surprisingly, little is known about the genomic processes underlying coevolution in an ecological context. However, species comprise genetically differentiated populations that interact with temporally variable abiotic and biotic environments. We discuss the recent advances in coevolutionary theory and genomics as well as shortcomings, to identify coevolving genes that take into account this spatial and temporal variability of coevolution, and propose a practical guide to understand the dynamic of coevolution using an ecological genomics lens.

What we know and do not know about coevolution

Species involved in symbioses have the potential to shape each other's evolutionary trajectory through reciprocal selection and adaptation (i.e., they have the potential to coevolve). Coevolution between two or several species is a fundamental mechanism shaping the organization and diversity of life [1–5]. Surprisingly, we know little about the genomic processes underlying coevolution, particularly in an ecological context. Genomic studies of coevolution have predominantly focused on only one of the partners, thus telling only half of the story (but see [6–9]). Furthermore, most genomic studies of coevolution have focused on single **populations** (see Glossary) [10–18] (but see [6,9]). However, species comprise genetically differentiated populations embedded in a network of interactions with complex and temporally varying abiotic and biotic environments. Local populations of species coevolve over short and long timescales in complex **habitats**. Thus, the coevolution of two (or more) species is expected to vary across space and time, and depends on the cumulated effects of ecological and evolutionary processes acting within and among populations. So far, the large set of population genomics and statistical association tools (applied to model and non-model species) available to understand the evolutionary processes involved in species diversification and adaptation [19–21] have seldom been used to investigate the genes underlying coevolution [22].

We argue that the lack of *ad hoc* inference and statistical approaches to study the complexity of coevolution at appropriate temporal and spatial scales hampers our understanding of the processes involved in coevolution and prevents the detection of genes underlying coevolution. This theoretical lag can be partly explained by the strong divide between the theoretical predictions of the **ecological dynamics** of coevolution over short timescales, which do not imply genomic perspectives [23–25], and the theory of the genomic consequences of **coevolutionary dynamics**, which would benefit from a deeper consideration of the ecological context (Boxes 1 and 2).

Here, we first summarize the main theoretical and empirical advances that allow the identification of the main processes shaping coevolution. We emphasize that understanding both the

Highlights

Coevolution is a fundamental process shaping species interactions and communities.

Coevolution has been mostly studied experimentally as an isolated process involving local reciprocal selection between two species. However, species comprise genetically differentiated populations across space in constant interaction with dynamic abiotic and biotic environments. Coevolution is therefore a dynamic equilibrium.

The genes and genomic processes underlying the complexity of coevolution over time and space remain poorly known.

A range of new theoretical developments, technological advances, and empirical approaches now allow coevolutionary dynamics to be investigated with genomic data from interacting species.

Recent advances in population genomics and genome-wide association studies will enable us to better understand the genetic basis of coevolutionary dynamics.

[1]Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE – Le Moulon, 91190 Gif-sur-Yvette, France
[2]Department of Environmental Sciences, Zoology, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
[3]Max Planck Institute for Terrestrial Microbiology, Max Planck Research Group, Fungal Biodiversity, Marburg, Germany
[4]Ecologie Systematique Evolution, Université Paris Saclay, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

[5]Department of Plant and Microbial Biology, 250 Biological Sciences, 1445 Gortner Ave., University of Minnesota, Saint Paul, MN 55108, USA
[6]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland
[7]Population Genetics, Department of Life Science Systems, Technical University of Munich, Liesel-Beckman-Str. 2, 85354 Freising, Germany

*Correspondence:
amandine.cornille@cnrs.fr (C. Amandine), daniel.croll@unine.ch (D. Croll), and aurelien.tellier@tum.de (A. Tellier).
@Twitter: @CornilleAmand (C. Amandine) and @danielcroll (D. Croll).

Box 1. Theoretical foundation of the genetic basis of coevolution

Theory predicts that the coevolutionary dynamics of allele frequencies in antagonistic interactions are situated between two extreme scenarios [2,22,38]. In 'arms race' dynamics, alleles at coevolving loci are repeatedly fixed, typically leading to signatures of positive selection. 'Trench warfare' dynamics, in which several alleles at coevolving loci are stably maintained, typically result in signatures of balancing selection. The genomic architecture of coevolution (i.e., the number, location, and effect of the genes involved) has been considered to be either Mendelian or quantitative [25]. In Mendelian models, the arms race occurs at one or few major genes. An arms race in a quantitative trait setup corresponds to the phenotypic difference model [25]. The quantitative model equivalent in trench warfare is the phenotypic matching model [25]. While these expectations provide a rationale for conducting selection scans on genes of major effect, the theoretical framework is ill suited for making inferences on the genomic footprint of coevolution in an ecological context. First, coevolutionary dynamics are affected by ecological and evolutionary factors [43,44,46,68]. This means that demographic processes due to finite population size [45] and changing population sizes due to eco-evolutionary feedback need to be accounted for when analyzing genome-wide patterns of polymorphism [43,44,46,68]. It has now been shown convincingly [44,48] that (negative indirect) frequency-dependent selection generated by host–parasite coevolution in itself does not guarantee the occurrence of long-term trench warfare dynamics [44,45]. Second, the complex and heterogeneous spatial structure of interacting species and the variable rates of gene flow and recombination across the genome need to be accounted for when detecting selection because these change the expectations regarding the footprints of coevolution (signatures of arms race or trench warfare) (reviewed in [22]).

demographic and the adaptive processes over space and time is critical to identify the genes involved in coevolution. Second, we review the recent advances in coevolutionary theory, as well as in population genomics and association mapping methods [i.e., genome-wide association studies (GWAS) and gene–environment-wide association studies (GEAS)] that have proved useful in understanding coevolutionary dynamics and the genes involved. We also discuss the need to develop suitable statistical methods to address coevolution in a more realistic manner (i.e., in its complexity) using genomic and phenotypic data. In a third part, we provide a practical guide (Box 3) for studies of coevolution using genomic data that aim to determine whether adaptations in interacting species truly stem from reciprocal selection [4,5,15,16]. We focus on host–parasite systems involving two species because data and theoretical expectations are primarily available for these systems. We believe that the theoretical and empirical developments currently tested on host–parasite systems, and presented below, could be applied to other symbiotic systems involving two or more species.

The three main processes shaping coevolution across space

Theory identifies three processes that shape coevolution (Box 1 and Figure 1, Key figure). The first is **demographic processes** [3,26,27] (Figure 1A), which include genetic drift affecting allele

Box 2. Co-demographic processes in host–parasite interaction

Coevolution is characterized by a variation in the hosts' and the parasites' fitness over time, which generates variation in population size within and between populations of a metapopulation (see Figure 1 and Box 1 in the main text [41,44,46,47,69]). Within a population, there is covariation in host and parasite population sizes over short timeframes due to the eco-evolutionary feedback (see Figure 1 in the main text), which is referred to as the co-demographic history [47]. The strength of the eco-evolutionary feedback is determined by the environmental effect on infection, epidemiology, and local density-dependent regulation [69]. Such rapid eco-evolutionary changes in population size (the N_e of both species) are observable using genomic data sampled in a time series and are thus likely to be relevant in the prediction of host–symbiont genome coevolution [58]. At the metapopulation level, field studies show that *Silene* or *Plantago lanceolata* populations exhibit strong metapopulation dynamics with frequent host and/or parasite extinction–recolonization influencing parasite persistence and population sizes [52]. To reveal these co-demographic events, a large dataset obtained across space and time is necessary; however, obtaining this may be difficult for many systems. The approach proposed in [68] may allow the fast reconstruction of the coevolutionary dynamics of interacting species using sequence data of species amenable to laboratory coevolution experiments or for which data can be sampled over a sufficient number of generations across several coevolutionary cycles. The sampling time scale is then defined by the generation time of hosts and parasites. For instance, time sampling is more amenable in annual or multiannual species than in perennials. In the future, genome-wide statistical methods should allow the inference of coevolving host and parasite allele trajectories across populations and habitats using samples at different time points (see Figure I in Box 3), while taking into account the past demography and spatial structure of each species and their co-demographic history. This would allow us to decipher the bases of G×G×E interactions using genomic scans of selection at high spatial and temporal resolution (see earlier).

frequencies within populations, **gene flow** between populations, and **metapopulation** dynamics (local population extinction and recolonization) [3]. In addition, interacting species may have different population structures and different levels of gene flow and be differentially affected by local (abiotic) environments. It is therefore crucial to consider the population structure of each interacting species when studying coevolution [28,29]. At the metapopulation level, populations of interacting species can also share elements of a past demographic history (e.g., common post-glacial expansion), which are relevant in understanding the processes of coevolution. Second, the outcome of interspecific interactions, that is, the genotype-by-genotype interaction (referred to as G×G hereafter; Figure 1B), defines directly the parameters of coevolution (strength of coevolution, disease severity, virulence of the parasite, fitness costs of resistance or infectivity). Third, within and across habitats, spatially variable selection (i.e., the **selection mosaic**) is characterized by the heterogeneity in biotic and abiotic interactions. This selection mosaic can, in its simplest form, be represented by a spatial matrix of **coevolutionary coldspots and hotspots** defined by the presence of one species (single-species evolution) or both interacting partners (coevolution), respectively (Figure 1B). For instance, in the mutualism between fig trees and fig wasps, there are only hotspots, as fig trees and fig wasps fully depend on each other. The spatial heterogeneity has also the potential to generate more complex patterns: genotype-by-genotype-by-environment interactions (referred to as G×G×E hereafter; Figure 1C). This means that the strength and speed of coevolution can vary across localities within and among habitats (Figure 1C). In other words, the importance of coevolution to the overall evolution of a population can vary across space. Low levels of gene flow can result in asynchrony of the coevolutionary dynamics between populations, while high levels of gene flow may homogenize gene pools and synchronize coevolutionary dynamics across space [24,30].

Given the complexity of variable selection and demographic processes (Figure 1D), it is crucial to assess the relative importance of demographic [i.e., the **effective population size (N_e)** and gene flow] versus adaptive (i.e., G×G, G×E, and G×G×E) processes in driving coevolution over space and time. Later, we summarize studies, which provide first glimpses into coevolutionary dynamics, that rely on two main approaches: (i) population genomics inference methods that take into account demographic processes to detect G×G interactions; and (ii) GWAS and GEAS, which provide insights into G×G×E interactions. We also discuss new or improved statistical methods in these fields that allow the study of more realistic, and hence more complex, scenarios.

How have population genomics and GWAS provided an insight into the dynamics of coevolution?

On demographic processes and co-spatial genetic structures

Inferring the demographic history of each species in each population and the spatial genetic structure of each species is important to draw accurate conclusions regarding the strength of coevolution [29]. This can be achieved using a population genomics approach based on a large set of genetic markers (preferably genome-wide polymorphism data) to simultaneously estimate N_e and its variation over time and space, the population genetic structure, and the extent of gene flow among populations of each species in interaction. The population structure of coevolving species has been characterized in only a few model systems. One is the anther smut fungus *Microbotryum violaceum* and its host, members of the plant genus *Silene* (specifically *Silene latifolia*). These species show strong spatial **genetic co-structure**, probably because the pollinator of *Silene* is responsible for both the long-distance dispersal of the smut fungus and long-distance gene flow in *Silene* [28,31]. In newts (*Taricha granulosa*) that are preyed on by garter snakes (*Thamnophis sirtalis*), the levels of newt (prey) toxin and garter snake (predator) resistance are tightly matched across the landscape. Although predator resistance is geographically structured according to signatures of **local adaptation** to prey, levels of prey toxin are structured following neutral population divergence [32]. This later study suggests that neutral processes, including gene flow, rather than reciprocal

Glossary

Coevolutionary dynamics: dynamics of genotype/trait/allele frequencies in interacting species due to the reciprocal nature of coevolution.
Coevolutionary hotspots/coldspots: locations where interspecific interactions are strong and reciprocal are defined as hotspots, whereas areas where population interactions are asymmetric or nonexistent are defined as coldspots.
Demographic processes: include population size fluctuations over time and space, which influences the variation in N_e and gene flow among populations. Changes in N_e (the demography history of a population) and gene flow influence the efficiency of selection and thus the strength of coevolution.
Ecological dynamics: changes in population size (or population density), here in the context of eco–evo feedbacks in host and parasite populations.
Effective population size (N_e): the number of individuals that effectively participate in producing the next generation. N_e determines the rate of change in the composition of a population caused by genetic drift (i.e., the random sampling of genetic variants in a finite population).
Gene flow: the migration of individuals, and thus of genes/alleles, between populations/demes.
Genetic co-structure: statistical congruence between the population genetic structures of interacting species.
Habitat: the set of local abiotic and biotic conditions experienced by one or several populations in a given geographic area.
Local adaptation: in spatially heterogeneous environments, evolution can lead to the adaptation of populations to their local environmental conditions. A pattern of local adaptation observed is when the mean fitness of a population in its home environment is higher than the mean fitness of populations from elsewhere [64].
Metapopulation: populations connected by gene flow. Individual populations may go extinct and new populations may be established by migrants.
Population: a group of individuals who are more genetically similar to each other than they are to individuals outside the subpopulation, as a result of genetic drift, migration, mutation, and selection; also called a 'deme' in the

adaptation explain most of the phenotypic variation of the two interacting species observed across the landscape. Two other studies that investigated the spatial co-structure of mutualist interactions [33] also detected spatial structure in either the host or the symbiont, but no spatial co-structure. (i) In [34], genetic differentiation in the legume *Medicago lupulina* was roughly concordant with the geographic turnover of its N-fixing bacterial symbionts *Ensifer* but only at the genus level. (ii) In [35], the leafcutter ant *Atta texana* showed variable levels of congruence with its two main fungal symbionts, most likely because the strength of drift and gene flow differs between the fungal partners. There is a need for additional studies of interacting species to quantify the importance of life-history traits such as host and parasite life span, parasite transmission, dormancy, selfing or clonal reproduction in shaping the spatial co-structure, and the coevolutionary dynamics within and between populations. New methods [36,37] can be used to infer divergence time, rate of gene flow, and N_e (and its variation in time; Box 2) while taking into account variable rates of gene flow along the genome. Such inferences are important in defining a neutral threshold (i.e., the consequences of demography) to identify the genes under selection in host and parasite genomes. Population genomics inference methods should also be developed to compare demographic histories, spatial structures, and gene flow in several species simultaneously (see recommendations in [19]; Box 3).

metapopulation framework and often assumed as a panmictic group inferred from Bayesian inference methods in population genomics studies [65–67].
Selection mosaic: spatial variation in the strength of coevolution (hot- and coldspots) among interacting species due to spatial variation in the biotic and selective pressure.

Inference of local selective process of coevolution *per se* (G×G)

Classical models that predict coevolving loci within populations rely on the identification of signatures of selective sweeps (i.e., arms race) or balanced polymorphisms (i.e., trench warfare dynamics, also called Red Queen dynamics) in the genomes [2,38] (Box 1). Few methods take into account the role of polygenic traits in coevolution [25,39,40]. Furthermore, the theoretical expectations regarding loci of major effect have been recently challenged by studies that take demographic processes (e.g., genetic drift) into account [41–45]. Specifically, trench warfare dynamics are predicted to be less likely than arms race dynamics when local population sizes are small (Box 1) [43–45]. This seems particularly true when population sizes vary over time due to eco-evolutionary feedback [43,44,46]. Nevertheless, trench warfare dynamics may occur in local populations as alleles from asynchronized populations are reintroduced by gene flow [47] or because of spatial heterogeneity among populations [48,49]. The signatures of trench warfare or arms races, especially when these dynamics occur over a brief period of time, are not necessarily detectable in the genome [43,45]. Theory also predicts that polymorphism signatures of coevolution within a population, especially signatures of selective sweeps due to arms race, are more likely to be detected in parasite than in host genomes, but only if parasites show sufficiently high rates of recombination [41]. Recently, host and parasite polymorphism data were jointly analyzed using a new framework: approximate Bayesian computation (ABC). This framework was used to infer the type of dynamics at coevolving loci, as well as to detect genes in coevolution and infer the parameters of coevolution (disease severity, effectiveness of resistance, etc.) [22]. This is an interesting new tool for understanding the dynamics of coevolution, but it also has drawbacks as it requires repeated sampling in space and time to infer the main parameters that define coevolution; namely, the reduction in host fitness due to infection [48,50].

Future approaches to identify coevolving genes from genomic data need to move beyond the analysis of single populations and the focus of coevolutionary dynamics only from major loci. New quantitative models could be based on phenotypic-matching and phenotypic-difference approaches and should take population heterogeneity into account (Box 1). At present, methods to detect genes under selection across populations rely on the identification of population differentiation outliers (e.g., F_{ST}, *XtX*). Statistical thresholds in such scans need to account for the neutral spatial structure and demographic history of each species. New approaches can provide more explicit demographic models based on the inferred co-structure and co-demographic history of coevolving pairs of species (Box 2).

Key figure

Major factors underlying coevolutionary dynamics in an ecological context

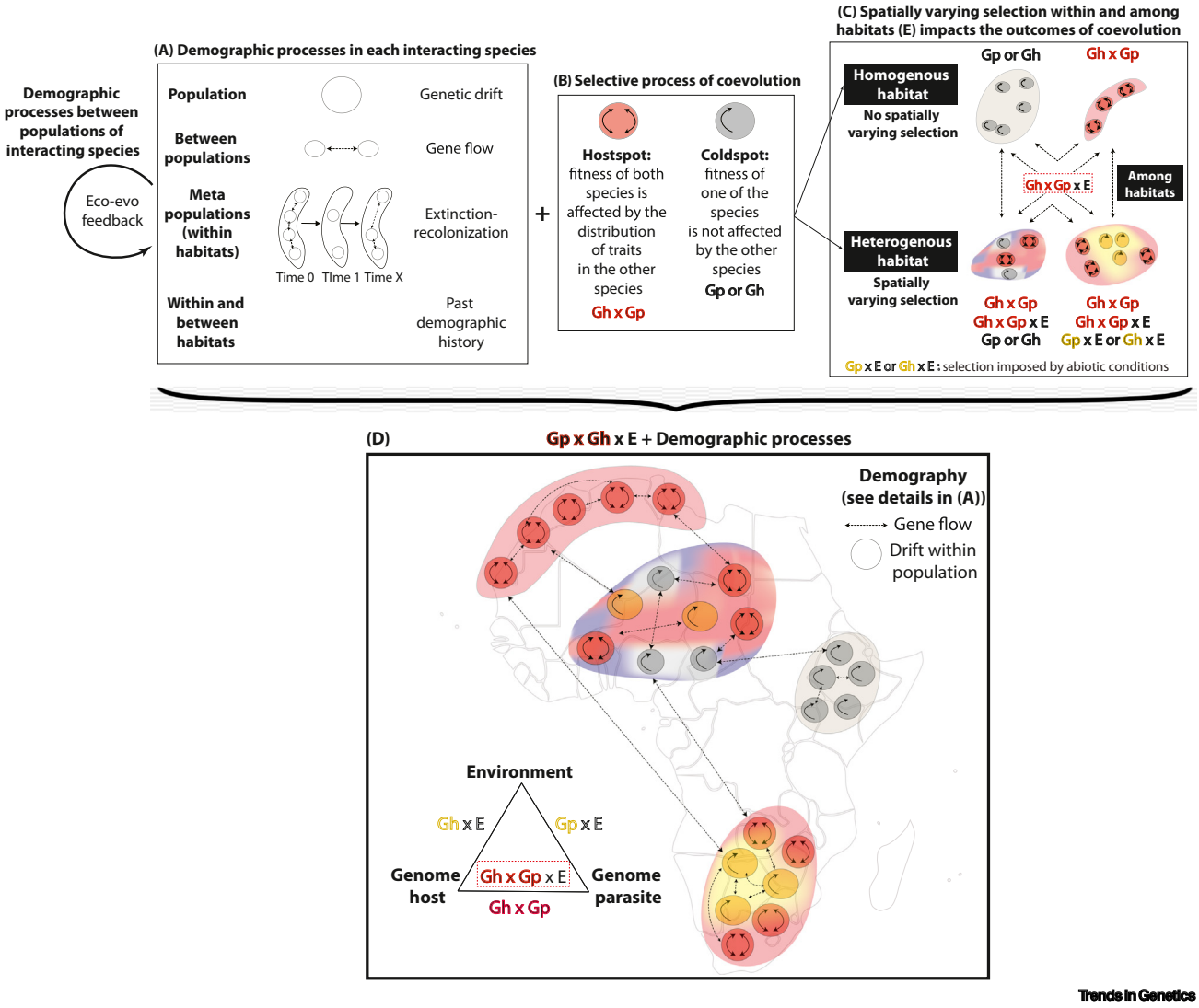

Figure 1. For simplicity, we focus on host–parasite (h-p) coevolution [3,83]. Species that interact in a given geographic area (i.e., habitat) are collected multiple times across a spatial gradient. Gene flow occurs among host or parasite populations and habitats. The three processes determining coevolutionary interactions in an ecological context using a population genomics lens are: (A) the demographic processes occurring within and between populations; (B) the process of coevolution (G×G) in its simplest representation – hotspot (G_h×G_p) or coldspot where the species co-occur but do not coevolve (G_p or G_h); (C) the outcomes of coevolution are impacted by spatially varying selection – in a habitat, the strength of the interaction (coldspot or hotspot) between host and parasite populations can vary depending on within-habitat heterogeneity (G_h×G_p, G_p, or G_h, G_h×E or G_h×E, G_p×G_h×E) as well as on between-habitat heterogeneity (G_p×G_h×E). Overall, the heterogeneity in biotic and abiotic interactions among populations and habitats (G×G×E) leads to a complex pattern of spatially heterogeneous selection. (D) Demographic processes [i.e., effective population size (N_e) and gene flow] and the adaptive processes driven by interactions or abiotic environments (i.e., G×G, G×E, and G×G×E) are major factors underlying co-evolutionary dynamics in an ecological context (i.e., over time and space).

Inference of spatially heterogeneous selection (G×G×E)

Selection pressure from abiotic and biotic conditions (e.g., temperature, humidity, food sources, competition) can be highly variable over space and time [3]. Hence, in addition to co-evolution depending on variation in the host and the parasite (G×G), spatial variation in species interactions can be affected by other sources of variation leading to G×G×E (Figure 1). The contribution of environmental conditions to overall selection depends on, among other factors, the closeness of the interaction. Intracellular parasites, for instance, are nearly exclusively exposed to selection pressures imposed by the host, although parasites of invertebrates and plants are also indirectly exposed to the outside environment since it determines the host's temperature. By contrast, parasites and mutualists with extended free-living periods face significant selection pressures other than factors associated with host resistance [4,51]. Parasites of annual

Box 3. A practical guide to the ecological genomics approaches for the investigation of coevolution

We propose the following steps to conduct in-depth coevolutionary studies (Figure I).

Hierarchical sampling (1 in Figure I)

Sample infected (and uninfected if present) hosts and their infecting parasites in several populations from the same habitat and in several different habitats. The information regarding which parasitic individual infects which host should be kept.

Sequencing (2 in Figure I)

Sequence individuals of both host and parasite and perform read mapping and variant calling. Ideally, high-quality reference genomes are available for analysis of gene families. Sequencing can involve full genomes or sequence capture of a few thousand genes including candidates for immunity, resistance, and pathogenicity.

Inferences of the demography and population structure for each species (3a in Figure I)

The variance of migration rate can be compared with information from recombination maps when available (or by simultaneous modeling of demographic history and inference of recombination maps using the integrative sequentially Markov coalescent (iSMC) method on full genome data [70]): correlation between the rate of gene flow and that of recombination depends on selection at introgressed genes (adaptive versus maladaptive, across populations and habitats). We also suggest using full-genome data to test for correlation between the demographic history of host populations and that of parasite populations using the comparative pairwise sequentially Markovian coalescent (C-SMC) method [71]. ABC [72] and machine-learning methods [73] can be used to infer the demographic history and the spatial structure parameters for more complex scenarios.

Selection scans (4a in Figure I)

Apply scans using the neutral demographic model and spatial population genetic structure inferred above to set statistical thresholds. It is possible to run selection scans on host (infected and noninfected) and parasite samples within a population, on a pool of populations from the same habitat, and on all populations from different habitats. This makes use of the power of hierarchical sampling to detect selection based on F_{ST} [74], *XtX* [75], or dimension reduction (e.g., principal component analyses [76], genotype matrices) methods as well as on classic selective sweep or balancing selection tests.

Sampling over time (3b and 4b in Figure I)

With appropriate sampling, it is possible to test the co-demographic history of hosts and parasites based on the site frequency spectrum of hosts and pathogens at different time points [68]. The number and periodicity of the time sampling depend on the life cycle of the host and the parasite. For instance, time sampling is more amenable in annual or multiannual species than in perennials. Based on simulations of neutral background changes in allele frequencies, it is possible to study genes under coevolution with outlier allele frequencies over time [77].

Co-GWAS and co-GEAS (5a in Figure I)

Using the full genomes or sequence capture data from hosts and parasites, we suggest conducting a co-GWAS for organisms for which it is doable [26,78], from which cross-species association indices can be used to identify genes under significant association [20,50], followed by the inference of coevolutionary parameters at major loci [74]. If experiments are possible, performing all possible pairwise infection tests and then using a co-GWAS is powerful in identifying the genes underlying the interaction [40]. It is advisable to perform co-GWAS under differing environmental conditions to account for G×G×E effects or to use GEAS to link genetic variation of hosts and parasites to environmental variables [59] (e.g., climate).

Functional validation (5b in Figure I)

If controlled experiments are possible, the number of candidate loci involved in coevolution can be obtained from co-expression analyses (joint RNA-seq [79] or proteomics) of host and symbiont samples. Coexpressed genes can be compared with those identified by co-GWAS and selection scans. Cross-species expression quantitative trait loci (eQTL) mapping-combined genomes and gene expression can identify polymorphisms in the host genome affecting parasite expression and vice versa [80]. New tools to determine protein structure (e.g., AlphaFold, RoSetta [81,82]) may also deepen our insights into the molecular basis of protein interactions and coevolution. Functional assays to validate these candidate genes and proteins at the population level in the host and the parasite is the final step that can be done together with population genomics approaches.

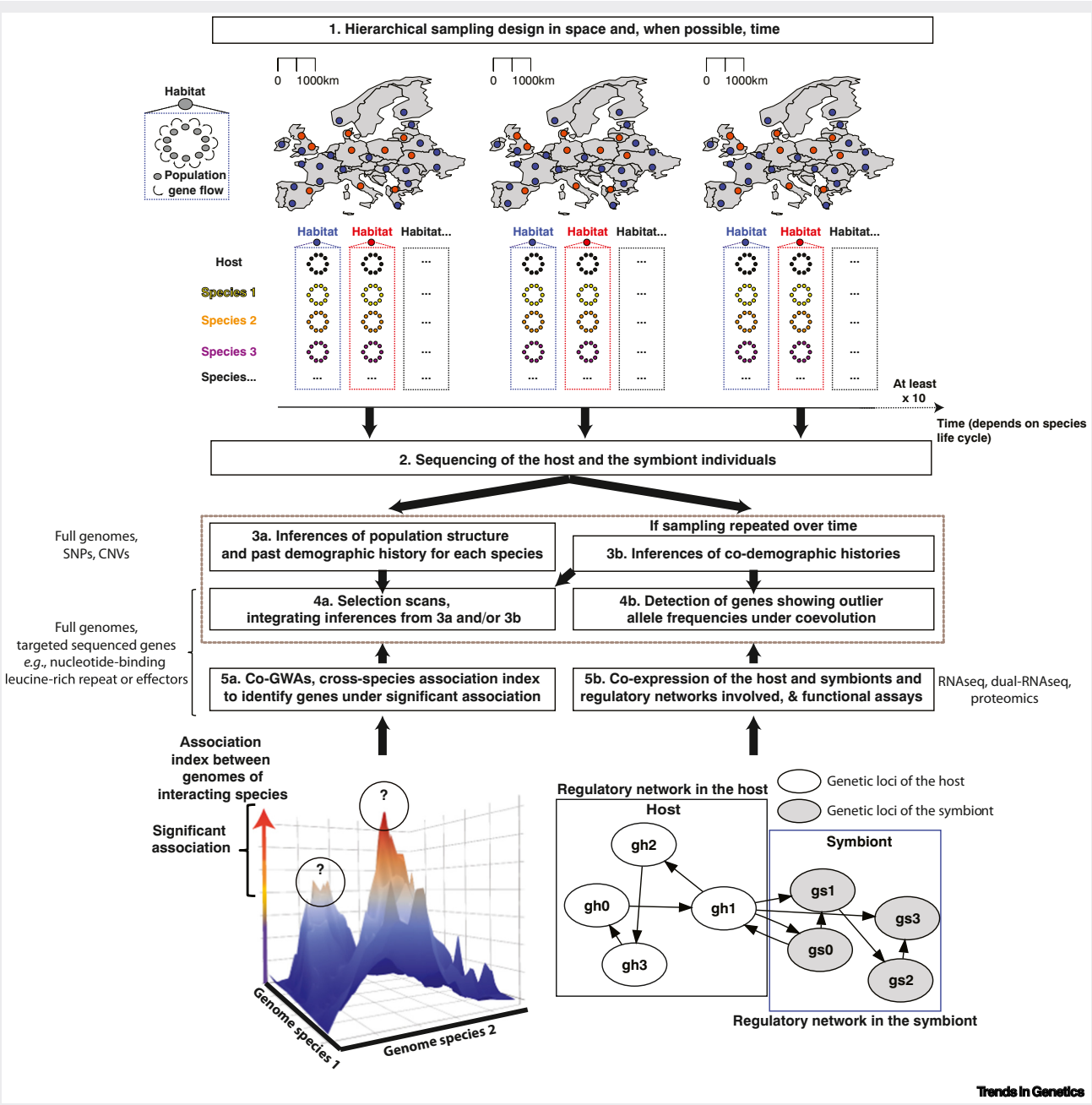

Figure I. Practical guide for future studies of coevolutionary dynamics using ecological genomics. Abbreviation: CNV, copy number variation; GWA, genome-wide association study; RNA-seq, RNA sequencing.

crops (and other annual plant species) must survive extended periods in a host-free environment before the host becomes available. Hence, tolerance to desiccation, solar radiation, and possibly alternative modes of nutrition that are independent of a host may be important [52]. Some parasites, such as fungal pathogens or aphids, also rely on an alternative host to complete their life cycle [53,54]. For instance, the devastating stem rust pathogen of wheat

requires an alternative host, barberry, for sexual reproduction, thus adding further host-dependent selection on the parasite. Crop pathogens also show local adaptation to climatic conditions and the presence of chemicals (i.e., fungicides applied to protect crops; see later).

G×G×E can generate local adaptations and potentially constrain the evolution of exploitative strategies (e.g., pathogenicity) and in turn alter the coevolutionary dynamics. Furthermore, even though a trade-off between traits in the host (or parasite) can occur in the absence of environmental variation, the magnitude and direction of the trade-offs can show large variance across habitats and thus potentially shed light on the genetic bases of G×G×E interactions. For instance, fungicides sprayed on crops select for pathogen populations with reduced growth rates in the absence of fungicides [55]. Hence, fungicides impose a shift towards a new trait optimum. Pathogen populations often vary in such trait optima [55]. GWAS are well suited to identify loci with genetic variants associated with these trait variations [56]. In a major wheat pathogen, a systematic investigation of environment-dependent trade-offs among adaptive traits across the life cycle revealed significant constraints [57]. Importantly, adaptation to different hosts showed trade-offs with environmental adaptation, in this case to temperature and fungicides [57]. Although traits related to pathogenicity on different host genotypes were largely correlated, the pathogen faced a specific trade-off between killing host plant tissue necessary to acquire nutrients and producing spores for dispersal. The exact molecular mechanisms underlying these trade-offs remain largely unknown. In a recent pioneering study, Roberts et *al.* [58] demonstrated that in the lepidopteran species *Plodia interpunctella* and its DNA virus PiGV, variation in the architecture of the resistance of the host to PiGV depends on the available resources (i.e., the abiotic environment). GWAS in controlled conditions can be used to reveal the genetic basis of the trade-offs between coevolving genes but also between coevolving genes and other metabolic pathways, while providing empirical data for the population genomic analyses. Furthermore, GEAS can identify genes and alleles of importance in G×G×E interactions, while taking as covariables environmental data (e.g., climate) across habitats [59]. However, in all association studies, distinguishing the effects of mutations that are directly responsible for the phenotype from correlated mutations (i.e., from linkage disequilibrium) remains an issue [60]. We argue that systematic investigation of G×G×E will prove crucial in deciphering the genetic architecture of trade-offs across habitats and in predicting the coevolutionary trajectories of interacting species across heterogeneous landscapes (Box 3).

Concluding remarks and future perspectives: a framework to decipher the genetic bases of coevolution

Advances in technology and statistical approaches will open new avenues to understand coevolution from an ecological and genetic perspective (see Outstanding questions). In the near future, methods to jointly analyze genomic data from hosts and parasites across populations and habitats should become available. We propose a practical guide to conducting such analyses based on existing methods (Box 3). Note that we focus on host–parasite interactions because these are the best-studied systems (but see [6]). However, research on coevolution is increasingly trying to understand how networks rather than just pairs of interacting species coevolve [1–5]. The next challenge will be to develop the statistical and theoretical toolbox that can integrate population genomics and association mapping in different ecological settings; that is, that can integrate a large parameter space with several species at the community level while retaining statistical power. Solving these issues will go a long way to address the question of whether adaptation in coevolving species truly stems from coevolution [4,5,25,26] and will improve our understanding of the evolution of complex communities [1,4,61–63].

Outstanding questions

What and how many genes are involved in coevolution (i.e., what is the genetic architecture underlying coevolution)?

What are the evolutionary trajectories of these genes (type of selection) within and between populations and habitats and can these be inferred from patterns of genomic polymorphism?

Are the same genes/types of genes involved across (i) host or parasite species, (ii) habitats of a given symbiont species (host or parasite), or (iii) antagonistic or mutualistic systems from the same geographic location (trophic network)? What is the genetic basis of trade-offs in host and parasite populations?

How can the theory and genomic tools developed for the study of two-species interactions be adapted to multispecies interactions at the community level?

Acknowledgments

We thank the European Society of Evolutionary Biology for funding the Special Topic Network 'Ecological genomics of coevolution' (https://eseb.org/prizes-funding/special-topic-networks/stn-current-networks/ecological-genomics-of-coevolutionary-interactions/). A.C. receives funding from the ATIP-Avenir CNRS Inserm (France) program. A.T. receives funding from the Deutsche Forschungsgemeinschaft grant TE809/3 (project 274542535) within the SPP 1819 'Rapid evolutionary adaptation: potential and constraints'. D.C. is supported by the Fondation Pierre Mercier pour la Science. D.C. and D.E. are supported by the Swiss National Science Foundation.

Declaration of interests

None are declared.

References

1. Koskella, B. *et al.* (2017) The microbiome beyond the horizon of ecological and evolutionary theory. *Nat. Ecol. Evol.* 1, 1606–1615
2. Ebert, D. and Fields, P.D. (2020) Host–parasite co-evolution and its genomic signature. *Nat. Rev. Genet.* 21, 754–768
3. Thompson, J.N. (2005) *The geographic mosaic of coevolution*, University of Chicago Press
4. Hall, A.R. *et al.* (2020) Measuring coevolutionary dynamics in species-rich communities. *Trends Ecol. Evol.* 35, 539–550
5. Bergelson, J. *et al.* (2021) Functional biology in its natural context: a search for emergent simplicity. *Elife* 10, e67646
6. Yoder, J.B. (2016) Understanding the coevolutionary dynamics of mutualism with population genomics. *Am. J. Bot.* 103, 1742–1752
7. Nallu, S. *et al.* (2018) The molecular genetic basis of herbivory between butterflies and their host plants. *Nat. Ecol. Evol.* 2, 1418–1427
8. Zhang, X. *et al.* (2020) Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* 183, 875–889.e17
9. Grillo, M.A. *et al.* (2016) Intrapopulation genomics in a model mutualist: population structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution* 70, 2704–2717
10. Moran, N.A. *et al.* (2008) Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190
11. Oliver, K.M. *et al.* (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* 55, 247–266
12. Wernegreen, J.J. (2015) Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann. N. Y. Acad. Sci.* 1360, 16–35
13. Sabater-Muñoz, B. *et al.* (2017) Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. *ISME J.* 11, 1291–1304
14. Wernegreen, J.J. (2017) In it for the long haul: evolutionary consequences of persistent endosymbiosis. *Curr. Opin. Genet. Dev.* 47, 83–90
15. Guyomar, C. *et al.* (2018) Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. *Microbiome* 6, 181
16. Williams, L.E. and Wernegreen, J.J. (2015) Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini. *PeerJ* 3, e881
17. Ravcheev, D.A. and Thiele, I. (2017) Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Front. Genet.* 8, 111
18. Delaux, P.-M. *et al.* (2014) Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet.* 10, e1004487
19. Weigel, D. and Nordborg, M. (2015) Population genomics for understanding adaptation in wild plant species. *Annu. Rev. Genet.* 49, 315–338
20. Bourgeois, Y.X.C. and Warren, B.H. (2021) An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Mol. Ecol.* 30, 6036–6071
21. Forester, B.R. *et al.* (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* 27, 2215–2233
22. Märkle, H. *et al.* (2021) Novel genomic approaches to study antagonistic coevolution between hosts and parasites. *Mol. Ecol.* 30, 3660–3676
23. Boots, M. *et al.* (2009) The role of ecological feedbacks in the evolution of host defence: what does theory tell us? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 27–36
24. Gandon, S. *et al.* (2008) Host–parasite coevolution and patterns of adaptation across time and space. *J. Evol. Biol.* 21, 1861–1866
25. Week, B. and Nuismer, S.L. (2019) The measurement of coevolution in the wild. *Ecol. Lett.* 22, 717–725
26. Bartoli, C. and Roux, F. (2017) Genome-wide association studies in plant pathosystems: toward an ecological genomics approach. *Front. Plant Sci.* 8, 763
27. Gomulkiewicz, R. *et al.* (2007) Dos and don'ts of testing the geographic mosaic theory of coevolution. *Heredity* 98, 249–258
28. Feurtey, A. *et al.* (2016) Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. *New Phytol.* 212, 668–679
29. Croll, D. and Laine, A.-L. (2016) What the population genetic structures of host and pathogen tell us about disease evolution. *New Phytol.* 212, 537–539
30. Nuismer, S.L. *et al.* (2017) Identifying coevolving loci using interspecific genetic correlations. *Ecol. Evol.* 7, 6894–6903
31. Hartmann, F.E. *et al.* (2020) Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants *Silene italica* and the *S. nutans* species complex. *Mol. Ecol.* 29, 1154–1172
32. Hague, M.T.J. *et al.* (2020) The geographic mosaic of arms race coevolution is closely matched to prey population structure. *Evol. Lett.* 4, 317–332
33. Daubech, B. *et al.* (2017) Spatio-temporal control of mutualism in legumes helps spread symbiotic nitrogen fixation. *Elife* 6, e28683
34. Harrison, T.L. *et al.* (2017) Geographically structured genetic variation in the *Medicago lupulina–Ensifer* mutualism. *Evolution* 71, 1787–1801
35. Smith, C.C. *et al.* (2019) Landscape genomics of an obligate mutualism: concordant and discordant population structures between the leafcutter ant *Atta texana* and its two main fungal symbiont types. *Mol. Ecol.* 28, 2831–2845
36. Roux, C. *et al.* (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14, e2000234
37. Haller, B.C. and Messer, P.W. (2019) SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* 36, 632–637
38. Moller, M. and Stukenbrock, E.H. (2017) Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* 15, 756–771
39. Bartha, I. *et al.* (2013) A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife* 2, e01123

40. Wang, M. *et al.* (2018) Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5440
41. Schenk, H. *et al.* (2020) How long do Red Queen dynamics survive under genetic drift? A comparative analysis of evolutionary and eco-evolutionary models. *BMC Evol. Biol.* 20, 8
42. Czuppon, P. and Gokhale, C.S. (2018) Disentangling eco-evolutionary effects on trait fixation. *Theor. Popul. Biol.* 124, 93–107
43. MacPherson, A. *et al.* (2018) Keeping pace with the Red Queen: identifying the genetic basis of susceptibility to infectious disease. *Genetics* 208, 779
44. MacPherson, A. *et al.* (2021) Feedback between coevolution and epidemiology can help or hinder the maintenance of genetic variation in host–parasite models. *Evolution* 75, 582–599
45. Tellier, A. *et al.* (2014) Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68, 2211–2224
46. Gokhale, C.S. *et al.* (2013) Lotka–Volterra dynamics kills the Red Queen: population size fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC Evol. Biol.* 13, 254
47. Thrall, P.H. and Burdon, J.J. (2002) Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathol.* 51, 169–184
48. Brown, J.K.M. and Tellier, A. (2011) Plant–parasite coevolution: bridging the gap between genetics and ecology. *Annu. Rev. Phytopathol.* 49, 345–367
49. Gavrilets, S. and Michalakis, Y. (2008) Effects of environmental heterogeneity on victim–exploiter coevolution. *Evolution* 62, 3100–3116
50. Märkle, H. and Tellier, A. (2020) Inference of coevolutionary dynamics and parameters from host and parasite polymorphism data of repeated experiments. *PLoS Comput. Biol.* 16, e1007668
51. Burghardt, L.T. *et al.* (2018) Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2425
52. Antonovics, J. (2004) Long-term study of a plant–pathogen metapopulation. In *Ecology, genetics and evolution of metapopulations* (Hanski, I. and Gaggiotti, O.E., eds), pp. 471–488, Academic Press
53. Simon, J.-C. *et al.* (2015) Genomics of adaptation to host–plants in herbivorous insects. *Brief. Funct. Genomics* 14, 413
54. Peterson, P.D. (2018) The barberry eradication program in Minnesota for stem rust control: a case study. *Annu. Rev. Phytopathol.* 56, 203–223
55. Lendenmann, M.H. *et al.* (2015) QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genet. Biol.* 80, 53–67
56. Sánchez-Vallet, A. *et al.* (2018) Nature's genetic screens: using genome-wide association studies for effector discovery. *Mol. Plant Pathol.* 19, 3–6
57. Dutta, A. *et al.* (2021) Mapping the adaptive landscape of a major agricultural pathogen reveals evolutionary constraints across heterogeneous environments. *ISME J.* 15, 1402–1419
58. Roberts, D. and Hamann, A. (2015) Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proc. Biol. Sci.* 282, 20142903
59. Frachon, L. *et al.* (2019) Adaptation to plant communities across the genome of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 36, 1442–1456
60. Chebib, J. and Guillaume, F. (2021) Pleiotropy or linkage? Their relative contributions to the genetic correlation of quantitative traits and detection by multitrait GWA studies. *Genetics* 219, iyab159
61. Henry, L.P. *et al.* (2019) Can the microbiome influence host evolutionary trajectories? *bioRxiv* Published online July 14, 2019. https://doi.org/10.1101/700237
62. Borer, E.T. *et al.* (2022) Disease-mediated nutrient dynamics: coupling host–pathogen interactions with ecosystem elements and energy. *Ecol. Monogr.* 92, e1510
63. O'Brien, A.M. *et al.* (2021) Whose trait is it anyways? Coevolution of joint phenotypes and genetic architecture in mutualisms. *Proc. Biol. Sci.* 288, 20202483
64. Williams, G. (1966) *Adaptation and natural selection*, Princeton University Press
65. Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959
66. Raj, A. *et al.* (2014) fastSTRUCTURE: variational inference of population structure in large SNP datasets. *Genetics* 197, 573–589
67. Alexander, D.H. *et al.* (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664
68. Zivković, D. *et al.* (2019) Neutral genomic signatures of host–parasite coevolution. *BMC Evol. Biol.* 19, 230
69. Ashby, B. *et al.* (2019) Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution. *J. Theor. Biol.* 464, 115–125
70. Barroso, V.G. *et al.* (2019) Inference of recombination maps from a single pair of genomes and its application to ancient samples. *PLoS Genet.* 15, e1008449
71. Hecht, L.B.B. *et al.* (2018) Comparative demography elucidates the longevity of parasitic and symbiotic relationships. *Proc. Biol. Sci.* 285, 20181032
72. Fraimout, A. *et al.* (2017) Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. *Mol. Biol. Evol.* 34, 980–996
73. Sanchez, T. *et al.* (2020) Deep learning for population size history inference: design, comparison and combination with approximate Bayesian computation. *Mol. Ecol. Resour.* 21, 2645–2660
74. de Villemereuil, P. and Gaggiotti, O.E. (2015) A new F_{ST}-based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* 6, 1248–1258
75. Gautier, M. (2015) Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 201, 1555–1579
76. Duforet-Frebourg, N. *et al.* (2015) Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Mol. Biol. Evol.* 33, 1082–1093
77. Foll, M. *et al.* (2015) WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol. Ecol. Resour.* 15, 87–98
78. Ansari, M.A. *et al.* (2017) Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat. Genet.* 49, 666–673
79. Westermann, A.J. *et al.* (2017) Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathog.* 13, e1006033
80. Guo, Y. *et al.* (2017) Networks underpinning symbiosis revealed through cross-species eQTL mapping. *Genetics* 206, 2175–2184
81. Baek, M. *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876
82. Jumper, J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589
83. Gomulkiewicz, R. *et al.* (2000) Hot spots, cold spots, and the geographic mosaic theory of coevolution. *Am. Nat.* 156, 156–174