



Unraveling coevolutionary dynamics using ecological genomics

Cornille Amandine, Dieter Ebert, Eva Stukenbrock, Ricardo Rodríguez de la Vega, Peter Tiffin, Daniel Croll, Aurélien Tellier

► To cite this version:

Cornille Amandine, Dieter Ebert, Eva Stukenbrock, Ricardo Rodríguez de la Vega, Peter Tiffin, et al.. Unraveling coevolutionary dynamics using ecological genomics. Trends in Genetics, 2022, 38 (10), pp.1003-1012. <10.1016/j.tig.2022.05.008>. <hal-03803855>

HAL Id: hal-03803855

<https://hal.science/hal-03803855v1>

Submitted on 6 Oct 2022

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.



HAL Authorization

Unraveling coevolutionary dynamics using ecological genomics

Cornille Amandine¹, Dieter Ebert², Eva Stukenbrock³, Ricardo C. Rodríguez de la Vega⁴, Peter Tiffin⁵, Daniel Croll⁶, Aurélien Tellier⁷

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, St 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 (@CornilleAmand), daniel.croll@unine.ch

(@danielcroll), tellier@wzw.tum.de

24 *Author ORCIDs*

25 Cornille Amandine ID <https://orcid.org/0000-0002-5348-7081>

26 Dieter Ebert ID <https://orcid.org/0000-0003-2653-3772>

27 Eva Stukenbrock ID <https://orcid.org/0000-0001-8590-3345>

28 Ricardo Rodriguez de la Vega ID <https://orcid.org/0000-0002-9852-6654>

29 Daniel Croll ID <https://orcid.org/0000-0002-2072-380X>

30 Peter Tiffin ID <https://orcid.org/0000-0003-1975-610X>

31 Aurélien Tellier ID <https://orcid.org/0000-0002-8895-0785>

32 **Abstract (101 words)**

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

41

42

43 **Keywords:** coevolution, population genomics, association mapping, species interaction, ecology,
44 adaptation.

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. Indeed, 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 [10–18] (but see [6,9]). However, species are composed of 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-time scales 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 for understanding 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 time scales, 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 demographic and adaptive processes over space and time is critical for identifying 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 proving useful for 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, Figure 1). The first one is **demographic processes** [3,26,27] (Figure 1a), which include genetic drift affecting allele frequencies within populations, gene flow between populations and **metapopulation** dynamics (local population extinction and recolonization) [3]. In addition, interacting species may have

91 different population structures, different levels of gene flow and be differentially affected by local
92 (abiotic) environments. It is therefore crucial to consider the population structure of each
93 interacting species when studying coevolution [28,29]. At the metapopulation level, populations
94 of interacting species can also share elements of a past demographic history (*e.g.*, common post-
95 glacial expansion), which are relevant for understanding the processes of coevolution. Second, the
96 outcome of inter-specific interactions, *i.e.*, the genotype-by-genotype interactions (referred to as
97 GxG hereafter, Figure 1b) defines directly the parameters of coevolution (strength of coevolution,
98 disease severity, virulence of the parasite, fitness costs of resistance or infectivity). Third, within
99 and across **habitats**, spatially variable selection, *i.e.*, the **selection mosaic**, is characterized by the
100 heterogeneity in biotic and abiotic interactions. This selection mosaic can, in its simplest form, be
101 represented by a spatial matrix of **coevolutionary coldspots and hotspots** defined by the presence
102 of one species (single species evolution) or both interacting partners (coevolution), respectively.
103 For instance, in the mutualism between fig trees and fig wasps, there are only hotspots as fig trees
104 and fig wasps fully depend on each other. The spatial heterogeneity has also the potential to
105 generate more complex patterns: genotype-by-genotype-by-environment interactions (referred to
106 as GxGxE hereafter, Figure 1c). This means that the strength and speed of coevolution can vary
107 across localities within and among habitats (Figure 1c). In other words, the importance of
108 coevolution to the overall evolution of a population can vary across space. Low levels of gene flow
109 can result in asynchrony of the **coevolutionary dynamics** between populations while high levels
110 of gene flow may homogenize gene pools and synchronize coevolutionary dynamics across space
111 [24,30].

112 Given the complexity of variable selection and demographic processes (Figure 1d), it is
113 crucial to assess the relative importance of demographic (*i.e.*, **effective population size N_e** and

gene flow) *vs.* adaptive (*i.e.*, GxG, GxE and GxGxE) processes in driving coevolution over space and time. Below, we summarize studies which provide first glimpses into coevolutionary dynamics, that rely on two main approaches: 1) population genomics inference methods that take into account demographic processes to detect GxG interactions, and 2) genome-wide association studies (GWAS and GEAS), which provide insights into GxGxE interactions. We also discuss new or improved statistical methods in these fields that allow the study of more realistic, hence more complex, scenarios.

How have population genomics and genome-wide association studies provided an insight into the dynamics of coevolution?

a. 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 the **effective population size** (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 only in a few model systems. One of these is the anther smut fungus *Microbotryum violaceum* and its host, members of the plant genus *Silene* (specifically *S. 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 adaptation explain most of the phenotypic variation of the two interacting species observed across the landscape. Two other studies that investigated spatial co-structure of mutualist interactions [33] also detected spatial structure in either the host or the symbiont, but no spatial co-structure: 1) 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; 2) 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 for defining a neutral threshold (*i.e.*, the consequences of demography) in order 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).

b. Inference of local selective process of coevolution *per se* (GxG)

159 Classical models that predict coevolving loci within populations rely on the identification of
160 signatures of selective sweeps (*i.e.*, arms race) or balanced polymorphisms (*i.e.*, trench warfare
161 dynamics, also called Red Queen dynamics) in the genomes [2,38] (Box 1). Few methods take into
162 account the role of polygenic traits in coevolution [25,39,40]. Furthermore, the theoretical
163 expectations regarding loci of major effect have been recently challenged by studies that take
164 demographic processes (*e.g.*, genetic drift) into account [41–45]. Specifically, trench warfare
165 dynamics are predicted to be less likely than arms race dynamics when local population sizes are
166 small (Box 1) [43–45]. This seems particularly true when population sizes vary over time due to
167 eco-evolutionary feedback [43,44,46]. Nevertheless, trench warfare dynamics may occur in local
168 populations as alleles from asynchronized populations are reintroduced by gene flow [47] or
169 because of spatial heterogeneity among populations [48,49]. The signatures of trench warfare or
170 arms races, especially when these dynamics occur over a brief period of time, are not necessarily
171 detectable in the genome [43,45]. Theory also predicts that polymorphism signatures of
172 coevolution within a population, especially signatures of selective sweeps due to arms race, are
173 more likely to be detected in parasite than in host genomes, but only if parasites show sufficiently
174 high rates of recombination [41]. Recently, host and parasite polymorphism data were jointly
175 analyzed using a new framework: approximate Bayesian computation (ABC hereafter). This
176 framework was used to infer the type of dynamics at coevolving loci, as well as detect genes in
177 coevolution and infer the parameters of coevolution (disease severity, effectiveness of resistance,
178 etc.) [22]. This is an interesting new tool for understanding the dynamics of coevolution, but it also
179 has drawbacks as it requires repeated sampling in space and time in order to infer the main
180 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).

c. Inference of spatially heterogeneous selection (GxGxE)

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 coevolution depending on variation in the host and the parasite (GxG), spatial variation in species interactions can be affected by other sources of variation leading to GxGxE (Figure 1). The contribution of environmental conditions to overall selection depends, amongst other factors, on 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 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

204 nutrition that are independent of a host may be important [52]. Some parasites, such as fungal
205 pathogens or aphids, also rely on an alternative host to complete their life cycle [53,54]. For
206 instance, the devastating stem rust pathogen of wheat requires an alternative host, barberry, for
207 sexual reproduction, thus adding further host-dependent selection on the parasite. Crop pathogens
208 also show local adaptation to climatic conditions and the presence of chemicals (*i.e.*, fungicides
209 applied to protect crops, see below).

210 GxGxE can generate local adaptations and potentially constrain the evolution of
211 exploitative strategies (*e.g.*, pathogenicity) and in turn alter the coevolutionary dynamics.
212 Furthermore, even though a trade-off between traits in the host (or parasite) can occur in the
213 absence of environmental variation, the magnitude and direction of the trade-offs can show large
214 variance across **habitats**, and thus potentially shed light on the genetic bases of GxGxE
215 interactions. For instance, fungicides sprayed on crops select for pathogen populations with
216 reduced growth rates in absence of fungicides [55]. Hence, fungicides impose a shift towards a
217 new trait optimum. Pathogen populations often vary in such trait optima [55]. GWAS are well
218 suited to identify loci with genetic variants associated with these trait variations [56]. In a major
219 wheat pathogen, a systematic investigation of environment-dependent trade-offs among adaptive
220 traits across the life-cycle revealed significant constraints [57]. Importantly, adaptation to different
221 hosts showed trade-offs with environmental adaptation, in this case to temperature and fungicides
222 [57]. Although traits related to pathogenicity on different host genotypes were largely correlated,
223 the pathogen faced a specific trade-off between killing host plant tissue necessary to acquire
224 nutrients and producing spores for dispersal. The exact molecular mechanisms underlying these
225 trade-offs remain largely unknown. In a recent pioneering study, Roberts *et al.* [58] demonstrated
226 that in the lepidopteran species (*Plodia interpunctella*) and its DNA virus (PiGV), variation in the

architecture of resistance of the host to PiGV depends on 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 studies can identify genes and alleles of importance in GxGxE interactions, while taking as co-variables 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 GxGxE will prove crucial for deciphering the genetic architecture of trade-offs across **habitats**, and for predicting the coevolutionary trajectories of interacting species across heterogeneous landscapes (Box 3, Figure 2).

A framework to decipher the genetic bases of coevolution

Advances in technology and statistical approaches will open new avenues understanding coevolution from an ecological and genetic perspective. 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 for carrying out such analyses based on existing methods (Box 3, Figure 2). 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, *i.e.*, 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 in addressing 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].

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/>). AC receives funding from the ATIP-Avenir CNRS Inserm program. AT receives funding from the Deutsche Forschungsgemeinschaft grant TE809/3 (project 274542535) within the SPP 1819 “Rapid evolutionary adaptation: potential and constraints”. DC is supported by the Fondation Pierre Mercier pour la science. DC and DE are supported by the Swiss National Science Foundation.

References

- 1 Koskella, B. *et al.* (2017) The microbiome beyond the horizon of ecological and evolutionary theory. *Nature Ecology & Evolution* 1, 1606–1615
- 2 Ebert, D. and Fields, P.D. (2020) Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics* 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 in Ecology & Evolution* DOI: <https://doi.org/10.1016/j.tree.2020.02.002>

272 5 Bergelson, J. *et al.* (2021) Functional biology in its natural context: A search for emergent
273 simplicity. *eLife* 10, e67646

274 6 Yoder, J.B. (2016) Understanding the coevolutionary dynamics of mutualism with
275 population genomics. *American Journal of Botany* 103, 1742–1752

276 7 Nallu, S. *et al.* (2018) The molecular genetic basis of herbivory between butterflies and their
277 host plants. *Nature Ecology & Evolution* 2, 1418–1427

278 8 Zhang, X. *et al.* (2020) Genomes of the banyan tree and pollinator wasp provide insights into
279 fig-wasp coevolution. *Cell* 183, 875–889.e17

280 9 Grillo, M.A. *et al.* (2016) Intrapopulation genomics in a model mutualist: Population
281 structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution*
282 70, 2704–2717

283 10 Moran, N.A. *et al.* (2008) Genomics and evolution of heritable bacterial symbionts. *Annual*
284 *Review of Genetics* 42, 165–190

285 11 Oliver, K.M. *et al.* (2010) Facultative symbionts in aphids and the horizontal transfer of
286 ecologically important traits. *Annu Rev Entomol* 55,

287 12 Wernegreen, J.J. (2015) Endosymbiont evolution: predictions from theory and surprises from
288 genomes. *Annals of the New York Academy of Sciences* 1360, 16–35

289 13 Sabater-Muñoz, B. *et al.* (2017) Chance and necessity in the genome evolution of
290 endosymbiotic bacteria of insects. *The ISME Journal* 11, 1291–1304

291 14 Wernegreen, J.J. (2017) In it for the long haul: evolutionary consequences of persistent
292 endosymbiosis. *Current Opinion in Genetics & Development* 47, 83–90

293 15 Guyomar, C. *et al.* (2018) Multi-scale characterization of symbiont diversity in the pea aphid
294 complex through metagenomic approaches. *Microbiome* 6, 181

295 16 Williams, L.E. and Wernegreen, J.J. (2015) Genome evolution in an ancient bacteria-ant
296 symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe
297 Camponotini. *PeerJ* 3, e881

298 17 Ravcheev, D.A. and Thiele, I. (2017) Comparative genomic analysis of the human gut
299 microbiome reveals a broad distribution of metabolic pathways for the degradation of host-
300 synthesized mucin glycans and utilization of mucin-derived monosaccharides. *Frontiers in*
301 *Genetics* 8, 111

302 18 Delaux, P.-M. *et al.* (2014) Comparative phylogenomics uncovers the impact of symbiotic
303 associations on host genome evolution. *PLOS Genetics* 10, 1–12

304 19 Weigel, D. and Nordborg, M. (2015) Population genomics for understanding adaptation in
305 wild plant species. *Annual Review of Genetics* 49, 315–338

306 20 Bourgeois, Y.X.C. and Warren, B.H. (2021) An overview of current population genomics
307 methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular*
308 *Ecology* 0, 1–36

309 21 Forester, B.R. *et al.* (2018) Comparing methods for detecting multilocus adaptation with
310 multivariate genotype–environment associations. *Molecular Ecology* 27, 2215–2233

311 22 Märkle, H. *et al.* (2021) Novel genomic approaches to study antagonistic coevolution
312 between hosts and parasites. *Molecular Ecology* 30, 3660–3676

313 23 Boots, M. *et al.* (2009) The role of ecological feedbacks in the evolution of host defence:
314 what does theory tell us? *Philosophical Transactions of the Royal Society B: Biological*
315 *Sciences* 364, 27–36

316 24 Gandon, S. *et al.* (2008) Host–parasite coevolution and patterns of adaptation across time
317 and space. *Journal of Evolutionary Biology* 21, 1861–1866

318 25 Week, B. and Nuismer, S.L. (2019) The measurement of coevolution in the wild. *Ecology*
319 *Letters* 22, 717–725

320 26 Bartoli, C. and Roux, F. (2017) Genome-wide association studies in plant pathosystems:
321 toward an ecological genomics Approach. *Frontiers in Plant Science* 8, 763

322 27 Gomulkiewicz, R. *et al.* (2007) Dos and don'ts of testing the geographic mosaic theory of
323 coevolution. *Heredity* 98, 249–258

324 28 Feurtey, A. *et al.* (2016) Strong phylogeographic co-structure between the anther-smut
325 fungus and its white campion host. *New Phytologist* 212, 668–679

326 29 Croll, D. and Laine, A.-L. (2016) What the population genetic structures of host and
327 pathogen tell us about disease evolution. *New Phytologist* 212, 537–539

328 30 Nuismer, S.L. *et al.* (2017) Identifying coevolving loci using interspecific genetic
329 correlations. *Ecology and Evolution* 7, 6894–6903

330 31 Hartmann, F.E. *et al.* (2020) Congruent population genetic structures and divergence
331 histories in anther-smut fungi and their host plants *Silene italica* and the *S. nutans* species
332 complex. *Molecular Ecology* 29, 1154–1172

333 32 Hague, M.T.J. *et al.* (2020) The geographic mosaic of arms race coevolution is closely
334 matched to prey population structure. *Evolution Letters* 4, 317–332

335 33 Daubech, B. *et al.* (2017) Spatio-temporal control of mutualism in legumes helps spread
336 symbiotic nitrogen fixation. *eLife* 6, e28683

337 34 Harrison, T.L. *et al.* (2017) Geographically structured genetic variation in the *Medicago*
338 *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. *Molecular Ecology* 28, 2831–2845
- 36 Roux, C. *et al.* (2016) Shedding Light on the grey zone of speciation along a continuum of genomic divergence. *PLOS Biology* 14, e2000234
- 37 Haller, B.C. and Messer, P.W. (2019) SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution* 36, 632–637
- 38 Moller, M. and Stukenbrock, E.H. (2017) Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* advance online publication, 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. *Proceedings of the National Academy of Sciences* 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 Evolutionary Biology* 20, 8
- 42 Csuppon, P. and Gokhale, C.S. (2018) Disentangling eco-evolutionary effects on trait fixation. *Theoretical Population Biology* 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

362 45 Tellier, A. *et al.* (2014) Speed of adaptation and genomic footprints of host-parasite
363 coevolution under arms race and trench warfare dynamics. *Evolution* 68, 2211–2224

364 46 Gokhale, C.S. *et al.* (2013) Lotka–Volterra dynamics kills the Red Queen: population size
365 fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC*
366 *Evolutionary Biology* 13, 254

367 47 Thrall, P.H. and Burdon, J.J. (2002) Evolution of gene-for-gene systems in metapopulations:
368 the effect of spatial scale of host and pathogen dispersal. *Plant Pathology* 51, 169–184

369 48 Brown, J.K.M. and Tellier, A. (2011) Plant-parasite coevolution: Bridging the gap between
370 genetics and ecology. *Annu. Rev. Phytopathol.* 49, 345–367

371 49 Gavrillets, S. and Michalakakis, Y. (2008) Effects of environmental heterogeneity on victim-
372 exploiter coevolution. *Evolution* 62, 3100–3116

373 50 Märkle, H. and Tellier, A. (2020) Inference of coevolutionary dynamics and parameters from
374 host and parasite polymorphism data of repeated experiments. *PLOS Computational Biology*
375 16, e1007668

376 51 Burghardt, L.T. *et al.* (2018) Select and resequence reveals relative fitness of bacteria in
377 symbiotic and free-living environments. *Proceedings of the National Academy of Sciences*
378 115, 2425

379 52 Antonovics, J. (2004) Long-term study of a plant-pathogen metapopulation. In *Ecology,*
380 *Genetics and Evolution of Metapopulations* (Hanski, I. and Gaggiotti, O. E., eds), pp. 471–
381 488, Academic Press

382 53 Simon, J.-C. *et al.* (2015) Genomics of adaptation to host-plants in herbivorous insects.
383 *Briefings in Functional Genomics* 14, 413

384 54 Peterson, P.D. (2018) The Barberry Eradication Program in Minnesota for Stem Rust
385 Control: A Case Study. *Annual Review of Phytopathology* 56, 203–223

386 55 Lendenmann, M.H. *et al.* (2015) QTL mapping of fungicide sensitivity reveals novel genes
387 and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genetics and*
388 *Biology* 80, 53–67

389 56 Sánchez-Vallet, A. *et al.* (2018) Nature’s genetic screens: using genome-wide association
390 studies for effector discovery. *Molecular Plant Pathology* 19, 3–6

391 57 Dutta, A. *et al.* (2021) Mapping the adaptive landscape of a major agricultural pathogen
392 reveals evolutionary constraints across heterogeneous environments. *The ISME Journal* 15,
393 1402–1419

394 58 Roberts, D. and Hamann, A. (2015) Glacial refugia and modern genetic diversity of 22
395 western North American tree species. *Proceedings. Biological sciences / The Royal Society*
396 282,

397 59 Frachon, L. *et al.* (2019) Adaptation to plant communities across the genome of *Arabidopsis*
398 *thaliana*. *Molecular Biology and Evolution* 36, 1442–1456

399 60 Chebib, J. and Guillaume, F. (2021) Pleiotropy or linkage? Their relative contributions to the
400 genetic correlation of quantitative traits and detection by multitrait GWA studies. *Genetics*
401 219, iyab159

402 61 Henry, L.P. *et al.* (2019) Can the microbiome influence host evolutionary trajectories?
403 *bioRxiv* DOI: 10.1101/700237

404 62 Borer, E.T. *et al.* (2022) Disease-mediated nutrient dynamics: Coupling host-pathogen
405 interactions with ecosystem elements and energy. *Ecological Monographs*
406 <https://doi.org/10.1002/ecm.1510>, e1510

407 63 O'Brien, A.M. *et al.* (2021) Whose trait is it anyways? Coevolution of joint phenotypes and
 408 genetic architecture in mutualisms. *Proceedings of the Royal Society B: Biological Sciences*
 409 288, 20202483
 410 64 Williams, G. (1966) *Adaptation and natural selection*, Princeton (NJ): Princeton University
 411 Press.
 412 65 Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data.
 413 *Genetics* 155, 945–959
 414 66 Raj, A. *et al.* (2014) fastSTRUCTURE: Variational inference of population structure in large
 415 SNP datasets. *Genetics* at
 416 <http://www.genetics.org/content/early/2014/04/14/genetics.114.164350.abstract>
 417 67 Alexander, D.H. *et al.* (2009) Fast model-based estimation of ancestry in unrelated
 418 individuals. *Genome Research* 19, 1655–1664
 419 68 Zivković, D. *et al.* (2019) Neutral genomic signatures of host-parasite coevolution. *BMC*
 420 *Evolutionary Biology* 19, 230
 421 69 Ashby, B. *et al.* (2019) Understanding the role of eco-evolutionary feedbacks in host-parasite
 422 coevolution. *Journal of Theoretical Biology* 464, 115–125
 423 70 Barroso, V.G. *et al.* (2019) Inference of recombination maps from a single pair of genomes
 424 and its application to ancient samples. *PLOS Genetics* 15, e1008449
 425 71 Hecht, L.B.B. *et al.* (2018) Comparative demography elucidates the longevity of parasitic
 426 and symbiotic relationships. *Proceedings of the Royal Society B: Biological Sciences* 285,
 427 20181032
 428 72 Fraimout, A. *et al.* (2017) Deciphering the routes of invasion of *Drosophila suzukii* by means
 429 of ABC random forest. *Molecular Biology and Evolution* 34, 980–996

430 73 Sanchez, T. *et al.* (2020) Deep learning for population size history inference: Design,
431 comparison and combination with approximate Bayesian computation. *Molecular Ecology*
432 *Resources* 0, 1–16

433 74 de Villemereuil, P. and Gaggiotti, O.E. (2015) A new F_{ST} -based method to uncover local
434 adaptation using environmental variables. *Methods in Ecology and Evolution* 6, 1248–1258

435 75 Gautier, M. (2015) Genome-wide scan for adaptive divergence and association with
436 population-specific covariates. *Genetics* 201, 1555–1579

437 76 Duforet-Frebourg, N. *et al.* (2015) Detecting genomic signatures of natural selection with
438 principal component analysis: Application to the 1000 genomes data. *Molecular Biology and*
439 *Evolution* 33, 1082–1093

440 77 Foll, M. *et al.* (2015) WFABC: a Wright–Fisher ABC-based approach for inferring effective
441 population sizes and selection coefficients from time-sampled data. *Molecular Ecology*
442 *Resources* 15, 87–98

443 78 Ansari, M.A. *et al.* (2017) Genome-to-genome analysis highlights the effect of the human
444 innate and adaptive immune systems on the hepatitis C virus. *Nature Genetics* 49, 666–673

445 79 Westermann, A.J. *et al.* (2017) Resolving host–pathogen interactions by dual RNA-seq.
446 *PLOS Pathogens* 13, e1006033

447 80 Guo, Y. *et al.* (2017) Networks underpinning symbiosis revealed through cross-species
448 eQTL Mapping. *Genetics* 206, 2175–2184

449 81 Baek Minkyung *et al.* (2021) Accurate prediction of protein structures and interactions using
450 a three-track neural network. *Science* 373, 871–876

451 82 Jumper, J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*
452 596, 583–589

83 Gomulkiewicz, R. *et al.* (2000) Hot spots, cold spots, and the geographic mosaic theory of
coevolution. *The American Naturalist* 156, 156–174

Highlight

- 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 are composed of 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 are still 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.

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 polymorphism?

- Are the same genes/types of genes involved across 1) host or parasite species, 2) habitats of a given symbiont species (host or parasite), or 3) 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 multi-species interactions at the community level?

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.

Genetic co-structure: statistical congruence between the population genetic structures of interacting species.

Demographic processes: include population size fluctuations over time and space, which influences the variation in **effective population size (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

499 population caused by genetic drift, *i.e.*, the random sampling of genetic variants in a finite
500 population.

501 **Habitat:** the set of local abiotic and biotic conditions experienced by one or several populations
502 in a given geographic area.

503 **Gene flow:** the migration of individuals, and thus of genes/alleles, between populations/demes.

504 **Local adaptation:** In spatially heterogeneous environments, evolution can lead to the adaptation
505 of populations to their local environmental conditions. Pattern of local adaptation occurs when the
506 mean fitness of a population in their home environment is higher than the mean fitness of
507 populations from elsewhere [64].

508 **Metapopulation:** populations connected by gene flow. Individual populations may go extinct and
509 new populations may be established by migrants.

510 **Population:** a group of individuals who are more genetically similar to each other than they are to
511 individuals outside the subpopulation, as a result of genetic drift, migration, mutation, and
512 selection. Also called “deme” in the metapopulation framework, and often assumed as a panmictic
513 group inferred from Bayesian inference methods in population genomics studies [65–67].

514 **Selection mosaic:** spatial variation in the strength of coevolution (hot- and cold- spots) among
515 interacting species due to spatial variation in the biotic and selective pressure.

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 set up 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]. Indeed, 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]).

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 (Figure 1 and Box 1, [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 (Figure 1), 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 effective population size, N_e , of both species) are observable using genomic data sampled in a time series, and are thus likely to be relevant for predicting 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 multi-annual 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, while taking into account for the past demography and spatial structure of each species and their co-demographic history. This would allow us to

562 decipher the bases of GxGxE interactions using genomic scans of selection at a high spatial and
563 temporal resolution (see above).

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

We propose the following steps for carrying out in-depth coevolutionary studies (Figure 2):

1) Hierarchical sampling: 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.

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

3a) Inferences of the demography and population structure for each species. The variance of migration rate can be compared to information from recombination maps when available (or by simultaneous modelling of demographic history and inference of recombination maps using the iSMC method on full genome data [70]). Indeed, 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.

4a) Selection scans: apply scans using the neutral demographic model and inferred spatial population genetic structure to set statistical thresholds. It is possible to run selection scans on host (infected and non-infected) 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] or genotype matrices) methods as well as on classic selective sweep or balancing selection tests.

3b and 4b) Sampling over time: with appropriate sampling, it is possible to test the 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 multi-annual 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].

5a) Co-genomics: using the full genomes or sequence capture data from hosts and parasites, we suggest carrying out 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 using co-GWAS is very powerful to identify the genes underlying the interaction [40]. It is advisable to perform co-GWAS under different environmental conditions to account for GxGxE effects, or use GEAS to link genetic variation of hosts and parasites to environmental variables [59], e.g., climate.

5b) Functional validation: if controlled experiments are possible, the number of candidate loci involved in coevolution can be obtained from **co-expression analyses** (joint RNAseq [79] or proteomics) of host and symbiont samples. Co-expressed genes can be compared to those identified by co-GWAS and selection scans. Cross-species eQTL mapping combined genomes and gene expression can identify polymorphisms in the host genome affecting parasite expression, and

610 *vice versa* [80]. New tools to determine protein structure (AlphaFold and RoSetta for instance
611 [81,82]) may also deepen our insights into the molecular basis of protein interactions and
612 coevolution. Functional assays to validate these candidate genes and proteins at the population
613 level in the host and the parasite is the final step that can be done together with population
614 genomics approaches.
615

Figure 1. Major factors underlying coevolutionary dynamics in an ecological context. 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 (GxG) in its simplest representation: hotspot ($G_h \times G_p$) 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 ($G_h \times G_p$, G_p or G_h , $G_h \times E$ or $G_h \times E$, $G_p \times G_h \times E$), as well as on between habitat heterogeneity ($G_p \times G_h \times E$). Overall, the heterogeneity in biotic and abiotic interactions among populations and habitats ($G_x G_x E$) leads to a complex pattern of spatially heterogeneous selection. **d.** Adding demographic processes (*i.e.*, **effective population size** N_e and gene flow) to the adaptive processes driven by interactions (*i.e.*, $G_x G$, $G_x E$ and $G_x G_x E$) helps to assess variation in selection pressure and the speed of coevolution over space and time.

Figure 2. Practical guide for future studies of coevolutionary dynamics using ecological genomics