Article (refereed) - postprint

Gugerli, Felix; Brandl, Roland; Castagneyrol, Bastien; Franc, Alain; Jactel, Hervé; Koelewijn, Hans-Peter; Martin, Francis; Peter, Martina; Pritsch, Karin ; Schroder, Hilke; Smulders, Marinus J.M.; Kremer, Antoine; Ziegenhagen, Birgit. 2013. **Community genetics in the time of next-generation molecular technologies.** *Molecular Ecology*, 22 (12). 3198-3207. 10.1111/mec.12300

Opinion

Community genetics in the time of next generation molecular technologies

FELIX GUGERLI[1], ROLAND BRANDL[2], BASTIEN CASTAGNEYROL[3], ALAIN FRANC[3], HERVÉ JACTEL[3], HANS-PETER KOELEWIJN[4], FRANCIS MARTIN[5], MARTINA PETER[1], KARIN PRITSCH[6], HILKE SCHRÖDER[7], MARINUS J.M. SMULDERS[8], ANTOINE KREMER[3], BIRGIT ZIEGENHAGEN[2] and EVOLTREE JERA3 CONTRIBUTORS*

[1] *WSL Swiss Federal Research Institute, 8903 Birmensdorf, Switzerland*

[2] *Fachbereich Biologie, Philipps-Universität Marburg, 35032 Marburg, Germany*

[3] *UMR1202 Biodiversity, Genes & Communities, INRA Pierroton, 33612 Cestas Cedex, France*

[4] *ALTERRA Centre for Ecosystem Studies, Wageningen UR, 6700 AA Wageningen, The Netherlands*

[5] *UMR "Interactions Arbres/Micro-Organismes", INRA Nancy, 54280 Champenoux, France*

[6] *Institute of Soil Ecology, Helmholtz Zentrum München, 85764 Neuherberg, Germany*

[7] *Institute for Forest Genetics, Johann Heinrich von Thuenen-Institute, 22927 Grosshansdorf, Germany*

[8] *Plant Research International, Wageningen UR, 6700 AA Wageningen, The Netherlands*

* The EC-supported Network of Excellence Evoltree (http://www.evoltree.eu) formed a group of scientists involved in and actively contributing to Jointly Executed Research Activities on community genetics in forest ecosystems. In addition to the main authors, the

group includes S. Augustin, M. Brändle, C. Burban, J. Burczyk, S. Cavers, I. Chybicki, C. Conord, E. Cremer, J. DeWoody, K. Donges, B. Fady, L. Karlinski, C. Kerdelhué, B. Kieliszewska-Rokicka, G. Kost, M. Kulczyk-Skrzeszewska, F. Lakatos, F. Lefèvre, S. Liepelt, S. Oddou-Muratorio, K.-H. Rexer, M. Rudawska, M. Schädler, G. Taylor, K. Tuba, M. Viger, F. Villani, M. Villar

Keywords: ectomycorrhizal symbiosis, gene-to-gene interactions, forest ecosystems, herbivorous insects, population genomics, quantitative trait analysis

Corresponding author: Felix Gugerli, WSL Swiss Federal Research Institute, 8903 Birmensdorf, Switzerland (fax: +41 (0)44 739 22 15; E-mail: gugerli@wsl.ch)

Running title: Community genetics and genomics

Word count: 7173

Abstract

Understanding the interactions of co-occurring species within and across trophic levels provides key information needed for understanding the ecological and evolutionary processes that underlie biological diversity. As genetics has only recently been integrated into the study of community-level interactions, the time is right for a critical evaluation of potential new, gene-based approaches to studying communities. Next generation molecular techniques, used in parallel with field-based observations and manipulative experiments across spatio-temporal gradients, are key to expanding our understanding of community-level processes. Here, we introduce a variety of "-omics" tools, with recent studies of plant–insect herbivores and of ectomycorrhizal systems providing a detailed example of how next generation approaches can revolutionize our understanding of interspecific interactions. We suggest ways that novel technologies may convert community genetics from a field that relies on correlative inference to one that reveals causal mechanisms of genetic co-variation and adaptations within communities.

Community genetics aims to understand how genetic variation within and among populations of host species affects the composition of associated organisms interacting with the host (Agrawal 2003; Whitham *et al.* 2006; Johnson & Stinchcombe 2007; Rowntree *et al.* 2011; Wymore *et al.* 2011). Empirical community genetics has been stimulated by pioneering work on poplars (*Populus* spp.), their genotype-based phenotypic variation, and associated communities (Whitham *et al.* 2006). However, community genetics has hitherto largely remained phenomenological, and the underlying genetic basis and processes involved in the interactions between host and associated organisms have not been studied in detail yet. Given the rapid development of molecular techniques (Rokas & Abbot 2009), it will soon be

feasible to characterize the genomes of numerous members of a community. With whole-genome sequences or other types of -omics data at hand (Nadeau & Jiggins 2010), community genetics will be able to establish a solid genetic framework in which to understand the interplay between ecological and evolutionary processes (Rokas & Abbot 2009). Here, we sketch possible avenues along which research in community genetics may proceed, focussing in particular on how -omics may improve our understanding of the role of gene variants in species interactions. First, we argue for exploring spatio-temporal variation to investigate the fundamental ecological and evolutionary aspects of community genetics. Second, we describe how genomic, transcriptomic, proteomic, and metabolomic research can improve understanding of the interactions between trees as focal species and ectomycorrhizal fungi or herbivorous insects, the key players in forest ecosystems.

Community genetics in a spatio-temporal perspective

Let us consider populations of a focal species that start to diverge genetically. Genetic drift and/or selection may induce shifts in allele frequencies, leading to changes in the phenotypic traits mediating interactions with associated species that use the focal species as a host. First, these genetic changes and changes in the associated traits may lead to shifts in the occurrence and abundance of species already associated with the host. Second, the new phenotypic traits of the focal species may allow new species from the regional species pool to colonize it. Finally, changes in the genetics of the host may induce evolutionary responses, including speciation events, in the associated organisms, which may feedback to evolutionary changes in the host.

If the above scenarios hold true, we expect the relatedness of host genotypes to co-vary with similarity among the communities of associated species (Bangert *et al.* 2006; Brändle &

Brandl 2006). Within species, such patterns have received considerable attention under the concept of the "extended phenotype". This concept was introduced by Richard Dawkins (1982) to describe effects of genes on an individual's environment including other organisms. Whitham *et al.* (2003; 2005; 2006) adopted this concept and developed a framework for community and ecosystem genetics, which includes a feedback where an individual's phenotype is dependent on the interaction with other species.

Community assembly (Kraft *et al.* 2007; Emerson & Gillespie 2008) is shaped by successive filters, including regional species pool, habitat area and isolation (biogeographical filters), local environmental constraints (abiotic, biophysical filters) and biological interactions such as competition or predation (biotic filters; Fig. 1). The host genotype, interacting with the environment, may affect the structure of associated communities at several filtering steps by controlling phenotypic traits that allow associated organisms to locate, select and exploit resources of their host (Johnson & Agrawal 2005; Bailey *et al.* 2009) (Fig. 1). Thus, spatial variation in the composition of associated communities has a strong regional component.

Despite many reports demonstrating a correlation between genotypes of a focal species and the composition of associated communities, the fundamental ecological, genetic and evolutionary processes that generate this correlation remain poorly explored and require consideration in future studies. In this regard, three aspects deserve special attention: spatial variation, temporal variation, and gene-to-gene interactions.

First, space needs to be better integrated into study designs. As noted above, the assembly of species depends on the regional species pool, whose phylogenetic and functional structure imposes a constraint on the emerging local communities (Fig. 1). A group of genotypes of a focal species in natural or experimental population is embedded in a landscape context that

may include forest patches, arable land, urban environment or other habitats type, each of which has different species pools that might interact with the focal species. As the associated community influences the fitness of the focal species, the relative fitness of these genotypes will vary across sites, even if the abiotic conditions are similar. However, in single common garden experiments, genotypes of a focal species are exposed only to one particular species pool. Therefore, regional replicates of such experiments are necessary to estimate the stability of relationships between genotypes of the focal species and communities of associated species. Such replicates would enable us to distinguish between mainly spatial effects and those that can be attributed to the interaction between host genotypes and associated organisms. Alternatively, one might set up more complex common gardens including particular treatments, for example through fertilization or irrigation. Such an approach would allow tests of the effect of genotype x environment interactions on the assemblage of associated species for each local species pool. Furthermore, replicated common garden experiments would further allow constructing reaction norms of different genotypes of the focal species. Do these genotypes respond differentially for their extended phenotypes to the changes of abiotic or biotic conditions across the testing sites? An initial step would be to identify the shape of the reaction norms (linear or quadratic) and then to estimate their variation among genotypes. Finally, the spatial context may also be dissected at the within-population level. For example, natural populations of trees usually exhibit strong spatial autocorrelation due to limited dispersal, which increases steadily over generations. On the other hand random spatial genetic structure is observed in recently planted forests. One would therefore expect very different spatial structures of extended phenotypes among these strongly contrasting cultural regimes.

Second, community genetics should consider temporal variation in species interactions, e.g. among seasons, among years along successional sequences, and other types of temporal gradients. Traits involved in plant–herbivore interactions are known to change during plant ontogeny (Boege & Marquis 2005; Holeski *et al.* 2009), which is why communities of insect herbivores – and herbivory pressure – on seedlings and mature individuals may differ (Le Corff & Marquis 1999; Basset 2001). Furthermore, although associated communities may change within and between years due to fluctuations in plant phenotypes, equally they may change due to differences in weather conditions. Thus, the phenotypic traits that are important for species interactions in a particular season or year may change within and between years, and drawing conclusions from short-term experiments may be misleading. Although such traits, and the underlying genes, are genuinely involved in community interactions, their relative importance compared to other genes may vary in time and can therefore only be established in long-term experiments. Hundreds of insect generations interact with a long-living host such as a tree during its lifetime, and each generation experiences different biophysical constraints and trophic interactions with other fungi, herbivores or predators. As a consequence, even though insect populations can adapt to individual host genotypes (Mopper *et al.* 2000), the strength and direction of these adaptations are likely to change over time (moving targets; Ruhnke *et al.* 2006).

Moreover, genetic processes underlie the formation of adaptive demes and co-evolution between host and associated organisms (Fig. 1). At present, the number and type of genes involved and the associated phenotypes of interacting species are largely unknown. Recent technological advances enable researchers to sequence whole genomes and to monitor gene expression of interacting species, offering the potential to identify the candidate genes mediating the interactions between focal and associated species. Such approaches will move

158 community genetics from studying anonymous genotype/phenotype effects to studying gene-
159 to-organism, gene-to-gene, and ultimately to genome-to-genome interactions. While current
160 research has focused on the few "genome-enabled" species (Ekblom & Galindo 2011), the
161 many ongoing whole-genome projects will widen the array of study systems applying
162 genomics data in the near future (e.g. http://www.arthropodgenomes.org/wiki/i5K,
163 http://1000.fungalgenomes.org/home/, http://pinegenome.org/pinerefseq/).
164 The following sections describe how the various types of -omics may stimulate community
165 genetics. and how they enable the genetic component of variation in community composition
166 to be addressed at the level of variants in adaptive genes and their differential expression.

167 **An example of functional genomics based on a complete genome sequence:**
168 **ectomycorrhizal symbiosis**
169 Ectomycorrhizae, the mutualistic symbiosis between tree roots and a cortege of soil fungal
170 partners, are the most widespread and species-rich associations in temperate and boreal
171 forests. Ectomycorrhizal fungi receive carbon from photosynthesis and, in turn, promote tree
172 growth, enhance the survival of seedlings and increase the fitness of their plant partners under
173 a wide range of environmental conditions. Despite the ecological significance of this
174 mutualistic interaction, we have only started to explore its role for community ecology.
175 A breakthrough was the release of the first two full-genome drafts of mycorrhizal fungi,
176 namely *Laccaria bicolor* (Basidiomycota) and *Tuber melanosporum*, the Périgord truffle
177 (Ascomycota; Martin *et al.* 2008; Martin *et al.* 2010). Comparative genomics of the two
178 mycorrhizal fungi indicated that they use different gene networks (‘molecular toolkits’) to
179 establish symbiosis (Martin *et al.* 2010). There are vast differences between these two
180 ectomycorrhizal genomes. *Laccaria bicolor* has a 65 Mb genome with more than 23 000

predicted proteins, which is the largest complement of genes known for any fungus, whereas *T. melanosporum* has the largest fungal genome so far with 125 Mb, but has only 7500 predicted genes, one of the smallest complement of proteins in any filamentous fungal genomes sequenced so far. Also, whereas the secretion of effector-like small secreted proteins seems to be crucial for the establishment of the symbiosis in *L. bicolor* (Plett *et al.* 2011), these so-called mycorrhiza-induced small secreted proteins (MiSSPs) are not present in the transcriptome of *T. melanosporum* symbiotic tissue (Martin *et al.* 2010) In spite of these differences,, some common features and some novelties emerged from the comparison with genomes of saprophytic and pathogenic fungi. Besides the loss of plant cell-wall degrading enzymes in ectomycorrhizae, an increase in the diversity and expression of nutrient transporters and signalling pathways (e.g. tyrosine kinases) in symbiotic tissues are hallmarks of mycorrhizal genomes (Martin *et al.* 2008; Kosti *et al.* 2010; Martin *et al.* 2010; Plett *et al.* 2011). These symbiosis-related genes are good candidates for gene expression studies of multi-species interactions in the field. On the tree side, it is not known how the host tree selects its symbiotic associates. Plant-encoded small secreted proteins may be required, as shown for nitrogen-fixing symbioses (Van de Velde *et al.* 2010). Genomic studies will probably be the only way to elucidate the mechanisms of interaction and to understand the effect of gene variants on this interplay. Therefore, we think that this system is an exciting model for community genetics in the -omics era.

Ectomycorrhizal fungi show a continuum of specialization to the host tree from strict specialists to generalists. Differences in the expansion of multigene families, in particular dynamic repertoires of genes encoding small secreted proteins and sugar-cleaving enzymes, might be responsible for the different host ranges of specialists, e.g. *T. melanosporum*, and generalists, e.g. *L. bicolor* (Martin *et al.* 2010). That is, the genome expansion observed in *L.*

bicolor might be driven by selection of the symbiont to exploit diverse substrates provided by multiple potential hosts and by diverse soils. As more genomes of mycorrhizal fungi are sequenced (Martin *et al.* 2011), this hypothesis will become testable.

In addition to the genomics of host–symbiont interactions, studies of geographical patterns of co-evolution add to our knowledge of processes leading to reciprocal adaptation and specialization. There are only a handful of studies reporting the structure of geographic variation and patterns of co-evolution in mycorrhizal interactions, indicating that these patterns are geographically highly variable (Hoeksema 2010; Hoeksema *et al.* 2012). To date, mostly higher-level traits, such as intensity of mycorrhizal colonization or growth of host trees, have been studied. Several of these studies found significant genetic variation in either the host plant or the mycorrhizal fungus in its ecological effect on the other partner. For example, the relationship between the colonization intensity of the ectomycorrhizal fungus *Thelephora terrestris* and the growth of its host, Lodgepole pine (*Pinus contorta*), depends on the tree's genotype (Karst *et al.* 2009). In poplar, both the intensity of colonization and the amount of enzymes secreted by poplar root tips colonized by *L. bicolor* are under the genetic control of the host (Courty *et al.* 2011). Similar findings come from arbuscular mycorrhizal systems, where host identity has a strong effect on the fitness of different strains of *Glomus intraradices* (Ehinger *et al.* 2009).

An increasing body of evidence shows that subtle intraspecific differences in the genome of host plants determine the composition of interacting communities in mycorrhizal fungi (e.g. Korkama *et al.* 2006; Whitham *et al.* 2006; Sthultz *et al.* 2009; Karliński *et al.* 2010; Leski *et al.* 2010; Hoeksema *et al.* 2012). We have experimental evidence that such as intraspecific genetic variation in the host also affects the composition of interacting mycorrhizal populations (Hoeksema & Thompson 2007), but this has not yet been tested

under natural conditions. To understand the links between structure and diversity of communities and ecosystem functioning, we need to know more about spatio-temporal patterns of genetic variation. There are indications that both interspecific (e.g. van der Heijden *et al.* 1998; Maherali & Klironomos 2007) and intraspecific (e.g. Johnson *et al.* 2012) diversity of mycorrhizal fungi can regulate productivity and ecosystem functioning. We advocate studies of community and population diversity in forests and combining them with functional field studies, involving both partners of ectomycorrhizal symbioses. Numerous new techniques are emerging for gene expression studies, marker gene evaluation using comparative genomics, and enzyme activity profiling of whole ectomycorrhizal assemblages (Courty *et al.* 2010). The rapid development of high-throughput sequencing technologies facilitates the survey and comparison of whole microbial communities (Buée *et al.* 2009), although analysis, interpretation, and publication of data still needs to be optimized (Henrik Nilsson *et al.* 2012). Nevertheless, combined genotypic and functional studies are now feasible and may be expanded to natural and experimental gradients. Several reports indicate that soil microbe and mycorrhizal diversity differentially affect ecosystem functioning under different environmental conditions, e.g. nutrient status (van der Heijden *et al.* 2008). We also know that plant-associated microorganisms are an important factor influencing plant responses to climate change (Courty *et al.* 2010; Pickles *et al.* 2012). Combined genotypic and functional studies in diverse environments will help to understand current patterns and to predict changes and effects in the future.

Associations between genes and traits: potential of next generation approaches in community genetics

An essential part of future studies in community genetics will be to identify the genes that underlie the traits of hosts that affect associated organisms. For this, sequencing of the complete genome of a host species is not sufficient. Rather, it is essential to link the presence or action of particular variants of genes or genomic regions of a host plant to the presence or abundance of associated organisms or arrays of their genes. There are basically two strategies for this, namely QTL mapping and genome-wide association studies (GWAS). We briefly outline and illustrate below the pros and cons of these two approaches for community genetics.

An example of QTL mapping of community traits of poplar is a study aimed at identifying genomic regions associated with susceptibility to insects (DeWoody *et al.* submitted). Parents and progeny of a poplar (*Populus trichocarpa* × *P. deltoides*) F2 mapping population were assessed for various categories of leaf damage, including chewers and skeletonizers. The damage levels significantly varied among offspring genotypes. Each category was treated as a quantitative trait in a QTL mapping approach and more than ten QTLs were detected. QTLs also varied seasonally, suggesting that the insect community responds to traits and the underlying genetic variation over time. This underlines the importance of considering temporal variation in studies of community genetics, as noted above.

Another example is a study on QTLs affecting ectomycorrhizal symbiosis in a *P. deltoides* × *P. trichocarpa* F1 population (Labbé *et al.* 2011). Four identified QTLs were associated with candidate genes, and differential transcript levels were assessed with the help of a whole-genome microarray. The transcripts with the highest overrepresentation were, based on their gene ontology, in the repress defense mechanisms and in pathogen resistance.

Relatively few mapping populations have been produced for long-lived tree species, due to the length of time needed to maintain and study them, and the high costs associated with it. As a single cross will not contain all alleles present in a large population of an outcrossing species, not all QTLs can be detected in a single cross, and most QTL interactions will go unnoticed. Hence, several populations are necessary, and producing them would be an important investment. Next to full-sib families it may be possible to use full or partial diallel designs with multiple parents, so that more alleles are included and many more allele combinations can be studied, similar to MAGIC populations (Kover *et al.* 2009) but without the need for selfing to multiply and maintain the population.

In the meantime, an elegant alternative for forest trees is to use existing progeny trials. Many of these have been established and often replicated at different locations, and phenotypic data are usually available for extensive periods of time. Many trials consist of half-sib families, in which the alleles from the mother segregate in the progeny. If only a limited number of fathers were involved, genotyping may even allow them to be split into a few interconnected full-sib families. Common garden experiments often include a sample of the diversity of an area. When these experiments are replicated at multiple sites, it may be possible to perform genome-wide association mapping with the advantage of multi-site / multi-year data.

An issue for community genetics, as mentioned above, is that the local species pool may be different between the locations of the trials. This can be tackled efficiently by replicating the populations and planting them in different locations. Replicated populations will also spread the risk of losing individual members of the populations.

After finding a QTL region based on the presence of an associated organism or, for example, damage caused by an insect species, the underlying mechanism can be unravelled,

297 in this case by measuring the secondary compound composition of all progeny trees and
298 locating such traits on the genetic map. Co-localization of a compound with a QTL would
299 suggest that it was responsible for the effect on the insects and that a structural or regulatory
300 gene involved in its synthesis is located in that genomic region. In some species, this can be
301 tested by mutant analysis, but it is not practical with trees. Alternatively, one could analyse
302 the naturally occurring genetic variation in a large set of unrelated trees with different
303 combinations of compounds and conduct association tests (i.e. GWAS).
304 GWAS assumes that, in the absence of population substructure, markers that are physically
305 linked to a gene associated with a phenotype of a trait can be distinguished from markers that
306 are not linked, as the latter are assumed to occur randomly in individuals of the population
307 regardless of the phenotype (Nordborg & Weigel 2008). There is no need to construct a
308 mapping population as in QTL detection, but a reference genome or a dense genetic map in
309 combination with sufficient linkage disequilibrium (LD) are required (Kim *et al.* 2007). LD
310 appears to be limited in tree species (Ingvarsson 2005; Heuertz *et al.* 2006; Pyhäjärvi *et al.*
311 2007), which implies that high-density genetic marker arrays are needed for applying
312 association mapping and that many more individuals need to be studied. For instance,
313 Fournier-Level *et al.* (2009) tested target candidate genes and identified the functional
314 variation responsible for the observed variation in anthocyanin variation in grape by
315 association analysis. The very low LD often encountered in natural tree populations (Neale &
316 Savolainen 2004) will assist in finding many of the possible combinations of compounds, thus
317 increasing the power of the association study. A new approach, becoming feasible because of
318 high-throughput sequencing technology, is to pool and sequence DNA from multiple
319 individuals within a population with clearly distinct phenotypes or habitat conditions (Turner
320 *et al.* 2010), and to identify those markers across the genome that display a large difference in

allelic frequency between the pooled groups (Holderegger *et al.* 2008). The advantage of this 'population resequencing' approach, which vaguely resembles bulked segregant analysis (BSA), is that no mapping population or extensive LD is necessary; the drawback is that an annotated genome is still needed for reference. Since annotated genome sequences are increasingly becoming available, this will be less of a problem in the future. The approach can be readily extended to polygenic traits (Heard *et al.* 2010). A potential application to community genetics in trees would be to pool the DNA from trees that host a particular insect with DNA from those that do not, and compare the sequenced genomes of the two groups.

Next generation methods now enable genotyping-by-sequencing (Baird *et al.* 2008). In the context of segregating populations, restriction-site associated DNA (RAD) markers or transcriptome sequencing enable direct mapping-by-sequencing, thus skipping marker development altogether (Hartwig *et al.* 2012; Zhu *et al.* 2012). In QTL mapping this solves the problem of generating dense maps, so that the limiting factor for high resolution is the number of recombinations or the size of the segregating population. As forest trees have very small LD, the ability to generate high volumes of genomic data is a very promising development for GWAS.

Gene expression profiling, a complementary approach to association genomics as a strategy for functional genomics, is also being revolutionized by developments in next generation technologies. Gene expression profiling has been applied to study stress response in trees, for example following insect attack where transcript analyses by cDNA microarray profiles have been combined with 2-D protein and protein spectrometric analyses (Lippert *et al.* 2007). In this pioneering work on pines and pine weevils, the authors identified interspecific cross-talking transcripts and their proteins. Next generation sequencing of tagged cDNA ends now enables researchers to quantify the number of transcripts from different

345 subsets of individuals. Given the availability of gene annotations, the transcripts will be
346 associated with gene models and their regulators using publicly available databases. We
347 expect that co-expression profiling will become feasible for populations as well as for
348 individual ontogenetic stages of interacting species. Such an approach may also be scaled up
349 from two-species interactions to multiple-species interactions, i.e. a true 'community
350 transcriptome' approach.

351 Proteomic approaches allow for an efficient and simultaneous detection of the proteins in a
352 sample. The proteome composition to some extent integrates fluctuations in expression over a
353 period of time, thus potentially being robust with regard to sampling time in the field. The
354 identification of peptides relies on either a large, high-quality RNA-seq dataset, a complete
355 set of alleles from a multigene family, or the genome sequence. An example is the use of
356 peptide identification (Q-TOF LC-MS^E) for fast screening of Bet v 1 isoforms in pollen of
357 various birch species, as it was possible to determine both presence and relative abundances
358 of individual isoforms (Schenk *et al.* 2009). For this, the mass spectra obtained from the
359 pollen were compared with a set of predicted peaks based on a complete set of isoforms
360 obtained by sequencing the genes. In species for which the genome sequence or a large
361 amount of transcriptome data is available, this prediction becomes a relatively simple
362 bioinformatics exercise.

363 Other -omics techniques, such as metabolomics, may be employed in similar experimental
364 schemes. Recent advances have increased the sensitivity and throughput of metabolomics and
365 proteomics assays (‘next-gen biochem’). Now, one can directly map QTL controlling the
366 metabolic profile of all offspring of a cross. For instance, untargeted GC-TOF-MS metabolite
367 profiling allowed mapping of 100 mQTLs (Carreno-Quintero *et al.* 2012). The main
368 drawbacks of metabolomics are the higher costs and the problem of interfering factors due to

the different growing conditions of the trees included in the association analysis. Moreover the samples cannot be all taken at the same time. On the other hand, the compounds measured are also the ones that affect the interaction with associated insect species. So if genetic variation in multiple genes affects the content of one important compound, the association of the compound with presence or absence of one or more insect species will be stronger than that of each of the underlying genes, and the association will also be more informative on the mechanism of the interaction. Even GWAS could be done in this way. In our example using a pool of trees including those that host a particular insect and those that do not, a comparison of compounds may be more straightforward than comparing DNA markers. In particular, if the insect is not always present on the same trees across years, the compounds present in each tree in each year could reveal a strong correlation, whereas the genes that enable the tree to produce the compounds would not.

If, as indicated above, a compound affects the presence of insect species, then one would expect, reciprocally, the presence of catabolites of the compound in insect species that tolerate the compound, when these insects are sampled on the trees that produce it. This can be used to experimentally validate the statistical associations between compounds in the tree and the presence of insect species or guilds, and for a starting point for understanding the mechanisms behind the interactions between trees and insects.

Perspectives

A suite of -omics approaches is available to pave the way for studying entire communities. Accordingly, we need to refine hypotheses and develop suitable study designs and statistical tools (Augustin *et al.* 2010; Ovaskainen *et al.* 2010), which will improve implementation once

391 reduced costs make these tools applicable to large-scale sampling of community-level
392 interactions (Table 1).
393 As outlined above, we see two main directions that should be followed in community
394 genetics to substantiate inference on the interplay of genes, organisms, communities, and their
395 respective environments. First, joint descriptive and experimental studies should include
396 spatial and temporal gradients to account for environmental variation in these dimensions
397 (Thompson 2005; Crutsinger *et al.* 2009; Tack *et al.* 2010). Second, researchers in community
398 genetics should make better use of the exponentially increasing genomic information
399 becoming available, which will requires solid expertise in bioinformatics. If this is achieved,
400 gene-to-gene interactions can be explored in individual-based associations and at the level of
401 entire communities and shift community genetics towards becoming community genomics.
402 Moreover, community genetics goes beyond the effects of genotypes in one species on the
403 community of associated organisms. We also need to consider the reciprocal effects of how
404 associated communities shape the genotypic composition of their hosts and of how the
405 genotypes of associated species affect host communities (Fig. 1). There are virtually no
406 studies available on this aspect of community interactions, which leaves a wide-open field of
407 empirical research for the future. Exploring reciprocal interactions might help to extrapolate
408 population genomics and quantitative genomics of focal species. We will then need to adopt a
409 community-based understanding of selection and drift as well as to include G x G x E
410 interactions into reaction norm calculations. However, elaborating on this subject goes
411 beyond the scope of the present article.
412 In conclusion, we believe that the amalgamation of traditional population genetics,
413 quantitative genetics and ecology, fostered by the advent of new genomic technologies, will

414 revolutionise our perception of community and ecosystem processes and push community
415 genetics into a new era.

416 **Acknowledgments**

417 Our sincere thanks go to Rolf Holderegger and the anonymous reviewers for helpful
418 comments on earlier versions of the manuscript. Karen A. Brune and Stephen Cavers
419 provided linguistic editing. This work was financed through the EC-supported Network of
420 Excellence Evoltree (GOCE-016322).

421 **References**

422 Agrawal AA (2003) Community genetics: new insights into community ecology by
423 integrating population genetics. *Ecology*, **84**, 543–544.

424 Bailey JK, Hendry AP, Kinnison MT *et al.* (2009) From genes to ecosystems: an emerging
425 synthesis of eco-evolutionary dynamics. *New Phytologist*, **184**, 746–749.

426 Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping
427 using sequenced RAD markers. *PLoS One*, **3**, e3376.

428 Bangert RK, Turek RJ, Rehill B, Wimp GM, Bailey JK, Whitham TG (2006) A genetic
429 similarity rule determines arthropod community structure. *Molecular Ecology*, **15**, 1379–
430 1391.

431 Basset Y (2001) Communities of insect herbivores foraging on saplings versus mature trees of
432 *Pourouma bicolor (*Cecropiaceae) in Panama. *Oecologia*, **129**, 253–260.

433 Boege K, Marquis RJ (2005) Facing herbivory as you grow up: the ontogeny of resistance in
434 plants. *Trends in Ecology & Evolution*, **20**, 441–448.

435 Brändle, Brandl R (2006) Is the composition of phytophagous insects and parasitic fungi
436 among trees predictable? *Oikos*, **113**, 296–304.

437 Buée M, Reich M, Murat C *et al.* (2009) 454 Pyrosequencing analyses of forest soils reveal
438 an unexpectedly high fungal diversity. *New Phytologist*, **184**, 449–456.
439 Carreno-Quintero N, Acharjee A, Maliepaard C *et al.* (2012) Untargeted metabolic
440 quantitative trait loci analyses reveal a relationship between primary metabolism and
441 potato tuber quality. *Plant Physiology*, **158**, 1306-1318.
442 Courty P-E, Buée M, Diedhiou AG *et al.* (2010) The role of ectomycorrhizal communities in
443 forest ecosystem processes: New perspectives and emerging concepts. *Soil Biology &*
444 *Biochemistry*, **42**, 679–698.
445 Courty PE, Labbé J, Kohler A *et al.* (2011) Effect of poplar genotypes on mycorrhizal
446 infection and secreted enzyme activities in mycorrhizal and non-mycorrhizal roots. *Journal*
447 *of Experimental Botany*, **62**, 249–260.
448 Crutsinger GM, Cadotte MW, Sanders NJ (2009) Plant genetics shapes inquiline community
449 structure across spatial scales. *Ecology Letters*, **12**, 285–292.
450 Crutsinger GM, Reynolds WN, Classen AT, Sanders NJ (2008) Disparate effects of plant
451 genotypic diversity on foliage and litter arthropod communities. *Oecologia*, **158**, 65–75.
452 Dawkins R (1982) *The Extended Phenotype: The Gene as the Unit of Selection*. Freeman,
453 Oxford.
454 DeWoody J, Viger M, Lakatos F, Tuba K, Taylor G, Smulders MJM (submitted) Quantitative
455 trait loci for insect community composition an d leaf development co-located in hybrid
456 poplar. *PLoS One*.
457 Ehinger M, Koch AM, Sanders IR (2009) Changes in arbuscular mycorrhizal fungal
458 phenotypes and genotypes in response to plant species identity and phosphorus
459 concentration. *New Phytologist*, **184**, 412–423.

460 Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology
461 of non-model organisms. *Heredity*, **107**, 1–15.
462 Emerson BC, Gillespie RG (2008) Phylogenetic analysis of community assembly and
463 structure over space and time. *Trends in Ecology & Evolution*, **23**, 619–630.
464 Fournier-Level A, Cunff LL, Gomez C *et al.* (2009) Quantitative genetic bases of anthocyanin
465 variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: A quantitative trait locus to
466 quantitative trait nucleotide integrated study. *Genetics*, **183**, 1127–1139.
467 Hartwig B, Velikkakam James G, Konrad KK, Schneeberger K, Turck F (2012) Fast isogenic
468 mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant*
469 *Physiology*, **160**, 591–600.
470 Heard E, Tishkoff S, Todd JA *et al.* (2010) Ten years of genetics and genomics: what have we
471 achieved and where are we heading? *Nature Reviews Genetics*, **11**, 723–733.
472 Henrik Nilsson R, Tedersoo L, Lindahl BD *et al.* (2012) Towards standardization of the
473 description and publication of next-generation sequencing datasets of fungal communities.
474 *New Phytologist*, **191**, 314–318.
475 Heuertz M, de Paoli E, Källmann T *et al.* (2006) Multilocus patterns of nucleotide diversity,
476 linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.)
477 Karst]. *Genetics*, **174**, 2095–2105.
478 Hoeksema JD (2010) Ongoing coevolution in mycorrhizal interactions. *New Phytologist*, **187**,
479 286–300.
480 Hoeksema JD, Hernandez JV, Rogers DL, Mendoza LL, Thompson JN (2012) Geographic
481 divergence in a species-rich symbiosis: Interactions between Monterey pines and
482 ectomycorrhizal fungi. *Ecology*, **in press**.

Hoeksema JD, Thompson JN (2007) Geographic structure in a widespread plant–mycorrhizal interaction: pines and false truffles. *Journal of Evolutionary Biology*, **20**, 1148–1163.

Holderegger R, Herrmann D, Poncet B *et al.* (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology and Diversity*, **1**, 273–283.

Holeski LM, Kearsley MJC, Whitham TG (2009) Separating ontogenetic and environmental determination of resistance to herbivory in cottonwood. *Ecology*, **90**, 2969–2973.

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics*, **169**, 945–953.

Johnson D, Martin F, Cairney J, Anderson IC (2012) The importance of individuals: intraspecific diversity of mycorrhizal plants and fungi in ecosystems. *New Phytologist*, **194**, 614–628.

Johnson MTJ, Agrawal AA (2005) Plant genotype and environment interaction to shape a diverse arthropod community on evening primrose (*Oenothera biennis*). *Ecology*, **86**, 874–885.

Johnson MTJ, Stinchcombe JR (2007) An emerging synthesis between community ecology and evolutionary biology. *Trends in Ecology & Evolution*, **22**, 250–257.

Karliński L, Rudawska M, Kieliszewska-Rokicka B, Leski T (2010) Relationship between genotype and soil environment during colonization of poplar roots by mycorrhizal and endophytic fungi. *Mycorrhiza*, **20**, 315-324.

Karst J, Jones M, Turkington R (2009) Ectomycorrhizal colonization and intraspecific variation in growth responses of lodgepole pine. *Plant Ecology*, **200**, 161–165.

506 Kim S, Plagnol V, Hu TT *et al.* (2007) Recombination and linkage disequilibrium in
507 *Arabidopsis thaliana*. *Nature Genetics*, **39**, 1151–1155.
508 Korkama T, Pakkanen A, Pennanen T (2006) Ectomycorrhizal community structure varies
509 among Norway spruce (*Picea abies*) clones. *New Phytologist*, **171**, 815–824.
510 Kover PX, Valdar W, Trakalo J *et al.* (2009) A multiparent advanced generation nter-cross to
511 fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics*, **5**, e1000551.
512 Kraft NJB, Cornwell WK, Webb CO, Ackerly DD (2007) Trait evolution, community
513 assembly, and the phylogenetic structure of ecological communities. *American Naturalist*,
514 **170**, 271–283.
515 Labbé J, Jorge V, Kohler A *et al.* (2011) Identification of quantitative trait loci affecting
516 ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of
517 genes in ectomycorrhizas of the two parents: *Populus deltoides* and *Populus trichocarpa*.
518 *Tree Genetics & Genomes*, **7**, 617–627.
519 Le Corff J, Marquis RJ (1999) Differences between understorey and canopy in herbivore
520 community composition and leaf quality for two oak species in Missouri. *Ecological*
521 *Entomology*, **24**, 46–58.
522 Leski T, Aučina A, Skridaila A, Pietras M, Riepšas E, Rudawska M (2010) Ectomycorrhizal
523 community structure of different genotypes of Scots pine under forest nursery conditions.
524 *Mycorrhiza*, **20**, 473–481.
525 Lippert D, Chowrira S, Ralph SG *et al.* (2007) Conifer defense against insects: Proteome
526 analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or
527 feeding by white pine weevils (*Pissodes strobi*). *Proteomics*, **7**, 248–270.
528 Maherali H, Klironomos JN (2007) Influence of phylogeny on fungal community assembly
529 and ecosystem functioning. *Science*, **316**, 1746–1748.

Martin F, Aerts A, Ahren D *et al.* (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, **452**, 88–92.

Martin F, Cullen D, Hibbett D *et al.* (2011) Sequencing the fungal tree of life. *New Phytologist*, **190**, 818–821.

Martin F, Kohler A, Murat C *et al.* (2010) Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, **464**, 1033–1038.

Mopper S, Stiling P, Landau K, Simberloff D, Van Zandt P (2000) Spatiotemporal variation in leafminer population structure and adaptation to individual oak trees. *Ecology*, **81**, 1577–1587.

Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, **26**, 484–492.

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, **9**, 325–330.

Nordborg M, Weigel D (2008) Next-generation genetics in plants. *Nature*, **456**, 720–723.

Novotny V, Basset Y, Miller SE *et al.* (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature*, **416**, 841–844.

Ovaskainen O, Nokso-Koivista J, Hottola J *et al.* (2010) Identifying wood-inhabiting fungi with 454 sequencing – what is the probability that BLAST gives the correct species? *Fungal Ecology*, **3**, 274–283.

Pickles BJ, Egger KN, Massicotte HB, Green DS (2012) Ectomycorrhizas and climate change. *Fungal Ecology*, **5**, 73–84.

Plett J, Kemppainen M, Kale S *et al.* (2011) A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology*, **21**, 1197–1203.

553 Pyhäjärvi T, Rosario García-Gil M, Knürr T, Mikkonen M, Wachowiak W, Savolainen O
554 (2007) Demographic history has influenced nucleotide diversity in European *Pinus*
555 *sylvestris* populations. *Genetics*, **177**, 1713–1724.
556 Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. *Trends in Ecology*
557 *and Evolution*, **24**, 192–200.
558 Rowntree JK, Shuker DM, Preziosi RF (2011) Forward from the crossroads of ecology and
559 evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**,
560 1322–1328.
561 Ruhnke H, Schädler M, Matthies D, Klotz S, Brandl R (2006) Are sawflies adapted to
562 individual host trees? A test of the adaptive deme formation hypothesis. *Evolutionary*
563 *Ecology Research*, **8**, 1039–1048.
564 Schenk MF, Cordewener JHG, America AHP, van't Westende WPC, Smulders MJM, Gilissen
565 LJWJ (2009) Characterization of PR-10 genes from eight *Betula* species and detection of
566 Bet v 1 isoforms in birch pollen. *BMC Plant Biology*, **29**, 24.
567 Schoonhoven LM, Loon JJAv, Dicke M (2005) *Insect–Plant Biology,* 2nd edn. Oxford
568 University Press, New York.
569 Sthultz CM, Whitham TG, Kennedy K, Deckert R, Gehring CA (2009) Genetically based
570 susceptibility to herbivory influences the ectomycorrhizal fungal communities of a
571 foundation tree species. *New Phytologist*, **184**, 657–667.
572 Tack AJM, Ovaskainen O, Pulkkinen P, Roslin T (2010) Spatial location dominates over host
573 plant genotype in structuring an herbivore community. *Ecology*, **91**, 2660–2672.
574 Thompson JN (2005) *The geographic mosaic of coevolution*. University of Chicago Press,
575 Chicago.

576 Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population
577 resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature*
578 *Genetics*, **42**, 260–263.
579 Van de Velde W, Zehirov G, Szatmari A *et al.* (2010) Plant peptides govern terminal
580 differentiation of bacteria in symbiosis. *Science*, **327**, 1122–1126.
581 van der Heijden MGA, Bardgett RD, Van Straalen NM (2008) The unseen majority: soil
582 microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology*
583 *Letters*, **11**, 296–310.
584 van der Heijden MGA, Klironomos JN, Ursic M *et al.* (1998) Mycorrhizal fungal diversity
585 determines plant biodiversity, ecosystem variability and productivity. *Nature*, **396**, 69–72.
586 Whitham TG, Bailey JK, Schweitzer JA *et al.* (2006) A framework for community and
587 ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics*, **7**, 510–523.
588 Whitham TG, Lonsdorf EV, Schweitzer JA *et al.* (2005) "All effects of a gene on the world":
589 Extended phenotypes, feedbacks, and multi-level selection. *Ecoscience*, **12**, 5–7.
590 Whitham TG, Young WP, Martinsen GD *et al.* (2003) Community and ecosystem genetics: a
591 consequence of the extended phenotype. *Ecology*, **84**, 559–573.
592 Wymore AS, Keeley ATH, Yturralde KM, Schroer ML, Propper CR, Whitham TG (2011)
593 Genes to ecosystems: exploring the frontiers of ecology with one of the smallest biological
594 units. *New Phytologist*, **191**, 19–36.
595 Zhu Y, Mang H-g, Sun Q, Qian J, Hipps A, Hua J (2012) Gene discovery using mutagen-
596 induced polymorphisms and deep sequencing: application to plant disease resistance
597 *Genetics*, **192**, 139–146.
598

599 **Table 1** From genes of focal species to traits of the extended phenotype – and back: questions and experimental considerations, related to
600 (a) spatio-temporal variation, (b) the application of -omics approaches, and (c) reciprocal effects to stimulate future studies in community
601 genetics
602

Theme	Questions	Experimental considerations
(a) Spatio-temporal variation	To what degree do regional species pools determine the composition of organisms associated to particular genotypes?	Assess naturally occurring spatial replicates of particular genotypes, e.g. agricultural, horticultural or silvicultural clones, and perform regionally replicated experiments using the same (set of) genotypes exposed to various regional species pools of potentially associated organisms.
	What is the relevance of phylogeographic structure in host species for the composition of associated communities?	Consider genetic structure and evolutionary lineages of the focal species.

	How do relationships between genotypes and associated organisms vary among seasons or among life stages?	Perform temporally replicated experiments or monitor natural communities across >1 year; establish long-term experiments with host plants from seedlings to mature adults.
	How does landscape configuration, e.g. differences in the relative abundance of, or connectivity among, particular habitat types, affect regional species pools and, thus, the communities of associated organisms in a focal species?	Include landscape characteristics when setting up experimental plots or assessing natural communities.
	To which degree does plasticity shape extended phenotypes?	Set up common garden experiments along ecological gradients including reciprocal transplants to test for genotype-by-environment interactions and reaction norms.
(b) -omics approaches	Which QTL relate to particular groups of associated organisms?	Establish various fullsib families or diallel crosses to include a wide range of allele variants.

What (classes of) compounds differ among host genotypes that are differentially affected by groups of associated organisms?	Genome/transcriptome sequencing of pools of host plants differing in their associated communities.
Do traits affecting community composition of associated species rely on single or multiple genes, and how large is their allelic variation within host populations?	Identify genes directly involved in the interaction, e.g. through QTL mapping, and quantify the degree of polymorphism using high-throughput, reduced-representation sequencing.
Does one gene of a focal species influence a single, a group or all associated species?	Use feeding (herbivores) or inoculation (ectomycorrhizae) experiments and perform co-expression profiling and subsequent protein annotation.
How many such genes exist, given that a focal species may interact with hundreds of associated species?	Perform gene expression studies of focal species that are experimentally associated with different single species or groups of species of associated organisms.

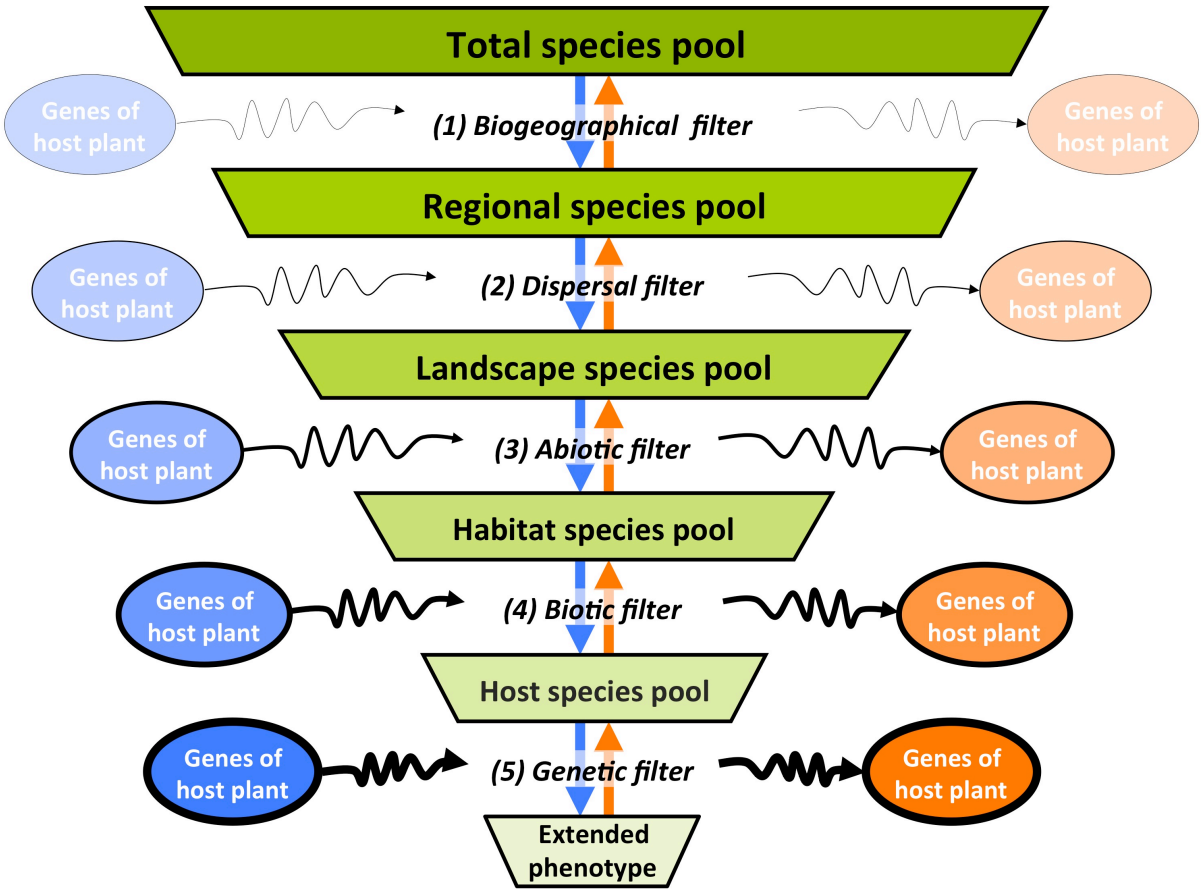
(c) Reciprocal effects	How do different groups of associated species induce changes in the phenotypic traits (and the underlying allele frequencies) of the host? What genes in host and associated species determine whether they interact as generalists or specialists?	Expose the same (set of) hosts to different (sets of) associated species and test for changes in traits and allele frequencies over time. Combine comparative genomics and expression profiling among generalists and specialists in both hosts and associated species.

603

Fig. 1 How host plant genes might shape assemblages of associated organisms (blue pathway on the left). Several ecological filters drive the structure of communities associated with one host plant. Among associated species co-occurring within a region and determined by evolutionary and biogeographical processes (1, Total species pool), local species assemblages depend on dispersal (2, Landscape species pool) and habitat filters (3, Habitat species pool). Dispersal filter refers to the ability of species to colonize the focal site. Habitat filters correspond to their capacity to develop and survive in a habitat given abiotic constraints. Biotic interactions with the host species contribute to the shaping of a host species pool (4, biotic filter). Finally, variation among host plant genotypes may further select different associated communities, shaping the extended phenotypes.

Genes of the focal host plant can interact with the four filters, as illustrated by the interaction between trees and associated insect herbivores: (1) There is evidence that pools of insect herbivore species of different tree families or genera are significantly different, probably owing to a long co-evolutionary process involving insect feeding traits and plant defence responses (Novotny *et al.* 2002); (2) insect herbivores use genetically controlled physical (e.g. shape, colour) and chemical cues (e.g. volatile organic compounds) provided by host plants to locate the plants; (3) trees can be seen as ecological engineers which can modify abiotic conditions that insects experience, e.g. wind, moisture, or light; (4) genes control plant phenotype and resistance traits that are deeply involved in interactions with insect herbivores (Schoonhoven *et al.* 2005); and (5) variants of host plant genes may ultimately induce quantitative changes in traits involved in plant–insect interactions with consequences for insect community structure (Crutsinger *et al.* 2008).

Presumed reciprocal effects, through which associated organisms feed back to the composition of host genes, are depicted by orange colors (right side).


Total species pool
Genes of host plant
(1) Biogeographical filter
Genes of host plant
Regional species pool
Genes of host plant
(2) Dispersal filter
Genes of host plant
Landscape species pool
Genes of host plant
(3) Abiotic filter
Genes of host plant
Habitat species pool
Genes of host plant
(4) Biotic filter
Genes of host plant
Host species pool
Genes of host plant
(5) Genetic filter
Genes of host plant
Extended phenotype

629