



Review in Advance first posted online
on March 3, 2011. (Changes may
still occur before final publication
online and in print.)

Evolution and Population Genetics of Exotic and Re-Emerging Pathogens: Novel Tools and Approaches

Niklaus J. Grünwald^{1,2,*} and Erica M. Goss^{1,3}

¹Horticultural Crops Research Laboratory, USDA Agricultural Research Service, Corvallis, Oregon 97330; email: Nik.Grunwald@ars.usda.gov

²Department of Botany and Plant Pathology and Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon 97330

³Current address: Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, Florida; email: emgoss@ufl.edu

Annu. Rev. Phytopathol. 2011. 49:5.1–5.19

The *Annual Review of Phytopathology* is online at
phyto.annualreviews.org

This article's doi:
10.1146/annurev-phyto-072910-095246

Copyright © 2011 by Annual Reviews.
All rights reserved

0066-4286/11/0908/0001\$20.00

*Corresponding author.

Keywords

coalescent, genealogy, invasive pathogen, population structure, maximum likelihood, Bayesian inference

Abstract

Given human population growth and accelerated global trade, the rate of emergence of exotic plant pathogens is bound to increase. Understanding the processes that lead to the emergence of new pathogens can help manage emerging epidemics. Novel tools for analyzing population genetic variation can be used to infer the evolutionary history of populations or species, allowing for the unprecedented reconstruction of the demographic history of pathogens. Specifically, recent advances in the application of coalescent, maximum likelihood (ML), and Bayesian methods to population genetic data combined with increasing availability of affordable sequencing and parallel computing has created the opportunity to apply these methods to a broad range of questions regarding the evolution of emerging pathogens. These approaches are particularly powerful when used to test multiple competing hypotheses. We provide several examples illustrating how coalescent analysis provides critical insights into understanding migration pathways as well as processes of divergence, speciation, and recombination.

Identity by state:
refers to alleles that
are identical but do
not share ancestry

Identity by descent:
refers to alleles that
came from the same
ancestral allele

ML: maximum
likelihood

**Population
bottleneck:**
a substantial and
transient reduction in
population size in the
ancestry of a
population or species

INTRODUCTION

In a world relying more and more on global trade, migration of exotic pathogens to new hosts or environments is bound to accelerate. Meanwhile, demands for increasing food production to sustain a growing human population will invariably result in intensification of agriculture. This, in turn, will facilitate an accelerated (re)emergence of pathogens because of increased movement of pathogens associated with shipments of food and goods, as well as human travel. The events leading to the emergence of a pathogen vary on a case-by-case basis, but will generally involve intense selection on the pathogen population. This selection causes substantial demographic changes to the populations that leave distinct signatures in the pattern of genetic variation within and among populations. Contemporary analytical tools for the analysis of population genetic variation can be used to infer the evolutionary history of populations or species, allowing for the unprecedented reconstruction of the history of pathogens. Understanding the evolutionary history of these newly or reemerging pathogens can inform management of emerging epidemics and suggest strategies to address future threats.

Population genetic analysis has had a long history in plant pathology (1, 32, 43, 44, 77–79, 82, 83, 90). Most recent research on the population biology of plant pathogens has relied on the use of electrophoretic markers, such as restriction fragment length polymorphisms (RFLP) (17, 35, 36, 80), amplified fragment length polymorphisms (AFLP) (48, 61), and simple sequence repeats (SSR) (2, 5, 27, 50, 60, 86, 93). Interpretation of such electrophoretic markers as loci or alleles is limited in scope by the fact that comigrating bands shared by two individuals do not necessarily reflect descent from a common ancestor; hence, identity by allelic state does not indicate identity by descent (19, 68, 75). Consequently, these markers are not optimal for phylogenetic and genealogical reconstruction. In contrast, nucleotide sequence data offer the possibility of reconstructing patterns of descent among genotypes

within a species, or among populations of one or more species (20, 37, 98). Ancestral and derived states can be distinguished from sequence data using a combination of coalescent analysis to infer gene genealogies and Bayesian or maximum likelihood (ML) approaches to determine distributions for population parameters of interest. These approaches were traditionally expensive in both computer and sequencing requirements. Now, the convergence of affordable sequencing and parallel computing allows the routine use of these analytical approaches, which provide unprecedented opportunities for characterizing the evolution of pathogens.

The novel approaches we discuss here, particularly those based on coalescent theory, allow for the testing of competing hypotheses regarding questions on the origin of pathogens. Is an emerging or reemerging pathogen population more likely to have been introduced through one or more migration events, or is it a resident population that emerged because of changes in agronomic practices? If a recent host shift is suspected, is there ongoing gene flow with the source population that could be generating genetic variation in the target population? If there is evidence for an expansion in population size associated with its emergence, is this period of expansion old or has it begun only recently? Is there evidence for a population bottleneck prior to this expansion, as might be expected with a host shift, the introduction of an exotic, or the introduction of a new fungicide? Is sexual recombination generating variation in the pathogen population with the potential to generate more fit or virulent offspring? The answers to these questions can have direct impacts on management, such as identification of migration pathways for mitigation, design of fungicide rotations for chemical control, mining of wild plant hosts for resistance genes, and targeting of particular pathogen populations (e.g., sexually reproducing, genetically variable, or rapidly expanding populations). In general, these studies will lead to a greater understanding of how pathogens emerge and reemerge, and suggest proactive approaches to mitigate future threats. This

review focuses on novel approaches derived from coalescent theory that can be employed to answer these kinds of evolutionary questions about the emergence of pathogens.

COALESCENT THEORY

Coalescent theory provides the mathematical foundation for modeling the genealogical history of a sample of genetic data (64–66, 99, 100, 102). Coalescent theory is now a mature and well-established body of theory and comes in many flavors that will only be touched on here. The reader is referred to many good review articles and books (23, 51, 57, 85, 88, 96, 101). Coalescent analysis, in its simplest form, starts by inferring the genealogical history of a population sample by building a tree going backward in time by pairwise coalescing of samples sharing a common ancestor until all samples coalesce to the most common recent ancestor (MCRA) for the overall sample (Figure 1). One of the basic tenets of coalescent theory is that only direct ancestors of the sample matter, because it follows a population sample backward in time.

One of the strengths of the coalescent approach is that predictions emerge about the population processes that generated the observed patterns of variability in a population sample. The ability to test hypotheses based on competing demographic scenarios, as illustrated in Figure 2, has obvious advantages over descriptive approaches that reveal patterns but do not explicitly test competing hypotheses on evolutionary processes. Although traditional population genetics test if, for example, populations are significantly differentiated based on F_{st} analysis, coalescent-based approaches can for the same populations determine directions and rates of migration, presence of bottlenecks, and admixture (Figure 2). Examples of the applications of these approaches are discussed in more detail below.

As is generally the case in population genetics, the standard coalescent as originally described by Kingman makes a number of assumptions, including panmixis, selectively neutral variation, constant effective population size, and lack of recombination (64–66). Deviations from some of these assumptions can be incorporated into the coalescent analysis.

MRCA: most common recent ancestor

Panmixis: random mating in a population

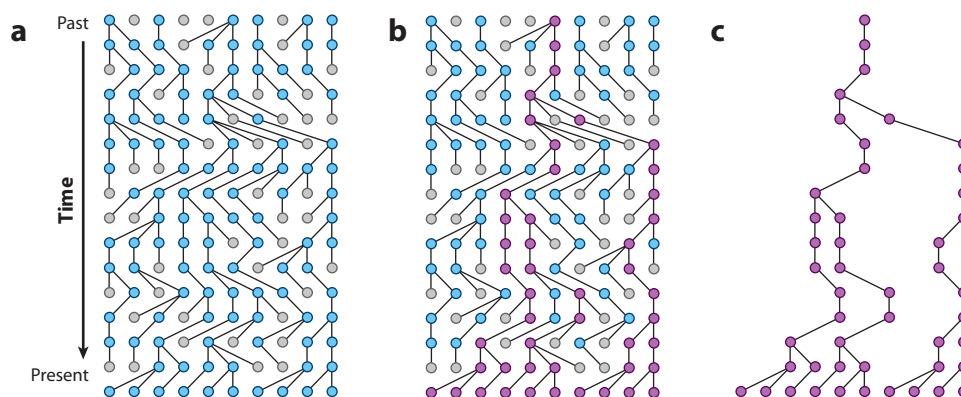


Figure 1

Illustration of the power of the coalescent process whereby haplotypes are coalesced backward in time to the most common recent ancestor (TMCRA). (a) Classical population genetics takes a forward-in-time perspective. In each generation, there are individuals that do not contribute offspring to the following generation, thus making the process of tracking a population through time inefficient. (b) Coalescent theory takes a backward-in-time perspective, only tracking the ancestors of the present-day sample, making the process of tracking the population through time much more efficient and resulting in (c) an inferred ancestry for a population sampled in the present. Simple coalescent simulations can be run at <http://www.coalescent.dk/>.

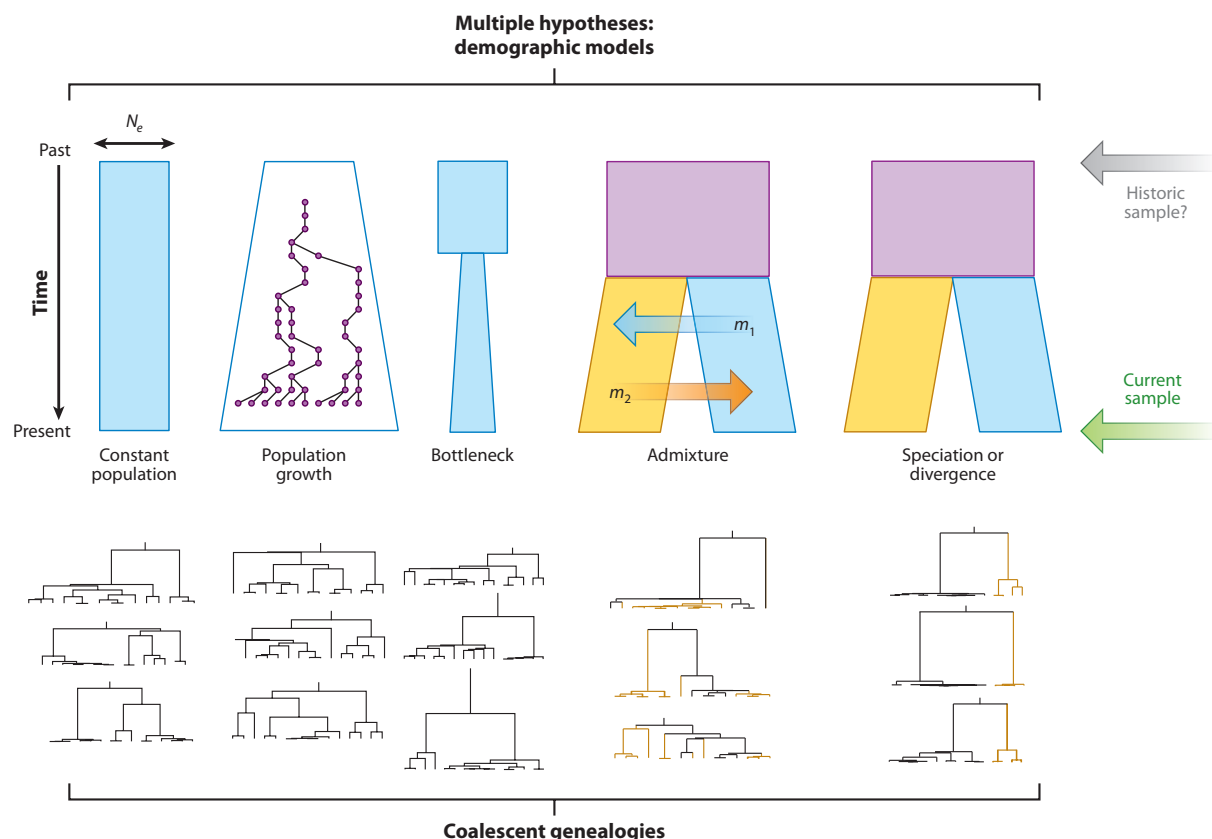
**Figure 2**

Illustration of select evolutionary scenarios that can be subjected to coalescent analysis for hypothesis testing. Typically, a population of individuals (current or historic sample) is obtained. For each individual, several genetic loci are characterized and subjected to coalescent analyses to obtain genealogies to infer the evolutionary processes that resulted in the patterns observed in the data. Shown are typical genealogies observed under each demographic scenario. Differences in genealogies generated under the same demographic model reflect the variation due to stochastic sampling effects. Note that one genealogy from a single locus would not provide conclusive insights. Genealogies were created using Hudson's MS program (58).

N_e : abstract parameter representing the size of an idealized population displaying the same level of genetic drift as the population sampled

The coalescent has important properties to consider. First, the larger the population size, the slower the rate of coalescence. Forward in time, genotypes are lost by genetic drift in small populations because of the sampling process. What is called genetic drift when moving forward in time, is the coalescent process when moving backward in time (101). Essentially, the timing of coalescent events is proportional to the effective population size (N_e), which is assumed to be large. Second, the topology of the coalescent tree is independent of the neutral mutational process (88). Thus, mutations are

mapped onto the genealogy after the genealogy is built because state (i.e., mutation) can be separated from descent (i.e., genealogy). The coalescent is a stochastic process, for which probabilities of coalescence or mutational events can be written down, but each iteration of the coalescent will be different, as is reflected by the replicate simulated genealogies shown under each scenario in **Figure 2**. Third, the coalescent process is quite robust regarding the effect of sample size on tree topology (**Figure 3**) (9, 30, 71). Looking backward in time, most of the coalescent events in a sample

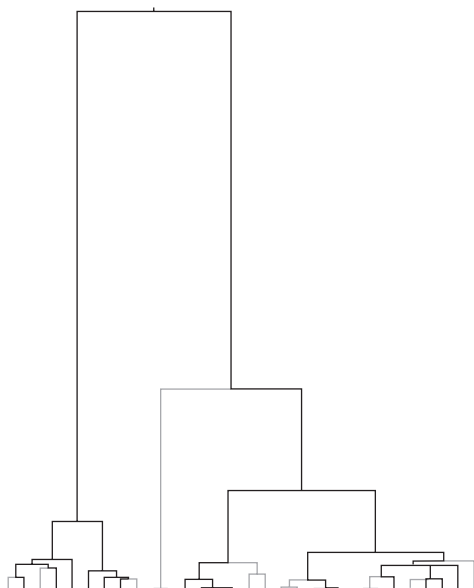


Figure 3

Moderate sample sizes provide accurate coalescent histories. The overall structure of the coalescent tree is retained when sample size is increased from 15 (black) to 30 sequences (gray). Note that most of the deep branches are present in the tree with 15 sequences, whereas some of the fine detail at the base of the tree is not. This genealogy was obtained for a simulation assuming the basic neutral model, constant population size, no recombination, panmixis, and the infinite sites model. Genealogies were created using Hudson's MS program (58).

will be relatively recent, whereas the last coalescent events will take up most of the time to the most recent common ancestor (TMRCA) of the sample. A moderate sample of about 15 sequences can provide the basic structure of the coalescent tree in which most of the deep branches are present, and additional sampling adds fine detail to the tips of the tree. This is one of the reasons the coalescent is so powerful.

Coalescent algorithms come in different forms (see below) and are typically employed to determine several population parameters and their distributions. These might include Θ , migration rates, TMCRA, recombination rates, and ages of specific mutations. The parameter Θ is critical to coalescent analyses because it is proportional to N_e , which cannot be estimated

directly but only through Θ . To infer parameter values, the coalescent process is simulated millions of times to obtain probability distributions for each parameter given the data and a specified demographic scenario. Two different but related statistical approaches are used to estimate parameters using coalescent-based methods, namely ML and Bayesian inference. ML approaches aim to maximize the likelihood of the data D given the parameters, $\Pr(D|P)$, where P is the vector of parameters and the genealogy (31, 41, 72). Bayesian inference aims to estimate the posterior probability distributions of the parameters given the data, $\Pr(P|D)$, which can be expressed as a function of the above likelihood and the prior probabilities of the parameters (6, 34, 87). Thus, Bayesian inference makes use of prior knowledge of the parameter values, which is generally expressed as a range or distribution of possible parameter values. The advantage of the Bayesian approach is that one obtains probability distributions rather than single point estimates with confidence intervals as in ML analysis. Priors can be designed to have little influence on the parameter estimates (posterior probability distributions) if the data are sufficiently informative.

Coalescent analysis is powerful, but its application to the analysis of observed data is analytically difficult and computationally intensive. Many of the algorithms that use the coalescent for parameter estimation make use of Markov chain Monte Carlo (MCMC) sampling to obtain distributions for the above likelihoods (71). Most of the programs discussed below use Metropolis-Hastings MCMC (31). MCMC presents challenges of its own, which are discussed elsewhere (11). Consequently, care must be used when using coalescent-based approaches. The user must understand the assumptions and limitations of these methods, and interpret the results accordingly.

COALESCENT-BASED TOOLS

The analysis of population genetic data under the coalescent is an iterative and nonlinear procedure (Figure 4). Careful consideration of the

Genealogy: a gene tree of ancestor to descendant relationships of descent among samples at a given molecular locus (typically a DNA sequence)

TMRCA: time to the most common recent ancestor

Θ : population scaled mutation rate; measure of population genetic diversity, where $\Theta = 4 N_e \mu$ and μ is mutation rate per generation

Bayesian inference: statistical framework that uses probability distributions to quantify uncertainty in inferences from data

Markov chain Monte Carlo (MCMC): algorithm to sample from a probability distribution of interest that cannot be obtained directly

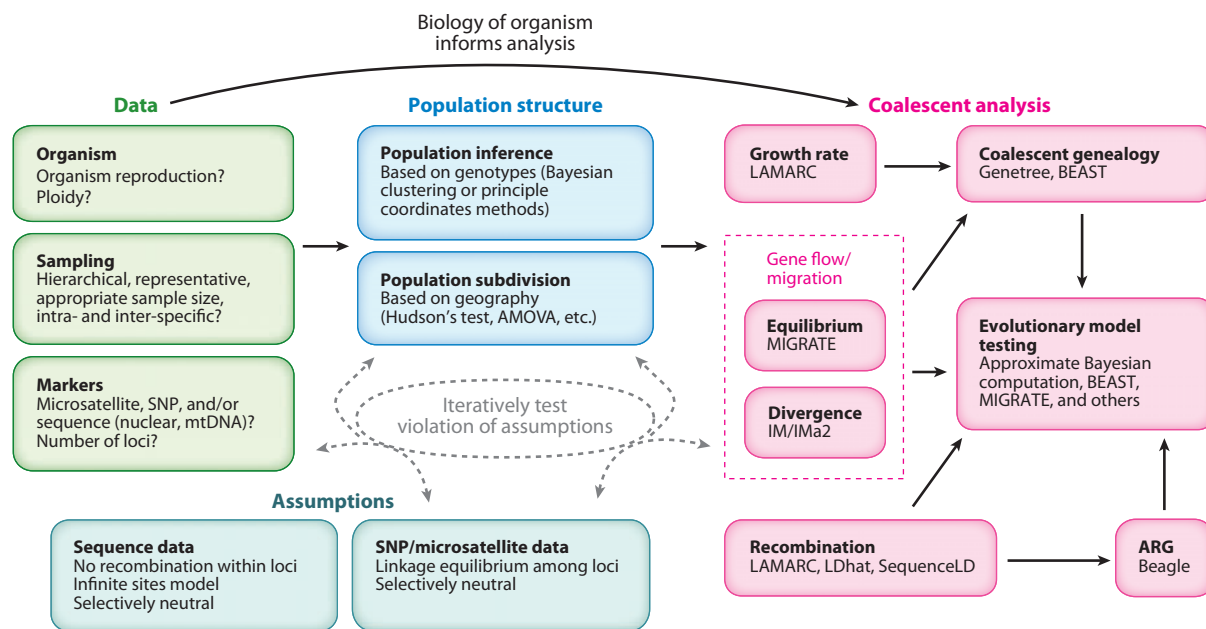


Figure 4

Illustration of the analytical process involved in inferring the evolutionary history of pathogens. Typically, this analysis is iterative, and there is no single best pathway. Furthermore, pathways of analysis have to be adapted for ploidy, mode of reproduction (sexual, asexual, or mixed), molecular marker system, and assumptions underlying the analysis.

biology of the organism and the nature of the genetic data collected cannot be overemphasized, particularly when using coalescent-based methods. The genetic and demographic models assumed by these methods must approximate the biology of the organism. For example, species that are clonal or that have mixed reproduction can complicate analyses that assume Hardy-Weinberg equilibrium (82). Likewise, coalescent analyses that assume a panmictic population are unlikely to be appropriate for a sample that includes isolates from more than one clonal lineage. Species that are polyploid may also require special consideration. An appropriate population sampling strategy is critical, and sampling design considerations are presented elsewhere (3, 56). The genetic markers used should in most cases be selectively neutral, although purifying/negative selection is acceptable, and reflect the evolutionary timescale of interest (Figure 5). Several of the coalescent-based programs discussed below

accommodate multiple loci and different types of markers because using multiple loci significantly increases the power of these analyses, serving as a means of replicating analysis on independent loci. Different loci have different evolutionary histories; thus, inferring the history of a population or species will require integrating information across different genealogies. Although data from multiple unlinked loci are advantageous, recombination within loci can be limiting, as most programs assume no recombination within loci because of the computational challenge posed by recombination (see below). Likewise, the coalescent process may be simplified by the assumption of the infinite sites model for sequence data or the infinite allele model for SSR data in some programs. As a result, some analyses may require removing nucleotide sites exhibiting homoplasy or more than two states. Once the data set is finalized, one must decide which evolutionary processes require parameterization

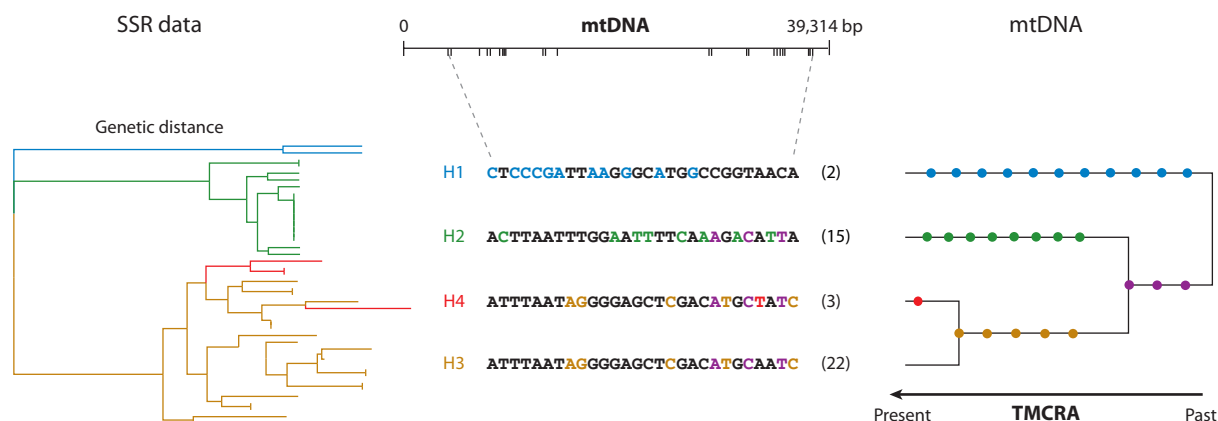


Figure 5

Illustration of molecular markers showing resolution at different evolutionary time scales. A typical coalescent analysis of a population of individuals may use multiple types of markers; for example, simple sequence repeat (SSR; *left*) and sequence based data (mtDNA; *right*) (data from 38, 45). In the center, mtDNA sequence data from *Phytophthora ramorum* is collapsed into haplotypes and frequencies of haplotypes are recorded (see brackets). The gene genealogy is then inferred from the haplotypes. Mutations are mapped on the tree (filled circles) and color coded by mtDNA haplotype lineage. The SSR markers reveal diversity within clonal mtDNA lineages. Mitochondrial haplotypes H1-H4 are assigned to *P. ramorum* clonal lineages as follows: H1 = NA2, H2 = EU1, and H4 and H3 = NA1.

(e.g., Θ , number of populations, migration, population growth, recombination, etc.) and the most appropriate method or program for inferring the needed parameters given the biology of the organism. The analysis should culminate in specifically testing alternate hypotheses of the evolution of the pathogen population (Figure 3). Parameter distribution estimation and the resulting hypothesis testing is where the power of the coalescent and the advantage of these methods over traditional descriptive approaches are most apparent.

Lineage Age and Origin

A major focus in the study of emerging plant pathogens is investigation of the origin of new pathogens. Are emerging pathogens of recent evolutionary origin (i.e., new species or subspecies), or are they just new to a host or environment? Are reemerging pathogens evolving from local populations or are they new introductions? Are variable populations the result of diversification since emergence or multiple events (e.g., migrations or host-shifts)? These types of questions can be

addressed using coalescent-based methods that date events, allowing estimation of the ages of populations, species, or evolutionary lineages. An early tool was the command line program Genetree (4), which can be used to estimate ML coalescent genealogies given sequence data from nonrecombining loci that are consistent with the infinite sites model of DNA sequence evolution (42). Specifically, many coalescent simulations, each with an associated likelihood, are used to estimate the ML topology. This topology can then be used to estimate relative mutation ages and TMRCA of the sample. The result is a genealogy that details the evolutionary relationships among alleles. When there are two (or more) populations represented, one can input or estimate migration rates between the populations and estimate the population of origin for each mutation and of the MRCA of each population and the entire sample. Note that migration rates will influence estimates of populations of origin and that migration rates are not generally obtained independently. Therefore, it is wise to examine the robustness of results to error or uncertainty in migration rate estimates. Population growth can also be

Infinite sites model: assumes that any one nucleotide site has a near zero probability of experiencing more than one mutation

incorporated into the genealogies (41), but there are limitations to the number of parameters for which ML values can reasonably be obtained. Multiple loci can be combined only when they share the same evolutionary history (i.e., are completely linked). An associated program draws the genealogies calculated by the Genetree program. The program SNAP Workbench provides a graphical interface to prepare data files for input and analysis in Genetree (20, 91). Genetree, unlike most of the following programs, is not under active development.

Another program that can be used to estimate the age of lineages is IM, short for Isolation with Migration (54, 55). This command line program implements a specific evolutionary model in which one lineage diverges over time into two lineages, with the possibility of migration during the period of divergence (87). This approach is appropriate for divergence at the species level or for incipient speciation, when populations or lineages are diverging but are not yet reproductively or geographically isolated, including pathogen ecotypes or cryptic species. This is in contrast to Genetree, which is most powerful in examining evolutionary relationships at the population level. IM may be a representative model for pathogens that emerge via host-shift speciation, particularly when some gene flow continues during the speciation process. One can estimate population sizes, migration rates during divergence, and divergence time. An advantage of this program is that multiple unlinked loci can be used, which makes for stronger parameter estimates that may also be more representative of the evolution of the lineages than estimates from single genes. An extension of the IM approach has recently been implemented in IMA2, which can accommodate three or more lineages when the phylogeny of the lineages is known (53).

A third tool for placing the evolution of populations or lineages on a timescale is the program BEAST (25). BEAST, short for Bayesian evolutionary analysis by sampling trees, is particularly powerful in its application of molecular clock models to determine the timing of evolutionary events.

For plant pathogens, which generally do not have fossil records, calibration dates could come from samples obtained from dated historical collections or dated samples from fast-evolving viruses. A compromise alternative that has been used to date genealogies for fungal plant pathogens is to use rough estimates of mutation rates for the molecular clock (for examples, see 97, 98, 103). BEAST has been developed with associated programs to prepare BEAST input files (BEAUti), to examine the output from Bayesian MCMC (Tracer), and to produce publication ready figures (FigTree), all of which have graphical user interfaces.

There are two major challenges in the dating of events for plant pathogens. Dating events in years is dependent on relatively accurate estimation of mutation rates per year or per generation (and therefore of generation time), and these are not well understood for many plant pathogens. Emerging plant pathogens may also lack genetic variation in introduced populations, which may be the only known populations. With little polymorphism, population genetic parameters can be difficult to estimate. This was the case with *Phytophthora ramorum*, in which there was no detectable variation among isolates within clonal lineages for the several nuclear genes that we sequenced (37). Therefore, we estimated lineage age by estimating the coalescent age of mutations at each locus. The youngest mutation shared between lineages was used as a proxy for the minimum age of the lineages. Although this method did not provide a population divergence date, we were able to determine that the lineages were thousands of years old and had not diverged since the emergence of the pathogen in North America and Europe.

When sufficiently polymorphic population data are available, coalescent tools can provide robust estimates of divergence times with confidence intervals (ML) or posterior probability distributions (Bayesian inference). A pertinent example comes from the wheat pathogen *Mycosphaerella graminicola*, which is thought to have emerged with the domestication of wheat by host shifting from wild grass to

wheat in the Fertile Crescent. This scenario was examined through coalescent analysis of DNA sequence data from several loci in a worldwide sample of *M. graminicola* and in a *Mycosphaerella* species found on uncultivated grasses in northwestern Iran (98). In particular, this scenario perfectly fits the isolation with migration model, in which two populations diverge despite continued gene flow between the populations. Genealogies confirmed that one of the populations collected from wild grasses (S1) was the closest relative to *M. graminicola*. The program IM was used to estimate time since divergence of the *M. graminicola* and S1 populations, and migration between populations (Figure 6). Migration was found to be an order of magnitude higher from S1 to *M. graminicola* than the reverse, suggesting an extended period of host shifting from wild grasses to wheat during speciation. IM assumes that the ancestral population splits into two populations and can be used to estimate the fraction of the ancestral population leading to each of the descendent populations. As expected, a small fraction, 2%, of the ancestral population shifted to wheat to become the contemporary *M. graminicola* population. The expansion of the *M. graminicola* population and small, perhaps shrinking, S1 population suggested by the IM results were confirmed by estimation of population size and growth rates using the LAMARC software. Using Kasuga et al.'s (63) estimation of substitution rates in fungi and a generation time of one year, the divergence of the populations was calculated to be approximately 10,500 years ago, which coincides with the expansion of agriculture and domestication of grasses. Note that estimates of substitution rates and generation times are necessarily averages over long time periods during which the life history and environment of the pathogen can change and are very difficult or impossible to experimentally quantify.

Migration

Gene flow among populations or regions is a central concern in plant pathogen epidemiol-

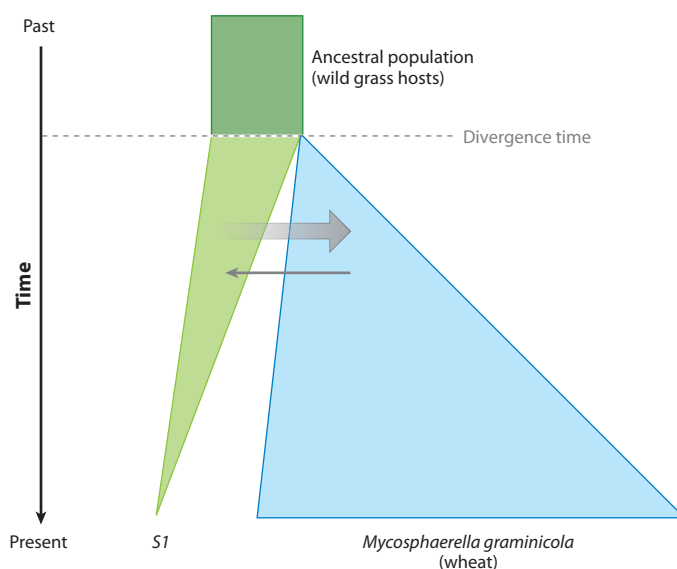


Figure 6

Hypothesis for the divergence of *Mycosphaerella* S1 and *Mycosphaerella graminicola*. The wheat pathogen *M. graminicola* is hypothesized to have evolved from an ancestral population on wild grasses related to the current population on these hosts, which has been designated S1. A very small proportion of the ancestral population is estimated to have founded the *M. graminicola* population followed by a period of continued gene flow mostly from S1 to *M. graminicola*. Widths of populations are proportional to estimated population sizes and widths of arrows are proportional to estimated migration rates.

ogy. Coalescent-based analysis of migration allows the testing of specific migration models rather than crudely estimating migration rates from F_{st} values. For emerging pathogens, we may be interested in whether there is evidence for ongoing migration from source populations or in identifying the direction of migration among populations or regions. When it is known that an emerging pathogen population is from a specific source population and that these populations are diverging (by drift or for a specific trait), the isolation with migration model used by IM may be used to estimate time since divergence and ongoing migration. When investigating equilibrium migration, such as migration within a species with little or no divergence among populations, a better choice may be the program MIGRATE (also known as MIGRATE-n) (10, 12, 13). A suggested rule of thumb for using MIGRATE over IM is that the TMRCA of the sample should be less than

Akaike information criterion (AIC): an index reflecting the goodness of fit of a given model

the divergence time of the populations (11). In the case of emerging pathogens, in which populations may be very young, it may be wise to compare results between MIGRATE and IM. However, an advantage of MIGRATE is that it can accommodate multiple populations and many different migration models, which can be evaluated using either Bayesian inference or ML. One can then test different models of migration using ad hoc approaches (11, 24).

An example comes from our work on the oomycete pathogen *P. ramorum*, which is responsible for the disease sudden oak death in coastal California, sudden larch death in the United Kingdom, and blight or dieback of numerous woody understory species and popular ornamentals, such as rhododendrons, camellias, and viburnums (16, 47, 95). *P. ramorum* was simultaneously discovered in Germany and California and is thought to be an exotic pathogen that was recently introduced to Europe and North America. Three clonal lineages of *P. ramorum* have been described, called EU1, NA1, and NA2, with limited genetic variation within lineages (38, 46). European isolates have been the EU1 lineage and California forest isolates have been NA1. In 2003, the EU1 lineage was found in horticultural nurseries on the West Coast of North America, and a third lineage was identified, NA2. The conventional wisdom is that the EU1 lineage was introduced into Europe and then migrated to North America in the nursery trade. We tested this hypothesis using SSR genotype data, given that SSRs evolve rapidly enough to show variation within clonal lineages, from EU1 isolates collected in Europe and North America (39). Using the program MIGRATE, we used Bayesian inference to estimate migration rate from Europe to North America and from North America to Europe (Figure 7). These estimates indicated that there was higher migration from Europe to North America. We then tested unidirectional migration against bidirectional migration using ML. The highest likelihood model was unidirectional migration from Europe to North America. This model was a significantly better fit to the data by a likelihood

ratio test and Akaike information criterion (AIC), which accounts for the different numbers of parameters in the models under comparison.

Another very different example of the application of the program MIGRATE was to examine species boundaries in *Claviceps purpurea* ecotypes (24), specifically to reject the hypothesis of a single species in favor of a hypothesis of three species with rare migration events among species.

Historical Population Sizes

Emerging plant pathogens may experience severe population bottlenecks during the process of emergence. Small initial population sizes may be the result of introduction to a new region or host, or a new virulent lineage emerging from a single clone or sexual event. Subsequent to this bottleneck, conditions may be conducive to rapid population expansion. Coalescent theory can be used to look for exponential growth in N_e and for the signature of a population bottleneck in DNA sequences. Changes in population size change the rate of coalescent events, thus changes in population size may be detected by departures of the distribution of coalescent events from that expected when population size is constant over time (73). The program LAMARC can be used to estimate population exponential growth rate (g), in the presence of migration and recombination if necessary, using data from a variety of genetic markers (70). BEAST can also be used to estimate changes in population size over time using Bayesian skyline plots (26). This method does not assume constant growth or decline in population size over time and thus can describe population bottlenecks. Of particular interest is the ability to produce Bayesian skyline plots from multigene analysis (52). Skyline plots visualize changes in population size over the time from the MCRA to the present. This is a highly parameterized and computationally intensive model.

BEAST was used to examine the evolutionary history of the rye and barley pathogen *Rhynchosporium secalis*. This pathogen has three host-associated lineages that were estimated to

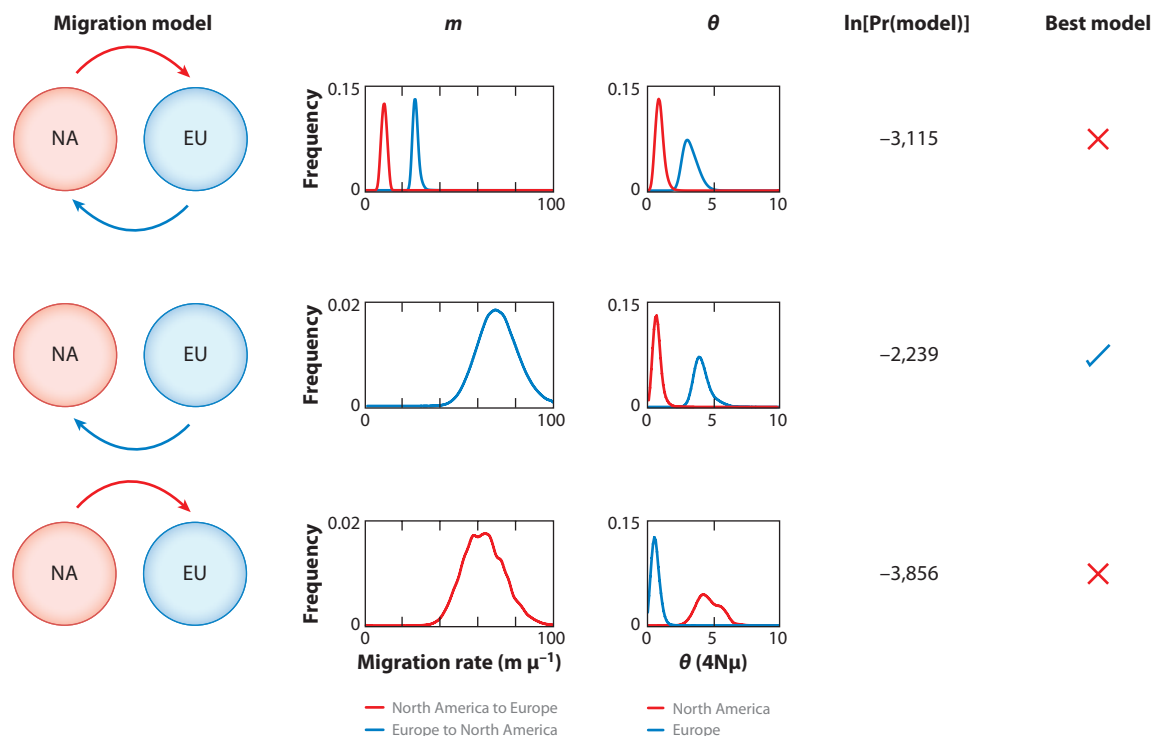


Figure 7

Testing multiple hypotheses of global migration scenarios for the sudden oak death pathogen *Phytophthora ramorum*. Three competing hypotheses are shown including bidirectional or unidirectional migration. Migration rates m and the population scaled mutation rate θ were estimated in several runs through the use of Bayesian approaches using the program MIGRATE (13) as described in 39. The highest Bayesian posterior probability was obtained for the model of unidirectional migration from Europe to North America supporting the hypothesis that *P. ramorum* was introduced to North America from Europe.

have diverged in the last few thousand years using time-calibrated multilocus phylogenies (103). Bayesian skyline plots, depicting the demographic history of the lineages, showed population bottlenecks associated with the presumed host shifts followed by rapid recovery of population sizes.

Recombination

One of the first questions regarding a new or re-emerging pathogen is whether there is sexual reproduction in the population. Sexual reproduction may allow pathogens to rapidly adapt to new environments via shuffling of variation by homologous recombination or introgression of locally adapted alleles by hybridization with resident species. Coalescent theory can be used

to estimate recombination rates in genealogies. However, recombination is a notoriously difficult problem. Many recombination events are not detectable, because the recombination is between very similar sequences, so recombination rates will often be underestimated. A challenge specific to coalescent methods is that with each past recombination there is a splitting event, creating additional ancestors that must be tracked through time. Thus, incorporating recombination into the coalescent is mathematically and computationally difficult. Several methods for estimating recombination rates using the coalescent have been developed (28, 40, 74, 81) and their performance compared (28). The program LAMARC is unique in that it can estimate recombination rates from microsatellite data (70). The programs SequenceLD (29)

SNP: DNA sequence variation in a single nucleotide among two haplotypes at the same locus

ARG: ancestral recombination graph

and LDhat (81), for sequence or single nucleotide polymorphism (SNP) data, use methods that approximate likelihoods and can thus handle larger data sets with reasonable computational speeds. These methods have been applied to detect recombination in the evolutionary history of the exotic and presently clonal sudden oak death pathogen, *P. ramorum* (37).

To visualize recombination in a gene genealogy, one must construct ancestral recombination graphs (ARGs). These graphs trace recombination events back in time following all ancestral sequences and can be quite complex. A minimal ARG includes only the recombination events that are required due to conflicting phylogenetic histories in the data thus facilitating interpretation. The program BEAGLE constructs minimal ARGs, showing an inferred order of recombination and coalescent events back to an inferred or designated root (76). The ARG can be used to infer relationships among recombination blocks (18) and the ancestry of recombinant sequences (37). However, the minimal ARG produced by BEAGLE is not scaled to time like a true coalescent genealogy.

Model Testing Using Approximate Bayesian Computation

We have discussed programs that use the coalescent to examine specific evolutionary models that include migration, recombination, population divergence, and population size change. Realistic evolutionary models for a given population may include most or all of these processes, yet we may not know which processes were involved, when they occurred, or the order in which they occurred (e.g., the ancestry of populations). We would like to be able to evaluate different evolutionary models to not only find the ML or highest probability parameter values, but also estimate the best model. In addition, the best model may not fit a predefined model. Approximate Bayesian computation (ABC) is an approach that addresses both of these needs, essentially by avoiding the evaluation of a complex likelihood function (7, 14, 22). This is an area of rapid development,

both in terms of theory and implementation (i.e., software development), and thus the applications of ABC in the present literature use methodology and computational pipelines that may be outdated or unnecessary in just a few years. The basic idea behind ABC is to test among a small set of evolutionary models using coalescent simulations. For example, one could explicitly test whether an emerging population experienced a severe bottleneck during introduction followed by gene flow with its population of origin or whether a model with no bottleneck or no migration is better supported by the data. The flow of the analysis is roughly as follows: (a) define evolutionary models (limited to processes that can be incorporated into coalescent simulations using existing software), (b) choose summary statistics that capture the structure of the data (e.g., variation within and between populations), (c) simulate millions of data sets that are equivalent to the observed data in number of populations and sampled individuals, calculating summary statistics for each simulated data set, (d) filter the simulations so that only subsets most similar in summary statistics to the observed data are retained for each model, (e) select the model with the highest probability based on the posterior probability of each model, (f) estimate parameter values, and (g) conduct various quality controls on the model selection and parameter estimation (14). ABC has played a major part in bringing statistically rigorous hypothesis testing to the field of phylogeography (8, 67) and will likely do the same for studies of the evolution and emergence of plant pathogens. Studies on the introduction of the western corn rootworm to Europe illustrate the power of ABC over traditional population genetic methods for reconstructing the movement of exotic pests (49, 84).

Cautionary Points

It cannot be overemphasized that the use of coalescent-based programs requires a comprehensive understanding of their assumptions and estimation methods. Data that violate



assumptions of the models may give misleading or nonsensical results. MCMC itself is challenging because it will give misleading results if chains are too short, if they are not efficiently exploring the state space (poor mixing), or if there is too little data to begin with. More problematic is that it may not be obvious that the results are poor unless one knows how to evaluate the performance of the MCMC (11). Similarly, for Bayesian analysis one needs to assess if results are sensitive to choice of priors. Results should always be interpreted with caution and in the context of the pathogen's biology.

Another area for caution is the temptation to convert coalescent time to real time to obtain, for example, years since divergence. For most plant pathogens, this conversion is not straightforward because mutation rates and generation times can be only roughly estimated. For many of these organisms, definition of a generation is not obvious given mixed modes of reproduction and overlapping generations. Furthermore, calculation of these times based on only one or two loci can be very misleading. Ideally, these calculations should be made using many unlinked loci and reported as ranges or with clearly defined confidence intervals.

OTHER LIKELIHOOD-BASED TOOLS

Bayesian Clustering

Other new tools using likelihood or Bayesian approaches have been valuable in the study of plant pathogen population genetic structure, such as Bayesian clustering methods (21, 33, 59, 92). These methods cluster individuals into underlying groups that may not correspond to geographically defined populations and can be used to help define panmictic populations, identify multiple introductions, uncover cryptic species, and otherwise define population structure without a priori knowledge of gene flow and migration patterns. The programs Structure and InStruct can also identify

admixture, referring to individuals in a population with parentage from different genetically distinct populations. Individuals are assigned to clusters assuming linkage equilibrium among loci and Hardy-Weinberg equilibrium within populations. These assumptions are violated in organisms that reproduce asexually. The program InStruct was designed for inbreeding species and does not assume Hardy-Weinberg equilibrium; however, individuals must be diploid and many plant pathogens are not diploid. Principal coordinates analysis has been used as a population genetic model-free alternative to Bayesian clustering (89, 94). However, it cannot be used to estimate the number of clusters and assign individuals to clusters. Discriminant analysis of principle components (DAPC) has recently been proposed as a powerful alternative to both Bayesian clustering and standard principal component analysis (62). This method does not assume a population genetic model and thus is applicable to a wide variety of organisms, population structures, and data types, and also provides a means of estimating the number of clusters and assign individuals to these clusters. However, DAPC cannot identify admixed individuals.

Testing For Selection

Plant pathogens can be subject to intense selection pressures that may affect specific genes for virulence, fungicide resistance, or other traits that can have a large role in host-pathogen interactions. Evolutionary analysis of confirmed or candidate genes for these traits may be informative for management. For example, one could infer specific sites under diversifying selection in a gene, which may identify amino acids or domains of functional importance. One can also identify lineages in which the gene has been subject to selection, which may suggest species for which the gene has been functionally important. There are now a variety of likelihood-based models available for testing for selection on coding sequences (15, 69).



SUMMARY POINTS

1. The genetic processes affecting genetic variability in a population can be thought of as consisting of two simultaneous and independent processes: coalescence and neutral mutation. When simulating the coalescent, mutations are mapped onto the tree after tree building is accomplished.
2. Evolutionary parameters, including relative population sizes, growth rates, and migration rates, and their distributions, can be estimated based on the patterns of variation in population samples using coalescent simulations.
3. Coalescent theory provides a framework for testing competing hypotheses on population genetic and/or evolutionary scenarios.
4. Application of coalescent-based evolutionary tools provides critical novel insights into forces shaping current population structure of exotic and reemerging pathogens that can inform management of emerging epidemics.

FUTURE ISSUES

1. Algorithms and software that currently only implement one or a few analytical techniques must become integrated, user-friendly, multiplatform, and iterative, allowing for testing multiple evolutionary scenarios under different demographic and genetic models.
2. Multilocus approaches will become the standard, allowing for integrated analysis and comparison across unlinked loci.
3. There remains a need to integrate the effect of natural selection on population dynamics will become standard.
4. The multi-species coalescent will become a standard tool for defining species boundaries providing insights well beyond traditional phylogenetics.
5. Coalescent analysis based on genome-wide, high-density SNP data or population genomic studies using high throughput sequencing data will provide yet another level of analysis not currently available.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are very thankful to Eva Stukenbrock and Linda Kohn for their comments on an earlier draft for this manuscript. This work was supported in part by funds from USDA ARS CRIS Project 5358-22000-034-00, and grants from the US Forest Service PSW Sudden Oak Death Research Program, USDA APHIS, and USDA ARS. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official



endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

LITERATURE CITED

1. Anderson JB, Kohn LM. 1995. Clonality in soilborne, plant-pathogenic fungi. *Annu. Rev. Phytopathol.* 33:369–91
2. Atallah ZK, Maruthachalam K, Toit Ld, Koike ST, Michael Davis R, et al. 2010. Population analyses of the vascular plant pathogen *Verticillium dahliae* detect recombination and transcontinental gene flow. *Fungal Genet. Biol.* 47:416–22
3. Avise JC. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43:1192–208
4. Bahlo M, Griffiths RC. 2000. Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* 57:79–95
5. Baumgartner K, Travadon R, Bruhn J, Bergemann SE. 2010. Contrasting patterns of genetic diversity and population structure of *Armillaria mellea* sensu stricto in the Eastern and Western United States. *Phytopathology* 100:708–18
6. Beaumont MA. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153:2013–29
7. Beaumont MA. 2010. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. S.* 41:379–406
8. Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, et al. 2010. In defence of model-based inference in phylogeography. *Mol. Ecol.* 19:436–46
9. Beerli P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol. Ecol.* 13:827–36
10. Beerli P. 2006. Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* 22:341–45
11. Beerli P. 2008. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In *Population Genetics for Animal Conservation*, ed. G Bertorelle, MW Bruford, HC Hauffe, A Rizzoli, C Vernesi, pp. 39–77. Cambridge, MA: Cambridge Univ. Press
12. Beerli P, Felsenstein J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–73
13. Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* 98:4563–68
14. Bertorelle G, Benazzo A, Mona S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* 19:2609–25
15. Bielawski JP, Yang Z. 2005. Maximum likelihood methods for detecting adaptive protein evolution. In *Statistical Methods in Molecular Evolution*, ed. R Nielsen, pp. 103–24. New York: Springer-Verlag
16. Brasier C, Webber J. 2010. Sudden larch death. *Nature* 466:824–25
17. Burdon JJ, Roberts JK. 1995. The population genetic structure of the rust fungus *Melampsora lini* as revealed by pathogenicity, isozyme and RFLP markers. *Plant Pathol.* 44:270–78
18. Carbone I, Jakobek JL, Ramirez-Prado JH, Horn BW. 2007. Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. *Mol. Ecol.* 16:4401–17
19. Carbone I, Kohn LM. 2004. Inferring process from pattern in fungal population genetics. In *Applied Mycology and Biotechnology Series*, ed. DK Arora, pp. 29–58. Amsterdam, The Netherlands: Elsevier Science
20. Carbone I, Liu Y, Hillman BI, Milgroom MG. 2004. Recombination and migration of *Cryphonectria hypovirus 1* as inferred from gene genealogies and the coalescent. *Genetics* 166:1611–29
21. Corander J, Marttinen P, Siren J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinforma.* 9:539
22. Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25:410–18

11. An excellent introduction to the application of coalescent theory to sampled data. A must-read for new users of MIGRATE.

19. A friendly introduction to contemporary evolutionary approaches with a particular emphasis on fungal populations.



37. An example of how the coalescent analysis can be used to assess divergence of lineages in an emerging pathogen.

23. Donnelly P, Tavaré S. 1995. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* 29:401–21
24. Douhan GW, Smith ME, Huyrn KL, Westbrook A, Beerli P, Fisher AJ. 2008. Multigene analysis suggests ecological speciation in the fungal pathogen *Claviceps purpurea*. *Mol. Ecol.* 17:2276–86
25. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214
26. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–92
27. Dutech C, Frabreguettes O, Capdevielle X, Robin C. 2010. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity* 105:220–28
28. Fearnhead P, Donnelly P. 2001. Estimating recombination rates from population genetic data. *Genetics* 159:1299–318
29. Fearnhead P, Donnelly P. 2002. Approximate likelihood methods for estimating local recombination rates. *J. R. Stat. Soc. B* 64:657–80
30. Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol. Biol. Evol.* 23:691–700
31. Felsenstein J, Kuhner MK, Yamato J, Beerli P. 1999. Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. In *Proc. of the AMS Summer Workshop Statistics in Molecular Biology*, ed. F Seillier-Moiseiwitsch, pp. 163–185. Providence, RI: Am. Math. Soc.
32. Fry WE, Goodwin SB, Matuszak JM, Spielman LJ, Milgroom MG, Drenth A. 1992. Population genetics and intercontinental migrations of *Phytophthora infestans*. *Annu. Rev. Phytopathol.* 30:107–30
33. Gao H, Williamson S, Bustamante CD. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–51
34. Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. *Bayesian Data Analysis*. London: Chapman & Hall/CRC Press
35. Giraud T, Fortini D, Levis C, Leroux P, Brygoo Y. 1997. RFLP markers show genetic recombination in *Botryotinia fuckeliana* (*Botrytis cinerea*) and transposable elements reveal two sympatric species. *Mol. Biol. Evol.* 14:1177–85
36. Goodwin SB, Cohen BA, Fry WE. 1994. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proc. Natl. Acad. Sci. USA* 91:11591–95
37. Goss EM, Carbone I, Grünwald NJ. 2009. Ancient isolation and independent evolution of the three clonal lineages of the exotic sudden oak death pathogen *Phytophthora ramorum*. *Mol. Ecol.* 18:1161–74
38. Goss EM, Larsen M, Chastagner GA, Givens DR, Grünwald NJ. 2009. Population genetic analysis infers migration pathways of *Phytophthora ramorum* in US nurseries. *PLoS Pathog.* 5:e1000583
39. Goss EM, Larsen M, Vercauteren A, Werres S, Heungens K, Grünwald NJ. 2011. *Phytophthora ramorum* in Canada: evidence for migration within North America and from Europe. *Phytopathology* 101:166–71
40. Griffiths RC, Marjoram P. 1996. Ancestral inference from samples of DNA sequences with recombination. *J. Comput. Biol.* 3:479–502
41. Griffiths RC, Tavaré S. 1994. Sampling theory for neutral alleles in a varying environment. *Philos. T. R. Soc. B* 344:403–10
42. Griffiths RC, Tavaré S. 1995. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.* 127:77–98
43. Grünwald NJ, Flier WG. 2005. The biology of *Phytophthora infestans* at its center of origin. *Annu. Rev. Phytopathol.* 43:171–90
44. Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE. 2003. Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* 93:738–46
45. Grünwald NJ, Goss EM. 2009. Genetics and evolution of the sudden oak death pathogen *Phytophthora ramorum*. In *Oomycete Genetics and Genomics: Biology, Interactions, and Research Tools*, ed. KH Lamour, S Kamoun, pp. 179–96. Hoboken, New Jersey: John Wiley & Sons
46. Grünwald NJ, Goss EM, Ivors K, Garbelotto M, Martin FN, et al. 2009. Standardizing the nomenclature for clonal lineages of the sudden oak death pathogen, *Phytophthora ramorum*. *Phytopathology* 99:792–95

47. Grünwald NJ, Goss EM, Press CM. 2008. *Phytophthora ramorum*: a pathogen with a remarkably wide host-range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Mol. Plant. Pathol.* 9:729–40
48. Grünwald NJ, Hoheisel G-A. 2006. Hierarchical analysis of diversity, selfing, and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*. *Phytopathology* 96:1134–41
49. Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A. 2010. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 104:88–99
50. Halkett F, Coste D, Rivas Platero GG, Zapater MF, Abadie C, Carlier J. 2010. Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Mol. Ecol.* 19:3909–23
51. Hein J, Schierup MH, Wiuf C. 2005. *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. New York: Oxford Univ. Press. 276 pp.
52. Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289
53. Hey J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol.* 27:905–20
54. Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–60
55. Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. USA* 104:2785–90
56. Hillis DM, Moritz C, Mable BK. 1996. *Molecular Systematics*. Sunderland, MA: Sinauer Associates
57. Hudson RR. 1990. Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7:1–44
58. Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model. *Bioinformatics* 18:337–38
59. Huelsenbeck J, Andolfatto P. 2007. Inference of population structure under a Dirichlet process model. *Genetics* 175:1787–802
60. Ivors K, Garbelotto M, Vries IDE, Ruyter-Spira C, Hekkert BT, et al. 2006. Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Mol. Ecol.* 15:1493–505
61. Ivors KL, Hayden KJ, Bonants PJM, Rizzo DM, Garbelotto M. 2004. AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycol. Res.* 108:378–92
62. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94
63. Kasuga T, White TJ, Taylor JW. 2002. Estimation of nucleotide substitution rates in Eurotiomycete fungi. *Mol. Biol. Evol.* 19:2318–24
64. Kingman JFC. 1982. On the genealogy of large populations. *J. Appl. Probab.* 19:27–43
65. Kingman JFC. 1982. Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, ed. G Koch, F Spizzichino, pp. 97–112. Amsterdam: North-Holland
66. Kingman JFC. 1982. The coalescent. *Stoch. Process. Appl.* 13:235–48
67. Knowles LL. 2009. Statistical phylogeography. *Annu. Rev. Ecol. Evol. S.* 40:593–612
68. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, et al. 2008. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* 40:1068–75
69. Kosakovsky Pond SL, Poon AFY, Frost SDW. 2009. Estimating selection pressures on alignments of coding sequences. In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, ed. P Lemey, M Salemi, A-M Vandamme, pp. 419–490. Cambridge, UK: Cambridge Univ. Press
70. Kuhner MK. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–70
71. Kuhner MK. 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol. Evol.* 24:86–93
72. Kuhner MK, Yamato J, Felsenstein J. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–30
73. Kuhner MK, Yamato J, Felsenstein J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429–34



88. A very readable review on coalescent theory with illustrative genealogies useful for gaining an intuitive understanding of the coalescent.

91. A helpful program that walks the user step-by-step through the analysis of population genetic data using some of the methods discussed in this review.

98. An example of the application of the IM model and other coalescent-based methods to a plant pathogen that domesticated with its host.

74. Kuhner MK, Yamato J, Felsenstein J. 2000. Maximum likelihood estimation of recombination rates from population data. *Genetics* 156:1393–401
75. Lynch M. 1988. Estimation of relatedness by DNA fingerprinting. *Mol. Biol. Evol.* 5:584–99
76. Lyngsø RB, Song YS, Hein J. 2005. Minimum recombination histories by branch and bound. *Lect. Notes Comput. Sci.* 3692:239–50
77. Malvarez G, Carbone I, Grünwald NJ, Subbarao KV, Schafer M, Kohn LM. 2007. New populations of *Sclerotinia sclerotiorum* from lettuce in California and peas and lentils in Washington. *Phytopathology* 97:470–83
78. McDonald BA. 1997. The population genetics of fungi: tools and techniques. *Phytopathology* 87:448–53
79. McDonald BA, Linde C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40:349–79
80. McDonald BA, Martinez JP. 1990. DNA restriction fragment length polymorphism among *Mycosphaerella graminicola* (anamorph *Septoria tritici*) isolates collected from a single wheat field. *Phytopathology* 80:1368–73
81. McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160:1231–41
82. Milgroom MG. 1996. Recombination and the multilocus structure of fungal populations. *Annu. Rev. Phytopathol.* 34:457–77
83. Milgroom MG, Peever TL. 2003. Population biology of plant pathogens—the synthesis of plant disease epidemiology and population genetics. *Plant Dis.* 87:608–17
84. Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, et al. 2005. Multiple transatlantic introductions of the western corn rootworm. *Science* 310:992
85. Mohle M. 2000. Ancestral processes in population genetics—the coalescent. *J. Theor. Biol.* 204:629–38
86. Montarry J, Andrivon D, Glais I, Corbiere R, Mialdea G, Delmotte F. 2010. Microsatellite markers reveal two admixed genetic groups and an ongoing displacement within the French population of the invasive plant pathogen *Phytophthora infestans*. *Mol. Ecol.* 19:1965–77
87. Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–96
88. Nordborg M. 2001. Coalescent theory. In *Handbook of Statistical Genetics*, ed. DJ Balding, MJ Bishop, C Cannings, pp. 179–212. Chichester, UK: John Wiley & Sons
89. Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6:288–95
90. Peever TL, Salimath SS, Su G, Kaiser WJ, Muehlbauer FJ. 2004. Historical and contemporary multilocus population structure of *Ascochyta rabiei* (teleomorph: *Didymella rabiei*) in the Pacific Northwest of the United States. *Mol. Ecol.* 13:291–309
91. Price EW, Carbone I. 2005. SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics* 21:402–4
92. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59
93. Prospero S, Hansen EM, Grünwald NJ, Winton LM. 2007. Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Mol. Ecol.* 16:2958–73
94. Reeves PA, Richards CM. 2009. Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS ONE* 4:e4269
95. Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST. 2002. *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Litocarpus densiflorus* in California. *Plant Dis.* 86:205–14
96. Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3:380–90
97. Rosendahl S, McGee P, Morton JB. 2009. Lack of global population genetic differentiation in the arbuscular mycorrhizal fungus *Glomus mosseae* suggests a recent range expansion which may have coincided with the spread of agriculture. *Mol. Ecol.* 18:4316–29
98. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. 2007. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol. Biol. Evol.* 24:398–411

99. Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–60
100. Tavaré S. 1984. Lines-of-descent and genealogical processes, and their applications in population-genetics models. *Adv. Appl. Probab.* 16:27–27
101. Wakeley J. 2009. *Coalescent Theory: An Introduction*. Greenwood Village, CO: Roberts & Company
102. Watterson GA. 1975. On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* 7:256–76
103. Zaffarano PL, McDonald BA, Linde CC. 2008. Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution* 62:1418–36

101. A thorough introduction to coalescent theory and its applications.

RELATED RESOURCES

- Lemey P, Salemi M, Vandamme A-M. 2009. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. New York: Cambridge University Press. 723 pp.
- Nielsen R. 2005. *Statistical methods in molecular evolution*. New York: Springer. 504 pp.

