# Population Genomics of Fungal and Oomycete Pathogens

Niklaus J. Grünwald,[1,2,4] Bruce A. McDonald,[3] and Michael G. Milgroom[4,*]

[1]Horticultural Crops Research Laboratory, USDA Agricultural Research Service, Corvallis, Oregon 97330; email: nik.grunwald@ars.usda.gov

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331

[3]Plant Pathology, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland; email: bruce.mcdonald@usys.ethz.ch

[4]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, New York 14853; email: mgm5@cornell.edu

Annu. Rev. Phytopathol. 2016. 54:323–46

First published online as a Review in Advance on June 8, 2016

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

This article's doi: 10.1146/annurev-phyto-080614-115913



*Corresponding author

## Keywords

adaptation, comparative genomics, evolution, genome-wide association, population genetics, QTL mapping

## Abstract

We are entering a new era in plant pathology in which whole-genome sequences of many individuals of a pathogen species are becoming readily available. Population genomics aims to discover genetic mechanisms underlying phenotypes associated with adaptive traits such as pathogenicity, virulence, fungicide resistance, and host specialization, as genome sequences or large numbers of single nucleotide polymorphisms become readily available from multiple individuals of the same species. This emerging field encompasses detailed genetic analyses of natural populations, comparative genomic analyses of closely related species, identification of genes under selection, and linkage analyses involving association studies in natural populations or segregating populations resulting from crosses. The era of pathogen population genomics will provide new opportunities and challenges, requiring new computational and analytical tools. This review focuses on conceptual and methodological issues as well as the approaches to answering questions in population genomics. The major steps start with defining relevant biological and evolutionary questions, followed by sampling, genotyping, and phenotyping, and ending in analytical methods and interpretations. We provide examples of recent applications of population genomics to fungal and oomycete plant pathogens.

## INTRODUCTION

**Population genetics:** the study of evolutionary changes in populations resulting from mutation, selection, gene flow, random genetic drift, and recombination

**Single-nucleotide polymorphisms (SNPs):** variation in a single nucleotide that occurs at a specific position in the genome

**Population genomics:** population genetics based on large-scale SNP typing, distributed throughout the genome and derived from whole-genome sequencing or reduced-representation library sequencing

**Resequencing:** genome sequencing of multiple individuals at relatively low read depth; sequences are assembled by alignment to a common reference genome sequence

Recent technological advances in both high-throughput sequencing and computational tools have made it possible to sequence and analyze whole genomes of many plant pathogens. These technologies continue to advance rapidly, and costs have declined to the point that it is becoming affordable to sequence genomes of many individuals within a species. Advanced sequencing technologies have enabled the current genomics era, in which we are learning a great deal about genome organization and structure at an unprecedented level of detail (90).

The genomics revolution has also impacted population genetics studies that seek to understand the ecological, evolutionary, and demographic processes acting on genomes and populations. The large number of markers now available on a whole-genome scale provides fine-grained detail regarding processes of divergence, differentiation, gene conversion, and recombination. New sequencing technologies also provide unprecedented scales of nucleotide sequence data for studying the evolution of specific genes and entire chromosomes. For example, such studies have led to a greater realization of the importance of horizontal gene transfer (HGT) and interspecific hybridizations in pathogen origins. HGT of pathogenicity gene clusters has featured prominently in the evolution of bacterial pathogens (100), but more recently HGT has also been well documented in fungi and oomycetes (35, 38, 77). All told, these types of population genetics studies have the potential to provide a deeper and finer-grained understanding of the biology, genetics, and evolution of plant pathogens.

Genome-scale sequence data enable advanced studies of nonmodel organisms that were inconceivable until recently because of limitations imposed by high costs and the lack of extensive genetic resources, such as reference genomes and collections of knockout mutants. The increasing availability of complete genome sequences, coupled with the ability to genotype thousands of single-nucleotide polymorphisms (SNPs) simultaneously or resequence whole genomes of many individuals within a species, makes it possible to go far beyond traditional population genetics analyses. Thus, population genomics represents a new area of research that portends significant conceptual breakthroughs in how we view the genetics, evolution, and emergence of plant pathogens.

We begin this review with a definition of population genomics and a discussion of the types of questions that can be addressed. This is followed by a discussion of the methodological challenges likely to be encountered in population genomics projects and by examples of applications of population genomics based on recent studies of plant-pathogenic fungi and oomycetes. We finish by speculating about likely future developments in this area. We believe that this review provides a useful entry point for anyone considering a population genomics project aiming to discover genetic mechanisms underlying evolutionary processes based on population-scale genome genotyping or resequencing of fungi and oomycetes.

## WHAT IS POPULATION GENOMICS?

Population genomics can be viewed as an extension of population genetics, with several fundamental methodological differences that radically alter and expand the types of questions that can be addressed. The simplest perspective is that population genomics is population genetics using a large number of markers with a known distribution throughout the genome (85). With this perspective, the questions addressed in population genomics are essentially identical to those in population genetics. As in population genetics, the types of questions addressed in population genomics fit into three broad categories (44): genome-wide evolutionary and demographic processes

affecting population structure; evolutionary processes affecting speciation and the divergence of closely related taxa; and locus-specific effects, particularly selection, acting on specific genes or chromosomes that affect adaptation or defined phenotypes (**Table 1**). To date, most studies of population genetics of fungal and oomycete plant pathogens have used neutral genetic markers with the aim of inferring genome-wide demographic effects caused by migration, bottlenecks, and mating systems (42, 70). The underlying principle is that evolutionary history and demography affect all neutral loci equally across the entire genome. Therefore, genotyping a very large number of loci (e.g., thousands of SNPs) or sequencing the genomes of many individuals (resequencing) makes inferences of genome-wide effects more robust compared to studies based on a limited number of genetic markers. However, the larger number of genetic markers alone does not fundamentally change the types of questions being asked about genome-wide effects.

By contrast, population genomics marks a shift in emphasis from traditional population genetics, in which genome-wide effects are inferred from a relatively small number of neutral markers, to an emphasis on both genome-wide and locus-specific effects. The availability of genome-wide measures of diversity, determined by genome scans of thousands of SNPs, makes it possible to address some population genetics questions—particularly those involving locus-specific effects—that were practically impossible previously. The shift in emphasis from genome-wide effects and demographic considerations to locus-specific effects and a functional understanding of diversity is a key distinction between traditional population genetics and population genomics. This distinction highlights the potential of population genomics to address conceptually novel questions, such as what are the genes underlying adaptation or the divergence of closely related taxa, and what is the genetic basis of phenotypic variation and evolutionary change? In plant pathology, a population genomics approach can be used for identifying candidate genes involved in pathogenicity, virulence (or aggressiveness), host specialization, fungicide resistance, and adaptation to different environments (87). Identifying such genes and elucidating their mechanisms are major goals in the study of plant-microbe interactions. Therefore, population genomics complements methods in molecular biology and functional genomics aimed at achieving these goals.

Three main types of questions are addressed in population genomics, each of which is described in subsections below. Research approaches to answering these questions vary depending on the specific questions and the biology of the organism studied. These research approaches include (*a*) detailed genetic analyses of demographic processes, (*b*) comparative genomics of closely related species, (*c*) identification of genes or genetic regions under selection, and linkage analyses in (*d*) genome-wide association studies (GWASs) or (*e*) segregating populations resulting from crosses to identify genetic loci that explain observed phenotypes, i.e., the genetic architecture of phenotypic variation (**Table 1**). Demography or traditional population genetics provides an understanding of basic pathogen biology that can inform many disease management strategies (42, 66, 70). Comparative population genomics of sister taxa uncovers genomic regions that diverged during or after speciation as a first-level identification of genomic regions worthy of further investigation. Genome scans for genomic regions under positive selection can identify genes involved in adaptation, both within and between closely related species, although the associated phenotype may not be known in advance. GWASs and linkage mapping can identify genomic regions associated with particular phenotypes, leading to identification of quantitative trait loci (QTLs) or genes associated with a measured phenotype. Thus, population genomics encompasses a continuum of research approaches with the ultimate goal of finding the causal genetic loci and allelic variants that explain adaptation or particular phenotypes and places them in the context of an organism's demographic and evolutionary history. These approaches are discussed in the following sections that are organized according to the major questions addressed.

**Genome scan:** several techniques whereby the entire genome sequence(s) of an individual or population is searched systematically for sequences associated with specific attributes, such as regions of unusually high or low diversity

**Genome-wide association studies (GWASs):** analyses that correlate SNPs with phenotypes in individuals drawn from large, random-mating populations to identify candidate genes explaining the trait variance

**Genetic architecture:** the genetic basis of a measured phenotype, for example, quantitative or Mendelian inheritance involving additive or dominant interactions among few or many genes

**Table 1 Summary of the types of questions, approaches, and methods used in population genomics**

| Approaches → / Steps ↓ | Demographic processes | Comparative population genomics | Genome scans for genes under selection | Genome-wide association studies (GWASs) | Quantitative trait locus (QTL) mapping using crosses |
|---|---|---|---|---|---|
| **Biological questions** | **What evolutionary and demographic processes (genome-wide effects) shape population structure?**<br>■ Demographic history<br>■ Gene flow; migration<br>■ Drift; bottlenecks; founder effects<br>■ Mating systems; recombination | **What processes affect the divergence of closely related taxa?**<br>■ Host specialization<br>■ Domestication<br>■ Horizontal gene transfer<br>■ Interspecific hybridization<br>■ Selective sweeps<br>■ Genome structure (indels, inversions, duplications, and translocations) | **Which traits and genes are under selection for adaptation?**<br>■ Selection at the locus level to find genes involved in adaptation, for example, to different host cultivars<br>■ Selective sweeps | **Which genes are responsible for specific phenotypes?**<br>■ Connect phenotype to genotype<br>■ Genetic architecture of traits | **Which genes are responsible for specific phenotypes?**<br>■ Connect phenotype to genotype<br>■ Genetic architecture of traits<br>■ Recombination maps<br>■ Genome structure (indels, inversions, duplications, and translocations) |
| **Sampling strategy** | Hierarchical sampling of natural populations across geographical or ecological gradients (hundreds to thousands of isolates) | Sample natural populations of sister species (tens of isolates), in sympatry where possible | Hierarchical sampling of natural populations across geographical or ecological gradients (hundreds to thousands of isolates) | Sample individuals differing in phenotypes from well-characterized, randomly mating populations (hundreds to thousands of isolates) | Sample sexual progeny from controlled crosses (hundreds to thousands of isolates) |
| **Genotyping strategy** | ■ Genome resequencing<br>■ Genotyping by sequencing (GBS) or restriction site–associated DNA sequencing (RADseq) | ■ De novo assembly of complete genome sequences<br>■ Assembly based on reference genome of one of the species | ■ Genome resequencing<br>■ GBS or RADseq | ■ Genome resequencing<br>■ GBS or RADseq | ■ Genome resequencing<br>■ GBS or RADseq |

| | | | | | |
|---|---|---|---|---|---|
| **Phenotyping strategy** | Automated image analysis of traits in hundreds to thousands of isolates; distribution of quantitative trait variance within and among populations, $Q_{ST}$ analysis (54) to populations | Automated image analysis or manual measurements of traits in several isolates from each species to determine host range, thermal tolerances, etc. | Automated image analysis or manual measurements of relevant traits in tens of isolates sampled from each geographical or ecological gradient | Automated image analysis of relevant traits in hundreds of individuals sampled from each geographical or ecological gradient | Automated image analysis of relevant traits in all offspring |
| **Analytical methods and inferences** | ■ Population differentiation $F_{ST}$, AMOVA (Analysis of MOlecular VAriance) to VAriance), $Q_{ST}$ analysis (54)<br>■ Migration rate<br>■ Genome-wide linkage disequilibrium | ■ $F_{ST}$ scans<br>■ $d_N/d_S$ ratios<br>■ $K_a/K_s$ test<br>■ Genome regions under diversifying, positive, and purifying (negative) selection | ■ $F_{ST}$-$Q_{ST}$ comparisons to determine type of selection operating on different traits (54)<br>■ $F_{ST}$ scans<br>■ Tajima's $D$<br>■ $d_N/d_S$ ratios<br>■ Hot spots of recombination and selection<br>■ Genome regions under diversifying, positive, and purifying (negative) selection | ■ Correct for underlying population structure<br>■ Quantitative trait nucleotide (QTN) associations with trait variance (Manhattan plots)<br>■ Candidate gene discovery<br>■ Recombination<br>■ Linkage disequilibrium (LD) | ■ Genome scans to identify QTLs<br>■ Recombination<br>■ LD<br>■ Candidate gene discovery |

## Demographic Processes and Genome-Wide Effects

The first major type of biological question is an extension of the demographic studies conducted previously in population genetics. As in traditional population genetics, population genomics can address fundamental questions such as what evolutionary and demographic processes (genome-wide effects) shape population structure. This type of question is addressed using neutral genetic markers and is usually independent of selection (see next section for a discussion of selection). Therefore, the evolutionary processes of most interest are migration, random genetic drift, and recombination. These processes affect all neutral loci more or less equally. Most studies to date in population genetics of pathogenic fungi and oomycetes deal with these types of genome-wide, demographic effects (42, 70). The availability of a large number of genetic markers distributed throughout the genome (e.g., SNPs) makes inferences more robust but also makes it possible to refine inferences by uncovering variation that was previously undetectable. For example, detecting fine-grained variation within lineages of clonal organisms makes it possible to make inferences about the migration or evolution of a specific lineage (72, 79) or to detect rare recombination events between lineages (71).

A conceptually more interesting advance in population genetics, made possible by the large number of markers, is to refine our understanding of the relative roles of migration and drift. For example, to estimate genetic differentiation between populations, $F_{ST}$ is estimated for each locus and averaged across loci to make inferences for the entire genome. Because the goal is to make inferences about migration and drift from $F_{ST}$, all loci included in an analysis are assumed to be selectively neutral, although this assumption is rarely tested. By doing an $F_{ST}$ scan of many SNPs across the genome, it is possible to identify loci that show much less or much greater differentiation (very low or very high $F_{ST}$) than expected from random variation among loci caused by drift. Such outlier loci are more likely to be under strong selection, as explained in the next section. Therefore, identifying and eliminating such outliers improves estimates of demographic parameters based on the remaining neutral loci.

Population genomics data sets can also be used to identify hot spots and cold spots of recombination (20, 93) and may prove useful for identifying possible hot spots of mutation as well. The concept of genomic regions that differ in their rates of evolution as a result of underlying differences in recombination, mutation, or genome structure has been encapsulated in the concept of a two-compartment genome (21, 76), similar to core and accessory genomes in bacteria. In this model, selection favors the sequestering of housekeeping genes with vital conserved functions into genomic regions that experience less recombination and mutation (e.g., GC-rich, gene-dense regions located near centromeres). By contrast, genes under diversifying selection that evolve rapidly (e.g., effector genes involved in arms-race coevolution) are selected to occur in more dynamic genomic regions. These regions are characterized by higher rates of recombination (e.g., in telomeric or subtelomeric regions). They are also populated by repeat-rich transposable elements or other repetitive DNA families prone to high mutation rates due to a combination of polymerase error, transposition, and mismatch recombination.

## Identifying Genes Under Selection

One of the most significant advances made possible in population genomics is the identification of traits and genes that are under positive selection for adaptation. Variation in genome sequences among individuals reflects their evolutionary histories, including signatures of selection. Genome scans of individuals within a population can detect regions with marked reductions in genetic diversity relative to the rest of the genome. Such local reductions are indicative of selective sweeps, often reflecting adaptation, in which a gene under strong positive selection goes to fixation. This

process also sweeps away diversity at closely linked loci because of hitchhiking selection (12). Linkage disequilibrium (LD) between the selected gene and the flanking regions decays over time in recombining populations; hence, the signatures of selective sweeps are transient in the evolution of chromosome segments.

**Linkage disequilibrium (LD):** the nonrandom association of DNA sequences as a result of an absence of recombination between those sequences

A complementary approach is to scan the genome for regions showing greater or less divergence (measured as $F_{ST}$) between populations or closely related taxa than expected for neutral loci. Outlier loci (i.e., those with $F_{ST}$ in the extreme tails of the distribution) are often interpreted as being under selection. Regions of less than expected divergence (outliers with low $F_{ST}$) may indicate regions under purifying selection that are conserved in all populations. Conversely, outliers with high values of $F_{ST}$ are interpreted as being under positive selection and therefore as candidate genes contributing to local adaptation. For example, if local adaptation favors one allele in one population and a different allele in another population, divergence at this locus would be greater than that caused by random genetic drift at neutral loci (31, 68, 82). Signatures of positive selection are often associated with gene-for-gene coevolution between pathogens and their hosts (10, 68, 104), including the emergence of host specialization. In addition to identifying candidate genes for further study of locus-specific effects, inferences of genome-wide effects based on neutral loci are improved by eliminating loci under selection (both high and low $F_{ST}$), as mentioned in the previous section.

Until recently, it was only possible to investigate candidate genes known a priori to be associated with phenotypes. In the candidate gene approach, researchers start by searching for homologs to genes with known functions in other organisms and testing their function or association with phenotypes in another organism. This approach has been applied successfully for understanding phenotypes that affect fitness, such as effectors or genes conferring fungicide resistance (10, 18, 34, 68, 79, 82). By contrast, a significant advantage of population genomics is the ability to search for genes under selection without any prior knowledge of traits they affect or their effects on fitness (e.g., 31, 88). This approach has been referred to as reverse genetics or reverse ecology (30, 31).

Genome scans for selection have the potential to identify genes involved in adaptation either within or between species. For example, comparative genomics of closely related pathogens can identify genes involved in speciation (**Table 1**), such as genes underlying host specialization or adaptation to domesticated crops (11, 88). Although the divergence and phylogenetic relationships of closely related species can be inferred by sequencing a relatively small number of genes, this approach cannot identify particular traits or genes contributing to speciation. Similarly, within-species genome scans among individuals sampled from different populations, such as different crop cultivars, have the potential to identify candidate genes involved in virulence or adaptation to different host genotypes or environments (18, 79).

Recombination rates can affect how much of a genome is affected by selection. When recombination rates are high, especially in natural populations of randomly mating species, selection may have very localized, locus-specific effects on the genome. In highly recombining pathogens, a recent selective sweep (characterized by regions with low diversity within populations) may be evident for a relatively small number of generations after a favorable allele has gone to fixation because LD between the favored allele and neighboring genes decays rapidly, causing the size of the chromosomal region affected by the sweep to decrease and eventually disappear. When LD decays over a scale of hundreds of base pairs, one needs a large number of markers distributed throughout the genome for a selection scan to be useful. The sequencing technologies used for population genomics provide the markers needed to detect genes under selection in highly recombined populations.

At the other extreme, because recombination is rare (or nonexistent) in clonal organisms, selection on any one locus in clonal populations leads to hitchhiking selection affecting the entire

genome, not just the neighboring genes, making it impossible to identify locus-specific effects by analyzing genome scans. An extreme example is the mitochondrial genome, which is typically nonrecombining (94–97). Knowledge of the mating system of a species is important to be able to differentiate locus-specific effects from genome-wide effects. This is a crucial consideration because many plant pathogens are clonal and lack regular sexual reproduction. Unfortunately, genome scans for finding genes or regions under selection are mostly irrelevant for clonal populations because of the extent of LD in the absence of repeated recombination.

**Quantitative trait nucleotides (QTNs):** In a GWAS, refers to SNPs statistically associated with a trait. In QTL mapping, refers to polymorphic nucleotides explaining the trait

## Identifying Genes Underlying Known Phenotypes

Another approach to studying locus-specific effects is to identify alleles associated with known phenotypes using GWASs or QTL mapping (**Table 1**). GWASs are conducted by sampling many individuals from a naturally recombining population and genotyping them with large numbers of markers distributed throughout the genome or by whole-genome resequencing. Genotyping is coupled with phenotyping of sampled individuals to enable statistical tests for associations between genotypes and phenotypes. When genotyping is done with SNPs in a GWAS, quantitative trait nucleotides (QTNs) refer to SNPs that are statistically associated with a trait. Linkage mapping, or QTL mapping, also aims to find genes associated with phenotypes, but this is accomplished by analyzing segregating progeny from controlled crosses instead of individuals sampled from nature. When genotyping is done with SNPs in this context, QTNs refer to the polymorphic nucleotides in the alternative parental alleles that are ultimately responsible for the observed trait variance in the offspring. A relatively small number of QTL mapping studies were conducted in fungal plant pathogens before the widespread availability of short-read deep sequencing (23, 49, 61). These studies found statistically significant QTLs, but because of limitations in genotyping and phenotyping technologies the identified regions contained too many candidate genes to be considered for functional studies. In general, the ability to identify relevant genes improves with increasing numbers of progeny and markers and with increasing precision of the phenotyping.

A pivotal difference of GWASs and QTL mapping compared to selection scans is that the former start with known phenotypes and look for underlying genes, whereas the latter identifies genes under selection, for which phenotypes may or may not be known. In plant pathology, we are often interested in identifying specific genes or alleles underlying adaptive phenotypes such as pathogenicity, virulence, fungicide resistance, and host specialization. Genes identified as being under selection or associated with phenotypes involved in adaptation become candidates that require further experimental testing to determine how they affect phenotypes, e.g., by conducting knockout or allele-swap experiments. One of the upcoming challenges will be to determine the effects of specific alleles on fitness and their functional role in adaptation in nature.

Depending on the biology of the organism being studied, GWASs and QTL mapping have different advantages. For many species, randomly mating populations have relatively little LD because of many generations of recombination involving many individuals. In this case, markers associated with a particular phenotype are likely to be in a region close to or in the functional gene underlying that phenotype. When LD is minimal, associations with phenotypes are not likely to be found unless genotyping is done with a very large number of markers distributed throughout the genome. Advanced sequencing technologies provide the necessary markers. With QTL mapping, linkage groups from controlled crosses are defined by the crossovers occurring between markers in the mapping populations. Markers with significant associations to phenotypes typically identify genomic regions too large to directly identify the functional gene(s). For this reason, it may be especially powerful to combine GWASs with QTL mapping (33), using a GWAS to determine the genetic architecture of a trait and identify individuals with relevant markers and phenotypes to cross for QTL mapping.

The choice between a GWAS and QTL mapping may be determined, in part, by whether it is possible to make controlled crosses in the laboratory. Some fungi and oomycetes may be randomly mating in nature but cannot be crossed easily in the lab. For both methods, large numbers of isolates may need to be analyzed to identify genes and genomic regions associated with small effects on overall trait variance (51). Another significant limitation is the effort needed to obtain accurate phenotypes for hundreds or thousands of individuals, especially for quantitative traits. A major limitation of QTL mapping is that only the polymorphic alleles segregating in a particular experimental population can be mapped and identified; gene sequences that do not differ between the parents are not mapped as a QTL even if the associated genes make an important contribution to the trait in natural populations. To overcome this limitation, it is common to make more than one set of crosses or to construct specialized mapping populations that include multiple parents. A disadvantage in GWASs is that only alleles present at relatively high frequencies are likely to be detected, unless thousands of strains are analyzed. An important caveat for a GWAS is the need to determine whether individuals are sampled from a population with an underlying genetic structure. Associations between markers and phenotypes can be caused by admixture if individuals are sampled from two or more genetically differentiated subpopulations. In this case, population structure needs to be corrected statistically, leading to a reduction in the statistical power to detect significant associations.

GWASs and QTL mapping require that the organisms being studied reproduce sexually, either in nature or in the lab, respectively. Unfortunately, as with genome scans for selection, these approaches to studying locus-specific effects are not relevant for clonal populations in which entire genomes are inherited asexually as a single, nonrecombining linkage group. For clonal organisms, a comparative genomics approach looking for associations between phenotypes and genetic markers among closely related lineages can complement candidate gene approaches and functional analyses of specific genes. This multifaceted approach has been used successfully with *Phytophthora infestans* (18), *Fusarium oxysporum* f. sp. *melonis* (79), and *Verticillium dahliae* (27).

## METHODOLOGICAL STEPS

Projects in population genomics can be designed along two axes. The first axis represents the sequence of steps needed to conduct a population genomics project, as outlined in the first column in **Table 1**. The second axis represents the desired research approaches associated with a specific project. These approaches range from demographic analysis and comparative genomics of closely related species to genome scans for selection, GWASs, and QTL mapping, all of which consider variation within and among populations as outlined in the top row (left to right) in **Table 1**. The choice of steps and approaches depends inextricably on the questions being addressed and the biology of the organism being studied. Below, we outline the steps necessary for successfully conducting a population genomics project.

### Biological Questions

Population genomics investigations should be grounded in testable, biological, and evolutionary questions. The research question, together with the known biology of the organism being studied, dictates much of the experimental design. The sections above and below provide examples of the kinds of questions that population genomics can address. After defining clear, testable hypotheses and identifying the most appropriate research approach, the next important step is designing a sampling strategy.

## Sampling

An appropriate sampling strategy answers the specific question being addressed. For example, controlled crosses segregating at the $F_1$ or $F_2$ generation might be most appropriate for a QTL mapping project seeking to identify genes responsible for phenotypic differences between parents, whereas individuals randomly drawn from a sexual population—lacking genetic structure—would be appropriate for GWAS analyses seeking to determine the genetic architecture of a trait in natural populations. Similarly, hierarchical sampling of field populations might be more suitable for studying differentiation of populations by location or environment. The two common denominators in all sampling designs are that (*a*) the sample size needs to be adequate to provide sufficient statistical power for hypothesis testing, and (*b*) the target population needs to be sampled randomly to approximate the true distribution of diversity in the population to which inferences are applied. By contrast, samples composed of individuals readily available in extant culture collections generally cannot be considered a random sample and are likely to poorly represent any natural population. Some applications, in particular GWASs, may require large sample sizes drawn from genetically unstructured populations to identify genes or genomic regions associated with phenotypes.

## Genotyping

Genotyping can be done by resequencing the whole genome or by reduced-representation sequencing. These genotyping technologies currently rely mainly on short-read, deep sequencing technologies, such as the Illumina platform (San Diego, CA). For de novo whole-genome sequencing, long-read technologies like Pacific Biosciences (Menlo Park, CA) or Oxford Nanopore Technologies (Oxford, United Kingdom) might be preferable. Most of this review focuses on short-read technologies (e.g., Illumina). After sequencing and assembling a reference genome, additional individuals of the same species can be resequenced at lower read depths by aligning reads to the completed reference genome.

An alternative to whole-genome sequencing is to use reduced-representation sequencing. This method is particularly useful for pathogens with larger genomes and an abundance of repetitive sequences, such as oomycetes and powdery mildew fungi (43, 81). It is also considerably cheaper than resequencing at the present time and therefore also useful for organisms with smaller genomes. The two most common methods are restriction site–associated DNA sequencing (RADseq) (4, 25) and genotyping by sequencing (GBS) (32). Instead of sequencing whole genomes, these techniques reduce the genome representation by sequencing only from the ends of restriction fragments. Genomic DNA is digested with a restriction enzyme, and fragments are barcoded with oligonucleotide sequences unique for each individual, pooled into a library, and then sequenced on a short-read platform (26). Read depth needs to be adjusted for genome size, ploidy, and choice of restriction enzyme. The choice of restriction enzyme reduces representation to appropriate ranges, such as 500–1,000 SNPs for genome-wide population genetics inferences or >5,000 SNPs for genome scans, GWASs, or linkage mapping. The resulting reads can be mapped against a reference genome using tools such as Bowtie 2 or BWA (53, 58). In the absence of a reference genome, SNPs can be called using *Stacks* (17) or the Universal Network Enabled Analysis Kit (UNEAK) pipeline in TASSEL (62). Once reads are mapped to the reference genome, several programs can be used to call SNPs, including TASSEL (5), GATK (29), *Stacks* (17), and SAMtools (59).

An essential requirement of reduced-representation sequencing is to filter the data extensively before conducting analyses. Data need to be filtered for missing data (a common issue in RADseq and GBS data sets), sequencing error, read depth, mapping quality, copy number variation, and

minimum allele frequency (2, 3). Each of the programs for identifying SNPs (variant callers) has different settings and options for filtering data. Unfortunately, there is not yet an agreed-upon set of best practices for filtering. As a starting point, loci exhibiting copy number variation or paralogs, or those with allele frequencies <5%, should be removed from most analyses. A panel of strains that is deeply sequenced can provide a reference panel of SNP loci that can further improve data quality. This type of extensive filtering is crucial for downstream analyses.

Genotype calling is also complex and currently lacks easily reproducible, standardized processes (2, 3). In our recent experience, variant callers such as FreeBAYES, SAMtools, GATK, and TASSEL differed in the number of SNPs recovered. For example, we used GBS on a *Phytophthora rubi* population and compared the results of these four variant callers. After achieving an average read depth of 21.5×, the total number of SNPs ranged from ~48,000 to 450,000, and the percentage of heterozygous sites ranged from 4% to 25% before filtering for quality (J.F. Tabima & N.J. Grünwald, unpublished results). Thus, data can be compared reliably only by using the same filtering criteria with the same version of a given variant caller. Furthermore, the number of SNPs shared among all four callers was unexpectedly low, ranging from 5% to 16% in pairwise comparisons using identical input data. At this point, comparisons of genotypes determined by different variant callers should be avoided until common standards emerge.

Data obtained for diploid or polyploid species can be phased to obtain haplotypes. Phasing involves estimation of haplotypes at each homologous locus in a genome based on the obtained sequence data using various statistical (e.g., multinomial) and/or genetic (e.g., coalescent) models (8). These models make a range of assumptions that need to be considered, varying from statistical assumptions (e.g., random sampling) to assuming a basic coalescent process or Hardy-Weinberg equilibrium (9). In our experience, phasing is robust only for diploids and is not straightforward for polyploids.

Imputation is a step in the variant calling process whereby missing genotype data are inferred statistically from a well-characterized reference panel of genotype data (7, 60). Imputation, briefly, copies haplotype stretches from the reference panel by assuming local linkage where data are missing. These methods were developed for the human HapMap project, in which reference panels are of high quality and data are phased to determine haplotypes. Programs such as BEAGLE assume that the input reference panel is phased before imputation (6). One crucial, underlying assumption is that the sample in the reference panel and the imputation set come from the same population. Some variant callers include imputation by default, invariably resulting in some SNP calls that are not supported by a single sequence read. In our recent experience, imputation does not work well except in populations with a high quality reference panel in which phasing is known (e.g., in controlled crosses with well-characterized parental genotypes). When imputation was applied to the diploid *P. rubi* data mentioned above, homozygosity differed by factors of two to ten among variant callers. Thus, imputation causes artifacts and should be avoided for natural populations lacking reference panels.

As with traditional molecular marker research, experimental controls should be included. For example, each sequencing run should include a panel of well-characterized reference isolates for quality assurance and as positive controls for reproducibility. These isolates should be genotyped using independent DNA extractions, barcoding, sequencing, and variant calling for each experiment.

In the era of population genomics, defining clone or lineage boundaries requires new tools. In contrast to traditional marker systems such as microsatellites, obtaining SNP data through deep sequencing inevitably results in missing data, erroneous allele calls, and other errors. Given data sets with thousands of SNPs, the definition of a multilocus genotype (MLG) becomes more nuanced. Is a difference of one SNP between strains sufficient to define a new genotype? We

expect that deep sequencing will lead to inflation of genotypic diversity estimates similar to the findings of meta-barcoding studies inferring species diversity (52). An analytical tool was recently developed in the statistical and programming language R that defines genotype boundaries based on user-defined levels of genetic distance and varying clustering algorithms that can be more or less stringent in defining MLGs and multilocus lineages (MLLs) (47). This is analogous to defining operational taxonomic units (OTUs) for biological species, in which boundaries among species are defined based on a predetermined level of sequence similarity (105).

## Phenotyping

For GWASs and QTL mapping analyses that seek to connect genotype and phenotype, phenotyping has become the limiting step to making progress relative to the ease of genotyping. After investigators choose a measurable trait that is relevant to the question of interest, the challenge is to accurately phenotype hundreds or thousands of individuals in the experimental populations. The measured traits may show Mendelian or quantitative inheritance. Quantitative traits present a greater phenotyping challenge and are more likely to be affected by the environment; however, they are more common than Mendelian traits. Investigators need to decide whether trait measurements should be made in the natural environment (i.e., in the field) or in an artificial environment (i.e., in a Petri dish, growth chamber, or greenhouse). Field measurements conducted under naturally fluctuating environmental conditions are more difficult to obtain and are more prone to experimental error compared to tightly controlled laboratory measurements. However, field measurements have the advantage of being more likely to identify candidate genes that are relevant under real-world field conditions (92). In all cases, phenotype measurements should ideally be precise, accurate, and high-throughput, and minimize human-associated measurement errors. For example, automated, computer-based analyses of digital images satisfy most of these criteria for estimating aggressiveness based on leaf symptoms or growth rates on Petri dishes (57, 82, 84).

## Analytical Methods and Inferences

Population genomics analyses vary depending on the biological questions and research approaches (**Table 1**). Analyses for demographic processes and genome-wide effects in plant pathogen populations have been reviewed extensively (42, 70) and are not discussed further. Selection in a genome can be studied using several statistical approaches (reviewed in 1, 73, 91, 103), including frequency-based spectrum methods (e.g., Tajima's *D*), LD analysis, and $F_{ST}$ scans that search for allelic polymorphisms enriched in one population relative to another. Analyses for GWASs and QTL mapping have also been reviewed extensively (106), and several software packages are already available to conduct these analyses (106). All of these approaches can identify candidate loci potentially involved in selection or adaptation, but further experimental work is needed to demonstrate causality.

One of the advantages of population genomics is that sequences or SNPs are the ultimate data. However, missing data or imputed SNPs add some uncertainty. Once the data are properly checked for quality, further analysis needs to include two major objectives: (*a*) assuring that genetic assumptions for analyses such as neutrality of markers or panmixia are valid and (*b*) testing the specific hypotheses under investigation. Typically, analysis uses SNP data in the variant call format (vcf) (**https://vcftools.github.io/specs.html**). Vcf files are text files that are machine readable and store genotypic data, including indels, SNPs, and more. Several open source tools exist to read and analyze these data (**https://vcftools.github.io/specs.html**). N.J.G. has recently developed more derived tools in R for reading and analyzing population genomics data from vcf files (47, 48, 50;

**https://github.com/grunwaldlab/vcfR**) that will be useful for analysis of population genomics data in a whole-genome context.

## APPLICATIONS OF POPULATION GENOMICS IN PLANT PATHOLOGY

Pathogen population genomics will find many applications in plant pathology, especially through improving our understanding of pathogen biology. It is also likely to inform resistance breeding practices by providing insight into optimal resistance gene deployment strategies. In general, knowledge of pathogen population genomics can inform disease management in the same ways as pathogen population genetics (42, 66, 67), but the information is more extensive and enables more conclusive interpretations on pathogen population biology. Beyond applications to disease management, many novel insights will result from population genomics investigations of plant pathogens. Ultimately, population genomics will enable the identification of genes affecting important pathogen traits. Examples of the types of studies done to date are given below.

### Demographic Processes and Genome-Wide Effects

As described above, studies that genotype a large number of neutral markers can be used to address questions concerning demographic processes and genome-wide effects, as previously done in population genetics. The large number of markers makes more robust inferences possible on genetic structure, gene or genotype flow, and mating systems, as shown in the following examples.

***Fusarium graminearum.*** A population genomics analysis was conducted using 213 strains of *Fusarium graminearum* sampled from 13 wheat fields across Germany (93). RADseq was used to generate 1,129 SNPs distributed across the genome, present at frequencies between 7% and 92%, and missing in no more than four strains. AMOVA (Analysis of MOlecular VAriance) revealed that 99.7% of the total genetic variance was within field populations and only 0.3% was among populations. An earlier analysis using the same isolates but based on 19 SSR (simple sequence repeat) loci found that 20% of the genetic variance was distributed among field populations. This comparison reveals how different types and densities of genetic markers can produce different outcomes. LD decayed within 1,000 bp on average across all four chromosomes, indicating that this population is highly recombined and undergoes regular recombination, despite the fact that *F. graminearum* is homothallic. Analysis of SNP markers at a fine scale revealed over 200 hot spots of recombination. These hot spots were fewer than ~650 bp in length and exhibited up to 0.2 recombinations/(bp · generation). They were also enriched in genes encoding DNA repair and membrane transport. The authors concluded that the German field populations of *F. graminearum* represent a single panmictic population with sufficient genetic diversity in each field population to enable rapid adaptation to changes in the local environment, including deployment of resistant cultivars or applications of fungicides. Another outcome of this analysis was to illustrate that this collection of isolates would be useful for a GWAS, as described later.

***Verticillium dahliae.*** Genotyping with a large number of SNPs in *Verticillium dahliae* illustrates two types of studies that were not previously possible. In the first study, genotyping by GBS produced more than 26,000 SNPs in a diverse collection of 141 isolates (71). SNPs were evenly distributed on all contigs of the reference genome and therefore ideally suited for population genomics analysis. *V. dahliae* has no known sexual stage, and SNP genotyping confirmed that populations have a highly clonal structure. Almost all SNP diversity was due to differences among clonal lineages; very few SNPs were found within lineages. Most genotypes clearly belonged to

well-defined lineages that correlate with vegetative compatibility groups (VCGs). However, several isolates and one newly discovered lineage had genotypes that were recombinant between known lineages. More extensive analysis indicated that nearly all clonal lineages arose by recombination. Although *V. dahliae* is ostensibly asexual, genes required for meiosis were found in the reference genome (71, 80), suggesting that recombination is most likely to have occurred by sexual reproduction. As a follow up to this study, additional SNP genotyping was conducted to infer the origin and migration of a highly virulent, defoliating lineage of *V. dahliae* (72). Coalescent analyses of these genotypes made it possible to infer five separate introductions of this lineage from North America into the Mediterranean basin. Because of the fine-grained genotyping afforded by GBS, enough variation was detected within this lineage to infer the ancestral haplotype as coming from North America, which is consistent with historical records.

## Comparative Population Genomics

Comparative population genomics provides insight into pathogen origins and the underlying processes driving pathogen speciation and emergence of novel clones or lineages (87), as well as the roles of migration (39, 64, 65), host tracking (88), host jumps (e.g., *Pyricularia graminis-tritici* causing wheat blast in Brazil) (16), hybridization (69, 89), selection (76), HGT (27, 36), and chromosome rearrangements (22, 28) in the emergence of new pathogens. Ultimately, these analyses may provide insight into the risk of emergence of novel pathogens in particular agroecosystems.

The most powerful comparative population genomics studies include multiple strains of each species because differences observed between species are more meaningful with accurate measurements of differences within species. A useful approach is to include wild pathogen populations sampled from ancestral or undomesticated hosts growing at or near the center of origin of the domesticated pathogen. This enables inference of the ancestral state and provides a measure of the degree of divergence. If possible, collections should be made in sympatry, with wild pathogen populations sampled at the same time, location, and spatial scale as the domesticated pathogen populations. Sympatric collections of wild and domesticated pathogen species enable strong inferences to be made regarding the underlying processes that led to pathogen domestication, including host specialization and adaptation to the corresponding host agroecosystem.

***Phytophthora infestans* and related species.** The potato late blight pathogen *P. infestans* and its close clade 1c relatives *Phytophthora ipomoeae* and *Phytophthora mirabilis* are known to occur at the putative center of origin in central Mexico (40, 41). Raffaele et al. (76) sequenced the genomes of these three taxa and found evidence for compartmentalized genomes, with repeat-rich regions of the genome showing higher evolutionary rates and transposon activity compared to more conserved regions that are enriched for housekeeping genes. The repeat-rich regions contain most of the pathogenicity genes, such as RxLR and Crinkler genes, that might be involved in host jumping (43, 98). Other work provided evidence that *Phytophthora andina* is a hybrid between *P. infestans* and another as yet unknown *Phytophthora* hybrid parent presumably of South American origin (38). Hybridization has been described for several *Phytophthora* species, including, for example, *Phytophthora alni* (46). Because of the hybrid nature of these genomes, it has not been possible to assemble genome sequences from short-read sequencing (i.e., Illumina). Long-read sequencing (e.g., Pacific Biosciences) will be necessary to solve this problem.

***Pyricularia* species on wheat and wild grasses.** The wheat blast disease caused by *P. graminis-tritici* and other *Pyricularia* species (16) was first reported in Brazil 30 years ago and has since spread across the wheat-growing region of Brazil and into neighboring countries (reviewed in 15).

It was long thought that the wheat blast disease emerged through a host jump from *Pyricularia* strains infecting rice, although recent population genetics analyses based on 11 SSR loci indicated that the rice- and wheat-infecting populations were genetically distinct (63). To further test this hypothesis and identify alternative original hosts for the wheat blast pathogen, *Pyricularia* strains were collected from wheat and rice fields across Brazil as well as from weeds and pasture grasses growing near those fields. Comparative population genomics enabled differentiation of many distinct lineages within the *Pyricularia* species complex and showed that the closest relatives of the wheat pathogen were found on the widely grown pasture grass *Urochloa* as well as other grasses growing in or near wheat fields (16, 19). By contrast, the *Pyricularia* population infecting rice was distantly related. These findings, coupled with analyses of population genetics based on SSRs and sequences of housekeeping genes in more than 500 strains (16, 63), indicated that the wheat blast pathogen likely emerged through a host jump from the *Pyricularia* population infecting *Urochloa* or other Brazilian grasses ~30 years ago. An important application of these population genomics comparisons is that they enabled a very rapid determination of the source of the wheat blast outbreak occurring in Bangladesh in 2016 (13). The genome sequences of the Bangladesh strains clustered tightly with the *P. graminis-tritici* genome sequences from Brazil, indicating that the Bangladesh outbreak most likely originated from imported Brazilian wheat (**https://github.com/crolllab/wheat-blast**).

***Blumeria graminis* on cereals.** Powdery mildew on triticale (*Blumeria graminis* f. sp. *triticale*) is another recently emerged pathogen that was first noticed ~15 years ago in Europe (102). Triticale became widely grown in Europe beginning in the 1960s. Comparative genomics indicated that the new, host-specialized triticale pathogen emerged following hybridization between *B. graminis* species specialized to infect either bread wheat (*B. graminis* f. sp. *tritici*) or rye (*B. graminis* f. sp. *secalis*), the two ancestors of triticale. Resequencing 40 *Blumeria* isolates from all three hosts showed that triticale powdery mildew likely emerged through two independent hybridization events that occurred in Europe between 7 and 47 years ago (69). Additional analyses that included six *B. graminis* f. sp. *dicocci* strains from tetraploid durum wheat traced the origin of wheat powdery mildew back to the domestication of wheat ~11,000 years ago, providing another example of pathogen emergence through host tracking.

## Identifying Genes Under Selection

The candidate gene approach to finding genes under selection has been best exemplified by studies on effectors. After candidate effector genes have been identified and validated, detailed analyses of effector profiles (effectoromics) can be conducted at the population level to guide resistance breeding strategies, as recently described for global populations of *Parastagonospora nodorum* (68), regional populations of *Leptosphaeria maculans* in Australia (99), and local populations of *P. infestans* (101). This practice is the modern equivalent of the race typing (pathotyping) conducted for many fungal and oomycete pathogens by earlier generations of plant pathology researchers. Population genomics analyses enable measurements of effector diversity across many loci simultaneously and can lead to the discovery of new effectors that may become relevant when breeders deploy new major resistance genes. Similarly, surveys of diversity in genes encoding proteins targeted by fungicides may be used in the future to predict the emergence of fungicide resistance in local pathogen populations and to formulate resistance management strategies.

The candidate gene approach, however, is limited to genes with already known functions. By contrast, genome scans for genes under selection may lead to the identification of novel genes not previously known to be associated with any phenotype.

***Zymoseptoria* species on wheat and wild grasses.** Genome scans in the wheat pathogen *Zymoseptoria tritici* illustrate the potential of this approach to identify genes previously unknown to be involved in host specialization. Iranian scientists sampled infected leaves from wild grasses growing near Iranian wheat fields to identify sympatric, wild *Zymoseptoria* species (86). Although the wild and domesticated species were sampled in sympatry, there was no evidence for recent movement of individuals between wild and domesticated hosts, and the wild pathogens appeared to be a different species from the wheat pathogen (91). Cross-inoculation experiments showed that *Z. tritici* had become specialized to infect wheat, unlike the wild *Zymoseptoria* species that infected multiple host species (88). Genome sequences were compared for ten isolates of *Z. tritici*, five isolates of *Zymoseptoria pseudotritici*, four isolates of *Zymoseptoria ardabiliae*, and one isolate of *Zymoseptoria passerinii* (a previously identified barley-specialized pathogen used as an outgroup). Sequencing several strains of each species enabled a comparison of diversity within and between species. These analyses identified 27 genes that diverged significantly more between species than the rest of the genes in the genome, a pattern expected for genes involved in adaptation (host specialization) and speciation (88). A coalescent analysis indicated that the domesticated wheat pathogen diverged from the closest wild pathogen species, *Z. pseudotritici*, ~11,000 years ago (86), around the same time that wheat was domesticated (78), supporting the hypothesis that *Z. tritici* emerged through host tracking (91). Four of the twenty-seven candidate genes for host specialization were functionally analyzed using allele swaps, and three of them were shown to have significant effects on the ability to infect and reproduce on wheat (75), suggesting that these highly diverged genes were involved in the emergence of host specialization toward wheat. Using a complementary candidate gene approach, Brunner et al. (11) conducted a targeted analysis of the 48 plant cell wall–degrading enzymes (based on the CAZy classification system) (14) found in these strains and identified three genes (two encoding cellulases and one encoding a cutinase) involved in host specialization, with three additional genes encoding a cutinase, a cellulase, and a hemicellulase showing an evolutionary pattern consistent with evading recognition by host receptors.

## Genome-Wide Association Studies

The overall objective of a GWAS is to establish a firm link between genotype and phenotype in natural populations. This analysis may eventually lead to identification of genes or genetic markers associated with important traits, such as host specialization, virulence (aggressiveness), fungicide sensitivity, thermal adaptation, pathogen reproduction, and pathogen growth rate. As explained earlier, an important limitation in a GWAS is to analyze individuals sampled from single, genetically undifferentiated populations. Otherwise, a GWAS needs to be statistically corrected for population structure, leading to a significant reduction in the ability to detect significant associations between phenotypes and genetic markers.

***Fusarium graminearum.*** A GWAS was conducted using the same 220 strains of *F. graminearum* described in the section on Demographic Processes and Genome-Wide Effects above (93). RAD-seq was used to generate ~29,000 SNP markers distributed across the genome. The traits aggressiveness and deoxynivalenol (DON) production were measured for 119 strains across four natural field environments, and sensitivity to the azole fungicide propiconazole was measured for all 220 strains using $EC_{50}$ measurements in a laboratory environment. A total of 50, 29, and 74 QTNs were significantly associated with aggressiveness, DON production, and propiconazole sensitivity, respectively, with three QTNs exceeding the Bonferroni threshold of significance located in genes not previously associated with azole sensitivity (92). Only ~40% of the detected QTNs caused nonsynonymous substitutions, representing candidate SNPs affecting gene function. The majority of QTNs were synonymous substitutions or located in intergenic regions. These QTNs

may represent regulatory elements, misannotated genes, or SNPs that are in LD with nearby genes or regulatory sequences that are responsible for the trait variance. A highly significant QTN explaining 24% of the variance in aggressiveness was found in a gene encoding a RAS-GTPase activating protein, implicating the RAS protein as an important contributor to aggressiveness in *F. graminearum*. Seventeen percent of the genetic variance in azole sensitivity was explained by each of the three QTNs that passed the Bonferroni significance threshold. One of the QTNs was in a gene encoding an acyl-CoA thioesterase, a protein that plays a role in converting activated fatty acids into their corresponding nonesterified fatty acids. This indicates that acyl-CoA thioesterase may play an important role in azole sensitivity, possibly through reprogramming the regulation of lipid metabolism in fungal cells exposed to azoles. The other two QTNs were located in genes representing completely novel contributors to fungicide resistance. These genes provide good candidates for future studies on the mechanism of azole resistance in field populations of pathogenic fungi.

***Heterobasidion annosum.*** Dalman et al. (24) sequenced 23 haploid isolates of the basidiomycete tree pathogen *Heterobasidion annosum* sampled from across Europe, with a mean read depth of 2.6–12.6× for the individual isolates, resulting in ~33,000 SNPs to include in a GWAS. Virulence for each strain was measured as both lesion length formed under the bark and fungal growth in the sapwood on Scots pine and Norway spruce. Twelve SNPs distributed across seven assembled contigs were significantly associated with at least one of the four measured traits. Six of the twelve SNPs were distributed among three tightly linked genes on one contig. One of these genes was predicted to encode calcineurin, a phosphatase regulated by $Ca^{2+}$ and calmodulin that is involved in calcium-dependent signal transduction pathways. Calcineurin was shown to affect virulence in several other fungi and this GWAS finding led the authors to postulate that natural variation in this gene affects the ability of *H. annosum* to grow in pine trees. Five of the eight candidate virulence genes identified by GWASs were found within QTLs identified in an earlier linkage mapping study (61). This example illustrates well how GWASs can identify novel candidate genes for important traits and lead to new insights into the biology of nonmodel pathogens.

## Quantitative Trait Locus Mapping Using Controlled Crosses

QTL mapping is often done with the same goals as GWASs, i.e., to identify genes or genomic regions associated with specific phenotypes. Relatively few QTL mapping studies have been done with plant-pathogenic fungi (23, 49, 61), perhaps because making controlled, segregating crosses is often difficult or impossible, and the only one we are aware of that uses population genomics technologies is described here in *Z. tritici*.

***Zymoseptoria tritici.*** Recent work with *Z. tritici* combined precise phenotype data generated using high-throughput digital-image analyses with dense genetic marker sets based on RADseq to determine the genetic architecture of several quantitative traits and identify candidate genes responsible for these traits. Two crosses were made among four Swiss strains of *Z. tritici*, resulting in ~350 offspring for each cross. Complete genome sequences for all four parents were assembled by combining Illumina and Pacific Biosciences sequencing technologies, enabling identification of the specific point mutations in each candidate gene that could be responsible for the observed trait variance. Approximately 8,500 SNPs were segregating in each cross to create genetic linkage maps and conduct QTL analyses. Several significant QTLs were identified for each trait, including degree of melanization, fungicide and temperature sensitivity, degree of hyphal growth, and a suite of virulence traits, including pycnidial density, pycnidial size, and percentage of leaf area covered by lesions. Several of the QTLs were associated with large LOD (logarithm of the odds) scores,

explained a large percentage of the overall trait variance, and had fewer than 10 genes within their 95% confidence intervals. By comparing the parental sequences of the genes in these QTLs and focusing on nonsynonymous substitutions that are likely to affect protein function, high priority candidate genes were identified for melanization (*PKS1*), fungicide sensitivity (*CYP51*), thermal adaptation (*MgPBS2*), and hyphal growth (a guanine nucleotide exchange factor) (55–57). In addition, it was possible to separate genes involved in the necrosis phenotype, reflecting host damage, from genes involved in forming pycnidia, reflecting pathogen reproduction (83). Candidate genes involved in host specialization were also identified (83). It is notable that all of these QTLs are based on natural genetic variation existing in the parents; hence, they represent the type of genetic variation on which selection can operate in natural field populations.

## SUMMARY POINTS

1. Population genomics is an extension of population genetics, but the ability to identify large numbers of SNPs distributed throughout the genome radically alters the types of questions that can be addressed.
2. The availability of fine-resolution genotyping has enabled a shift in emphasis toward understanding locus-specific effects, such as discovering the genetic mechanisms underlying pathogenicity, virulence, fungicide resistance, and host specialization.
3. Three main types of questions are addressed in population genomics related to understanding the genome-wide demographic process (as in traditional population genetics), identifying genes under selection and involved in adaptation to different environments, and determining the genetic architecture underlying known phenotypes.
4. Research approaches in population genomics include: genetic analyses of natural populations, comparative genomic analyses of closely related species, identification of genes under selection, GWASs in natural populations, and QTL mapping in segregating populations resulting from crosses.
5. This rapidly developing field has not yet established standardized best practices for either genotyping methods or analyses, and therefore great care is needed to avoid artifacts that lead to inappropriate conclusions.
6. Pathogen population genomics will find many applications in plant pathology, especially through improving our understanding of pathogen biology, but also in contributing to disease management and deployment of disease resistance.

## FUTURE ISSUES

1. Additional advances in genomic technology are likely to continue driving down the costs for genome sequencing. What will be the effects on plant pathology in 10 years when a fungal or oomycete genome sequence can be routinely obtained at a cost affordable by any lab? We expect that data analysis and interpretation will remain limiting factors for progress, even though automated analysis software will be available for addressing some specific questions and applications. However, major improvements in throughput, reproducibility, and accuracy of phenotyping will be needed to firmly establish the connections between genotype and phenotype, especially for quantitative traits.

2. As families of effector genes encoding host-specific toxins and avirulence are identified and validated in field experiments for their role in aggressiveness and virulence, molecular phenotyping is likely to replace traditional pathotype analysis based on measuring strain virulence on sets of differential hosts in the greenhouse. Similarly, traditional measures of fungicide sensitivity such as $EC_{50}$ may be replaced by molecular phenotyping for pathogens with well-established mutations that can be associated with fungicide resistance in the field.
3. In the field of human medicine, molecular epidemiology is already applied for important pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) (74), HIV-1 (45), and the Ebola virus (37), providing examples of what can be achieved when sufficient resources are brought to bear on an urgent disease outbreak. Similar applications will become commonplace in plant pathology, as illustrated recently for the wheat blast outbreak in Bangladesh (13, **https://github.com/crolllab/wheat-blast**). After sequencing an appropriate number of pathogen strains in key locations to establish the baseline data sets needed to minimize ascertainment bias, molecular epidemiology—using the tools of population genomics—can be applied to identify new emerging clones, races, or pathotypes that exhibit novel properties (e.g., changes in virulence, fungicide sensitivity, or thermal adaptation) in real time to identify the source of the new pathogen population and its likely pathway of migration.
4. Connected to these developments, we also foresee the implementation of virtual collections of pathogen isolates based on archived genome sequences stored in databases on computer servers that can be accessed by plant pathologists worldwide, regardless of quarantine restrictions on the movement of pathogens across borders. These virtual collections would complement existing collections of living pathogen specimens that occupy physical space and require regular and expensive upkeep. As costs for sequencing and data storage continue to fall, we can imagine that the number of pathogen strains maintained in virtual collections will quickly exceed the number of strains maintained in physical collections, although the traditional, physical culture collection will remain an indispensable biological resource far into the future.
5. Regardless of whether any of these predictions come to pass, it is clear that research in the field of pathogen population genomics will make major contributions to plant pathology for the foreseeable future.

## 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

B.A.M. acknowledges receipt of a fellowship from the OECD Co-operative Research Program: Biological Resource Management for Sustainable Agricultural Systems in 2015 as well as research funding from the Swiss National Science Foundation (31003A_134755 and 31003A_155955). N.J.G. was partially supported by funds from USDA-ARS CRIS Project 2027-22000-039-00 and USDA National Institute of Food and Agriculture Grant 2011-68004-30154. M.G.M. was

partially supported by Grant AGL2011-24935 (Vertigen project) from Ministerio de Economía y Competitividad, Spain, and co-financed via the European Regional Development Fund of the European Union. 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. Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect. Genet. Evol.* 9(4):656–70
2. **Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17(2):81–92**
3. Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. *Mol. Ecol.* 23(7):1661–67
4. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* 3(10):e3376
5. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–35
6. Browning BL, Browning SR. 2016. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98(1):116–26
7. Browning SR. 2008. Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum. Genet.* 124(5):439–50
8. Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81(5):1084–97
9. Browning SR, Browning BL. 2011. Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* 12(10):703–14
10. Brunner PC, Stefanato FL, McDonald BA. 2008. Evolution of the *CYP51* gene in *Mycosphaerella graminicola*: evidence for intragenic recombination and selective replacement. *Mol. Plant Pathol.* 9(3):305–16
11. Brunner PC, Torriani SFF, Croll D, Stukenbrock EH, McDonald BA. 2013. Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. *Mol. Biol. Evol.* 30(6):1337–47
12. Brunner PC, Torriani SFF, Croll D, Stukenbrock EH, McDonald BA. 2014. Hitchhiking selection is driving intron gain in a pathogenic fungus. *Mol. Biol. Evol.* 31(7):1741–49
13. Callaway E. 2016. Devastating wheat fungus appears in Asia for first time. *Nature* 532:421–22
14. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37:D233–38
15. Castroagudín VL, Ceresini PC, de Oliveira SC, Reges JTA, Maciel JLN, et al. 2015. Resistance to QoI fungicides is widespread in Brazilian populations of the wheat blast pathogen *Magnaporthe oryzae*. *Phytopathology* 105(3):284–94
16. Castroagudín VL, Moreira SI, Pereira DAS, Moreira SS, Brunner PC, et al. 2016. *Pyricularia graminis-tritici* sp. nov., a new *Pyricularia* species causing wheat blast. *Persoonia*. In press
17. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH, De Koning D-J. 2011. *Stacks*: building and genotyping loci de novo from short-read sequences. *G3* 1(3):171–82
18. Cooke DEL, Cano LM, Raffaele S, Bain RA, Cooke LR, et al. 2012. Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLOS Pathog.* 8(10):e1002940
19. Croll D, Ceresini P, Maciel J, McDonald BA. 2016. The origin of wheat blast in Bangladesh. **https://github.com/crolllab/wheat-blast**
20. **Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics* 201(3):1213–28**

**2. Discusses the challenges of SNP genotyping and analysis from reduced-representation sequencing.**

**20. RADseq markers enabled identification of rapidly evolving recombination hot spots in two crosses of *Zymoseptoria tritici*.**

21. Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLOS Pathog.* 8:e1002608
22. Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLOS Genet.* 9(6):e1003567
23. Cumagun CJR, Bowden RL, Jurgenson JE, Leslie JF, Miedaner T. 2004. Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. *Phytopathology* 94(5):520–26
24. Dalman K, Himmelstrand K, Olson Å, Lind M, Brandström-Durling M, Stenlid J. 2013. A genome-wide association study identifies genomic regions for virulence in the non-model organism *Heterobasidion annosum* s.s. *PLOS ONE* 8:e53525
25. Davey JW, Blaxter ML. 2010. RADseq: next-generation population genetics. *Brief. Funct. Genom.* 9(5–6):416–23
**26. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12(7):499–510**
27. de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *PNAS* 109(13):5110–15
28. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* 23(8):1271–82
29. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43(5):491–98
**30. Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29(1):51–63**
31. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, et al. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *PNAS* 108(7):2831–36
32. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* 6(5):e19379
33. Famoso AN, Zhao K, Clark RT, Tung C-W, Wright MH, et al. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLOS Genet.* 7(8):e1002221
34. Frenkel O, Cadle-Davidson L, Wilcox WF, Milgroom MG. 2015. Mechanisms of resistance to an azole fungicide in the grapevine powdery mildew fungus, *Erysiphe necator*. *Phytopathology* 105(3):370–77
35. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, et al. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* 38(8):953–56
36. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, et al. 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLOS Pathog.* 8(9):e1002952
37. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, et al. 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345(6202):1369–72
38. Goss EM, Cardenas ME, Myers K, Forbes GA, Fry WE, et al. 2011. The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. *PLOS ONE* 6(9):e24543
39. 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(9):e1000583
40. Goss EM, Tabima JF, Cooke DEL, Restrepo S, Fry WE, et al. 2014. The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *PNAS* 111(24):8791–96
41. Grünwald NJ, Flier WG. 2005. The biology of *Phytophthora infestans* at its center of origin. *Annu. Rev. Phytopathol.* 43:171–90
42. Grünwald NJ, Goss EM. 2011. Evolution and population genetics of exotic and re-emerging pathogens: novel tools and approaches. *Annu. Rev. Phytopathol.* 49:249–67
43. Haas BJ, Kamoun S, O'Neill MC, Jiang RHY, Handsaker RE, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461(7262):393–98

**26. Overview of methods for SNP genotyping by RADseq and GBS.**

**30. Excellent summary of the opportunities and challenges in population genomics for nonmodel organisms.**

44. Hartl DL, Clark AG. 2007. *Principles of Population Genetics*. Sunderland, MA: Sinauer Assoc. 652 pp. 4th ed.
45. Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2011. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 25(5):679–89
46. Ioos R, Andrieux A, Marçais B, Frey P. 2006. Genetic characterization of the natural hybrid species *Phytophthora alni* as inferred from nuclear and mitochondrial DNA analyses. *Fungal Genet. Biol.* 43(7):511–29
**47. Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208**
48. Kamvar ZN, Tabima JF, Grünwald NJ. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281
49. Kema GHJ, Goodwin SB, Hamza S, Verstappen ECP, Cavaletto JR, et al. 2002. A combined amplified fragment length polymorphism and randomly amplified polymorphism DNA genetic linkage map of *Mycosphaerella graminicola*, the Septoria tritici leaf blotch pathogen of wheat. *Genetics* 161(4):1497–505
50. Knaus BJ, Grünwald NJ 2016. VcfR: an R package to manipulate and visualize VCF format data. *Mol. Ecol. Resour*. In press
**51. Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9(1):29**
52. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12(1):118–23
53. Langmead B, Salzberg SL. 2012. Fast gapped–read alignment with Bowtie 2. *Nature Methods* 9(4):357–59
54. Leinonen T, McCairns RJS, O'Hara RB, Merilä J. 2013. $Q_{ST}$-$F_{ST}$ comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* 14(3):179–90
55. Lendenmann MH, Croll D, McDonald BA. 2015. QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genet. Biol.* 80:53–67
56. Lendenmann MH, Croll D, Palma-Guerrero J, Stewart EL, McDonald BA. 2016. QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity* 116(4):384–94
**57. Lendenmann MH, Croll D, Stewart EL, McDonald BA. 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. *G3* 4(12):2519–33**
58. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–60
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–79
60. Li Y, Willer C, Sanna S, Abecasis G. 2009. Genotype imputation. *Annu. Rev. Genom. Hum. Genet.* 10:387–406
61. Lind M, Dalman K, Stenlid J, Karlsson B, Olson Å. 2007. Identification of quantitative trait loci affecting virulence in the basidiomycete *Heterobasidion annosum* s.l. *Curr. Genet.* 52(1):35–44
62. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLOS Genet.* 9(1):e1003215
63. Maciel JLN, Ceresini PC, Castroagudin VL, Zala M, Kema GHJ, McDonald BA. 2014. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. *Phytopathology* 104(1):95–107
64. Martin MD, Ho SYW, Wales N, Ristaino JB, Gilbert MTP. 2014. Persistence of the mitochondrial lineage responsible for the Irish potato famine in extant New World *Phytophthora infestans*. *Mol. Biol. Evol.* 31(6):1414–20
65. Martin MD, Vieira FG, Ho SYW, Wales N, Schubert M, et al. 2016. Genomic characterization of a South American *Phytophthora* hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. *Mol. Biol. Evol.* 33(2):478–91
66. McDonald BA. 2015. How can research on pathogen population biology suggest disease management strategies? The example of barley scald (*Rhynchosporium commune*). *Plant Pathol.* 64(5):1005–13
67. McDonald BA, Mundt CC. 2016. How knowledge of pathogen population biology informs management of Septoria tritici blotch. *Phytopathology* **http://dx.doi.org/10.1094/PHYTO-03-16-0131-RVW**

**47. Significant extensions to the R package *poppr* for analyzing large, genome-wide SNP data in clonal organisms.**

**51. Excellent discussion of the biological and statistical considerations required for conducting GWASs.**

**57. QTL analysis combining RADseq SNP markers with automated image analysis identified a candidate QTN affecting melanization in the *PKS1* gene.**

68. McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA. 2013. Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytol.* 199(1):241–51
69. Menardo F, Praz CR, Wyder S, Ben-David R, Bourras S, et al. 2016. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nat. Genet.* 48(2):201–5
**70. Milgroom MG. 2015. *Population Biology of Plant Pathogens: Genetics, Ecology, and Evolution*. St. Paul, MN: APS Press**
71. Milgroom MG, Jiménez-Gasco MM, Olivares García C, Drott MT, Jiménez-Díaz RM. 2014. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. *PLOS ONE* 9(9):e106740
72. Milgroom MG, Jiménez-Gasco MM, Olivares-García C, Jiménez-Díaz RM. 2016. Clonal expansion and migration of a highly virulent, defoliating lineage of *Verticillium dahliae*. *Phytopathology*. **http://dx.doi.org/10.1094/PHYTO-11-15-0300-R**
73. Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218
74. Otter JA, French GL. 2010. Molecular epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in Europe. *Lancet Infect. Dis.* 10(4):227–39
75. Poppe S, Dorsheimer L, Happel P, Stukenbrock EH. 2015. Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLOS Pathog.* 11(7):e1005055
76. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, et al. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330(6010):1540–43
77. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr. Biol.* 16(18):1857–64
78. Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the Near East. *Nat. Rev. Genet.* 3(6):429–41
79. Schmidt SM, Lukasiewicz J, Farrer R, van Dam P, Bertoldo C, Rep M. 2016. Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *fom-2* resistance gene in melon. *New Phytol.* 209(1):307–18
80. Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao KV. 2014. Maintenance of sex-related genes and the co-occurrence of both mating types in *Verticillium dahliae*. *PLOS ONE* 9(11):e112145
81. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330(6010):1543–46
82. Stefansson TS, Willi Y, Croll D, McDonald BA. 2014. An assay for quantitative virulence in *Rhynchosporium commune* reveals an association between effector genotype and virulence. *Plant Pathol.* 63(2):405–14
83. Stewart EL, Croll D, Lendenmann MH, Sanchez-Vallet A, Hartmann FE, et al. 2016. QTL mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. bioRxiv. **http://dx.doi.org/10.1101/051169**
84. Stewart EL, McDonald BA. 2014. Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology* 104(9):985–92
85. Stinchcombe J, Hoekstra H. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100(2):158–70
86. 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(2):398–411
87. Stukenbrock EH, Bataillon T. 2012. A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. *PLOS Pathog.* 8(9):e1002893
**88. Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, et al. 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Res.* 21(12):2157–66**
89. Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *PNAS* 109(27):10954–59
90. Stukenbrock EH, Croll D. 2014. The evolving fungal genome. *Fungal Biol. Rev.* 28(1):1–12

**70. A comprehensive introduction to population genetics of fungi and oomycetes.**

**88. Comparative population genomics analysis illustrating several analytical methods to identify candidate genes involved in host specialization and speciation.**

91. Stukenbrock EH, McDonald BA. 2009. Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Mol. Plant-Microbe Interact.* 22:371–80
92. Talas F, Kalih R, Miedaner T, McDonald BA. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol. Plant-Microbe Interact.* 29:417–30
93. Talas F, McDonald BA. 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. *BMC Genom.* 16:996
94. Torriani SFF, Brunner PC, McDonald BA. 2011. Evolutionary history of the mitochondrial genome in *Mycosphaerella* populations infecting bread wheat, durum wheat and wild grasses. *Mol. Phylogenet. Evol.* 58(2):192–97
95. Torriani SF, Brunner PC, McDonald BA, Sierotzki H. 2009. QoI resistance emerged independently at least 4 times in European populations of *Mycosphaerella graminicola*. *Pest Manag. Sci.* 65(2):155–62
96. Torriani SFF, Goodwin SB, Kema GHJ, Pangilinan JL, McDonald BA. 2008. Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genet. Biol.* 45(5):628–37
97. Torriani SFF, Penselin D, Knogge W, Felder M, Taudien S, et al. 2014. Comparative analysis of mitochondrial genomes from closely related *Rhynchosporium* species reveals extensive intron invasion. *Fungal Genet. Biol.* 62:34–42
98. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313(5791):1261–66
99. Van de Wouw AP, Marcroft SJ, Ware A, Lindbeck K, Khangura R, Howlett BJ. 2014. Breakdown of resistance to the fungal disease, blackleg, is averted in commercial canola (*Brassica napus*) crops in Australia. *Field Crops Res.* 166:144–51
100. Vinatzer BA, Monteil CL, Clarke CR. 2014. Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annu. Rev. Phytopathol.* 52:19–43
101. Vleeshouwers VGAA, Raffaele S, Vossen JH, Champouret N, Oliva R, et al. 2011. Understanding and exploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathol.* 49:507–31
102. Walker AS, Bouguennec A, Confais J, Morgant G, Leroux P. 2011. Evidence of host-range expansion from new powdery mildew (*Blumeria graminis*) infections of triticale (× *Triticosecale*) in France. *Plant Pathol.* 60(2):207–20
**103. Weigel D, Nordborg M. 2015. Population genomics for understanding adaptation in wild plant species. *Annu. Rev. Genet.* 49:315–38**
104. Wolfe MS, McDermott JM. 1994. Population genetics of plant pathogen interactions: the example of the *Erysiphe graminis–Hordeum vulgare* pathosystem. *Annu. Rev. Phytopathol.* 32:89–113
105. Wooley JC, Godzik A, Friedberg I. 2010. A primer on metagenomics. *PLOS Comput. Biol.* 6(2):e1000667
106. Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *Plant Genome* 1(1):5

**103. An excellent introduction to population genomics oriented mainly around plants, particularly *Arabidopsis*.**