Harnessing Population Genomics to Understand How Bacterial Pathogens Emerge, Adapt to Crop Hosts, and Disseminate

Boris A. Vinatzer,[1,*] Caroline L. Monteil,[1,2] and Christopher R. Clarke[1]

[1]Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, Virginia 24061; email: vinatzer@vt.edu, gtg681r@vt.edu

[2]INRA, UR0407 Pathologie Végétale, 84143 Montfavet Cedex, France; email: caroline.monteil@avignon.inra.fr

Annu. Rev. Phytopathol. 2014. 52:19–43

First published online as a Review in Advance on April 28, 2014

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

This article's doi: 10.1146/annurev-phyto-102313-045907

*Corresponding author.

Keywords

population genomics, next-generation sequencing, environmental reservoirs, disease emergence, phylogeography, plant immunity, disease diagnostics

Abstract

Crop diseases emerge without warning. In many cases, diseases cross borders, or even oceans, before plant pathologists have time to identify and characterize the causative agents. Genome sequencing, in combination with intensive sampling of pathogen populations and application of population genetic tools, is now providing the means to unravel how bacterial crop pathogens emerge from environmental reservoirs, how they evolve and adapt to crops, and what international and intercontinental routes they follow during dissemination. Here, we introduce the field of population genomics and review the population genomics research of bacterial plant pathogens over the past 10 years. We highlight the potential of population genomics for investigating plant pathogens, using examples of population genomics studies of human pathogens. We also describe the complementary nature of the fields of population genomics and molecular plant-microbe interactions and propose how to translate new insights into improved disease prevention and control.

INTRODUCTION

The goals of phytobacteriology are to describe the bacterial agents that cause plant diseases, identify the sources/reservoirs of these pathogens, unravel the modalities and routes of dissemination, study the mechanisms of infection, and, finally, develop efficient methods based on the acquired knowledge to prevent and control bacterial plant diseases.

The technology available to plant pathologists and clinical microbiologists to identify and characterize bacterial pathogens has changed dramatically over the years, transitioning from phenotypic characterization to molecular analysis. Until recently, molecular analysis was limited to comparison of patterns of DNA sequence variation in a very small subset of a bacterial genome. Consequently, the resolution at which bacteria could be identified and compared was very low. Only during the past 10 years has innovation in DNA sequencing technology started to allow us to extend the analysis of bacterial pathogens from a few genes to entire genomes and from a single pathogen isolate to representative samples of entire pathogen populations. This innovation is revolutionizing all aspects of clinical microbiology and phytobacteriology. Here, we focus on how DNA sequencing of samples that represent pathogen populations is changing our ability to investigate bacterial plant pathogen emergence, crop adaptation, and dissemination, and we share our vision of how DNA sequencing could transform plant disease prevention and control in the future.

BACTERIAL PLANT PATHOGEN POPULATIONS

What Constitutes a Bacterial Plant Pathogen Population?

In population genetics, a population consists of those members of a species that can interbreed because they are located in the same geographic area. Given that bacteria do not sexually reproduce like plants and animals, the definition cannot strictly be applied to bacteria, and phytobacteriologists use the term "pathogen population" quite loosely. It is sometimes even used to refer to all members of a species or an entire complex of related species.

However, since John Maynard Smith's landmark paper "How Clonal Are Bacteria?" (122), it has become clear that bacteria do not only reproduce clonally by passing DNA from one generation to the next and that mutations are not the only source of genetic variation. To the contrary, bacteria exchange DNA among each other by a process called homologous recombination, whereby bacteria acquire DNA from other bacteria by transformation, conjugation, or transduction and then replace a region of their own chromosome with the acquired DNA (39). This mechanism is distinct from site-directed recombination, a mechanism by which a recipient bacterium acquires new genes without replacement of its own genes through insertion of gene clusters at specific sites by bacteriophages or various other mobile genetic elements (67). The more general term horizontal gene transfer (HGT) refers to the transfer of genes by either mechanism and also includes the transfer of plasmids that do not integrate into the main bacterial chromosome.

Although acquisition of genes by site-directed recombination even occurs between distantly related genera, homologous recombination occurs most frequently between closely related bacteria because donor and recipient DNA need to be very similar to each other to allow efficient recombination. Therefore, similar to sexually reproducing organisms, bacterial populations can be delineated on the basis of the extent of DNA exchange, whereby members of the same population are characterized by extensive homologous recombination and members of different populations are characterized by very limited homologous recombination. It has even been proposed that bacterial

species could be defined on the basis of the extent of homologous recombination (51) and that, as in sexually reproducing organisms, ecological differentiation can lead to speciation when few genomic regions sweep through a population that resides in one habitat and at the same time gene flow to and from populations residing in other habitats is interrupted (119).

What Types of Pathogen Populations Can Be Distinguished?

The relative contribution of recombination and mutation to the diversity that exists among members of bacterial species varies greatly from species to species (44). Bacterial populations and species in which members vary mainly because of mutations are defined as clonal, and bacterial populations and species in which variation is mainly generated by recombination are defined as freely recombining, or panmictic (41).

By identifying the relative effects of mutation and recombination in a pathogen population, we can classify pathogen populations into two main groups. A pathogen is considered to have an epidemic population structure when individual members of a recombining pathogen population expand rapidly and become the founders of a new clonal population (122). In this case, all diseased hosts during an epidemic are colonized by the same clonal population while the majority of the recombining source population is hiding in apparently asymptomatic hosts, other host species, or the environment. Over time, the clonal population may change: A new clone may arise from the source population and become the founder of a new clonal population, or a member of the clonal population may expand and replace the other members within the pre-existing clonal population in a process called clonal replacement. A pathogen is considered to have an endemic population structure if it emerged a relatively long time ago and diverged into many different clones. In this case, diseased hosts are colonized by many different clones. Because these clones share the same host environment, they are likely to recombine at some frequency with each other (41).

These concepts are very important for crop disease management because the population structure determines how likely it is that new pathogen variants arise in response to intervention (e.g., the introduction of a new disease-resistant cultivar or a new pesticide). In fact, the existence of a freely recombining and genetically diverse source population can be expected to make it likely that a crop pathogen against which a new crop cultivar is resistant is easily replaced by another clone to which the cultivar is susceptible. Contrastingly, a crop cultivar may display more durable resistance against a clonal pathogen population characterized by very little genetic diversity.

THE FIELD OF POPULATION GENOMICS

What Is Population Genomics?

The epidemiological and evolutionary processes that lead to disease emergence and spread leave their signatures in the genomes of pathogen populations. Therefore, these processes can be inferred by sequencing genes or genomes of representative samples of pathogen populations and by comparing patterns of genetic variation through application of appropriate statistical tools. This approach is called population genomics.

The First Step in Any Population Genomics Analysis: Sampling

Because it is impractical (or almost impossible) to exhaustively sample a population, inferences necessarily need to be made from projections that are based on a relatively small population

sample. Therefore, incorrect sampling leads to incorrect inferences. Obviously, a sample can never precisely represent an entire population, but we can extrapolate the structure of the population as long as we employ appropriate sampling strategies and statistical methods.

First, the sample size must be large enough to detect the genetic variation that exists in the pathogen population as a result of selection or as a result of the ecological, geographical, or epidemiological context. More fundamentally, frequencies of individuals in the sample must be representative of the population structure we want to describe. Therefore, comparing a number of bacterial genomes with each other does not per se constitute a population genomics analysis. In fact, if the genomes that are being compared do not represent the population structure, we are performing a comparative genomics analysis. Only if the genomes that are compared are representative of the population structure can we extrapolate the results to the population scale and define the analysis as population genomics.

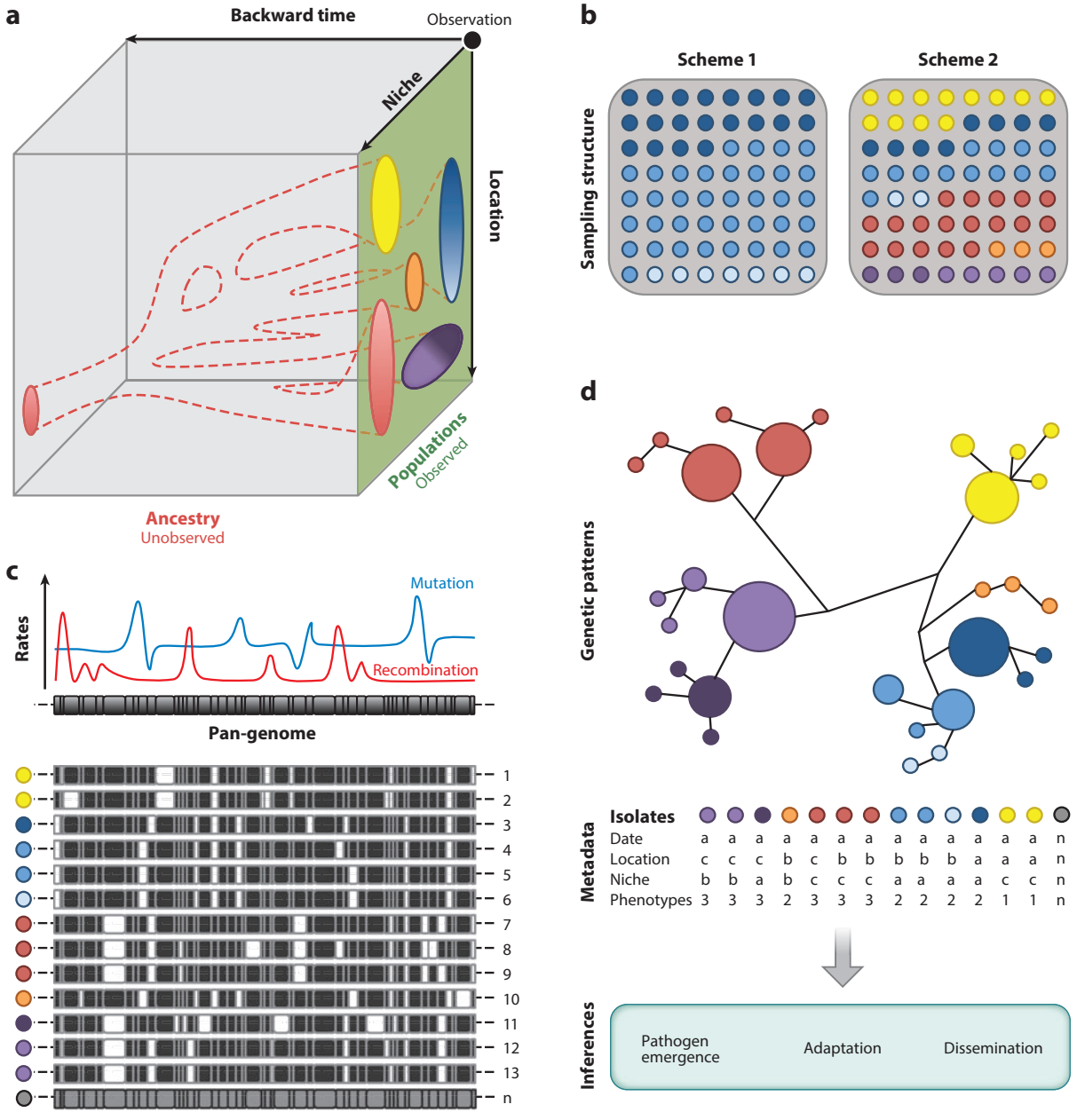
Moreover, the sampling strategy has to be tailored to the space scale, niche, and timescale in which effects of the investigated evolutionary processes are visible. Therefore, depending on the questions we ask, samples must represent a different diversity of niches, dates, geographic origins, phenotypes, and/or genetic relatedness (**Figure 1**). For example, studies of adaptation to different hosts or environments must be based on samples that are representative of the diversity of the pathogen in each of those different hosts and environments. Studies to investigate pathogen dissemination need to be based on samples that are representative of the diversity of the pathogen in each of the geographic areas that are affected by the respective disease. Inference of yearly mutation rates needs to be based on samples that are representative of the diversity across the relevant time frame. Finally, using an example of the analysis of population structure, if a pathogen has an epidemic population structure, then that structure can only be determined by including isolates of the source population from which the epidemic clones arose. The source population, however, may not be present on the host species that is affected by the epidemic clone. Therefore, the result could be misleading if only the affected host species is sampled. In this scenario, the results would be interpreted as a clonal population structure instead of an epidemic population structure. The challenge is that the existence of the source population may not even be known. Therefore, it is extremely difficult to sample correctly and very easy to be misguided while developing a sampling strategy based on the limited a priori knowledge of a pathogen's diversity. In this context, Morris et al. (96) highlight the importance of sampling beyond agriculture if we want to truly understand a plant pathogen's ecology and evolutionary history because the source population may in fact be present in nonagricultural reservoirs, as is the case for *Pseudomonas syringae* (see below).

Figure 1

Population genomics workflow applied to pathogenic bacteria. (*a*) Populations diverge over time, adapting to different niches and/or migrating to new locations. Here, populations are symbolized by different colors. Different shades of the same color symbolize populations within which members are diverging from each other. (*b*) Inference of evolutionary histories requires analysis of samples that are representative of the population structure over time, niches, and geography, and sampling needs to be tailored to the evolutionary scale of the processes that are being described. For example, microevolution is investigated by comparing isolates having diverged very recently (*sampling scheme 1*; *different shades of the same color*), whereas more ancient evolutionary events are investigated on the basis of more distantly related isolates (*sampling scheme 2*; *different colors*). (*c*) Population genomics aims at identifying genomic variation within and between populations. Allelic variation due to mutation and recombination and presence and absence of genes is found by comparing genomes and/or genes with each other or with the known gene repertoire of the species, i.e., the pan-genome. Here, black and white boxes represent genes that are present or absent, respectively. The structure of genomic variation that is being revealed allows the identification of genes under selection, for example, during adaptation to specific crops. (*d*) Finally, correlating signatures of natural selection, presence/absence of genes, whole-genome phylogenies, and metadata allows inferences about pathogen emergence, crop adaptation, and dissemination.

Population Genomics Analysis Using a Small Number of Genes

Until recently, sequencing of bacterial genomes was too expensive to apply to a representative sample of a pathogen population. Therefore, sequencing was necessarily limited to a small number of loci (typically four to seven) in approaches called multilocus sequence typing (MLST) (79) and multilocus sequence analysis (MLSA) (63). MLST usually refers to the analysis of allele distribution among members of a population (or species), whereas MLSA refers to the phylogenetic analysis



of strains belonging to more than one species and using the concatenated DNA sequences from the chosen loci. However, the terms are often used interchangeably.

Although DNA sequence information provided by MLST is minimal compared with that provided by sequencing entire genomes (approximately 1/1,000), MLST represented immense progression compared with earlier techniques. Most importantly, earlier techniques were not based on qualitative DNA sequences but on banding patterns (e.g., restriction fragment length polymorphism, random amplified polymorphic DNA, and amplified fragment length polymorphism) that are difficult to compare between laboratories. MLST allowed, for the first time, establishment of databases that could be accessed by any laboratory around the world with an internet connection. This not only permits users to compare their pathogen isolates with the isolates already in an MLST database without actually transferring isolates or DNA between laboratories, but it even allows users to contribute their own isolate data to these databases to continuously expand them. In fact, some MLST databases of human pathogens today contain data for thousands of pathogen isolates from around the world. In regard to plant pathogens, the Plant-Associated Microbial Database (PAMDB) (3) allows straightforward addition of MLST schemes of any bacterial plant pathogen. Users can add MLST data of isolates they submit to the database as well as images of disease symptoms caused by the same isolates. The database currently contains MLST schemes for *P. syringae*, *Xanthomonas* species, *Ralstonia solanacearum*, *Clavibacter michiganensis*, and *Acidovorax*.

Population Genomics Analysis Using Whole-Genome Sequences

Levels of genetic differentiation between genomes vary from locus to locus (100). Consequently, determining genomic variation on the basis of a small number of MLST loci might not give a good approximation of the true evolutionary history of a population (1). For instance, some evolutionary processes occur within time frames that are too short to affect individual genes, i.e., isolates that are identical in all analyzed MLST loci may have diverged significantly at other loci. Today, next-generation sequencing and new computational methods for genome comparisons and genome-wide population genetic tests overcome these issues, enabling the characterization of patterns of genetic variation within and between populations at the whole-genome scale.

Several sequencing technologies are available today [see the review by Shendure & Ji (120)]. These technologies allow reliable identification, at low costs, of base pair substitutions [usually referred to as single nucleotide polymorphisms (SNPs)] and the determination of the presence or absence of genes. Various bioinformatics pipelines are available to do this. The main difference between pipelines is that they either include or do not include assembly of the individual sequencing reads (which are typically between 100 and a few hundred base pairs long) into contiguous sequences (contigs), whereby each contig may include between one and a few hundred genes. If sequencing reads are not assembled into contigs, they are simply aligned against a complete genome sequence of a closely related reference isolate. The latter allows reliable detection of SNPs and the determination of presence or absence of those genes present in the reference isolate. In other words, this kind of analysis includes the core genome of a species (i.e., the genes present in every single sequenced genome of a species) but only a small fraction of the species' pan-genome (i.e., all genes of a species, including those present only in a subset of strains). Therefore, this kind of analysis severely limits determination of the presence or absence of genes, and thus the investigation of niche adaptation. However, using assemblies instead of reads may increase the risk of including sequencing errors in the analysis. Therefore, a combination of the two approaches is advisable.

After SNPs and the presence or absence of genes have been determined, phylogeny, recombination and mutation rates, and signatures of natural selection can be inferred individually for each

gene and for the whole genome. Combining these genomic data with metadata (e.g., substrate, geographic location, and year of isolation) and/or phenotypic data (e.g., virulence or host range) permits inference of pathogen population structure, adaptation, and dissemination (**Figure 1**).

Pioneering Population Genomics Studies of Human and Animal Diseases

Population genomics studies of human disease have already significantly advanced our understanding of human pathogens, and results are starting to be translated into improved epidemiology and diagnostics (38). Here, we list a few examples that stand out for the new insights they provided.

The ability to measure recombination at the whole-genome level enabled reliable determination of the effects of clinical intervention on human pathogen populations. Beautiful examples are provided by the demonstration that drug resistance was acquired several times in the evolutionary history of a *Staphylococcus aureus* lineage (102) and that *Streptococcus pneumoniae* can adapt to the introduction of new vaccines through population shifts and by switching out, through recombination, the genomic region encoding the targeted antigen (28, 29).

Research performed on *Yersinia pestis*, the causal agent of plague, provides an impressive example of the use of population genomics to investigate the geographic origin of an infectious disease and its historic routes of dissemination (30, 95). Moreover, the resolution provided by comparing whole genomes even allowed resolving dissemination over short times frames from country to country or even hospital to hospital (83, 101, 102). Notably, in these cases, genomes of isolates belonging to a chain of transmission with mutations accumulating along the chain were sequenced. These data made it possible to infer yearly mutation rates by interpolating the genetic distance from the ancestor with the time of isolation. Additionally, modes of transmission can also be inferred; for example, Holt et al. (64) found evidence for the importance of asymptomatic carriers in the epidemiology of *Salmonella enterica* serovar Typhi.

Along with being geographical and epidemiological, the investigated context can also be ecological. For example, host adaptation or host switching can be inferred by comparing populations from different niches or reservoirs. In fact, Sheppard et al. (121) contributed to the identification of the genetic basis of host specificity in *Campylobacter* through genome-wide association studies that extended population genomics to phenotypic characterization. On an even finer scale, genome sequencing can reveal adaptation during infection of a single host; for example, Young et al. (137) succeeded in identifying the small number of SNPs and genes responsible for the transition from nasal carriage to fatal bloodstream infection within a host population of methicillin-sensitive *S. aureus*.

In conclusion, population genomics using whole genomes makes it possible to attribute a geographic, epidemiological, and ecological context to the emergence and spread of bacterial diseases and unravel the underlying adaptation of the causal agents to different hosts and environmental niches.

POPULATION GENOMICS INSIGHTS INTO BACTERIAL PLANT PATHOGEN EVOLUTION, ECOLOGY, AND DISSEMINATION

Pseudomonas syringae

P. syringae comprises a genetically very diverse group of bacteria that includes strains that are as distantly related to each other as *Escherichia coli* and *Salmonella typhimurium* (3). In fact, strains identified as *P. syringae* on the basis of common phenotypic characteristics could be assigned to nine separate species on the basis of DNA similarity measurements (52). Intriguingly though, it

has so far been impossible to find any consistent phenotypic differences that distinguish most of these DNA similarity groups from each other (which are thus referred to as numbered genomospecies 1–9 and not as named species). What could be the reason for this? Could it be that all strains belong to a single population and that HGT is so extensive among its members that it interferes with independent evolution of separate populations within this group? On the basis of MLST studies representing the diversity of *P. syringae* strains that cause various crop diseases (116), homologous recombination appears limited. Also, in another MLST study, it was found that most genomospecies correspond to statistically well-supported phylogenetic groups (18), which is consistent with the idea that genomospecies represent separate populations. However, these studies were based on only 7 (116) or even 4 of the ~6,000 *P. syringae* genes (18). Therefore, at the whole-genome level, HGT of genes that determine phenotypic differences may very well have occurred frequently enough during *P. syringae*'s evolutionary history to explain the absence of consistent phenotypic differences among genomospecies. Moreover, there is experimental evidence of the transformation of *P. syringae* followed by site-directed recombination in planta (78), which further suggests that HGT among *P. syringae* strains in planta may have occurred frequently enough during its evolution to homogenize phenotypes.

Each genomospecies identified by Gardan and colleagues (52) consists of several pathovars (pvs.), i.e., groups of strains that cause distinct disease symptoms on distinct host species (40). But not all *P. syringae* strains can be easily assigned to pathovars. In fact, after many years of collecting and analyzing *P. syringae* strains from nonagricultural environments, in particular, compartments of the water cycle (clouds, precipitation, snow pack, rivers, lakes, and irrigation water), Morris and colleagues (37, 92, 93, 96–98) demonstrated that the number of genetic lineages of *P. syringae* that causes disease on crops is very small compared with the diversity of *P. syringae* strains that exist in nonagricultural and nonplant environments. Many, but not all, of these strains have been found to cause disease in at least some tested crop plants (25, 37, 90, 92, 97). However, it is impractical to test their host range so extensively as to assign them to pathovars. Therefore, pathovar classification has become impractical, and population genomics may be the only possible approach to comprehensively describe and name the diversity within *P. syringae* in the future.

MLST studies that included *P. syringae* pathovars and environmental isolates revealed that strains belonging to the same genetic lineage based on MLST can be isolated from compartments of the water cycle and from crop plants (92, 98). This suggests that in addition to the known inoculum sources of *P. syringae* diseases (seed, other plant material, and weeds), *P. syringae* present in rain, surface water, and irrigation water may also cause disease outbreaks. However, being identical in a small number of genes does not exclude the fact that the most recent common ancestor of the identified water and crop isolates could have already existed hundreds of years ago. Therefore, an epidemiological link between water and crop isolates could only be inferred if they are shown to be almost identical at the whole-genome level, suggesting a very recent common ancestor.

The high diversity of *P. syringae* in the environment compared with the relatively small number of crop pathogenic lineages isolated from diseased crops also suggests the existence of a large *P. syringae* reservoir in the environment (i.e., in compartments of the water cycle, wild plants, and leaf litter). Crop pathogenic lineages may have emerged from these reservoirs in the past and may emerge again in the future. In the case of *P. syringae* pv. *tomato* (*Pto*), one genetic lineage, referred to as T1, has been isolated from tomato fields all over the world (20, 136). Isolates belonging to this lineage are extremely similar to each other and differ by less than 100 nucleotides in more than 3 million base pairs of analyzed genome. *P. syringae* pv. *aesculi* (*Pae*) strains from a recent outbreak of bleeding canker of chestnut in the United Kingdom (59) and *P. syringae* pv. *actinidiae* (*Psa*) strains from recent outbreaks of canker of kiwifruit in Europe, New Zealand, and Chile (19, 82, 85) are similarly closely related to each other. This similarity suggests that these pathogens

only recently emerged from environmental reservoirs. Moreover, Monteil and colleagues (92), using MLST, identified recombination events between the T1 lineage and environmental strains, whereas McCann and colleagues (85) found evidence for recombination with unknown *P. syringae* strains in the genome sequences of *Psa* strains. Both results suggest that *P. syringae* crop pathogens may comprise clonal lineages that emerged from populations that recombine in environmental reservoirs. This would suggest an epidemic population structure. This hypothesis needs to be addressed rigorously by a population genomics analysis based on whole genomes of representative crop and environmental strains. Such analysis will also reveal the extent to which environmental strains can donate new virulence genes to clonal crop pathogenic strains through HGT. In fact, Monteil and colleagues found preliminary evidence for this intriguing hypothesis (92).

Several highly virulent host-specific human pathogens, for example, *Salmonella typhi* or *Bordetella pertussis*, have close relatives with wider host range (*S. typhimurium* and *Bordetella broncoseptica*, respectively) (55, 114). It appears that these human-specific pathogens evolved from their ancestors with a wider host range only a few thousand years ago and after the human population reached a size and complexity sufficient to provide a large supply of susceptible hosts. It has been hypothesized that agriculture played a similar role in the emergence of highly virulent host-specific crop pathogens from ancestors that also had a wider host range (21, 126). In fact, although the ancestors of today's crop pathogens probably needed a broad host range to be fit in diverse plant communities, today's agricultural practice provides large monocultures of susceptible individuals that belong to a single host species. Agriculture, thus, dramatically changed the selection pressure on pathogens, possibly allowing pathogens to adapt to a single host species and to become highly virulent in that single species. This scenario is supported by the finding that the *Pto* lineage T1 has a restricted host range, whereas some of its closest relatives found in the environment have a larger host range (92). However, it is also possible that narrow-host-range strains predated agriculture but simply existed at low frequency and then underwent clonal expansion once they came in contact with fields or orchards of tomato, kiwifruit, or other crops. Again, extensive sampling of crop and environmental *P. syringae* populations, combined with population genomics analyses and host range tests, will be necessary to address this hypothesis.

After a *P. syringae* crop pathogen emerges, how fast does it spread around the world? The similarity of genome sequences of *Psa* and *Pto* from different continents suggests that a clonal lineage of *P. syringae* can spread around the world within 10 to 50 years (19, 20). In particular, over 50 years, epidemic clones of *Pto* replaced each other several times simultaneously in Europe and North America, suggesting frequent transfer of *Pto* between these continents (20). Thus, local endemic *Pto* populations may only play a minor role in the epidemiology of bacterial speck disease of tomato. Moreover, the current data suggest that clonal replacement was not due to the emergence of new clonal lineages from environmental reservoirs but that clones within the same lineage continued adapting to tomato and replaced each other over time (20).

Interestingly, sampling in various regions of Colombia revealed a *Pto* subpopulation distinct from that in North America and Europe (20). This could be due to less intense seed traffic involving South America as compared with traffic between North America and Europe. However, using MLST, the similarity of populations of *P. syringae* in headwaters of rivers on different continents supports the hypothesis that *P. syringae* can travel long distances through the atmosphere (97). Therefore, we can speculate that hypothetical long-distance movement of *Pto* between Europe and North America occurs through the water cycle and is favored by the jetstream, whereas air masses are less frequently exchanged between North and South America. Such atmospheric patterns could keep the *Pto* populations in the Northern and Southern Hemispheres relatively separate.

The geographic origin of *Pto* has not yet been identified. Contrastingly, population genomics and epidemiological data suggest China to be the geographic origin of the *Psa* lineage that causes

the current outbreak of bacterial canker of kiwifruit (19, 82). Genome analysis performed so far linked the worldwide populations to China but not to each other. This suggests that vegetative material exported from China independently to Europe, New Zealand, and Chile may be the vehicle of worldwide dissemination (19, 82). Again, genome sequencing must be extended to larger samples that represent the diversity in each country in order to support more definitive conclusions.

MLST (73) also provided insight into the pathogen population that overcame *PTO* resistance in California: almost all strains were found to belong to the aforementioned T1 lineage (20). MLST (132) and genome sequencing (106) revealed that two separate genetic lineages with very different virulence gene repertoires cause hazelnut decline. Moreover, virulence and host range in *P. syringae* were further addressed by combining MLST with the analysis of presence and absence of virulence genes and phenotypes related to virulence (65, 115). Similar studies are described in detail in a section below and illustrate how comparison of genomes of isolates of the same, or closely related, pathogen populations can aid in understanding crop adaptation and give new insight into basic molecular plant-microbe interactions.

Xanthomonas Species

MLST has revealed that the genetic diversity of strains in the genus *Xanthomonas* is very similar to the diversity among strains classified as *P. syringae* (3). MLST studies of *Xanthomonas* to date were mainly designed to confirm or revise species descriptions (e.g., 2, 17, 138), investigate relatedness between species (62), or determine relatedness between pathovars within a single species (43), but not to address population genomics questions.

As with *P. syringae*, the overall evolutionary relationships among *Xanthomonas* species are the same, independent of the genes used. Therefore, recombination appears to be relatively infrequent between distantly related strains (138). To the contrary, evidence for recombination within closely related subgroups of *Xanthomonas* has been found. Fargier and colleagues (43) inferred that mutation and recombination contributed equally to diversity within *Xanthomonas campestris*, and Mhedbi-Hajri et al. (87) found that recombination occurred as frequently as point mutation in the evolutionary history of *Xanthomonas axonopodis*. Putative recombination events were also detected between *Xanthomonas* species that infect tomato and pepper and *Xanthomonas* strains that infect other plants (62).

Specific evidence for the importance of HGT in the evolution of *Xanthomonas* strains was detected in regard to the gene cluster for the biosynthesis of lipopolysaccharides (LPS) (107). Experimental evidence for homologous recombination in *Xanthomonas* by conjugation in planta was also obtained (11), and there is strong evidence that HGT by homologous and/or site-directed recombination played an important role in host-range evolution in *Xanthomonas*. For example, some phylogenetically distinct strains within the *X. axonopodis* group have similar repertoires of virulence-associated genes and very similar host ranges, whereas their close relatives have different repertoires of virulence-associated genes and different host ranges (61, 86, 87). This suggests that HGT of virulence-associated genes between different genetic lineages heavily contributed to the host range of today's *Xanthomonas* strains. Mhedbi-Hajri and colleagues (87) also attempted to estimate the timeline of evolution of the *X. axonopodis* group by correlating mutations in the four analyzed MLST loci with year of strain isolation. They inferred that the most-recent common ancestor of this group existed approximately 250,000 years ago and did not adapt to specific hosts until after the advent of agriculture, when large populations of susceptible crop hosts became available. They further estimated that recent HGT events occurred over the past few hundred years because agriculture provided the context for contact of different pathovars. This is in agreement with what we proposed above for crop adaptation of *P. syringae*. However, it was shown for human pathogens

that deep sampling of a pathogen population over several years combined with whole-genome sequencing is necessary to reliably infer yearly mutation rates (e.g., 94, 101). Moreover, *P. syringae* isolates that differ by only approximately 100 mutations in 3,000 genes were isolated 40 years apart (20). This suggests that, similar to human and animal pathogens, bacterial plant pathogens evolve much too slowly to reliably estimate a yearly mutation rate based on MLST; only whole-genome sequencing of a large number of strains from the same pathogen population, collected over dozens of years, will enable a reliable estimate of yearly mutation rates for bacterial plant pathogens.

Although many comparative genomic studies in *Xanthomonas* have provided insights into the basis of virulence and host range (e.g., see 12, 31, 109), population genomics studies based on whole genomes have been rare. In one recent study, Bart and colleagues (10) sequenced 65 isolates of the cassava bacterial blight pathogen *X. axonopodis* pv. *manihotis* collected over 70 years from South America, Africa, and South East Asia. Although this collection is a treasure for population genomics studies, the authors' goal was identification of genes for targeted plant breeding (see below). Multiple genomes of the same pathovar were also sequenced for *X. campestris* pv. *musacearum* (133), the pathogen that causes banana *Xanthomonas* wilt in East Africa. In this case, the authors resolved two separate genetic lineages, one causing the disease in Ethiopia and one causing the disease in the neighboring countries. This disproved an earlier hypothesis, based solely on timing of the outbreaks, that Ethiopia was the geographic origin of the outbreaks in the neighboring countries. Such insight is crucial to unravel the sources of disease outbreaks and to reduce the risk of future pathogen introductions into previously unaffected areas.

Ralstonia solanacearum

Similar to *P. syringae* and *Xanthomonas*, *R. solanacearum* encompasses a genetically diverse group of plant pathogens (3). Contrastingly, whereas *P. syringae* and *Xanthomonas* mainly invade plants from aerial parts, *R. solanacearum* is a soilborne pathogen that invades through the roots (54). The *R. solanacearum* complex was divided into five clades using MLST (22, 134). Interestingly, there appears to be a correlation between phylogeny and geographic origin of *R. solanacearum* strains (110, 134), whereas no such correlation appears to exist for *P. syringae* or *Xanthomonas*. We hypothesize that the foliar pathogens *P. syringae* and *Xanthomonas* have been subject to frequent long-distance migration in preagricultural times (possibly through the water cycle), and this may have interfered with the establishment of geographically restricted populations. Being a soilborne pathogen, *R. solanacearum* populations may instead disperse less frequently in the absence of agriculture. International trade of potato tubers and other vegetative material may have been the main avenue of recent long-distance pathogen movement after the advent of agriculture.

Using MLST (134), it was found that recombination played an important role in the diversification of some *R. solanacearum* lineages. Interestingly, the *fliC* gene coding for flagellin seems to be more frequently exchanged among strains than are housekeeping genes, a trend also observed for *P. syringae* (92), with possible reasons discussed below. As with *P. syringae* (78) and *Xanthomonas* (11), there is experimental evidence for recombination among *R. solanacearum* strains in planta (27), and many comparative genomics studies have addressed the molecular and evolutionary basis of host range and virulence commonalities and differences between *R. solanacearum* strains (e.g., 112, 113). However, population genomics studies based on whole genomes have yet to be published for *R. solanacearum*.

Enterobacteriaceae

The Enterobacteriaceae include human, animal, and plant pathogens. Several plant pathogenic species in this family have been described using MLST, in particular, species within the genera

Pantoea (14, 15, 36) and *Pectobacterium* (71, 105). Comparative genomic studies have been performed to obtain insight into virulence mechanisms of enterobacterial plant pathogens (e.g., 56, 81, 124, 125). No population-level studies have been performed using genomes, but genome-derived VNTR (variable number tandem repeat) markers have proven successful in analyzing the worldwide population structure of the fire blight pathogen *Erwinia amylovora* (80). VNTRs are among the most variable regions in bacterial genomes and are thus better suited than MLST for the analysis of genetically monomorphic pathogens, such as *E. amylovora*. Among several other conclusions from this VNTR study, a higher diversity at VNTR loci among strains from North America compared with other regions supports the previous assumption that North America is the geographic origin of this pathogen (16).

Xylella fastidiosa

Xylella fastidiosa has emerged as a detrimental pathogen in the United States and Brazil. This pathogen is closely related to the genus *Xanthomonas* but has a much smaller genome, probably reflecting its ecological niche, which is restricted to life in the plant xylem and the insect vectors (mainly grasshoppers) that transmit it from plant to plant (111). Many different subspecies with different host specificity have been described, and MLST has revealed their phylogenetic relationships (117, 118).

Interestingly, recombination has been implied in the emergence and further evolution of two lineages within *X. fastidiosa* subsp. *pauca*, which cause citrus variegated chlorosis and coffee leaf scorch in Brazil (4, 104). Neither coffee nor citrus are native to Brazil. Although both crops have been cultivated there for hundreds of years, the host-specific, highly virulent *X. fastidiosa* subsp. *pauca* lineages emerged only recently. Nunney and colleagues (104) hypothesize that close relatives of these pathogens existed (and still exist) in native Brazilian plants and may constitute the reservoir from which these lineages emerged. Evidence for recombination in *X. fastidiosa* has also been found among strains isolated in North America, for example (117, 139), and has been experimentally observed in vitro (72).

A beautiful phylogeographic study inferred the probable geographic origin of *X. fastidiosa* subsp. *fastidiosa* in the United States (103). An MLST analysis of strains from the United States and from Central America showed how the genetically monomorphic lineage that causes Pierce's Disease on grapes in the United States is nested among a much wider diversity of strains isolated in Guatemala. Moreover, reported importation of coffee plants from Guatemala to Southern California just before the first known outbreak of the disease in California is strong circumstantial evidence supporting the conclusions from the population genomics evidence.

Other Bacterial Pathogens

MLST has also been used in other important plant pathogens to investigate evolutionary relationships and to describe genetic diversity within and across species, for example, in the causative agent of bacterial canker of tomato, *Clavibacter michiganensis* (68, 89); in the causative agent of bacterial fruit blotch of watermelon and melon, *Acidovorax avenae* subsp. *citrulli* (48); in *Pseudomonas viridiflava* (57); and in various *Phytoplasma* species (32, 34) for which interspecies recombination was inferred (32). Similar to *E. amylovora* (16), genome-derived VNTR markers were successful in determining and comparing Candidatus Liberibacter asiaticus populations in Asia, Brazil, and Florida, revealing, for example, multiple introductions of citrus greening from Asia into Florida (66).

In conclusion, the insight gained from MLST and whole-genome analyses of bacterial plant pathogen populations over the past 10 years is impressive. However, the one limitation of the listed

studies is that sample size and sample representativeness were limited. Thus, conclusions from these studies need to be considered as preliminary. In other words, population genomics studies have so far provided enough data to develop new hypotheses on pathogen emergence, adaptation, and dissemination, but these hypotheses still need to be confirmed using whole-genome sequencing of larger and more representative samples.

AT THE INTERFACE OF POPULATION GENOMICS AND MOLECULAR PLANT-MICROBE INTERACTIONS

In this section, we explore how the fields of population genomics and molecular plant-microbe interactions can reciprocally complement each other. We show that population genomics analysis is an excellent tool for furthering the understanding of molecular plant-microbe interactions. Correspondingly, we propose that population genomics analysis of plant pathogens is most efficacious when informed by the knowledge of molecular plant-microbe interactions. Leveraging the complementary nature of these two fields enables identification of signatures of crop adaptation in genomes, which will lead to a better understanding of plant-microbe interactions and, consequently, aid in the identification of new sources of genetic resistance for plant breeding.

In order to cause disease, plant pathogens must subvert a well-armed plant immune system. The active defenses of the plant immune system are typically delineated into two branches. The first branch mediates a defense response following recognition of conserved pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) via plant pattern recognition receptors (PRRs), leading to the defense response known as pattern- or PRR-triggered immunity (PTI) (70). Successful bacterial pathogens overcome PTI by deploying immunity-suppressing, type III secretion system (T3SS)-secreted effectors (T3Es) (47). However, T3Es also have the potential to betray the presence of the pathogen. In fact, the second branch of the active plant immune system entails specific detection of T3Es, traditionally referred to as avirulence proteins. Detection by plant resistance (R) proteins culminates in a strong defense response known as effector-triggered immunity (ETI) (24, 58). Plant pathogens can further overcome ETI through deployment of yet more T3Es, leading to a molecular arms race between pathogen virulence weapons (i.e., T3Es) and plant R proteins (13). Many excellent reviews are available on this subject (33, 88, 91). Relatively unconsidered, however, is the potential for population genomics approaches to expedite identification of the details of the molecular arms race and, potentially, to predict emergence of new sources of virulence and the corresponding breakdown of disease resistance (*R*) genes. Moreover, population genomics analyses have revealed that allelic diversification in both MAMPs and T3Es is another important component of the molecular arms race, as we briefly review below.

Diversification in Type III Secretion System–Secreted Effectors as an Important Pathogen Virulence Strategy

For compatible interactions (i.e., successful colonization and disease progression), the majority of bacterial plant pathogens deploy a suite of T3Es sufficient to suppress the plant immune system. A T3E is potentially both a virulence component and an avirulence component for the pathogen depending on the context of the *R*-gene repertoire of the host plant (e.g., 49, 76). Therefore, the repertoire of T3Es is thought to be a determinant of the host range of bacterial plant pathogens (60). Accordingly, plant pathogens have evolved myriad mechanisms to refine their T3E repertoire. For example, loss of avirulence effectors through excision of pathogenicity islands, transposon insertions, and frameshift or point mutations have all been observed (6). It was even shown that

the frequency of virulent members in an experimental *P. syringae* population can change from 0% to 100% through a combination of gene loss and selection in planta (108). Moreover, loss of an avirulence effector from *P. syringae* has been observed in the field in a single growing season (135). In other examples, bacterial plant pathogens have been shown to evade recognition by crop *R* genes by either reducing avirulence protein expression (73) or by mutating avirulence genes, which allows escape from recognition by *R* genes without losing virulence activity on susceptible plants (53, 73).

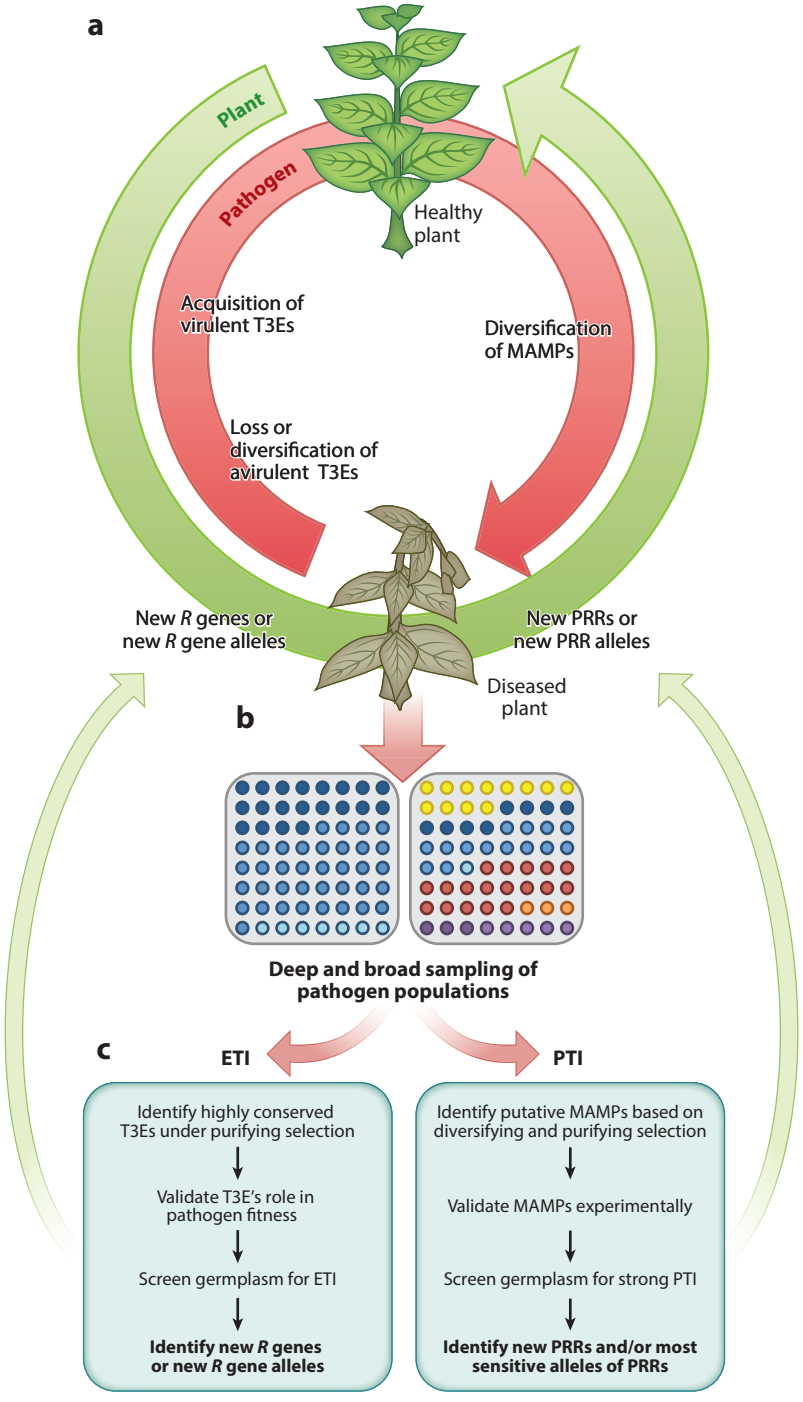
Population genomics can now provide a broader understanding of the diversity and stability of repertoires of T3Es in pathogen populations, with implications for preventing and predicting loss of *R* gene–mediated resistance, as we detail below. Sequencing of plant pathogen populations both broadly (i.e., sampling from different populations) and deeply (i.e., sampling many members of the same population) is informative to this end (**Figure 2**). An example of the power of broad sequencing was the comparison of 19 genomes representing the genetic diversity of *P. syringae* crop strains (9), which revealed the core repertoire of *P. syringae* effectors and the dynamics of adaptation to different crops. An example of deep sequencing was the comparison of 65 *Xanthomonas axonopodis* pv. *manihotis* isolates to identify the least polymorphic T3Es in terms of both their presence and sequence. This information was used to propose the best T3Es to use in screens to identify cognate *R* genes that the pathogen would not easily overcome (10). Sequencing pathogen populations deeply also has the potential to provide novel insights about the role of T3Es in the molecular arms race. For example, several truncation variants of the T3E *hopM1* were identified through sequencing multiple members of the *Pto* population (20). The truncations in *hopM1* were associated with loss of its cell death–eliciting activity (20), which had previously been associated with virulence (8, 35). However, evidence of strong selection for loss of the activity indicates that, in at least some environments, it also has a fitness cost, another intriguing finding only possible through population genomics analysis.

Diversification in Microbe-Associated Molecular Patterns as an Important Pathogen Virulence Strategy

In contrast to T3Es, MAMPs are typically considered highly conserved as a result of being essential for specific microbial functions. However, recent sequencing efforts that sampled within and across populations suggest that MAMPs are not under as strong a purifying selection as previously

Figure 2

Population genomics and molecular plant-microbe interactions work together to unravel the molecular plant-pathogen arms race to inform plant breeders on how to give crops a boost in this race. (*a*) Pathogens acquire, lose, or diversify type III effectors (T3Es) and diversify microbe-associated molecular patterns (MAMPs) to suppress and/or evade plant immunity, and plants acquire and diversify resistance (*R*) genes and pattern recognition receptor (PRR) genes during the molecular plant-pathogen arms race. (*b*) Sampling of plant pathogen populations both deeply within a population (*different shades of blue*) and broadly across different populations (*different colors*) allows identification of highly conserved T3Es and putative MAMPs based on signatures of natural selection. (*c*) Highly conserved effectors are identified on the basis of purifying selection. Their essential contribution to fitness is experimentally validated, germplasm of the crop species and/or related species is screened for effector-triggered immunity (ETI), and, finally, new *R* genes or new *R* gene alleles are cloned and introduced into crops. MAMPs are predicted using purifying selection when sampling across populations and diversifying selection when sampling within populations. Putative MAMPs are experimentally validated, germplasm of the crop species and/or related species is screened for strong PAMP (pathogen-associated molecular pattern)-triggered immunity (PTI), and new PRR genes or new PRR gene alleles are cloned and introduced into crops.


a
Plant
Pathogen
Healthy plant
Acquisition of virulent T3Es
Diversification of MAMPs
Loss or diversification of avirulent T3Es
New R genes or new R gene alleles
New PRRs or new PRR alleles
Diseased plant
b
Deep and broad sampling of pathogen populations
c
ETI
PTI
Identify highly conserved T3Es under purifying selection
Validate T3E's role in pathogen fitness
Screen germplasm for ETI
Identify new R genes or new R gene alleles
Identify putative MAMPs based on diversifying and purifying selection
Validate MAMPs experimentally
Screen germplasm for strong PTI
Identify new PRRs and/or most sensitive alleles of PRRs

presumed (20, 127). Results indicate that allelic diversification in MAMPs is a viable pathogen virulence strategy to avoid PTI and is complementary to deployment of PTI-suppressing T3Es.

Several MAMPs were originally identified on the basis of their high degree of conservation between pathogen populations. In fact, flg22, the first region of flagellin identified to be a MAMP, is the most conserved region of flagellin (46). The bacterial MAMPs elf18 (74) and cold shock protein (CSP) (45) were similarly identified based on relatively high conservation. Mutations in MAMPs were only occasionally identified between distantly related pathogens that diverged over relatively long evolutionary times [for example, in elf18 (75), hrpZ (42, 129), and flg22 (26, 99)]. Accordingly, the current paradigm (69) dictates that the purifying selection to maintain microbial function of the MAMP is greater than the diversifying selection exerted by the fitness cost of MAMP perception by host plants.

However, after deeper sampling within pathogen populations, this view is being challenged. Sun and colleagues (127) identified polymorphisms in flagellin that attenuate PTI within the *Xanthomonas euvesicatoria* population. Also, only certain alleles of flagellin from *Acidovorax avenae* (previously *Pseudomonas avenae*) elicit an immune response in rice cell cultures (23, 128). Tellingly, the flagellin alleles from incompatible strains of *A. avenae* are active elicitors of PTI, highlighting the significance of MAMP diversification for disease outcome. Because flagellin perception is also an important aspect of animal innate immunity (7, 123) and mammalian pathogens also escape flagellin-triggered immunity through allelic diversification (5), escape of host recognition through MAMP diversification appears to be a converged virulence strategy of both animal and plant pathogens.

This new evidence that diversifying selection of MAMPs is an important virulence strategy of plant pathogens has inspired searches for genomic evidence of diversifying selection to thereby identify novel MAMPs. This approach is best exemplified by the work of Cai and colleagues (20), in which a region of flagellin, distinct from flg22, was identified as being under diversifying selection within the *Pto* population and confirmed as an important MAMP, flgII-28, in the *Pto*-tomato pathosystem. The identification of the flgII-28 MAMP depended upon both population genomics analyses and prior knowledge from the molecular plant-microbe interactions field that flagellin is an elicitor of PTI. Intriguingly, different alleles of flgII-28 also elicit varying degrees of PTI in a host-dependent manner (26), providing a further clear example of MAMP diversification as a pathogen virulence strategy. Importantly, none of the different flagellin alleles compromised swimming motility when expressed in a flagellin-deficient background, strongly undermining the dogma that MAMPs (e.g., flagellin) are under robust purifying selection to maintain microbial function (e.g., swimming motility) (26).

McCann and colleagues (84) also confirmed that genomic evidence of diversifying selection in a pathogen population enables identification of novel MAMPs. On the basis of the observation that MAMPs, such as flg22 or flgII-28, are relatively conserved across pathogen populations but relatively variable within pathogen populations, we propose that the most appropriate combination of strategies to identify novel MAMPs using signatures of natural selection appears to be identification of genes under purifying selection across broad pathogen populations but under diversifying selection within pathogen populations (**Figure 2**).

The new understanding that MAMPs undergo allelic diversification to avoid PTI also has important implications for the recently proposed strategy of transferring PRRs (receptors for MAMPs) to improve agricultural disease resistance. The transfer of *EFR*, the corresponding PRR for the elf18 MAMP, from *Brassicaceae* into *Solanaceae* (75) and the transfer of *Ve1*, a PRR associated with *Verticillium* resistance, from *Solanaceae* into *Brassicaceae* (50) have both been shown to improve disease resistance. It is postulated that PRRs will prove to be more durable than *R* genes in the field as a result of MAMPs being under stronger purifying selection than T3Es in pathogen populations.

However, as the recent population-level analyses described above show, this assumption may not be valid.

It is also important to call attention to examples of diversification in PRRs. For example, flg22 perception varies dramatically among *Arabidopsis* ecotypes, although the variation is caused by differences in protein expression and downstream signaling components in addition to allelic differences in the *FLS2* sequence (130). Similarly, the MAMPs flg22, chitin, and elf18 elicit significantly different PTI responses among various *Brassica napus* cultivars, suggesting PRR polymorphism within the species (77). Also, some alleles of flgII-28 elicit a stronger immune response in pepper than in tomato, suggesting that the recently emerged flgII-28 alleles encode stronger ligands for the corresponding allele of the receptor in pepper, relative to tomato (26). Therefore, we postulate a MAMP-PRR arms race driven by both the presence/absence and allelic diversification of MAMPs and PRRs, similar to the T3E–*R* genes arms race between plants and their pathogens (**Figure 2**). Equipped with an understanding of a multifaceted arms race between plants and their pathogens, it should be possible to design novel population genomics–enabled approaches to identify new pathogen resistance determinants, as we propose below.

TRANSLATING POPULATION GENOMICS FINDINGS INTO MORE EFFECTIVE DISEASE DIAGNOSTICS, PREVENTION, AND CONTROL

The cost of sequencing a bacterial genome dropped to less than $100 in 2013 and is thus already competitive with the price for performing MLST. Because the price will continue to drop, and analysis tools will continue to improve, we predict that soon bacterial pathogens will be identified by whole-genome sequencing and genome sequencing will be as widely used as polymerase chain reaction (PCR) is today.

How can we take advantage of the resultant wealth of data? It has recently been proposed that genome sequencing could entirely replace traditional identification of human pathogens and molecular epidemiological analysis in clinical microbiology (38). We think that the same is true for plant pathogens. The main challenge is the establishment of a database that will allow integration of genome sequences of pathogen isolates with respective metadata, such as plant species of isolation, disease symptoms, and date and geographic location of isolation. These data would then need to be further integrated with complete information on well-characterized reference strains and our knowledge of pathogenicity and virulence genes.

If such a database were developed and isolates from plant disease outbreaks worldwide were routinely added, it would be possible to easily and precisely identify pathogens (on the basis of genome comparison with reference strains), to infer the likely geographic source of outbreak strains (on the basis of comparison with similar strains that were recently isolated at other geographic locations), to determine the availability of cultivars resistant to the outbreak strain (on the basis of effector gene repertoires), and to determine susceptibility to antibiotics or other pesticides (on the basis of the presence of antibiotic or copper resistance genes, for example). Of course, this need not be limited to bacterial pathogens. Sequencing of RNA extracted directly from samples of diseased plants could allow fast and precise identification of all plant pathogens, including viruses.

Because of the above described arms race between MAMPs and PRRs, a promising crop improvement strategy would be to identify not only novel PRRs to transfer between plant families (75) but also to identify (or engineer) better alleles of PRRs for such transfer. Additionally, the described large-scale genomic monitoring of plant pathogen populations in the field (and possibly in nonagricultural environments) should enable the identification of new MAMP alleles in the plant-pathogen arms race. It may, therefore, be possible to screen for new alleles of pathogen MAMPs, identify alleles of the corresponding PRR that are better at recognizing the new MAMPs,

and deploy these alleles in the field before a pathogen epidemic occurs. Of course, the same can be done with regard to ETI by screening crop and field isolates for loss or acquisition of T3Es for early detection of *R* gene breakdown (**Figure 2**). Furthermore, parallel identification of novel *R* genes that recognize highly conserved pathogen T3Es will facilitate rapid deployment of new resistance determinants, as previously proposed (10, 88, 131).

Therefore, extending next-generation sequencing and population genomics analysis to plant diagnostic clinics and border controls around the globe would allow early detection of emerging plant pathogens and of new highly virulent pathogen variants that have overcome disease *R* genes. This would in turn permit early intervention to deploy effective eradication programs, rigorous import/export controls, and applications of appropriate pesticides, as well as to develop and plant new resistant crop cultivars based on a combination of ETI and PTI.

SUMMARY POINTS

1. For effective population genomics analyses, next-generation sequencing needs to be applied to samples representative of pathogen populations and tailored to the evolutionary processes that are being investigated.
2. The combination of population genomics and molecular plant-microbe interactions gives deeper insight into disease emergence, crop adaptation, and the molecular plant-pathogen arms race than does either field separately.
3. Plant disease prevention and control will greatly benefit from the integration of population genomics, molecular plant-microbe interactions, and worldwide genome-based plant disease monitoring.

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

We thank John McDowell (Virginia Tech) for critical reading of the manuscript and for helpful suggestions.

LITERATURE CITED

1. Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, et al. 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8:e1002776
2. Ah-You N, Gagnevin L, Grimont PA, Brisse S, Nesme X, et al. 2009. Polyphasic characterization of xanthomonads pathogenic to members of the Anacardiaceae and their relatedness to species of *Xanthomonas*. *Int. J. Syst. Evol. Microbiol.* 59:306–18
3. Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, et al. 2010. PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 100:208–15
4. Almeida RP, Nascimento FE, Chau J, Prado SS, Tsai CW, et al. 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Appl. Environ. Microbiol.* 74:3690–701
5. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, et al. 2005. Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc. Natl. Acad. Sci. USA* 102:9247–52
6. Arnold DL, Jackson RW, Waterfield NR, Mansfield JW. 2007. Evolution of microbial virulence: the benefits of stress. *Trends Genet.* 23:293–300

7. Ausubel FM. 2005. Are innate immune signaling pathways in plants and animals conserved? *Nat. Immunol.* 6:973–79
8. Badel JL, Shimizu R, Oh H-S, Collmer A. 2006. A *Pseudomonas syringae* pv. *tomato avrE1/hopM1* mutant is severely reduced in growth and lesion formation in tomato. *Mol. Plant-Microbe Interact.* 19:99–111
9. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Shahid Mukhtar M, et al. 2011. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* 7:e1002132
10. Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, et al. 2012. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. USA* 109:E1972–79
11. Basim H, Stall RE, Minsavage GV, Jones JB. 1999. Chromosomal gene transfer by conjugation in the plant pathogen *Xanthomonas axonopodis* pv. *vesicatoria*. *Phytopathology* 89:1044–49
12. Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angiuoli SV, et al. 2011. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* 193:5450–64
13. Boller T, He SY. 2009. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* 324:742–44
14. Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J, Coutinho T. 2008. Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst. Appl. Microbiol.* 31:447–60
15. Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, et al. 2009. *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59:2339–45
16. Buhlmann A, Dreo T, Rezzonico F, Pothier JF, Smits TH, et al. 2014. Phylogeography and population structure of the biologically invasive phytopathogen *Erwinia amylovora* inferred using minisatellites. *Environ. Microbiol.* In press. doi: 10.1111/1462-2920.12289
17. Bui Thi Ngoc L, Verniere C, Jouen E, Ah-You N, Lefeuvre P, et al. 2010. Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of *Xanthomonas citri* pv. *citri* and *Xanthomonas campestris* pv. *bilvae*. *Int. J. Syst. Evol. Microbiol.* 60:515–25
18. Bull CT, Clarke CR, Cai R, Vinatzer BA, Jardini TM, Koike ST. 2011. Multilocus sequence typing of *Pseudomonas syringae* sensu lato confirms previously described genomospecies and permits rapid identification of *P. syringae* pv. *coriandricola* and *P. syringae* pv. *apii* causing bacterial leaf spot on parsley. *Phytopathology* 101:847–58
19. Butler MI, Stockwell PA, Black MA, Day RC, Lamont IL, Poulter RTM. 2013. *Pseudomonas syringae* pv. *actinidiae* from recent outbreaks of kiwifruit bacterial canker belong to different clones that originated in china. *PLoS ONE* 8:e57465
20. Cai R, Lewis J, Yan S, Liu H, Clarke CR, et al. 2011. The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathog.* 7:e1002130
21. Cai R, Yan S, Liu H, Leman S, Vinatzer BA. 2011. Reconstructing host range evolution of bacterial plant pathogens using *Pseudomonas syringae* pv. *tomato* and its close relatives as a model. *Infect. Genet. Evol.* 11:1738–51
22. Castillo JA, Greenberg JT. 2007. Evolutionary dynamics of *Ralstonia solanacearum*. *Appl. Environ. Microbiol.* 73:1225–38
23. Che F-S, Nakajima Y, Tanaka N, Iwano M, Yoshida T, et al. 2000. Flagellin from an incompatible strain of *Pseudomonas avenae* induces a resistance response in cultured rice cells. *J. Biol. Chem.* 275:32347–56
24. Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124:803–14
25. Clarke CR, Cai R, Studholme DJ, Guttman DS, Vinatzer BA. 2010. *Pseudomonas syringae* strains naturally lacking the classical *P. syringae hrp/hrc* locus are common leaf colonizers equipped with an atypical type III secretion system. *Mol. Plant-Microbe Interact.* 23:198–210

26. Clarke CR, Chinchilla D, Hind SR, Taguchi F, Miki Y, et al. 2013. Allelic variation in two distinct *Pseudomonas syringae* flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. *New Phytol.* 200:847–60
27. Coupat-Goutaland B, Bernillon D, Guidot A, Prior P, Nesme X, Bertolla F. 2010. *Ralstonia solanacearum* virulence increased following large interstrain gene transfers by natural transformation. *Mol. Plant-Microbe Interact.* 24:497–505
28. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, et al. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* 45:656–63
29. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, et al. 2011. Rapid pneumococcal evolution in response to clinical interventions. *Science* 331:430–44
30. Cui Y, Yu C, Yan Y, Li D, Li Y, et al. 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* 110:577–82
31. da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, et al. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459–63
32. Danet JL, Balakishiyeva G, Cimerman A, Sauvion N, Marie-Jeanne V, et al. 2011. Multilocus sequence analysis reveals the genetic diversity of European fruit tree phytoplasmas and supports the existence of inter-species recombination. *Microbiology* 157:438–50
33. Dangl JL, Horvath DM, Staskawicz BJ. 2013. Pivoting the plant immune system from dissection to deployment. *Science* 341:746–51
34. Davis RE, Zhao Y, Dally EL, Lee I-M, Jomantiene R, Douglas SM. 2013. "*Candidatus Phytoplasma pruni*," a novel taxon associated with X-disease of stone fruits, *Prunus* spp.: multilocus characterization based on 16S rRNA, secY, and ribosomal protein genes. *Int. J. Syst. Evol. Microbiol.* 63:766–76
35. DebRoy S, Thilmony R, Kwack Y-B, Nomura K, He SY. 2004. A family of conserved bacterial effectors inhibits salicylic acid–mediated basal immunity and promotes disease necrosis in plants. *Proc. Natl. Acad. Sci. USA* 101:9927–32
36. Deletoile A, Decre D, Courant S, Passet V, Audo J, et al. 2009. Phylogeny and identification of *Pantoea* species and typing of *Pantoea agglomerans* strains by multilocus gene sequencing. *J. Clin. Microbiol.* 47:300–10
37. Diallo MD, Monteil CL, Vinatzer BA, Clarke CR, Glaux C, et al. 2012. *Pseudomonas syringae* naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. *ISME J.* 6:1325–35
38. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13:601–12
39. Didelot X, Maiden MCJ. 2010. Impact of recombination on bacterial evolution. *Trends Microbiol.* 18:315–22
40. Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev. Plant Pathol.* 59:153–68
41. Dykhuizen D, Kalia A. 2007. The population structure of pathogenic bacteria. In *Evolution in Health and Disease*, ed. SC Stearns, JC Koella, pp. 185–98. Oxford, UK: Oxford Univ. Press.
42. Engelhardt S, Lee J, Gäbler Y, Kemmerling B, Haapalainen M-L, et al. 2009. Separable roles of the *Pseudomonas syringae* pv. *phaseolicola* accessory protein HrpZ1 in ion-conducting pore formation and activation of plant immunity. *Plant J.* 57:706–17
43. Fargier E, Fischer-Le Saux M, Manceau C. 2011. A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. *Syst. Appl. Microbiol.* 34:156–65
44. Feil EJ, Holmes EC, Bessen DE, Chan M-S, Day NPJ, et al. 2001. Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* 98:182–87
45. Felix G, Boller T. 2003. Molecular sensing of bacteria in plants: the highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. *J. Biol. Chem.* 278:6201–8

46. Felix G, Duran JD, Volko S, Boller T. 1999. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* 18:265–76
47. Feng F, Zhou J-M. 2012. Plant–bacterial pathogen interactions mediated by type III effectors. *Curr. Opin. Plant Biol.* 15:469–76
48. Feng J, Schuenzel EL, Li J, Schaad NW. 2009. Multilocus sequence typing reveals two evolutionary lineages of *Acidovorax avenae* subsp. *citrulli*. *Phytopathology* 99:913–20
49. Ferrante P, Clarke CR, Cavanaugh KA, Michelmore RW, Buonaurio R, Vinatzer BA. 2009. Contributions of the effector gene *hopQ1-1* to differences in host range between *Pseudomonas syringae* pv. *phaseolicola* and *P. syringae* pv. *tabaci*. *Mol. Plant Pathol.* 10:837–42
50. Fradin EF, Abd-El-Haliem A, Masini L, van den Berg GCM, Joosten MHAJ, Thomma BPHJ. 2011. Interfamily transfer of tomato ve1 mediates verticillium resistance in *Arabidopsis*. *Plant Physiol.* 156:2255–65
51. Fraser C, Hanage WP, Spratt BG. 2007. Recombination and the nature of bacterial speciation. *Science* 315:476–80
52. Gardan L, Shafik H, Belouin S, Broch R, Grimont F, Grimont PAD. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int. J. Syst. Bacteriol.* 49:469–78
53. Gassmann W, Dahlbeck D, Chesnokova O, Minsavage GV, Jones JB, Staskawicz BJ. 2000. Molecular evolution of virulence in natural field strains of *Xanthomonas campestris* pv. *vesicatoria*. *J. Bacteriol.* 182:7053–59
54. Genin S, Denny TP. 2012. Pathogenomics of the *Ralstonia solanacearum* species complex. *Annu. Rev. Phytopathol.* 50:67–89
55. Gerlach G, von Wintzingerode F, Middendorf B, Gross R. 2001. Evolutionary trends in the genus *Bordetella*. *Microbes Infect.* 3:61–72
56. Glasner JD, Marquez-Villavicencio M, Kim HS, Jahn CE, Ma B, et al. 2008. Niche-specificity and the variable fraction of the pectobacterium pan-genome. *Mol. Plant-Microbe Interact.* 21:1549–60
57. Goss EM, Kreitman M, Bergelson J. 2005. Genetic diversity, recombination and cryptic clades in *Pseudomonas viridiflava* infecting natural populations of *Arabidopsis thaliana*. *Genetics* 169:21–35
58. Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL. 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60:425–49
59. Green S, Studholme DJ, Laue BE, Dorati F, Lovell H, et al. 2010. Comparative genome analysis provides insights into the evolution and adaptation of *Pseudomonas syringae* pv. *aesculi* on *Aesculus hippocastanum*. *PLoS ONE* 5:e10224
60. Guttman DS, Vinatzer BA, Sarkar SF, Ranall MV, Kettler G, Greenberg JT. 2002. A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science* 295:1722–26
61. Hajri A, Brin C, Hunault G, Lardeux F, Lemaire C, et al. 2009. A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* 4:e6632
62. Hamza AA, Robene-Soustrade I, Jouen E, Lefeuvre P, Chiroleu F, et al. 2012. Multilocus sequence analysis– and amplified fragment length polymorphism–based characterization of xanthomonads associated with bacterial spot of tomato and pepper and their relatedness to *Xanthomonas* species. *Syst. Appl. Microbiol.* 35:183–90
63. Hanage WP, Fraser C, Spratt BG. 2006. Sequences, sequence clusters and bacterial species. *Philos. Trans. R. Soc. B* 361:1917–27
64. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, et al. 2008. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* 40:987–93
65. Hwang MS, Morgan RL, Sarkar SF, Wang PW, Guttman DS. 2005. Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 71:5182–91
66. Islam M-S, Glynn J, Bai Y, Duan Y-P, Coletta-Filho H, et al. 2012. Multilocus microsatellite analysis of "*Candidatus Liberibacter asiaticus*" associated with citrus Huanglongbing worldwide. *BMC Microbiol.* 12:39
67. Jackson RW, Vinatzer B, Arnold DL, Dorus S, Murillo J. 2011. The influence of the accessory genome on bacterial pathogen evolution. *Mobile Genet. Elem.* 1:55–65

68. Jacques MA, Durand K, Orgeur G, Balidas S, Fricot C, et al. 2012. Phylogenetic analysis and polyphasic characterization of *Clavibacter michiganensis* strains isolated from tomato seeds reveal that nonpathogenic strains are distinct from *C. michiganensis* subsp. *michiganensis*. *Appl. Environ. Microbiol.* 78:8388–402
69. Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444:323–29
70. Katagiri F, Tsuda K. 2010. Understanding the plant immune system. *Mol. Plant-Microbe Interact.* 23:1531–36
71. Kim HS, Ma B, Perna NT, Charkowski AO. 2009. Phylogeny and virulence of naturally occurring type III secretion system-deficient *Pectobacterium* strains. *Appl. Environ. Microbiol.* 75:4539–49
72. Kung SH, Almeida RP. 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 77:5278–84
73. Kunkeaw S, Tan S, Coaker G. 2010. Molecular and evolutionary analyses of *Pseudomonas syringae* pv. *tomato* race 1. *Mol. Plant-Microbe Interact.* 23:415–24
74. Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G. 2004. The N terminus of bacterial elongation factor tu elicits innate immunity in *Arabidopsis* plants. *Plant Cell* 16:3496–507
75. Lacombe S, Rougon-Cardoso A, Sherwood E, Peeters N, Dahlbeck D, et al. 2010. Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat. Biotechnol.* 28:365–69
76. Lin N-C, Martin GB. 2005. An *avrPto/avrPtoB* mutant of *Pseudomonas syringae* pv. *tomato* DC3000 does not elicit Pto-mediated resistance and is less virulent on tomato. *Mol. Plant-Microbe Interact.* 18:43–51
77. Lloyd SR, Schoonbeek H-j, Trick M, Zipfel C, Ridout CJ. 2013. Methods to study PAMP-triggered immunity in *Brassica* species. *Mol. Plant-Microbe Interact.* 27:286–95
78. Lovell HC, Mansfield JW, Godfrey SAC, Jackson RW, Hancock JT, Arnold DL. 2009. Bacterial evolution by genomic island transfer occurs via DNA transformation in planta. *Curr. Biol.* 19:1586–90
79. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* 95:3140–45
80. Malnoy M, Martens S, Norelli JL, Barny MA, Sundin GW, et al. 2012. Fire blight: applied genomic insights of the pathogen and host. *Annu. Rev. Phytopathol.* 50:475–94
81. Mann RA, Smits THM, Buhlmann A, Blom J, Goesmann A, et al. 2013. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS ONE* 8:e55644
82. Mazzaglia A, Studholme DJ, Taratufolo MC, Cai R, Almeida NF, et al. 2012. *Pseudomonas syringae* pv. *actinidiae* (PSA) isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. *PLoS ONE* 7:e57464
83. McAdam PR, Templeton KE, Edwards GF, Holden MTG, Feil EJ, et al. 2012. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* 109:9107–12
84. McCann HC, Nahal H, Thakur S, Guttman DS. 2012. Identification of innate immunity elicitors using molecular signatures of natural selection. *Proc. Natl. Acad. Sci. USA* 109:4215–20
85. McCann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, et al. 2013. Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. *PLoS Pathog.* 9:e1003503
86. Mhedbi-Hajri N, Darrasse A, Pigne S, Durand K, Fouteau S, et al. 2011. Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evol. Biol.* 11:67
87. Mhedbi-Hajri N, Hajri A, Boureau T, Darrasse A, Durand K, et al. 2013. Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS ONE* 8:e58474
88. Michelmore RW, Christopoulou M, Caldwell KS. 2013. Impacts of resistance gene genetics, function, and evolution on a durable future. *Annu. Rev. Phytopathol.* 51:291–319
89. Milijašević-Marčić S, Gartemann K-H, Frohwitter J, Eichenlaub R, Todorović B, et al. 2012. Characterization of *Clavibacter michiganensis* subsp. *michiganensis* strains from recent outbreaks of bacterial wilt and canker in Serbia. *Eur. J. Plant Pathol.* 134:697–711
90. Mohr TJ, Liu H, Yan S, Morris CE, Castillo JA, et al. 2008. Naturally occurring nonpathogenic isolates of the plant pathogen *Pseudomonas syringae* lack a type III secretion system and effector gene orthologues. *J. Bacteriol.* 190:2858–70

91. Monaghan J, Zipfel C. 2012. Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.* 15:349–57
92. Monteil CL, Cai R, Liu H, Mechan Llontop ME, Leman S, et al. 2013. Non-agricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytol.* 199:800–11
93. Monteil CL, Guilbaud C, Glaux C, Lafolie F, Soubeyrand S, Morris CE. 2012. Emigration of the plant pathogen *Pseudomonas syringae* from leaf litter contributes to its population dynamics in alpine snowpack. *Environ. Microbiol.* 14:2099–112
94. Morelli G, Didelot X, Kusecek B, Schwarz S, Bahlawane C, et al. 2010. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet.* 6:e1001036
95. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, et al. 2010. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* 42:1140–43
96. Morris CE, Monteil CL, Berge O. 2013. The life history of *Pseudomonas syringae*: linking agriculture and Earth system processes. *Annu. Rev. Phytopathol.* 51:85–104
97. Morris CE, Sands DC, Vanneste JL, Montarry J, Oakley B, et al. 2010. Inferring the evolutionary history of the plant pathogen *Pseudomonas syringae* from its biogeography in headwaters of rivers in North America, Europe, and New Zealand. *mBio* 1:1–10
98. Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, et al. 2008. The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J.* 2:321–34
99. Mueller K, Chinchilla D, Albert M, Jehle AK, Kalbacher H, et al. 2012. Contamination risks in work with synthetic peptides: flg22 as an example of a pirate in commercial peptide preparations. *Plant Cell* 24:3193–97
100. Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18:375–402
101. Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, et al. 2010. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.* 6:e100855
102. Nübel U, Roumagnac P, Feldkamp M, Song J-H, Ko KS, et al. 2008. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* 105:14130–35
103. Nunney L, Yuan X, Bromley R, Hartung J, Montero-Astua M, et al. 2010. Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the U.S. *PLoS ONE* 5:e15488
104. Nunney L, Yuan X, Bromley RE, Stouthamer R. 2012. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Appl. Environ. Microbiol.* 78:4702–14
105. Nykyri J, Niemi O, Koskinen P, Nokso-Koivisto J, Pasanen M, et al. 2012. Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *Pectobacterium wasabiae* SCC3193. *PLoS Pathog.* 8:e1003013
106. O'Brien H, Thakur S, Gong Y, Fung P, Zhang J, et al. 2012. Extensive remodeling of the *Pseudomonas syringae* pv. *avellanae* type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol.* 12:141
107. Patil P, Bogdanove A, Sonti R. 2007. The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evol. Biol.* 7:243
108. Pitman AR, Jackson RW, Mansfield JW, Kaitell V, Thwaites R, Arnold DL. 2005. Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. *Curr. Biol.* 15:2230–35
109. Potnis N, Krasileva K, Chow V, Almeida NF, Patil PB, et al. 2011. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* 12:146
110. Prior P, Fegan M. 2005. Recent development in the phylogeny and classification of *Ralstonia solanacearum*. In *Proceedings of the First International Symposium on Tomato Diseases*, ed. T Momol, JB Jones, pp. 127–36. Leuven, Belgium: ISHS-Acta Horticulturae
111. Purcell AH, Hopkins DL. 1996. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* 34:131–51

112. Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, et al. 2010. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics* 11:379
113. Remenant B, de Cambiaire J-C, Cellier G, Jacobs JM, Mangenot S, et al. 2011. *Ralstonia syzygii*, the blood disease bacterium and some asian *R. solanacearum* strains form a single genomic species despite divergent lifestyles. *PLoS ONE* 6:e24356
114. Sabbagh SC, Forest CG, Lepage C, Leclerc J-M, Daigle F. 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. *FEMS Microbiol. Lett.* 305:1–13
115. Sarkar SF, Gordon JS, Martin GB, Guttman DS. 2006. Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* 174:1041–56
116. Sarkar SF, Guttman DS. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl. Environ. Microbiol.* 70:1999–2012
117. Scally M, Schuenzel EL, Stouthamer R, Nunney L. 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Appl. Environ. Microbiol.* 71:8491–99
118. Schuenzel EL, Scally M, Stouthamer R, Nunney L. 2005. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 71:3832–39
119. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, et al. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* 336:48–51
120. Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26:1135–45
121. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, et al. 2013. Genome-wide association study identifies vitamin B-5 biosynthesis as a host specificity factor in *Campylobacter*. *Proc. Natl. Acad. Sci. USA* 110:11923–27
122. Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* 90:4384–88
123. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, et al. 2003. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat. Immunol.* 4:1247–53
124. Smits THM, Rezzonico F, Duffy B. 2011. Evolutionary insights from *Erwinia amylovora* genomics. *J. Biotechnol.* 155:34–39
125. Smits THM, Rezzonico F, Kamber T, Blom J, Goesmann A, et al. 2010. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. *Mol. Plant-Microbe Interact.* 23:384–93
126. Stukenbrock EH, McDonald BA. 2008. The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.* 46:75–100
127. Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF. 2006. Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2–dependent defenses. *Plant Cell* 18:764–79
128. Tanaka N, Che F-S, Watanabe N, Fujiwara S, Takayama S, Isogai A. 2003. Flagellin from an incompatible strain of *Acidovorax avenae* mediates H_2O_2 generation accompanying hypersensitive cell death and expression of PAL, Cht-1, and PBZ1, but not of LOX in rice. *Mol. Plant-Microbe Interact.* 16:422–28
129. Tsunemi K, Taguchi F, Marutani M, Watanabe-Sugimoto M, Inagaki Y, et al. 2011. Degeneration of *hrpZ* gene in *Pseudomonas syringae* pv. *tabaci* to evade tobacco defence: an arms race between tobacco and its bacterial pathogen. *Mol. Plant Pathol.* 12:709–14
130. Vetter MM, Kronholm I, He F, Häweker H, Reymond M, et al. 2012. Flagellin perception varies quantitatively in *Arabidopsis thaliana* and its relatives. *Mol. Biol. Evol.* 29:1655–67
131. 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
132. Wang PW, Morgan RL, Scortichini M, Guttman DS. 2007. Convergent evolution of phytopathogenic pseudomonads onto hazelnut. *Microbiology* 153:2067–73

133. Wasukira A, Tayebwa J, Thwaites R, Paszkiewicz K, Aritua V, et al. 2012. Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pv. *musacearum*. *Genes* 3:361–77
134. Wicker E, Lefeuvre P, de Cambiaire JC, Lemaire C, Poussier S, Prior P. 2012. Contrasting recombination patterns and demographic histories of the plant pathogen *Ralstonia solanacearum* inferred from MLSA. *ISME J.* 6:961–74
135. Willis DK, Kinscherf TG. 2009. Population dynamics of *Pseudomonas syringae* pv. *tomato* strains on tomato cultivars Rio Grande and Rio Grande-Pto under field conditions. *J. Phytopathol.* 157:219–27
136. Yan S, Liu H, Mohr TJ, Jenrette J, Chiodini R, et al. 2008. Role of recombination in the evolution of the model plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000, a very atypical tomato strain. *Appl. Environ. Microbiol.* 74:3171–81
137. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, et al. 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc. Natl. Acad. Sci. USA* 109:4550–55
138. Young JM, Park DC, Shearman HM, Fargier E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31:366–77
139. Yuan X, Morano L, Bromley R, Spring-Pearson S, Stouthamer R, Nunney L. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. *Phytopathology* 100:601–11