Population Genomics of Bacterial Plant Pathogens

Christina Straub,^{1,2} Elena Colombi,^{3,4} and Honour C. McCann^{5,6,†}

¹ Institute of Environmental Science and Research, Health and Environment, Auckland, New Zealand

² Genomics Aotearoa, New Zealand

³ Curtin Health Innovation Research Institute (CHIRI), Curtin University, Perth, Western Australia, Australia

⁴ School of Pharmacy and Biomedical Sciences, Curtin University, Perth, Western Australia, Australia

⁵ New Zealand Institute for Advanced Study, Massey University, Albany, New Zealand

⁶ Max Planck Institute for Developmental Biology, Tübingen, Germany

Accepted for publication 5 November 2020.

ABSTRACT

Population genomics is transforming our understanding of pathogen biology and evolution, and contributing to the prevention and management of disease in diverse crops. We provide an overview of key methods in bacterial population genomics and describe recent work focusing on three topics of critical importance to plant pathology: (i) resolving pathogen origins and transmission pathways during outbreak events, (ii) identifying the genetic basis of host specificity and virulence, and (iii) understanding how pathogens evolve in response to changing agricultural practices.

Keywords: bacterial evolution, bacterial genomics, bacterial pathogens, disease emergence, diseases in natural plant populations, evolution, plant–pathogen interactions, population genomics

Bacterial plant pathogens cause disease outbreaks with devastating economic costs and severe consequences for regional food security. Challenges to agricultural production will increase as agricultural practices favoring the cultivation of single crop varieties over large areas enhance opportunities for pathogen emergence and spread, and globalized trade in plant material for breeding or cultivation purposes provides avenues for intercontinental dissemination (McDonald and Stukenbrock 2016; Savary et al. 2019). Population genomics contributes to the prevention and management of disease by transforming our understanding of host–pathogen interactions and evolution.

Earlier microbial population genetics studies were performed using multilocus enzyme electrophoresis (MLEE), which used variable banding patterns produced by enzymatic digests of conserved housekeeping genes to characterize the amount of genetic variation in a population. MLEE was superseded by multilocus sequence analysis (MLSA), which involves partial sequencing of multiple housekeeping

genes, and multilocus sequence typing (MLST), where sequence data are converted to numbered sequence types representing unique allele combinations. These methods were widely adopted due to ease of use, reproducibility and the creation of standardized typing schemes (Maiden et al. 1998). MLST is still in use today, and extensions of typing schemes have been proposed to include larger core gene sets (Silva et al. 2018). Whole genome sequencing provides an unparalleled view of the amount and source of genetic variation in natural populations, offering far greater resolution of phylogenetic relationships and contributing to our understanding of how microbes evolve to exploit plants as hosts.

The decreasing cost of whole genome sequencing paired with the increasing accessibility of analytical tools provides us with the opportunity to address critically important topics in plant pathology, from resolving pathogen origins and transmission pathways during outbreak events; to identifying the genetic basis of host specificity and virulence; and understanding how agricultural practices influence pathogen evolution (Laflamme et al. 2020; McCann et al. 2017; Quibod et al. 2020). We review the latest progress in bacterial plant pathogen population genomics focused on these three domains, beginning with an overview of methods commonly used to address key questions in microbial population genomics.

METHODS IN BACTERIAL POPULATION GENOMICS

While the particular workflow used depends on the pathosystem of interest, available sequence data, and questions pursued, there are specific approaches that can be used to attain common goals in the

[†]Corresponding author: H. McCann; honour.mccann@tuebingen.mpg.de

Funding: Support was provided by Max-Planck-Gesellschaft and the Royal Society of New Zealand (Marsden Fast-Start (MAU1709)).

First and second authors contributed equally to this work.

The author(s) declare no conflict of interest.

analysis of pathogen populations. Most studies begin with short-read sequencing of single colony isolates from sampled plant material, followed by a read-mapping based workflow, an assembly-based workflow, or both (Fig. 1). Although we do not discuss sampling here, a sampling scheme appropriate for the question of interest is an essential component of any study. For example, sequencing isolates from a limited number of hosts sampled across multiple countries may reveal the global population structure of a broadly distributed pathogen, but provides little insight into the genetic diversity and evolutionary processes operating at local scales.

Populations are often operationally defined based on the question under study, for example as the set of strains collected from a particular host. Bacterial populations may be grouped according to the host or environment of isolation, at both small and large spatial scales. While the chosen division between populations may be meaningful according to the question under study, investigation of the underlying population structure may reveal little genetic variation between them (Balloux 2010, 2018). The population structure—or the amount and organization of genetic variation in the population—will vary according to the processes generating and redistributing that variation. Bacteria undergoing little recombination may exhibit a clonal population structure, while those that recombine freely may be described as panmictic. The presence of recombination does not preclude the occasional expansion of clonal lineages, and a spectrum of population structures is observed in natural populations (Spratt and Maiden 1999).

Recombination refers to a variety of mechanisms responsible for the transfer of genetic material from one organism to

another. These mechanisms include RecA-mediated homologous recombination of DNA taken up from the environment as well as the acquisition of mobile elements like a bacteriophage or plasmid. Variation arising from recombination (whether homologous or heterologous) is horizontally acquired, in contrast to vertically inherited variation transmitted during binary fission of a cell into two daughter cells (Hanage 2016). The amount of recombination detected in the population will necessarily inform which methods are best suited for its analysis. For example, admixture analyses are more suitable for highly recombining bacteria where reliable phylogenies cannot be obtained (Cheng et al. 2013; Lawson et al. 2012). Recombination is a profoundly important process that can have varying impacts on the structure of bacterial populations and their associations with plant hosts.

The first goal of many bacterial population genomics studies is to understand the evolutionary history and relationships between individuals sampled from a population. For more clonal bacterial species, this can be inferred using standard phylogenetic approaches. These typically rely on core genome alignments obtained by read-mapping to a complete, well-annotated reference genome (Langmead and Salzberg 2012; Li and Durbin 2010), followed by variant calling and filtering (Garrison and Marth 2012; Li et al. 2009). The core genome alignment will vary according to the reference and strain set chosen, as well as the amount of diversity in the population. The selection of an appropriate reference genome is therefore important. For example, if the purpose of the study is the identification of recent dissemination

FIGURE 1

Representation of a population genomics workflow. After collection of infected plant material, bacterial isolates are cultured and sequenced. The adapter-trimmed and quality filtered sequencing reads can be used for two general approaches: **A**, an assembly-based workflow or **B**, a read-mapping workflow. **A**, The reads are assembled into draft genomes, which are annotated and can be further used for identification of virulence genes, pan-genome analyses and genome-wide association studies. **B**, Read-mapping is performed against a high-quality reference genome, which can also be generated using a combination of long- and short-read sequencing. After variant calling and filtering to remove low-quality calls, a core genome alignment is generated. This serves as the basis for phylogenetic analysis, association studies, and inference of population structure. Together, these approaches allow us to address important questions in plant pathology, from the genomic basis of host specificity and virulence to pathogen origins and evolution in agricultural environments.

events, the reference genome chosen must be sufficiently closely related to the strains under study to resolve fine relationships in the tree. In the absence of a suitable reference genome, a complete (or “closed”) assembly of a strain selected as reference can be obtained using a combination of short- and long-read sequencing (Bankevich et al. 2012; Souvorov et al. 2018; Wick et al. 2017).

The core genome alignment obtained after read mapping and variant calling includes both variant and invariant sites shared between all isolates included in the study. Alignment curation is then performed to remove positions where mobile elements are known to be present in the reference genome or positions that are present in only a fraction of strains and absent in the remainder. The identification and removal of regions in the core genome alignment subject to horizontal transfer or homologous recombination is standard practice, since these sequences have distinct evolutionary histories compared with the vertically inherited portion of the genome (Didelot and Wilson 2015). While recombination and horizontal transfer are important processes in their own right, their presence in a core genome alignment used for phylogenetic inference will distort or obscure the relationships in the tree. Vertically inherited single nucleotide polymorphisms (SNPs) are thus relied on for phylogenetic inference, while polymorphism introduced via recombination is removed. Maximum likelihood methods use alignments of vertically inherited core genome SNPs (or nonrecombinant core genome alignments) along with a model of sequence substitution to return a tree that maximizes the likelihood of observing the alignment, given the model selected. Bootstrapping is performed to provide a measure of support for the relationships displayed in the tree, given the data available (Guindon et al. 2010; Stamatakis 2014).

Reliable estimation of the evolutionary rate of a bacterial pathogen can provide an approximation of when outbreak lineages emerged and the timing of introduction events into new geographic areas. Rate estimation can be performed using a nonrecombinant core genome alignment provided the sampling time for individual isolates is known, the sampling period is sufficiently broad, and there is a temporal signal in the data. The presence of a temporal signal, which indicates substitutions are accumulating in a clock-like manner, can be determined by first using a regression analysis of the genetic distance in a tree with the sampling time. A significant positive correlation between the amount of divergence and date of sampling indicates there is a temporal signal in the data, allowing inference of divergence times and substitution rates (Drummond et al. 2012; Rambaut et al. 2016; To et al. 2016). In a survey of human-associated bacterial pathogens, Duchêne et al. (2016) found most species with sampling periods between 10 and 100 years had measurable rates of evolution, though rates vary between lineages and rate estimates are higher over short time-scales than longer ones. Expanding the sampling period may provide more reliable rate estimates, though sampling periods of even hundreds of years may not be sufficient for the identification of a temporal signal among slowly evolving bacteria.

While reconstructing a core genome alignment may reveal a lot about the evolutionary history and genetic structure of a population, the accessory genome represents a pool of genes whose composition may be linked with fitness in different hosts or environmental contexts. Pangenome analyses are now a standard component of many population genomics studies, subsequent to genome assembly (Bankevich et al. 2012; Wick et al. 2017). The composition of the core and accessory genome will vary depending on the level of diversity in the strain set of interest. All methods are sensitive to assembly quality and annotation method: omission of highly fragmented assemblies and reannotation of assemblies from public databases may be required for some analyses (Aziz et al. 2008; Page et al. 2015; Seemann 2014). The acquisition or diversification of particular genes may be responsible for the emergence of new pathogen lineages with altered host ranges. The identification of secreted proteins, mobile elements, and virulence factors of interest

can be performed using a wide variety of tools and databases, some of which are dedicated to specific pathogen species.

The primary goal of much research in modern plant pathology is identifying the genetic basis of virulence and specificity and understanding the molecular mechanisms involved in microbial pathogenesis and plant disease resistance. Population genomics can contribute to these goals by identifying key genes or variants linked with particular phenotypes of interest. Correlations between genetic variation and quantitative or binary phenotypic traits may be identified using genome-wide association studies (GWAS). GWAS, frequently used to find variants linked with disease in eukaryotes, is increasingly used in bacterial genomics for the identification of variants linked with phenotypes like virulence and antibiotic resistance (Farhat et al. 2013; Kachroo et al. 2019). Controlling for population structure is particularly important: if a phenotype is strongly associated with a particular lineage, causal variants associated with this phenotype will be vertically inherited along with other background (lineage-specific) variants and cannot be easily distinguished from them. Variants acquired via horizontal transfer or arising via convergent evolution may be easier to interpret (Corander et al. 2019). Tools developed for bacterial GWAS account for population structure in different ways while linking SNPs and gene content variation with phenotypic variation (Collins and Didelot 2018; Earle et al. 2016; Lees et al. 2016). These approaches hold considerable promise for the identification of variants linked with host specificity and virulence in bacterial plant pathogens. A new approach for joint analysis of host and pathogen genomes allows for the identification of interacting loci in both organisms at once (Wang et al. 2018).

CROP PATHOGEN EMERGENCE

Understanding the origins of infectious disease outbreaks and the evolutionary processes linked with pathogen emergence and spread are critical areas of inquiry in microbial population biology. Emergence events can occur as a result of adaptation to novel hosts, long-distance dissemination of pathogens to new areas, or introduction of host plants to areas where pathogens are endemic. Novel crop pathogens may emerge from host-associated microbiota within agricultural environments, from wild relatives of domesticated plants, or from nonhost associated environmental reservoirs (McCann 2020). Environmental conditions favoring infection and transmission contribute to the likelihood of epidemic spread after emergence. Human activity plays an important role in disease emergence, both by establishing favorable conditions for pathogen spread and via trade in plant resources.

The emergence of a plant pathogen in agricultural settings can have catastrophic consequences, particularly in settings where large-scale single-variety cultivation occurs (Savary et al. 2019). There is increasing interest in the genomic epidemiology of bacterial plant pathogens, since the identification of pathogen origins and routes of transmission can guide efforts to prevent future incursions. Multiple studies sample pathogen diversity across broad geographic areas, outside agricultural regions and in alternate hosts, to assess spillover risk and identify links between outbreaks and disease reservoirs (McCann et al. 2017; Monteil et al. 2016; Perez-Quintero et al. 2020).

Perez-Quintero et al. (2020) sought to identify the origin of a rapidly expanding outbreak of *Xanthomonas vasicola* pv. *vasculorum* causing bacterial leaf streak of corn (*Zea mays*) in the United States. *Xanthomonas vasicola* pv. *vasculorum* was first described in South Africa in 1949, but after an initial report in Nebraska in 2014, it has now been documented in most corn-growing regions of the United States as well as Argentina and Brazil. Isolates from South Africa, North America, and South America group together in a clade with a sister group of exclusively South African isolates, indicating a likely South African outbreak origin. A tip-dated Bayesian phylogeny suggests there were two separate introduction

events to North and South America 3 to 13 years apart, followed by mixing between both subpopulations. The authors note the use of winter nurseries for breeding and seed production in South America by U.S. corn breeders may have been an avenue for the rapid spread and exchange of *Xanthomonas vasicola* pv. *vasculorum* (Perez-Quintero et al. 2020).

A pandemic of bacterial canker on kiwifruit (*Actinidia* sp.), caused by *Pseudomonas syringae* pv. *actinidiae* was similarly facilitated by trade in plant material between kiwifruit growing regions around the world. In an effort to determine the origin of repeated *Pseudomonas syringae* pv. *actinidiae* outbreaks over the last few decades, McCann et al. (2017) collected isolates from China and New Zealand and included sequence data from outbreaks in Italy, Portugal, South Korea, and Japan. The phylogeny of 80 isolates revealed global outbreak isolates form a sublineage nested within more diverse Ps-3 isolates from multiple provinces in China (McCann et al. 2017). There is little phylogeographic pattern: Ps-3 was circulating rapidly across multiple provinces in China prior and subsequent to the emergence of the pandemic sublineage. A higher resolution tree of Ps-3 alone revealed multiple independent transmission events were responsible for outbreaks in New Zealand and Italy; clustering of Japanese and New Zealand Ps-3 indicates pandemic Ps-3 was likely introduced to Japan from New Zealand. Interestingly, all *Pseudomonas syringae* pv. *actinidiae* isolated in China fall into a single lineage (Ps-3), while isolates from South Korea and Japan reflect the broader diversity of *Pseudomonas syringae* pv. *actinidiae*, suggesting *Pseudomonas syringae* pv. *actinidiae* lineages may emerge outside of China. These studies highlight the potential of broad range geographical sampling to investigate outbreak origins and global spread.

Expanding the scope of sampling beyond the focal agricultural host may reveal alternate hosts or environmental reservoirs of strains with epidemic potential. *Pseudomonas syringae* isolates closely related to pathogenic lineages have been isolated from environmental sources. Some environmental isolates have similar Type 3 secreted effector (T3SE) repertoires and displayed in planta growth within two orders of magnitude to pathogenic strains (Monteil et al. 2013). Lineages of agricultural pathogens in two *Pseudomonas syringae* phylogroups are more closely related to environmental isolates than isolates from cultivated plants, indicating frequent transfer between the environment and agricultural areas may occur (Monteil et al. 2016). Close relatives of known hosts may also serve as disease reservoirs. For example, southern cutgrass (*Leersia hexandra*), a close relative of rice (*Oryza sativa*), is frequently found growing adjacent to rice paddies. *L. hexandra* harbors both *Xanthomonas oryzae* pv. *oryzae*, which causes bacterial blight in rice, and *Xanthomonas oryzae* pv. *leersiae*, which can also colonize rice (Lang et al. 2019). This uncultivated grass is likely to play an important role in the evolution of bacterial blight in rice, an important agricultural pathogen.

A single introduction of *Xylella fastidiosa* subsp. *fastidiosa* into California over a century ago was responsible for destructive outbreaks on grapevine (*Vitis vinifera*) (Vanhove et al. 2020). This well-established pathogen population displays strong geographic clustering and evidence of local adaptation driven by abiotic factors, with variants linked with climatic variables such as temperature, altitude and precipitations. Estimation of the divergence time of each subpopulation (mid-1900s) was not consistent with the recorded presence of the pathogen (early 1900s), suggesting sweeps or the establishment of novel genotypes may have occurred along with the expansions of grapevine cultivation in the 1970s.

GENOMICS OF HOST SPECIFICITY AND VIRULENCE

Population genomics accelerates the identification of genes linked with host specificity and virulence. For example, Laflamme et al. (2020) identified 5,127 T3SEs in 494 different strains of

Pseudomonas syringae, and screened 529 alleles for effector-triggered immunity (ETI) on *Arabidopsis thaliana*. Approximately 11% of the alleles tested were recognized and 97% of *Pseudomonas syringae* genomes were predicted to harbor at least one ortholog highly similar to an ETI-eliciting effector (Laflamme et al. 2020). Linking variation in virulence factor repertoires (including toxins, phytohormones, and secreted virulence proteins) with phenotypic data like host resistance or disease severity is a popular method of identifying putative causal variants prior to experimental confirmation. The development of a method for the joint analysis of host and pathogen genomes by Wang et al. (2018) allows for the identification of interacting loci in both organisms. Notably, SNPs and insertion-deletion polymorphisms in both the core and accessory genome are included in the association mapping. This method was applied to interactions between 130 inbred lines of *Arabidopsis thaliana* and 22 strains of *Xanthomonas arboricola*, with whole genome sequencing performed for all of these. All combinations were phenotyped by inoculating each host genotype with each pathogen strain and scoring for symptom production. The analysis revealed that while there is a lack of *Arabidopsis thaliana* genes conferring broad-spectrum resistance to *Xanthomonas arboricola*, there are variants with pathogen-specific interactions (Wang et al. 2018). The identification of variants in interacting organisms' genomes will allow us to characterize novel components of host-pathogen interactions. The development of methods allowing the precise identification of genes linked with virulence will continue to be an important avenue of research, since large-scale screens are more difficult and time-consuming in nonmodel hosts like long-lived woody perennials.

Pathogenicity of *Pseudomonas syringae* on cherry (*Prunus avium*) is a convergent trait, which makes it an excellent system to identify genes conferring host specificity. Hulin et al. (2018) sequenced a collection of *Pseudomonas syringae* from three different phylogroups and identified four T3SEs over-represented among strains virulent on cherry (*hopAR1*, *hopBB1*, *hopBF1*, and *hopH1*) and two T3SEs that were absent or truncated in cherry pathogens (*avrPto* and *hopAB*). Expression of *hopAB* alleles elicited a defense response and limited bacterial growth, suggesting recognition by cherry selected for the loss or pseudogenization of *hopAB* in different clades. One of the overrepresented T3SEs (*hopAR1*) was found in association with two different prophages (Hulin et al. 2018). Mobile elements like phage, plasmids, and integrative and conjugative elements (ICEs) can facilitate the transfer of pathogenicity genes between otherwise distantly related strains occupying the same niche, contributing to the convergent evolution of specificity and virulence.

Many population genomic studies highlight the importance of mobile elements in pathogen evolution. A family of ICEs is associated with the emergence of the pandemic sublineage of *Pseudomonas syringae* pv. *actinidiae* infecting kiwifruit and with the adaptation of *Pseudomonas syringae* phylogroup 2 strains on cucurbits (McCann et al. 2013; Newberry et al. 2019). All ICEs in this family share a conserved set of backbone genes involved in the ICE lifecycle, as well as a variable set of accessory genes. For instance, the different ICEs associated with Ps-3 in New Zealand, Chile, and Italy carried an identical transposon referred to as the "enolase region" putatively linked with virulence in *Pseudomonas syringae* pv. *actinidiae*, the ICEs found in cucurbit isolates carry T3SEs and a pyoverdine biosynthetic cluster (McCann et al. 2013; Newberry et al. 2019).

Mobile elements may even confer transitions between a nonpathogenic and pathogenic microbial lifestyle. The acquisition of a pPATH plasmid encoding a complete Type 3 secretion system, T3SEs and an indole-3-acetic acid biosynthesis gene cluster transforms *Pantoea agglomerans* from an epiphytic bacterium to a pathogen of *Gypsophila* sp. and sugar beet (*Beta vulgaris*) (Barash and Manulis-Sasson 2009). A linear virulence plasmid encoding *att* (biosynthesis and secretion of arginine and β -lactam-like

FIGURE 2
Large-scale deployment of a single host resistance gene drove a population shift in the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. Quibod et al. (2020) sequenced 91 isolates collected from rice in the Philippines over a 40 year period. A maximum likelihood phylogeny based on a core genome alignment is shown at left, with the year of isolation shown in the middle panel. Colors of terminal nodes and year of isolation indicate cluster assignment based on population structure analysis with BAPS. The outcome of infection assays on Xa4 variety IRBB4 is shown at right, where dark green indicates host resistance and light green indicates host susceptibility. As cultivation of Xa4 varieties increased (rising to 91.5% of all rice cultivation areas by the early 1990s), virulent sublineages emerged, increasing from 20 to 84% of isolates between 1972 and 1990. PX-A1 now dominates the contemporary *Xanthomonas oryzae* pv. *oryzae* population in the Philippines. Adapted from Quibod et al. (2020) with permission.

compounds) and *fas* genes (synthesis and modification of cytokinins) was identified among virulent *Rhodococcus* strains isolated from ornamental *Veronica* sp.; nonpathogenic *Rhodococcus* strains notably differed by the absence of this plasmid. The acquisition of this plasmid transforms *Rhodococcus* from *Nicotiana benthamiana* root hair growth promotion to leafy gall formation (Savory et al. 2017).

Plasmid acquisition in *Agrobacterium* sp. is also linked with different disease phenotypes: conjugative oncogenic tumor-inducing (Ti) plasmids cause crown galls and root-inducing (Ri) plasmids cause hairy root disease. *Agrobacteria* are a diverse, polyphyletic group of bacteria in the agrobacteria–rhizobacteria complex. Weisberg et al. (2020) used multiple methods to classify and explore the evolutionary history of Ti and Ri virulence plasmids across multiple lineages, including *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. Plasmid T-DNA regions are transferred into plant host cells where they recombine into the host genome, a well-known mechanism of transformation widely used in plant molecular biology. T-DNA regions include genes involved in the synthesis of plant-growth promoting hormones auxin and cytokinin, which confer abnormal growth phenotypes in the host, and genes synthesizing opine, which serve as nutrients for the bacteria. There are different classes of opines, and the cognate uptake and catabolism genes are located on the Ti or Ri plasmid, outside the T-DNA region.

Weisberg et al. (2020) built a time-calibrated phylogeny using eight neutrally evolving genes and found *Agrobacterium rhizogenes* diverged more recently (1.6 ± 0.5 mya) than the distantly related *Agrobacterium tumefaciens* (48 ± 10 mya). Multiple approaches were then used to classify a set of 143 plasmids into nine distinct lineages (six Ti types and three Ri types) and assess how gene composition and synteny varies across plasmid types. The authors used this classification scheme as a framework for understanding Ti and Ri plasmid evolution; these plasmids were previously classified using core genes that serve as a major source of variation. T-DNA regions in particular are highly recombinogenic; short invariant border sequences promote gene exchange and chimerization without affecting transformation. There are constraints to exchange however: opine synthesis and uptake/catabolism genes in the same functional class are retained, forming interdependent modules on the same plasmid. Weisberg et al. (2020) developed a model of oncogenic plasmid evolution, proposing Ti and Ri plasmids are derived from an ancestral proto-oncogenic plasmid with opine catabolic activity. This proto-oncogenic plasmid had a T-DNA region with an opine synthesis gene, and developed into either Ti or Ri before recombining with a separate *repABC* plasmid to generate the other class. The independent acquisition of tumorigenic oncogenes is then thought to have occurred, resulting in the emergence of virulence plasmids that may be readily transmitted between saprophytic soil agrobacteria.

Studies of host-specific symbioses among plant-associated bacteria contribute to our understanding of the role of horizontal transfer in host adaptation. The transition between a soil-dwelling lifestyle to being a nitrogen-fixing symbiont of a plant host in the family Fabaceae is one of the most dramatic and well-studied transformations driven by horizontal gene transfer (HGT). Typically, these transitions are mediated by the acquisition of plasmids or ICEs carrying genes for the establishment of N₂-fixing symbiosis (Remigi et al. 2016). Horizontally transferred elements have distinct evolutionary histories compared with the host bacterial genomes and therefore they cannot be used to infer strain phylogeny. However, ICEs may have distinct patterns of association with host genomes, as was found by Greenlon et al. (2019) in an expansive study of *Mesorhizobium* species carrying symbiosis ICEs (ICESym). The authors sampled root nodules from chickpea (*Cicer arietinum*) and wild relatives in natural and agricultural environments in the crop's center of origin in the Fertile Crescent as well as areas where chickpea cultivation was

introduced in ancient (Ethiopia and India) and recent (United States, Canada, and Australia) times. The bacterial symbionts include 28 different species, and there are striking differences in the distribution and abundance of symbionts across all areas sampled. Approximately one third of individual plants sampled were colonized by two *Mesorhizobium* species; strains from multiple clades were frequently isolated from different plants from the same field, from different nodules on the same plant and in some cases even from the same nodule. There were notable differences between symbionts isolated from *C. reticulatum*, the wild ancestor of chickpea; symbionts isolated from cultivated chickpea in areas of long-term traditional cultivation (e.g., India and Ethiopia); and symbionts isolated from areas where chickpea cultivation was recently introduced (e.g., Canada and Australia), where rhizobial inoculants are used. Wild *C. reticulatum* is colonized by two groups of *Mesorhizobium*, while cultivated chickpea in areas of traditional cultivation form nodules with phylogenetically distinct *Mesorhizobium* spp., which display strong geographic structure. The ICESyms in turn display different patterns of association with the symbiont genomes they reside in. Phylogenetically diverse *Mesorhizobium* strains that occur in the same geographic region are more likely to share recently transferred ICESyms compared with closely related strains found in different geographic regions. The authors suggest the bacterial symbionts coevolving with chickpea might have dispersed along with the crop from the center of origin to India and Ethiopia, and subsequently transferred the ICESyms to locally adapted *Mesorhizobium* that ultimately outcompeted the dispersed symbionts. In a separate study, symbiosis plasmids in *Rhizobium* do not appear to have coevolved with the chromosomes or other accessory plasmids; however, they appear to have coevolved with the host plant (Li et al. 2020).

The importance of HGT to bacterial evolution and host–microbe interactions is undeniable, yet the dynamics of mobile elements are mainly addressed through the lens of a particular lineage or pathosystem of interest. Mobile elements have their own evolutionary trajectory however, and may form only transient associations that confer symbiosis or pathogenesis phenotypes to host genomes. Studying the evolutionary dynamics of mobile elements in plant- and soil-associated bacterial populations will shed light into the secret lives of these fascinating elements.

PATHOGEN EVOLUTION IN RESPONSE TO AGRICULTURAL PRACTICES

When disease prevention fails and pathogens become established in agricultural areas, control efforts become critically important. These include the removal and destruction of inoculum sources, the use of biological control or antimicrobial compounds, and the introduction of resistant plant varieties. A clear understanding of the diversity and virulence of established pathogen populations can contribute to the development of location-specific control efforts (Prokhorchik et al. 2020). Monitoring pathogen populations after changes in agricultural management can also provide early warning of the acquisition of antimicrobial resistance genes or the loss of recognized T3SEs.

A recent paper by Quibod et al. (2020) highlights how consistent sampling programs can yield insight into how pathogen populations evolve in response to changing agricultural practices. In the early 20th century, rice cultivation in the Philippines involved a variety of locally adapted landraces. *Xanthomonas oryzae* pv. *oryzae* causes bacterial blight on rice and became the target of breeding efforts soon after the establishment of the International Rice Research Institute (IRRI) in the Philippines in 1960. The first resistant rice variety (IR20) developed by the IRRI encoded *Xa4*, a cell-wall kinase promoting cellulose synthesis after infection (Quibod et al. 2020). IR20 was released for commercial cultivation in 1967, and introgression of the *Xa4* locus into multiple varieties ensued. By

1985, almost 82% of all rice-cultivation areas in the Philippines were planted with *Xa4* varieties. *Xa4* variety cultivation reached 91.5% of all areas planted by the early 1990s. The dramatic increase in the allele frequency of a single resistance gene in a host population drove a significant population shift in *Xanthomonas oryzae* pv. *oryzae*, selecting for strains that could overcome *Xa4*-mediated resistance.

The authors performed a phylogenomic analysis using a collection of *Xanthomonas oryzae* pv. *oryzae* in the Philippines spanning 40 years, and screened 91 isolates for pathogenicity on seven rice lines with seven different resistance genes (Quibod et al. 2020). As cultivation of the *Xa4* varieties increased, two virulent sublineages of *Xanthomonas oryzae* pv. *oryzae* emerged (PX-A1 and PX-B1) and increased in frequency from 20 to 84% between 1972 and 1990, with the greatest increase occurring between 1985 and 1990. First, the PX-B1 sublineage became dominant in the 1980s, before being supplanted by the recent clonal expansion of a PX-A1 sublineage, which now dominates the contemporary pathogen population in the Philippines (Fig. 2). Temporal dating indicates the recent clonal expansion of PX-A1 linked with the loss of *Xa4*-mediated resistance began around 1992. The specific mechanism underlying PX-A1's ability to overcome *Xa4* resistance remains to be clarified, though a GWAS analysis identified 31 SNPs associated with virulence on a *Xa4* host variety. The SNPs are proximal to 14 genes linked with secretion, cell-wall degradation, reactive oxygen species detoxification and lipopolysaccharide production. These data provide a compelling example of a pathogen population shift in response to selection imposed by the widespread adoption of resistant host varieties. There are also clear recommendations for contemporary rice cultivation, which should benefit from the use of other resistance genes that were identified as capable of controlling PX-A1.

Adaptation in crop pathogens is frequently linked with variation in T3SE repertoires. *Xanthomonas perforans* strains carrying *avrXv3* (now *xopAF*) were first identified in Florida tomato crops in 1991. *avrXv3* is recognized by the *Xv3* resistance gene, but in this case, a population shift was detected prior to the widespread commercial deployment of *Xv3* in tomato crops (Wang et al. 2011). In 1998, *Xanthomonas perforans* isolates with null mutations in *avrXv3* were identified (Schwartz et al. 2015). These isolates were pathogenic on tomato and also carried *avrBsT*, an effector not previously documented in *Xanthomonas perforans*. *avrBsT* suppresses defense responses and confers higher fitness on tomato in field settings, but at the cost of reduced virulence in pepper (*Capsicum annuum*). *Xanthomonas perforans* strains with disrupted *avrXv3* and functional *avrBsT* increased from 75% in 2006 to 100% in 2012. The subsequent loss of *avrBsT* was then linked with a host range expansion of *Xanthomonas perforans* onto pepper (Abrahamian et al. 2018; Jibrin et al. 2018; Schwartz et al. 2015; Timilsina et al. 2020).

Antimicrobial applications represent another important selective pressure in agricultural environments. Foliar antimicrobial sprays (and even intravenous antibiotics) are often applied to prevent or manage bacterial infections. The emergence and spread of antimicrobial resistance is a well-documented phenomenon in plant and human pathogens alike (Klemm et al. 2018; Sundin and Wang 2018). As in human pathogens, the acquisition of antimicrobial resistance genes can facilitate the emergence and spread of a plant pathogen. For instance, the seventh pandemic of *Vibrio cholerae* was shaped by the acquisition of the SXT family of ICEs encoding antibiotic resistance genes (Mutreja et al. 2011). Similarly, *Xanthomonas arboricola* pv. *juglandis* strains responsible for a recent pandemic of vertical oozing canker of walnut (*Juglans* spp.) in France have undergone clonal expansion with the acquisition of copper resistance genes encoded by the *Xanthomonas arboricola* pv. *juglandis* ICE (Cesbron et al. 2015). Monitoring for such acquisition events is crucial because the frequency of resistance will increase in a pathogen population subject to

continuous exposure to antibiotics, ultimately compromising the efficacy of the antimicrobial agent used. Population genomics offers us the ability to monitor antimicrobial resistance in pathogen populations.

Copper-based sprays have been used in agriculture since 1882, long before the discovery of the first antibiotic. The sublineage of Psa-3 introduced to New Zealand during the global outbreak of kiwifruit canker disease were initially susceptible to copper, but within 4 years the first copper resistant (CuR) Psa-3 strains were identified (Colombi et al. 2017). Sampling the area where CuR strains were first identified revealed 27% of isolated Psa-3 were CuR. Resistance was acquired via acquisition of three distinct copper resistant ICEs and one plasmid harboring different resistance loci (*czc/cusABC* and *copABCD* systems). Strains collected from kiwifruit leaves in New Zealand in 1991 and 2010 (*Pseudomonas marginalis* and *Pseudomonas syringae* pv. *actinidifoliorum*, respectively) harbored identical CuR encoding ICEs, indicating that *Pseudomonas syringae* pv. *actinidiae* had acquired those elements from the co-occurring species which had already evolved resistance to copper sprays (Colombi et al. 2017).

CONCLUSION

The first sequenced bacterial phytopathogen genome was *Xylella fastidiosa* 9a5c, isolated in 1992 from Valencia orange in São Paulo, Brazil displaying symptoms of citrus variegated chlorosis (Simpson et al. 2000). Published in 2000, this genome was the product of the combined efforts of 116 people in 34 different labs and research centers. Twenty years later, a single individual can now sequence, assemble, and annotate a bacterial genome in a matter of hours in the lab or even the field using a portable sequencing device. The accessibility of sequencing and proliferation of well-supported and freely available analytical tools will deepen our understanding of pathogen evolution and host-microbe interactions in diverse contexts. Though outside the scope of this review, exciting developments in the study of host-associated microbiomes will provide a fuller understanding of the community context of disease emergence. Commensals are known to impact pathogen growth, either directly or via the modulation of host immunity (Innerebner et al. 2011; Straub et al. 2018), and infections may be caused by pathogenic consortia rather than single lineages (Bartoli et al. 2018). The analysis of plant-associated microbiomes can also illuminate the hidden lives of mobile elements outside focal pathogen populations, revealing how their spread impacts microbial communities and ultimately plant health. Given the pace with which the field develops, it is difficult to predict what the next 20 years will hold, but our hope is that they will deepen our understanding of microbial evolution and plant-microbe interactions while contributing to long-term agricultural resilience.

ACKNOWLEDGMENTS

We thank Ian Lorenzo Quibod and Ricardo Oliva for sharing data published in Quibod et al. (2020).

LITERATURE CITED

- Abrahamian, P., Timilsina, S., Minsavage, G. V., Kc, S., Goss, E. M., Jones, J. B., and Vallad, G. E. 2018. The Type III effector *AvrBsT* enhances *Xanthomonas perforans* fitness in field-grown tomato. *Phytopathology* 108: 1355-1362.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Balloux, F. 2010. Demographic influences on bacterial population structure. In: *Bacterial Population Genetics in Infectious Disease*. D. A. Robinson, D. Falush, and E. J. Feil, eds. Wiley, Hoboken.

- Baltrus, D. A. 2018. What is a *Pseudomonas syringae* population? Pages 99-121 in: Population Genomics: Microorganisms. M. Polz and O. Rajora, eds. Springer, Cham.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.
- Barash, I., and Manulis-Sasson, S. 2009. Recent evolution of bacterial pathogens: The gall-forming *Pantoea agglomerans* case. *Annu. Rev. Phytopathol.* 47:133-152.
- Bartoli, C., Frachon, L., Barret, M., Rigal, M., Huard-Chauveau, C., Mayjonade, B., Zanchetta, C., Bouchez, O., Roby, D., Carrere, S., and Roux, F. 2018. In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. *ISME J.* 12:2024-2038.
- Cesbron, S., Briand, M., Essakhi, S., Gironde, S., Boureau, T., Manceau, C., Fischer-Le Saux, M., and Jacques, M.-A. 2015. Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* 6:1126.
- Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., and Corander, J. 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30:1224-1228.
- Collins, C., and Didelot, X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Comput. Biol.* 14:e1005958.
- Colombi, E., Straub, C., Künzel, S., Templeton, M. D., McCann, H. C., and Rainey, P. B. 2017. Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* through acquisition of integrative conjugative elements and plasmids. *Environ. Microbiol.* 19:819-832.
- Corander, J., Croucher, N. J., Harris, S. R., Lees, J. A., and Tonkin-Hill, G. 2019. Bacterial population genomics. Pages 997-1020 in: Handbook of Statistical Genomics. D. Balding, I. Moltke, and J. Marioni, eds. Wiley, Hoboken, NJ.
- Didelot, X., and Wilson, D. J. 2015. ClonalFrameML: Efficient inference of recombination in whole bacterial genomes. *PLOS Comput. Biol.* 11: e1004041-e18.
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969-1973.
- Duchêne, S., Holt, K. E., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D. J., Fourment, M., and Holmes, E. C. 2016. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* 2:e000094.
- Earle, S. G., Wu, C.-H., Charlesworth, J., Stoesser, N., Gordon, N. C., Walker, T. M., et al. 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat. Microbiol.* 1:16041.
- Farhat, M. R., Shapiro, B. J., Kieser, K. J., Sultana, R., Jacobson, K. R., Victor, T. C., et al. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 45: 1183-1189.
- Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907.
- Greenlon, A., Chang, P. L., Damte, Z. M., Muleta, A., Carrasquilla-Garcia, N., Kim, D., et al. 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc. Natl. Acad. Sci.* 116:15200-15209.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307-321.
- Hanage, W. P. 2016. Not so simple after all: Bacteria, their population genetics, and recombination. *Cold Spring Harb. Perspect. Biol.* 8:a018069.
- Hulin, M. T., Armitage, A. D., Vicente, J. G., Holub, E. B., Baxter, L., Bates, H. J., Mansfield, J. W., Jackson, R. W., and Harrison, R. J. 2018. Comparative genomics of *Pseudomonas syringae* reveals convergent gene gain and loss associated with specialization onto cherry (*Prunus avium*). *New Phytol.* 219:672-696.
- Innerebner, G., Knief, C., and Vorholt, J. A. 2011. Protection of *Arabidopsis thaliana* against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. *Appl. Environ. Microbiol.* 77: 3202-3210.
- Jibrin, M. O., Potnis, N., Timilsina, S., Minsavage, G. V., Vallad, G. E., Roberts, P. D., Jones, J. B., and Goss, E. M. 2018. Genomic inference of recombination-mediated evolution in *Xanthomonas euvesicatoria* and *X. perforans*. *Appl. Environ. Microbiol.* 84:e00136-e18.
- Kachroo, P., Erasó, J. M., Beres, S. B., Olsen, R. J., Zhu, L., Nasser, W., et al. 2019. Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nat. Genet.* 51:548-559.
- Klemm, E. J., Wong, V. K., and Dougan, G. 2018. Emergence of dominant multidrug-resistant bacterial clades: Lessons from history and whole-genome sequencing. *Proc. Natl. Acad. Sci.* 115:12872-12877.
- Laflamme, B., Dillon, M. M., Martel, A., Almeida, R. N. D., Desveaux, D., and Guttman, D. S. 2020. The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science* 367:763-768.
- Lang, J. M., Pérez-Quintero, A. L., Koebnik, R., DuCharme, E., Sarra, S., Doucoure, H., et al. 2019. A pathovar of *Xanthomonas oryzae* infecting wild grasses provides insight into the evolution of pathogenicity in rice agroecosystems. *Front. Plant Sci.* 10:507.
- Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Lawson, D. J., Hellenenthal, G., Myers, S., and Falush, D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet.* 8:e1002453.
- Lees, J. A., Vehkala, M., Välimäki, N., Harris, S. R., Chewapreecha, C., Croucher, N. J., et al. 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* 7: 12797.
- Li, H., and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589-595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Li, X., Wang, H., Tong, W., Feng, L., Wang, L., Rahman, S. U., Wei, G., and Tao, S. 2020. Exploring the evolutionary dynamics of *Rhizobium* plasmids through bipartite network analysis. *Environ. Microbiol.* 22:934-951.
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M., and Spratt, B. G. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci.* 95:3140-3145.
- McCann, H. C. 2020. Skirmish or war: The emergence of agricultural plant pathogens. *Curr. Opin. Plant Biol.* 56:147-152.
- McCann, H. C., Li, L., Liu, Y., Li, D., Pan, H., Zhong, C., Rikkersink, E. H. A., Templeton, M. D., Straub, C., Colombi, E., Rainey, P. B., and Huang, H. 2017. Origin and evolution of the kiwifruit canker pandemic. *Genome Biol. Evol.* 9:932-944.
- McCann, H. C., Rikkersink, E. H. A., Bertels, F., Fiers, M., Lu, A., Rees-George, J., 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.
- McDonald, B. A., and Stukenbrock, E. H. 2016. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. B.* 371:20160026.
- Monteil, C. L., Cai, R., Liu, H., Llongtop, M. E. M., Leman, S., Studholme, D. J., Morris, C. E., and Binatzer, B. A. 2013. Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytol.* 199:800-811.
- Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Méric, G., Swingle, B., Morris, C. E., Vinatzer, B. A., and Sheppard, S. K. 2016. Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas syringae* pathogens. *Microb. Genom.* 2:e000089.
- Mutreja, A., Kim, D. W., Thomson, N. R., Connor, T. R., Lee, J. H., Kariuki, S., et al. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477:462-465.
- Newberry, E. A., Ebrahim, M., Timilsina, S., Zlatković, N., Obradović, A., Bull, C. T., Goss, E. M., Huguet-Tapia, J. C., Paret, M. L., Jones, J. B., and Potnis, N. 2019. Inference of convergent gene acquisition among *Pseudomonas syringae* strains isolated from watermelon, cantaloupe, and squash. *Front. Microbiol.* 10:270.
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A., and Parkhill, J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691-3693.
- Perez-Quintero, A. L., Ortiz-Castro, M., Lang, J. M., Rieux, A., Wu, G., Liu, S., Chapman, T. A., Chang, C., Ziegler, J., Peng, Z., White, F. F., Plazas, M. C., Leach, J. E., and Broders, K. 2020. Genomic acquisitions in emerging populations of *Xanthomonas vasculorum* pv. *vasculorum* infecting corn in the United States and Argentina. *Phytopathology* 110:1161-1173.
- Prokhorchik, M., Pandey, A., Moon, H., Kim, W., Jeon, H., Jung, G., Jayaraman, J., Poole, S., Segonzac, C., Sohn, K. H., and McCann, H. 2020. Host adaptation and microbial competition drive *Ralstonia solanacearum* phylogeny I evolution in South Korea. *Microb. Genom.* doi.org/10.1099/mgen.0.000461
- Quibod, I. L., Atienza-Grande, G., Oreiro, E. G., Palmos, D., Nguyen, M. H., Coronejo, S. T., Aung, E. E., Nugroho, C., Roman-Reyna, V., Burgos, M. R., Capistrano, P., Dossa, S. G., Onaga, G., Saloma, C., Vera Cruz, C., and Oliva, R. 2020. The green revolution shaped the population structure of the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. *ISME J.* 14:492-505.

- Rambaut, A., Lam, T. T., Carvalho, L. M., and Pybus, O. G. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vev007.
- Remigi, P., Zhu, J., Young, J. P. W., and Masson-Boivin, C. 2016. Symbiosis within symbiosis: Evolving nitrogen-fixing legume symbionts. *Trends Microbiol.* 24:63-75.
- Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. 2019. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3:430-439.
- Savory, E. A., Fuller, S. L., Weisberg, A. J., Thomas, W. J., Gordon, M. I., Stevens, D. M., Creason, A. L., Belcher, M. S., Serdani, M., Wiseman, M. S., Grunwald, N. J., Putnam, M. L., and Chang, J. H. 2017. Evolutionary transitions between beneficial and phytopathogenic *Rhodococcus* challenge disease management. *eLife* 6:e30925.
- Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patanã, J., Martins, J., Minsavage, G. V., Dahlbeck, D., Akhunova, A., Almeida, N., Vallad, G. E., Barak, J. D., White, F. F., Miller, S. A., Ritchie, D., Goss, E., Bart, R. S., Setubal, J. A. C., Jones, J. B., and Staskawicz, B. J. 2015. Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:208-217.
- Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.
- Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., Ramirez, M., and Carrico, J. A. 2018. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 4:e000166.
- Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bové, J. M., Briones, M. R. S., Bueno, M. R. P., Camargo, A. A., Camargo, L. E. A., Carraro, D. M., Carrer, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C. R., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A. P., Ferreira, A. J. S., Ferreira, V. C. A., Ferro, J. A., Fraga, J. S., França, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H. S., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C. C., Lemos, E. G. M., Lemos, M. V. F., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M. B. N., Madeira, H. M. F., Marino, C. L., Marques, M. V., Martins, E. A. L., Martins, E. M. F., Matsukuma, A. Y., Menck, C. F. M., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L. T. O., Netto, L. E. S., Nhani, A., Jr., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A. G., Pereira, H. A., Jr., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa, V. E., Jr., de Sá, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C. R., da Silva, A. M., da Silva, F. R., Silva, W. A., da Silveira, J. F., Silvestri, M. L. Z., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshakho, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J., and Setubal, J. C. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406:151-157.
- Souvorov, A., Agarwala, R., and Lipman, D. J. 2018. SKESA: Strategic k-mer extension for scrupulous assemblies. *Genome Biol.* 19:153.
- Spratt, B. G., and Maiden, M. C. J. 1999. Bacterial population genetics, evolution and epidemiology. *Philos. Trans. R. Soc. B.* 354:701-710.
- Stamatakis, A. 2014. RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Straub, C., Colombi, E., Li, L., Huang, H., Templeton, M. D., McCann, H. C., and Rainey, P. B. 2018. The ecological genetics of *Pseudomonas syringae* from kiwifruit leaves. *Environ. Microbiol.* 20:2066-2084.
- Sundin, G. W., and Wang, N. 2018. Antibiotic resistance in plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* 56:161-180.
- Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathirana, P., Iruegas-Bocardo, F., White, F. F., Goss, E. M., and Jones, J. B. 2020. *Xanthomonas* diversity, virulence and plant-pathogen interactions. *Nat. Rev. Microbiol.* 8:415-427.
- To, T.-H., Jung, M., Lycett, S., and Gascuel, O. 2016. Fast dating using least-squares criteria and algorithms. *Syst. Biol.* 65:82-97.
- Vanhove, M., Sicard, A., Ezennia, J., Leviten, N., and Almeida, R. P. P. 2020. Population structure and adaptation of a bacterial pathogen in California grapevines. *Environ. Microbiol.* 22:2625-2638.
- Wang, H., Hutton, S. F., Robbins, M. D., Sim, S.-C., Scott, J. W., Yang, W., Jones, J. B., and Francis, D. M. 2011. Molecular mapping of hypersensitive resistance from tomato 'Hawaii 7981' to *Xanthomonas perforans* race T3. *Phytopathology* 101:1217-1223.
- Wang, M., Roux, F., Bartoli, C., Huard-Chauveau, C., Meyer, C., Lee, H., Roby, D., McPeck, M. S., and Bergelson, J. 2018. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc. Natl. Acad. Sci. USA* 115:E5440-E5449.
- Weisberg, A. J., Davis, E. W., Tabima, J., Belcher, M. S., Miller, M., Kuo, C. H., Loper, J. E., Grünwald, N. J., Putnam, M. L., and Chang, J. H. 2020. Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* 368(6495).
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* 13:e1005595.