Interaction between climatic variation and pathogen diversity shape endemic disease dynamics in the agricultural settings

Rishi Bhandari*[1], Amanpreet Kaur*[1], Ivory Russell[1], Ogonnaya Michael Romanus[2], Destiny Brokaw[1], Anthony P Keinath[3], Zachary Snipes[4], Philip Rollins[4], Nettie Baugher[5], Inga Meadows[5], Ty Nicholas Torrance[6], Bhabesh Dutta[7], Edward Sikora[1], Roberto Molinari[2], Samuel Soubeyrand[8], Neha Potnis[1]

Affiliations

[1] Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA

[2] Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA

[3] Department of Plant and Environmental Sciences, Clemson University, Charleston, SC 29414, USA

[4] Clemson Cooperative Extension Service, Clemson University, Charleston, SC, 29401, USA

[5] NC State Cooperative Extension, Mountain Research Station, North Carolina State University, Waynesville, NC 28786, USA

[6] UGA Cooperative Extension, Tift County Extension, University of Georgia, Tifton, GA, 31793, USA

[7] Department of Plant Pathology, University of Georgia, Tifton CAES Campus, Tifton, GA, 31793, USA

[8] INRAE, BioSP, 84914, Avignon, France

*These authors contributed equally to this work.

Short title: Climate - Pathogen diversity interaction shape disease dynamics

Corresponding Author: Neha Potnis, nzp0024@auburn.edu

Abstract

Recurring outbreaks caused by endemic pathogens continue to pose problems in managing plant, wildlife, livestock and human health. Understanding how these outbreaks unfold and what drives the variability in disease epidemics across space and time is less understood, especially in the agricultural settings. In this study, we investigated the contribution of pathogen genetic diversity, climatic variation and their interaction towards disease dynamics, with an integrative approach grounded on multitype, high resolution sequencing data and analysis techniques. This investigation was carried out for bacterial spot disease epidemic by surveying tomato fields for bacterial pathogen (*Xanthomonas perforans*) across southeastern US over a span of three years. The strength of epidemic severity varied across space and time in the agricultural fields. Disease severity was positively associated with strain diversity, and was linked to environmental fluctuations, specifically, large variation and extreme changes in certain climatic factors. Strain-resolved metagenomics approach revealed that co-existence of multiple pathogen lineages was common in the individual fields, although accompanied by differential lineage dynamics. The co-occurring lineages displayed environmentally dependent fitness contributions. By tracing allelic frequencies in pathogen populations across temporal scales, we find evidence for asynchronous allele cycling across seasons, hinting at the presence of adaptive single-nucleotide polymorphisms (SNPs) being polymorphic in space in response to seasonality. Despite this pathogen heterogeneity, we identified positively selected loci under parallel evolution, which may explain the nature of selection pressures experienced by the pathogen. While single pathogen lineage is assumed to dominate the host in the agricultural settings, our findings challenge this notion by demonstrating genetic diversity in the pathogen population observed within a single field and linking it to the disease dynamics. Our results explain the role of pathogen genetic diversity, climate-dependent

compositional dynamics, and differential fitness contributions in dictating the variability of disease epidemics in the agricultural settings. Such findings will be invaluable for building predictive models in disease epidemiology. Our high-resolution combinatorial approach exploiting high-resolution sequence data, metadata types and analysis tools, is general enough to finely investigate disease epidemics at large scales in diverse case-studies concerning plant, animal and human health.

Introduction

Predicting the disease epidemics in the face of climate change is a major challenge in epidemiological studies of diseases impacting human, wildlife and plant health (1–4). Disease epidemics result from complex host-pathogen-environment interactions where ecological and evolutionary processes shape host and pathogen populations and environment can directly impact pathogen fitness or indirectly impact host susceptibility (5–9). These complexities involving interdependent interactions pose a challenge in dissecting the contribution of each component towards variability in disease epidemics. In addition, while controlled experiments have provided insights into specific environmental parameters in shaping disease outbreaks, we have little experimental evidence on how frequent shifts in these climatic factors can impact disease dynamics and whether climate change has already led to a surge of plant disease outbreaks over the past century (10–15).

Among different factors contributing to variable incidences and trajectories of disease epidemics, pathogen diversity plays a major role in this equation. While molecular markers and next-

generation sequencing approaches have accelerated characterization of pathogen diversity, the diversity studies have been focused on informing pathogen population structure and evolution of the pathogen (16–19). Integral role of population genetics into epidemiology and disease management has been much appreciated with the rapid next-generation sequencing efforts (20,21). Time-series isolate-genome sequencing has transformed our views on population genetics going beyond conventional diagnostic tools informing races, pathotypes, or species (17), although isolation-based approaches present a bias of choosing the most dominant pathogen genotype for sequencing and thus, may suppose pathogen homogeneity in a diseased sample, leaving out the details of low abundant pathogen variants that may be co-occurring. Such approaches may miss the "true genetic diversity" existing within single field, or a single organ, or a single lesion. In case of agricultural systems, majority of the focus has been on the dominant pathogen genotypes for designing management strategies, specifically breeding efforts, and characterization of pathogen spatial genetic structure studies at the micro-geographical scale are limited. Two recent studies have indicated presence of polyclonal infections in a single disease lesion (22), and epidemics with multiple co-existing pathogen genotypes (23) in a single field. These findings underscore the need for characterizing spatio-temporal pathogen population structure and the strain dynamics at the micro-geographical scale as disease epidemics unfold (24).

While much of the work in understanding the disease dynamics and pathogen diversity across space and time has been conducted in the natural settings with wild plant pathogens (25,26), studies linking pathogen diversity and dynamics to explaining disease epidemics in the agricultural settings are limited (27–29). Many factors associated with modern agricultural practices such as global trade, use of chemicals, or intensified monoculture system can predispose the system susceptible to the disease epidemics. For example, in the case of seed-borne pathogens, the global

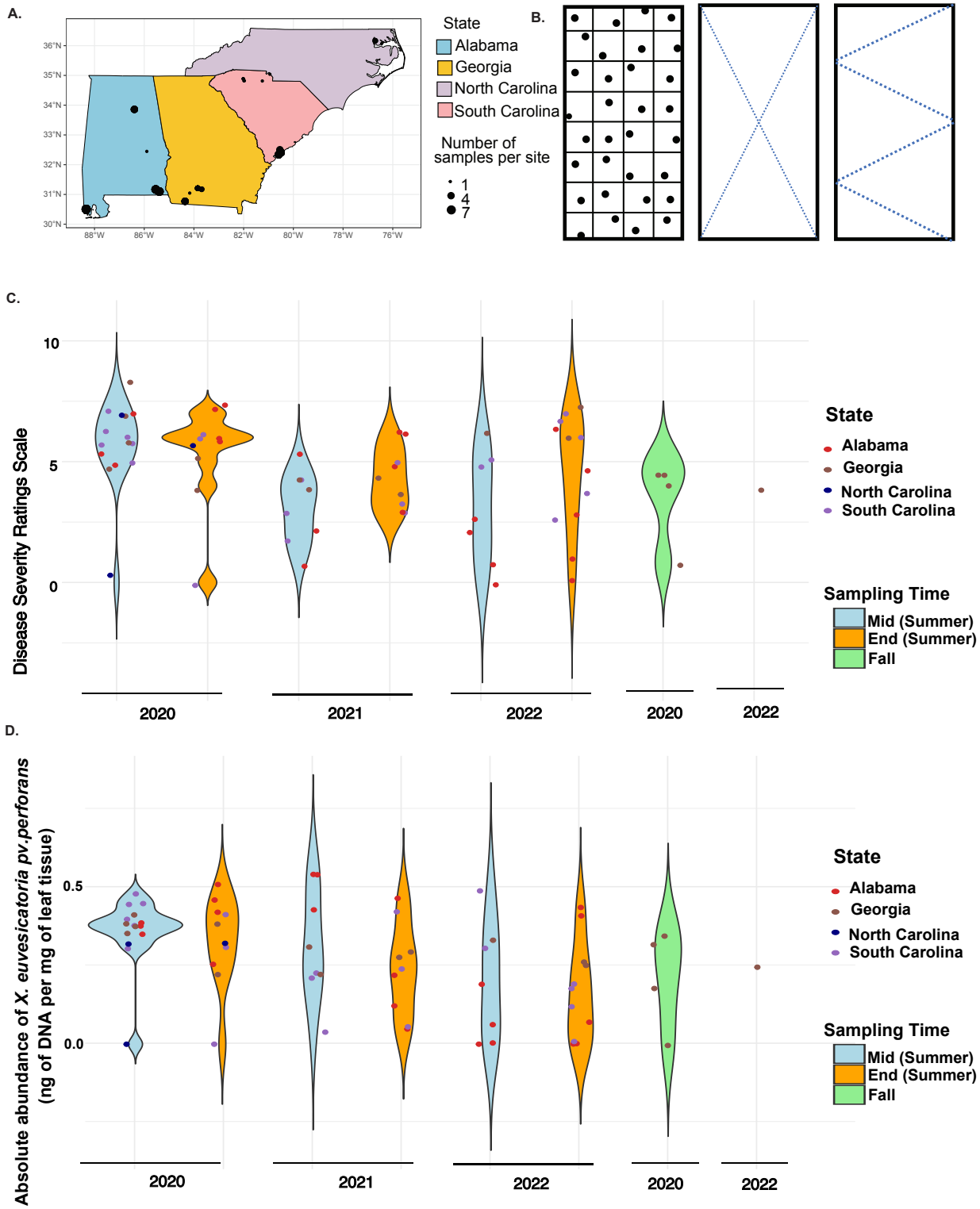
95 trade facilitating the continuous input of different pathogen genotypes on seeds or plant materials
96 circulated worldwide has been a driving force in the pathogen diversity observed under field
97 conditions (30). Thus, the initial strain composition of pathogen population entering the
98 agricultural fields can be variable, depending on the source of the inoculum on seeds, or
99 transplants, or surviving population from the previous epidemic season in the fields (31–33). Upon
100 establishment of these initial pathogen populations, variation in the pathogen dynamics across
101 different fields depends on many factors. These include differential fitness levels of pathogen
102 genotypes on the host considering both virulence and aggressiveness, interactions between
103 pathogen and host genotype (host specificity), strain competition, climate sensitivity of individual
104 pathogen genotypes, and other processes such as dispersal that may introduce new pathogen
105 genotypes on to the plant during the course of the season (34–39). In addition, local conditions
106 such as neighboring plants, may affect the strain assembly processes (40). Grower's practices such
107 as use of copper sprays, or other chemical pesticides can also impact strain composition,
108 facilitating dominance of chemically resistant population (41–43), while managing fitness trade-
109 offs with virulence. The field landscape can influence pathogen population structure (44). In
110 addition to the ecological dynamics, pathogen populations also undergo diversification through a
111 variety of genetic processes such as mutation or recombination, although studies characterizing
112 these genetic processes within a span of single or consecutive growing seasons are limited (39,45).
113 Higher genetic diversity in the pathogen population has been hypothesized to reflect higher
114 evolutionary potential (46), to facilitate genetic exchange (22) and was shown to contribute to
115 epidemic growth (24) and in some cases, higher genetic diversity was shown to affect transmission
116 dynamics (26).

117 The pathosystem of bacterial leaf spot (BLS) on tomatoes caused by *Xanthomonas perforans (Xp)*
118 (also referred to as *Xanthomonas euvesicatoria pv. perforans*) is an example of an endemic disease
119 where no known host resistance genes have been deployed in commercial tomato cultivars and
120 disease epidemics continue to be observed across all the tomato producing regions (47). Genome
121 sequencing of time-series collection of field isolates have informed that extensive shifts in the
122 pathogen population, including species displacement and extensive diversification have been
123 observed in the past decades (23,31,33,47–50). Tomato production system that introduces multiple
124 pathogen genotypes into the fields via transplants has been shown to shape pathogen population
125 structure in the field (31). In addition, inoculum from the previous growing season may also be
126 additional source of introduction, depending on the cultural practices in the field. However,
127 assuming variable probabilities of introductions of pathogen genotypes into the field, the following
128 questions still remains unanswered: Does a single pathogen genotype dominate the agricultural
129 fields as it is assumed for implementing management strategies? What factors explain variation in
130 disease dynamics observed across the fields? How does pathogen respond to abrupt changes in the
131 climatic patterns? We addressed these questions in the present study by specifically focusing on
132 the contribution of climatic variation, pathogen diversity and strain dynamics in explaining
133 different bacterial spot epidemic trajectories in the agricultural field outbreaks. Using strain-
134 resolved metagenomics to characterize pathogen diversity within a single field (23), we surveyed
135 tomato fields in the southeastern US over the course of growing season in three consecutive years
136 and found that multiple (up to 5) pathogen genotypes co-exist and their abundance varies across
137 the fields. While at least three different pathogen genotypes seemed to dominate the southeastern
138 US, their differential abundance across fields and across seasons was explained by differential
139 impact of abrupt climatic changes on the fitness of the three genotypes. The findings from this

140 work on strain dynamics and fitness will be useful in diversity-informed epidemiological
141 predictions, especially under changing climate.
142
143 **Results**
144 **Varying levels of BLS disease severity exist within the Southeastern US**
145 A total of 66 tomato phyllosphere samples were gathered over three years (2020-2022) from
146 Alabama, Georgia, North Carolina, and South Carolina (Fig 1A; S1A Table). The sampling
147 method was determined based on the field size and the amount of disease present (Fig 1B).
148 Metagenomic DNA from the leaf samples was subjected to shotgun metagenomics to obtain a
149 high-resolution pathogen population structure in the southeastern US. Apart from five samples
150 showing no BLS symptoms, the sampled fields showed variable levels of disease severity (S1B
151 Table) across consistently sampled individual fields (S1 Fig), across the season and within and
152 across states (Figs 1C and S2A). We observed that average disease severity in samples from
153 Alabama was lower during mid-season compared to other states. But the average disease severity
154 during the end season was higher in Alabama compared to other states (Figs 1C and S2A).
155 However, disease severity was not consistently higher as the season progressed, as seen with
156 samples from Georgia and South Carolina (Figs 1C and S2A). We used the SNV profiling method,
157 strainEST, to detect pathogen lineages at finer resolution including those with low abundance
158 levels. For this, we screened metagenomic reads for the presence of BLS pathogenic species, *Xp*,
159 *Xeu (Xanthmonas euvesicatoria pv. euvesicatoria)*, and *Xeu*-related pathovars (*Xeu* sister clades).
160 We observed the absence of *Xeu* in sampled tomato fields suggesting that *Xp* is the dominant
161 pathogen of BLS in tomatoes in the southeastern United States. Using correlation analysis as an
162 exploratory tool, we found that the absolute abundance of *Xp* (Figs 1D and S2B) positively

correlated with disease severity (mid-season: $R^2 = 0.381$, $p < 0.001$, end-season: $R^2 = 0.283$, $p < 0.01$). A similar observation was made with the relative abundance of *Xp* (S2C Fig) (mid-season: $R^2 = 0.508$, $p < 0.001$, end-season: $R^2 = 0.378$, $p < 0.001$). The pathogen population was significantly associated with disease severity during the early season compared to the late season based on R^2 value. Although small-scale farms showed slightly higher disease severity, this difference compared to commercial-scale farms was not statistically significant ($p = 0.45$, see S3A Fig).

170


A.
36°N
35°N
34°N
33°N
32°N
31°N
30°N
88°W
86°W
84°W
82°W
80°W
78°W
76°W
State
Alabama
Georgia
North Carolina
South Carolina
Number of samples per site
1
4
7
B.
C.
Disease Severity Ratings Scale
10
5
0
2020
2021
2022
2020
2022
State
Alabama
Georgia
North Carolina
South Carolina
Sampling Time
Mid (Summer)
End (Summer)
Fall
D.
Absolute abundance of X. euvesicatoria pv.perforans
(ng of DNA per mg of leaf tissue)
0.5
0.0
2020
2021
2022
2020
2022
State
Alabama
Georgia
North Carolina
South Carolina
Sampling Time
Mid (Summer)
End (Summer)
Fall

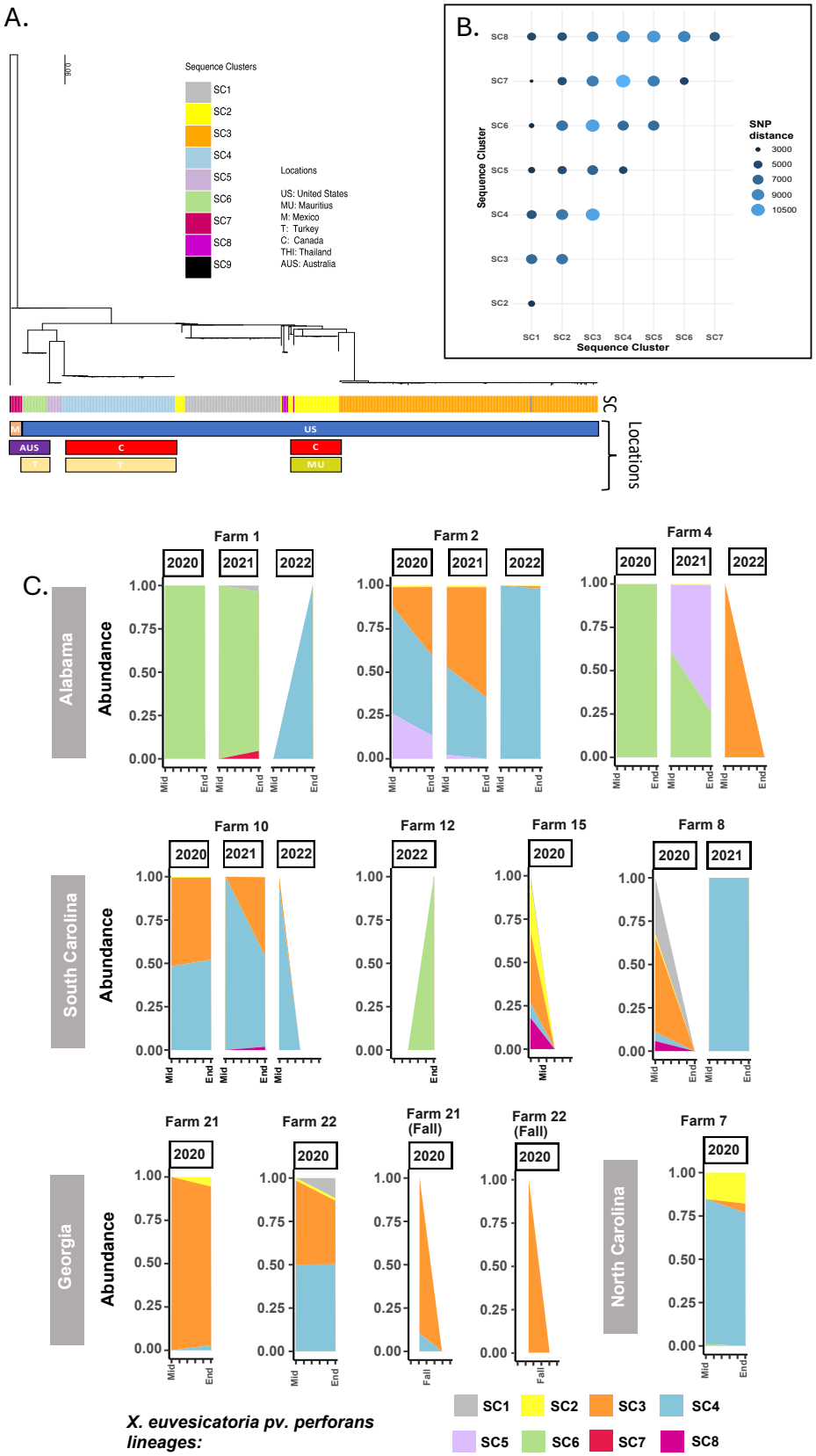
171

Fig 1. *Xanthomonas euvesicatoria* pv. *perforans* (*Xp*) continues to be a significant pathogen in the tomato fields across *the* southeastern US, although with variable disease severity in the region. (A) Map of the study area from Alabama, North Carolina, South Carolina, and Georgia from Southeast United States. The dots in the figures represent the sampled farm, and the size of the dot represents the number of samples collected during the years 2020, 2021, and 2022. The samples were collected from two different growing seasons: summer (with mid and end-season sampling points) and fall season. **(B)** Sampling methods varied by field size. In small farms, stratified random sampling was used to collect representative leaflets (left). In large fields, samples were taken along an X (mid) or W (right) pattern. Split-violin plots showing **(C)** disease severity ratings for all collected samples estimated using the Horsfall-Barratt scale, which ranges from 1 to 12, with the scale of 1 being no disease and 12 being 100% defoliation (51) and **(D)** the absolute abundance of *Xp* collected from different states during 2020, 2021, and 2022. The absolute abundance for each sample was estimated by multiplying the amount of DNA per ng of the sample with relative abundance of *Xp* present in it.

Xanthomonas abundance reveals a heterogeneous spatial distribution of pathogen lineages, and the increased pathogen diversity was associated with higher disease severity.

Given our observation of varying levels of epidemics, we hypothesized that pathogen heterogeneity is responsible for varying disease dynamics. Thus, we assessed the heterogeneity in the pathogen population by studying the extent of diversity of *Xp* in individual fields. Our approach involved mapping the metagenome reads to the SNP matrices of known pathogen lineages or sequence clusters (SC) of *Xp* and calculating relative abundance for each SC. To understand the current global population structure of *Xp*, we leveraged the availability of extensive genome resources for *Xp* collected worldwide over the last three decades (473 genomes as of March 2023). Based on the phylogeny of these publicly available global *Xp* collections, determined by core genome single nucleotide polymorphisms (SNPs), eight distinct lineages were identified within *Xp* strains referred to hereafter as sequence clusters (SC) or lineages (Fig 2A). This SNP-level resolution into different lineages was a baseline for our analysis to identify diversity and co-occurrence of pathogen lineages across field metagenomes. Surprisingly, we identified the presence of eight lineages in the field samples. Distantly related SC3 and 4 (SNP distance of 10,825 SNPs) co-existed in majority of the fields (Fig 2B). Three lineages, SC3, SC4, and SC6 were in higher abundance in these fields, although their abundance varied from field to field, and their

203 dynamics varied across neighboring fields (Figs 2C and S4). Notably, SC1, SC7, and SC8 were in
204 lower abundance, appearing in only eight samples (Figs 2C and S4). The co-occurrence of different
205 *Xp* lineages were observed in the samples, as noted in our previous limited metagenome survey
206 (23). Fields had a minimum of 1 to a maximum of 5 lineages present at a given sampling time, and
207 this diversity of co-occurring lineages was different across different years, and samples from
208 neighboring states (Figs 2C and S4). Next, we investigated whether multiple co-occurring lineages
209 of *Xp* within individual fields influenced disease severity. Fields with more than two co-occurring
210 SCs displayed higher average nucleotide diversity, and showed significantly higher disease
211 severity compared to those with less than 2 SCs (S5 Fig). Correlation analysis between the
212 Shannon diversity index of the pathogen population and disease severity revealed a positive
213 association ($R^2 = 0.52$, $p < 0.001$) (S6 Fig). This suggests that *Xp* population diversity is positively
214 associated with disease severity. Changes in strain composition were not only evident across
215 different fields but also within individual fields throughout the growing season, spanning three
216 years. These dynamics in strain composition were marked by strain fluctuations, where the
217 abundance of lineages shifted from one season to the other, strain invasion where a new lineage
218 likely infected the plants during the growing season, and differential strain dynamics (Figs 2C and
219 S4). Overall, the diversity of *Xp* found globally is circulating in Southeastern US tomato fields,
220 and strain diversity and composition is significantly associated with disease severity.


A.
Sequence Clusters
SC1
SC2
SC3
SC4
SC5
SC6
SC7
SC8
SC9
Locations
US: United States
MU: Mauritius
M: Mexico
T: Turkey
C: Canada
THI: Thailand
AUS: Australia
SC
Locations
B.
Sequence Cluster
SNP distance
3000
5000
7000
9000
10500
C.
Alabama
South Carolina
Georgia
North Carolina
Abundance
Farm 1
Farm 2
Farm 4
Farm 10
Farm 12
Farm 15
Farm 8
Farm 21
Farm 22
Farm 21 (Fall)
Farm 22 (Fall)
Farm 7
2020
2021
2022
Mid
End
Fall
X. euvesicatoria pv. perforans lineages:
SC1
SC2
SC3
SC4
SC5
SC6
SC7
SC8

221

Fig 2. Co-occurrence of multiple lineages of the BLS pathogen observed in the southeastern United States, which leads to increased BLS disease severity. (A) Midpoint-rooted maximum-likelihood phylogeny based on a concatenated alignment of core genome SNPs from 473 *Xp* strains. The tips are color-coded according to the lineages identified in the first level of the HierBAPS hierarchy. SC9 is a single-strain lineage exclusively found in Nigeria. **(B)** Dot plot shows the SNP differences across different SCs, where size and colour of each point refer to the number of SNPs differences present while comparing one SC (on x-axis) comparison to another SC (on y-axis). **(C)** Stacked bar plot depicting the co-occurrence of multiple *Xp* lineages, spatial and temporal variations, the introduction of new lineages, turnover, and dominance shifts in individual fields across various states during the mid and end of summer seasons and fall seasons for the years 2020, 2021, and 2022. Representative fields are shown in this figure. **S4 Fig** is showing data for all fields is included in the supplementary file.

Climatic shifts and extremes explain variation in BLS epidemics across the Southeastern United States

Given the findings of variable disease dynamics across neighboring states, we assessed the contribution of climatic factors and their spatial and temporal variations as drivers of disease epidemics. We used regression models to analyze how disease severity and pathogen abundance (response variables) are related to predictor variables such as climatic factors, sampling time, year, and farm scale (treating the latter three as categorical variables). While the data on disease severity, pathogen abundance, and pathogen diversity were collected for specific timepoints during the growing season, the datasets obtained from NASA-POWER for weather related variables (S2 Table) were at a much finer spatio-temporal resolution with daily record (0.5 x 0.625 latitude-longitude grid cell resolution). In addition to considering routine mean or median values of climatic variables for analyses, we also considered characteristics of climatic predictors using indexes that capture the spread (standard deviation), asymmetries, and tail-heaviness (skewness and kurtosis). This approach accounts for within-season climatic variations, considering interseasonal weather events, that may present highly conducive conditions for bacterial diseases. Considering the sample size, this approach, however, considerably increased the number of predictors so, to achieve reasonable statistical conclusions, for all regressions we ran a preliminary predictor-selection method (i.e. Lasso) and successively applied the appropriate regression on the selected

252 predictors for each response (see method section for Statistical analysis). Having selected the
253 relevant predictors for each response, we ran the logistic ordinal regression model to understand
254 drivers for disease severity which indicated that the Shannon diversity of various SCs of *Xp* present
255 in the field (t value = 4.15, $p < 0.05$) strongly influenced BLS disease severity as suggested in the
256 earlier result (S6A1 Fig). Among the climatic factors, the standard deviation of Wet Bulb
257 Temperature at 2 meters (t-value = 2.06, $p < 0.05$) and the standard deviation of the average of the
258 wind direction at 10 meters above the surface of the earth (t-value = 2.73, $p < 0.05$) significantly
259 increase BLS disease severity (Tables 1A and S3A), indicating that large variation in wet bulb
260 temperature (adiabetic saturation temperature) and wind direction values led to higher disease
261 severity. Next, a beta regression model predicting the drivers of *Xp* abundance identified one key
262 predictor: the standard deviation of wind direction at 10 meters above the earth's surface. This
263 variable has a positive association with the absolute (z-value = 2.08, $p < 0.05$)) (Tables 1B and
264 S3B) and relative (z-value = 2.69, $p < 0.05$) abundance of *Xp* (S3C Table). Additionally, kurtosis
265 and skewness of certain climatic parameters were observed as significant, suggesting extreme and
266 frequent shifts of climatic variables influence disease severity and pathogen abundance (S3 Table).
267 The complete analysis providing information about other significant climatic variables to explain
268 disease severity and pathogen abundance is available in the supplementary text (Supplementary
269 Text).
270
271 Finally, we used Dirichlet compositional regression to evaluate the influence of climatic factors in
272 explaining the differential dynamics of pathogen lineages or SCs across different fields. This type
273 of regression ensures that dependence between cluster abundance is accounted for (i.e. the
274 proportion of clusters must sum to one). We found that the variables that appeared to be significant

across pathogen lineages were the standard deviation of specific humidity at 2 meters and mid-season (compared to end-season). In particular, these appeared to diminish abundance across all clusters except for SC6 (S4 Table). These findings suggest that frequent shifts in the climatic conditions can influence individual lineages and their abundance. Specific variables that impact individual lineages are discussed in detail in the supplementary text (Supplementary Text).

Table 1. Various climatic factors influence disease dynamics and *Xp* abundance. (A) Significant climatic variables from the ordinal regression model predict the degree of correlation between various environmental factors and BLS disease severity. **(B)** Significant climatic variables from the beta regression model, showing the degree of correlation between various environmental variables and the absolute abundance of *Xp.* The Value represents the estimated increase in the log odds of the outcome for a one-unit increase in the predictor variable. The Std. Error measures the precision of this estimate. The t/z-value indicates that the effect of the predictor variable is statistically significant if the values exceed 2 or fall below -2.

Table 1A			
Factors	**Value**	**Std. Error**	**t-value**
Shannon diversity	2.37409	0.57175	4.1523
Standard deviation of Wet Bulb Temperature at 2 Meters	0.96797	0.46843	2.0664
Standard deviation of Wind Direction at 10 Meters	0.09547	0.03492	2.7338
Skewness of Clear Sky Surface PAR Total	-0.94701	0.36778	-2.5749
Kurtosis of Temperature at 2 Meters Range (C)	-1.25559	0.59055	-2.1262

Table 1B				
Factors	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
Standard deviation of Wind Direction at 10 Meters	0.03934	0.01885	2.087	0.0368
Skewness of Surface Pressure	-1.70069	0.52146	-3.261	0.0011
Kurtosis of Relative humidity at 2 Meters (%)	-0.66357	0.26293	-2.524	0.0116

Pathogen heterogeneity and variable fitness levels influence disease dynamics.

We observed differential strain dynamics with spatial and temporal variations in dominant lineages (Fig 2C). The variation in strain dynamics can result from variability in fitness levels of different pathogen lineages and/or variable effects of environment on host-pathogen genotype interactions. This variation in fitness levels can stem from differential virulence and transmission capabilities

294 or direct or indirect (host-mediated) competition among co-occurring pathogen lineages. We asked
295 how pathogen heterogeneity and variable dynamics could contribute to the explanation of
296 fluctuating disease pressures during outbreaks. We used complementary data from the greenhouse
297 experiment and field data provided here to evaluate the hypothesis that different disease pressures
298 are caused by different fitness levels of co-existing pathogen lineages. The greenhouse experiment
299 involved infecting tomato plants with individual strains representative of different lineages or in
300 mixed infections. Overall, mixed infections led to higher disease severity than individual infections
301 and individual lineages differed in their disease severity levels (S7 Fig). For example, disease
302 severity level of mixed infection of SC3 and SC4 was significantly higher than that of individual
303 infections (S7B Fig).
304 Next, we used field metagenome data with the relative abundance of pathogen lineages and disease
305 severity data to assess pathogen fitness under the field condition. We used StrainRanking, a
306 regression-based method for ranking of pathogen strains with respect to their contribution to
307 epidemics. This allowed us to combine pathogen genetic data with epidemiological monitoring of
308 disease progression to investigate differences among pathogen lineages in their performance in the
309 field. The fitness of an SC is the contribution of SC to the variation in the observed disease severity
310 between the two sampling points, for example, mid and end season in this study, and is measured
311 by a parameter estimated with the StrainRanking approach. A positive value for this parameter
312 indicates that corresponding SC leads to an increase in the disease severity, and a negative value
313 corresponds to a decrease of the disease severity. Lineages SC4 and SC6 tend to have a higher
314 fitness than SC3 and other low frequency SCs with respect to their contribution to the field
315 epidemics, as indicated by their positive contribution towards disease severity from mid to end
316 season in the theoretical epidemic curves (Fig 3A). Equal weighted means of the genetic

317 frequencies observed during the mid- and end-seasons (p=0.5 for each) were taken into
318 consideration for this analysis. Even considering the weightage of genetic frequencies high (p=0.9)
319 at the mid-season (S8A Fig) or at the end of the season (p=0.10) (S9A Fig), SC6 and SC4
320 contributed significantly in terms of fitness towards the growth of disease severity. Because SC6
321 was not predominant lineage across our samples, so we did not consider SC6 for further analysis.
322 When strain interactions were evaluated for their fitness contributions, lineages SC4 (>20%
323 abundance) and SC3 (≤20% abundance) combination showed positive fitness, whereas interaction
324 corresponding to SC4 (≤20% abundance) whatever SC3 frequency led to negative fitness
325 contribution. When SC4 (>20% abundance) and SC3 (>20%) coexisted, we observed near zero
326 fitness contribution (Figs 3B, S8B and S9B). Next, we evaluated fitness levels of individual SCs
327 when strain-environment interactions were considered. It's interesting to note that we found that
328 various contextual factors significantly impacted SC4 fitness levels but less in SC3. When average
329 surface pressure and standard deviation of wind speed were less than mean values, but average
330 temperature and skewness of wet bulb temperature were more than mean values, SC4
331 demonstrated a larger positive fitness contribution (Fig 3C). We found that specific environmental
332 conditions that are observed in the middle or end of the season have different effects on the fitness
333 contributions of SC3 and SC4, particularly when taking into account genetic frequencies that are
334 reported at the end of the season (p=0.10) (S9C Fig). While lower values of the parameters promote
335 better fitness of SC3, higher values of the clear sky PAR standard deviation, kurtosis of surface
336 pressure, and standard deviation of temperature at 2 meters support higher fitness of SC4.
337 Conversely, lower levels of average clear sky PAR favor SC4 fitness (S9C Fig), whereas higher
338 values enhance SC3 fitness. However, when genetic frequencies which are reported in the mid of

the season were taken into account, SC4 was more fit in the across different environmental parameters than SC3 (S8C Fig).

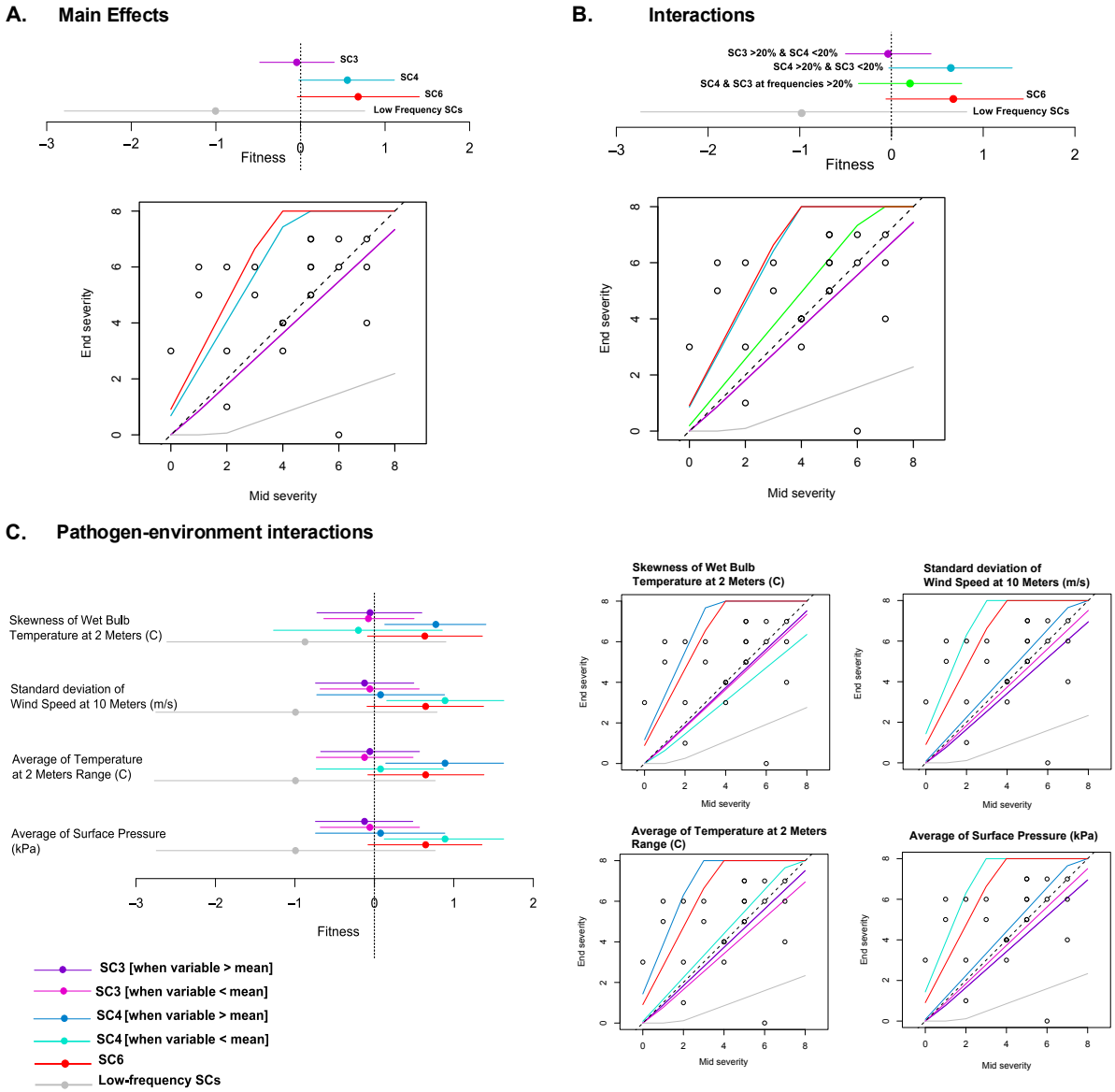


Fig 3. Different lineages have variable fitness levels and resilience towards different climatic factors and competition with co-occurring lineages. Using StrainRanking **(A)** Comparisons of different SCs with fitness estimates and their confidence intervals under main effects were made. These estimates focused on SC3, SC4, and SC6 (which were more abundant in samples), and a group of SCs (SC1, SC2, SC5, SC7, and SC8) that were categorized as low frequency (due to their lower relative abundance, as estimated in Fig. 2C). The bottom plot illustrates the growth curves of these SCs, estimating their contributions to the progression of disease severity from mid and end of the summer season. **(B)** Pathogen-pathogen interactions, specifically between the two dominant SCs (SC3 and SC4), were analyzed based on their combined frequencies being either below or above 20%. Fitness estimates with confidence intervals were

compared when the summed frequencies of SC3 and SC4 were taken into account. Different interaction scenarios were considered, such as SC3 > 20% & SC4 < 20%, SC4 > 20% & SC3 < 20%, and both SC3 & SC4 at equal frequencies, along with SC6 and the low-frequency SCs. The growth curves in the bottom plot highlight how these interactions contributed to the progression of disease during the season. **(C)** Pathogen-environment interactions were examined by applying the frequencies of SC3 or SC4 to different environmental variables, depending on whether each variable was above or below its mean at the sampling point. The analysis assumed an equal-weighted mean of genetic frequencies to calculate fitness levels: p x mid-frequency + (1 - p) x end-frequency, where (p = 0.5) in this figure. All the colour keys are specific in each section of the figure and a similar colour scheme is followed for their corresponding growth curve plots.

Tracing the allelic frequencies across seasons suggested presence of conditional SNPs that oscillate across temporal scales. While pathogen diversity analysis based on globally known pathogen lineages revealed strain dynamics driven by climatic factors and variable strain-fitness levels, we next investigated the extent of genetic polymorphism in the pathogen population (without having a bias of known pathogen lineages) by tracing allelic frequencies within and across the fields. For this analysis, a pangenome reference was constructed for *X. perforans* and allelic frequencies for different sites within the genome were obtained by mapping metagenomic reads to the pangenome. These allelic frequencies were monitored within and across seasons and fields to study the maintenance of genetic polymorphism and allele frequency dynamics. The influence of season and year on pathogen diversity led us to identify variants selected across pathogen populations over time and seasons. We looked for alleles whose frequencies changed within and across seasons and parallel across fields. For this analysis, we grouped all samples from each respective sampling time (i.e., summer or fall of a specific year) into a single metapopulation (Fig 4A). We then tracked the major and minor allele frequencies across three sampling points over three years. We observed a strong influence of seasonality on the allelic frequencies. Allelic frequencies at most sites differed across the mid/end of the summer and fall seasons. For example, several of the alleles (with f >=0.8) found in the middle or end of the summer growing season

381 showed a significant change in frequency (with $f <= 0.2$) during the fall seasons for both the years
382 of 2020 and 2022 (Fig 4B). Similarly, the alleles with their frequencies less than 0.8 during mid-
383 2020 remained in the population at intermediate frequencies during end seasons or mid-seasons in
384 the subsequent years but changed to either 1 or 0 during the fall seasons of the 2020 and 2022
385 (Figs 4C and S10). These polymorphisms that undergo dramatic rapid shifts in allelic frequency
386 and oscillate asynchronously between seasons repeatedly over subsequent years can be defined as
387 conditional or seasonal SNPs. These seasonal SNPs would most likely be adaptively responding
388 to selection pressures that vary across seasons.
389 Given the influence of seasonality on oscillating patterns of polymorphisms, we focused on
390 identifying alleles that were seen in parallel across at least 50% of farms with $f <= 0.2$ (called a
391 minor) during the first season and became $f >= 0.8$ (considered a major) in the following season
392 (Figs 4A and 4C). We found that 237 of the alleles that were minor/low-frequency alleles in mid-
393 2020 turned major (i.e., reached high frequency) during fall 2020, with 157 of those alleles
394 reappearing as major during fall 2022 but were not observed as high-frequency alleles in any other
395 season of the following years. Similarly, 1234 minor alleles that existed at the end of Summer
396 2020 reached high frequency (i.e., major alleles) in the fall of 2020. Of these, 138 alleles were
397 observed at high frequency in the subsequent fall season of 2022 (S11A Fig). Furthermore, we
398 observed 3598 allelic sites with allele frequency less than 0.2 during the fall of 2020 became major
399 during mid-season (summer) of the year 2021 and most of the sites (1025) appeared back as major
400 alleles in the mid-season of the year 2022 (S11B Fig).

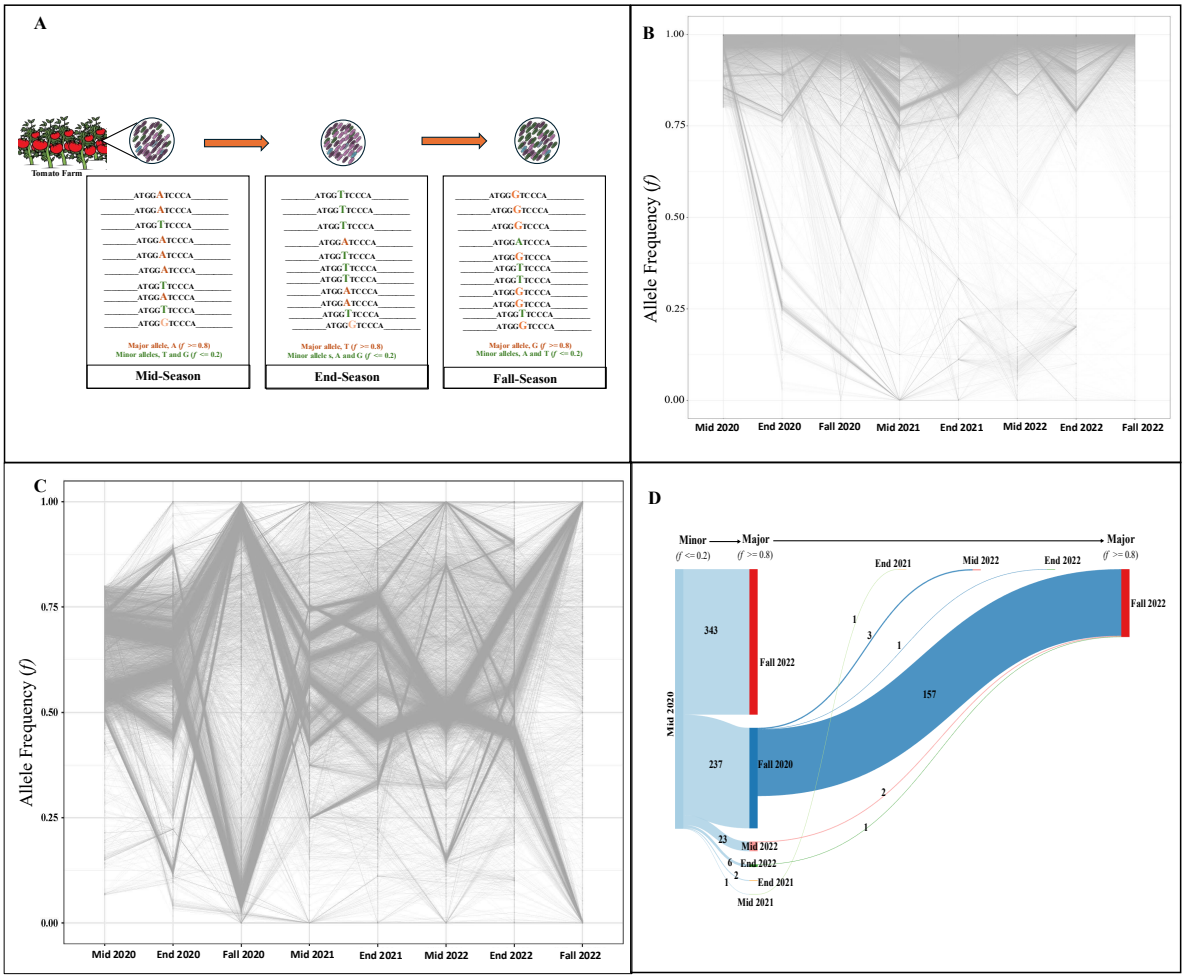


Fig 4. Seasonal oscillations of allelic frequencies with the signature of parallel evolution. (A) A graphical representation of seasonal variation, where during mid-season (of summer), allele A was the major allele (allele frequency > 0.8) in the population. By the end of the season (of summer), the A was replaced with T, and allele A became minor in population. However, by the fall season, allele G became a major allele, and the rest were present in low abundance in the pathogen population. The line plot shows alleles with their average frequencies of **(B)** $f >= 0.8$ & **(C)** $f < 0.8$ in the pathogen population during mid-season (of summer) of year 2020. The allele frequencies were tracked for different sites of *Xp* during the following seasons over three years. **(D)** The Sankey plot presenting the counts of those alleles which were present in parallel across at least 50% of the farms sampled during the mid-season (of summer) of 2020 with frequency <= 0.2 (with read depth of at least 10) and became major allele ($f >= 0.8$) in the population in the next season and counted those same alleles which stayed as major allele during the next seasons of following years.

Despite variability in disease dynamics, signatures of selection and parallel evolution were evident in the pathogen population.

418 Given the varying disease pressure observed across different farms, we sought to determine if
419 any genes in the pathogen genome were consistently under selection across these locations.
420 Using Metapop package, we estimated the pN/pS (non-synonymous to synonymous ratios) of
421 different genes of *Xp* and pooled out the genes with pN/pS > 1 values. We filtered out the genes
422 if they were present in at least 50 % of farms in a particular sampling point to provide evidence
423 for parallel evolution. We found eight genes under selection were parallel across farms (at least
424 50 % of farms of specific sampling time) and seasons. These genes comprised type VI secretion
425 system (T6SS) protein ImpG, ClpV1 family T6SS ATPase, S9 family peptidase,
426 aminotransferase, tonB-dependent receptors, S9 family proteins (S12 Fig and S5 Table).
427 However, none of these eight genes were clustered and are present at distinct locations based on
428 the reference genome of *Xp* LH3. Interestingly, two proteins associated with the T6SS were
429 under positive selection, suggesting the microbiome's potential role in selection pressure and
430 shaping the pathogen genome. We also identified a gene, class 1 fructose-bisphosphatase, under
431 positive selection only in the populations from both fall seasons of 2020 and 2022 (S5 Table).
432 Overall, these results suggest evidence of parallel evolution across farms and seasons in the *Xp*
433 pathogen population.
434

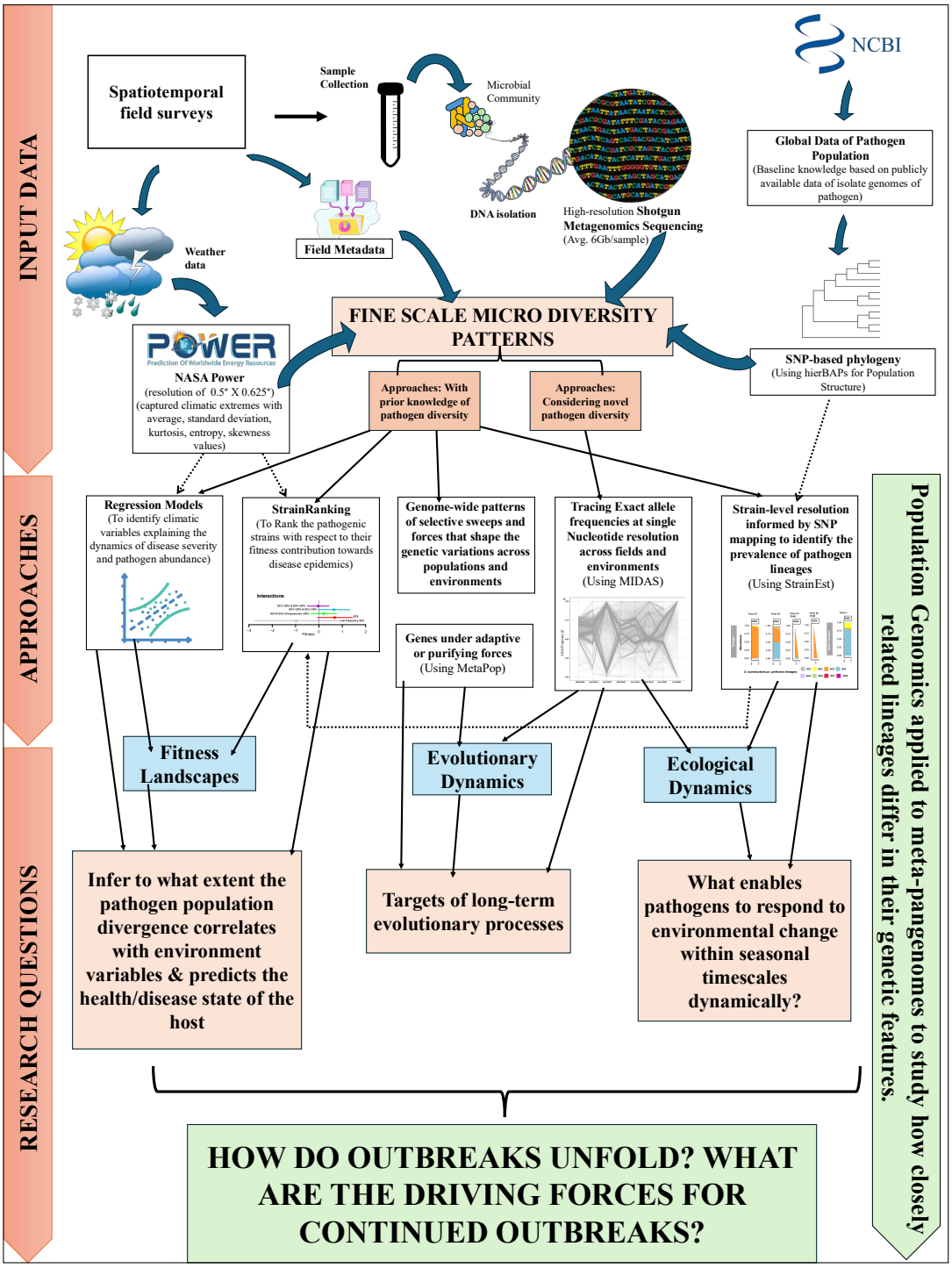


Fig 5. Integrated workflow to study pathogen outbreaks incorporating large scale field survey data, high-resolution metagenomics data and subsequent analyses to understand eco-evolutionary pathogen dynamics.

Discussion

The significance of strain diversity, strain fitness and compositional dynamics towards disease epidemics has been highlighted in wild plant pathogens but the importance of genetic polymorphism in explaining variability in disease outbreaks is less understood in the agricultural settings, partly due to an assumption of genetically monomorphic populations under monoculture system and the emphasis on the use of methods that lack resolution of pathogen population structure at microgeographic level. In this work, we developed an approach to study the extent of genetic polymorphism in the pathogen population within individual fields over the course of a growing season and over three years, linked the pathogen dynamics to climatic variables to explain variability of disease dynamics in the fields in neighboring states (Fig. 5). We found that as many as five pathogen lineages can co-exist within a single agricultural field. Prior studies have documented that under natural systems, no single lineage takes over and multiple lineages co-exist at low abundance but in contrast, single dominant lineage is assumed to dominate in the agricultural fields (25). We found that at least three distinct lineages dominate across different fields and at times, two distantly related lineages can co-exist within an individual field at high abundance throughout the growing season (Fig 2). The differential strain dynamics was observed across neighboring fields and was linked to differential climate sensitivity of the lineages and differential fitness contributions towards epidemics (Fig 3). When we conducted analysis of pathogen dynamics without having bias of known lineages, but by tracing the allelic variants over three years, the analysis revealed a considerable influence of seasonality on pathogen dynamics, with distinct seasonal alleles strongly favored throughout specific seasons in the consecutive years (Fig 4). This genetic heterogeneity with differential and environmentally dependent fitness contributions might be an adaptation strategy that enables pathogen to dynamically respond to

462 environmental fluctuations within seasonal timescales. These complex interdependent pathogen
463 genotype-genotype-environment interactions may complicate the selective pressures during a
464 host-pathogen arms race in case of endemic diseases in agricultural systems. Understanding the
465 roles of conditional seasonal alleles that we observed in this work and the role of loci under
466 selection in dictating the pathogen fitness represent an important avenue for future work.
467 Identifying the drivers of the disease dynamics and pathogen diversification will be critical in
468 reducing the frequency of pathogen outbreaks (Fig 5).

469 Regression models applied to link environmental data, disease outcomes and pathogen
470 diversity in this study revealed that extreme climatic events were significant drivers of disease
471 dynamics and pathogen diversity across different farms. More variations and extreme events such
472 as changes in wind direction, photosynthetically active radiation, temperature, humidity, and
473 surface pressure (Tables 1 and S3) can directly influence pathogen dispersal (52–54) and plant
474 host physiology (55–57), or indirectly alter host-pathogen, strain-strain, and pathogen-microbiome
475 interactions (39,58,59). While empirical evidence for how recent climatic shifts have contributed
476 to disease outbreaks is limited, studies on the impact of extreme weather events on the virulence
477 of fungal pathogens like *Hemileia vastatrix* (coffee rust), *Puccinia striiformis* (wheat stripe rust),
478 and *Phytophthora cinnamomi* (Mediterranean oak decline) have highlighted the need for these
479 studies (60–62). Our approach of capturing the climatic extremes using measures such as standard
480 deviation, kurtosis, skewness of climatic variables observed for each farm across seasonal
481 timescales allowed us to assess the extent to which pathogen genetic divergence correlates with
482 climatic shifts. The findings from these models highlighted the importance of capturing these
483 climatic extremes in the epidemiological modeling, rather than average climatic variables. For
484 example, parameters that we identified as significant (negative impact on disease severity and

485 pathogen abundance) were more extreme values of temperature and humidity. Our observation of
486 more variation in wind direction being significantly associated with higher disease severity and
487 higher pathogen abundance align with previous observation of importance of dispersal in the
488 epidemic growth for bacterial spot disease. The aerosolization and long-distance dispersal of <=2m
489 was shown for this pathogen under humid transplant house conditions (63). Later a single effector,
490 XopJ2, was shown to provide increased fitness by increased dispersal in the field (64).

491 Our findings of co-existence also challenge the traditional view that a single lineage of an
492 abundant pathogen dominates the host population in agricultural systems. Two lineages, SC3 and
493 SC4 co-existed, both at large densities, but neither showed a strong dominance across different
494 fields in all three examined growing seasons (Figs 2C, S4). First, this observation is intriguing,
495 because it certainly challenges the competitive exclusion principle, which states that two species
496 occupying the same ecological niche cannot coexist indefinitely (65). Second, it reveals an
497 apparent paradox because, based on our analyses, SC4 seems to be fitter than SC3. SC6 also seems
498 to be as fit as SC4 but is present at a lower level with distribution limited to only certain fields in
499 Alabama than SC3 and SC4, thus, its contribution in the lineage dynamics could not be
500 systematically assessed. Here, we sought explanations for this apparently unexpected co-existence
501 of SC3 and SC4 and their ecological dynamics. Prior studies on *E. coli* have demonstrated the
502 prevalence and persistence of season-specific strains due to fluctuating temperatures throughout
503 the year (66–68). However, we have yet to observe any clear patterns of early or late-season
504 specialization of individual SCs, indicating that time partitioning is unlikely to explain their co-
505 existence. Since we did not sample outside the host, current data cannot explain whether these
506 lineages have high fitness outside the host or neighboring plant species (non-host). In addition,
507 lack of data on host genotypes or field landscapes do not allow to evaluate whether host

508 heterogeneity and landscape heterogeneity in the field supports their co-existence (44). Another
509 limitation of our approach which used pooled leaf samples representative of a field for
510 metagenome sequencing, was that we couldn't exclude spatial partitioning or microhabitat
511 difference (for example, leaves of different ages, or patchy distributions within individual fields)
512 that can support co-existence of two or more pathogen lineages. Resource specialization cannot be
513 ruled out either because the dominant lineages coexisting in the fields, SC3 and SC4, differ in type
514 III effector content and may differ in within-host multiplication rate and transmission rate, or they
515 may prefer different niches, such as leaf surface vs apoplast. Our finding that sudden seasonal
516 events or climatic fluctuations had a significant impact on defining individual lineages' fitness
517 contributions may potentially explain their coexistence. Two independent approaches assessing
518 compositional dynamics (S4 Table) and fitness dynamics highlighted the influence of
519 environmental parameters on individual lineages (Figs 3, S8 and S9), specifically SC3 and SC4,
520 specifically climatic fluctuations as indicated by skewness, kurtosis and standard deviation of
521 different parameters, being significant in dictating abundance of SC3 or SC4. We observed that
522 field-level fitness contribution by SC4 seemed to be impacted by other biotic (SC3) and abiotic
523 factors, whereas that by SC3 did not. The varying performance of SC4 depending on different
524 environmental conditions (from very fit to intermediately fit: fitness = 0) may contribute to the
525 similarity of its overall frequency with the SC3 frequency (more generalist with respect to the
526 values of tested factors and the presence of other SC). Thus, maintaining co-existence of lineages
527 that show overall environmental resilience as well as environmentally dependent fitness might be
528 a strategy that pathogen adopts to cope up with the sudden climatic shifts. All the arguments made
529 above that may contribute to explain the co-existence of SC3 and SC4 at relatively high and similar
530 levels, and the estimation of their relative fitness, can be summarized by making a parallel with

531 the effective reproduction number, Rt, classically considered in human and animal epidemiology.
532 Rt is the average number of new infections caused by a single infected individual at time t in the
533 partially susceptible population. Rt varies in time because of the variation of the number of
534 susceptible individuals but may also vary because of temporal heterogeneity in the eco-evo-
535 determinants of the transmission (e.g., because of lockdown measures during the covid-19
536 pandemic; (69)). Rt can obviously be considered at the lineage level. Broadly, the epidemic or the
537 lineage dynamics tends to keep going on if Rt ≥ 1. In terms of fitness, as estimated by the
538 StrainRanking analysis in this study, it is equivalent to a fitness value ≥ 0 (the fitness here is an
539 indicator that integrates diverse processes and variables and is therefore an aggregated
540 measurement of the capacity of the lineage to "multiply" in real conditions). Thus, SC4 with fitness
541 >0 and SC3 with fitness of around 0 indicate Rt ≥ 1, suggesting that both SC3 and SC4 are fit to
542 keep epidemic going or to expand. As Rt evolves in time, the fitness (viewed as an aggregated
543 measurement) may also evolve in time, and an interesting research area would be to finely
544 characterize the eventual variation in time of the SC's fitness, evaluation of lineage-specific genes
545 in contribution towards fitness, and to determine how the possible processes evoked above
546 contribute to this variation. This is because possibility of transitory state with co-existence as
547 opposed to equilibrium state cannot be ruled out, especially, given the limited geographical region
548 studied here (70) and that genetic divergence has been observed in this pathosystem (Potnis et al.
549 2021). Thus, continuing such high-resolution lineage fitness tracking across tomato producing
550 regions and disease outcomes will be important outputs to predict future outbreaks, and to design
551 disease management strategies that are tailored to concurrent pathogen population structure (i.e.
552 breeding for disease resistance and climate resilience).

553 Although seasonality was not evident at the individual lineage level, analyzing the data via
554 variant tracing indicated the presence of season-specific allelic pathogen variants (Fig 4),
555 particularly unique to the summer vs fall growing season over the three consecutive years of
556 sampling (Fig 5). These seasonal variants might be indicative of a trade-off between in-season
557 within-host multiplication, in-season transmission, and inter-season survival, where variants that
558 contribute as primary inoculum (the form under which pathogen survives host absence) in fall vs
559 summer might differ due to the variable inter-season survival periods experienced (71). One
560 approach to test this notion is to see if there is a trade-off between survival outside the host (with
561 varied overwintering durations) and infection efficiency (72). Periodic host absence has been
562 claimed to cause evolutionary divergence (also known as evolutionary branching), however,
563 contradictory evidence exists (73–75). The absence of year-round production in Alabama, South
564 Carolina, North Carolina, and Georgia may drive evolutionary divergence. The seasonality could
565 also mean differing day lengths altering host physiological responses to the pathogen, thus
566 impacting the fitness contributions of different variants towards epidemics (76) or differing light
567 intensities impacting pathogen behavior and virulence, as seen with differing motilities and
568 differing epiphytic fitness in the case of *Pseudomonas syringae* strains (77). The findings of
569 seasonal oscillations of genome-wide allelic frequencies may suggest evidence of fluctuation
570 selection contributing towards the genetic variation observed in this population.

571 We also looked at signatures of parallel evolution by identifying loci under positive
572 selection across > 50% of the fields (S5 Table). Notably, we identified certain genes under positive
573 selection over three years, including the TonB-dependent receptor gene located adjacent to the
574 PepSy domain-containing protein. The widespread use of heavy fertilizers and pesticides in
575 agricultural systems (such as mancozeb applications to manage bacterial spot contain Mn and Zn)

576 may have been responsible for the selection of this TonB-dependent receptor likely involved in
577 heavy metal transport (78,79). Furthermore, multiple genes belonging to two separate Type VI
578 secretion system clusters were found to be under selection, suggesting that the microbiome may
579 play a significant role in driving the selection within the pathogen genome (80–82).

580 Consistent with the assumptions of seedborne diseases and their global spread (30,83–85)
581 in the agricultural fields, our survey indicated that almost all known *X. perforans* lineages
582 identified in the global population are circulating in the southeastern US fields. Although multiple
583 lineages co-existed in the individual fields, no such apparent link between genetic relatedness of
584 co-occurring lineages and disease severity was observed, as was previously shown to be important
585 through the mechanism of competition or increased virulence (86). This could be because co-
586 occurrence of SC3 and SC4 was predominant in majority of the cases, with other low frequency
587 lineages not particularly contributing towards fitness and we lacked consistent sampling of the
588 fields throughout the growing season. However, higher diversity (Shannon as well as average
589 nucleotide diversity) was correlated with higher disease severity (S6 Fig). Such observation could
590 be explained by ability of diverse strains to manipulate the host defenses in a variety of ways
591 (87,88), or through resource-mediated competition or through sharing of public goods, such as
592 quorum sensing signals (89). While our survey collected information on host cultivars, the lack of
593 information on host genetics or levels of tolerance/susceptibility did not allow evaluation of the
594 role of host genotypes in driving disease dynamics. However, age-related host resistance (90–92)
595 or the prevalence of other diseases later in the season might contribute towards seasonal disease
596 pressure and thus, seasonal dynamics that we observed. The regression analysis also included scale
597 of the farm as one of variables. While disease severity was not significantly different across small
598 vs commercial farms, more pathogen diversity was observed in the commercial farms compared

599 to small farms (S3 Fig). This intriguing observation brings about apparent contradiction, i.e. small-
600 scale farms often have transplants, seeds from a variety of sources, thus, possibility of introduction
601 of multiple pathogen genotypes. However, crop/genotype diversity in the small-scale farms with
602 intercropping may also lead to reduction in pathogen abundance and diversity (93). These
603 observations demand further investigation of whether such intercropping might be a resilient
604 strategy for disease control.

605 Taken together, we find that plant-pathogen interactions in agricultural environments are
606 much more complex than previously thought. Our study revealed that maintaining pathogen
607 heterogeneity and variable fitness with seasonal allele fluctuations in response to extreme weather
608 events are the strategies employed by the pathogen. This work also presents empirical evidence on
609 the influence of recent climatic changes on altering disease outcomes. We suggest that breeding
610 programs need to consider multiple pathogen lineages that show positive fitness contributions
611 based on large scale field surveys. Tracing these evolutionary changes can predict which lineages
612 will be a better fit in the long run, especially in the face of sudden climatic shifts, aiding in
613 developing more effective breeding strategies. Future work based on the results of this study
614 should focus on identifying the roles of specific fitness factors in pathogen biology, as this could
615 provide valuable insights into pathogen adaptation strategies and the selection pressures they face.
616 Finally, we present an integrated workflow utilizing metapangenomics to infer fine scale
617 microdiversity patterns in the pathogen population, coupled with epidemiological and statistical
618 modeling to link climatic variables to the eco-evolutionary processes impacting disease risk and
619 to study disease epidemics at large scales (Fig 5). This approach can allow disentangling targets
620 of long-term eco-evolutionary processes and for studying fundamental principles involved in

621 response of microbes in dynamically responding to environmental shifts. Such approach can be
622 applied to different case studies in plant, animal, and human health.
623

624 **Materials and Methods**

625 **Reconstruction of the *X. perforans* core genome using publicly available *Xp* genomes from**
626 **NCBI**

627 Sequencing reads for 467 *Xp* strains downloaded from NCBI (as of 08/15/2023) were used
628 to study the global diversity of *Xp* isolate genomes. To ensure high-quality and minimally biased
629 genomes, the genomes were de-replicated using dRep (v3.2.2) (94) using a 95% minimum genome
630 completeness and 5% maximum contamination. We manually discarded the genomes with more
631 than 400 contigs to ensure better alignment with the reference genome and avoid bias while
632 creating the SNV profile. Four hundred sixty-seven quality-controlled genomes collected from
633 tomato and pepper (with one eggplant isolate) around the globe were individually aligned against
634 the completed genome of *Xp* strain LH3. Single-nucleotide polymorphisms (SNPs) common to all
635 genomes were extracted to generate a concatenated set of high-quality core-genome SNPs using
636 Parsnp (95). The phylogenetic tree was visualized and annotated using iTol (v.6.7.5) (96)

637 **Sample collection and disease severity estimation**

638 To study the diversity of BLS *Xanthomonas* in the southeastern United States, we collected
639 66 leaf samples from 22 fields, designated as either commercial-scale or small-scale (Table S1).
640 These samples were collected from symptomatic and asymptomatic tomato plants grown in
641 Alabama, Georgia, South Carolina, and North Carolina over 2020, 2021, and 2022 (Fig. 1A).
642 Samples were taken during mid-season, end-season of the summer growing season, and when

643 available, in the fall growing season (S1A Table) to capture the temporal dynamics of the pathogen
644 population. While sampling, two to three leaflets were randomly taken from individual plants in
645 the field using variable sampling methods based on the scale of the field (Fig. 2B). Multiple plants
646 were sampled across the field to obtain a representative 50-100 grams leaf tissue sample.
647 Assessments of the disease were made by estimating the percent of disease symptoms and
648 defoliation caused by bacterial spots (S1B Table) using the Horsfall-Barratt scale, which ranges
649 from 1 to 12, with scale of 1 being no disease and 12 being 100% defoliation (51).

650 **DNA extraction and sequencing**

651 Metagenomic DNA extraction, quantification, and sequencing of the samples were done
652 as described earlier (23). Approximately 40 grams of the leaf sample was taken in a Ziploc bag
653 (Ziploc®) with 50 ml 0.1% wash buffer (0.05M PBS, 8.5 g NaCl, and 0.2 ml tween 20 per liter of
654 water), sonicated for 20 minutes, followed by transfer of the buffer to a 50 ml falcon tube with the
655 help of a sterile plastic serological pipette. The tubes were centrifuged at 4000 rpm, 4°C for 20
656 minutes (Eppendorf® 5418 R) to collect all the cells. The cells were then washed twice with sterile
657 deionized water, and the metagenomic DNA extraction was done using Wizard® Genomic DNA
658 Purification Kit (Promega) as per the manufacturer's instructions with the addition of a phenol-
659 chloroform step to remove protein contamination. The extracted DNA was quantified using a
660 Qubit® fluorometer and dsDNA high-sensitivity assay kit (Life Technologies, Carlsbad, CA, USA)
661 and kept at -80^{0}C until further processing. Sequencing was performed on the Illumina HiSeq 4000
662 platform. The raw reads were processed and trimmed for the adapter using BBDuk (Bushnell,
663 2014). DynamicTrim command of SolexaQa was used to trim the reads based on quality values
664 with Qphred less than 20, followed by filtering the reads shorter than 50bp (97). Host reads were

665 separated and removed using tomato cv. Heinz 1706 as a reference (GCA_022405115.1) from the
666 metagenomic samples using KneadData tools (98) as a reference.

667 **Taxonomic profiling and diversity estimation**

668 The relative abundance of *Xanthomonas* sp. in quality-controlled and host-decontaminated
669 reads was performed using Kraken2 (v2.1.2) (99) against a standard Kraken2 standard database.
670 The database contained RefSeq libraries of archaeal, bacterial, human, and viral sequences as of
671 March 1, 2022 (100). The resulting Kraken2 report files were used as inputs for Bayesian re-
672 estimation of abundance using Kraken (Bracken) (v2.6.2) (101). We also calculated the absolute
673 abundance estimates for *Xp* by combining their relative abundance with the total DNA content in
674 each sample. Microbiota density, expressed as total DNA (ng) per mg of fresh sample, was
675 computed for each sample. This information was utilized to determine the absolute abundance of
676 *Xp* by multiplying the amount of DNA per mg of the sample with its relative abundance.

677 To explore the intraspecific diversity and dynamics of *Xp* in the metagenomic samples, we
678 employed the StrainEst pipeline (102). StrainEst is a reference-based method that utilizes single-
679 nucleotide variant (SNV) profiles from carefully selected reference genomes. This approach
680 enables us to estimate strains' presence and relative abundance within a given sample. Initially, we
681 conducted SNV profiling of representative strains from *Xp*, *Xeu*, and other closely related
682 pathovars belonging to *Xeu* species complex to determine their presence. A second round of
683 profiling was done to identify the diversity of *Xp* strains. A representative genome from each *Xp*
684 SC (Fig 2) was selected and aligned against the completed reference genome of *Xp* strain 91-118
685 using the MUMmer algorithm (103). This genome alignment was used to construct a SNV matrix,
686 where each row corresponded to a variable position in the reference genome, and columns

687 contained allelic variants present in the reference strains. This matrix was used as a reference in
688 modeling strain level abundance.

689 To assess the diversity of *Xp* SCs within the samples, we computed the Shannon diversity
690 index (Shannon-Wiener diversity index) using the "diversity()" function from the R package *vegan*
691 (v2.6-4) (104). The index considers the count of SCs present (richness) and their proportional
692 distribution (evenness). A higher index value indicates greater species diversity in the habitat.
693 When the Shannon diversity index of a sample is equal to 0, it implies that only a single SC is
694 detected in that sample.

695 **Weather data**

696 Daily climate data spanning three years for all 22 fields were compiled from the NASA
697 Power database using the nasapower (v4.0.12) package (105) in R. Data were collected during
698 mid-season, end-season of summer, and fall samples as per the sampling time in Table S1. The
699 gridded weather data used in this analysis are based on satellite observations and models and are
700 offered globally at a horizontal resolution of 0.5° X 0.625° latitude–longitude grid cell. Weather
701 variables, such as temperature, precipitation, solar radiation, relative humidity, wind, and others
702 (S2 Table), were consolidated over time and space. The intervening period between the planting
703 and sampling times was accounted for in the analysis for each weather parameter by calculating
704 averages. For instance, the mid-season average was determined by averaging daily weather data
705 from the planting date to the mid-season sampling time, which usually occurred 60 days after
706 transplanting. The end-season average was calculated from the planting date to the end-of-season
707 sampling time, typically 30 days after the mid-season sampling. To account for interseason
708 climatic events, variability among the weather parameters was measured using descriptive
709 statistics (standard deviation, skewness, entropy, and kurtosis) for all the climatic variables to see

710 if the extreme weather events during the growing season impact disease severity, pathogen
711 diversity, and abundance. Standard deviation measures the deviation from the mean value. A
712 higher standard deviation indicates greater weather parameter variability, while a lower standard
713 deviation suggests a more consistent pattern. Skewness measures the degree and direction of
714 asymmetry. A positive skewness implies more extreme high weather parameters, while a negative
715 skewness suggests more extreme low, with zero suggesting a symmetric distribution. Kurtosis is
716 a measure of tail extremity reflecting the presence of outliers in a distribution (106). Higher
717 positive kurtosis implies extreme weather values or outliers, while negative kurtosis indicates a
718 flatter distribution with less extreme weather values. Similarly, entropy quantifies the level of
719 unpredictability in the weather values. The higher the entropy, the more unpredictability in the
720 weather fluctuations will be.

721

722 **Statistical analysis via regression models**

723 We quantified the relative abundance of various *Xp* lineages across the samples and analyzed their
724 distribution patterns in relation to different climatic factors. This analysis was designed to explain
725 three different types of responses based on climatic and other variables: (i) relative abundance, (ii)
726 absolute abundance, and (iii) disease severity. Before running these regressions, we chose to run a
727 predictor-selection procedure for each of these (given the high number of predictors we considered
728 for this work) by using Least Absolute Shrinkage and Selection Operator (Lasso) regression, which
729 jointly aims at correctly predicting the response and selecting the most associated variables (107).
730 As no implemented or stable versions of the Lasso directly address the types of responses we have
731 considered in our work, we decided to use proxy responses to run the Lasso and then use the
732 selected variables from this step to run the specific regressions and inference described above.

More specifically, we use a logistic Lasso to select the variables for the (relative abundance) where we classify low and high abundance based on the 50% threshold. In contrast, we use standard Lasso regression on the severity scales (assuming they can be considered as continuous) and the Shannon entropy (a continuous scale proxy for the distribution of abundance between clusters). Predictor-selection was performed on the variables that still have non-zero coefficients after the shrinkage process using the "cv.glmnet" function to prevent the overfitting of the model. This study determined the tuning parameter for regularization using 10-fold cross-validation. The path length (min_lambda/max-lambda) was chosen to be 0.001, and 100 default values along the regularization path were tested to find the best lambda value. Once a selection was run for all responses (proxies) of interest, we ran the appropriate regressions for each response. More specifically, since disease severity is measured on a discrete ordered scale in the 0-12 range, we used an ordinal logistic regression, which accounts for the order of severity and the discrete scale of the response (108). On the other hand, as the abundance (both absolute and relative) of *Xp* is measured on a 0-1 range, we employed a Beta regression (109) using the *betareg* package (110). The beta regression model is employed for non-normally distributed data with a range between 0 and 1, addressing heteroskedasticity and asymmetry issues (109), and is specifically chosen here due to the non-normal distribution of *Xp* relative abundance in the sample, which conforms to a beta distribution. To analyze the distribution of abundance among different Xp lineages, we used a compositional regression approach, which ensures that the regression model explains and preserves the property that the total abundance among the clusters should sum to one (indeed, a separate regression for each cluster would not consider the dependence between clusters). We used the Dirichlet regression approach, using the *DirichletReg* package in R, to describe how the distribution of various *Xp* lineages changes according to changes in these potential predictors

756 (111). As for the previous response variables, the use of the Lasso for this type of regression is
757 computationally unstable. Therefore, we chose to use the Shannon diversity of *Xp* as the response
758 when using the Lasso: this is a reasonable proxy since Shannon diversity is related to how the data
759 is distributed (in this case, across the lineages). Once this was obtained, 18 variables were selected
760 to run the Dirichlet regression.

761 **Greenhouse Experiment**

762 We evaluated the effects of different SCs of *Xp* on 4-5 week old tomato cultivar FL47 plants grown
763 under greenhouse conditions. Plants were dip-inoculated for approximately 30 seconds in bacterial
764 suspensions containing 10^6cfu/ml of representative strains from different SCs (SC3 (Xp2010), SC4
765 (AL57), SC5 (AL37), and SC6 (AL65)). These SCs were chosen based on their observed
766 abundance in samples from our previous study (23). Treatments included a control (0.01M
767 $MgSO_4$), individual inoculations with SC3, SC4, SC5, and SC6, as well as mixed co-infections of
768 SC4 + SC3, SC5 + SC6, SC3 + SC6, SC3 + SC5, and SC3 + SC5 + SC6. All treatments were
769 supplemented with 0.00025% Swilvet and replicated five times, with plants randomly arranged in
770 the greenhouse. Plants were watered daily, and disease severity was assessed on days 7 and 14,
771 with additional evaluations on days 9 and 12 for plants exhibiting delayed symptoms. Disease
772 severity was measured using the Horsfall-Barratt scale (51). To minimize bias, two observers
773 recorded the disease severity, and the average score was used. Raw disease severity indices were
774 then used to calculate the area under the disease progress curve (AUDPC), which correlates
775 multiple observations of disease progression into a single value (112). AUDPC values were plotted
776 for each treatment across seven experimental batches using the "dplyr" package in R Studio and
777 built linear and linear mixed-effects models using "lme4" and "lmerTest" (113,114). For individual
778 strains a generalized linear hypothesis test was used while for multiple comparisons Tukey's

779 method was used. Linear mixed effect models were used to assess the development of disease
780 severity slopes across treatments.

781 **Strain Ranking**

782 StrainRanking (115) was used to estimate and compare fitness differences between SCs (main
783 effects), between SCs considered alone or in association (pathogen-pathogen interaction), and
784 between SCs under specific environmental conditions (pathogen-environment interaction).
785 StrainRanking requires input data consisting of the frequencies of the SCs and the measurement
786 of disease growth or decline for each sampling location (i.e., fields), and estimates the fitness as
787 the relative contributions of SCs to the growth or the decline of disease severity between two
788 sampling times (i.e., Mid and End sampling times). SC frequencies were computed as a weighted
789 mean of Mid and End SC frequencies: $p \times$ Mid frequencies $+ (1 - p) \times$ End frequencies, where
790 p is equal to 0.1, 0.5 or 0.9 to consider different representativeness of Mid and End frequencies in
791 terms of disease dynamics between Mid and End sampling times of summer season. (Because of
792 low sample size, we did not include Fall season samples in this analysis.) Significance tests
793 performed in StrainRanking are based on randomization, which adequately accounts for the
794 relatively small sample sizes encountered when distributing data at the resolution of SCs or even
795 at a higher resolution (when one considers pathogen-pathogen or pathogen-environment
796 interactions). To account for the specificity of the severity measurement in the present study, which
797 is a mark ranging between 0 and 8, we added to the StrainRanking method a saturation level of the
798 growth and the decline at +8 and -8, respectively. In addition, we developed a function providing
799 confidence intervals for estimated fitness in addition to the pairwise comparison tests embedded
800 in the original StrainRanking R package.

801 For the study of main effects, StrainRanking was applied to SC3, SC4, SC6 and the group of SCs
802 with low frequencies (SC1, 2, 5, 7, and 8), say SCLowFreq.

803 For the study of pathogen-pathogen interaction, only the interaction between the two dominant
804 SCs was accounted for, namely SC3 and SC4. We consider SC3xSC4 interaction depending on
805 whether SC3 and SC4 are lower or larger than 20%. Thus, StrainRanking was applied to the
806 summed frequency freq(SC3)+freq(SC4) when freq(SC3)>20% and freq(SC4)>20%, and to the
807 frequencies of SC3 when freq(SC3)≤20% or freq(SC4)≤20%, SC4 when freq(SC4)≤20% or
808 freq(SC3)≤20%, SC6 and SCLowFreq. Not that the interaction has to be understood at the field
809 level here since we do not have information indicating whether or not SC3 and SC4 were collected
810 on the same plants.

811 For the study of pathogen-environment interaction, only the interaction with the two dominant SCs
812 was accounted for. This step incorporated only the significant environmental parameters identified
813 through statistical analyses using the regression model, along with their standard deviation,
814 skewness, entropy, and kurtosis. Hence, for each factor X (e.g., average temperature, kurtosis of
815 wind speed, etc.), StrainRanking was applied to the frequencies of SC3factor+, i.e. the frequency
816 of SC3 when $X > \bar{X}(\text{SC3})$ (where $\bar{X}(\text{SC3})$ is the mean of X at sampling points where SC3 is present,
817 i.e. freq(SC3)>0), SC3factor- (when $X \leq \bar{X}(\text{SC3})$), SC4factor+ (when $X > \bar{X}(\text{SC3})$), SC4factor-
818 (when $X \leq \bar{X}(\text{SC4})$), SC6 and SCLowFreq. The value of X is the weighted mean of X observed at
819 Mid and End sampling times: $p \times [X \text{ at Mid point}] + (1 - p) \times [X \text{ at End point}]$, where p takes
820 the same value as for the weighted average of genetic frequencies.

821 **Construction of *Xanthomonas* pan-genome**

822 Initially, we downloaded 287 strains of *Xp (Xanthomonas euvesicatoria* pv. *perforans)* and *Xeu*
823 *(Xanthomonas euvesicatoria* pv. *euvesicatoria)* from the NCBI RefSeq database (as of February
824 16th, 2022). Using SuperPang (v0.9.4 beta1) (116), we constructed a non-redundant pangenome
825 as a reference for the subsequent analyses.
826 To avoid read stealing and donating of *Xanthomonas* with other microbiome species, we
827 implemented an additional filter to remove genes with more than 97% similarity in our
828 metagenomic reads. These genes were identified similarly to the method described by Kaur et al.,
829 2024. We used the "run_midas.py species" command to identify the predominant species in our
830 samples, utilizing the MIDAS (v1.3.2) default database as a reference, which consists of 31,007
831 bacterial reference genomes organized into 5,952 species groups (117). We then built a custom
832 database of non-redundant pangenomes (*Xp* and *Xeu*) using MIDAS and subjected it to BLAST
833 (blast+) with the concatenated gene sequences from species with a mean abundance > 0 from the
834 previous step. Genes with 97% similarity to those in the database were removed from all samples
835 using BBDuk (v37.36) (118). These reads were used for our subsequent analyses.
836
837 **Genes under selection**
838 We discarded the samples with zero *Xp* abundance in the following evolutionary analyses. To
839 calculate the genetic variation in our samples, we first mapped the samples to the non-redundant
840 pangenome using BWA-MEM (v0.7.12) (119). Next, we removed low-quality alignments and
841 duplicate reads using samtools (v1.11) (120) and picard (v1.79) (121), respectively. The resulting
842 bam files were then processed with MetaPop (v1.0) (122) to calculate the ratio of non-synonymous
843 to synonymous polymorphisms (pN/pS) for each gene. In order to find the genes under positive
844 selection, we chose the genes with pN/pS ratios >1 and present in at least 50% of the samples in a

845 particular season. Average nucleotide diversity values for each farm were also estimated using this
846 package.
847
848 **Seasonal variants**
849 To calculate the allele frequencies of SNV sites across the pangenome, we used the MIDAS
850 "midas.py snps" script (117). We treated all samples from different states in the southeastern USA
851 as a single population for each season. To increase the accuracy of results, we had additional filters
852 for removing SNV sites with a read depth of less than 10. For seasonal oscillations, we averaged
853 the allele frequencies of major alleles during the season. To identify parallel variants in the
854 population, we first extracted SNV sites where the frequency of minor alleles was less than 0.2
855 (considered minor alleles) and changed to greater than 0.8 (considered major alleles) in the
856 following season and remained major in subsequent seasons. For example, we identified sites with
857 minor alleles ($f <= 0.2$) during mid-season 2020 that became major alleles ($f >= 0.8$) by the end of
858 the 2020 season. We retained only those SNVs that exhibited this trend in at least 50% of the
859 samples in a given season.
860

861

862 **Acknowledgments**
863 The climatic data were obtained from the NASA Langley Research Center (LaRC) POWER
864 Project funded through the NASA Earth Science/Applied Science Program. This work was made
865 possible in part by a grant of high-performance computing resources and technical support from
866 the Alabama Supercomputer Authority. We thank growers in the southeastern US for supporting
867 this project, allowing sampling their fields, providing metadata and communicating with us their

concerns in tomato production. This helped us better understand the nature of the outbreaks. We thank the Alabama Agricultural Experiment Station, and Auburn University for providing resources for conducting this work.

Data availability

Sequence data generated from this work have been deposited in the SRA (Sequencing Read Achieve) database under the BioProject accessions number PRJNA1129844. All other data and code used in this study are publicly and freely available through GitHub (https://github.com/Potnislab/pathogendiversity).

Funding

This work was funded by the Foundation for Food and Agricultural Research (FFAR) New Innovator award to NP (NIA19 0050), USDA-NIFA hatch project funding to NP (7005375), and BEYOND project (Grant ANR-20-PCPA-0002; French National Research Agency (ANR)) to SS.

Author contributions

NP conceptualized and designed the study. RB, AK, and IR processed the samples. RB, AK, and IR conducted analyses; specifically, RB and IR conducted ecological dynamics analysis. RB and IR, OR, and RM conducted statistical modeling analyses. AK worked on building pangenome and removal of blacklisted genes and the evolutionary analyses. DB conducted the greenhouse experiment. SS conducted analysis of pathogen fitness contributions towards epidemics using StrainRanking. RB, AK, and NP worked on data interpretation. APK, ZS, PR, IM, TT, BD, and

ES helped collect samples and metadata from different tomato farms across four states. RB, AK, and NP wrote the manuscript with contributions from all the authors.

References:

1. King KC, Hall MD, Wolinska J. Infectious disease ecology and evolution in a changing world. Philos Trans R Soc B Biol Sci. 2023 Feb 6;378(1873):20220002.

2. Campbell AM, Hauton C, Baker-Austin C, van Aerle R, Martinez-Urtaza J. An integrated eco-evolutionary framework to predict population-level responses of climate-sensitive pathogens. Curr Opin Biotechnol. 2023 Apr 1;80:102898.

3. Singh BK, Delgado-Baquerizo M, Egidi E, Guirado E, Leach JE, Liu H, et al. Climate change impacts on plant pathogens, food security and paths forward. Nat Rev Microbiol. 2023 Oct;21(10):640–56.

4. Pfenning-Butterworth A, Buckley LB, Drake JM, Farner JE, Farrell MJ, Gehman ALM, et al. Interconnecting global threats: climate change, biodiversity loss, and infectious diseases. Lancet Planet Health. 2024 Apr 1;8(4):e270–83.

5. Engering A, Hogerwerf L, Slingenbergh J. Pathogen–host–environment interplay and disease emergence. Emerg Microbes Infect. 2013 Feb 6;2(2):e5.

6. Burdon J, Thrall P, Ericson L. Genes, communities & invasive species: understanding the ecological and evolutionary dynamics of host–pathogen interactions. Curr Opin Plant Biol. 2013 Aug 1;16(4):400–5.

7. Parratt SR, Numminen E, Laine AL. Infectious Disease Dynamics in Heterogeneous Landscapes. Annu Rev Ecol Evol Syst. 2016 Nov 1;47(Volume 47, 2016):283–306.

8. Höckerstedt L, Numminen E, Ashby B, Boots M, Norberg A, Laine AL. Spatially structured eco-evolutionary dynamics in a host-pathogen interaction render isolated populations vulnerable to disease. Nat Commun. 2022 Oct 13;13(1):6018.

9. Barrat-Charlaix P, Neher RA. Eco-evolutionary dynamics of adapting pathogens and host immunity. eLife [Internet]. 2024 Oct 18 [cited 2024 Oct 20];13. Available from: https://elifesciences.org/reviewed-preprints/97350

10. Koelle K, Pascual M, Yunus M. Pathogen adaptation to seasonal forcing and climate change. Proc Biol Sci. 2005 May 7;272(1566):971–7.

11. Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE. Climate Change Effects on Plant Disease: Genomes to Ecosystems. Annu Rev Phytopathol. 2006 Sep 1;44(1):489–509.

923 12. Lafferty KD. The ecology of climate change and infectious diseases. Ecology.
924 2009;90(4):888–900.

925 13. Chakraborty S. Migrate or evolve: options for plant pathogens under climate change. Glob
926 Change Biol. 2013;19(7):1985–2000.

927 14. Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD. Climate change and infectious
928 diseases: from evidence to a predictive framework. Science. 2013 Aug 2;341(6145):514–9.

929 15. Velásquez AC, Castroverde CDM, He SY. Plant-Pathogen Warfare under Changing Climate
930 Conditions. Curr Biol CB. 2018;28(10):R619–34.

931 16. Sarkar SF, Guttman DS. Evolution of the core genome of Pseudomonas syringae, a highly
932 clonal, endemic plant pathogen. Appl Environ Microbiol. 2004 Apr;70(4):1999–2012.

933 17. Vinatzer BA, Monteil CL, Clarke CR. Harnessing population genomics to understand how
934 bacterial pathogens emerge, adapt to crop hosts, and disseminate. Annu Rev Phytopathol.
935 2014 Aug 4;52(1):19–43.

936 18. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic
937 analysis of diversity, population structure, virulence, and antimicrobial resistance in
938 Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci. 2015 Jul
939 7;112(27):E3574–81.

940 19. Garfinkel AR, Coats KP, Sherry DL, Chastagner GA. Genetic analysis reveals unprecedented
941 diversity of a globally-important plant pathogenic genus. Sci Rep. 2019 Apr 30;9(1):6671.

942 20. Milgroom MG. Population biology of plant pathogens : genetics, ecology, and evolution
943 [Internet]. APS Press,; 2015 [cited 2019 Oct 3]. 399 p. Available from:
944 http://www.sidalc.net/cgi-
945 bin/wxis.exe/?IsisScript=earth.xis&method=post&formato=2&cantidad=1&expresion=mfn
946 =040539

947 21. Everhart S, Gambhir N, Stam R. Population Genomics of Filamentous Plant Pathogens—A
948 Brief Overview of Research Questions, Approaches, and Pitfalls. Phytopathology®. 2021
949 Jan;111(1):12–22.

950 22. Pruvost O, Boyer K, Ravigné V, Richard D, Vernière C. Deciphering how plant pathogenic
951 bacteria disperse and meet: Molecular epidemiology of Xanthomonas citri pv. citri at
952 microgeographic scales in a tropical area of Asiatic citrus canker endemicity. Evol Appl.
953 2019;12(8):1523–38.

954 23. Newberry E, Bhandari R, Kemble J, Sikora E, Potnis N. Genome-resolved metagenomics to
955 study co-occurrence patterns and intraspecific heterogeneity among plant pathogen
956 metapopulations. Environ Microbiol. 2020 Jul 2;22(7):2693–708.

957 24. Eck JL, Barrès B, Soubeyrand S, Sirén J, Numminen E, Laine AL. Strain Diversity and Spatial
958 Distribution Are Linked to Epidemic Dynamics in Host Populations. Am Nat. 2022
959 Jan;199(1):59–74.

960 25. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. Arabidopsis
961 thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary
962 Timescales. Cell Host Microbe. 2018 Jul 11;24(1):168-179.e4.

963 26. Susi H, Sallinen S, Laine AL. Coinfection with a virus constrains within-host infection load
964 but increases transmission potential of a highly virulent fungal plant pathogen. Ecol Evol.
965 2022;12(3):e8673.

966 27. Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, et al. Coexistence of
967 Multiple Endemic and Pandemic Lineages of the Rice Blast Pathogen. mBio. 2018 Apr
968 3;9(2):10.1128/mbio.01806-17.

969 28. Roques L, Desbiez C, Berthier K, Soubeyrand S, Walker E, Klein EK, et al. Emerging strains
970 of watermelon mosaic virus in Southeastern France: model-based estimation of the dates
971 and places of introduction. Sci Rep. 2021 Mar 29;11(1):7058.

972 29. Ali S, Gladieux P, Ravel S, Adreit H, Meusnier I, Milazzo J, et al. Evolution of the rice blast
973 pathogen on spatially structured rice landraces maintains multiple generalist fungal
974 lineages. Mol Ecol. 2023;32(10):2519–33.

975 30. Elmer WH. Seeds as vehicles for pathogen importation. Biol Invasions. 2001;3:263–71.

976 31. Klein-Gordon JM, Xing Y, Garrett KA, Abrahamian P, Paret ML, Minsavage GV, et al.
977 Assessing Changes and Associations in the Xanthomonas perforans Population Across
978 Florida Commercial Tomato Fields Via a Statewide Survey. Phytopathology. 2021
979 Jun;111(6):1029–41.

980 32. Klein-Gordon JM, Timilsina S, Xing Y, Abrahamian P, Garrett KA, Jones JB, et al. Whole
981 genome sequences reveal the Xanthomonas perforans population is shaped by the tomato
982 production system. ISME J. 2022 Feb;16(2):591–601.

983 33. Potnis N. Harnessing Eco-Evolutionary Dynamics of Xanthomonads on Tomato and Pepper
984 to Tackle New Problems of an Old Disease. Annu Rev Phytopathol. 2021 Aug 25;59(Volume
985 59, 2021):289–310.

986 34. Anderson AR, Moore LW. Host specificity in the genus Agrobacterium. Phytopathology.
987 1979;69(4):320–3.

988 35. Mitchell SE, Rogers ES, Little TJ, Read AF. Host-Parasite and Genotype-by-Environment
989 Interactions: Temperature Modifies Potential for Selection by a Sterilizing Pathogen.
990 Evolution. 2005;59(1):70–80.

991 36. Laine A -L. Pathogen fitness components and genotypes differ in their sensitivity to
992 nutrient and temperature variation in a wild plant–pathogen association. J Evol Biol. 2007
993 Nov 1;20(6):2371–8.

994 37. Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J. Continua of specificity and
995 virulence in plant host–pathogen interactions: causes and consequences. New Phytol.
996 2009;183(3):513–29.

997 38. Suffert F, Goyeau H, Sache I, Carpentier F, Gélisse S, Morais D, et al. Epidemiological trade-
998 off between intra- and interannual scales in the evolution of aggressiveness in a local plant
999 pathogen population. Evol Appl. 2018;11(5):768–80.

1000 39. Kaur A, Russell I, Liu R, Holland A, Bhandari R, Potnis N. Within-host adaptation of a foliar
1001 pathogen, Xanthomonas, on pepper in presence of quantitative resistance and ozone
1002 stress [Internet]. bioRxiv; 2024 [cited 2024 Aug 14]. p. 2024.02.29.582737. Available from:
1003 https://www.biorxiv.org/content/10.1101/2024.02.29.582737v1

1004 40. Meyer KM, Porch R, Muscettola IE, Vasconcelos ALS, Sherman JK, Metcalf CJE, et al. Plant
1005 neighborhood shapes diversity and reduces interspecific variation of the phyllosphere
1006 microbiome. ISME J. 2022 May;16(5):1376–87.

1007 41. Thompson JN, Burdon JJ. Gene-for-gene coevolution between plants and parasites.
1008 Nature. 1992 Nov;360(6400):121–5.

1009 42. Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. Threats to global food
1010 security from emerging fungal and oomycete crop pathogens. Nat Food. 2020
1011 Jun;1(6):332–42.

1012 43. Castledine M, Padfield D, Sierocinski P, Soria Pascual J, Hughes A, Mäkinen L, et al. Parallel
1013 evolution of Pseudomonas aeruginosa phage resistance and virulence loss in response to
1014 phage treatment in vivo and in vitro. eLife. 2022 Feb 21;11:e73679.

1015 44. Papaix J, Carpentier F, Monod H, Lannou C. The role of spatial heterogeneity in shaping
1016 pathogen populations in agricultural landscapes. 2024 Jan 31 [cited 2024 Sep 25];
1017 Available from: https://www.authorea.com/users/735225/articles/711720-the-role-of-
1018 spatial-heterogeneity-in-shaping-pathogen-populations-in-agricultural-landscapes

1019 45. Sharma A, Timilsina S, Abrahamian P, Minsavage GV, Jones J, Vallad GE, et al. Bacterial
1020 mutation during seasonal epidemics. Mol Plant-Microbe Interactions® [Internet]. 2023 Dec
1021 17 [cited 2024 Feb 11]; Available from:
1022 https://apsjournals.apsnet.org/doi/abs/10.1094/MPMI-10-23-0164-SC

1023 46. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable
1024 resistance. Annu Rev Phytopathol. 2002 Sep 28;40(1):349–79.

1025 47. Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW. Reclassification of the Xanthomonads
1026 Associated with Bacterial Spot Disease of Tomato and Pepper. Syst Appl Microbiol. 2004
1027 Dec 15;27(6):755–62.

1028 48. Potnis N, Krasileva K, Chow V, Almeida NF, Patil PB, Ryan RP, et al. Comparative genomics
1029 reveals diversity among xanthomonads infecting tomato and pepper. BMC Genomics.
1030 2011 Dec 11;12(1):146.

1031 49. Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J, et al. Phylogenomics of
1032 Xanthomonas field strains infecting pepper and tomato reveals diversity in effector
1033 repertoires and identifies determinants of host specificity. Front Microbiol. 2015;6:535.

1034 50. Newberry EA, Minsavage GV, Holland A, Jones JB, Potnis N. Genome-Wide Association to
1035 Study the Host-Specificity Determinants of Xanthomonas perforans. Phytopathology®.
1036 2023 Mar;113(3):400–12.

1037 51. Horsfall JG, Barratt RW. An improved grading system for measuring plant diseases.
1038 Phytopathology. 1945;35:655.

1039 52. El Jarroudi M, Karjoun H, Kouadio L, El Jarroudi M. Mathematical modelling of non-local
1040 spore dispersion of wind-borne pathogens causing fungal diseases. Appl Math Comput.
1041 2020 Jul 1;376:125107.

1042 53. Myers TA, Scott RC, Zelinka MD, Klein SA, Norris JR, Caldwell PM. Observational
1043 constraints on low cloud feedback reduce uncertainty of climate sensitivity. Nat Clim
1044 Change. 2021 Jun;11(6):501–7.

1045 54. Lahlali R, Taoussi M, Laasli SE, Gachara G, Ezzouggari R, Belabess Z, et al. Effects of climate
1046 change on plant pathogens and host-pathogen interactions. Crop Environ. 2024 Sep
1047 1;3(3):159–70.

1048 55. Janda M, Lamparová L, Zubíková A, Burketová L, Martinec J, Krčková Z. Temporary heat
1049 stress suppresses PAMP-triggered immunity and resistance to bacteria in Arabidopsis
1050 thaliana. Mol Plant Pathol. 2019;20(7):1005–12.

1051 56. Desaint H, Aoun N, Deslandes L, Vailleau F, Roux F, Berthomé R. Fight hard or die trying:
1052 when plants face pathogens under heat stress. New Phytol. 2021;229(2):712–34.

1053 57. Chaloner TM, Gurr SJ, Bebber DP. Plant pathogen infection risk tracks global crop yields
1054 under climate change. Nat Clim Change. 2021 Aug;11(8):710–5.

1055 58. Bilgin DD, Aldea M, O'Neill BF, Benitez M, Li M, Clough SJ, et al. Elevated Ozone Alters
1056 Soybean-Virus Interaction. Mol Plant-Microbe Interactions®. 2008 Oct;21(10):1297–308.

1057 59. Bhandari R, Sanz-Saez A, Leisner CP, Potnis N. Xanthomonas infection and ozone stress
1058 distinctly influence the microbial community structure and interactions in the pepper
1059 phyllosphere. ISME Commun. 2023 Mar 27;3(1):1–13.

1060 60. Scherm,H., Yang, X. B. Interannual variations in wheat rust development in China and the
1061 United States in relation to the El Niño/Southern Oscillation. Phytopathol 970–976.
1062 1995;(19951000069, English, Journal article).

1063 61. Brasier CM. Phytophthora cinnamomi and oak decline in southern Europe. Environmental
1064 constraints including climate change. Ann Sci For. 1996;53(2–3):347–58.

1065 62. Corredor-Moreno P, Saunders DGO. Expecting the unexpected: factors influencing the
1066 emergence of fungal and oomycete plant pathogens. New Phytol. 2020;225(1):118–25.

1067 63. Abrahamian P, Sharma A, Jones JB, Vallad GE. Dynamics and Spread of Bacterial Spot
1068 Epidemics in Tomato Transplants Grown for Field Production. Plant Dis. 2021
1069 Mar;105(3):566–75.

1070 64. Sharma A, Timilsina S, Abrahamian P, Minsavage GV, Colee J, Ojiambo PS, et al. Need for
1071 speed: bacterial effector XopJ2 is associated with increased dispersal velocity of
1072 Xanthomonas perforans. Environ Microbiol. 2021;23(10):5850–65.

1073 65. Gause GF. Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle
1074 for Existence. Science. 1934 Jan 5;79(2036):16–7.

1075 66. Byappanahalli MN, Whitman RL, Shively DA, Sadowsky MJ, Ishii S. Population structure,
1076 persistence, and seasonality of autochthonous Escherichia coli in temperate, coastal forest
1077 soil from a Great Lakes watershed. Environ Microbiol. 2006;8(3):504–13.

1078 67. Jang J, Unno T, Lee SW, Cho KH, Sadowsky MJ, Ko G, et al. Prevalence of season-specific
1079 Escherichia coli strains in the Yeongsan River Basin of South Korea. Environ Microbiol.
1080 2011;13(12):3103–13.

1081 68. Jang J, Di DYW, Lee A, Unno T, Sadowsky MJ, Hur HG. Seasonal and Genotypic Changes in
1082 Escherichia coli Phylogenetic Groups in the Yeongsan River Basin of South Korea. PLOS
1083 ONE. 2014 Jul 7;9(7):e100585.

1084 69. Roques L, Klein EK, Papaïx J, Sar A, Soubeyrand S. Impact of Lockdown on the Epidemic
1085 Dynamics of COVID-19 in France. Front Med [Internet]. 2020 Jun 5 [cited 2024 Sep 25];7.
1086 Available from:
1087 https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2020.00274/full

1088 70. Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, et al. Testing
1089 Differences Between Pathogen Compositions with Small Samples and Sparse Data.
1090 Phytopathology®. 2017 Oct;107(10):1199–208.

1091 71. Van Den Berg F, Gilligan CA, Bailey DJ, Van Den Bosch F. Periodicity in Host Availability
1092 Does Not Account for Evolutionary Branching as Observed in Many Plant Pathogens: An
1093 Application to *Gaeumannomyces graminis* var. *tritici*. Phytopathology®. 2010
1094 Nov;100(11):1169–75.

1095 72. Penczykowski RM, Walker E, Soubeyrand S, Laine AL. Linking winter conditions to regional
1096 disease dynamics in a wild plant–pathogen metapopulation. New Phytol [Internet]. 2015
1097 [cited 2024 Aug 19];205(3):1142–52. Available from:
1098 https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.13145

1099 73. van den Berg F, Gilligan CA, Bailey DJ, van den Bosch F. Periodicity in host availability does
1100 not account for evolutionary branching as observed in many plant pathogens: an
1101 application to Gaeumannomyces graminis var. tritici. Phytopathology. 2010
1102 Nov;100(11):1169–75.

1103 74. van den Berg F, Bacaer N, Metz JAJ, Lannou C, van den Bosch F. Periodic host absence can
1104 select for higher or lower parasite transmission rates. Evol Ecol [Internet]. 2011 Jan 1
1105 [cited 2024 Aug 19];25(1):121–37. Available from: https://doi.org/10.1007/s10682-010-
1106 9387-0

1107 75. Hamelin FM, Castel M, Poggi S, Andrivon D, Mailleret L. Seasonality and the evolutionary
1108 divergence of plant parasites. Ecology. 2011 Dec;92(12):2159–66.

1109 76. Ballaré CL. Light regulation of plant defense. Annu Rev Plant Biol. 2014;65:335–63.

1110 77. Beattie GA, Hatfield BM, Dong H, McGrane RS. Seeing the Light: The Roles of Red- and
1111 Blue-Light Sensing in Plant Microbes. Annu Rev Phytopathol [Internet]. 2018 Aug 25 [cited
1112 2024 Aug 19];56(Volume 56, 2018):41–66. Available from:
1113 https://www.annualreviews.org/content/journals/10.1146/annurev-phyto-080417-
1114 045931

1115 78. Yeats C, Rawlings ND, Bateman A. The PepSY domain: a regulator of peptidase activity in
1116 the microbial environment? Trends Biochem Sci. 2004 Apr 1;29(4):169–72.

1117 79. Gallois N, Alpha-Bazin B, Bremond N, Ortet P, Barakat M, Piette L, et al. Discovery and
1118 characterization of UipA, a uranium- and iron-binding PepSY protein involved in uranium
1119 tolerance by soil bacteria. ISME J. 2022 Mar;16(3):705–16.

1120 80. Coyne MJ, Comstock LE. Type VI Secretion Systems and the Gut Microbiota. Microbiol
1121 Spectr. 2019 Mar;7(2):10.1128/microbiolspec.psib-0009–2018.

1122 81. Allsopp LP, Bernal P, Nolan LM, Filloux A. Causalities of war: The connection between type
1123 VI secretion system and microbiota. Cell Microbiol. 2020 Mar;22(3):e13153.

1124 82. Serapio-Palacios A, Woodward SE, Vogt SL, Deng W, Creus-Cuadros A, Huus KE, et al. Type
1125 VI secretion systems of pathogenic and commensal bacteria mediate niche occupancy in
1126 the gut. Cell Rep. 2022 Apr 26;39(4):110731.

1127 83. Thri Murty VS, Devadath S. Role of seed in survival and transmission of Xanthomonas
1128 campestris *pv.* oryzae. J Phytopathol. 1984 May 1;110(1):15–9.

1129 84. Gitaitis R, Walcott R. The Epidemiology and Management of Seedborne Bacterial Diseases.
1130 Annu Rev Phytopathol. 2007 Sep 8;45(Volume 45, 2007):371–97.

1131 85. Chalam VC, Deepika DD, Abhishek GJ, Maurya AK. Major Seed-Borne Diseases of
1132 Agricultural Crops: International Trade of Agricultural Products and Role of Quarantine. In:
1133 Kumar R, Gupta A, editors. Seed-Borne Diseases of Agricultural Crops: Detection, Diagnosis
1134 & Management [Internet]. Singapore: Springer; 2020 [cited 2024 Oct 28]. p. 25–61.
1135 Available from: https://doi.org/10.1007/978-981-32-9046-4_2

1136 86. Koskella B, Giraud T, Hood ME. Pathogen relatedness affects the prevalence of within-host
1137 competition. Am Nat. 2006 Jul;168(1):121–6.

1138 87. Thrall PH, Burdon JJ. Effect of resistance variation in a natural plant host–pathogen
1139 metapopulation on disease dynamics. Plant Pathol. 2000;49(6):767–73.

1140 88. Susi H, Vale PF, Laine AL. Host Genotype and Coinfection Modify the Relationship of within
1141 and between Host Transmission. Am Nat. 2015 Aug;186(2):252–63.

1142 89. Alizon S, de Roode JC, Michalakis Y. Multiple infections and the evolution of virulence. Ecol
1143 Lett. 2013 Apr;16(4):556–67.

1144 90. Hu L, Yang L. Time to Fight: Molecular Mechanisms of Age-Related Resistance.
1145 Phytopathology®. 2019 Sep;109(9):1500–8.

1146 91. Whalen MC. Host defence in a developmental context. Mol Plant Pathol. 2005 May
1147 1;6(3):347–60.

1148 92. Sharabani G, Shtienberg D, Borenstein M, Shulhani R, Lofthouse M, Sofer M, et al. Effects
1149 of plant age on disease development and virulence of Clavibacter michiganensis subsp.
1150 michiganensis on tomato. Plant Pathol. 2013;62(5):1114–22.

1151 93. Huss CP, Holmes KD, Blubaugh CK. Benefits and Risks of Intercropping for Crop Resilience
1152 and Pest Management. J Econ Entomol. 2022 Oct 1;115(5):1350–62.

1153 94. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
1154 comparisons that enables improved genome recovery from metagenomes through de-
1155 replication. ISME J [Internet]. 2017 Dec 1 [cited 2021 Nov 12];11(12):2864. Available from:
1156 /pmc/articles/PMC5702732/

1157 95. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome
1158 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol.
1159 2014 Nov 19;15(11):524.

1160 96. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display
1161 and annotation. Bioinforma Oxf Engl. 2007 Jan 1;23(1):127–8.

1162 97. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina
1163 second-generation sequencing data. BMC Bioinformatics. 2010 Sep 27;11(1):485.

1164 98. The Huttenhower Lab. Kneaddata [Internet]. 2020. Available from:
1165 https://github.com/biobakery/kneaddata

1166 99. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol.
1167 2019 Nov 28;20(1):257.

1168 100. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference
1169 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional
1170 annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733-745.

1171 101. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in
1172 metagenomics data. PeerJ Comput Sci. 2017 Jan 2;3:e104.

1173 102. Albanese D, Donati C. Strain profiling and epidemiology of bacterial species from
1174 metagenomic sequencing. Nat Commun. 2017 Dec 22;8(1):2260.

1175 103. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and
1176 open software for comparing large genomes. Genome Biol. 2004;5(2):R12.

1177 104. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci [Internet].
1178 2003 [cited 2022 Mar 21];14(6):927–30. Available from:
1179 https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1654-1103.2003.tb02228.x

1180 105. Sparks AH. nasapower: a NASA POWER global meteorology, surface solar energy and
1181 climatology data client for R [Journal of Open Source Software, vol. 3, no. 30, 1035
1182 [Internet]. Baltimore, Maryland, United States: NumFOCUS; 2018 [cited 2023 Oct 18].
1183 Available from: http://joss.theoj.org/papers/10.21105/joss.01035

1184 106. Westfall PH. Kurtosis as Peakedness, 1905 - 2014. R.I.P. Am Stat. 2014;68(3):191–5.

1185 107. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B Methodol
1186 [Internet]. 1996 Jan 1 [cited 2023 Oct 22];58(1):267–88. Available from:
1187 https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

108. McCullagh P. Regression Models for Ordinal Data. J R Stat Soc Ser B Methodol [Internet]. 1980 Jan 1 [cited 2024 Aug 20];42(2):109–27. Available from: https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

109. Ferrari S, Cribari-Neto F. Beta Regression for Modelling Rates and Proportions. J Appl Stat [Internet]. 2004 [cited 2023 Oct 26];31(7):799–815. Available from: https://econpapers.repec.org/article/tafjapsta/v_3a31_3ay_3a2004_3ai_3a7_3ap_3a799-815.htm

110. Cribari-Neto F, Zeileis A. Beta Regression in R. J Stat Softw [Internet]. 2010 Apr 5 [cited 2023 Oct 25];34:1–24. Available from: https://doi.org/10.18637/jss.v034.i02

111. Hijazi RH, Jernigan RW. Modelling Compositional Data Using Dirichlet Regression Models. Journal of Applied Probability & Statistics. 2009;4(1):77–91.

112. Simko I, Piepho HP. The area under the disease progress stairs: calculation, advantage, and application. Phytopathology. 2012 Apr;102(4):381–9.

113. Bates, D, Sarkar, D, Bated, M. D., Matrix, L. The lme4 package. R package. 2007.

114. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw. 2017 Dec 6;82:1–26.

115. Soubeyrand S, Tollenaere C, Haon-Lasportes E, Laine AL. Regression-Based Ranking of Pathogen Strains with Respect to Their Contribution to Natural Epidemics. PLOS ONE. 2014 Jan 31;9(1):e86591.

116. Puente-Sánchez F, Hoetzinger M, Buck M, Bertilsson S. Exploring intra-species diversity through non-redundant pangenome assemblies [Internet]. bioRxiv; 2022 [cited 2023 Feb 26]. p. 2022.03.25.485477. Available from: https://www.biorxiv.org/content/10.1101/2022.03.25.485477v1

117. Nayfach S. Metagenomic Intra-Species Diversity Analysis System (MIDAS) [Internet]. 2022 [cited 2022 Nov 9]. Available from: https://github.com/snayfach/MIDAS

118. Bushnell B. BBDuk. Jt Genome Inst Available Online Httpsjgi Doe Govdata--Toolsbbtoolsbb-Tools-Userguidebbduk-Guid 25 June 2020. 2020;

119. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv; 2013 [cited 2024 Aug 27]. Available from: http://arxiv.org/abs/1303.3997

120. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

121. Broad Institute. Picard Toolkit [Internet]. 2019. Available from: https://broadinstitute.github.io/picard/

122. Gregory AC, Gerhardt K, Zhong ZP, Bolduc B, Temperton B, Konstantinidis KT, et al. MetaPop: a pipeline for macro- and microdiversity analyses and visualization of microbial and viral metagenome-derived populations. Microbiome. 2022 Mar 15;10(1):49.

Supporting Information:

Supplementary Text:

S1 Appendix. Climatic shifts and extremes explain variation in BLS epidemics across the Southeastern United States

Supplementary Figures:

S1 Fig. A total of 8 fields consistently sampled over all three years displayed variable disease pressures across three years. Disease severity score is based on percent leaf area affected and estimated using the Horsfall-Barratt scale, which ranges from 1 to 12, with scale of 1 being no disease and 12 being 100% defoliation (Horsfall and Barratt 1945).

S2 Fig. (**A)** Disease severity ratings for different sampling time points across different states of Southeastern US. **(B)** the absolute abundance and **(C)** Relative abundance of *Xp* collected from different states during 2020, 2021, and 2022.

S3 Fig. Plots are comparing the samples based on the farm size (commercial vs small scale farms) **(A)** disease severity ratings; **(B)** Shannon diversity & **(C)** Number of Sequence clusters (SCs) or pathogen lineages present in the samples. The pathogen diversity was significantly higher in commercial fields compared to small-scale farms, although disease severity values did not significantly vary.

S4 Fig. Stacked bar plot depicting the co-occurrence of multiple *Xp* lineages, spatial and temporal variations, the introduction of new lineages, turnover, and dominance shifts in individual fields across various states during the mid and end of the season for the years 2020, 2021, and 2022

S5 Fig. Boxplots showing **(A)** disease severity ratings **& (B)** Average nucleotide diversity for farms having presence of less than and more than 2 Sequence clusters (pathogen lineages).

S6 Fig. (A) Correlation plot showing the relationship between Shannon diversity of *Xp* lineages and disease severity across all samples. The presence of more *Xp* lineages in the field correlates

with higher BLS disease severity; **(B)** Boxplot showing the Shannon diversity across the samples with low and high disease severity conditions. High DS includes the samples with disease severity scale ratings of more than and equal to 5, whereas low DS includes the farms with disease severity scale ratings of less than 5.

S7 Fig. Evaluating the fitness contributions of individual vs mixed infections by strains belonging to different sequence clusters under greenhouse conditions. Plants were dip-inoculated with ~1 X 10^6 of cell suspensions for each treatment. **(A)** Boxplot presenting the disease severity ratings evaluated using Horsfall-Barratt scale at day 7 and 15 of post-inoculation. The mixed infections showed higher disease severity compared to single infections**. (B)** pairwise-comparisons of AUDPC values across different pairs of treatments. Differences in means of disease development i.e. AUDPC for different treatment pairs are shown. Mean values of disease development with a 95% confidence interval were contrasted between treatments. Difference in means of disease development between treatments with an estimate 0 have no difference in means of disease development among pair of strains. Negative and positive values represent statistically significant treatments based on Tukey's HSD test.

S8 Fig. Using StrainRanking, **(A)** Comparisons of different SCs with fitness estimates and their confidence intervals under main effects were made. These estimates focused on SC3, SC4, and SC6 (which were more abundant in samples), and a group of SCs (SC1, SC2, SC5, SC7, and SC8) that were categorized as low-frequency (due to their lower abundance). The bottom plot illustrates the growth curves of these SCs, estimating their contributions to the progression of disease severity from mid and end of summer**. (B)** Pathogen-pathogen interactions, specifically between the two dominant SCs (SC3 and SC4), were analyzed based on their combined frequencies being either below or above 20%. Fitness estimates with confidence intervals were compared when the summed frequencies of SC3 and SC4 were taken into account. Different interaction scenarios were considered, such as SC3 > 20% & SC4 < 20%, SC4 > 20% & SC3 < 20%, and both SC3 & SC4 at equal frequencies, along with SC6 and the low-frequency SCs. The growth curves in the bottom plot highlight how these interactions contributed to the progression of disease during the season. **(C)** Pathogen-environment interactions were examined by applying the frequencies of SC3 or SC4 to different environmental variables, depending on whether each variable was above or below its mean at the sampling point. The analysis assumed weighted mean of genetic frequencies to calculate fitness levels as: p x mid-frequency + (1 - p) x end-frequency, where (p = 0.90).

S9 Fig. Using StrainRanking: **(A)** Comparisons of different SCs with fitness estimates and their confidence intervals under main effects were made. These estimates focused on SC3, SC4, and SC6 (which were more abundant in samples), and a group of SCs (SC1, SC2, SC5, SC7, and SC8) that were categorized as low-frequency (due to their lower abundance). The bottom plot illustrates the growth curves of these SCs, estimating their contributions to the progression of disease severity from mid and end of summer**. (B)** Pathogen-pathogen interactions, specifically between the two dominant SCs (SC3 and SC4), were analyzed based on their combined frequencies being either below or above 20%. Fitness estimates with confidence intervals were compared when the summed frequencies of SC3 and SC4 were taken into account. Different interaction scenarios were considered, such as SC3 > 20% & SC4 < 20%, SC4 > 20% & SC3 < 20%, and both SC3 & SC4 at equal frequencies, along with SC6 and the low-frequency SCs. The growth curves in the bottom plot highlight how these interactions contributed to the progression of disease during the season.

(C) Pathogen-environment interactions were examined by applying the frequencies of SC3 or SC4 to different environmental variables, depending on whether each variable was above or below its mean at the sampling point. The analysis assumed weighted mean of genetic frequencies to calculate fitness levels as: p x mid-frequency + (1 - p) x end-frequency, where (p = 0.10).

S10 Fig. plot showing the average allele frequencies for those alleles having **(A)** $f >= 0.8$ & **(B)** $f < 0.8$ during End-season 2020; **(C)** $f >= 0.8$ & **(D)** $f < 0.8$ during Fall-season 2020 and traced for the frequency changes during the previous and following seasons of three years.

S11 Fig. Sankey plot shows counts of the alleles present in parallel across farms during **A)** End-Season 2020 & **B)** Fall-season of 2020 with frequency less than and equal to 0.2 and then counts for those alleles which stayed as major during the next seasons.

S12 Fig. In circos plot is presenting LH3 genome with all the colored bars indicating different genes in different scaffolds, where genes under positive selection are shown with the red colored asterisk sign in the grey colored ring and the red colored arrows are showing those genes which were present in parallel across farms and seasons.

Supplementary Tables:

S1A Table. Farm details and time of sampling during three sampling years

S1B Table. Details of farm and disease severity across season and sampling years used in the study

S2 Table. Climatic parameters used in the study and their meaning

S3A Table. Influence of various factors in BLS disease severity
S3B Table. Influence of various factors in absolute abundance of X. euvesicatoria pv. perforans
S3C Table. Influence of various factors in relative abundance of X. euvesicatoria pv. perforans

S4 Table. Coefficients Estimates and p-values for different Sequence Clusters

S5 Table. Gene under positive selection pressure (Annotations according LH3 Genome)