Climatic fluctuations modulate eco-evolutionary processes associated with pathogen dynamics and continue to fuel bacterial spot epidemics in tomato

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: Risk factors for bacterial spot disease epidemics

Corresponding Author: Neha Potnis, nzp0024@auburn.edu

Abstract

Recurring outbreaks caused by endemic pathogens present a significant challenge to agricultural systems. Thus, understanding the risk factors involved in fueling the continued outbreaks and pathogen evolution is a priority. Isolate genome sequencing efforts have largely guided our past understanding of the pathogen population structure. However, this approach can overlook the importance of co-occurring pathogenic genera, species, or even lineages of the same pathogenic species in shaping disease dynamics. Here, we aimed to monitor pathogen population dynamics at a finer resolution, tap into the genetic variation existing and emerging within and across fields, and understand the determinants of this diversity in the *Xanthomonas*-tomato pathosystem. Using strain-resolved metagenomics, we found that pathogen heterogeneity with multiple co-occurring lineages is common, although accompanied by differential lineage dynamics and that higher disease severity is associated with higher pathogen diversity. Considering these observations, we used response-specific regression models to investigate the roles of environmental variables on driving these differential dynamics. We find that climatic fluctuations can modify the endemic disease risk and that the pathogen adapts to these climatic shifts by maintaining diversity of co-occurring lineages, each with a varying fitness contribution. We identified signatures of seasonal adaptation by monitoring genome-wide allele frequencies in pathogen. The observation of seasonal oscillations in allelic frequencies depicted evidence for fluctuating selection contributing to the patterns of genetic variation. We also identified positively selected loci under parallel evolution such as type VI secretion system genes and TonB-dependent receptors, which may explain the nature of selection pressures experienced by the pathogen. The findings from this study reveal fitness strategies adopted by endemic pathogens and how pathogens can evolve under the changing climate. Our high-resolution combinatorial approach exploiting a series of sequence data

50 and metadata types and analysis tools, is general enough to finely investigate eco-evolutionary
51 dynamics of pathogens at large scales in diverse case-studies concerning plant health, but also
52 animal and human health.

53 **Introduction**

54 Plants are engaged in an ongoing battle against quickly evolving pathogens, termed the 'co-
55 evolutionary arms race' (P. S. Teng *et al.*, 1984; Dixon, 2012; Savary *et al.*, 2012). This arms race
56 pushes the host plant to employ mechanisms of defense against pathogens that are simultaneously
57 co-evolving to evade or overpower those mechanisms (Kareiva, 1999; Anderson *et al.*, 2010).
58 Studies have shown a gene-for-gene link between pathogen virulence genes and corresponding
59 plant resistance genes, where selection imposed by the host plant can push a pathogenic population
60 to diversify linked virulence genes and vice versa (Ma *et al.*, 2006; Karasov *et al.*, 2014). Our view
61 of this arms race is heavily focused on selective pressures arising from binary interactions of the
62 pathogen with the host. However, along with the host selection pressure, pathogens must overcome
63 various biotic (competition with the resident microflora) and abiotic stressors (osmotic stress,
64 temperature, humidity, etc.). The resident microbiome members can mediate direct (synergistic or
65 antagonistic) or indirect interactions (host-mediated, i.e., via induction of host immunity) with the
66 pathogens. Co-occurring pathogens can also influence the disease dynamics through niche
67 competition with the pathogen or induced plant defenses (Sadhukhan *et al.*, 2024). While the
68 impact of pathogens on resident microbial community structure has been studied (Gao *et al.*, 2021;
69 Bhandari *et al.*, 2023a), how resident microbial communities influence pathogen population
70 dynamics during infection and evolution over time is still in question.

71 Epidemiological studies have identified a range of climatic factors involved in disease outbreaks.
72 However, 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. Furthermore, factors associated with modern agricultural ecosystems that may alter the plant-pathogen arms race include changes in agronomic practices, increased global trade, an intensified monoculture system, and heavy use of chemical control strategies (McDonald & Stukenbrock, 2016). These practices have been hypothesized to predispose the system to a significant challenge of emerging diseases and increased incidences of endemic disease outbreaks. In the case of seed-borne pathogens, the global trade facilitating the continuous input of different pathogen genotypes on seeds or plant materials circulated worldwide has been a driving force in the pathogen diversity observed under field conditions (Elmer, 2001). The evolution of chemical-resistant pathogens in response to chemical pesticides has been extensively studied (Thompson & Burdon, 1992; Fones *et al.*, 2020; Castledine *et al.*, 2022). Yet, the link between the cost of chemical resistance and compromise in the pathogen's virulence is unclear. Similarly, how chemical overuse could alter host susceptibility towards pathogens is poorly understood. Given this nuance and considering the complex interdependencies associated with the disease pyramid, the host-pathogen arms race needs to be re-evaluated. Understanding the variable contribution of different biotic and abiotic factors, individually and collectively, towards pathogen diversity and disease severity is crucial to pinpoint how and why outbreaks continue to occur. As pathogen genetic diversity enhances fitness, adaptability, and evolution, allowing pathogens to evade management, studying pathogen diversity is crucial for understanding infectivity, evolution, and adaptability in a changing environment, and for predicting pathogen and plant response under future climatic conditions (Bartoli et al. 2016; Desaint et al. 2020; McDonald and Linde 2002; Sundin and Wang 2018; Wichmann et al. 2005; Zhan et al. 2002).

95 The pathosystem of bacterial leaf spot (BLS) on tomatoes and peppers is an example of an endemic
96 disease where no known host resistance genes have been deployed in commercial tomato cultivars.
97 Yet, extensive shifts in the pathogen population, including species displacement and extensive
98 diversification have been observed in the past decades. For over a century, BLS has been
99 considered an endemic problem in the United States, particularly in the southeast (Jones *et al.*,
100 2004). Although *Xanthomonas euvesicatoria* pv. *euvesicatoria* (*Xeu*, hereafter) was the dominant
101 pathogen on tomatoes until the mid-1990s, *Xanthomonas euvesicatoria* pv. *perforans* (previously
102 known as *X. perforans*, *Xp*, hereafter) has been the predominant species in the southeastern US for
103 the past three decades (Potnis, 2021) . Over the past 32 years following its isolation in Florida, not
104 only has the *Xp* displaced the previously dominant *Xeu*, but studies have also shown a rapid change
105 in pathogen genotype and phenotype where sequencing of the isolates within farm fields routinely
106 discovers more diversity (Newberry *et al.*, 2019; Klein-Gordon *et al.*, 2021).The description of
107 this diversity has been changing since 2014, where the intraspecific diversity was first described
108 as 3 'groups' identified based on core genome phylogeny before being adapted to 6 different
109 'sequence clusters', referred to as 'SC' (synonymous to pathogen lineages based on core SNP-
110 based hierarchical clustering) in 2018 (Newberry *et al.*, 2019, 2020a). Newberry et al. (2019)
111 reported two new lineages, SC5 and SC6 in Alabama, not previously described in other
112 southeastern states. This study hinted at recent diversification events in the pathogen population
113 and unexplored diversity essentially just waiting to be identified. Multiple recombination and
114 horizontal transfer events were proposed to be responsible for the emergence of new *Xp* lineages
115 over the past three decades (Timilsina *et al.*, 2019; Newberry *et al.*, 2019).
116 Since isolate genome sequencing may present a bias of sequencing the dominant pathogen isolates,
117 we recently optimized the high-resolution metagenome sequencing approach on BLS-infected

118 phyllosphere field tomato and pepper samples. This approach allowed us to tap into the diversity
119 in an individual field at a given time, including low abundant pathogen genotypes. The finding of
120 multiple co-occurring pathogen genotypes in a single field was worth noting. It added another
121 variable to consider when studying pathogen dynamics and evolution, i.e., heterogeneity in
122 pathogen population (Newberry *et al.*, 2020a). This small-scale metagenome study in Alabama
123 fields and our previous research led to more questions surrounding pathosystem dynamics in farms
124 outside of Florida (1) Was the pathogen population heterogeneity common across southeastern US
125 tomato fields? (2) Are there persistent pathogen lineages that tend to exist in fields despite variable
126 factors of climate or cultural practices? And (3) whether shifts in pathogen lineages over the course
127 of the growing season could be linked to disease severity outcomes? We addressed these questions
128 in this study using a high-resolution metagenomics approach on phyllosphere tomato samples from
129 the fields across the southeastern states of Alabama, Georgia, North Carolina, and South Carolina
130 for three consecutive growing seasons from 2020 to 2022. While efforts have been geared to keep
131 up with newfound diversity, a lack of apparent resistance within tomatoes has led to questions
132 surrounding the driving factor in this diversification and the continued disease outbreaks
133 worldwide (Newberry *et al.*, 2019; Potnis, 2021). Each biotic and abiotic factor, individually and
134 in combination, affects pathogen and disease dynamics at each step along the tomato production
135 chain, starting with seed-production areas, transplant houses, and farm fields (Potnis, 2021; Klein-
136 Gordon *et al.*, 2022). Production complexity with these steps during the production chain aids in
137 the continuous input of different pathogen genotypes on seeds or plant materials circulated
138 worldwide and increases opportunities for genetic exchange and diversification. Among these
139 factors, the tomato production system has partially credited its structure in shaping genetic
140 variation within *Xp* (Klein-Gordon *et al.*, 2022).

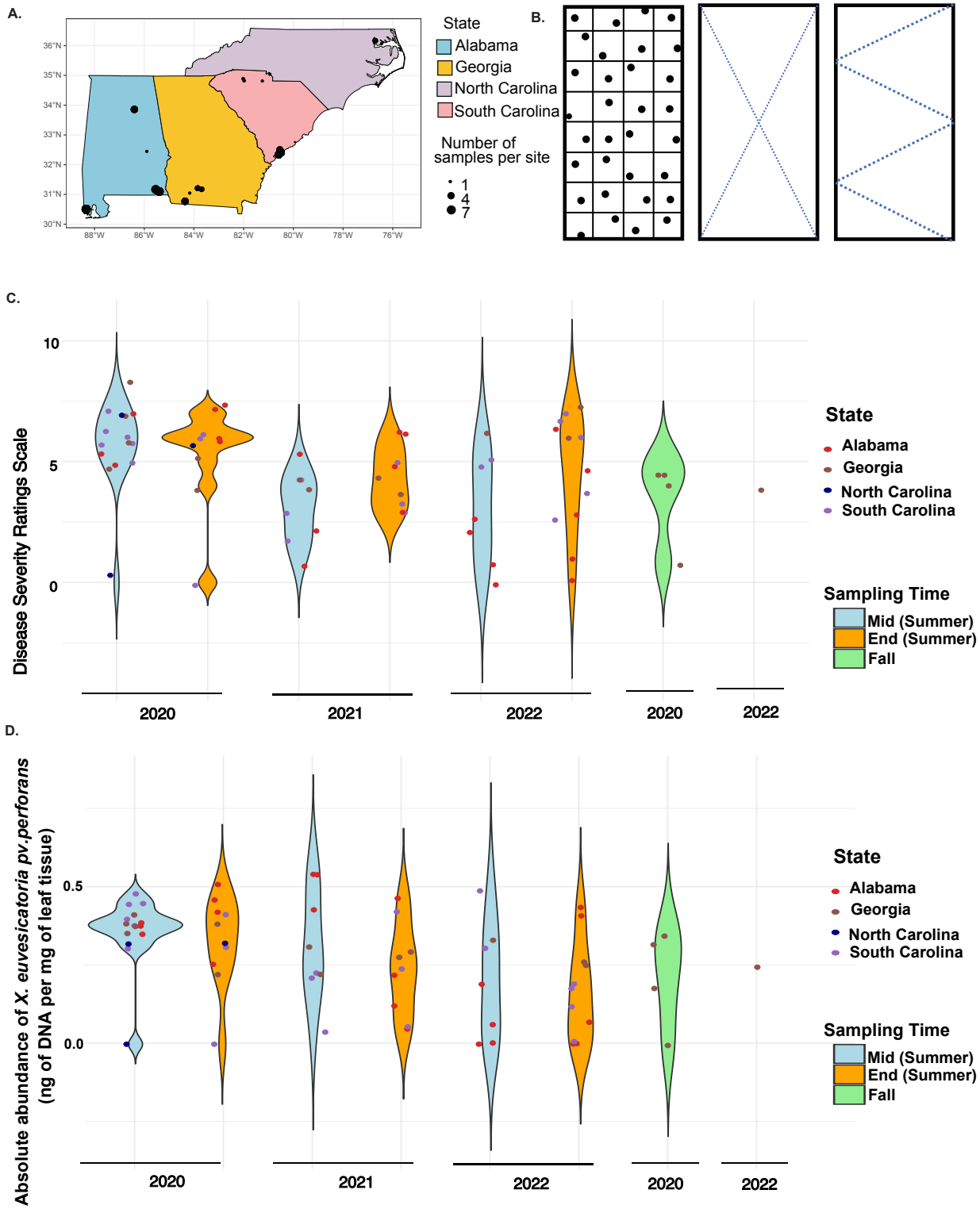
141 In this study, we focused on identifying the drivers of pathogen population dynamics and
142 disease dynamics at the field level, assuming that multiple pathogen genotypes are introduced into
143 the field. We specifically focused on the contribution of climatic factors, production scale, and
144 pathogen intraspecies diversity in explaining a broad spectrum of disease severities observed for
145 this endemic disease across neighboring states. These factors were also evaluated as drivers of
146 pathogen dynamics, explaining the temporal shifts observed in the pathogen lineages across fields.
147 We find that climatic fluctuations can modify the endemic disease risk. The diversity of *Xp* found
148 globally is circulating in tomato fields of the southeastern US, and multiple co-occurring lineages
149 are common. We found evidence for different ecological processes involved in pathogen dynamics
150 such as strain fluctuation, strain invasion, and differential strain turnover across neighboring fields.
151 Our findings demonstrate that the variable fitness levels of pathogen lineages and the climatic
152 anomalies in combination can drive differential pathogen dynamics and disease severities observed
153 across the fields. The seasonal oscillations in allelic frequencies observed across spatio-temporal
154 field samples suggest adaptive dynamics where different pathogen variants with variable fitness
155 levels across different climatic conditions are maintained in the population. The evidence of
156 positive selection at certain loci in the pathogen genome despite variability in field environments,
157 host genotypes, or geographical regions, indicates the role of these loci in pathogen fitness. This
158 study demonstrates the eco-evolutionary processes involved in within-field disease dynamics and
159 emphasizes the relevance of evaluating the role of climatic shifts in influencing plant-pathogen
160 interactions.
161
162 **Results**
163 **Varying levels of BLS disease pressure exist within the Southeastern US**

164 A total of 66 tomato phyllosphere samples were gathered over three years (2020-2022) from
165 Alabama, Georgia, North Carolina, and South Carolina (Fig 1A, Table S1A, S1B). The sampling
166 method was determined based on the field size and the amount of disease present (Fig 1B).
167 Metagenomic DNA from the leaf samples was subjected to shotgun metagenomics to obtain a
168 high-resolution pathogen population structure in the southeastern US. Apart from six samples
169 showing no BLS symptoms, the sampled fields showed variable levels of disease severity across
170 consistently sampled individual fields (Fig. S1), across the season and within and across states (Fig
171 1C, Fig. S2A). We observed that average disease severity in samples from Alabama was lower
172 (2.7) during mid-season compared to other states. But the average disease severity during the end
173 season was higher in Alabama (6) compared to other states (Fig. 1C, Fig. S2A). However, disease
174 severity was not consistently higher as the season progressed, as seen with samples from Georgia
175 and South Carolina (Fig 1C, Fig. S2A). We used the SNV profiling method, strainEST, to detect
176 pathogen lineages at finer resolution including those with low abundance levels. For this, we
177 screened metagenomic reads for the presence of BLS pathogenic species, *Xp*, *Xeu*, and X*eu*-related
178 pathovars (*Xeu* sister clades). We observed the absence of *Xeu* in sampled tomato fields suggesting
179 that *Xp* is the dominant pathogen of BLS in tomatoes in the southeastern United States. Using
180 correlation analysis as an exploratory tool, we found that the absolute abundance of *Xp* (Fig. 1D,
181 Fig. S2B) positively correlated with disease severity (mid-season: $R^2 = 0.381$, $p < 0.001$, end-
182 season: $R^2 = 0.283$, $p < 0.01$). A similar observation was made with the relative abundance of *Xp*
183 (mid-season: $R^2 = 0.508$, $p < 0.001$, end-season: $R^2 = 0.378$, $p < 0.001$) (Fig. S2C). The pathogen
184 population was significantly associated with disease severity during the early season compared to
185 the late season based on R^2 value. Although commercial-scale farms showed slightly higher

186 disease severity, this difference compared to small-scale farms was not statistically significant (p

187 $= 0.45$, Fig. S3A).

188


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

189

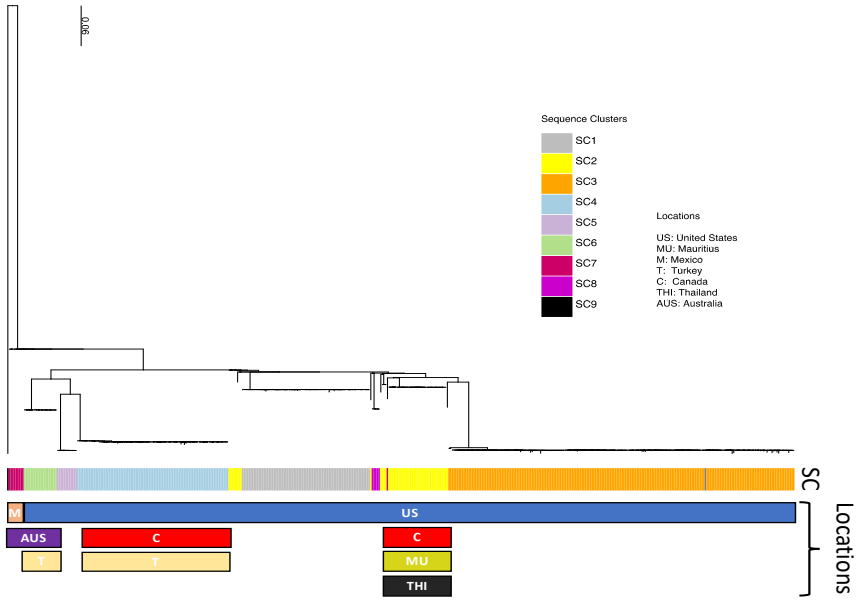
Figure 1: *Xanthomonas euvesicatoria* pv. *perforans* (*Xp*) continues to be a significant pathogen in the tomato fields across *the* southeastern US, although with variable disease pressure in the region. (A) Map of the study area from Alabama, North Carolina, South Carolina, and Georgia from Southeast 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. (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) the average disease severity and (D) the absolute abundance of *Xp* collected from different states during 2020, 2021, and 2022.

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

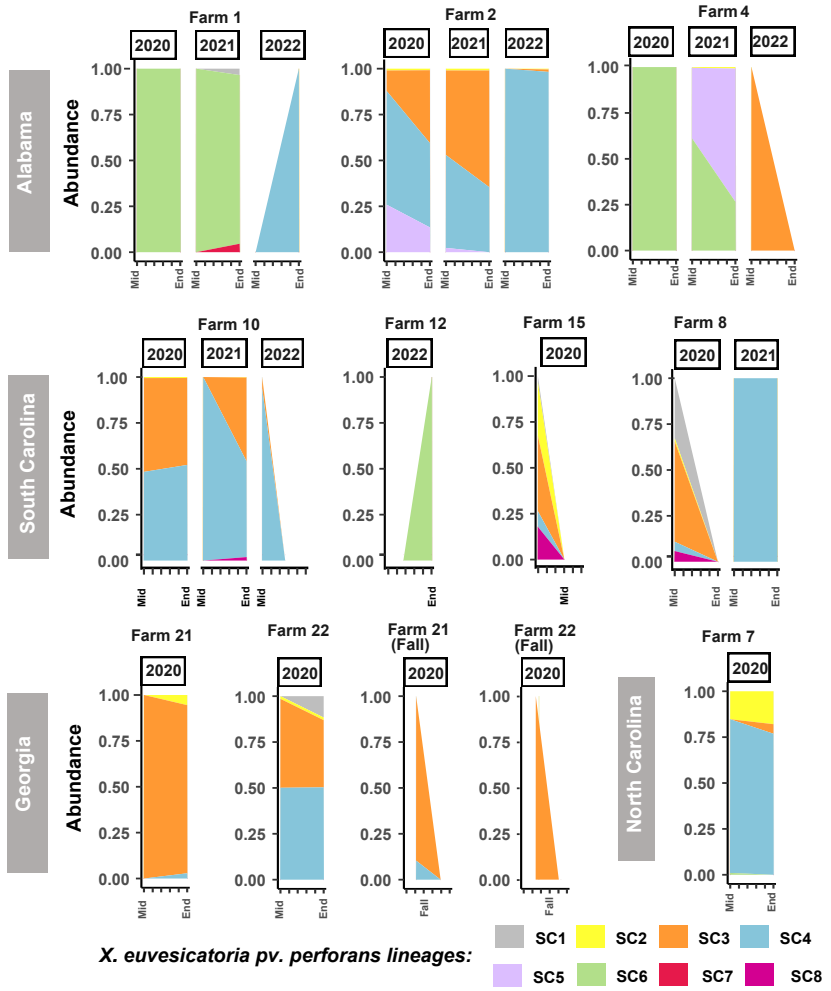
Given our observation of varying levels of epidemics, we hypothesized that pathogen heterogeneity is responsible for varying disease dynamics. Thus, we assessed the heterogeneity in the pathogen population by assessing the extent of diversity of *Xp* in individual fields. Our approach involved mapping the metagenome reads to the SNP matrices of known lineages or sequence clusters (SC) of *Xp* and calculating 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 here 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 (Fig. 2B). Surprisingly, we identified the presence of eight lineages in the field samples. We found that three lineages, SC3, SC4, and SC6, were in higher abundance in these fields, although their abundance varied from field to field, and their dynamics varied across neighboring fields. Notably, SC1, SC7, and SC8 were in lower abundance, appearing in only eight samples (Fig. 2B, Fig. S4). The co-occurrence of different *Xp* lineages were observed in the samples, as noted in our previous limited metagenome survey (Newberry et

218 al. 2020). Next, we investigated whether multiple co-occurring lineages of *Xp* within individual
219 fields influenced disease severity. We estimated the Shannon diversity index that captures
220 evenness indicative of the extent of diversity of *Xp*. Correlation analysis between the Shannon
221 diversity index of the pathogen population and disease severity revealed a positive association (R^2
222 $= 0.52$, $p < 0.001$) (Fig. S5). This suggests that *Xp* population diversity is positively associated
223 with disease severity. Looking at the diversity of *Xp* lineages across the states, fields had a
224 minimum of 1 to a maximum of 5 lineages present at a given sampling time, and this diversity of
225 co-occurring lineages was different across different years, and samples from neighboring states
226 (Fig. 2B). Changes in strain composition were not only evident across different fields but also
227 within individual fields throughout the growing season, spanning three years. These dynamics in
228 strain composition were marked by strain fluctuations, where the abundance of lineages shifted
229 from one season to the other, strain invasion where a new lineage likely infected the plants during
230 the growing season, and differential strain dynamics (Fig. 2B). Commercial and small-scale farm
231 types differed in the Shannon diversity, with commercial farms showing higher pathogen diversity
232 ($p = 0.008$) and higher number of SCs compared to small scale farms (Fig. S3B, S3C). Overall,
233 the diversity of *Xp* found globally is circulating in Southeastern US tomato fields, and co-
234 occurrence of multiple *Xp* lineages is common in this region.

A.


90.0
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
M
US
AUS
C
C
T
T
MU
THI

B.


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

235

Figure 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) 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. A figure showing data for all fields is included in the supplementary file Fig. S4.

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). In addition to routine climatic variables and their mean or median values, we also considered other measures, such as standard deviation or kurtosis, for example, allowing to account for other characteristics of climatic predictors using indexes that capture the spread (standard deviation), asymmetries, and tail-heaviness (skewness and kurtosis). 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 predictors for each response (see Statistical analysis section). Having selected the relevant predictors for each response, we ran the logistic ordinal regression model to understand drivers for disease severity which indicated that the Shannon diversity of various lineages of *Xp* present in the field (t value = 4.15, $p < 0.05$) strongly influenced BLS disease severity as suggested in the earlier result (Fig. S4). Among the climatic factors, the standard deviation of Wet Bulb Temperature at 2 meters (t-value = 2.06, $p < 0.05$) and

the standard deviation of the average of the wind direction at 10 meters above the surface of the earth (t-value = 2.73, $p < 0.05$) significantly increases BLS disease severity (Table 1A, Table S3A), indicating that extreme values of wet bulb temperature (adiabetic saturation temperature) and frequent changes in wind direction led to higher disease severity. In contrast, the skewness of clear sky surface photosynthetically active radiation (PAR) total (t-value = -2.57, $p < 0.05$, negative skewness indicative of more data points accumulating in the upper range of clear sky PAR) and kurtosis of Temperature at 2 Meters Range (t-value = -2.12, $p < 0.05$, negative kurtosis indicative of when tail min temperature relatively close to typical minimum temperature) decrease the BLS disease severity (Table 1A, Table S3A). More specifically, for the skewness predictor, this implies that as the majority of clear sky PAR measurements accumulate towards the lower range (and hence have fewer extremes in this range), the less the BLS disease severity occurs; for the kurtosis predictor we also see that when the temperatures at 2 meters range have more extreme behaviors, the less is the BLS disease severity.

Next, a beta regression model predicting the drivers of *Xp* abundance identified one key predictor: the standard deviation of wind direction at 10 meters above the earth's surface. This variable has positive association with the absolute (z-value = 2.08, $p < 0.05$)) and relative (z-value = 2.69, $p < 0.05$) abundance of *Xp*, contributing to disease severity. In addition to this, the skewness of surface pressure (z $value_{abs}$ = -3.261, z $value_{rel}$ = -2.949, $p < 0.05$) and kurtosis of relative humidity at 2 meters (z $value_{abs}$ = -2.524, z $value_{rel}$ = -2.744, $p < 0.05$) showed a significant negative effect both on the relative and absolute abundance of *Xp* (Table 1B, Table S3B, S3C). Thus, frequent shifts in wind direction and extreme changes in surface pressure and relative humidity all appear to influence *Xp* abundance. Other significant predictors of *Xp* relative abundance included farm scale, with commercial farms having a larger abundance (z = 2.875, $p < 0.05$) (Table S3C).

Finally, we used Dirichlet compositional regression to evaluate the influence of climatic factors in explaining the differential dynamics of pathogen lineages across different fields. This type of regression ensures that dependence between cluster abundance is accounted for (i..e. the 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. Also, kurtosis of relative humidity appears to have a negative impact on abundance in clusters SC1 to SC4 and SC8. In general, the contribution of the variables appears to be more significant in clusters SC3 and SC4 (and to a lesser extent SC2). The standard deviation of relative humidity appears to increase SC3 abundance. In contrast, variables standard deviation of minimum to the maximum temperature range at 2 meters, skewness of temperature at 2 meters, skewness of wind speed at 10 meters, kurtosis of relative humidity, and kurtosis of all-sky PAR appears to decrease the abundance of SC3. The variables, standard deviation of relative humidity, skewness of temperature at 2m, and Commercial scale farms (compared to small farms), appear to increase the abundance of SC4 (Table S4).

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 dominating lineages (Fig. 2B). 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 or direct or indirect (host-mediated) competition among co-occurring pathogen lineages. We asked how pathogen heterogeneity and variable dynamics could contribute to the explanation of fluctuating disease pressures during outbreaks. We used complementary data from the greenhouse experiment and field data provided here to evaluate the hypothesis that different disease pressures are caused by different fitness levels of co-existing pathogen lineages. The greenhouse experiment involved infecting tomato plants with individual strains representative of different lineages or in mixed infections. Overall, mixed infections led to higher disease severity than individual infections (Fig. S6A). Individual lineages differed in fitness levels (AUDPC) (Fig. S6B). For example, fitness contributions towards disease severity of mixed infection of SC3 and 4 were significantly higher than that of individual infections (Fig. S6B).

Next, we used field metagenome data with the relative abundance of pathogen lineages and disease severity data to assess pathogen fitness under the field condition. We used StrainRanking, a regression-based method for ranking of pathogen strains with respect to their contribution to

330 epidemics. This allowed us to combine pathogen genetic data with epidemiological monitoring of
331 disease progression to investigate differences among pathogen lineages in their performance in the
332 field. Lineages SC4 and SC6 tend to have a higher fitness than SC3 and other low frequency SCs
333 with respect to their contribution to the field epidemics, as indicated by their positive contribution
334 towards disease severity from mid to end season in the theoretical epidemic curves (Fig 3A). Equal
335 weighted means of the genetic frequencies observed during the mid- and end-seasons (p=0.5 for
336 each) were taken into consideration for this analysis. Even considering the weightage of genetic
337 frequencies high (p=0.9) in the mid-season (Fig. S7A) or high (p=0.9) at the end of the season,
338 (Fig. S8A), SC6 and SC4 contributed significantly in terms of fitness towards the growth of disease
339 severity. However, SC6 was not predominant lineage across our samples, so we did not consider
340 SC6 for further analysis.
341 When strain interactions were evaluated for their fitness contributions, lineages SC4 (>20%
342 abundance) and SC3 (≤20% abundance) combination showed positive fitness, whereas interaction
343 corresponding to SC4 (≤20% abundance) whatever SC3 frequency led to negative fitness
344 contribution. When SC4 (>20% abundance) and SC3 (>20%) coexisted, we observed near zero
345 fitness contribution (Fig 3B, Fig. S7B, Fig S8B). Next, we evaluated fitness levels of individual
346 SCs when strain-environment interactions were considered. It's interesting to note that we found
347 that various contextual factors significantly impacted SC4 fitness levels but less in SC3. When
348 average surface pressure and standard deviation of wind speed were less than mean values, but
349 average temperature and skewness of wet bulb temperature were more than mean values, SC4
350 demonstrated a larger positive fitness contribution (Fig. 3C). We found that specific environmental
351 conditions that are observed in the middle or end of the season have different effects on the fitness
352 contributions of SC3 and SC4, particularly when taking into account genetic frequencies that are

reported at the end of the season (Fig. S8C). While lower values of the parameters promote better fitness of SC3, higher values of the clear sky PAR standard deviation, kurtosis of surface pressure, and standard deviation of temperature at 2 meters support higher fitness of SC4. Conversely, lower levels of average clear sky PAR favor SC4 fitness (Fig. S8C), whereas higher 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 (Fig. S7C).

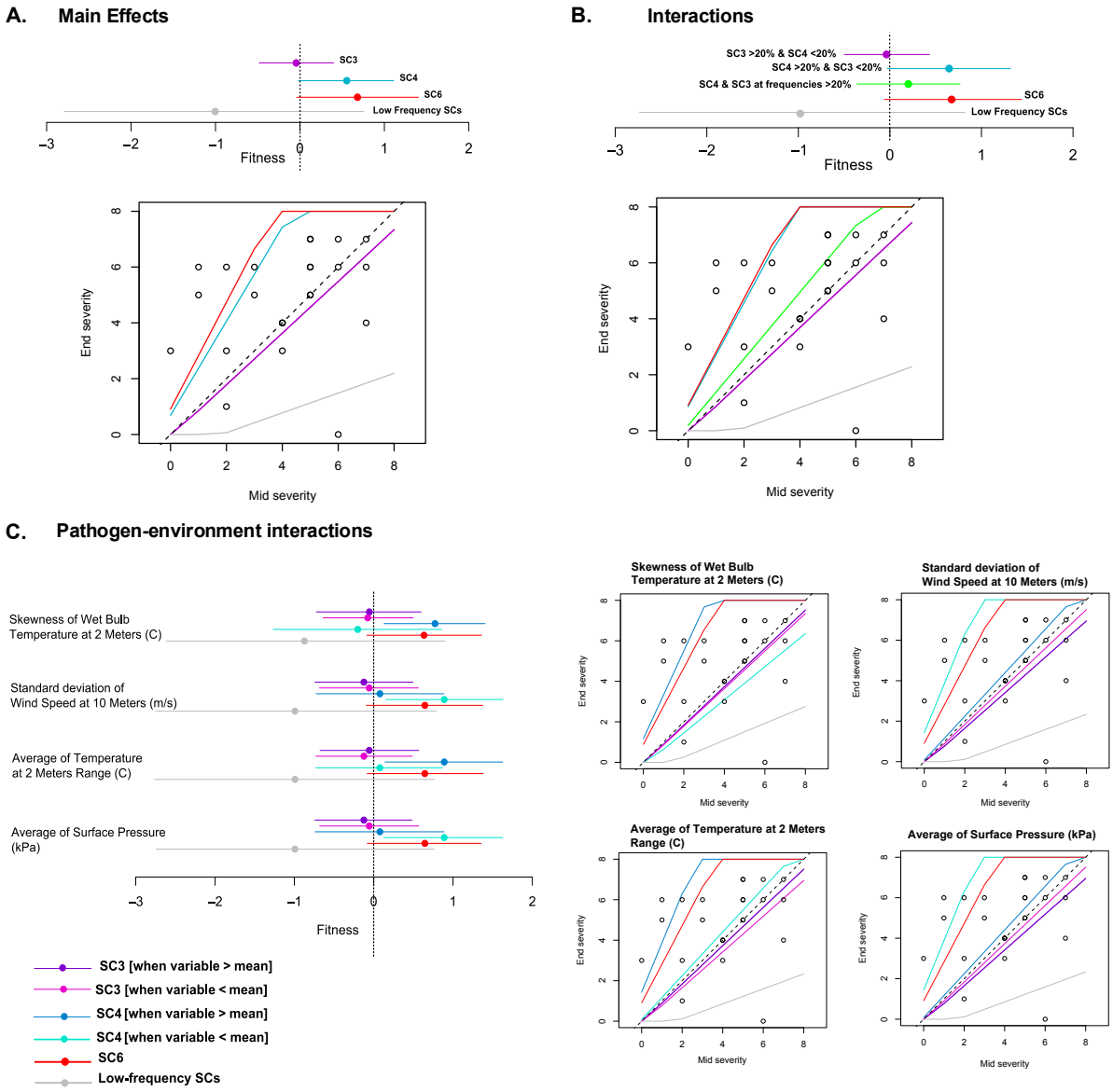


Figure 3. Different lineages have variable fitness levels and resilience towards different climatic factors and competition with co-occurring lineages. Using StrainRanking: A) Comparisons of different SCs with fitness estimates and their confidence intervals under main effects were made. These estimates 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 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

375 variable was above or below its mean at the sampling point. The analysis assumed an equal-weighted mean
376 of genetic frequencies to calculate fitness levels: p x mid-frequency + (1 - p) x end-frequency, where (p =
377 0.5).
378
379
380
381 **Seasonal preference for alleles indicative of conditional fitness for certain alleles was**
382 **observed.**
383 The influence of season and year on pathogen diversity led us to identify variants selected across
384 pathogen populations over time and seasons. We looked for alleles whose frequencies changed
385 within and across seasons and parallel across fields. For this analysis, we grouped all samples
386 from each respective sampling time (i.e., summer or fall of a specific year) into a single
387 metapopulation (Fig. 4A). We then tracked the major and minor allele frequencies across three
388 sampling points over three years. We observed a strong influence of seasonality on the allelic
389 frequencies. Allelic frequencies at most sites differed across the mid/end of the summer and fall
390 seasons. For example, a significant change in frequency (with f <= 0.2) was observed throughout
391 the fall season for several of the main alleles (f >=0.8) found in the middle or end of the summer
392 growing season (Fig. 4B). Similarly, the alleles that ranged in frequencies between 0.2 to 0.8
393 during mid-2020 remained in the population at intermediate frequencies during end seasons or
394 mid-seasons in the subsequent years but changed to either 1 or 0 during the fall seasons of the
395 subsequent years (Fig. 4C, Fig. S9). These polymorphisms that undergo dramatic rapid shifts in
396 allelic frequency and oscillate between seasons repeatedly over subsequent years can be defined
397 as conditional or seasonal SNPs (Fig. S9). These seasonal SNPs would most likely be adaptively
398 responding to selection pressures that vary across seasons.
399 Given the influence of seasonality on oscillating patterns of polymorphisms, we focused on
400 identifying alleles that were seen in parallel across at least 50% of farms with f <= 0.2 (called a
401 minor) during the first season and became f >= 0.8 (considered a major) in the following season

402 (Fig. 4A, 4C). We found that 237 of the alleles that were minor/low-frequency alleles in mid-2020
403 turned major (i.e., reached high frequency) during fall 2020, with 157 of those alleles reappearing
404 as major during fall 2022 but were not observed as high-frequency alleles in any other season of
405 the following years. Similarly, 1234 minor alleles that existed at the end of 2020 reached high
406 frequency (i.e., major alleles) in the fall of 2020. Of these, 138 alleles were observed at high
407 frequency in the subsequent fall season of 2022 (Fig. S10A). Furthermore, we observed 3598
408 allelic sites with allele frequency less than 0.2 during the fall of 2020 became major during mid-
409 season (summer) of the year 2021 and most of the sites (1025) appeared back as major alleles in
410 the mid-season of the year 2022 (Fig. S10B).

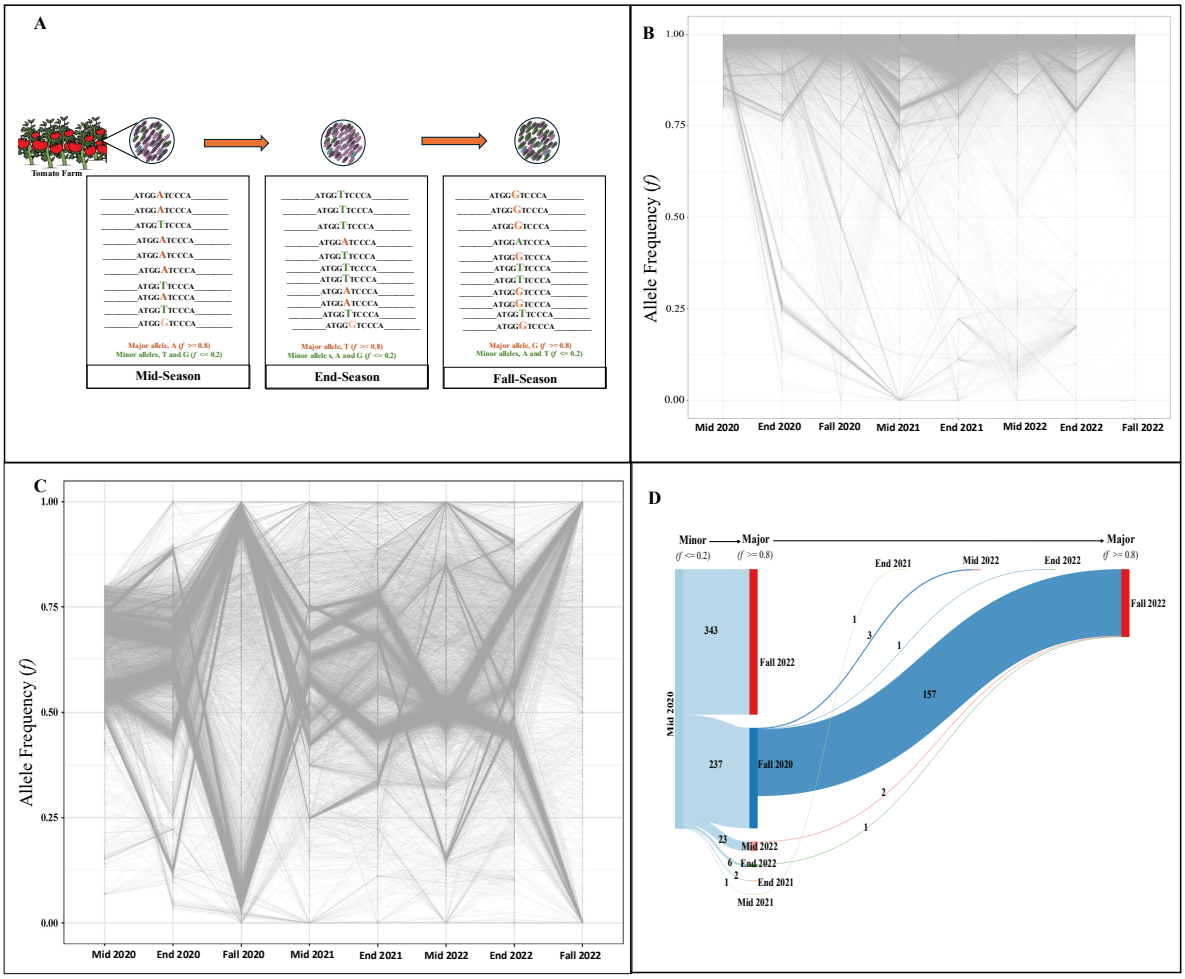


Figure 4. Seasonal oscillations of allelic frequencies with the signature of parallel evolution. A) A graphical representation of seasonal variation, where during mid-season (of summer), allele A was the major allele (allele frequency > 0.8) in the population. By the end of the season (of summer), the 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. B) The plot shows alleles with average frequencies of >= 0.8 in the pathogen population during mid-season (of summer) of year 2020, were oscillating to the following seasons with changes in their average frequencies over three years. C) The plot shows alleles with average frequencies of <= 0.8 in the pathogen population during mid-season 2020 (of summer), were changing in their average frequencies in the next seasons over three years. D) The Sankey plot shows all the alleles present in parallel in 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 in the population in the next season and how many of them stayed as major during the next seasons of following years.

427 **Despite variability in disease dynamics, signatures of selection and parallel evolution were**
428 **evident in the pathogen population.**
429 Given the varying disease pressure observed across different farms, we sought to determine if
430 any genes in the pathogen genome were consistently under selection across these locations.
431 Using Metapop package, we estimated the pN/pS (non-synonymous to synonymous ratios) of
432 different genes of *Xp* and pooled out the genes with pN/pS > 1 values. We filtered out the genes
433 if they were present in at least 50 % of farms in a particular sampling point to provide evidence
434 for parallel evolution. We found eight genes under selection were parallel across farms (at least
435 50 % of farms of specific sampling time) and seasons. These genes comprised type VI secretion
436 system (T6SS) protein ImpG, ClpV1 family T6SS ATPase, S9 family peptidase,
437 aminotransferase, tonB-dependent receptors, S9 family proteins (Fig. S11, Table S5). However,
438 none of these eight genes were clustered and present at distinct locations based on the reference
439 genome of *Xp* LH3. Interestingly, two proteins associated with the T6SS were under positive
440 selection, suggesting the microbiome's potential role in selection pressure and shaping the
441 pathogen genome. We also identified a gene, class 1 fructose-bisphosphatase, under positive
442 selection in both populations during the fall seasons (Table S5). Overall, these results suggest
443 evidence of parallel evolution across farms and seasons in the *Xp* pathogen population.
444

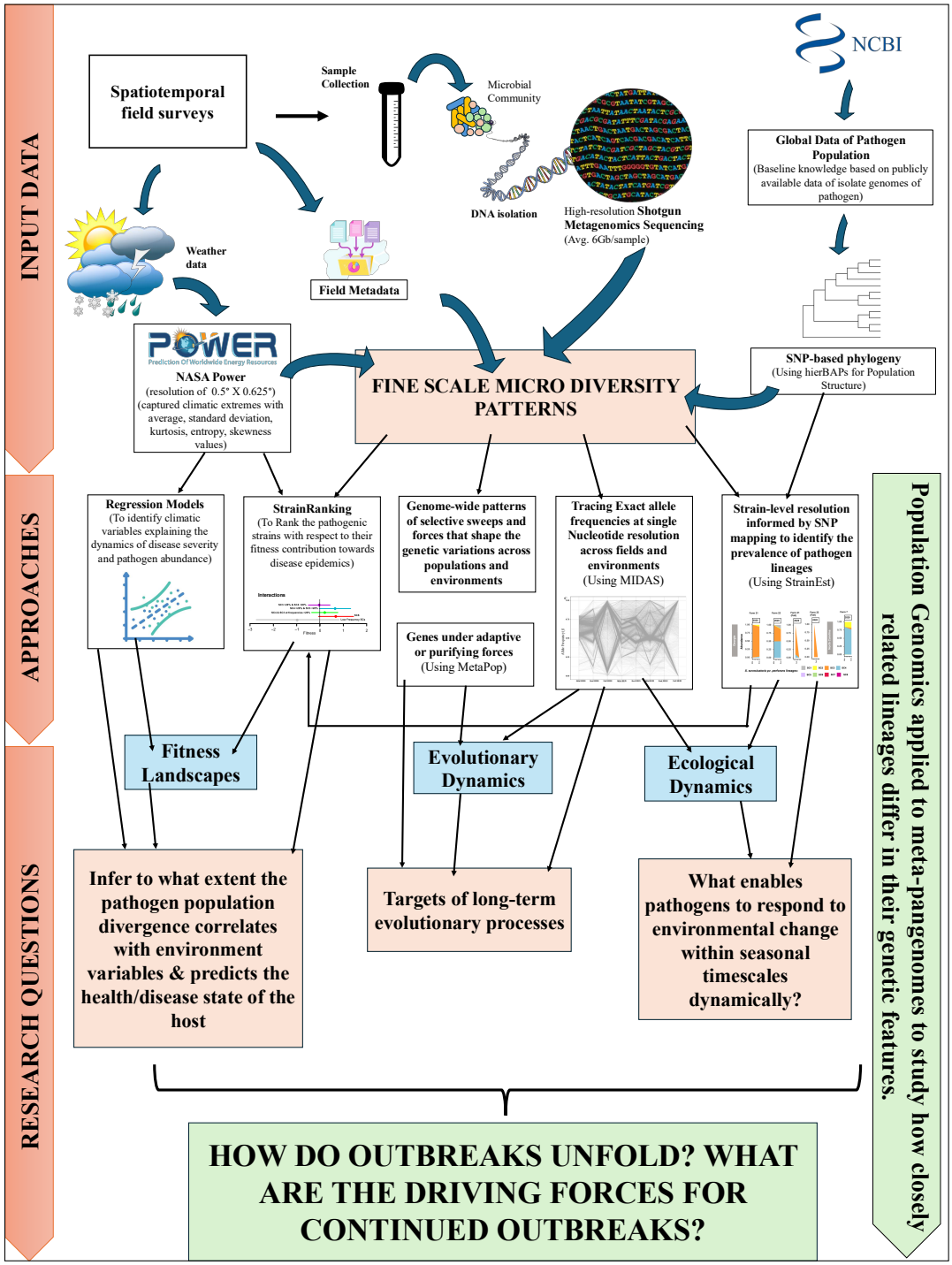


Figure 5. Integrated workflow to study pathogen outbreaks incorporating large scale field survey data, high-resolution metagenomics data and subsequent analyses to understand eco-evolutionary pathogen dynamics.

449 **Discussion**

450 This study investigated the factors contributing to continued disease outbreaks caused by an
451 endemic pathogen, identified eco-evolutionary processes involved in shaping the within-field
452 disease dynamics, inferred the extent to which these processes are linked to environmental
453 variables to predict health/disease state of the host (Fig. 5). We addressed this goal by conducting
454 a spatio-temporal survey of tomato fields in four neighboring states of the Southeastern US and by
455 applying combinatorial pathogen population genomics, strain-resolved metagenomics, and
456 epidemiological modeling approaches. Field surveys revealed variable disease pressure across
457 seasons and fields and *Xanthomonas* abundance levels alone could not explain this variability (Fig.
458 1). This variability could stem from host pressure, interactions among different pathogen lineages,
459 environmental factors, or a combination thereof. Here we systematically evaluated contribution of
460 pathogen-pathogen interactions and pathogen-environment interactions towards continued disease
461 outbreaks. While host genotype would be an important factor to consider in explaining disease
462 dynamics, the lack of known bacterial spot resistance genes in tomato cultivars and the difficulty
463 in obtaining this metadata during the survey did not provide enough information to systematically
464 evaluate the role of host genotypes in driving disease dynamics. However, age-related host
465 resistance (Whalen, 2005; Sharabani *et al.*, 2013; Hu & Yang, 2019) or the prevalence of other
466 diseases later in the season might contribute towards seasonal disease pressure. Epidemiological
467 modeling using high-resolution NASA POWER data for each sampling site coupled with disease
468 severity and pathogen population genomics data enabled evaluation of individual climatic
469 variables and the role of seasonal variation, sudden shifts in climate, and climatic extremes in
470 explaining disease outbreaks. We observed that sudden and abrupt changes in climatic variables,
471 such as wind direction, relative humidity, temperature, and surface pressure, can significantly

472 impact disease severity levels. Next, we characterized fine scale microdiversity patterns within
473 pathogen population to understand the ecological and evolutionary processes that drive the disease
474 dynamics. In contrast to the traditional view of agricultural systems selecting for a single dominant
475 pathogenic strain/lineage, we found that the coexistence of multiple lineages is common within
476 the current population structure and variable fitness contributions of different pathogen genotypes
477 can explain differential disease dynamics. We quantified the effect of intraspecific diversity in
478 pathogen populations on epidemic rates and found a strong positive correlation between pathogen
479 diversity and disease severity. To explain the within-host strain dynamics and their contribution to
480 the epidemic, we evaluated field-level fitness contributions of pathogen lineages individually and
481 in combination and observed that within-field disease dynamics are driven by variable fitness
482 levels of pathogens and climatic shifts (deviation in climatic variables from the ambient). Tracing
483 allelic variants over three years revealed a considerable influence of seasonality on pathogen
484 dynamics, with distinct seasonal alleles strongly favored throughout specific seasons. This
485 observation shows genetic heterogeneity comprising of co-existing lineages with differential and
486 environmentally dependent fitness contributions towards epidemic as an adaptation strategy that
487 enables pathogen to dynamically respond to environmental fluctuations within seasonal
488 timescales. These complex interdependent interactions complicate the selective pressures during a
489 host-pathogen arms race in case of endemic diseases in agricultural systems. Our study highlights
490 the value of high-resolution shotgun metagenomics in understanding within-host and across-host
491 pathogen dynamics. Identifying the drivers of the disease dynamics and pathogen diversification
492 will be critical in reducing the frequency of pathogen outbreaks (Fig. 5).

493 Regression models applied to link environmental data, disease outcomes and pathogen
494 diversity revealed that extreme climatic events were significant drivers of disease dynamics and

495 pathogen diversity across different farms. Frequent extreme events such as changes in wind
496 direction, photosynthetically active radiation, temperature, humidity, and surface pressure (Table
497 1) can directly influence pathogen dispersal (El Jarroudi *et al.*, 2020; Myers *et al.*, 2021; Lahlali
498 *et al.*, 2024) and plant host physiology (Janda *et al.*, 2019; Desaint *et al.*, 2021; Chaloner *et al.*,
499 2021), or indirectly alter host-pathogen, strain-strain, and pathogen-microbiome interactions
500 (Bilgin *et al.*, 2008; Bhandari *et al.*, 2023b; Kaur *et al.*, 2024). While empirical evidence for how
501 recent climatic shifts have contributed to disease outbreaks is limited, studies on the impact of
502 extreme weather events on the virulence of fungal pathogens like *Hemileia vastatrix* (coffee rust),
503 *Puccinia striiformis* (wheat stripe rust), and *Phytophthora cinnamomi* (Mediterranean oak decline)
504 have highlighted the need for these studies (Scherm,H. & Yang, X. B., 1995; Brasier, 1996;
505 Corredor-Moreno & Saunders, 2020). Our approach of capturing the climatic extremes using
506 measures such as standard deviation, kurtosis, skewness of climatic variables observed for each
507 farm across seasonal timescales allowed us to assess the extent to which pathogen genetic
508 divergence correlates with climatic shifts. The findings from these models highlighted the
509 importance of capturing these climatic extremes in the epidemiological modeling, rather than
510 average climatic variables.

511 Our findings of co-existence also challenge the traditional view that a single lineage of an
512 abundant pathogen dominates the host population in agricultural systems. However, unlike natural
513 ecosystems where multiple sub-lineages can persist at low levels over long periods (Karasov *et*
514 *al.*, 2014), we observed that a handful of pathogen lineages impart strong positive fitness
515 contributions towards disease epidemics. Interestingly, two lineages, SC3 and SC4 co-existed,
516 both at large densities, but neither showed a strong dominance across different fields in all three
517 examined growing seasons. First, this observation is intriguing, because it certainly challenges the

518 competitive exclusion principle, which states that two species occupying the same ecological niche
519 cannot coexist indefinitely (Gause, 1934). Second, it reveals an apparent paradox because, based
520 on our analyses, SC4 seems to be fitter than SC3. SC6 also seems to be as fit as SC4 but is present
521 at a lower level with distribution limited to only certain fields in Alabama than SC3 and SC4, thus,
522 its contribution in the lineage dynamics could not be systematically assessed. Here, we sought
523 explanations for this apparently unexpected co-existence of SC3 and SC4 and their ecological
524 dynamics. Prior studies on *E. coli* have demonstrated the prevalence and persistence of season-
525 specific strains due to fluctuating temperatures throughout the year (Byappanahalli *et al.*, 2006;
526 Jang *et al.*, 2011, 2014). However, we have yet to observe any clear patterns of early or late-season
527 specialization of individual SCs, indicating that time partitioning is unlikely to explain their co-
528 existence. Since we did not sample outside the host, current data cannot explain whether these
529 lineages have high fitness outside the host or neighboring plant species (non-host). In addition,
530 lack of data on host genotypes or field landscapes do now allow to evaluate whether host
531 heterogeneity and landscape heterogeneity in the field supports their co-existence (Papaix *et al.*,
532 2024). Another limitation of our approach which used pooled samples representative of a field for
533 metagenome sequencing, was that we couldn't exclude spatial partitioning or microhabitat
534 difference (for example, stem vs leaf vs fruit, leaves of different ages, or patchy distributions within
535 individual fields) that can support co-existence of two or more pathogen lineages. Resource
536 specialization cannot be ruled out either because the dominant lineages coexisting in the fields,
537 SC3 and SC4, differ in type III effector content and may differ in within-host multiplication rate
538 and transmission rate, or they may prefer different niches, such as leaf surface vs apoplast. Our
539 finding that sudden seasonal events or climatic fluctuations had a significant impact on defining
540 individual lineages' fitness contributions may potentially explain their coexistence. Two

541 independent approaches assessing compositional dynamics and fitness dynamics highlighted the
542 influence of environmental parameters on individual lineages, specifically SC3 and SC4,
543 specifically climatic fluctuations as indicated by skewness, kurtosis and standard deviation of
544 different parameters, being significant in dictating abundance of SC3 or SC4. We observed that
545 field-level fitness contribution by SC4 seemed to be impacted by other biotic (SC3) and abiotic
546 factors, whereas that by SC3 did not. The varying performance of SC4 depending on different
547 environmental conditions (from very fit to intermediately fit: fitness = 0) may contribute to the
548 similarity of its overall frequency with the SC3 frequency (more generalist with respect to the
549 values of tested factors and the presence of other SC). Thus, maintaining co-existence of lineages
550 that show overall environmental resilience as well as environmentally dependent fitness might be
551 a strategy that pathogen adopts to cope up with the sudden climatic shifts. All the arguments made
552 above that may contribute to explain the co-existence of SC3 and SC4 at relatively high and similar
553 levels, and the estimation of their relative fitness, can be summarized by making a parallel with
554 the effective reproduction number, Rt, classically considered in human and animal epidemiology.
555 Rt is the average number of new infections caused by a single infected individual at time t in the
556 partially susceptible population. Rt varies in time because of the variation of the number of
557 susceptible individuals but may also vary because of temporal heterogeneity in the eco-evo-
558 determinants of the transmission (e.g., because of lockdown measures during the covid-19
559 pandemic; (Roques *et al.*, 2020)). Rt can obviously be considered at the lineage level. Broadly, the
560 epidemic or the lineage dynamics tends to keep going on if Rt ≥ 1. In terms of fitness, as estimated
561 by the StrainRanking analysis in this study, it is equivalent to a fitness value ≥ 0 (the fitness here
562 is an indicator that integrates diverse processes and variables and is therefore an aggregated
563 measurement of the capacity of the lineage to "multiply" in real conditions). Thus, SC4 with fitness

564 >0 and SC3 with fitness of around 0 indicate Rt ≥ 1, suggesting that both SC3 and SC4 are fit to
565 keep epidemic going or to expand. As Rt evolves in time, the fitness (viewed as an aggregated
566 measurement) may also evolve in time, and an interesting research area would be to finely
567 characterize the eventual variation in time of the SC's fitness, evaluation of lineage-specific genes
568 in contribution towards fitness, and to determine how the possible processes evoked above
569 contribute to this variation. This is because possibility of transitory state with co-existence as
570 opposed to equilibrium state cannot be ruled out, especially, given the limited geographical region
571 studied here (Soubeyrand *et al.*, 2017) and that genetic divergence has been observed in this
572 pathosystem (Potnis et al. 2021). Thus, continuing such high-resolution lineage fitness tracking
573 across tomato producing regions and disease outcomes will be important outputs to predict future
574 outbreaks, and to design disease management strategies that are tailored to concurrent pathogen
575 population structure (i.e. breeding for disease resistance and climate resilience).

576 Although seasonality was not evident at the individual lineage level, analyzing the data via
577 variant tracing indicated the presence of season-specific allelic pathogen variants, particularly
578 unique to the summer vs fall growing season over the three consecutive years of sampling (Fig 5).
579 These seasonal variants might be indicative of a trade-off between in-season within-host
580 multiplication, in-season transmission, and inter-season survival, where variants that contribute as
581 primary inoculum (the form under which pathogen survives host absence) in fall vs summer might
582 differ due to the variable inter-season survival periods experienced (Van Den Berg *et al.*, 2010).
583 One approach to test this notion is to see if there is a trade-off between survival outside the host
584 (with varied overwintering durations) and infection efficiency (Penczykowski *et al.*, 2015).
585 Periodic host absence has been claimed to cause evolutionary divergence (also known as
586 evolutionary branching), however, contradictory evidence exists (van den Berg *et al.*, 2010, 2011;

587 Hamelin *et al.*, 2011). The absence of year-round production in Alabama, South Carolina, North
588 Carolina, and Georgia may drive evolutionary divergence. The seasonality could also mean
589 differing day lengths altering host physiological responses to the pathogen, thus impacting the
590 fitness contributions of different variants towards epidemics (Ballaré, 2014) or differing light
591 intensities impacting pathogen behavior and virulence, as seen with differing motilities and
592 differing epiphytic fitness in the case of *Pseudomonas syringae* strains (Beattie *et al.*, 2018). The
593 findings of seasonal oscillations of genome-wide allelic frequencies may suggest evidence of
594 fluctuation selection contributing towards the genetic variation observed in this population.

595 We also looked at signatures of parallel evolution by identifying loci under positive
596 selection across > 50% of the fields. Notably, we identified certain genes under positive selection
597 over three years, including the TonB-dependent receptor gene located adjacent to the PepSy
598 domain-containing protein. The widespread use of heavy fertilizers and pesticides in agricultural
599 systems (such as mancozeb applications to manage bacterial spot contain Mn and Zn) may have
600 been responsible for the selection of this TonB-dependent receptor likely involved in heavy metal
601 transport (Yeats *et al.*, 2004; Gallois *et al.*, 2022). Furthermore, multiple genes belonging to two
602 separate Type VI secretion system clusters were found to be under selection, suggesting that the
603 microbiome may play a significant role in driving the selection within the pathogen genome
604 (Coyne & Comstock, 2019; Allsopp *et al.*, 2020; Serapio-Palacios *et al.*, 2022).

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 during arms race. This work also presents
609 empirical evidence on the influence of recent climatic changes on altering disease outcomes. We

610 suggest that breeding programs need to consider multiple pathogen lineages that show positive
611 fitness contributions based on large scale field surveys. Tracing these evolutionary changes can
612 predict which lineages will be a better fit in the long run, especially in the face of sudden climatic
613 shifts, aiding in developing more effective breeding strategies. Future work based on the results of
614 this study should focus on identifying the roles of specific fitness factors in pathogen biology, as
615 this could provide valuable insights into pathogen adaptation strategies and the selection pressures
616 they face. 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) (Olm *et al.*, 2017) using a 95%
630 minimum genome completeness and 5% maximum contamination. We manually discarded the
631 genomes with more than 400 contigs to ensure better alignment with the reference genome and
632 avoid bias while creating the SNV profile. Four hundred sixty-seven quality-controlled genomes

633 collected from tomato and pepper (with one eggplant isolate) around the globe were individually
634 aligned against the completed genome of *Xp* strain LH3. Single-nucleotide polymorphisms (SNPs)
635 common to all genomes were extracted to generate a concatenated set of high-quality core-genome
636 SNPs using Parsnp (Treangen *et al.*, 2014). The phylogenetic tree was visualized and annotated
637 using iTol (v.6.7.5) (Letunic & Bork, 2007)

638 **Sample collection and disease severity estimation**

639 To study the diversity of BLS *Xanthomonas* in the southeastern United States, we collected
640 66 samples from 22 fields, designated as either commercial-scale or small-scale (Table S1). These
641 samples were collected from symptomatic and asymptomatic tomato plants grown in Alabama,
642 Georgia, South Carolina, and North Carolina over 2020, 2021, and 2022 (Fig. 1A). Samples were
643 taken during mid-season, end-season of the summer growing season, and when available, in the
644 fall growing season (Table S1A) to capture the temporal dynamics of the pathogen population.
645 While sampling, two to three leaflets were randomly taken from individual plants in the field using
646 variable sampling methods based on the scale of the field (Fig. 2B). Multiple plants were sampled
647 across the field to obtain a representative 50-100 grams leaf tissue sample. Assessments of the
648 disease were made by estimating the percent of disease symptoms and defoliation caused by
649 bacterial spots (Table S1B) using the Horsfall-Barratt scale, which ranges from 1 to 12, with scale
650 of 1 being no disease and 12 being 100% defoliation (Horsfall & Barratt, 1945).

651 **DNA extraction and sequencing**

652 Metagenomic DNA extraction, quantification, and sequencing of the samples were done
653 as described earlier (Newberry *et al.*, 2020b). Approximately 40 grams of the leaf sample was
654 taken in a Ziploc bag (Ziploc®) with 50 ml 0.1% wash buffer (0.05M PBS, 8.5 g NaCl, and 0.2

655 ml tween 20 per liter of water), sonicated for 20 minutes, followed by transfer of the buffer to a 50
656 ml falcon tube with the help of a sterile plastic serological pipette. The tubes were centrifuged at
657 4000 rpm, 4˚C for 20 minutes (Eppendorf® 5418 R) to collect all the cells. The cells were then
658 washed twice with sterile deionized water, and the metagenomic DNA extraction was done using
659 Wizard® Genomic DNA Purification Kit (Promega) as per the manufacturer's instructions with
660 the addition of a phenol-chloroform step to remove protein contamination. The extracted DNA
661 was quantified using a Qubit® fluorometer and dsDNA high-sensitivity assay kit (Life
662 Technologies, Carlsbad, CA, USA) and kept at -80^{0}C until further processing. Sequencing was
663 performed on the Illumina HiSeq 4000 platform. The raw reads were processed and trimmed for
664 the adapter using BBDuk (Bushnell, 2014). DynamicTrim command of SolexaQa was used to trim
665 the reads based on quality values with Qphred less than 20, followed by filtering the reads shorter
666 than 50bp (Cox et al. 2010). Host reads were separated and removed using tomato cv. Heinz 1706
667 as a reference (GCA_022405115.1) from the metagenomic samples using KneadData tools
668 (https://bitbucket.org/biobakery/kneaddata/wiki/Home) as a reference.

669 **Taxonomic profiling and diversity estimation**

670 The relative abundance of *Xanthomonas* sp. in quality-controlled and host-decontaminated
671 reads was performed using Kraken2 (v2.1.2) (Wood *et al.*, 2019) against a standard Kraken2
672 standard database. The database contained RefSeq libraries of archaeal, bacterial, human, and viral
673 sequences as of March 1, 2022 (O'Leary *et al.*, 2016). The resulting Kraken2 report files were
674 used as inputs for Bayesian re-estimation of abundance using Kraken (Bracken) (v2.6.2) (Lu *et*
675 *al.*, 2017). We also calculated the absolute abundance estimates for *Xp* by combining their relative
676 abundance with the total DNA content in each sample. Microbiota density, expressed as total DNA
677 (ng) per mg of fresh sample, was computed for each sample. This information was utilized to

678 determine the absolute abundance of *Xp* by multiplying the amount of DNA per mg of the sample
679 with its relative abundance.

680 To explore the intraspecific diversity and dynamics of *Xp* in the metagenomic samples, we
681 employed the StrainEst pipeline (Albanese & Donati, 2017). StrainEst is a reference-based method
682 that utilizes single-nucleotide variant (SNV) profiles from carefully selected reference genomes.
683 This approach enables us to estimate strains' presence and relative abundance within a given
684 sample. Initially, we conducted SNV profiling of representative strains from *Xp*, *Xeu*, and other
685 closely related pathovars belonging to *Xeu* species complex to determine their presence. A second
686 round of profiling was done to identify the diversity of Xp strains. A representative genome from
687 each *Xp* SC (Fig. 2) was selected and aligned against the completed reference genome of *Xp* strain
688 91-118 using the MUMmer algorithm (Kurtz *et al.*, 2004). This genome alignment was used to
689 construct a SNV matrix, where each row corresponded to a variable position in the reference
690 genome, and columns contained allelic variants present in the reference strains. This matrix was
691 used as a reference in modeling strain level abundance.

692 To assess the diversity of *Xp* SCs within the samples, we computed the Shannon diversity
693 index (Shannon-Wiener diversity index) using the "diversity()" function from the R package *vegan*
694 (v2.6-4) (Dixon, 2003). The index considers the count of SCs present (richness) and their
695 proportional distribution (evenness). A higher index value indicates greater species diversity in the
696 habitat. When the Shannon diversity index of a sample is equal to 0, it implies that only a single
697 SC is detected in that sample.

698 **Weather data**

699 Daily climate data spanning three years for all 22 fields were compiled from the NASA
700 Power database using the nasapower (v4.0.12) package (Sparks, 2018) in R. Data were collected

during mid-season, end-season of summer, and fall samples as per the sampling time in Table S1. The gridded weather data used in this analysis are based on satellite observations and models and are offered globally at a horizontal resolution of 0.5° X 0.625° latitude–longitude grid cell. Weather variables, such as temperature, precipitation, solar radiation, relative humidity, wind, and others (Table S2), were consolidated over time and space to generate data points for each sampling time, resulting in an average value. The average for mid-season was calculated as an average from the time of planting to mid-season sampling time, which is usually 60 days after transplanting, and the average for end-season is calculated as the average from the time of planting to end-of-season sampling time, which is generally 30 days after mid-season sampling.

Variability among the weather parameters was measured using descriptive statistics (standard deviation, skewness, entropy, and kurtosis) for all the climatic variables to see if the extreme temperature events during the growing season impact disease severity, pathogen diversity, and abundance. Standard deviation measures the deviation from the mean value. A higher standard deviation indicates greater weather parameter variability, while a lower standard deviation suggests a more consistent pattern. Skewness measures the degree and direction of asymmetry. A positive skewness implies more extreme high weather parameters, while a negative skewness suggests more extreme low, with zero suggesting a symmetric distribution. Kurtosis is a measure of tail extremity reflecting the presence of outliers in a distribution (Westfall, 2014). Higher positive kurtosis implies extreme weather values or outliers, while negative kurtosis indicates a flatter distribution with less extreme weather values. Similarly, entropy quantifies the level of unpredictability in the weather values. The higher the entropy, the more unpredictability in the weather fluctuations will be.

724 **Statistical analysis via regression models**

725 We quantified the relative abundance of various *Xp* lineages across the samples and analyzed their
726 distribution patterns in relation to different climatic factors. This analysis was designed to explain
727 three different types of responses based on climatic and other variables: (i) relative abundance, (ii)
728 absolute abundance, and (iii) disease severity. Before running these regressions, we chose to run a
729 predictor-selection procedure for each of these (given the high number of predictors we considered
730 for this work) by using Least Absolute Shrinkage and Selection Operator (Lasso) regression, which
731 jointly aims at correctly predicting the response and selecting the most associated variables
732 (Tibshirani, 1996). As no implemented or stable versions of the Lasso directly address the types
733 of responses we have considered in our work, we decided to use proxy responses to run the Lasso
734 and then use the selected variables from this step to run the specific regressions and inference
735 described above. More specifically, we use a logistic Lasso to select the variables for the (relative
736 abundance) where we classify low and high abundance based on the 50% threshold. In contrast,
737 we use standard Lasso regression on the severity scales (assuming they can be considered as
738 continuous) and the Shannon entropy (a continuous scale proxy for the distribution of abundance
739 between clusters). Predictor-selection was performed on the variables that still have non-zero
740 coefficients after the shrinkage process using the "cv.glmnet" function to prevent the overfitting
741 of the model. This study determined the tuning parameter for regularization using 10-fold cross-
742 validation. The path length (min_lambda/max-lambda) was chosen to be 0.001, and 100 default
743 values along the regularization path were tested to find the best lambda value. Once a selection
744 was run for all responses (proxies) of interest, we ran the appropriate regressions for each response.
745 More specifically, since disease severity is measured on a discrete ordered scale in the 0-12 range,
746 we used an ordinal logistic regression, which accounts for the order of severity and the discrete

747 scale of the response (McCullagh, 1980). On the other hand, as the abundance (both absolute and
748 relative) of *Xp* is measured on a 0-1 range, we employed a Beta regression (Ferrari & Cribari-Neto,
749 2004) using the *betareg* package (Cribari-Neto & Zeileis, 2010). The beta regression model is
750 employed for non-normally distributed data with a range between 0 and 1, addressing
751 heteroskedasticity and asymmetry issues (Ferrari & Cribari-Neto, 2004), and is specifically chosen
752 here due to the non-normal distribution of *Xp* relative abundance in the sample, which conforms
753 to a beta distribution. To analyze the distribution of abundance among different Xp lineages, we
754 used a compositional regression approach, which ensures that the regression model explains and
755 preserves the property that the total abundance among the clusters should sum to one (indeed, a
756 separate regression for each cluster would not consider the dependence between clusters). We used
757 the Dirichlet regression approach, using the *DirichletReg* package in R, to describe how the
758 distribution of various *Xp* lineages changes according to changes in these potential predictors
759 (Hijazi & Jernigan, 2009). As for the previous response variables, the use of the Lasso for this type
760 of regression is computationally unstable. Therefore, we chose to use the Shannon diversity of *Xp*
761 as the response when using the Lasso: this is a reasonable proxy since Shannon diversity is related
762 to how the data is distributed (in this case, across the lineages). Once this was obtained, 18
763 variables were selected to run the Dirichlet regression.

764 **Greenhouse Experiment**

765 We evaluated the effects of different SCs of *Xp* on 4-5 week old tomato cultivar FL47 plants grown
766 under greenhouse conditions. Plants were dip-inoculated for approximately 30 seconds in bacterial
767 suspensions containing 10^6cfu/ml of representative strains from different SCs (SC3 (Xp2010), SC4
768 (AL57), SC5 (AL37), and SC6 (AL65)). These SCs were chosen based on their observed
769 abundance in samples from our previous study (Newberry et al., 2020). Treatments included a

770 control (0.01M $MgSO_4$), individual inoculations with SC3, SC4, SC5, and SC6, as well as mixed
771 co-infections of SC4 + SC3, SC5 + SC6, SC3 + SC6, SC3 + SC5, and SC3 + SC5 + SC6. All
772 treatments were supplemented with 0.00025% Swilvet and replicated five times, with plants
773 randomly arranged in the greenhouse. Plants were watered daily, and disease severity was assessed
774 on days 7 and 14, with additional evaluations on days 9 and 12 for plants exhibiting delayed
775 symptoms. Disease severity was measured using the Horsfall-Barratt scale (Hollis, 1984). To
776 minimize bias, two observers recorded the disease severity, and the average score was used. Raw
777 disease severity indices were then used to calculate the area under the disease progress curve
778 (AUDPC), which correlates multiple observations of disease progression into a single value
779 (Simko & Piepho, 2012). AUDPC values were plotted for each treatment across seven
780 experimental batches using the "dplyr" package in R Studio and built linear and linear mixed-
781 effects models using "lme4" and "lmerTest" (Bates, D et al., 2007; Kuznetsova et al., 2017). For
782 individual strains a generalized linear hypothesis test was used while for multiple comparisons
783 Tukey's method was used. Linear mixed effect models were used to assess the development of
784 disease severity slopes across treatments.

785 **Strain Ranking**

786 StrainRanking was used to estimate and compare fitness differences between SCs (main effects),
787 between SCs considered alone or in association (pathogen-pathogen interaction), and between SCs
788 under specific environmental conditions (pathogen-environment interaction). StrainRanking
789 requires input data consisting of the frequencies of the SCs and the measurement of disease growth
790 or decline for each sampling location (i.e., fields), and estimates the fitness as the relative
791 contributions of SCs to the growth or the decline of disease severity between two sampling times
792 (i.e., Mid and End sampling times). SC frequencies were computed as a weighted mean of Mid

793 and End SC frequencies: $p \times$ Mid frequencies $+ (1 - p) \times$ End frequencies, where p is equal to
794 0.1, 0.5 or 0.9 to consider different representativeness of Mid and End frequencies in terms of
795 disease dynamics between Mid and End sampling times of summer season. (Because of low
796 sample size, we did not include Fall season samples in this analysis.) Significance tests performed
797 in StrainRanking are based on randomization, which adequately accounts for the relatively small
798 sample sizes encountered when distributing data at the resolution of SCs or even at a higher
799 resolution (when one considers pathogen-pathogen or pathogen-environment interactions). To
800 account for the specificity of the severity measurement in the present study, which is a mark
801 ranging between 0 and 8, we added to the StrainRanking method a saturation level of the growth
802 and the decline at +8 and -8, respectively. In addition, we developed a function providing
803 confidence intervals for estimated fitness in addition to the pairwise comparison tests embedded
804 in the original StrainRanking R package.

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

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

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

825 **Construction of *Xanthomonas* pan-genome**

826 Initially, we downloaded 287 strains of *Xanthomonas euvesicatoria* pv. *perforans (Xp)* and
827 *Xanthomonas euvesicatoria* pv. *euvesicatoria (Xeu)* from the NCBI RefSeq database (as of
828 February 16th, 2022). Using SuperPang (v0.9.4 beta1) (Puente-Sánchez *et al.*, 2022), we
829 constructed a non-redundant pangenome as a reference for the subsequent analyses.
830 To avoid read stealing and donating of *Xanthomonas* with other microbiome species, we
831 implemented an additional filter to remove genes with more than 97% similarity in our
832 metagenomic reads. These genes were identified similarly to the method described by Kaur et al.,
833 2024. We used the "run_midas.py species" command to identify the predominant species in our
834 samples, utilizing the MIDAS (v1.3.2) default database as a reference, which consists of 31,007
835 bacterial reference genomes organized into 5,952 species groups (Nayfach, 2022). We then built
836 a custom database of non-redundant pangenomes (*Xp* and *Xeu*) using MIDAS and subjected it to
837 BLAST (blast+) with the concatenated gene sequences from species with a mean abundance > 0

838 from the previous step. Genes with 97% similarity to those in the database were removed from all
839 samples using BBDuk (v37.36) (Bushnell, 2020). These reads were used for our subsequent
840 analyses.
841
842 **Genes under selection**
843 We discarded the samples with zero *Xp* abundance in the following evolutionary analyses. To
844 calculate the genetic variation in our samples, we first mapped the samples to the non-redundant
845 pangenome using BWA-MEM (v0.7.12) (Li, 2013). Next, we removed low-quality alignments
846 and duplicate reads using samtools (v1.11) (Li *et al.*, 2009) and picard (v1.79)
847 (https://broadinstitute.github.io/picard/) , respectively. The resulting bam files were then processed
848 with MetaPop (v1.0) (Gregory *et al.*, 2022) to calculate the ratio of non-synonymous to
849 synonymous polymorphisms (pN/pS) for each gene. In order to find the genes under positive
850 selection, we chose the genes with pN/pS ratios >1 and present in at least 50% of the samples in a
851 particular season.
852
853 **Seasonal variants**
854 To calculate the allele frequencies of SNV sites across the pangenome, we used the MIDAS
855 "midas.py snps" script (Nayfach, 2022). We treated all samples from different states in the
856 southeastern USA as a single population for each season. To increase the accuracy of results, we
857 had additional filters for removing SNV sites with a read depth of less than 10. For seasonal
858 oscillations, we averaged the allele frequencies of major alleles during the season. To identify
859 parallel variants in the population, we first extracted SNV sites where the frequency of minor
860 alleles was less than 0.2 (considered minor alleles) and changed to greater than 0.8 (considered

major alleles) in the following season and remained major in subsequent seasons. For example, we identified sites with minor alleles (f <= 0.2) during mid-season 2020 that became major alleles (f>= 0.8) by the end of the 2020 season. We retained only those SNVs that exhibited this trend in at least 50% of the samples in a given season.

Acknowledgments

The climatic data were obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program. This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority. We thank growers in the southeastern US for supporting 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.

907 **References**

908 **Albanese D, Donati C**. **2017**. Strain profiling and epidemiology of bacterial species from
909 metagenomic sequencing. *Nature Communications* **8**: 2260.

910 **Allsopp LP, Bernal P, Nolan LM, Filloux A**. **2020**. Causalities of war: The connection between
911 type VI secretion system and microbiota. *Cellular Microbiology* **22**: e13153.

912 **Anderson JP, Gleason CA, Foley RC, Thrall PH, Burdon JB, Singh KB**. **2010**. Plants versus
913 pathogens: an evolutionary arms race. *Functional plant biology : FPB* **37**: 499–512.

914 **Ballaré CL**. **2014**. Light regulation of plant defense. *Annual Review of Plant Biology* **65**: 335–363.

915 **Bartoli C, Roux F, Lamichhane JR**. **2016**. Molecular mechanisms underlying the emergence of
916 bacterial pathogens: an ecological perspective. *Molecular Plant Pathology* **17**: 303–310.

917 **Beattie GA, Hatfield BM, Dong H, McGrane RS**. **2018**. Seeing the Light: The Roles of Red- and
918 Blue-Light Sensing in Plant Microbes. *Annual Review of Phytopathology* **56**: 41–66.

919 **van den Berg F, Bacaer N, Metz JAJ, Lannou C, van den Bosch F**. **2011**. Periodic host absence
920 can select for higher or lower parasite transmission rates. *Evolutionary Ecology* **25**: 121–137.

921 **van den Berg F, Gilligan CA, Bailey DJ, van den Bosch F**. **2010**. Periodicity in host availability
922 does not account for evolutionary branching as observed in many plant pathogens: an
923 application to Gaeumannomyces graminis var. tritici. *Phytopathology* **100**: 1169–1175.

924 **Bhandari R, Sanz-Saez A, Leisner CP, Potnis N**. **2023a**. Xanthomonas infection and ozone stress
925 distinctly influence the microbial community structure and interactions in the pepper
926 phyllosphere. *ISME Communications* **3**: 1–13.

927 **Bhandari R, Sanz-Saez A, Leisner CP, Potnis N**. **2023b**. Xanthomonas infection and ozone stress
928 distinctly influence the microbial community structure and interactions in the pepper
929 phyllosphere. *ISME Communications* **3**: 1–13.

930 **Bilgin DD, Aldea M, O'Neill BF, Benitez M, Li M, Clough SJ, DeLucia EH**. **2008**. Elevated Ozone
931 Alters Soybean-Virus Interaction. *Molecular Plant-Microbe Interactions®* **21**: 1297–1308.

932 **Brasier CM**. **1996**. Phytophthora cinnamomi and oak decline in southern Europe. Environmental
933 constraints including climate change. *Annales des Sciences Forestières* **53**: 347–358.

934 **Bushnell B**. **2020**. BBDuk. *Jt Genome Inst. Available online: https://jgi. doe. gov/data-and-*
935 *tools/bbtools/bb-tools-userguide/bbduk-guide/(accessed on 25 June 2020)*.

936 **Byappanahalli MN, Whitman RL, Shively DA, Sadowsky MJ, Ishii S**. **2006**. Population structure,
937 persistence, and seasonality of autochthonous Escherichia coli in temperate, coastal forest soil
938 from a Great Lakes watershed. *Environmental Microbiology* **8**: 504–513.

Castledine M, Padfield D, Sierocinski P, Soria Pascual J, Hughes A, Mäkinen L, Friman V-P, Pirnay J-P, Merabishvili M, de Vos D, *et al.* 2022. Parallel evolution of Pseudomonas aeruginosa phage resistance and virulence loss in response to phage treatment in vivo and in vitro. *eLife* **11**: e73679.

Chaloner TM, Gurr SJ, Bebber DP. **2021**. Plant pathogen infection risk tracks global crop yields under climate change. *Nature Climate Change* **11**: 710–715.

Corredor-Moreno P, Saunders DGO. **2020**. Expecting the unexpected: factors influencing the emergence of fungal and oomycete plant pathogens. *New Phytologist* **225**: 118–125.

Coyne MJ, Comstock LE. **2019**. Type VI Secretion Systems and the Gut Microbiota. *Microbiology Spectrum* **7**: 10.1128/microbiolspec.psib-0009–2018.

Cribari-Neto F, Zeileis A. **2010**. Beta Regression in R. *Journal of Statistical Software* **34**: 1–24.

Desaint H, Aoun N, Deslandes L, Vailleau F, Roux F, Berthomé R. **2021**. Fight hard or die trying: when plants face pathogens under heat stress. *New Phytologist* **229**: 712–734.

Dixon P. **2003**. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**: 927–930.

Dixon GR. **2012**. Climate change – impact on crop growth and food production, and plant pathogens. *Canadian Journal of Plant Pathology* **34**: 362–379.

El Jarroudi M, Karjoun H, Kouadio L, El Jarroudi M. **2020**. Mathematical modelling of non-local spore dispersion of wind-borne pathogens causing fungal diseases. *Applied Mathematics and Computation* **376**: 125107.

Elmer WH. **2001**. Seeds as vehicles for pathogen importation. *Biological Invasions* **3**: 263–271.

Ferrari S, Cribari-Neto F. **2004**. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics* **31**: 799–815.

Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. **2020**. Threats to global food security from emerging fungal and oomycete crop pathogens. *Nature Food* **1**: 332–342.

Gallois N, Alpha-Bazin B, Bremond N, Ortet P, Barakat M, Piette L, Mohamad Ali A, Lemaire D, Legrand P, Theodorakopoulos N, *et al.* 2022. Discovery and characterization of UipA, a uranium- and iron-binding PepSY protein involved in uranium tolerance by soil bacteria. *The ISME journal* **16**: 705–716.

Gao M, Xiong C, Gao C, Tsui CKM, Wang M-M, Zhou X, Zhang A-M, Cai L. **2021**. Disease-induced changes in plant microbiome assembly and functional adaptation. *Microbiome* **9**: 187.

970 **Gause GF**. **1934**. Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle
971 for Existence. *Science* **79**: 16–17.

972 **Gregory AC, Gerhardt K, Zhong Z-P, Bolduc B, Temperton B, Konstantinidis KT, Sullivan MB**.
973 **2022**. MetaPop: a pipeline for macro- and microdiversity analyses and visualization of microbial
974 and viral metagenome-derived populations. *Microbiome* **10**: 49.

975 **Hamelin FM, Castel M, Poggi S, Andrivon D, Mailleret L**. **2011**. Seasonality and the evolutionary
976 divergence of plant parasites. *Ecology* **92**: 2159–2166.

977 **Hijazi RH, Jernigan RW**. **2009**. Modelling Compositional Data Using Dirichlet Regression Models.
978 **4**: 77–91.

979 **Horsfall JG, Barratt RW**. **1945**. An improved grading system for measuring plant diseases.
980 *Phytopathology* **35**: 655.

981 **Hu L, Yang L**. **2019**. Time to Fight: Molecular Mechanisms of Age-Related Resistance.
982 *Phytopathology®* **109**: 1500–1508.

983 **Janda M, Lamparová L, Zubíková A, Burketová L, Martinec J, Krčková Z**. **2019**. Temporary heat
984 stress suppresses PAMP-triggered immunity and resistance to bacteria in Arabidopsis thaliana.
985 *Molecular Plant Pathology* **20**: 1005–1012.

986 **Jang J, Di DYW, Lee A, Unno T, Sadowsky MJ, Hur H-G**. **2014**. Seasonal and Genotypic Changes
987 in Escherichia coli Phylogenetic Groups in the Yeongsan River Basin of South Korea. *PLOS ONE* **9**:
988 e100585.

989 **Jang J, Unno T, Lee SW, Cho KH, Sadowsky MJ, Ko G, Kim JH, Hur H-G**. **2011**. Prevalence of
990 season-specific Escherichia coli strains in the Yeongsan River Basin of South Korea.
991 *Environmental Microbiology* **13**: 3103–3113.

992 **Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW**. **2004**. Reclassification of the Xanthomonads
993 Associated with Bacterial Spot Disease of Tomato and Pepper. *Systematic and Applied*
994 *Microbiology* **27**: 755–762.

995 **Karasov TL, Horton MW, Bergelson J**. **2014**. Genomic variability as a driver of plant-pathogen
996 coevolution? *Current opinion in plant biology* **18**: 24–30.

997 **Kareiva P**. **1999**. Coevolutionary arms races: Is victory possible? *Proceedings of the National*
998 *Academy of Sciences* **96**: 8–10.

999 **Kaur A, Russell I, Liu R, Holland A, Bhandari R, Potnis N**. **2024**. Within-host adaptation of a
1000 foliar pathogen, Xanthomonas, on pepper in presence of quantitative resistance and ozone
1001 stress. : 2024.02.29.582737.

1002 **Klein-Gordon JM, Timilsina S, Xing Y, Abrahamian P, Garrett KA, Jones JB, Vallad GE, Goss EM**.
1003 **2022**. Whole genome sequences reveal the Xanthomonas perforans population is shaped by
1004 the tomato production system. *The ISME journal* **16**: 591–601.

1005 **Klein-Gordon JM, Xing Y, Garrett KA, Abrahamian P, Paret ML, Minsavage GV, Strayer-Scherer**
1006 **AL, Fulton JC, Timilsina S, Jones JB, *et al.* 2021**. Assessing Changes and Associations in the
1007 Xanthomonas perforans Population Across Florida Commercial Tomato Fields Via a Statewide
1008 Survey. *Phytopathology* **111**: 1029–1041.

1009 **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL**. **2004**.
1010 Versatile and open software for comparing large genomes. *Genome biology* **5**: R12.

1011 **Lahlali R, Taoussi M, Laasli S-E, Gachara G, Ezzouggari R, Belabess Z, Aberkani K, Assouguem**
1012 **A, Meddich A, El Jarroudi M, *et al.* 2024**. Effects of climate change on plant pathogens and
1013 host-pathogen interactions. *Crop and Environment* **3**: 159–170.

1014 **Letunic I, Bork P**. **2007**. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree
1015 display and annotation. *Bioinformatics (Oxford, England)* **23**: 127–128.

1016 **Li H**. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

1017 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,**
1018 **1000 Genome Project Data Processing Subgroup**. **2009**. The Sequence Alignment/Map format
1019 and SAMtools. *Bioinformatics* **25**: 2078–2079.

1020 **Lu J, Breitwieser FP, Thielen P, Salzberg SL**. **2017**. Bracken: estimating species abundance in
1021 metagenomics data. *PeerJ Computer Science* **3**: e104.

1022 **Ma W, Dong FFT, Stavrinides J, Guttman DS**. **2006**. Type III Effector Diversification via Both
1023 Pathoadaptation and Horizontal Transfer in Response to a Coevolutionary Arms Race. *PLOS*
1024 *Genetics* **2**: e209.

1025 **McCullagh P**. **1980**. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society:*
1026 *Series B (Methodological)* **42**: 109–127.

1027 **McDonald BA, Linde C**. **2002**. Pathogen population genetics, evolutionary potential, and
1028 durable resistance. *Annual Review of Phytopathology* **40**: 349–379.

1029 **McDonald BA, Stukenbrock EH**. **2016**. Rapid emergence of pathogens in agro-ecosystems:
1030 global threats to agricultural sustainability and food security. *Philosophical transactions of the*
1031 *Royal Society of London. Series B, Biological sciences* **371**.

1032 **Myers TA, Scott RC, Zelinka MD, Klein SA, Norris JR, Caldwell PM**. **2021**. Observational
1033 constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nature Climate*
1034 *Change* **11**: 501–507.

1035 **Nayfach S**. **2022**. Metagenomic Intra-Species Diversity Analysis System (MIDAS).

1036 **Newberry E, Bhandari R, Kemble J, Sikora E, Potnis N**. **2020a**. Genome-resolved metagenomics
1037 to study co-occurrence patterns and intraspecific heterogeneity among plant pathogen
1038 metapopulations. *Environmental Microbiology* **22**: 2693–2708.

1039 **Newberry E, Bhandari R, Kemble J, Sikora E, Potnis N**. **2020b**. Genome-resolved metagenomics
1040 to study co-occurrence patterns and intraspecific heterogeneity among plant pathogen
1041 metapopulations. *Environmental Microbiology* **22**: 2693–2708.

1042 **Newberry EA, Bhandari R, Minsavage GV, Timilsina S, Jibrin MO, Kemble J, Sikora EJ, Jones JB,**
1043 **Potnis N**. **2019**. Independent Evolution with the Gene Flux Originating from Multiple
1044 Xanthomonas Species Explains Genomic Heterogeneity in Xanthomonas perforans. *Applied and*
1045 *Environmental Microbiology* **85**: e00885-19.

1046 **O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B,**
1047 **Smith-White B, Ako-Adjei D, *et al.* 2016**. Reference sequence (RefSeq) database at NCBI:
1048 current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**:
1049 D733-745.

1050 **Olm MR, Brown CT, Brooks B, Banfield JF**. **2017**. dRep: a tool for fast and accurate genomic
1051 comparisons that enables improved genome recovery from metagenomes through de-
1052 replication. *The ISME Journal* **11**: 2864.

1053 **P. S. Teng, W.W. Shane, David R. MacKenzie**. **1984**. Crop losses due to plant pathogens: Critical
1054 Reviews in Plant Sciences: Vol 2, No 1.

1055 **Papaix J, Carpentier F, Monod H, Lannou C**. **2024**. The role of spatial heterogeneity in shaping
1056 pathogen populations in agricultural landscapes.

1057 **Penczykowski RM, Walker E, Soubeyrand S, Laine A-L**. **2015**. Linking winter conditions to
1058 regional disease dynamics in a wild plant–pathogen metapopulation. *New Phytologist* **205**:
1059 1142–1152.

1060 **Potnis N**. **2021**. Harnessing Eco-Evolutionary Dynamics of Xanthomonads on Tomato and
1061 Pepper to Tackle New Problems of an Old Disease. *Annual Review of Phytopathology* **59**: 289–
1062 310.

1063 **Puente-Sánchez F, Hoetzinger M, Buck M, Bertilsson S**. **2022**. Exploring intra-species diversity
1064 through non-redundant pangenome assemblies. : 2022.03.25.485477.

1065 **Roques L, Klein EK, Papaïx J, Sar A, Soubeyrand S**. **2020**. Impact of Lockdown on the Epidemic
1066 Dynamics of COVID-19 in France. *Frontiers in Medicine* **7**.

1067 **Sadhukhan S, Jacques M-A, Potnis N**. **2024**. Influence of Co-occurring Weakly Pathogenic
1068 Bacterial Species on Bacterial Spot Disease Dynamics on Tomato. *Plant Disease* **108**: 190–199.

Savary S, Ficke A, Aubertot J-N, Hollier C. **2012**. Crop losses due to diseases and their implications for global food production losses and food security. *Food Security* **4**: 519–537.

Scherm,H., Yang, X. B. 1995. Interannual variations in wheat rust development in China and the United States in relation to the El Niño/Southern Oscillation. *Phytopathology, (970–976)*.

Serapio-Palacios A, Woodward SE, Vogt SL, Deng W, Creus-Cuadros A, Huus KE, Cirstea M, Gerrie M, Barcik W, Yu H, *et al.* 2022. Type VI secretion systems of pathogenic and commensal bacteria mediate niche occupancy in the gut. *Cell Reports* **39**: 110731.

Sharabani G, Shtienberg D, Borenstein M, Shulhani R, Lofthouse M, Sofer M, Chalupowicz L, Barel V, Manulis-Sasson S. **2013**. Effects of plant age on disease development and virulence of Clavibacter michiganensis subsp. michiganensis on tomato. *Plant Pathology* **62**: 1114–1122.

Soubeyrand S, Garreta V, Monteil C, Suffert F, Goyeau H, Berder J, Moinard J, Fournier E, Tharreau D, Morris CE, *et al.* 2017. Testing Differences Between Pathogen Compositions with Small Samples and Sparse Data. *Phytopathology®* **107**: 1199–1208.

Sparks AH. **2018**. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R [Journal of Open Source Software, vol. 3, no. 30, 1035.

Sundin GW, Wang N. **2018**. Antibiotic Resistance in Plant-Pathogenic Bacteria. *Annual Review of Phytopathology* **56**: 161–180.

Thompson JN, Burdon JJ. **1992**. Gene-for-gene coevolution between plants and parasites. *Nature* **360**: 121–125.

Tibshirani R. **1996**. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**: 267–288.

Timilsina S, Kara S, Jacques MA, Potnis N, Minsavage GV, Vallad GE, Jones JB, Saux MFL. **2019**. Reclassification of Xanthomonas gardneri (ex Šutič 1957) Jones et al. 2006 as a later heterotypic synonym of Xanthomonas cynarae Trébaol et al. 2000 and description of X. cynarae pv. cynarae and X. cynarae pv. gardneri based on whole genome analyses. *International Journal of Systematic and Evolutionary Microbiology* **69**: 343–349.

Treangen TJ, Ondov BD, Koren S, Phillippy AM. **2014**. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology* **15**: 524.

Van Den Berg F, Gilligan CA, Bailey DJ, Van Den Bosch F. **2010**. Periodicity in Host Availability Does Not Account for Evolutionary Branching as Observed in Many Plant Pathogens: An Application to *Gaeumannomyces graminis* var. *tritici*. *Phytopathology®* **100**: 1169–1175.

Westfall PH. **2014**. Kurtosis as Peakedness, 1905 - 2014. R.I.P. *The American Statistician* **68**: 191–195.

1103 **Whalen MC**. **2005**. Host defence in a developmental context. *Molecular Plant Pathology* **6**:
1104 347–360.

1105 **Wichmann G, Ritchie D, Kousik CS, Bergelson J**. **2005**. Reduced genetic variation occurs among
1106 genes of the highly clonal plant pathogen Xanthomonas axonopodis pv. vesicatoria, including
1107 the effector gene avrBs2. *Applied and environmental microbiology* **71**: 2418–32.

1108 **Wood DE, Lu J, Langmead B**. **2019**. Improved metagenomic analysis with Kraken 2. *Genome*
1109 *Biology* **20**: 257.

1110 **Yeats C, Rawlings ND, Bateman A**. **2004**. The PepSY domain: a regulator of peptidase activity in
1111 the microbial environment? *Trends in Biochemical Sciences* **29**: 169–172.

1112 **Zhan J, Mundt CC, Hoffer ME, McDonald BA**. **2002**. Local adaptation and effect of host
1113 genotype on the rate of pathogen evolution: an experimental test in a plant pathosystem.
1114 *Journal of Evolutionary Biology* **15**: 634–647.

1115