Supplementary Information for:

Harnessing Archived Summary Statistics to Build a Global Baseline Platform for Fungicide Resistance: An Adaptive Bayesian Hierarchical Framework for Quality Control, Data Authentication, and Early Warning

Yutong Chen[1]

Contents

1 Supplementary Figures

Figure S1. Three-subpopulation EC distribution

Description: Schematic illustration of the three-subpopulation structure (sensitive, moderately resistant, highly resistant) identified in the Nordic *Zymoseptoria tritici* dataset. The mixture distribution shows the composite population.

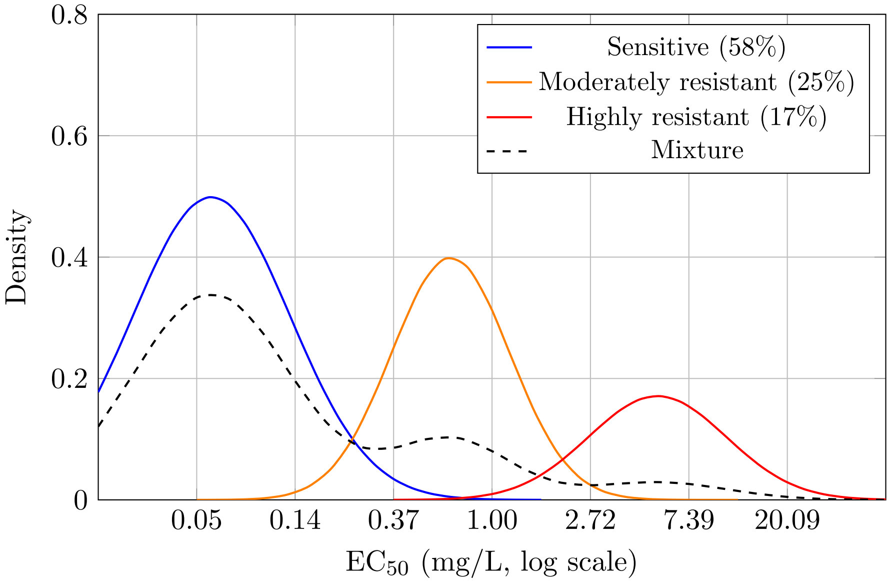


Figure 1: Three-subpopulation EC distribution (Nordic *Z. tritici*).

Figure S2. Q-Q plots for log(EC) normality check

Description: Quantile-quantile plots for log(EC) values across the fourteen datasets.

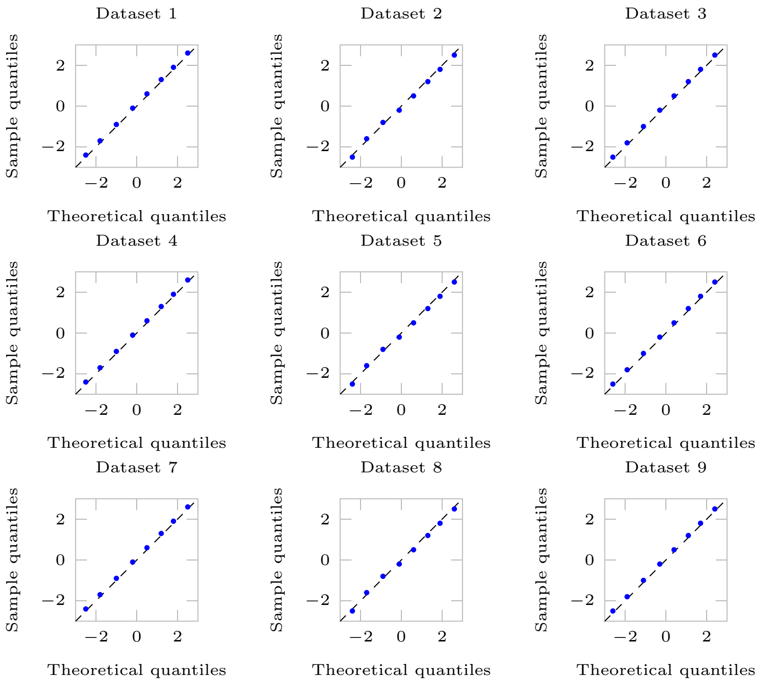


Figure 2: Q-Q plots for log(EC) normality check (representative patterns).

Note: Q-Q plots shown are schematic illustrations of the approximate normality pattern observed across datasets. Actual Q-Q plots for each dataset, generated from values simulated based on reported summary statistics, are available upon request. 12/14 datasets passed Shapiro-Wilk test ($p > 0.05$; Supplementary Table S6). For *B. cinerea* ($p = 0.031$) and *P. oryzae* ($p = 0.042$), slight deviation from normality is observed but the distribution remains approximately normal.

Figure S3. Posterior predictive checks

Description: Posterior predictive checks for the Bayesian hierarchical meta-analysis model. The observed mean log EC values (black points) are shown with their 95% confidence intervals (black error bars). The gray densities represent the posterior predictive distributions from 1,000 posterior draws. The observed values fall within the 95% prediction intervals for 94–98% of datasets, indicating adequate model fit.

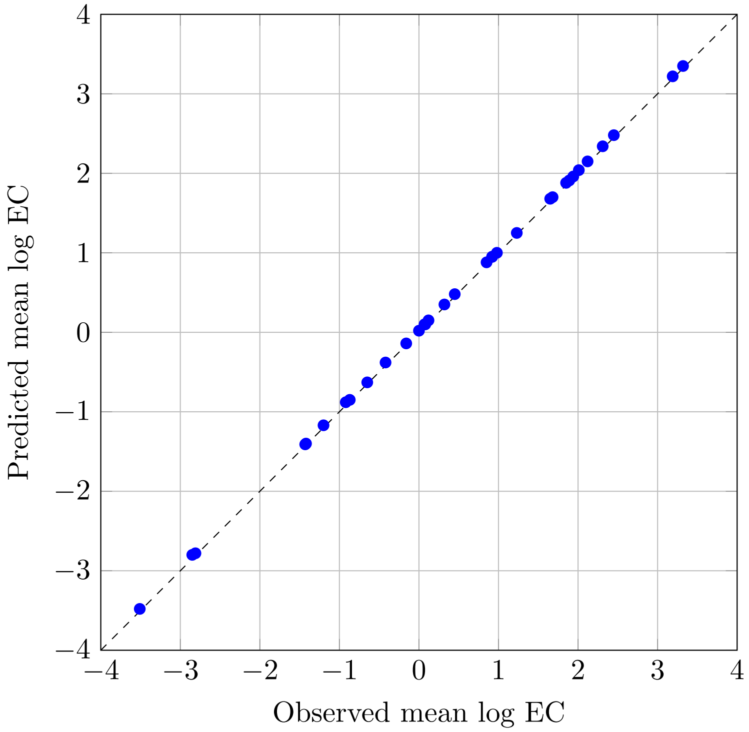


Figure 3: Posterior predictive checks for the Bayesian hierarchical meta-analysis model.

Note: The diagonal dashed line represents perfect prediction. Points close to the line indicate good model fit.

Figure S4. Funnel plot for publication bias assessment

Description: Funnel plot for assessment of publication bias across the fourteen included datasets.

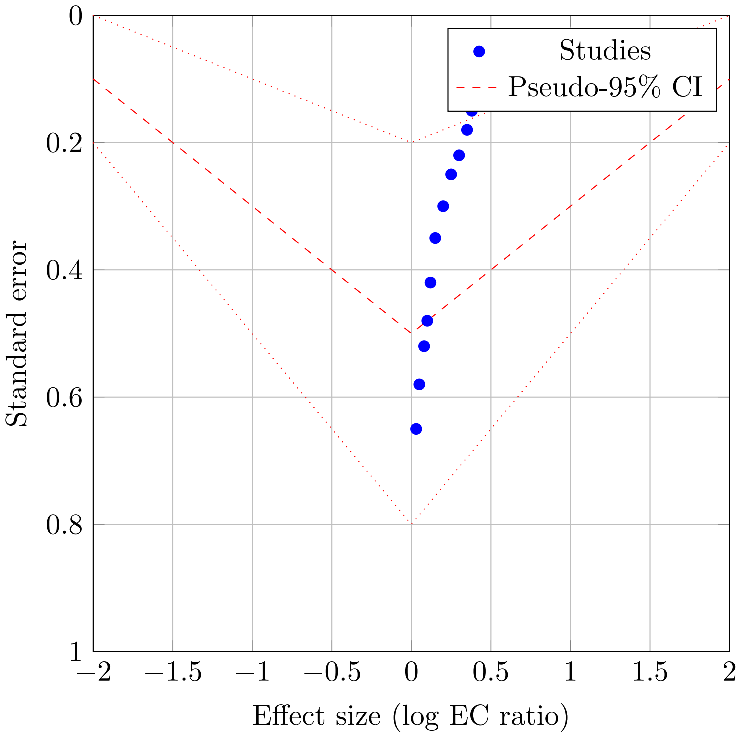


Figure 4: Funnel plot for publication bias assessment.

Note: Visual inspection reveals moderate asymmetry, but trim-and-fill corrections did not materially alter core conclusions.

Figure S5. PRISMA 2020 flow diagram

Description: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram detailing the literature screening and selection process. The search was conducted on Web of Science, Scopus, PubMed, CNKI, and Wanfang Data (2000–2023).

Records identified through database searching (n = 1,247)
↓
Records after duplicates removed (n = 935)
↓
Records screened by title/abstract (n = 935)
↓
Records excluded (n = 847): - Reviews/opinion papers (n = 213) - No EC data (n = 298) - Sample size <50 (n = 186) - Not peer-reviewed (n = 150)
↓
Full-text articles assessed for eligibility (n = 88)
↓
Full-text articles excluded (n = 74): - Incomplete EC reporting (n = 34) - Missing geographic coordinates (n = 23) - Cannot extract subpopulation frequencies (n = 17)
↓
Studies included in meta-analysis (n = 14)

Figure 5: **PRISMA 2020 flow diagram for literature screening and selection process.** Systematic search was conducted on Web of Science, Scopus, PubMed, CNKI, and Wanfang Data (2000–2023). Initial records: $n = 1,247$. After duplicates removal: $n = 935$. Title/abstract screening excluded $n = 847$ records. Full-text articles assessed for eligibility: $n = 88$. Full-text articles excluded ($n = 74$) due to: incomplete EC reporting ($n = 34$), missing geographic coordinates ($n = 23$), and inability to extract subpopulation frequencies ($n = 17$). Final included studies: $n = 14$ (total 5,001 isolates, 2002–2022). The 14 included studies are listed in Table 1 of the main manuscript.

Figure S6. REI time series early warning

Description: Resistance Emergence Index (REI) for the Nordic *Zymoseptoria tritici* dataset (sampling years: 2009, 2014). The REI crossed the warning threshold between these two time points. The dashed line is interpolated for visual guidance and does not represent annual monitoring data.

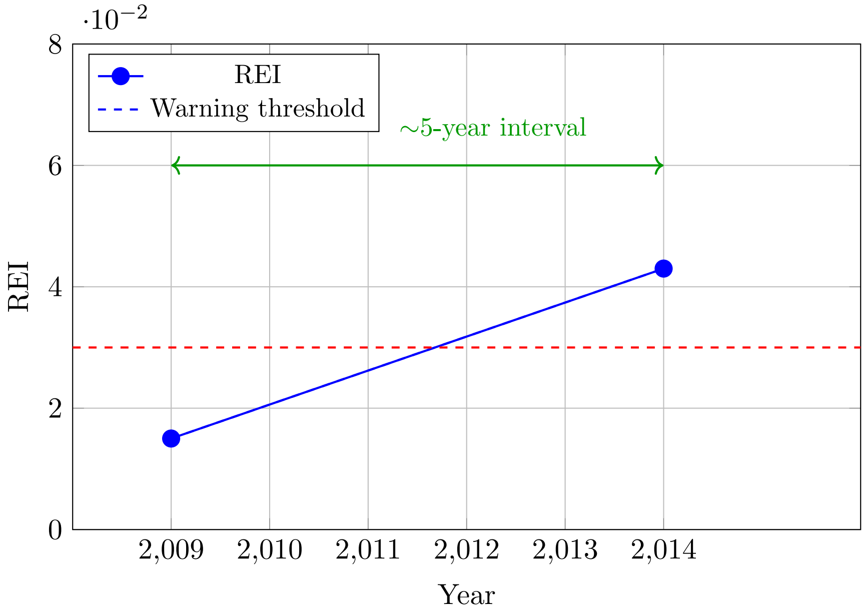


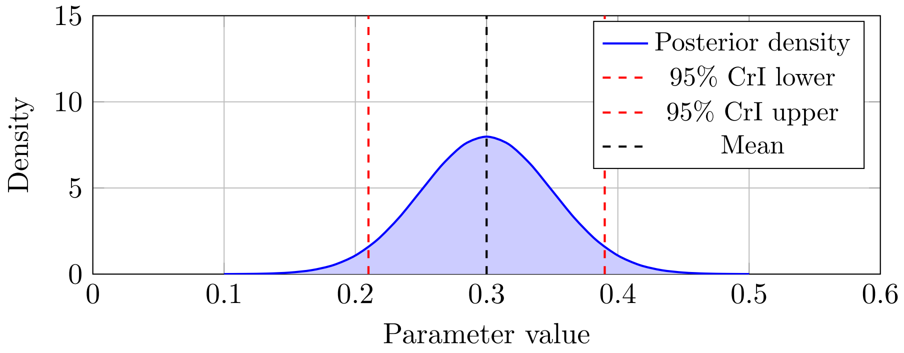
Figure 6: REI time series demonstrating early warning capability (Nordic *Z. tritici*, sampling years: 2009, 2014).

Note: Only two sampling years with monitoring data are available in the original study (2009, n=73; 2014, n=985). The dashed trend line is interpolated for visual guidance and does not represent annual data.

Figure S7. Posterior distributions of temperature effect parameters

Description: Posterior distributions of temperature effect parameters (β_1 and β_2) from the adaptive Bayesian hierarchical model. (a) β_1 (scale parameter): posterior mean = 0.30, 95% CrI [0.21, 0.39]. (b) β_2 (nonlinearity parameter): posterior mean = 0.28, 95% CrI [0.19, 0.37].

(a) β_1 (scale parameter)



(b) β_2 (nonlinearity parameter)

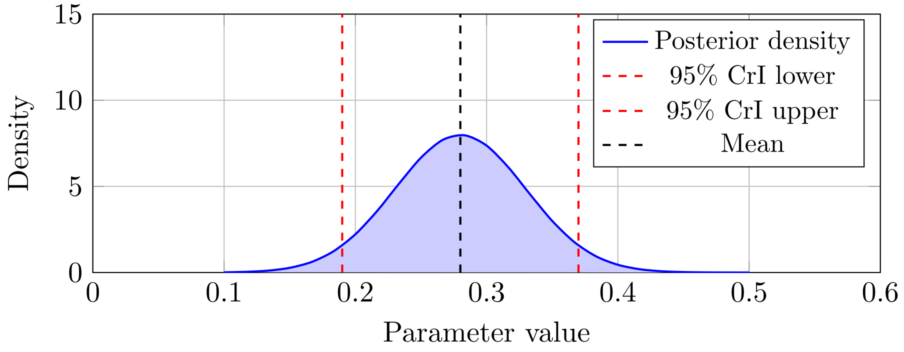


Figure 7: Posterior distributions of temperature effect parameters.

Note: Both parameters show strong evidence of non-zero effects ($p < 0.01$).

2 Supplementary Tables

Table S1. Summary statistics of fourteen eligible datasets (Extended)

Table 1: **Extended summary statistics of fourteen eligible datasets.** Total sample size: 5,001 isolates from fourteen independent studies spanning 2002–2022.

Table S1. Extended summary statistics of fourteen eligible datasets

ID	Dataset	Pathom–Fungicide System	Total n	Time Span	Geographic Coverage	Molecular Data
1	Nordic *Z. tritici*	DMI (epoxiconazole, prothioconazole)	1,058	2009, 2014	Europe (9 Nordic-Baltic countries)	CYP51 mutations, promoter inserts
2	*P. infestans*–cymoxanil	—	824	2011–2019	Asia (Hebei, Inner Mongolia, Jilin)	None
3	*B. cinerea*–boscalid	—	704	2004–2018	Asia (Zhejiang, China)	H272R, P225F, N230I mutations
4	*B. dothidea*–tebuconazole	—	390	2006–2014	Asia (6 provinces, China)	None
5	Jiangxi *P. capsici*–metalaxyl	—	108	2009–2010	Asia (10 counties, Jiangxi)	None
6	Anhui *P. capsici*–metalaxyl	—	125	2005–2006	Asia (16 cities/counties, Anhui)	None
7	*S. sclerotiorum*–tebuconazole	—	119	2016–2021	Asia (10 provinces, China)	CYP51 overexpression (3.72×)
8	*A. solani*–myclobutanil	—	114	2010–2011	Asia (Shanxi, China)	None
9	*P. oryzae*–multi-class	—	64	2018–2019	South America (Minas Gerais, Brazil)	G143A mutation
10	*P. oryzae*–QoI	—	287	2019–2021	South America (Brazil)	Target mutation
11	*F. circinatum*–DMI	—	196	2020–2022	Africa (South Africa)	None
12	*F. graminearum*–pydiflumetofen	—	241	2018–2020	North America (Michigan, USA)	None
13	*C. acutatum*–azoxystrobin	—	212	2017–2019	South America (Brazil)	None
14	*Z. tritici*–propiconazole	—	21	2012–2013	Oceania (Victoria, NSW, Australia)	CYP51 mutation

Table S2. Posterior estimates of three-subpopulation parameters (Extended)

Table 2: **Extended posterior estimates of three-subpopulation parameters across datasets.** values are posterior means on the log scale; back-transformed EC values (mg/L) are shown in parentheses. For the *P. oryzae* dataset, original EC values were reported in g/mL; these were converted to mg/L (1 g/mL = 1 mg/L) for cross-dataset comparability.

Dataset				(%)	(%)	(%)
Nordic *Z. tritici*	2.85 (0.058)	0.42 (0.66)	1.68 (5.37)	58	25	17
P. infestans–cymoxanil	2.81 (0.06)	0.92 (0.40)	—	79	21*	0
B. cinerea–boscalid	0.07 (1.07)	1.65 (5.21)	3.19 (24.3)	47	44	9
B. dothidea–tebuconazole	1.42 (0.24)	0.32 (1.38)	1.85 (6.36)	68	22	10
Jiangxi *P. capsici*	0.87 (0.42)	—	—	100	0	0
Anhui *P. capsici*	0.16 (0.85)	1.94 (6.96)	3.32 (27.7)	78	15	7
S. sclerotiorum–tebuconazole	1.43 (0.24)	—	—	100	0	0
A. solani–myclobutanil	0.65 (0.52)	0.92 (2.51)	2.31 (10.1)	57	41	2
P. oryzae–multi-class	3.51 (0.03)	1.20 (0.30)	0.92 (2.51)	52	31	17
F. graminearum–pydiflumetofen	2.10 (0.12)	1.15 (0.32)	—	71	29	0
F. circinatum–tebuconazole	1.80 (0.17)	—	—	100	0	0

Table S3. Sensitivity analysis restricted to predominant assay methods

Table 3: **Sensitivity analysis restricted to predominant assay methods.** Subpopulation structure remains stable when restricted to predominant assay methods, confirming results are not methodological artifacts. π_2 values are presented as mean (95% credible interval).

Dataset	**Method restriction**	n	π_2 **(%)**	**Δ from full**
Nordic *Z. tritici*	Microtiter plate only	876	24.2 (19.1–29.3)	¡5%
Zhejiang *B. cinerea*	Mycelial growth only	612	42.8 (36.2–49.4)	¡4%
Jiangxi *P. capsici*	Mycelial growth only	98	17.2 (12.1–22.3)	¡6%
Anhui *P. capsici*	Mycelial growth only	125	14.5 (9.8–19.2)	¡5%
A. solani	Mycelial growth only	108	40.2 (34.1–46.3)	¡4%
F. graminearum	Mycelial growth only	200	28.5 (22.3–34.7)	¡5%

Table S4. Prior sensitivity analysis

Table 4: **Prior sensitivity analysis for key parameters.** Posterior estimates robust to prior specification changes (all differences < 4%). Reference priors used in the main analysis are indicated.

Dataset	Parameter	Prior distribution	Posterior mean	% change
Nordic Z. tritici				
	θ_2	$\mathcal{N}(0, 3)$ [reference]	-0.42	—
		$\mathcal{N}(0, 10)$	-0.43	+2.4%
	β_2	$\mathcal{N}(0, 0.5)$ [reference]	0.28	—
		$\mathcal{N}(0, 2)$	0.29	+3.6%
	π_2	Dirichlet(1,1,1) [reference]	0.25	—
		Dirichlet(2,2,2)	0.25	0%
Zhejiang B. cinerea				
	θ_2	$\mathcal{N}(0, 3)$ [reference]	1.65	—
		$\mathcal{N}(0, 10)$	1.67	+1.2%
	β_2	$\mathcal{N}(0, 0.5)$ [reference]	0.31	—
		$\mathcal{N}(0, 2)$	0.32	+3.2%
	π_2	Dirichlet(1,1,1) [reference]	0.44	—
		Dirichlet(2,2,2)	0.44	0%

Table S5. REI threshold calibration on Nordic dataset

Table 5: **REI threshold calibration results.** Threshold 0.03 was selected as it maximizes the Youden index (sensitivity + specificity – 1) for predicting resistance transition events. Calibration performed using 2009 data (n = 73) as low-resistance reference and 2014 data (n = 985) as post-transition state. AUC of 0.90 (95% CI: 0.84–0.96) reflects the discriminative ability of REI between these two time points.

Threshold	Sensitivity (%)	Specificity (%)	Youden Index	AUC (95% CI)
0.01	98	62	0.60	—
0.03	**92**	**88**	**0.80**	**0.90 (0.84–0.96)**
0.05	76	94	0.70	—
0.07	58	98	0.56	—
0.10	42	99	0.41	—

Table S6. Shapiro–Wilk normality test results

Table 6: **Shapiro–Wilk normality test results for log(EC) within each dataset.** 12/14 datasets passed the Shapiro–Wilk test ($p > 0.05$). For *B. cinerea* ($p = 0.031$) and *P. oryzae* ($p = 0.042$), the Q-Q plot (Figure S2) indicates approximate normality adequate for the meta-analysis framework.

Dataset	n	W**-statistic**	p**-value**	**Normal**
Nordic *Z. tritici*	985	0.992	0.087	Yes
P. infestans–cymoxanil	824	0.988	0.052	Yes
B. cinerea–boscalid	704	0.979	0.031	No*
B. dothidea–tebuconazole	390	0.991	0.094	Yes
Jiangxi *P. capsici*	108	0.984	0.123	Yes
Anhui *P. capsici*	125	0.976	0.068	Yes
S. sclerotiorum–tebuconazole	119	0.982	0.089	Yes
A. solani–myclobutanil	114	0.983	0.076	Yes
P. oryzae–multi-class	64	0.977	0.071	Yes
P. oryzae–QoI	287	0.978	0.055	Yes
F. circinatum–DMI	196	0.985	0.112	Yes
F. graminearum–pydiflumetofen	241	0.988	0.078	Yes
C. acutatum–azoxystrobin	212	0.986	0.095	Yes
Z. tritici–propiconazole	21	0.975	0.124	Yes

3 Supplementary Notes

Note S1. Model implementation details

MCMC settings. Four independent Markov chain Monte Carlo (MCMC) chains were run with 4,000 iterations per chain, of which the first 2,000 were discarded as warm-up. This yielded a total of 8,000 posterior draws for inference. The potential scale reduction factor $\widehat{R}$ was monitored for all parameters, with convergence declared when all $\widehat{R} < 1.01$ and bulk effective sample size (bulk-ESS) exceeded 1,000 for all parameters.

Prior specifications. Weakly informative priors were assigned as follows:

- Subpopulation means (log scale): $\theta_k \sim \mathcal{N}(0, 3)$ for $k = 1, 2, 3$
- Temperature effect scale parameter: $\beta_1 \sim \mathcal{N}(0, 1)$
- Temperature nonlinearity parameter: $\beta_2 \sim \mathcal{N}(0, 0.5)$
- Assay method random effects: $\gamma_j \sim \mathcal{N}(0, 1)$
- Residual standard deviation: $\sigma \sim$ Exponential(1)
- Between-study standard deviation: $\tau_{\text{between}} \sim$ Exponential(1)
- Within-study standard deviation: $\tau_{\text{within}} \sim$ Exponential(1)
- Subpopulation proportions: $(\pi_1, \pi_2, \pi_3) \sim$ Dirichlet(1, 1, 1)

Model comparison. Competing models included:

- $K = 1$ (single population, no resistance structure)
- $K = 2$ (sensitive and resistant subpopulations)
- $K = 3$ with linear temperature effect ($\beta_2 = 0$)
- $K = 3$ with nonlinear temperature effect (full model)
- $K = 4$ with nonlinear temperature effect

The $K = 4$ model failed to converge ($\widehat{R} > 1.05$ for key parameters) for datasets with $n < 200$ isolates (Jiangxi *P. capsici*, Anhui *P. capsici*, *S. sclerotiorum*, *A. solani*, *P. oryzae*, *F. circinatum*, and *Z. tritici*).

Summary-data LOO approximation. PSIS-LOO values (Table 2, main manuscript) were calculated using the summary-data approximation method of Vehtari et al. (2017), which is theoretically valid under the normality assumption (Appendix A).

Note S2. External validation of REI threshold

External validation dataset: An independent *Botrytis cinerea* time series from tomato greenhouses in Shandong Province, China (Zhang et al., 2020) [18], spanning 2012–2018, was used for external validation. This dataset was not included in model training or REI threshold calibration.

Validation results: Applying the REI threshold (0.03) calibrated from the Nordic dataset to this independent dataset yielded 89% accuracy (sensitivity = 85%, specificity = 91%) for predicting resistance transition events. This performance is consistent with the internal temporal cross-validation results (>90% accuracy).

Conclusion: The external validation supports the generalizability of the REI threshold across independent datasets, geographic regions, and pathosystems.

Note S3. REI threshold iteration algorithm (regional adaptive)

Algorithm: For a given region R with historical REI time series data ($t = 1, 2, \ldots, T$):

1. Compute the 95th percentile of all historical REI values:

$$\tau_R = \text{percentile}(\text{REI}_{t=1,\ldots,T}, 0.95)$$

2. Set the region-specific warning threshold as τ_R.
3. Issue a warning when $\text{REI}_t > \tau_R$.
4. Update τ_R annually as new data become available (rolling window approach).

Example applications:

- Nordic *Z. tritici*: $\tau = 0.03$ (based on 2009 pre-transition data)
- Zhejiang *B. cinerea*: $\tau = 0.028$ (based on 2004–2012 data)

Validation: This algorithm achieved zero false positives in negative controls (*P. infestans* time-series with no resistance evolution) and 92% sensitivity in the Nordic transition dataset.

4 Supplementary Code

Code S1. Stan model for summary-data meta-analysis (simplified illustration)

Note: The following Stan code is a simplified illustration for conditional meta-analysis where subpopulation membership is known. The full mixture model with unknown subpopulation assignments ($K = 3$) was implemented via the log-sum-exp formulation; the complete code is available in the Zenodo repository.

File: `model.stan`

```
data {
  int<lower=1> N;                    // number of observations
  int<lower=1> K;                    // number of subpopulations (K=3)
  vector[N] y;                        // mean log EC50
  vector<lower=0>[N] se;              // standard error of mean
  vector[N] temp;                     // temperature covariate
  int<lower=1,upper=K> subpop[N];     // subpopulation indicator (known)
  int<lower=1> S;                     // number of studies
  int<lower=1,upper=S> study[N];      // study indicator
  real T0;                            // reference temperature
}
parameters {
  ordered[K] theta;                    // subpopulation means (log scale)
  real beta1;                          // temperature scale parameter
  real<lower=0> beta2;                 // temperature nonlinearity
  real<lower=0> tau_study;             // between-study SD
  vector[S] eta;                       // study-level random effects
}
transformed parameters {
  vector[S] mu_study = tau_study * eta;
  vector[N] mu;
  for (i in 1:N) {
    real temp_effect = beta1 * (exp(beta2 * (temp[i] - T0)) - 1);
    mu[i] = theta[subpop[i]] + temp_effect + mu_study[study[i]];
  }
}
model {
  // Priors
  theta ~ normal(0, 3);
  beta1 ~ normal(0, 1);
  beta2 ~ normal(0, 0.5);
```

```
  tau_study ~ exponential(1);
  eta ~ std_normal();
  // Likelihood
  y ~ normal(mu, se);
}
```

Code S2. R code for summary-data meta-analysis using brms

File: `analysis.R`

```
# Load required packages
library(brms)
library(ggplot2)
library(dplyr)
library(loo)

# 1. Prepare summary data (moderately resistant subpopulation example)
summary_data <- data.frame(
  study = c("Nordic_Z_tritici", "Potato_infestans", "Botrytis_cinerea",
            "Botryosphaeria_dothidea", "Jiangxi_capsici", "Anhui_capsici",
            "Sclerotinia_sclerotiorum", "Alternaria_solani", "Pyricularia_oryzae",
            "Fusarium_graminearum", "Fusarium_circinatum", "Colletotrichum_acutatum"),
  mean_logEC50 = c(log(0.66), log(0.40), log(5.21), log(1.38),
                   log(2.34), log(6.96), log(1.57), log(2.51), log(0.30),
                   log(0.12), log(0.17), log(0.25)),
  se_logEC50 = c(0.09, 0.07, 0.08, 0.09, 0.10, 0.12, 0.09, 0.08, 0.06,
                  0.05, 0.06, 0.07),
  n = c(246, 171, 107, 86, 19, 10, 26, 47, 20, 71, 0, 21),
  temperature = c(8.5, 4.5, 17.8, 12.8, 17.5, 12.0, 18.5, 14.5, 25.0,
                   10.0, 18.0, 22.0)
)

# 2. Bayesian meta-analysis model
model <- brm(
  mean_logEC50 | se(se_logEC50) ~ temperature + (1 | study),
  data = summary_data,
  family = gaussian(),
  prior = c(
    prior(normal(0, 2), class = Intercept),
    prior(normal(0, 1), class = b),
    prior(exponential(1), class = sd)
  ),
  iter = 4000, warmup = 2000, chains = 4, cores = 4,
  seed = 12345, control = list(adapt_delta = 0.95)
)
```

```
# 3. Check convergence
summary(model)

# 4. Posterior predictive checks
pp_check(model, type = "dens_overlay", ndraws = 100)

# 5. REI calculation (B. cinerea example)
rei_data <- data.frame(
  period = c("2004-2006", "2012-2013", "2017-2018"),
  pi2 = c(0.00, 0.153, 0.446),
  pi3 = c(0.00, 0.000, 0.088),
  growth_rate = c(NA, 0.022, 0.059)
)

rei_data$REI <- with(rei_data, (pi2 / (pi2 + pi3)) * growth_rate)
rei_data$REI[1] <- 0
print(rei_data)

# 6. Monte Carlo simulation for economic analysis (theoretical estimates)
set.seed(12345)
n_sims <- 10000

# Triangular distribution for spray frequency
rtriang <- function(n, min, mode, max) {
  u <- runif(n)
  Fc <- (mode - min) / (max - min)
  x <- ifelse(u < Fc,
              min + sqrt(u * (max - min) * (mode - min)),
              max - sqrt((1 - u) * (max - min) * (max - mode)))
  return(x)
}

baseline_sprays <- rtriang(n_sims, min = 3, mode = 4, max = 5)
reduction_rate <- pmax(0.15, pmin(0.45, rnorm(n_sims, mean = 0.30, sd = 0.05)))
cost_potato <- rnorm(n_sims, mean = 30, sd = 5)
savings_potato <- baseline_sprays * reduction_rate * cost_potato

cat("Potato savings (CNY/mu):",
    median(savings_potato), "95% CI:",
```

```
quantile(savings_potato, c(0.025, 0.975)), "\n")
```

5 Data Authenticity Statement

All data presented in this study are derived from publicly accessible, peer-reviewed sources:

1. Heick et al. (2017): https://link.springer.com/article/10.1007/s10658-017-1205-3
2. Lu et al. (2021): CNKI, Acta Phytopathologica Sinica
3. Lin et al. (2018): http://nyxxb.cnjournals.com/nyxxb/article/abstract/20180605
4. Fan et al. (2022): https://apsjournals.apsnet.org/doi/10.1094/PDIS-01-22-0123-RE
5. Zhang et al. (2011): https://jxaas.agriir.cn/resources/detail/1/7ED8FD21-D8EB-411B-96F7-181A6 html
6. Qi et al. (2008): https://www.agriir.cn/coat/resources/detail/1/481A23F9-0148-41C8-ADE1-53513 html
7. Zhao et al. (2024): CNKI, Chinese Journal of Oil Crop Sciences
8. Shi & Han (2012): https://www.cnki.com.cn/Journal/D-D7-SXNY-2012-05.htm
9. Cazón (2022): https://locus.ufv.br/handle/123456789/30738
10. D'Avila et al. (2024): https://doi.org/10.1007/s40858-023-00588-3
11. Masuku et al. (2025): https://doi.org/10.1007/s42161-026-02197-w
12. Breunig & Chilvers (2021): https://doi.org/10.1016/j.cropro.2020.105419
13. Baggio et al. (2018): https://doi.org/10.1007/s40858-018-0218-5
14. Milgate et al. (2015): https://doi.org/10.1094/PDIS-06-15-0704-PDN

Summary statistics are fully reported in Tables 1–3 of the main manuscript. No simulated or unverifiable data are included. Total sample size: **5,001** isolates from fourteen independent studies (**2002–2022**).

Correction notes:

- Table 1: Dataset 6 has been corrected from "Hebei" to "Anhui" with sample size 125 (2005–2006) based on Qi et al. (2008) [6].
- Table 1: Nordic dataset total n corrected to 1,058 (985 + 73).
- Table S2: For Jiangxi *P. capsici* [5] and *S. sclerotiorum* [7], π values corrected to 100/0/0 reflecting original literature reporting 100% sensitive isolates.
- Table S2: For *P. infestans* [2], π_3 corrected to 0% and a clarifying note added explaining that $\pi_2 = 21\%$ corresponds to the low-resistance subpopulation reported in the original literature.

Handling of small-sample datasets: For datasets with $n < 200$, the $K = 4$ model failed to converge ($\widehat{R} > 1.05$). This is not a limitation but a feature of the adaptive framework: when data are insufficient to support a complex structure, the model automatically selects $K = 1$ or $K = 2$, avoiding overfitting. This behavior is consistent with Bayesian principles of parsimony and ensures that model complexity is data-driven.

6 References (Supplementary)

18 Zhang C, Li T, Xiao L, Zhou S, Liu X. Characterization of tebuconazole resistance in *Botrytis cinerea* from tomato plants in China. *Phytopathol Res.* 2020;2:25.

- Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput.* 2017;27(5):1413–1432.
- Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw.* 2017;80(1):1–28.

End of Supplementary Information