Estimation of Relative Parasitic Fitness of Pathogens in Heterogeneous Host Populations

Ping Sun and Shimai Zeng

Associate professor and professor, Department of Plant Protection, Beijing Agricultural University, Beijing 100094, P. R. China. This research was supported in part by the National Science Foundation of China (grant 39230260). We thank the reviewers for their valuable suggestions. Accepted for publication 21 February 1995.

There have been many studies on relative parasitic fitness in plant pathology since Leonard's (5) pioneering work in 1969. Measurement of relative parasitic fitness offers an efficient way to quantitatively analyze host-pathogen interactions (2,3,6,12,13). The relationship between parasitic fitness and other related concepts also has been discussed (1,9).

The commonly used concept of relative fitness of race A to race B on a host can be described by its relative reproductive success. If \( p \) and \( q \) are the frequencies of races A and B, respectively, \( p_0 \) and \( q_0 \) are the initial frequencies, and \( w \) is the relative fitness of race A to race B on the host, then

\[
w' = \frac{(p/p_0)(q/q_0)}{(p/q)(q_0/p_0)},
\]

(1)

in which parameter \( t \) is the number of generations or another unit of time, and the more fit race is always defined as race B so \( w \) is not larger than 1 (4,7,8,10).

Many plant pathogens have only one generation per growing season. If more than one generation is produced, \( t \) is difficult to measure because the pathogen generations overlap. We suggest that the following equation can be used to estimate relative fitness for multiple generations:

\[
f = \frac{(p/q)(q_0/p_0)},
\]

(2)

in which \( p \) and \( q \) are frequencies of races A and B, respectively, at the end of a host's growing season and \( p_0 \) and \( q_0 \) are the initial frequencies for races A and B, respectively. To avoid confusion with \( w \) as defined by equation 1, \( f \) can be called seasonal fitness. If the pathogen has only one generation per growing season, then \( f = w \); otherwise the values of \( w \) and \( f \) may be different; \( f \) can reflect the relative parasitic ability of race A to race B on the host plant. Regardless of the complexity of the interaction between the two races, the final frequencies, \( p \) and \( q \), and the initial frequencies, \( p_0 \) and \( q_0 \), express this interaction. In practice, a pathogen population usually consists of several races, and there are usually multiple cultivars grown in an area. Equation 2 can be extended to increase its usefulness. Suppose that \( m \) cultivars are grown in a field, with percents \( c_1, c_2, \ldots, c_m \), respectively, and there are \( n \) races in the pathogen population, we use \( f_{ijk} \) to denote the relative seasonal fitness of race \( i \) to race \( j \) on cultivar \( k \); \( i \) and \( j \) = 1 to \( n \), and \( k \) = 1 to \( m \). The frequencies of the \( n \) races at the end of an epidemic are \( p_1, p_2, \ldots, p_m \) and the initial frequencies are \( p_0 \), \( p_0 \). The overall fitness of the \( i \)th race relative to the \( j \)th race can be expressed by the following model:

\[
(p/p_i)(p_0/p_0) = c_1 f_{ij1} + c_2 f_{ij2} + \ldots + c_m f_{ijn},
\]

(3)

in which \( i \) differs from \( j \) and \( 1 \leq i \leq n \).

The race with the greatest fitness, if known, is designated the \( j \)th race. For example, if it is race 1, we use the following general equations:

\[
(p/p_1)(p_0/p_0) = c_1 f_{i11} + c_2 f_{i12} + \ldots + c_m f_{i1m},
\]

(4)

for \( i = 2 \) to \( n \). Specifically, with \( i = 2 \) to \( n \), we have:

\[
(p/p_i)(p_0/p_0) = c_1 f_{i21} + c_2 f_{i22} + \ldots + c_m f_{i2m},
\]

(5)

\[
(p/p_i)(p_0/p_0) = c_1 f_{i31} + c_2 f_{i32} + \ldots + c_m f_{i3m},
\]

\[
\ldots \ldots
d...

\[
(p/p_i)(p_0/p_0) = c_1 f_{i51} + c_2 f_{i52} + \ldots + c_m f_{i5m}.
\]

In the above equations, the unknown parameter \( f_{ijk} (i = 2 \text{ to } n; k = 1 \text{ to } m) \), can be estimated by arranging field experiments. If there are \( s \) sampling fields, \( c_{ik} \) denotes the percentage of cultivar \( k \) used in sampling field \( t \). In each field we can obtain the value of \((p/p_i)(p_0/p_0)\) for \( i = 2 \) to \( n \). We use \( y_{it} = [(p/p_i)(p_0/p_0)]_t \) to denote the value of \((p/p_i)(p_0/p_0)\) obtained from the field, \( t = 1 \) to \( s \). For a constant \( i \), the following linear model applies:

\[
y_{it} = c_{i1} f_{i11} + c_{i2} f_{i12} + \ldots + c_{im} f_{i1m},
\]

(6)

in which \( t = 1 \) to \( s \).

Thus, the system of equations (for constant \( i, k = 1 \text{ to } m, \text{ and } t = 1 \text{ to } s \)) can be represented by the following matrix operation:

\[
\begin{bmatrix}
    c_{i1} & c_{i2} & \ldots & c_{im} \\
    c_{i1} & c_{i2} & \ldots & c_{im} \\
    \vdots & \vdots & \ddots & \vdots \\
    c_{i1} & c_{i2} & \ldots & c_{im}
\end{bmatrix}
\begin{bmatrix}
    f_{i11} \\
    f_{i12} \\
    \vdots \\
    f_{i1m}
\end{bmatrix}
= y_{i1}
\]

or \( C \cdot F_{ii} = Y_{i} \), in which \( F_{ii} \) is the \( m \times 1 \) vector, \( C \) is the \((c_{i,k})_{x\times m}\) matrix; and \( Y_{i} \) is the \( s \times 1 \) vector.

When \( s \geq m \), the least squares estimator (LSE) of \( F_{ii} \) is

\[
\hat{F}_{ii} = (C'C)^{-1}C'Y_{i},
\]

(7)

in which symbol \( C' \) denotes the transpose of matrix \( C \), \( \hat{F}_{ii} = (\hat{f}_{i11}, \hat{f}_{i12}, \ldots, \hat{f}_{i1m})' \); \( f_{i1k} \) denotes the LSE of unknown parameter \( f_{i1k} \) for \( k = 1 \) to \( m \).

For \( i = 2 \) to \( n \), by substituting \( C', C \), and \( Y_{i} \) into equation 7, we
can obtain the corresponding LSE of \( F_{i1} \). Our purpose is to obtain the nonnegative LSE of \( f_{i1} \) (\( i = 2 \) to \( n \); \( k = 1 \) to \( m \)) because fitness is nonnegative. A related method was reported by Sun and Zeng (11).

When \( m = 1 \) and \( n = 2 \), i.e., only one cultivar is used and two races exist, equation 4 becomes

\[
y_{21} = (p_2/p_1)(p_{10}/p_{20}) = f_{21}.
\]

This is the case defined by equation 2, in which \( f_{21} \) is the relative seasonal fitness of race 2 to race 1 on the cultivar. Equation 4, therefore, can be considered a logical extension of equation 2. When \( m = 1 \) and \( n > 2 \),

\[
y_{i1} = (p_2/p_1)(p_{10}/p_{20}) = f_{i1}, \quad \text{for} \quad i = 2 \text{ to } n.
\]

If there are \( s \) observations, the mean \((1/s)(y_1 + y_2 + \cdots + y_s)\) is the LSE of \( f_{i1} \), the relative seasonal fitness of race \( i \) to race 1 on that cultivar for \( i = 2 \) to \( n \).

The following is a numerical example for \( n = m = 2 \) for estimating unknown parameters \( f_{11} \) and \( f_{21} \) and the application of regression analysis after \( f_{21} \) and \( f_{12} \) have been estimated. The corresponding model is as follows:

\[
(p_2/p_1)(p_{10}/p_{20}) = c_1 f_{11} + (1 - c_1) f_{12}.
\]

Suppose four experiments (\( s = 4 \)) were conducted to estimate the relative seasonal fitness of the two races (\( n = 2 \)) on the two cultivars (\( m = 2 \)). Then the model is

\[
y_{21} = c_1 f_{21} + (1 - c_1) f_{22},
\]

in which \( y_{21} = (p_2/p_1)(p_{10}/p_{20}), \) for \( i = 1 \) to \( 4 \).

For simplicity, suppose the initial frequencies of races 1 and 2 are 0.5 and 0.5 and also suppose that the proportions of cultivar 1 in the four fields are 0.0, 0.4, 0.7, and 0.9, respectively, and the final ratios of \( p_2/p_1 \) in the corresponding fields are 0.95, 0.55, 0.35, and 0.25, respectively. Then \( Y_2 \) is a \( 4 \times 1 \) vector, the transpose of \( Y_2 \) is \( Y' = (y_{21}, y_{22}, y_{23}, y_{24}) = (0.95, 0.55, 0.35, 0.25) \),

\[
C = \begin{bmatrix}
0 & 0.4 & 0.6 & 0.7 \\
0.0 & 0.6 & 0.3 & 0.9 \\
0.0 & 0.6 & 0.3 & 0.9 \\
0.0 & 0.6 & 0.3 & 0.9
\end{bmatrix}, \quad C' = \begin{bmatrix}
0 & 0.4 & 0.7 & 0.9 \\
0.0 & 0.6 & 0.3 & 0.9
\end{bmatrix}.
\]

and the LSE of \( F_{21} \) is as follows:

\[
\hat{F}_{21} = (C'C)^{-1} C'Y_2,
\]

in which \( \hat{F}_{21} = (\hat{f}_{211}, \hat{f}_{212}) \) and \( \hat{F}_{21} \) is the LSE of the relative seasonal fitness of race 2 to race 1 on cultivar \( k = 1 \) or 2.

By substituting matrix \( C' \), \( C' \), and vector \( Y_2 \) into equation 9 and solving the equations, we obtain \( f_{11} = 0.134, f_{12} = 0.916 \) and the regression equation:

\[
(p_2/p_1)(p_{10}/p_{20}) = 0.134 c_1 + 0.916 (1 - c_1).
\]

Equation 10 can be used to predict the effect of host composition on race frequencies. Assuming the initial frequencies of the two races are equal \( p_{10} = p_{20} = 0.5 \), equation 10 becomes:

\[
(p_2/p_1) = 0.134 c_1 + 0.916 (1 - c_1).
\]

For \( 0 \leq c_1 \leq 1 \), we can estimate \( p_2/p_1 \) by equation 11. For example, if \( c_1 = 0 \), i.e., only cultivar 2 was used, \( p_2/p_1 = f_{211} = 0.916 \); if \( c_1 = 1 \), i.e., only cultivar 1 was used, \( p_2/p_1 = f_{212} = 0.134 \).

If \( p_2/p_1 \) needs to be fixed, i.e., we want to control the ratio of the two races, we can calculate \( c_1 \) by rearranging equation 11:

\[
c_1 = \frac{(0.916 - (p_2/p_1))/(0.916 - 0.134)}.
\]

As long as \( 0.134 \leq (p_2/p_1) \leq 0.916 \), we can find a \( c_1 \), \( 0 \leq c_1 \leq 1 \), that satisfies equation 12. When \( n \) and \( m \) are larger than 2, the equations can be solved by matrix algebra.

In estimating relative seasonal fitness, it is important to carefully define the meaning of "the end of a growing season" because it is closely related to the value of the final frequencies of races. We think "the end of a growing season" should be defined as the time when propagation of the pathogen population has stopped. For example, in the study of wheat stripe rust in China, a daily average temperature of 22°C is set as an indicator of the end of the epidemic season because no further infection and sporulation of the fungus can occur when the daily temperature is above 22°C. In some winter-wheat regions in China, the end of the epidemic is about 2 weeks before harvest, whereas in some mountain regions the end of the epidemic is at harvest when the daily average temperature is seldom higher than 22°C.

Relative parasitic fitness is a measure of the relative success of one pathogenic genotype in a given host population. We have presented a method to estimate fitness in the case of multiple races and cultivars. Although our models need to be tested by field-plot experiments, we foresee potential application of the approach in quantitative analysis of host-pathogen interactions and in predicting the effects of host mixtures on pathogen population.

LITERATURE CITED