In this supplementary information, we do three things: derive the fitness function Equation (10) used in the main paper; calculate the ESS for the fitness function in Equation (10); more generally, demonstrate the method in solving for an ESS and ‘closing a model’ (writing relatedness in terms of model parameters).

0.1 Deriving the Fitness Function in the Text

Here, we derive the fitness function in Equation (10). If an individual survives (with probability $k$), its fitness is 1; otherwise, with probability $(1 - k)$, its fitness is equal to its offsprings’ fitnesses. A proportion $d$ compete globally, and therefore have fitness relative to the global average, which we assume to be 1 (the population is neither growing nor shrinking). A proportion $1 - d$ remain locally, and their fitness is relative to the average fitness in the social group. After diffusion, the number of individuals on a patch is equal to the number of individuals produced on a patch that stay (with probability $1 - d$) plus the number of individuals arriving from elsewhere ($dN$) ([? ]). So, the total number of offspring on a patch after diffusion is:

$$(1 - d) [N + N (bZ (1 - Z) - cZ)] + dN = N[1 + (1 - d) (bZ (1 - Z) - cZ)].$$  (1)

From this, we can write the fitness function in the text, Equation (10):

$$w(x, y, Z) = (1 - k) (d) (x (1 - c + by) + (1 - x) (1 + by)) +$$

$$(1 - d) \frac{x (1 - c + by) + (1 - x) (1 + by)}{1 + (1 - d) (bZ (1 - Z) - cZ)} + k.$$  (2)
0.2 Calculating the ESS

To calculate the ESS for Equation (10), we use the Taylor–Frank approach. However, the solution is not analytically tractable. Instead, if we assume \( b \) and \( c \) to be small, we can write a first-order approximation of Equation (10) as

\[
w(x, y) = (1 - k) [1 - cx + (1 - x) by - (bZ (1 - Z) - cZ) (1 - d)^2] + k. \tag{3}
\]

We previously showed that relaxing the assumption of small \( b \) and \( c \) does not qualitatively alter the results ([? ]). From this, we can use the Taylor–Frank (1996) approach to solve for the ESS ([? ]). We take the derivative of fitness with respect to phenotype, solving for \( x = y = x^* \), and candidate ESSs occur where:

\[
\frac{dw}{dy} = \frac{\partial w}{\partial x} + R \frac{\partial w}{\partial y}
\]

\[
= -c - b x - (1 - d)^2 \left( -\frac{c}{n} - \frac{b (x + (-1 + n) x)}{n^2} + \frac{b \left(1 - \frac{x+(-1+n)x}{n}\right)}{n} \right) +

r \left( b (1 - x) - (1 - d)^2 \left( -\frac{c (1 + n)}{n} - \frac{b (1 + n) (x + (-1 + n) x)}{n^2} + \frac{b (1 + n) \left(1 - \frac{x+(-1+n)x}{n}\right)}{n} \right) \right)
\]

\[
= 0. \tag{4}
\]

Solving for \( x^* \), we get:
\[ x^* = \frac{cN - bNR + b(1 - d)^2 - c(1 - d)^2 - bR(1 - d)^2 + cR(1 - d)^2 + bNR(1 - d)^2 - cNR(1 - d)^2}{b \left( -N - NR + 2(1 - d)^2 - 2R(1 - d)^2 + 2NR(1 - d)^2 \right)}, \]

which is the ESS value of cooperation.

### 0.3 Writing Relatedness in Terms of Model Parameters

Equation (S5) gives the ESS in terms of \( R \) and other model parameters, but we expect \( R \) to depend on \( d \), \( k \), and \( N \). Here, we calculate \( R \) in terms of those parameters, though more generally any parameters may impact \( R \), and the following approach readily extends to such cases. We start by determining the relatedness, at equilibrium, of a focal RNA molecule to a random molecule in its social group, including itself. This is known as whole-group relatedness (denoted by \( R \)), because it includes the focal individual, in contrast to others-only relatedness (\( R \)), which does not include the focal individual ([? ?]). Our model requires others-only relatedness, because \( y \) is the average of the individuals on the patch, excluding the focal individual. \( R \) is the relatedness between two individuals drawn randomly from a local group with replacement. We can write this as the probability that those two individuals are the same individual (\( 1/N \)), and thus have relatedness 1, plus the probability that those two individuals are not the same (\( (N-1)/N \)), and thus have the relatedness of two random individuals drawn without replacement, or others-only relatedness, \( R \):

\[ R = \frac{1}{N} + \frac{N-1}{N}R \]

Equation (S6) is a general equation for relatedness, for any infinite population of individuals subdivided into \( N \) social groups. Now we take two individuals (without replacement) on the same patch with relatedness \( R \), and determine the
relatedness of their representatives in the previous generation. With chance $k^2$, they are both survivors from the previous generation, in which case their relatedness is the same ($R$). With chance $2k(1-k)$, one is a survivor and the other is a new offspring, which is native with probability $(1-d)$, in which case their relatedness is $R$. Otherwise, with chance $(1-k)^2$, they are both new offspring, are both native with probability $(1-d)^2$, and thus have relatedness $R$. Others-only relatedness between two individuals in the current generation is equal to

$$R_t = k^2 R_{t-1} + 2k(1-k)(1-d) R_{t-1} + (1-k)^2 (1-d)^2 R_{t-1}$$

(7)

Here, $R_t$ is relatedness in the current generation, or time step, and $R_{t-1}$ and $R_{t-1}$ are others-only and whole-group relatednesses, respectively, in the previous one. Note that Equation (S7) was derived assuming that the only population processes affecting relatedness were survival and diffusion, but a similar recursion could be written taking into account any parameters that affect relatedness. Setting $R_t = R_{t-1}$, we find the equilibrium others-only relatedness. Plugging into equation (A5), we find the equilibrium value of whole-group relatedness, $R^*$, to be:

$$R^* = \frac{1 + k}{N + kN + 2k(1-d) - 2kN(1-d) + (1-d)^2 - k(1-d)^2 - N(1-d)^2 + kN(1-d)^2}$$

(8)

This equation for relatedness was identified by Taylor and Irwin 2000 ([? ]). However, our model in Equation (10) in the text is modelling an others only-trait, and thus requires others-only relatedness, $R$. $\mathbb{N}$ gives us the number of relatives on our patch. Subtracting the focal individual, and dividing by the total number of remaining individuals ($N-1$), gives us $R^*$.
\[ R^* = \frac{N(1+k)}{(1+k)N-(N-1)[(1-d)(1-d)]} - 1 \] (9)

Plugging into Equation (S5) gives us Equation (11) in the text.