In the .rmd file, we provide the code for replicating our results. Below is a data dictionary naming the code files and explaining their contents in order of appearance in this file; this file is in order of empirical data appearance

in the main text, then in the supplementary information, then has modelling code at the end.

Files, uses, and variable names:

**SI\_transmission.csv** contains the data from our Transmission Rate Experiment and the data from

Stephenson et. al 2017 which we used in creating Figure 3A. The variables are

**full\_dwormattrans**: natural log of the number of parasites the donor had at the point of transmission

**full\_rspeedtrans**: the transmission rate, calculated following the details in the main text

**full\_line**: the identity of the isogenic line: Gt3 is the line used by Stephenson et. al 2017, while our

previously unpublished data uses different lines such as “GyTu\_CA\_L3”, which means *Gyrodactylus*

*turnbulli line CA\_L3*; this also corresponds to the CA\_L3 line in SI\_parasitetraits.csv.

**SI\_tradeoff.csv** contains data averages and identifying information from 22 lines. This is a summary and

subset of the data in SI\_parasitetraits.csv.

**linename**: the identifying name of the isogenic line

**meanintensity**: the mean intensity recorded on fish infected with that line

**meandeath**: the mean death rate of fish infected with that line, as explained in the main text.

**meantransmission**: the mean transmission rate of each isogenic line, estimated from their individual

measurements of intensity and the relationship in Fig. 3A

**origin**: whether the isogenic line originated from a guppy collected in the wild (‘Wild’), or obtained from a

commercial supplier (‘Domestic’)

NA in a row indicates that the donor fish did not transmit its infection in the 5 days of the experiment.

**SI\_Behav.csv** contains data from our Field Survey of Shoaling Rate, which contains the following

variables:

**Sex**: sex of the fish

**River**: the river of origin of the fish

**Course**: the course of the river of origin - upper or lower

**Length**: length of the fish (mm)

**Infec**: whether or not the fish was infected

**Prop.Shoal200**: the proportion of time the fish spent shoaling during a behavioral trial

**DOT**: date the data were collected in YYYYMMDD format

**ShoalID**: identity of the shoal used - each was used more than once

**ShoalSide**: side of the enclosure that contained the shoal - to control for side bias

**enclosure**: which enclosure was used for the trial

**Lighting**: whether the trial was conducted in relatively light or dark lighting

**Lit\_behavior.csv** is data on shoaling rate taken from the literature, which contains:

**river**: the river of origin of the fish

**course**: the course of the river of origin - upper or lower

**value**: the proportion of time the fish spent shoaling during a behavioral trial

**year**: the year in which the data were collected

**source**: the source of the data - either the paper from which they were extracted, or ‘Lab’, denoting that

2

we collected them ourselves during our Field Survey of Shoaling Rate in Trinidad in March 2020.

**Lit\_prevalence.csv**, which contains the following variables:

**river**: the river of origin of the fish

**course**: the course of the river of origin - upper or lower

**prev.yr**: the year in which the data were collected

**prev.n**: the number of fish sampled for that prevalence estimate

**prev**: the proportion of hosts that were found infected with *Gyrodactylus* spp.

**source**: the source of the data - either the paper from which they were extracted, or ‘Lab’, denoting that

we collected them ourselves during our Field Survey of Prevalence in Trinidad in March 2020.

**Lit\_density.csv** is data from the literature on guppy density and contains the following variables:

**river**: the river of origin of the fish

**course**: the course of the river of origin - upper or lower

**year.collected**: the year in which the data were collected

**value**: the estimated number of fish per square meter

**paper.source**: the source of the data

**SI\_parasitetraits.csv**, which gives data per individual fish per row, and includes the following variables:

**line**: the identity of the isogenic line

**species**: the species identity of the line with t for turnbulli, b for bullatarudis, and NA for unknown as not

all lines had an individual checked for species ID by molecular methods

**date**: the calendar date on which the observation was made in YYYYMMDD

**linf**: the total number of days the worm line was maintained in the lab

**lineday**: the number of days since the line was established in the lab

**lnldy**: lineday nested in line: gives each line/lineday combination a unique label to control for psuedoreplication

in our analysis

**gyro**: the number of gyrodactylus found on the fish at that date. NA for fish that were dead.

**type**: categorical variable describing whether the fish was dead, uninfected or infected

**course**: whether the isogenic line was initiated from a worm wild-caught in the upper or lower courses of a

Trinidadian river, or from a commercially available lab stock fish (‘Lab’)

**river**: the river (or “Lab” meaning a domestic line) from which the worm line originated

**lagint**: the mean intensity (total number of parasites/total number of infected fish) at the previous

observation

**daysls**: the number of days since the previous observation

**lagdeathinf**: the number of dead fish at this observation/the number of infected fish at the previous

observation [note this can theoretically be larger than 1 because new uninfected fish added at the previous

timepoint would not be included in the denominator here, but may become infected and die by the

subsequent timepoint]

**nfish**: the number of fish in the tank at the observation (excluding newly added uninfected fish)

**SI\_ParasiteCounts2020.csv** holds counts of parasites in wild populations taken in 2020. These are

used for comparing laboratory intensities to field intensities (Fig. S2) and for estimating the proportion of

infected fish with just one worm to determine wild coinfection rates (Table S1). Each row is one fish. The

relevant columns are

**Course2** This states whether a sample was from an upper coure or lower course population

**River** The river that was sampled

**Gyro** The number of *Gyrodactylus* spp. on that fish.

**SI\_ParasiteCountsOther.csv** holds data from Stephenson et. al 2015 to help establish what is a normal

percentage of fish with more than one worm by providing additional power, thus contributing to Table S1.

Each row is one fish. The relevant columns are

3

**course** This states whether a sample was from an upper coure or lower course population

**river** The river that was sampled

**gyro** The number of *Gyrodactylus* spp. on that fish.

**SI\_coinfection.csv** holds data from the Field Survey of Coinfection Rates. Each row is one fish. The

columns are

**Fish.ID** A unique ID for each fish.

**River** The river the fish was sampled from.

**Course** The course the fish was sampled from.

**Number.of.GB.genotyped** The number of worms sampled from the fish that were found to be G.B.

**Number.of.GB.genotypes** The number of different G.B. genotypes found on the fish.

**Number.of.GT.genotyped** See corresponding column for G.B.

**Number.of.GT.genotypes** See corresponding column for G.B.