

Supplementary Information of Translating microbial kinetics into quantitative responses and testable hypotheses using Kinbiont

Fabrizio Angaroni¹, Alberto Peruzzi¹, Edgar Z. Alvarenga², and Fernanda Pinheiro^{*1}

¹Computational Biology Research Centre, Human Technopole, Milan, Italy

²Fuzuê Tech, Kraków, Poland

Contents

1	List of models implemented in Kinbiont	1
1.1	Non-linear function models	1
1.2	Ordinary differential equation models	3
1.3	ODE Systems	6
1.4	Cybernetic Models for Bacterial Growth	9
1.5	Reaction Networks	9
2	List of error functions	11
3	Comparison of parameter inference between ODE and NL fit	12
4	Comparison of exponential growth rate estimates using a model-based or a non-parametric method	14
5	Change point detection inference	16
6	Supporting information for the parameter inference of complex microbial kinetics	18
7	Supporting information and tables for the automatic inference of empirical laws	20
8	Supporting information for the characterization of chemical mixture landscapes	25
8.1	Decision tree analysis of <i>Drosophila</i> gut microbiome co-culture dataset	25

1 List of models implemented in Kinbiont

The list of models hard-coded in Kinbiont is given below, together with a description of all model parameters.

1.1 Non-linear function models

- Exponential:

$$N(t) = N_0 e^{\mu t}, \tag{S1}$$

where μ is the growth rate, and N_0 it is the initial population size.

*Correspondence to fernanda.pinheiro@fht.org

- Gompertz:

$$N(t) = N_{\max} e^{-e^{-\mu(t-t_L)}}, \quad (\text{S2})$$

where μ is the growth rate, $N_{\max}$ is the maximum biomass concentration, and t_L the lag time.

- Logistic:

$$N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0} - 1\right) \exp(-\mu t)}, \quad (\text{S3})$$

where μ is the growth rate, N_0 it is the initial population size, and $N_{\max}$ the maximum biomass concentration.

- Richards model:

$$N(t) = \frac{N_{\max}}{[1 + \nu e^{-\mu(t-t_L)}]^{\frac{1}{\nu}}}, \quad (\text{S4})$$

where μ is the growth rate, $N_{\max}$ is the maximum biomass concentration, t_L the lag time, and ν a shape constant.

- Weibull:

$$N(t) = N_{\max} - (N_{\max} - N_0) e^{-(\mu t)^\nu}, \quad (\text{S5})$$

where μ is the growth rate, N_0 it is the initial population size, $N_{\max}$ the maximum biomass concentration and ν a shape constant.

- Morgan:

$$N(t) = \frac{N_0 K^\nu + N_{\max} t^\nu}{K^\nu + t^\nu}, \quad (\text{S6})$$

N_0 it is the initial population size, $N_{\max}$ the maximum biomass concentration, K is the time where half-maximal growth is achieved, and ν a shape constant.

- Bertalanffy:

$$N(t) = N_0 + (N_{\max} - N_0) (1 - e^{-\mu t})^{\frac{1}{\nu}}, \quad (\text{S7})$$

N_0 is the starting condition, $N_{\max}$ the maximum biomass concentration, μ the growth rate, and ν a shape constant.

- Piecewise Linear-Logistic

$$\begin{cases} N(t) = N_0, & t < t_L \\ N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0} - 1\right) \exp(-\mu(t - t_L))} & t_L \leq t, \end{cases} \quad (\text{S8})$$

N_0 is the starting condition, $N_{\max}$ the maximum biomass concentration, μ the growth rate, and t_L the lag time.

- Piecewise Exponential-Logistic:

$$\begin{cases} N(t) = N_0 \exp(\mu_0 t), & t < t_L \\ N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0 \exp(\mu_0 t_L)} - 1\right) \exp(-\mu(t - t_L))} & t_L \leq t, \end{cases} \quad (\text{S9})$$

N_0 is the starting condition, $N_{\max}$ the maximum biomass concentration, μ the growth rate, t_L the lag time, and μ_0 the growth during the lag phase.

1.2 Ordinary differential equation models

- Exponential:

$$\frac{dN(t)}{dt} = \mu N(t), \quad (\text{S10})$$

where μ is the growth rate.

- Hyper Gompertz [1]:

$$\frac{dN(t)}{dt} = \mu N(t) \log \left(\frac{N_{\max}}{N(t)} \right)^{(1+n)}, \quad (\text{S11})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration and n a shape constant.

- Hyper Logistic [1]:

$$\frac{dN(t)}{dt} = \frac{\mu}{N_{\max}} N(t)^{(1-n)} (N_{\max} - N(t))^{(1+n)}, \quad (\text{S12})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration and n a shape constant.

- Bertalanffy-Richards:

$$\frac{dN(t)}{dt} = \frac{\mu N(t)}{N_{\max}^n} (N_{\max}^n - N^n(t)), \quad (\text{S13})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, and n a shape constant.

- Logistic [1]:

$$\frac{dN(t)}{dt} = \mu \left(1 - \frac{N(t)}{N_{\max}} \right) N(t), \quad (\text{S14})$$

where μ is the growth rate, and $N_{\max}$ the maximum biomass concentration.

- Adjusted Logistic:

$$\frac{dN(t)}{dt} = \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^n \right) N(t), \quad (\text{S15})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration and n a shape constant.

- Gompertz [1]:

$$\frac{dN(t)}{dt} = \mu N(t) \log \left(\frac{N_{\max}}{N(t)} \right), \quad (\text{S16})$$

where μ is the growth rate, and $N_{\max}$ the maximum biomass concentration.

- Baranyi-Richards [2]:

$$\frac{dN(t)}{dt} = \frac{t^n}{t^n + t_L^n} \mu \left(1 - \frac{N(t)}{N_{\max}} \right) N(t), \quad (\text{S17})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, t_L is the lag time and n a shape constant.

- Baranyi-Roberts [3]:

$$\frac{dN(t)}{dt} = \frac{t^n}{t^n + t_L^n} \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^m \right) N(t), \quad (\text{S18})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, t_L is the lag time, n and m are shape constants.

- Piecewise Adjusted Logistic:

$$\frac{dN(t)}{dt} = \begin{cases} c N(t) & \text{for } t < t_L \\ \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^m \right) N(t) & \text{for } t \geq t_L, \end{cases} \quad (\text{S19})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, t_L is the lag time, m is shape constant, and c the growth rate during the lag phase (it can be 0).

- Triple Piecewise Adjusted Logistic:

$$\frac{dN(t)}{dt} = \begin{cases} c_1 N(t) & \text{for } t < t_L \\ \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^m \right) N(t) & \text{for } t_L \leq t < t_S \\ c_2 N(t) & \text{for } t \geq t_S, \end{cases} \quad (\text{S20})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, t_L is the lag time, m is a shape constant, c_1 the growth rate during the lag phase (it can be 0), t_S the time when the saturation phase starts, and c_2 the growth rate after the theoretical saturation point (e.g., to account for a linear increase in the time series when the growth curve should in principle be in the stationary phase).

- Triple Piecewise:

$$\frac{dN(t)}{dt} = \begin{cases} c_1 N(t) & \text{for } t < t_L \\ \mu N(t) & \text{for } t_L \leq t < t_S \\ c_2 \left(1 - \frac{N(t)}{N_{\max}} \right) N(t) & \text{for } t \geq t_S, \end{cases} \quad (\text{S21})$$

where μ is the growth rate, $N_{\max}$ the maximum biomass concentration, t_L is the lag time, c_1 is the growth rate during the lag phase (it can be 0), t_S is the time when stationary phase starts, and c_2 is the growth rate after the theoretical saturation point (e.g., to account for a linear increase in the time series when the growth curve should in principle be in the stationary phase).

- Triple Piecewise Exponential:

$$\frac{dN(t)}{dt} = \begin{cases} c_1 N(t) & \text{for } t < t_L \\ \mu N(t) & \text{for } t_L \leq t < t_S \\ c_2 N(t) & \text{for } t \geq t_S, \end{cases} \quad (\text{S22})$$

where μ is the growth rate, t_L is the lag time, c_1 is the growth rate during the lag phase (it can be 0), t_S is the time when stationary phase starts, and c_2 is the growth rate after the theoretical saturation point (e.g., to account for a linear increase in the time series when the growth curve should in principle be in the stationary phase).

- Four Piecewise Exponential:

$$\frac{dN(t)}{dt} = \begin{cases} c_1 N(t) & \text{for } t < t_1 \\ \mu N(t) & \text{for } t_1 \leq t < t_2 \\ c_2 N(t) & \text{for } t_2 \leq t < t_3 \\ c_3 N(t) & \text{for } t \geq t_3, \end{cases} \quad (\text{S23})$$

where μ is the growth rate, t_1 is the lag time, c_1 is the growth rate during the lag phase (which can be 0), t_2 is the time when growth occurs after the exponential phase, c_2 is the growth rate during this phase, t_3 marks the start of the stationary phase, and c_3 is the growth rate during the stationary phase.

- Heterogeneous Population Model (HPM) [4]:

$$\begin{cases} N(t) = N_1(t) + N_2(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) \left(1 - \frac{N_1(t) + N_2(t)}{N_{\max}} \right), \end{cases} \quad (\text{S24})$$

where N_1 is the population of dormant cells, N_2 is the population of active cells capable of duplicating, μ is the growth rate, $N_{\max}$ is the maximum biomass concentration, and r_L is the lag rate, defined as the rate of transition between the N_1 and N_2 populations. Here, we assume that all cells are in the dormant state at the start (i.e., $N_1(t=0) = \text{OD}(t=0)$, and $N_2(t=0) = 0$).

- Exponential Heterogeneous Population Model:

$$\begin{cases} N(t) = N_1(t) + N_2(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t), \end{cases} \quad (\text{S25})$$

where similarly to the HPM model, N_1 and N_2 refer to the populations of dormant and active cells, respectively. μ is the growth rate, and the lag rate r_L denotes the transition between the N_1 and N_2 populations. Here, we also assume that all cells are in the dormant state at the start (i.e., $N_1(t=0) = \text{OD}(t=0)$, and $N_2(t=0) = 0$).

- Adjusted Heterogeneous Population Model:

$$\begin{cases} N(t) = N_1(t) + N_2(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) \left(1 - \left(\frac{N_1(t) + N_2(t)}{N_{\max}} \right)^m \right), \end{cases} \quad (\text{S26})$$

where similarly to all variants of the HPM model, N_1 and N_2 refer to the populations of dormant and active cells, respectively. μ is the growth rate, $N_{\max}$ the maximum biomass concentration, r_L is the lag rate (i.e. the rate of transition between $N_1(t)$ and $N_2(t)$) and m a shape constant. Here, we also assume that all cells are in the dormant state at the start (i.e., $N_1(t=0) = \text{OD}(t=0)$, and $N_2(t=0) = 0$).

- Heterogeneous Population Model with Inhibition:

$$\begin{cases} N(t) = N_1(t) + N_2(t) + N_3(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) - r_I N_2(t) \\ \frac{dN_3(t)}{dt} = r_I N_2(t), \end{cases} \quad (\text{S27})$$

where N_1 , N_2 , and N_3 refer to the populations of dormant, active, and inhibited cells. In contrast to the population of actively duplicating cells in N_2 , cells in the N_3 inhibited state are active but do not duplicate. The other parameters are the growth rate μ , the lag rate for the transition between N_1 and N_2 , r_L , and the lag rate for the transition between N_2 and N_3 , r_I . As in the cases above, we assume that all cells start in the dormant state (i.e., $N_1(t=0) = \text{OD}(t=0)$, $N_2(t=0) = N_3(t=0) = 0$).

- Heterogeneous Population Model with Inhibition and Death:

$$\begin{cases} N(t) = N_1(t) + N_2(t) + N_3(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) - r_I N_2(t) \\ \frac{dN_3(t)}{dt} = -r_D N_3(t) + r_I N_2(t), \end{cases} \quad (\text{S28})$$

where N_1 , N_2 , and N_3 refer to the populations of dormant, active, and inhibited cells. In contrast to the population of actively duplicating cells in N_2 , cells in the N_3 inhibited state are active but do not duplicate. The other parameters are the growth rate μ , the lag rate for the transition between N_1 and N_2 , r_L , the lag rate for the transition between N_2 and N_3 , r_I , and the rate at which cells die, r_D . As in the cases above, we assume that all cells start in the dormant state (i.e., $N_1(t=0) = \text{OD}(t=0)$, $N_2(t=0) = N_3(t=0) = 0$).

- Heterogeneous Population Model with Inhibition, Death and Resistance:

$$\begin{cases} N(t) = N_1(t) + N_2(t) + N_3(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) - r_I N_2(t) \\ \frac{dN_3(t)}{dt} = -r_D N_3(t) \left(1 - \frac{N_3(t)}{N_R}\right) + r_I N_2(t), \end{cases} \quad (\text{S29})$$

where μ is the growth rate, r_L is the lag rate (i.e. the rate of transition between $N_1(t)$ and $N_2(t)$), r_I is the rate of which cell are inhibited (i.e. the rate of transition between $N_2(t)$ and $N_3(t)$), r_D is the rate of which cell are die, and N_R it the number of cell that will become inactive but do not die.

1.3 ODE Systems

Kinbiont ships with several hard-coded ODE systems that serve both as ready-to-use models and as templates for user-defined extentions. Additional models can be supplied by the user in analytic form or as Julia functions.

- SIR Model (Susceptible-Infected-Recovered)

$$\begin{cases} \frac{dS}{dt} = -\beta SI, \\ \frac{dI}{dt} = \beta SI - \gamma I, \\ \frac{dR}{dt} = \gamma I. \end{cases} \quad (\text{S30})$$

Here S , I , R are the susceptible, infected, and recovered populations, β is the infection rate and γ the recovery rate.

- SIR model with birth and death

$$\begin{cases} \frac{dS}{dt} = -\beta SI + bS - dS, \\ \frac{dI}{dt} = \beta SI - \gamma I - dI, \\ \frac{dR}{dt} = \gamma I - dR, \end{cases} \quad (\text{S31})$$

where S , I , R are the susceptible, infected, and recovered populations, β is the infection rate, γ is the recovery rate. and b and d are per-capita birth and death rates.

- SIS Model (Susceptible-Infected-Susceptible)

$$\begin{cases} \frac{dS}{dt} = -\beta SI + \gamma I, \\ \frac{dI}{dt} = \beta SI - \gamma I, \end{cases} \quad (\text{S32})$$

where S and I are the susceptible and infected populations, β is the infection rate, and γ is the recovery rate.

- Lotka-Volterra predator-prey Model

$$\begin{cases} \frac{dP}{dt} = \alpha P - \beta PC, \\ \frac{dC}{dt} = -\delta C + \gamma CP, \end{cases} \quad (\text{S33})$$

where P and C denote the prey and predator populations, α is the prey birth, β is the predation rate, δ is the predator death rate, and γ is the predator efficiency.

- Lotka-Volterra with substrate limitation

$$\begin{cases} \frac{dP}{dt} = \alpha P \frac{S}{S+K} - \beta PC, \\ \frac{dC}{dt} = -\delta C + \gamma CP \\ \frac{dS}{dt} = -\alpha P \frac{S}{S+K}, \end{cases} \quad (\text{S34})$$

where P and C denote the prey and predator populations, α is the prey birth, β is the predation rate, δ is the predator death rate, and γ is the predator efficiency, S is the limiting substrate, and K is the half-saturation constant.

- Monod chemostat model

$$\begin{cases} \frac{dX}{dt} = \mu X - DX, \\ \frac{dS}{dt} = D(S_{\text{in}} - S) - \frac{\mu X}{Y} - mX, \end{cases} \quad (\text{S35})$$

with

$$\mu = \mu_m \frac{S}{K_s + S} \quad (\text{S36})$$

where X is biomass, S is the inflow substrate concentration, K_s denotes substrate affinity, D is the dilution rate, m is the maintenance coefficient, Y is yield, and μ_m is the maximal specific growth rate.

- Synthetic Chemostat Model [5]

The synthetic chemostat model tracks three state variables: biomass N , substrate concentration S , and a reporter R that summarizes the cell's macromolecular composition. The core assumption is that this composition can be partitioned into two mutually exclusive sectors, P and U , subject to the conservation constraint $P + U = 1$. Under this constraint, the combined sector content is written as an affine (linear-plus-constant) function of the scalar reporter R , parametrized by three fixed coefficients.

Bacterial growth can then be described by the following system of equations:

$$\begin{cases} \frac{dS}{dt} = D(s_r - S) - q_s \cdot N, \\ \frac{dN}{dt} = \mu \cdot N - D \cdot N, \\ \frac{dR}{dt} = \mu \cdot \left(\frac{S}{k_r + S} - R \right), \end{cases} \quad (\text{S37})$$

where

$$q_s = R \cdot \frac{QS}{k_s + S} + (1 - R) \cdot \frac{Q'S}{k'_s + S}, \quad (\text{S38})$$

$$\mu = Y \cdot q_s - a_0 \cdot R. \quad (\text{S39})$$

The synthetic chemostat model has the following parameters:

- D : Dilution rate, which represents the rate at which the medium is replaced in the chemostat.
- s_r : Inflow substrate concentration, representing the constant inflow concentration of substrate.
- q_s : Substrate consumption rate. This rate depends on R and on the substrate concentration S . The rate is described by the expression:

$$q_s = R \cdot \frac{QS}{k_s + S} + (1 - R) \cdot \frac{Q'S}{k'_s + S},$$

where R modulates the shift between two terms: one for sector P and the other for sector U .

- μ : Biomass growth rate. This rate is influenced by q_s and the parameter a_0 that penalizes the value of R . It is given by:

$$\mu = Y \cdot q_s - a_0 \cdot R,$$

where Y represents the yield coefficient (biomass per substrate consumed), and a_0 is a constant that adjusts the growth rate based on the reporter variable R .

- k_r : The half-saturation constant for the growth rate, which determines the substrate concentration at which the growth rate is half its maximum value.
- k_s, k'_s : The half-saturation constants for substrate consumption, defining the concentration at which the consumption rate of substrate is half its maximum.
- Q, Q' : Scaling factors for the substrate consumption rate, representing how effectively the substrate is utilized by the biomass for growth.
- Y : Yield coefficient for biomass production, which determines the amount of biomass produced per unit of substrate consumed.
- a_0 : A constant that penalizes the reporter variable R , adjusting the growth rate in response to the cellular composition.

- Synthetic batch model

A modification of the previous model without dilution and nutrient inflow to simulate batch cultures:

$$\begin{cases} \frac{dS}{dt} = -q_s \cdot N, \\ \frac{dN}{dt} = \mu \cdot N, \\ \frac{dR}{dt} = \mu \cdot \left(\frac{S}{k_r + S} - R \right), \end{cases} \quad (\text{S40})$$

where

$$q_s = R \frac{QS}{k_s + S} + (1 - R) \frac{Q'S}{k'_s + S}, \quad (\text{S41})$$

$$\mu = Y q_s - a_0 R. \quad (\text{S42})$$

1.4 Cybernetic Models for Bacterial Growth

Kinbiont also supports cybernetic models in which users supply a structured definition of substrates, enzymes, and allocation rules; the package then generates the ODE system automatically. Given a substrate array S_i and associated enzymes P_i , the default cybernetic formulation is:

$$\begin{cases} \frac{dS_i}{dt} = -\frac{V_{S_i} P_i S_i}{k_{S_i} + S_i} \cdot N, & \forall i = 1, \dots, n, \\ \frac{dP_i}{dt} = a_i \cdot \text{alloc}_i \cdot k_{S_i} - b_i P_i N, \\ \frac{dN}{dt} = -\sum_{i=1}^n Y_{S_i} \frac{dS_i}{dt}. \end{cases} \quad (\text{S43})$$

Where the state variables are:

- N : Time-dependent biomass concentration.
- S : Time-dependent quantity of the i -th substrate.
- P_i : The protein concentration required for the assimilation of the i -th substrate.

The parameters are:

- V_{S_i} : The rate at which the i -th substrate is consumed.
- k_{S_i} : The half-saturation constant for the i -th substrate.
- a_i : The synthesis rate for the i -th protein, which determines how fast the protein is produced based on available substrates.
- b_i : The degradation constant for the i -th protein, representing how fast the protein is degraded over time.
- alloc_i : The resource allocation rule for the i -th substrate or protein. It determines how resources are distributed among different compounds based on various factors, such as substrate availability and protein thresholds. Users can customize this function freely.
- Y_{S_i} : The yield coefficient, which describes the efficiency of biomass production per unit of substrate consumption for the i -th substrate.

1.5 Reaction Networks

Users can also define biochemical systems as reaction networks and input the in Kinbiont pipeline. To obtain this, we leverage `Catalyst.jl` [6] to automatically generate and solve the corresponding ODEs. `Catalyst.jl` allows for easy definition of chemical reaction networks using a domain-specific language (DSL) in Julia, where reaction rates and species are specified intuitively. For example, a simple Michaelis-Menten model can be defined as follows:

```
u0 = [:S => 301, :E => 100, :SE => 0, :P => 0]
ps = [:kB => 0.00166, :kD => 0.0001, :kP => 0.1]

model_Michaelis_Menten = @reaction_network begin
    kB, S + E --> SE
    kD, SE --> S + E
    kP, SE --> P + E
end
```

By using this syntax, `Catalyst.jl` automatically extracts the species and parameters and formulates the system of ODEs.

Equation Name	Equation	Parameters
Exponential	$N(t) = N_0 e^{\mu t}$	N_0, μ
Gompertz	$N(t) = N_{\max} e^{-e^{-\mu(t-t_L)}}$	$N_{\max}, \mu, t_L$
Logistic	$N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0} - 1\right) \exp(-\mu t)}$	$N_{\max}, \mu, N_0$
Richards model	$N(t) = \frac{N_{\max}}{[1 + \nu e^{-\mu(t-t_L)}]^{\frac{1}{\nu}}}$	$N_{\max}, \mu, t_L, \nu$
Weibull (growth)	$N(t) = N_{\max} - (N_{\max} - N_0) e^{-(\mu t)^{\nu}}$	$N_{\max}, N_0, \mu, \nu$
Morgan	$N(t) = \frac{N_0 K^{\nu} + N_{\max} t^{\nu}}{K^{\nu} + t^{\nu}}$	$N_0, N_{\max}, \nu, K$
Bertalanffy	$N(t) = N_0 + (N_{\max} - N_0) (1 - e^{-\mu t})^{\frac{1}{\nu}}$	$N_0, N_{\max}, \mu, \nu$
Piecewise linear-logistic	$\begin{cases} N(t) = N_0, & t < t_L \\ N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0} - 1\right) \exp(-\mu(t-t_L))}, & t_L \leq t \end{cases}$	$N_0, t_L, N_{\max}, \mu$
Piecewise exponential-logistic	$\begin{cases} N(t) = N_0 \exp(\mu_0 t), & t < t_L \\ N(t) = \frac{N_{\max}}{1 + \left(\frac{N_{\max}}{N_0 \exp(\mu_0 t_L)} - 1\right) \exp(-\mu(t-t_L))}, & t_L \leq t \end{cases}$	$N_0, \mu_0, N_{\max}, \mu, t_L$

Equation Name	Equation	Parameters
Exponential	$\frac{dN(t)}{dt} = \mu N(t)$	μ
Hyper Gompertz	$\frac{dN(t)}{dt} = \mu N(t) \left(\log \frac{N_{\max}}{N(t)} \right)^{(1+n)}$	$\mu, N_{\max}, n$
Hyper Logistic	$\frac{dN(t)}{dt} = \frac{\mu}{N_{\max}} N(t)^{(1-n)} (N_{\max} - N(t))^{(1+n)}$	$\mu, N_{\max}, n$
Bertalanffy-Richards	$\frac{dN(t)}{dt} = \frac{\mu}{N_{\max}^n} N(t)^n (N_{\max} - N(t)^n)$	$\mu, N_{\max}, n$
Logistic	$\frac{dN(t)}{dt} = \mu \left(1 - \frac{N(t)}{N_{\max}} \right) N(t)$	$\mu, N_{\max}$
Adjusted Logistic	$\frac{dN(t)}{dt} = \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^n \right) N(t)$	$\mu, N_{\max}, n$
Gompertz	$\frac{dN(t)}{dt} = \mu N(t) \log \left(\frac{N_{\max}}{N(t)} \right)$	$\mu, N_{\max}$
Baranyi Richards	$\frac{dN(t)}{dt} = \frac{t^n}{t^n + t_L^n} \mu \left(1 - \frac{N(t)}{N_{\max}} \right) N(t)$	$\mu, N_{\max}, t_L, n$
Baranyi Roberts	$\frac{dN(t)}{dt} = \frac{t^n}{t^n + t_L^n} \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^m \right) N(t)$	$\mu, N_{\max}, t_L, n, m$
Piecewise Adjusted Logistic	$\frac{dN(t)}{dt} = \begin{cases} c N(t) & t < t_L \\ \mu \left(1 - \left(\frac{N(t)}{N_{\max}} \right)^m \right) N(t) & t \geq t_L \end{cases}$	$\mu, N_{\max}, t_L, m, c$
Heterogeneous Population Model (HPM)	$\begin{cases} N(t) = N_1(t) + N_2(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) \left(1 - \frac{N_1(t) + N_2(t)}{N_{\max}} \right) \end{cases}$	$\mu, N_{\max}, r_L$
Exponential Heterogeneous Population Model	$\begin{cases} N(t) = N_1(t) + N_2(t) \\ \frac{dN_1(t)}{dt} = -r_L N_1(t) \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) \end{cases}$	μ, r_L
Heterogeneous Population Model with Inhibition	$\begin{cases} N(t) = N_1(t) + N_2(t) + N_3(t), \\ \frac{dN_1(t)}{dt} = -r_L N_1(t), \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) - r_1 N_2(t), \\ \frac{dN_3(t)}{dt} = r_1 N_2(t), \end{cases}$	μ, r_L, r_1
Heterogeneous Population Model with Inhibition and Death	$\begin{cases} N(t) = N_1(t) + N_2(t) + N_3(t), \\ \frac{dN_1(t)}{dt} = -r_L N_1(t), \\ \frac{dN_2(t)}{dt} = r_L N_1(t) + \mu N_2(t) - r_1 N_2(t), \\ \frac{dN_3(t)}{dt} = -r_D N_3(t) + r_1 N_2(t), \end{cases}$	μ, r_L, r_1, r_D

Supplementary Table 1: **Examples of microbial dynamic models implemented in Kinbiont.** Summary table showing a representative subset of nonlinear (NL) and ordinary differential equation (ODE)-based models implemented in Kinbiont, commonly used to fit single-dimensional microbial kinetics data. Although parameter interpretations may vary slightly across models, here N_0 denotes the initial population size, μ the exponential growth rate, $N_{\max}$ the saturation level or maximum biomass concentration, ν typically represents an adjustable shape parameter, and t_L the lag time [1, 7]. Section 1 of the Supplementary Information provides detailed descriptions of all model-specific parameters. The complete list of models supported by Kinbiont is also available on the accompanying GitHub repository [8].

2 List of error functions

Kinbiont includes the following expressions for the loss function ($\mathcal{L}(P)$ where P is a set of model specific parameters):

1. $L2$ norm of the difference between the numerical solution (i.e., $\hat{N}(t_i, P)$ for $i = 1, \dots, n$, where n is the number of data points) of the desired model and the given data (i.e., $N(t_i)$ for $i = 1, \dots, n$):

$$\mathcal{L}(P) = \frac{1}{n} \sum_{i=1}^n \left(N(t_i) - \hat{N}(t_i, P) \right)^2, \quad (\text{S44})$$

where n is the number of data points.

2. Relative error between the solution and data:

$$\mathcal{L}(P) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \left(1 - \frac{N(t_i)}{\hat{N}(t_i, P)} \right)^2, \quad (\text{S45})$$

where n is the number of data points.

3. $L2$ norm of the difference between the specific growth rate of the numerical solution of the desired model and the corresponding derivatives of the data:

$$\mathcal{L}(P) = \frac{1}{n} \sum_{i=1}^n \left(\frac{dN(t_i)}{dt} - \frac{d\hat{N}(t_i, P)}{dt} \right)^2, \quad (\text{S46})$$

where n is the number of data points.

4. Logarithm of the $L2$ norm of the difference between the numerical solution of the desired model and the given data:

$$\mathcal{L}(P) = \log \left(\frac{1}{n} \sum_{i=1}^n \left(N(t_i) - \hat{N}(t_i, P) \right)^2 \right), \quad (\text{S47})$$

where n is the number of data points.

5. Logarithm of the relative error between the solution and data:

$$\mathcal{L}(P) = \log \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{2} \left(1 - \frac{N(t_i)}{\hat{N}(t_i, P)} \right)^2 \right), \quad (\text{S48})$$

where n is the number of data points.

6. $L2$ norm of the difference between the numerical solution of the desired model and the data, weighted by the standard deviation of empirical blank data:

$$\mathcal{L}(P) = \frac{1}{n} \sum_{i=1}^n \left(\frac{N(t_i) - \hat{N}(t_i, P)}{\text{std blank}} \right)^2, \quad (\text{S49})$$

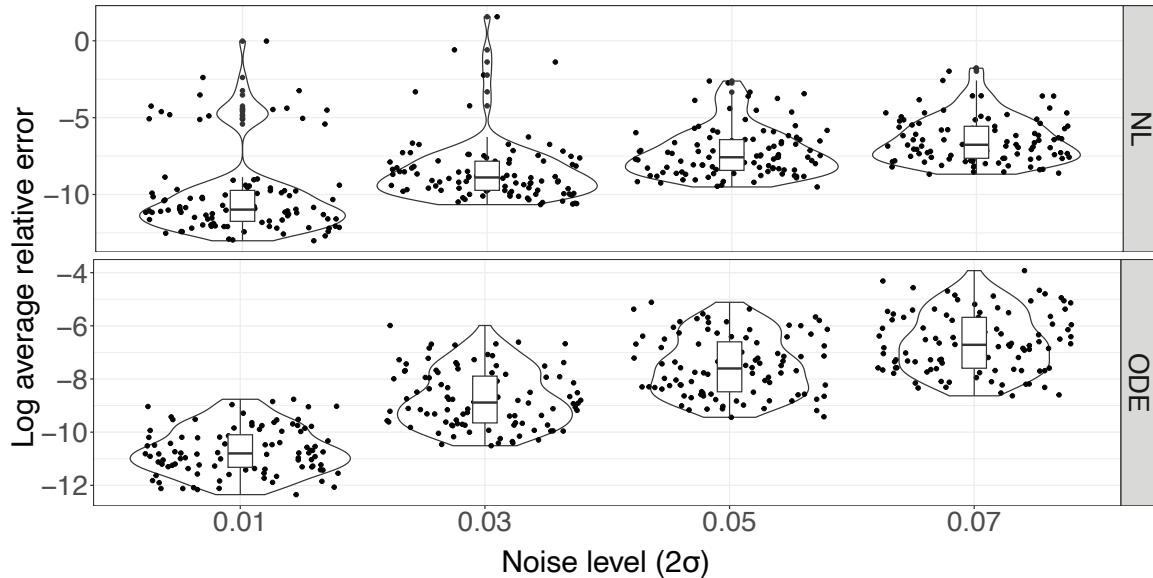
where std blank is the standard deviation of the empirical blank data, and n is the number of data points.

3 Comparison of parameter inference between ODE and NL fit

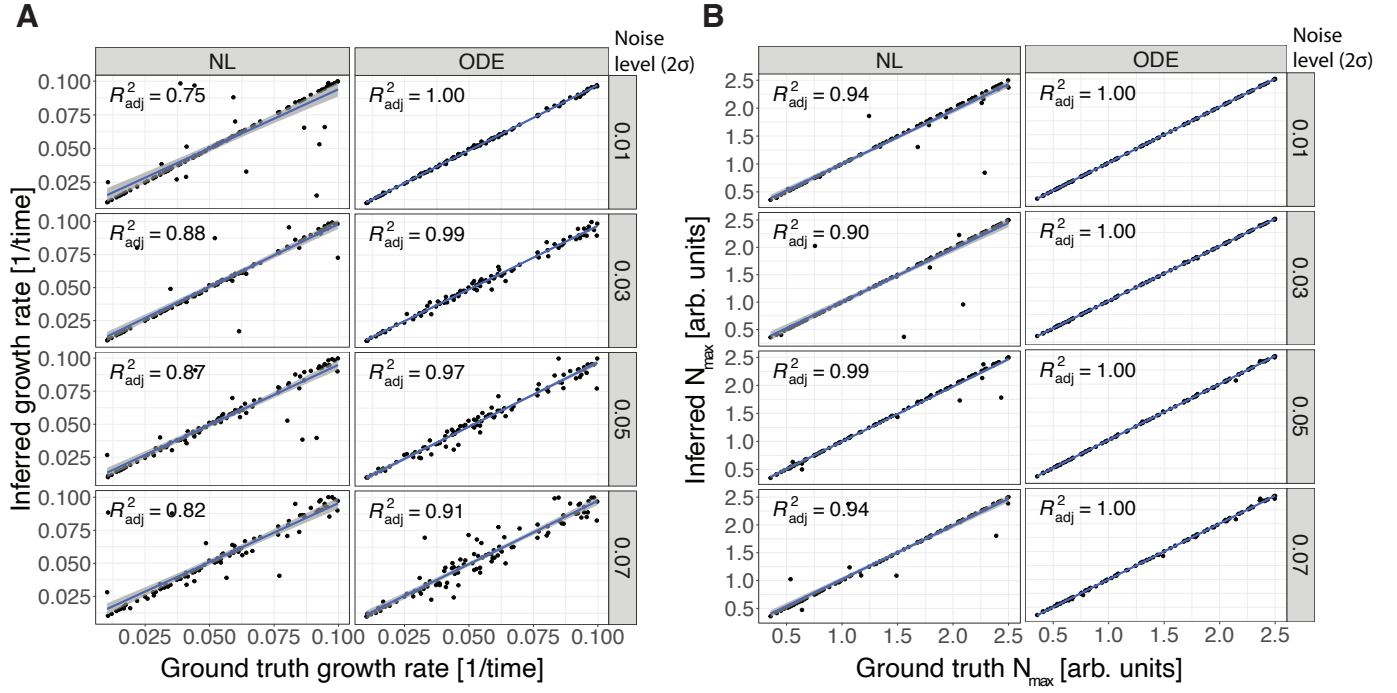
To compare the performance of the nonlinear (NL) and ordinary differential equation (ODE) fitting schemes, we carried out an analysis using synthetic data. We used the logistic model, which is a default option in Kinbiont for both fitting methods Eqs. S3, and S14. Importantly, the NL version of the model includes an additional parameter, as the initial value of the function is not constrained to the observed measurement, as it is in the ODE fits.

Using Kinbiont, we generated 100 synthetic logistic curves with random parameters sampled from uniform distributions: $\mu \sim \mathcal{U}(0.01, 0.1)$, $N_{\max} \sim \mathcal{U}(0.3, 2.5)$, and $N_0 \sim \mathcal{U}(0.1, N_{\max}/2)$. Gaussian noise with a mean of 0 and $2\sigma = \{0.01, 0.03, 0.05, 0.07\}$ was added to each point of the curves. The resulting curves were then fit using the corresponding ODE and NL methods. The loss function (average relative error) and inferred model parameters are presented in Supplementary Figures 1 and 2.

As shown in Supplementary Figure 1, the ODE model consistently yields a lower loss function when estimating parameters. Differences in parameter estimates, likely due to the additional degree of freedom in the NL model, can affect the stability of the results. However, these differences can be mitigated by adjusting the optimization scheme, such as increasing the number of iterations, making more informed initial parameter guesses or using a multistart approach. Supplementary Figure 2 further corroborates these findings by comparing the inferred parameters to the ground truth, demonstrating that both methods produce accurate inferences of microbial dynamics parameters. These results also emphasize the reliability of ODE model fits, particularly in scenarios where closed-form solutions for NL models are unavailable.



Supplementary Figure 1: **Distributions of average relative error for NL and ODE logistic model fits.** Distribution of the logarithm of the average relative error obtained by fitting 100 synthetic logistic curves with four different levels of Gaussian noise using the NL (upper panel) and ODE (lower panel) fitting schemes. While both methods perform well on synthetic data with noise, fitting the model in ODE form consistently yields lower average loss function values. Differences in parameter estimates provided by these two methods should not affect biological conclusions (see Supplementary Figure 2) and, as discussed in the text, can be mitigated by adjusting fit parameters (e.g., the chosen optimizer, number of iterations, tolerance, stopping criteria, or improved initial condition guesses). Here, each distribution contains 100 samples, and the boxplots represent the 0.25, 0.50, and 0.75 quantiles of the distributions.



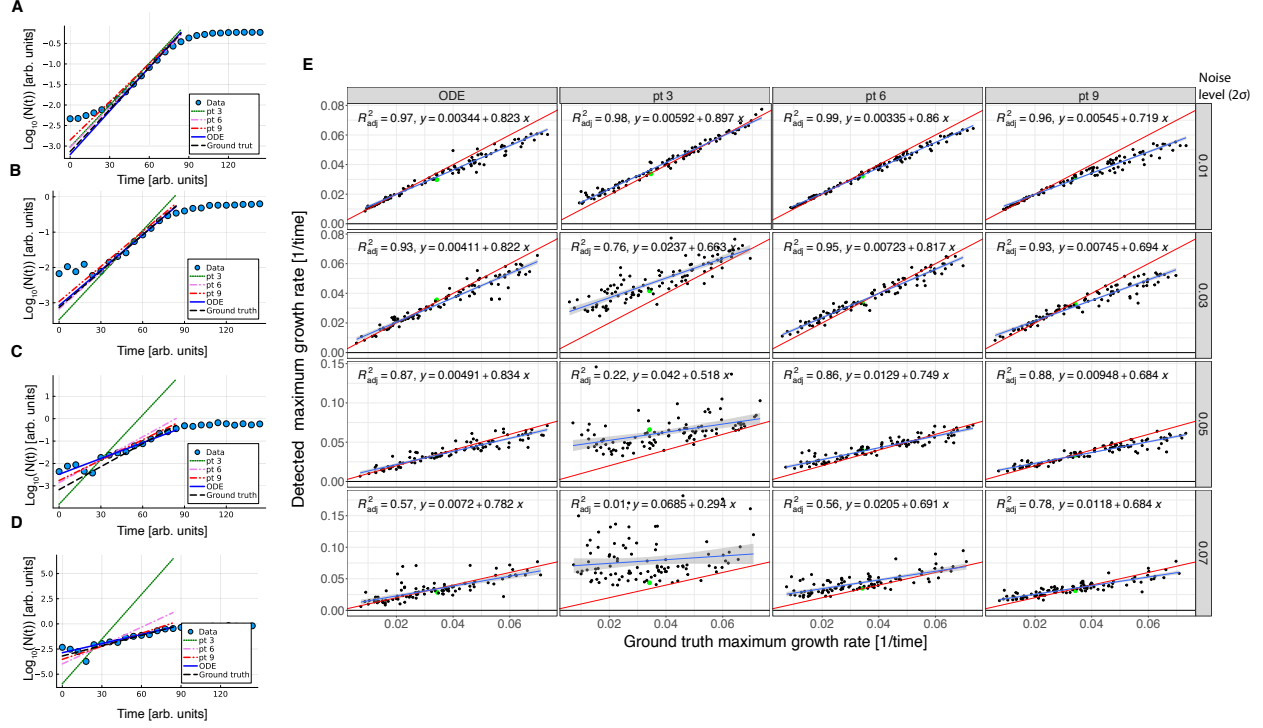
Supplementary Figure 2: **Parameters inference benchmark.** Scatter plots comparing the ground truth parameters with those inferred from fitting 100 synthetic logistic curves using NL and ODE fitting schemes under four different levels of Gaussian noise. **A** Logistic growth parameter μ . **B** N_{max} . Each plot includes a linear fit (blue line) with a shaded 95% confidence interval (gray), and the adjusted R^2 correlation coefficient is annotated.

4 Comparison of exponential growth rate estimates using a model-based or a non-parametric method

Given that different models interpret the growth rate μ differently, direct comparisons of μ values across models are not always straightforward [9]. A common practice in bacterial kinetics analysis is to focus on the maximum exponential growth rate, expressed as $\mu_{\max} = \max\left(\frac{d}{dt} \log(N(t)/N_0)\right)$, where N is the measurement and N_0 is the initial value. In batch culture, $\mu_{\max}$ is usually obtained from the specific growth rate, which is typically determined through a linear fit to the logarithm of the growth curve. Applying this method within a sliding window allows for tracking the dynamics of the exponential growth rate over time and identifying its maximum value or average.

In Supplementary Figure 3, we compare the ground truth maximum exponential/specific growth rate with estimates obtained using both an ODE-based approach and the sliding window method with different window sizes. This comparison involved generating 100 synthetic curves using the adjusted HPM model (Eq. S26) with parameters sampled from uniform distributions: $\mu \sim \mathcal{U}(0.01, 0.08)$, $N_{\max} \sim \mathcal{U}(0.4, 1.5)$, $n \sim \mathcal{U}(0.07, 5)$, and $r_L \sim \mathcal{U}(0.02, 0.09)$. Gaussian noise with a mean of 0 and $2\sigma = 0.01, 0.03, 0.05, 0.07$ was added to each data point to evaluate the effect of noise on parameter inference accuracy.

The results indicate that the ODE-based approach consistently provides more accurate maximum growth rate estimates than the sliding window method. Since the fitted model is inherently noise-free, this method eliminates the need for hyperparameters, such as the smoothing window size, making it a robust alternative for obtaining maximum growth rates. This approach also enhances the detection and standardization of growth rate observables. For completeness, Kinbiont reports both the growth rates of the selected model and the maximum exponential growth rate evaluated on the growth curve generated by the inferred model.

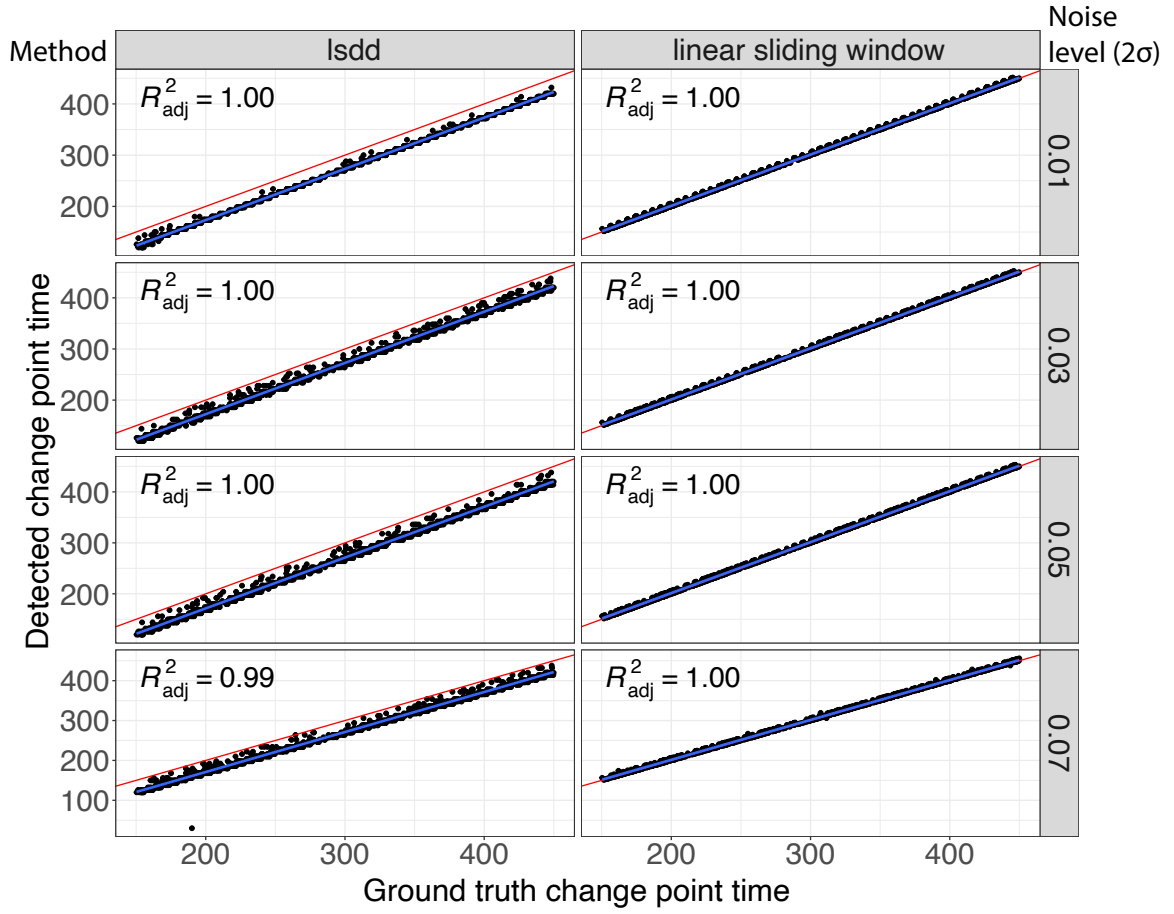


Supplementary Figure 3: Detection of maximum exponential growth rate. Comparison of the maximum growth rate estimates obtained using both model-based and non-parametric methods on a synthetic dataset (100 curves) generated with the aHPM model under different levels of Gaussian noise. The synthetic data were used to fit the aHPM model to each curve, and the maximum growth rate was subsequently evaluated using the inferred parameters (ODE) and log-linear fits on a sliding window with varying sizes (3 to 9 points, in steps of 3). **A - D** show examples of the estimated maximum growth rate alongside the data (dots) in a semi-logarithmic scale, illustrating the results from different methods. In **E**, scatter plots compare the inferred values to the ground truth for the maximum exponential growth rate across the 100 synthetic curves. The red line represents the identity line (perfect fit), the blue line shows the linear fit, the gray area indicates the 95% confidence interval, and the green dots correspond to the parameter values for the linear plots in **A - D**. In particular, the linear fits and the correlation coefficient show that ODE fits outperform the sliding window approach in the detection of maximum growth rate at any noise level. While a larger sliding window works better in comparison to a smaller ones at higher noise, it underestimates higher values of the growth rate. When using the latter method, users should set the size of the window according to the noise level and rate of data acquisition.

5 Change point detection inference

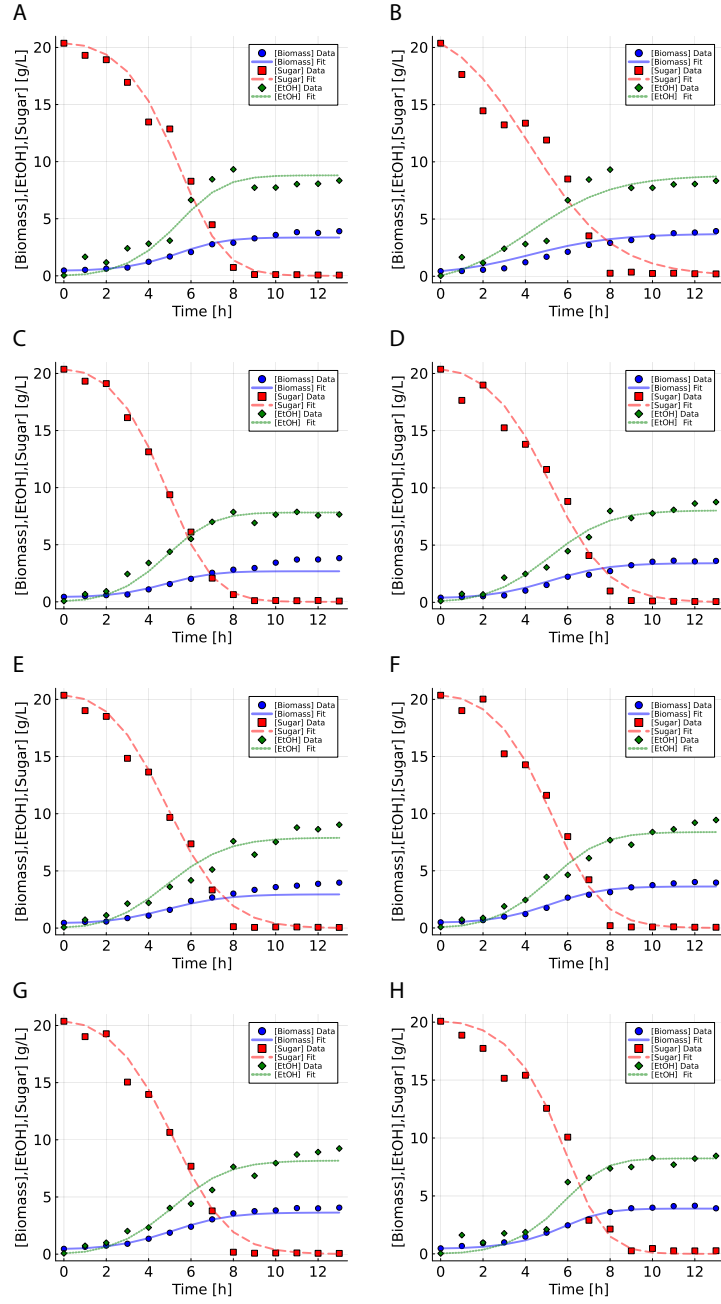
To benchmark the accuracy of the signal processing methods integrated in Kinbiont for detecting changes in microbial growth dynamics, we performed an *in silico* test. We generated 1000 time series, each containing a change point, using the following procedure: First, we simulated the HPM model (Eq. S24) with parameters $\mu_{\max} = 0.06$, $r_L = 0.01$, and $N_{\max} = 1.01$ over the time interval $(0, t_{\text{cp}})$, where the change point time t_{cp} was drawn from a uniform distribution $t_{\text{cp}} \sim \mathcal{U}(150, 450)$. To ensure continuity, we then simulated a second logistic model (cf. Supplementary Table 1) with parameters $\mu_{\max} = 0.08$ and $N_{\max} = 2$ over the time span $(t_{\text{cp}}, 600)$. Gaussian noise with a mean of 0 and $2\sigma = 0.01, 0.03, 0.05, 0.07$ was added to each generated curve.

We applied two change point detection methods implemented in Kinbiont: the least square density difference (LSDD) method [10] and the sliding window with linear fitting approach (described in the Methods section). In Supplementary Figure 4, we compared the inferred change point times with the ground truth values. Both methods demonstrated strong performance and were effective in detecting changes in microbial growth dynamics. However, a systematic error was observed in the change point estimates using the LSDD method due to the implementation in `ChangePointDetection.jl`, where the dissimilarity curve’s value is assigned to the start of the window rather than the middle. This deviation can be easily corrected by shifting the estimate by half the window size, according to the user’s needs.

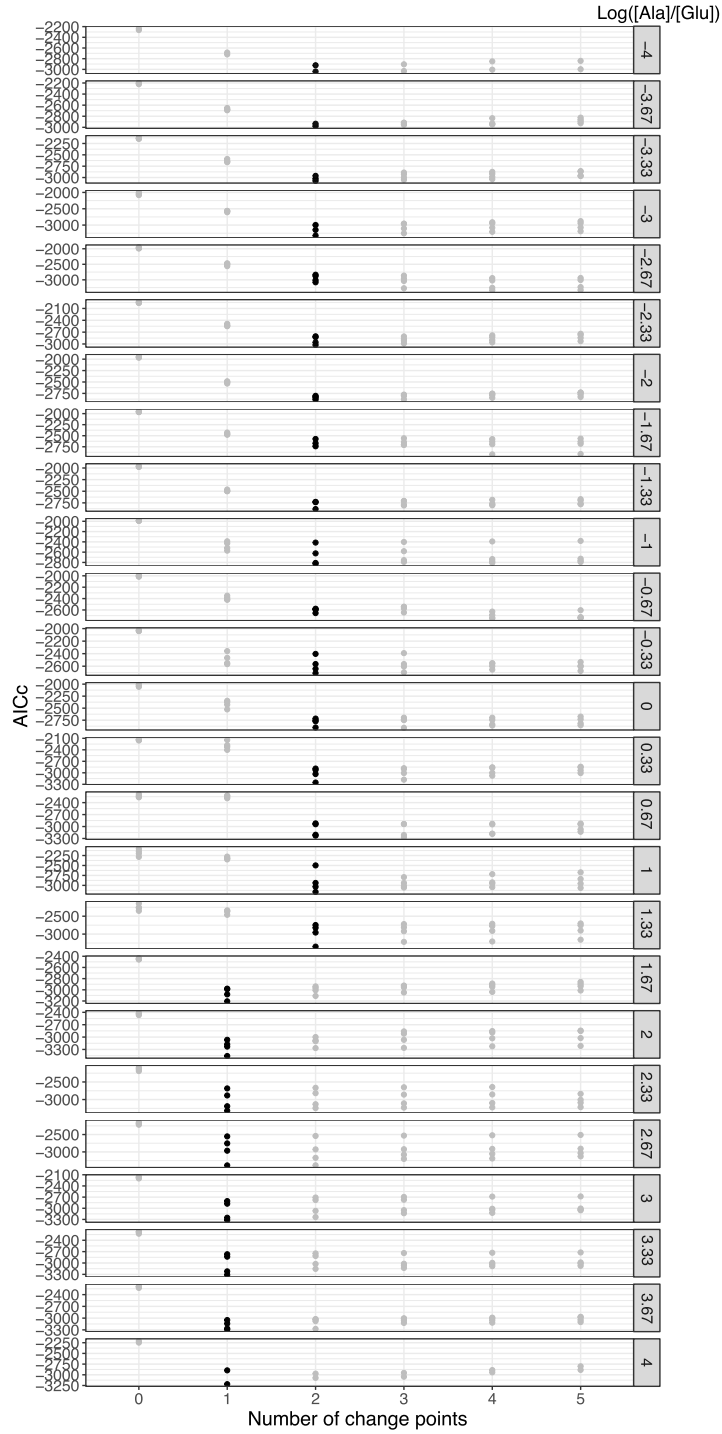


Supplementary Figure 4: **Change point detection benchmark.** Comparison of change point detection algorithms using 1000 synthetic time series. The x-axis represents the ground truth change point, and the y-axis shows the inferred value. Blue lines indicate a linear fit, with gray areas representing the 95% confidence interval, and the adjusted R^2 correlation coefficient is noted. Each row corresponds to a different noise level, and the two columns compare the change point detection methods implemented in Kinbiont. As discussed in the Supplementary Information, the systematic difference between the LSDD and linear window methods arises from the choice to assign the value of the dissimilarity curve to the start of the window in the LSDD method, and to the window mid-point when using the linear window method.

6 Supporting information for the parameter inference of complex microbial kinetics

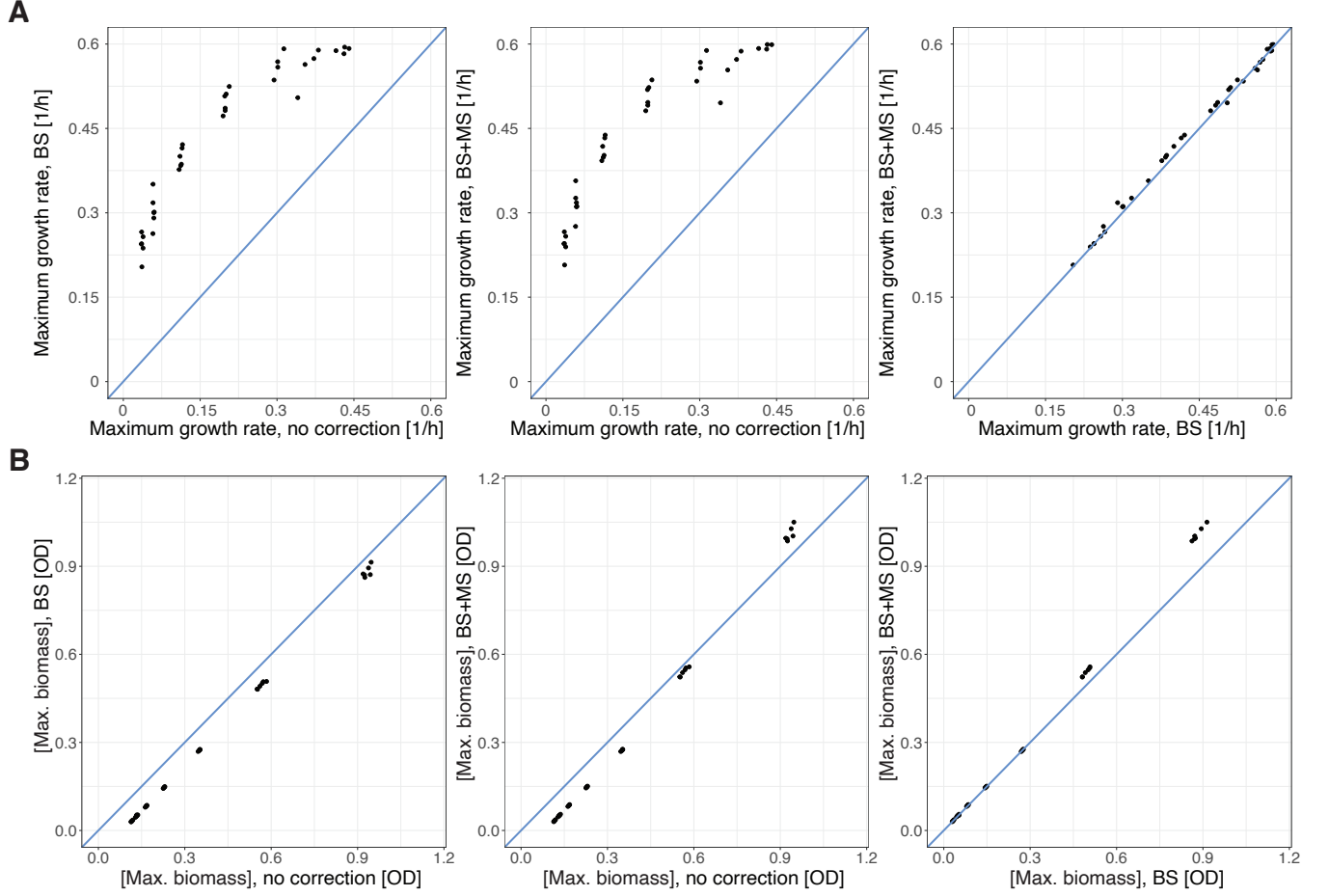


Supplementary Figure 5: **Fitting the Monod-Ierusalimsky model.** Fits for all the curves of [11]. The fitted parameters are available in the Kinbiont GitHub repository.

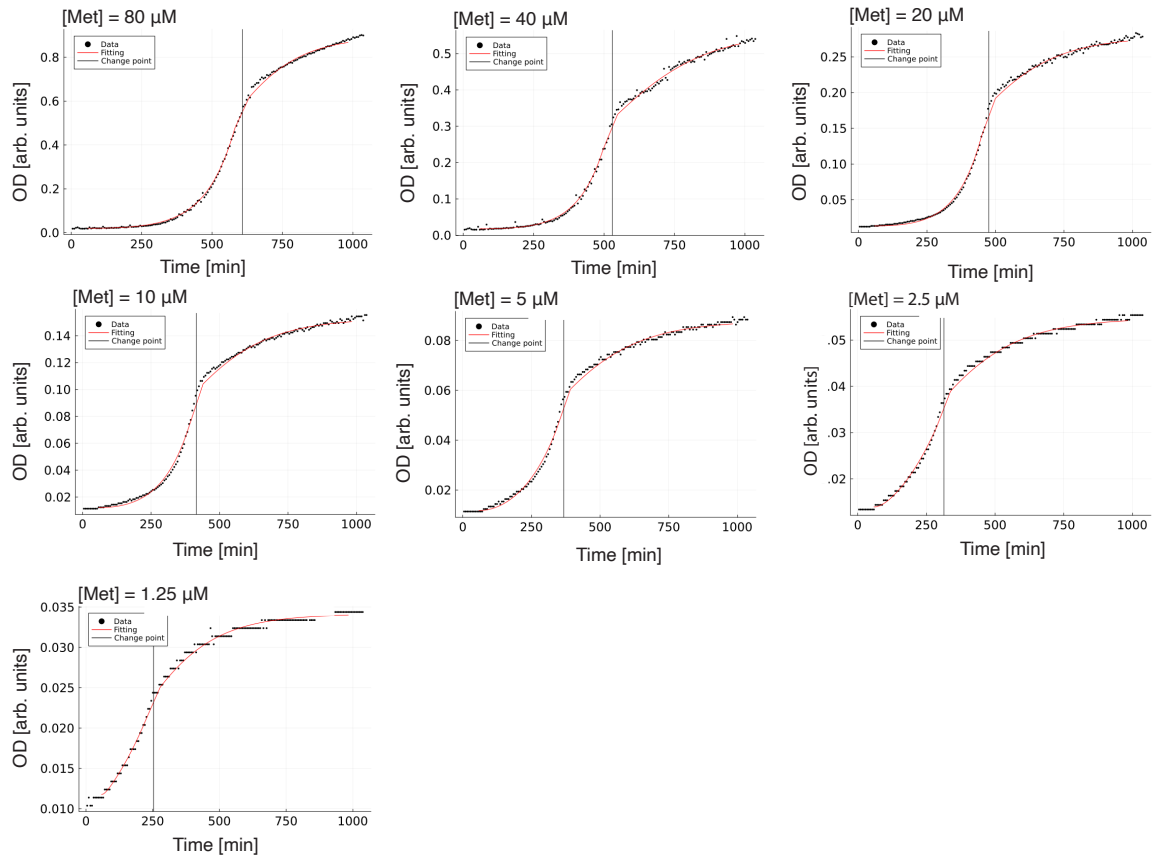


Supplementary Figure 6: **Segmented fits for the diauxic growth dataset.** AICc values for segmented fits of the dataset from Ref. [12] with varying numbers of change points. Black dots mark the solution used in the main text analyses. In particular, we observe a sharp transition between 1 and 2 change points at $\text{Log}([\text{Ala}]/[\text{Glu}]) = 1.67$.

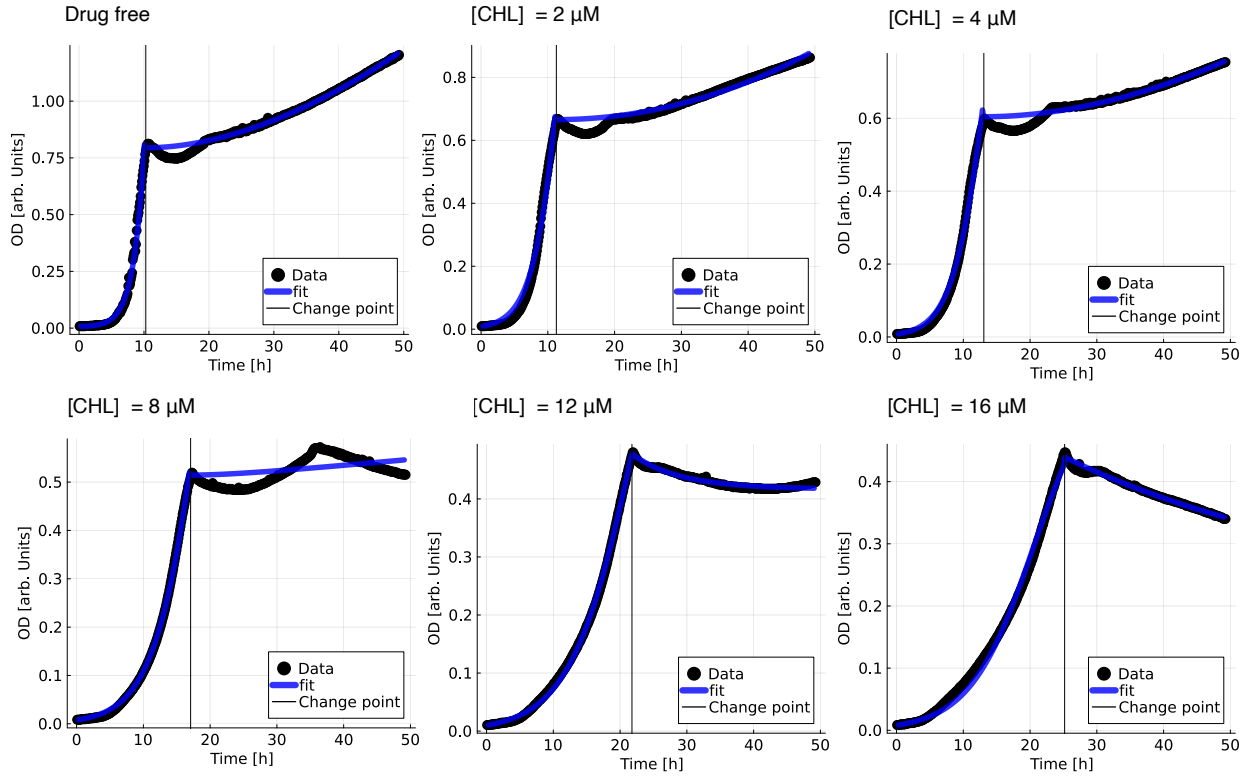
7 Supporting information and tables for the automatic inference of empirical laws



Supplementary Figure 7: **Effects of multiple scattering and background subtraction corrections on inferred parameters.** Scatter plots comparing parameters obtained by fitting the data of methionine auxotrophs following pre-processing with different methods: *(i)* no pre-processing, *(ii)* blank subtraction (BS) using the average value of the blank wells, and *(iii)* blank subtraction followed by correction for multiple scattering effects (BS+MS). The solid blue line represents the identity line. **A** The maximum growth rate estimates are significantly influenced by the background value. In particular, failure to account for background effects leads to an underestimation of growth rates. In contrast, neglecting multiple scattering correction has minimal impact, as the exponential phase typically occurs at OD values within the linear detection range [13]. **B** The maximum biomass concentration estimate suffers from a small but systematic overestimation due to the absence of blank subtraction. In addition, it is affected by multiple scattering correction, particularly at OD values beyond the linear detection range, where values are underestimated.



Supplementary Figure 8: **Model-based inference of auxotroph growth curves.** Measured growth curves of the $\Delta metA$ knockout (dots) together with segmented fit results (lines) for all growth experiments with supplemented methionine. Segmented fits were performed using one change point, with the exponential HPM (Eq. S25) and logistic models (Eq. S14) as candidates (ODE representation fits). The first segment was used to determine the exponential growth rate, while the second segment was used to estimate the saturation level or maximum biomass concentration.



Supplementary Figure 9: **Model-based inference of chloramphenicol dose-response.** Measured growth curves of *E. coli* (dots) together with segmented fit results (lines) for all growth experiments with different drug concentration. Segmented fits were performed using one change point, with the exponential HPM (Eq. S25) and logistic models (Eq. S14) as candidates (ODE representation fits). The first segment was used to determine the exponential growth rate, used in figure 6D.

Equation ID	Equation	Complexity	MSE
I	0.5006	1	0.0409
II	$\frac{[\text{Met}]}{[\text{Met}] - (-6.9190)}$	5	0.0242
III	$\frac{\frac{1.4060}{[\text{Met}]} + 2.2387}{([\text{Met}] + 2.2387)}$	7	0.0043
IV	$\frac{\left(\frac{[\text{Met}]}{1.3067}\right)}{([\text{Met}] - (-1.9690)) - 0.0525}$	9	0.0042
V	$\frac{\left(\frac{[\text{Met}]}{1.3067}\right)}{(([\text{Met}] - (-2.3203)) - 0.4808) - 0.0682}$	11	0.0041
VI	$\frac{\frac{1.4059}{[\text{Met}]} + \frac{2.2388}{[\text{Met}] + 0.0375} - (-1.3425)}{([\text{Met}] + \frac{1.4059}{[\text{Met}] + 0.0375} - (-1.3425))}$	13	0.0032
VII	$\frac{\frac{1.4059}{[\text{Met}]} + \frac{2.2388}{[\text{Met}] + 0.0375} - (-1.3425)}{\left([\text{Met}] + \frac{2.2388}{[\text{Met}] + 0.0375} - (-1.3425)\right)}$	15	0.0032
VIII	$\frac{\frac{1.4059}{[\text{Met}]} + \frac{2.2388}{[\text{Met}] + 0.0375} - (-1.3425)}{(([\text{Met}] \times 1.0996 + 1.6748) - \left(0.7466 + \frac{-0.9192}{-0.9192 + [\text{Met}]}\right))}$	17	0.0026
IX	$\frac{\frac{1.4059}{[\text{Met}]} + \frac{2.2388}{[\text{Met}] + 0.0375} - (-1.3425)}{(([\text{Met}] \times 1.0997 + 1.6748) - \left(0.7466 + \frac{-0.9193}{-0.9190 + [\text{Met}]}\right) - 0.3987)}$	19	0.0025

Supplementary Table 2: Hall of fame from the symbolic-regression search for the growth rate of an *E. coli* $\Delta metA$ knockout expressing GFP. Each candidate equation relates the growth-rate observable to the methionine concentration [Met] and is listed with its mean-squared error and expression complexity. The top-ranked expressions are plotted in Fig. 6 of main text.

ID	Equation	Complexity	MSE
X	0.3108	1	0.1250
XI	$[\text{Met}] \times 0.0138$	3	0.0004
XII	$(-1.5411 - [\text{Met}]) \times -0.01340$	5	0.0001
XIII	$(0.0206 - ([\text{Met}] \times 0.9860)) + [\text{Met}]$	7	0.0001
XIV	$\left([\text{Met}] + 0.7543 + \frac{[\text{Met}]}{[\text{Met}] + 1.4107}\right) \times 0.0134$	11	0.0001
XV	$(-0.1083 - [\text{Met}]) \times ((0.0010 - [\text{Met}]) \times 0.0001 + (-0.0143)) - (-0.0127)$	13	0.0001

Supplementary Table 3: Hall of fame from the symbolic-regression search for the total growth of an *E. coli* $\Delta metA$ knockout expressing GFP. Each candidate equation relates the total growth to the methionine concentration [Met] and is listed with its mean-squared error and expression complexity. The top-ranked expressions are plotted in Fig. 6 of main text.

Equation	Complexity	MSE
0.5377	1	0.02660
$[\text{Met}] - ([\text{Met}] + (-0.5377))$	5	0.02660
$\left(\frac{[\text{Met}]}{[\text{Met}] + 1.1924}\right) \times 0.6814$	7	0.0007
$\left(\frac{[\text{Met}] - 0.0056}{[\text{Met}] + 1.1924}\right) \times 0.6810$	9	0.0007
$\left(\frac{[\text{Met}] - \left(\frac{0.0056}{1.4040}\right)}{[\text{Met}] + 1.195}\right) \times 0.6814$	11	0.0007
$\frac{[\text{Met}]}{\left(\frac{[\text{Met}]}{0.6804} + 1.5218\right) + \left(\frac{0.6804}{[\text{Met}] + 0.7551}\right)}$	13	0.0007
$\frac{[\text{Met}]}{\left(\frac{[\text{Met}]}{0.6804} + (0.4959 + 0.9799)\right) + \left(\frac{0.8987}{[\text{Met}] + 1.0506}\right)}$	15	0.0007
$\left(\frac{[\text{Met}]}{\left(\frac{[\text{Met}] + (-0.9038)}{1.9396 - \left(\frac{-0.1607}{[\text{Met}] + (-0.9038)}\right)}\right)}\right) \times 0.6814$	17	0.0004
$\left(\frac{[\text{Met}]}{\left(\frac{[\text{Met}] + (-0.9038)}{1.9038 - \left(\frac{-0.2427}{[\text{Met}] - 1.9397 + 0.8696}\right)}\right)}\right) \times 0.6814$	19	0.0004

Supplementary Table 4: Hall of fame of symbolic regression for the growth rate for an *E. coli* $\Delta metA$ knockout expressing a red fluorescent marker (mCherry). [Met] represents the methionine concentration.

Equation	Complexity	MSE
0.2957	1	0.1098
$[\text{Met}] \times 0.0130$	3	0.0005
$([\text{Met}] - (-1.8942)) \times 0.01$	5	0.0002
$(([\text{Met}] - (-0.0234)) \times 1.0120) - [\text{Met}]$	7	0.0002
$((([\text{Met}] + 0.9123) - 0.9364) \times (-0.9874)) + [\text{Met}]$	9	0.0002
$(([\text{Met}] - (([\text{Met}] - (-0.7090)) \times 1.0)) - (-0.7240)) \times [\text{Met}]$	11	0.0001
$\frac{(([\text{Met}] - (-0.4853)) \times 0.0107) - ((([\text{Met}] - (-0.4853)) \times [\text{Met}]) \times 0.0001)}{0.7190}$	15	0.0001
$\left(\frac{([\text{Met}] + 0.3920)}{7.84 + \left(\frac{[\text{Met}]}{7.849 + (-1.911)}\right) + 0.3921} \times 0.1021\right) \times \frac{0.1020}{0.8741}$	19	0.0001

Supplementary Table 5: Hall of fame of symbolic regression for the total growth for an *E. coli* $\Delta metA$ knockout expressing a red fluorescent marker (mCherry). [Met] represents the methionine concentration.

ID	Equation	Complexity	MSE
XVI	0.3504	1	0.0193
XVII	$(([\text{CHL}] \times (-0.0229)) - (-0.0510))$	5	0.0026
XVIII	$\left(\frac{-0.7109}{([\text{CHL}] \times (-0.1629)) + (-1.2307)}\right)$	7	0.0008
XIX	$\left(\frac{-0.5985}{([\text{CHL}] \times (-0.3773)) \times (-0.0899)}\right) + (-0.9843)$	11	0.0008
XX	$\left(\frac{0.8584}{([\text{CHL}] - 0.2431) \times (-0.0975) + 0.0079}\right) + (-1.1715)$	13	0.0008
XXI	$\left(\frac{-0.6326}{(((((\text{CHL}] + (-0.6701)) \times 0.3318) - (-0.0293)) + [\text{CHL}]) \times (-0.1089)) + (-1.1161))}\right)$	15	0.0008
XXIII	$\left(\frac{-0.6378}{((([\text{CHL}] \times (-0.1066)) + ((\frac{[\text{CHL}] + 0.6336}{-0.1066 - [\text{CHL}]}) \times (-0.0106)))}\right) + (-1.1591)$	19	0.0008

Supplementary Table 6: Hall of fame from the symbolic regression search for the chloramphenicol dose-response experiment. Each candidate equation relates the growth rate to the chloramphenicol concentration [CHL] and is listed with its mean-squared error and expression complexity. The top ranked expressions are plotted in Fig. 6 of main text.

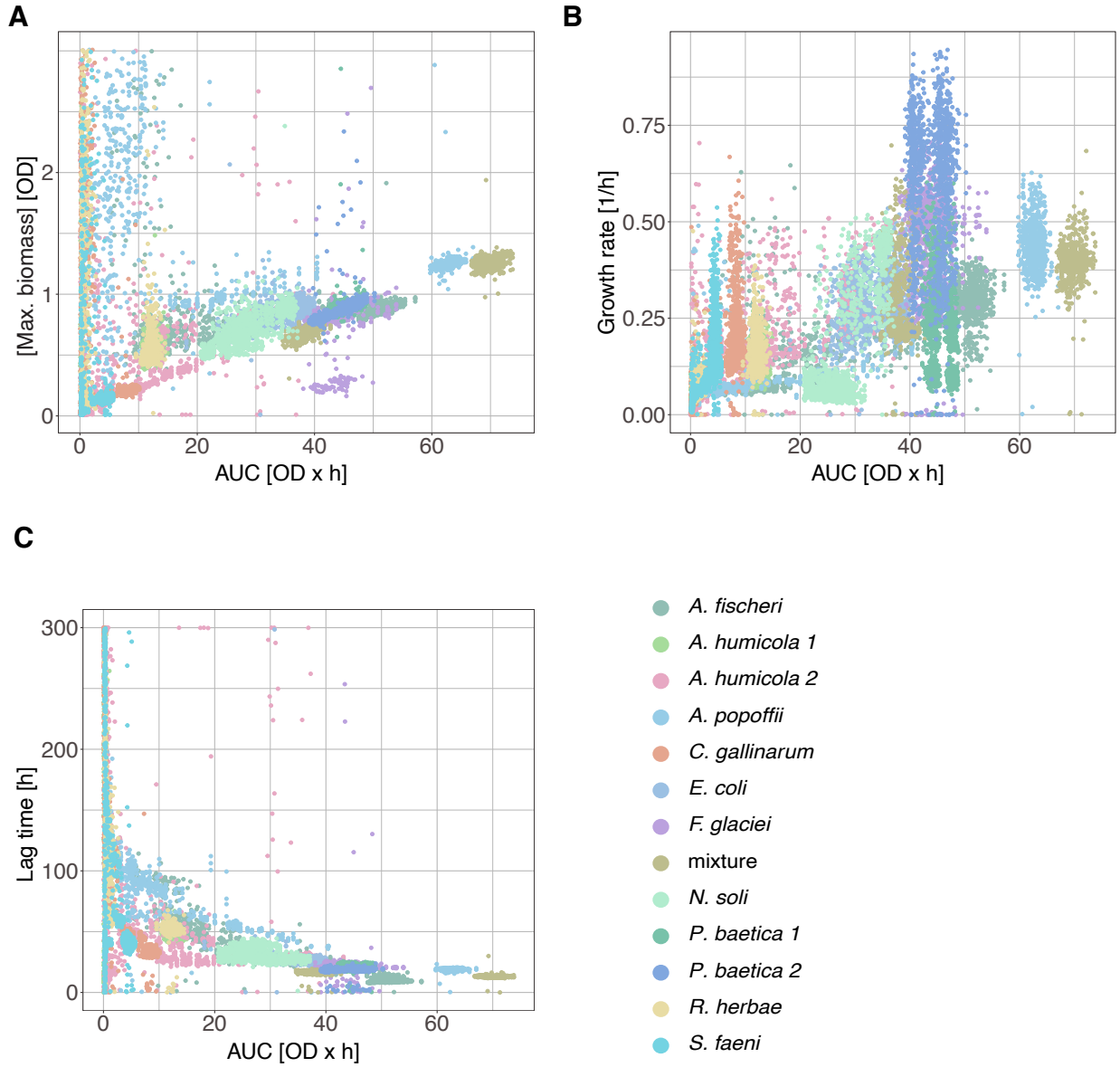
8 Supporting information for the characterization of chemical mixture landscapes

8.1 Decision tree analysis of *Drosophila* gut microbiome co-culture dataset

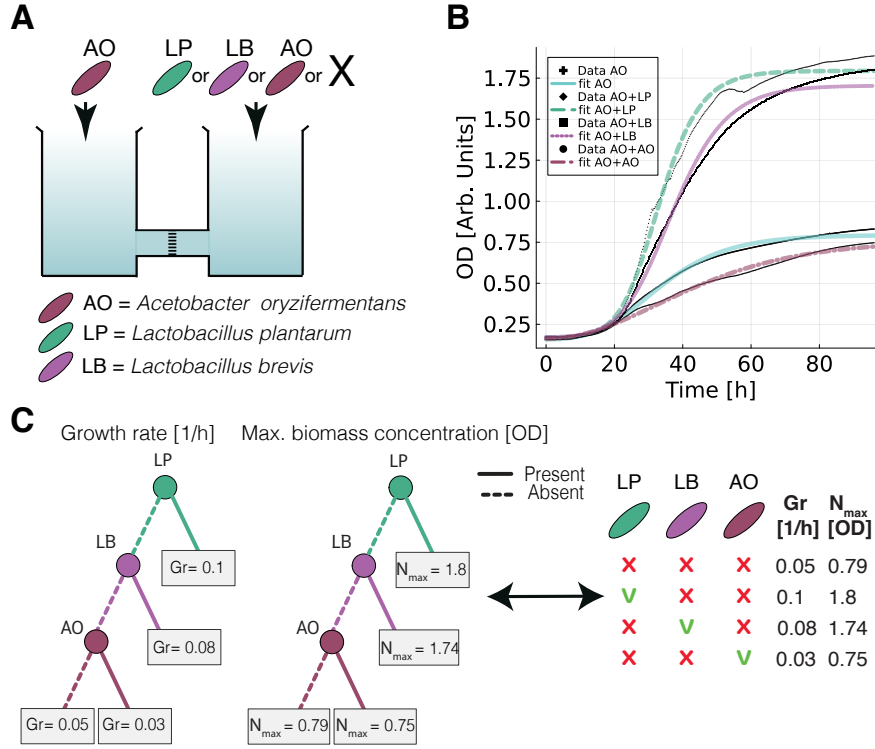
We analyzed the co-culture dataset from Ref. [14], where isolates from the *Drosophila* gut microbiome were cultured in minimal media using a BioMe plate. In this setup, pairs of wells were separated by porous membranes, permitting nutrient exchange between adjacent wells while preventing direct contact between species.

In particular, we focused on the scenario where *Acetobacter oryzafermentans* was cultured in one well, paired either with the same strain (self-paired), minimal media alone, *Lactobacillus plantarum*, or *Lactobacillus brevis* in the adjacent well (see Supplementary Figure 11A). After averaging replicates of optical density (OD) measurements for *Acetobacter oryzafermentans*, we fitted the resulting growth curves using the aHPM model (see Supplementary Figure 11B and Section 1 of Supplementary Information for the mathematical description of the model).

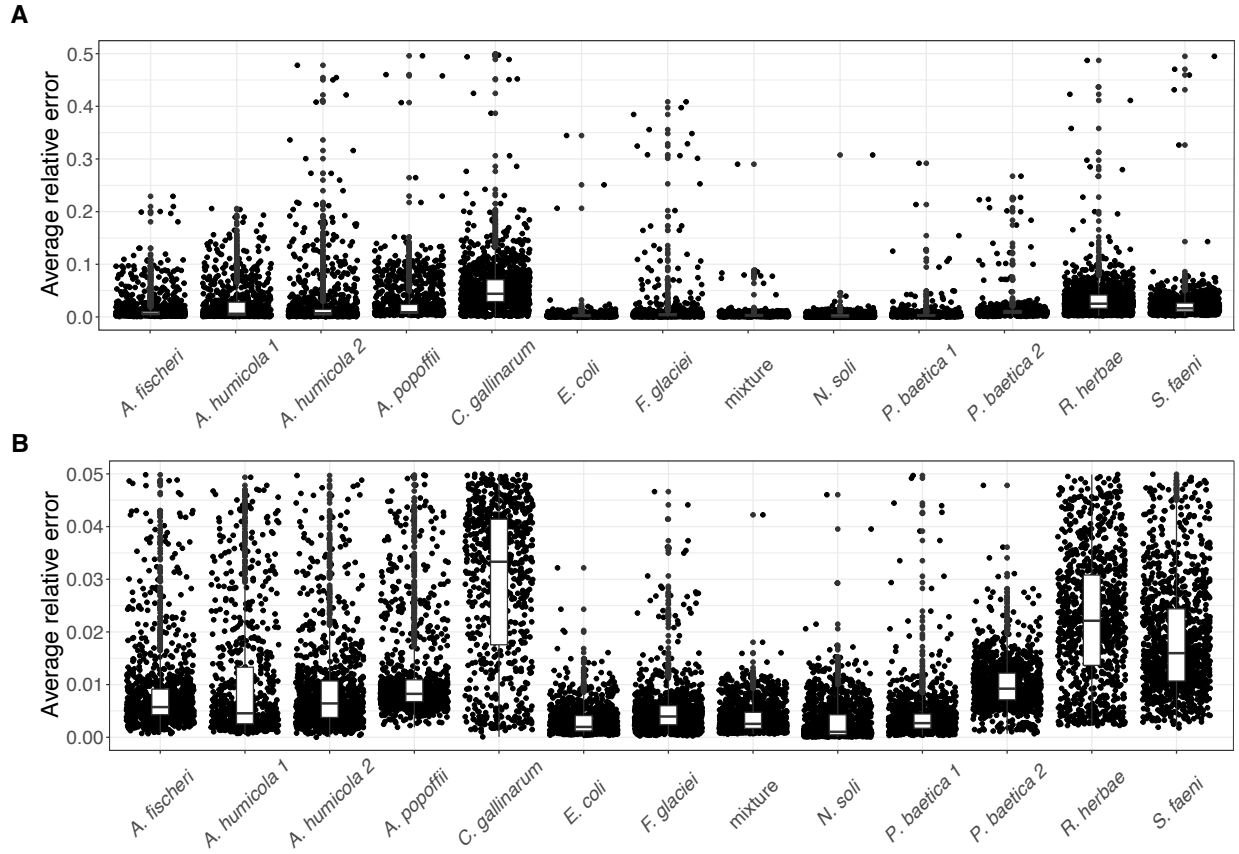
We then performed decision tree regression analyses to explore how the presence or absence of another bacterial strain affected two growth parameters – growth rate and maximum biomass ($N_{\max}$). The resulting graphical decision tree models (Supplementary Figure 11C) clearly indicated how each growth parameter varied according to the adjacent well condition. For instance, the decision tree revealed a synergistic interaction between *Acetobacter oryzafermentans* and the *Lactobacillus* strains, consistent with previous metabolic interaction studies [15].



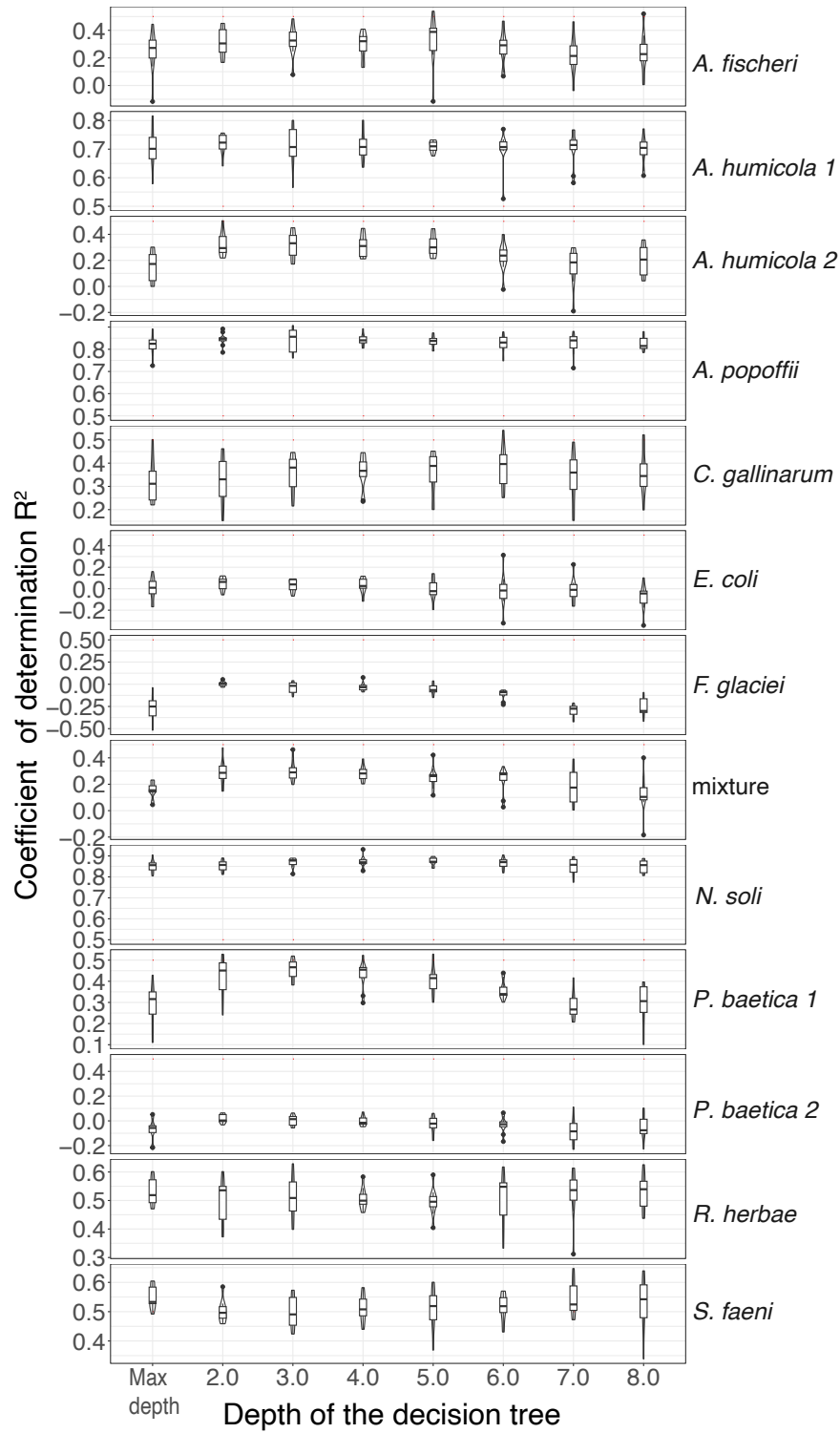
Supplementary Figure 10: **AUC measure vs. model-based fit parameters.** AUC vs. **A** the maximum biomass concentration, **B** the growth rate, and **C** the lag time inferred with the Richards model (Eq. S4) for all growth curves of Ref. [16]. These analyses suggest that while AUC may serve as a proxy for maximum biomass concentration, it should not be considered a reliable summary variable for growth, as the same AUC value can correspond to significantly different growth rates and lag times. AUC values used in these analyses were taken from Ref. [16].



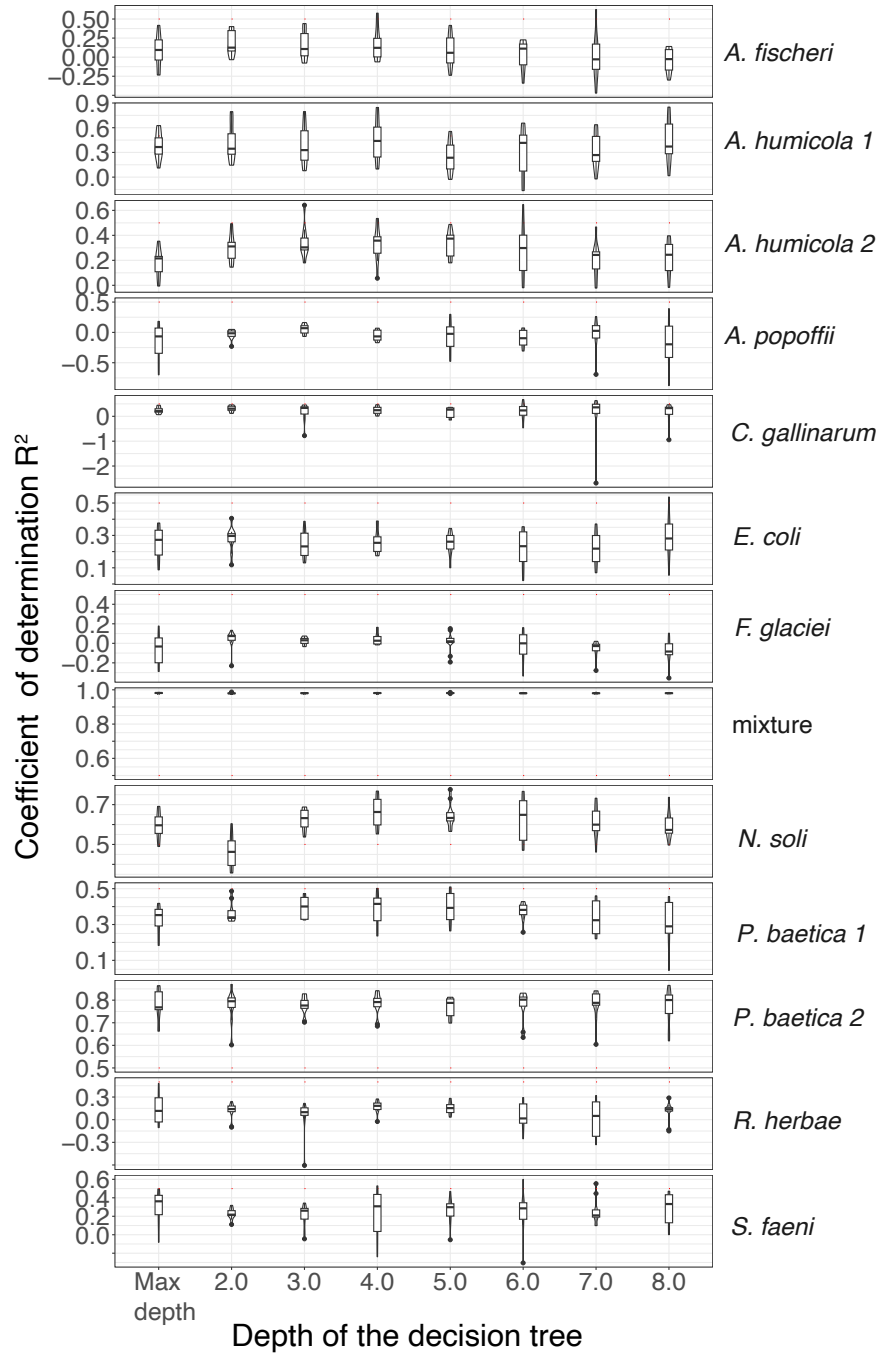
Supplementary Figure 11: **Illustrative decision tree analysis of bacterial co-culture experiments.** To illustrate the rationale and practical utility of decision tree analyses in revealing interpretable patterns from microbial growth data, we re-analyze an experiment involving bacterial isolates from the *Drosophila* gut microbiome (Ref. [14]). **A** Schematic of the experimental setup adapted from Ref. [14]. *Acetobacter oryzafermentans* (AO) is inoculated in one well, paired with either the same strain (AO), medium alone (X), *Lactobacillus plantarum* (LP), or *Lactobacillus brevis* (LB) in the adjacent well. Wells are separated by porous membranes that permit nutrient exchange but prevent direct microbial contact. **B** Averaged optical density (OD) measurements over time for *Acetobacter oryzafermentans* under the four experimental conditions. Growth curves were fitted using the adjusted Heterogeneous Population Model (Eq. S26). **C** Decision tree regression analysis of growth rate (Gr) and maximum OD ($N_{\max}$), using the presence or absence of neighboring bacterial strains as predictive features. The resulting graphical decision tree highlights condition-specific variations in growth parameters, clearly indicating a previously documented improvement in *Acetobacter oryzafermentans* growth when co-cultured with *Lactobacillus* strains [15].



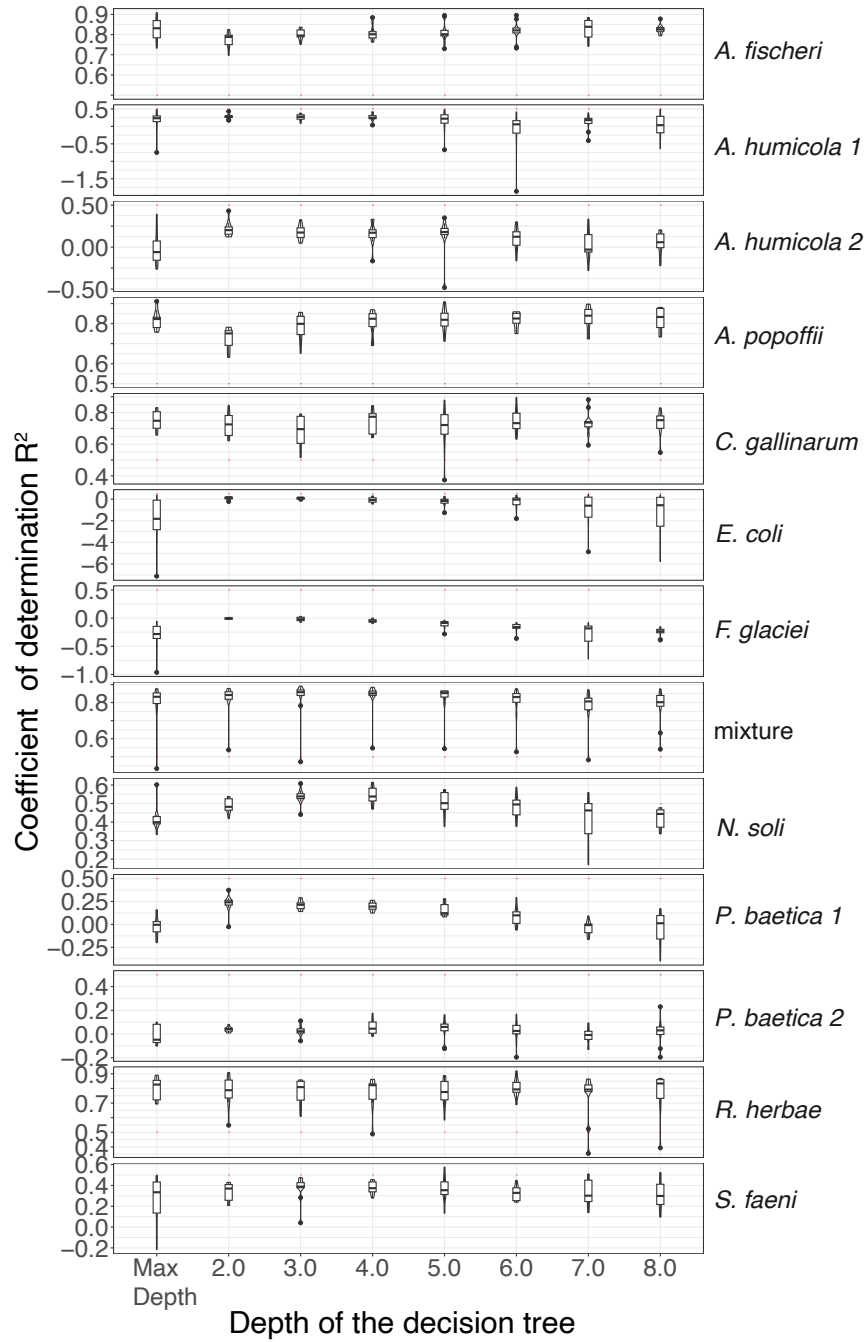
Supplementary Figure 12: **Average relative error for the fits of the dataset of Ref. [16].** **A** Distributions of the average relative error by strain (number of samples: 15120). **B** Distributions of the average relative error by strain after excluding misfits with an average relative error greater than 5% (number of samples: 13753). The fit results from these remaining samples were used in the analyses presented in the main text. Here, the boxplots represent the 0.25, 0.50, and 0.75 quantiles of the distributions.



Supplementary Figure 13: **Decision tree cross validation for the growth rate.** Box and violin plots show the coefficients of determination R^2 for the 10-fold cross validation (10 runs) of decision tree applied to the growth rates inferred using the Richards model (Eq. S4) for the growth curves of all strains in the ecotoxicological dataset analyzed from Ref. [16]. Here, each distribution contains 10 samples, and the boxplots represent the 0.25, 0.50, and 0.75 quantiles of the distributions.



Supplementary Figure 14: **Decision tree cross validation for the maximum biomass concentration.** Box and violin plots show the coefficients of determination R^2 for the 10-fold cross validation (10 runs) of decision tree applied to the maximum biomass concentration inferred using the Richards model (Eq. S4) for the growth curves of all strains in the ecotoxicological dataset analyzed from Ref. [16]. Here, each distribution contains 10 samples, and the boxplots represent the 0.25, 0.50, and 0.75 quantiles of the distributions.



Supplementary Figure 15: **Decision tree cross validation for the lag time.** Box and violin plots show the coefficients of determination R^2 for the 10-fold cross validation (10 runs) of decision tree applied to the lag time inferred using the Richards model (Eq. S4) for the growth curves of all strains in the ecotoxicological dataset analyzed from Ref. [16]. Here, each distribution contains 10 samples, and the boxplots represent the 0.25, 0.50, and 0.75 quantiles of the distributions.

References

1. Turner Jr, M. E., Bradley Jr, E. L., Kirk, K. A. & Pruitt, K. M. A theory of growth. *Mathematical Biosciences* **29**, 367–373 (1976).
2. Baranyi, J., Roberts, T. & McClure, P. A non-autonomous differential equation to model bacterial growth. *Food microbiology* **10**, 43–59 (1993).
3. Baranyi, J. & Roberts, T. A. A dynamic approach to predicting bacterial growth in food. *International journal of food microbiology* **23**, 277–294 (1994).
4. McKellar, R. A heterogeneous population model for the analysis of bacterial growth kinetics. *International Journal of Food Microbiology* **36**, 179–186 (1997).
5. Panikov, N. S. Genome-scale reconstruction of microbial dynamic phenotype: successes and challenges. *Microorganisms* **9**, 2352 (2021).
6. Loman, T. *et al.* Catalyst: Fast and flexible modeling of reaction networks. *bioRxiv*. <https://europepmc.org/article/PPR/PPR527106> (2022).
7. López, S., Prieto, M., Dijkstra, J., Dhanoa, M. S. & France, J. Statistical evaluation of mathematical models for microbial growth. *International journal of food microbiology* **96**, 289–300 (2004).
8. Angaroni, F., Peruzzi, A., Alvarenga, E. Z. & Pinheiro, F. *Translating microbial kinetics into quantitative responses and testable hypotheses using Kinbiont* <https://doi.org/10.5281/zenodo.15488336>. 2025.
9. Perni, S., Andrew, P. W. & Shama, G. Estimating the maximum growth rate from microbial growth curves: definition is everything. *Food microbiology* **22**, 491–495 (2005).
10. Sugiyama, M. *et al.* Density-difference estimation. *Neural Computation* **25**, 2734–2775 (2013).
11. Salazar, Y. *et al.* Mechanistic modelling of biomass growth, glucose consumption and ethanol production by *Kluyveromyces marxianus* in batch fermentation. *Entropy* **25**, 497 (2023).
12. Bloxham, B., Lee, H. & Gore, J. Diauxic lags explain unexpected coexistence in multi-resource environments. *Molecular Systems Biology* **18**, e10630 (2022).
13. Stevenson, K., McVey, A. F., Clark, I. B., Swain, P. S. & Pilizota, T. General calibration of microbial growth in microplate readers. *Scientific reports* **6**, 38828 (2016).
14. Jo, C., Bernstein, D. B., Vaisman, N., Frydman, H. M. & Segrè, D. Construction and modeling of a coculture microplate for real-time measurement of microbial interactions. *Msystems* **8**, e00017–21 (2023).
15. Consuegra, J. *et al.* Metabolic cooperation among commensal bacteria supports *Drosophila* juvenile growth under nutritional stress. *Iscience* **23** (2020).
16. Smith, T. P. *et al.* High-throughput characterization of bacterial responses to complex mixtures of chemical pollutants. *Nature Microbiology*, 1–11 (2024).