

UNIVERSITY OF BIRMINGHAM

Research at Birmingham

Advancing microbial sciences by individual-based modelling

Hellweger, Ferdi L.; Clegg, Robert; Clark, James R.; Plugge, Caroline M.; Kreft, Jan-Ulrich

DOI:

[10.1038/nrmicro.2016.62](https://doi.org/10.1038/nrmicro.2016.62)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Hellweger, FL, Clegg, RJ, Clark, JR, Plugge, CM & Kreft, J-U 2016, 'Advancing microbial sciences by individual-based modelling', *Nature Reviews Microbiology*. <https://doi.org/10.1038/nrmicro.2016.62>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1 **Advancing microbial sciences by individual-based modelling**

2 *Ferdi L. Hellweger¹, Robert J. Clegg², James Clark³, Caroline M. Plugge⁴, Jan-Ulrich Kreft²*

3 ¹*Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston,*
4 *MA 02115, USA.*

5 ²*Centre for Systems Biology & Institute of Microbiology and Infection, School of Biosciences, University of Birming-*
6 *ham, Birmingham, B15 2TT, UK.*

7 ³*Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK.*

8 ⁴*Laboratory of Microbiology, Wageningen University, 6703 HB Wageningen, The Netherlands.*

9
10 *Correspondence to J.U.K. e-mail: j.kreft@bham.ac.uk*

11

12 Abstract | Remarkable technological advances uncover ever more properties and behaviors of individual microor-
13 ganisms, but the novel data generated are not yet fully exploited. We explain how individual-based models
14 (IBMs), built on the findings of such techniques, help explore competitive and cooperative interactions hidden in
15 the data. Insights into self-organized spatial patterns from biofilms to the world's oceans, into phage-CRISPR dy-
16 namics, and into other emergent phenomena, are rewards already gained through this approach. Thus, combin-
17 ing individual-based observations with individual-based modeling can advance our understanding on both the
18 individual and population levels, leading to the new approach of microbial individual-based ecology (μ IBE). We
19 argue that the wider deployment of μ IBE has the potential to generate mechanistic understanding and models of
20 unprecedented predictive power.

21 Recent technological advances¹⁻¹², e.g., in microscopy, flow cytometry, microfluidics, spectroscopy, isotope and
22 molecular probes, have brought us much closer to the holy grail of microbial ecology: observing and understand-
23 ing who does what, when, where, and next to whom. We need no longer envy plant and animal ecologists who
24 have studied individual organisms for over a century. In fact, we are in a better situation now as it is much easier
25 to manipulate the environment of microbes in the laboratory and mix species together into synthetic communi-
26 ties of defined composition. What is more, the rich data from these technologies facilitate a mechanistic descrip-
27 tion of the behaviour of microbial individuals not yet feasible for larger organisms.

28 Complementing the experimental approach with mathematical modelling has, in all areas of science, pro-
29 vided valuable insights that would be difficult to obtain through experimentation alone¹³. A model consolidates
30 our knowledge of the system gathered from a variety of experiments, tests the consequences of our assumptions,
31 and exposes gaps and inconsistencies in our knowledge and understanding. Once validated, a mathematical mod-
32 el can be used to make predictions; for example, of system dynamics under conditions not yet investigated in the
33 laboratory or field, or of system properties that may be difficult to observe directly.

34 Traditionally, models of microbial systems have been constructed at the population level (FIG. 1). Popula-
35 tion-level models (PLMs) are typically “strategic models made to be as simple as possible to reveal general expla-
36 nations”¹⁴, and have proven to be of immense value, both in microbial ecology and ecology in general. PLMs are
37 good choices when the goal is to find general principles that apply across a broad range of organisms, such as the
38 tendency of predator-prey systems to generate oscillatory population dynamics; or as a first step to studying a
39 particular, complex system in detail. For an environment that is assumed to be spatially homogenous, PLMs can
40 be formulated in terms of difference equations if time is treated as being discrete or in terms of ordinary differen-
41 tial equations (ODEs) if time is treated as being continuous (FIG. 1). When considering spatially structured envi-
42 ronments, it is traditional to model the time-evolution of population densities (biomass per unit space) using par-
43 tial differential equations (PDEs) (FIG. 1). For an executive summary of different modelling approaches see Sup-
44plementary Information S1 (text).

45 Spatially explicit PLMs that simulate the temporal evolution of density distributions, e.g., in biofilms¹⁵, are
46 a valuable resource. However, they make several assumptions that are increasingly at odds with our growing
47 knowledge of microbial systems. Firstly, PLMs ignore the ever more apparent phenotypic heterogeneity between
48 individuals within a population, and the role these differences play in determining system level properties (e.g.,
49 population growth rates). At the same time, they can make little direct use of the information contained within
50 data describing the state and behaviour of individual microbes. Secondly, PLMs do not resolve the broad range of
51 interactions between individual organisms and their local biotic or abiotic environment. Thirdly, PLMs do not re-
52 solve adaptive processes at the level of individuals. Thus, while PLMs can simulate the dynamics of a system (e.g.,
53 changes in the spatial distribution of a microbial population), it is impossible to trace such changes back to the
54 behaviour of individual organisms.

55 An alternative to PLMs are individual-based models (IBMs) (FIG. 1), which can potentially overcome all of
56 these limitations^{16,17}. The defining characteristic of IBMs is that they model the properties, activities and interac-
57 tions of each individual within a population^{14,18-20}. Properties may include the biomass, size or physiological state

58 of the individual; activities may include the uptake of substrates from the environment, or the synthesis of new
59 biomass; and interactions may include competitive, synergistic or parasitic interactions between individuals with-
60 in a population or community²¹⁻²⁴. Such processes may be described as continuous and equation-based (e.g.,
61 growth) or discrete and rule-based (e.g., division). In IBMs, the collective action of each individual determines
62 population or community level properties. Feedbacks between the behaviour of individuals and the population as
63 a whole emerge automatically; as does fitness, since it depends on what other individuals do, and how this
64 changes the environment. Furthermore, IBMs can make direct use of single-cell data during their construction,
65 and of both individual and population level data during validation. An IBM therefore mimics the natural system it
66 models (FIG. 1). However, care must be taken to avoid the model becoming too complex to be useful. Precisely
67 because various approaches have their advantages and disadvantages (for a deeper discussion see REF¹⁴), it is
68 particularly beneficial to use them in conjunction²⁵⁻²⁷ (FIG. 1). For example, by comparing ODEs to PDEs of the
69 same system, the effects of local interactions in a spatially structured environment can be revealed. Likewise,
70 comparing PDEs to IBMs reveals the effects of individuality and adaptive behaviour.

71 In parallel to the technological advances for single-cell observations mentioned at the outset, the tech-
72 nology for developing, running and analysing individual-based models has also progressed significantly, making it
73 easier for scientists to use IBMs to help decipher and understand patterns within experimental data²⁵. For most
74 potential users, generic open-source platforms for individual-based modelling will be the best choice. A generic
75 platform is a software tool that allows the user to create models of a range of systems. This is done by selecting
76 the physical processes (e.g., diffusion, convection and mechanical interactions) together with the environment
77 (e.g., liquid culture, agar plate, biofilm flow cell or gut) and the set of species and their biological processes (e.g.,
78 growth, cell-cell communication or motility). Being able to select from a range of processes, one can readily iden-
79 tify those processes that affect the behaviour under investigation. Current platforms are approaching this goal,
80 and their further community-based development is the most effective way to implement an ever-wider range of
81 processes. However, specific applications may be better served by software that is more specialized. For recom-
82 mendations of software, see Supplementary Information S2 (box) and S3 (table).

83 In this Opinion article, we argue that individual-based models complement individual-based observations
84 perfectly. Joining these new developments into the approach of microbial individual-based ecology (μ IBE) will
85 become central for advancing the microbial sciences, as it makes the data from these new technologies available
86 for modelling. Here we discuss several examples in which individual-based modelling has advanced our under-
87 standing of interactions between microbes and their environment, e.g., the emergence of spatial structure and
88 feedbacks between individual and population behaviour. These advances would not have been possible without
89 IBMs, due to the complexity of the systems. While such insight is rightly valued, the ability of IBMs to make pre-
90 dictions deserves similar status: it is essential for rational engineering and management of microbial ecosystems
91 and proper testing of our models. Physicists and engineers for example have long used models to make predic-
92 tions, nicely illustrated by the prediction and later experimental verification of the existence of the Higgs boson.
93 Based on this discussion, we propose that μ IBE has the potential to revolutionize the microbial sciences.

94

95 **Characteristics of individual-based modelling**

96 The IBM approach has advantages over PLMs when simulating (i) individual heterogeneity (ii), local interactions
97 and (iii) adaptive behaviour^{14,18–20} (BOX 1), features increasingly recognized as important in microbial ecology.
98 Mounting evidence from flow cytometry²⁸ and single cell observations^{1,5,6,29,30} demonstrates the existence of indi-
99 vidual heterogeneity, even in clonal laboratory cultures^{12,31–33}. Local interactions are important because most
100 ecosystems are spatially structured and individuals only interact with neighbours. For example, even in well-
101 mixed marine or fresh water environments, mixing at the microbial scale is limited enough for hot spots of nutri-
102 ents excreted by phytoplankton to persist long enough to attract and nourish chemotactic bacteria^{34,35}. Adaptive
103 behaviour is prevalent in microbes since practically everything they do is in response to their environment, e.g.,
104 their genomes typically contain between 50 and 200 genes for two-component systems for sensing and respond-
105 ing to conditions³⁶. While rarely considered by modellers, adaptive behaviour is common in nature, of great fit-
106 ness benefit, and easy to implement in IBMs^{14,18–20}. IBMs are flexible, enabling behaviour to be specified in a vari-
107 ety of ways. In the simplest cases, rules like “if oxygen concentration below threshold, switch from aerobic to
108 anaerobic metabolism” can be implemented; such rules can be made stochastic. Kinetic equations with some
109 oxygen inhibition term for the rates of aerobic and anaerobic metabolism would lead to a smoother transition.
110 Incorporating a gene regulatory network submodel for oxygen sensing would replace the phenomenological with
111 a mechanistic description, see FIG. 1. Individual differences, local interactions, and adaptive behaviour may in-
112 teract in ways that are difficult to foresee without using an IBM to include and exclude these effects systematical-
113 ly (BOX 1).

114
115 **Microbial individuality and its consequences.** Phenotypic differences between individuals have consequences
116 for both the population and the ecosystem. Many factors contribute to phenotypic heterogeneity; these include
117 stochastic gene expression³¹, stochastic metabolism and growth³⁷, epigenetically-regulated modifications such as
118 phase variation³¹, phase in the cell cycle or biological clock³¹, asymmetric cell division³⁸, and differences in the
119 environment or stochastic sensing of the environment³⁹. Finally yet importantly, the interactions between the
120 above factors in a particular chronological order can affect the current state of an individual; this is simulated in
121 an IBM tracking changes of the local environment and cellular state. For example, some cells may have chanced
122 upon a nutrient-rich patch in the past and therefore downregulated their high-affinity transporters, later making
123 them less acclimated to nutrient-poor environments.

124 An important and surprising consequence of individual variation is that population averages can be mis-
125 leading if the functional relationship between an explanatory variable, e.g., substrate concentration, and a re-
126 sponse variable, e.g., specific growth rate, is non-linear. Non-linearities are prevalent in biology, e.g. they are
127 found in Monod kinetics, Droop kinetics and most other observed relationships between ‘dose’ and ‘response’.
128 Thus, awareness of the averaging fallacy is important¹⁷. FIG. 2a shows an example of Droop kinetics, here the non-
129 linear increase of specific growth rate of the marine cyanobacterium *Synechococcus* WH8103 with its intracellular
130 phosphorus content; this is typical for phototrophic microbes⁴⁰. As can be seen from the figure, the growth rate
131 that a population of cells with average P content would have is higher than the average growth rate of individual

132 cells with their various P contents. An IBM sums the population growth rate from the growth rates of all individu-
133 als, calculated from their measured P content. Hence, IBMs are able to predict the effects of individual variation,
134 local conditions, and adaptive behaviour on the population and ecosystem level, as well as any feedbacks the
135 changes in the ecosystem may have on individuals, taking care of any non-linearities in organisms' responses to
136 the environment.

137 IBMs are also ideal for incorporating rare events, like mutations or phenotypic switches, since they sud-
138 denly change the behaviour of one particular individual (BOX 1). For example, *Pseudomonas aeruginosa* cells in
139 laboratory flow-cell biofilms form 'mushroom' structures under specific conditions because a subpopulation of
140 immotile cells forms mushroom 'stalks' through growth and division, while the motile subpopulation forms 'caps'
141 on top of these stalks^{41,42}. These subpopulations are reminiscent of castes in social insects. Contrary to expecta-
142 tions, an IBM based on these experimental observations showed that surface-bound twitching motility cannot
143 explain the formation of mushroom caps⁴³; it was later shown experimentally that cap formation requires flagella-
144 driven swarming and chemotaxis⁴⁴.

145 Much of the large phenotypic heterogeneity observed in nature is likely due to environmental differ-
146 ences: the expression of most genes in *Escherichia coli* does not show any bursts⁴⁵, selection tends to minimize
147 noise in gene expression for most functions^{46,47}, and individual cells in a microfluidic device maintaining a strictly
148 constant environment showed reduced variation of specific growth rates³⁰. Nevertheless, non-heritable pheno-
149 typic differences that are intrinsically generated, and therefore independent of environmental conditions, are
150 important as the basis for bet-hedging and division-of-labour strategies^{32,48,49}. A division-of-labour example is the
151 segregation of the population into motile and immotile cells discussed earlier. A typical example of bet-hedging
152 strategies is the differentiation of some cells into non-growing resting stages. Less fit under benign conditions,
153 these have increased chances of survival under stress. For example, some cells of the filamentous cyanobacterium
154 *Anabaena flos-aquae* differentiate into akinetes, which sink to the sediment bed to germinate and re-emerge
155 under favourable conditions. It is clear that these dormant cells help the population survive adverse conditions,
156 but is this trait required to survive annual challenges or more irregular and extreme events? Using an IBM, a het-
157 erogeneous population of cells was simulated in a reservoir model that also tracked environmental conditions
158 (light, temperature, nutrients)⁵⁰. When the akinete differentiation trait was "knocked out" in the model, the
159 knockout population did not survive the first winter. As an unexpected insight from the simulations, akinetes pro-
160 vided an additional benefit by taking up nutrients while in the sediment bed⁵⁰. Although experiments with mu-
161 tants are straightforward in the laboratory, we cannot introduce genetically engineered mutants into the field for
162 ethical reasons so will have to rely on modelling approaches.

164 **Predicting complex spatial patterns**

165 **Predicting unobserved gradients:** FIG. 2b shows that PDEs or IBMs can predict the hard-to-measure concentra-
166 tion gradients of substrates forming in biofilms. Such a prediction requires three ingredients: (i) spatial distribu-
167 tion of biomass (for PDEs) or cells (for IBMs) acquired from a confocal image of the biofilm, (ii) laws of diffusion,
168 and (iii) measured substrate uptake kinetics of the cells. Wherever possible, these predictions should be validated

169 with microelectrodes; see the example of nitrifying granules below. Such validated models can then be used to
170 visualize the unobserved concentration gradients throughout the imaged region and in other, comparable envi-
171 ronments where it is unfeasible to take direct measurements.

172
173 **Predicting complex spatial patterns.** A microbe, whether growing in a well-mixed liquid or a spatially structured
174 biofilm, has the same genome and therefore the same potential to sense and react to the environment. Indeed,
175 microbes respond to nutrient-poor conditions within biofilms in much the same way as they would to similar con-
176 ditions in a chemostat^{51,52} or stationary-phase batch culture^{53,54}. It should therefore be possible to predict the
177 behaviour of a microbe anywhere if we fully understand how it influences, and is influenced by, its environment⁵⁵.
178 FIG. 2c explains the approach of using a chemostat to measure the dependence of population growth rate on
179 substrate concentration. PDEs or IBMs including such growth kinetics can then calculate the consumption and
180 diffusion of substrate in order to predict the resulting substrate concentration and growth rate gradients in a bio-
181 film. This then tracks how, over time, the changing growth rates of the cells give rise to the emerging spatial struc-
182 ture of the biofilm. This approach has been surprisingly successful given the simplifications involved; see the ex-
183 ample of nitrifying granules below. However, if such predictions fail, we can conclude that the simplifications are
184 inappropriate or that other factors may play a role, e.g., individual variation in kinetics of growing cells³³, the
185 presence of persisters^{56,57} or changes in gene expression upon attachment^{58,59} that affect growth kinetics⁶⁰ or
186 induce virulence⁵⁹. Refining the model by including such effects may then better match experimental results. Such
187 refinements would be straightforward to implement in IBMs but difficult to include in PDEs.

188 An example where feedbacks between substrate concentration gradients, growth rates and biofilm struc-
189 ture can lead to spontaneous formation of clusters from initially small stochastic perturbations is shown in Fig. 2d.
190 Once spontaneously arisen, clusters of slowly growing cells that utilize resources more economically grow faster
191 than clusters of cells that grow faster at any given substrate concentration⁶¹. This counterintuitive result is due to
192 the locally higher substrate concentration in the clusters of economical cells. Thus, complex spatial patterns can
193 emerge from stochastic positioning of cells.

194
195 **Predicting interactions in mixed cultures.** Microbial species are often studied in isolation, yet in their natural
196 environment they interact with many other species in a variety of ways; most interactions are indirect, mediated
197 by diffusible compounds such as metabolites, autoinducers or toxins. Using modelling to predict interactions is
198 extremely useful, as the number of potential interactions increases exponentially with the number of species in a
199 habitat, e.g., at 5 or 6 species there are already 31 or 63 potential interactions¹⁷, making it difficult to study all
200 possible interactions experimentally.

201 In mixed species biofilms, additional mechanisms and phenomena of pattern formation come into play.
202 Strains of *Saccharomyces cerevisiae* can be engineered to depend on each other for growth by producing a me-
203 tabolite that the other requires⁶². Engineering another strain requiring one metabolite without reciprocating adds
204 a cheater to the cooperating pair of obligate cross-feeding strains. When randomly placed cells start to grow into
205 colonies, mutually cross-feeding colonies that happen to be close by will grow well towards each other, forming

206 large areas of contact. In contrast, the cheater strain becomes squeezed out, as it does not facilitate the growth
207 of the strain that it depends on (FIG. 3a)⁶². The general insight from this and other IBMs is that spatial patterns in
208 microbial communities are to some extent self-organized as they emerge from different types of metabolic inter-
209 actions: cooperation consisting in restraint from competition⁶¹, cross-feeding^{63,64}, interspecies hydrogen trans-
210 fer^{65,66} and the combination of particular trophic interactions with motility of cells on hydrated rough surfaces⁶⁷.
211 The most rigorous demonstration to date of the ability of IBMs to predict solute gradients and spatial distribu-
212 tions of interacting metabolic/functional groups from kinetics measured in batch and chemostat cultures has
213 been carried out for nitrifying biofilms⁶⁸ and granules⁶⁹. These are assemblages of a few types of autotrophic am-
214 monia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB), plus a few types of heterotrophic denitrifiers,
215 forming a food-web that is utilized in wastewater treatment to convert ammonia to dinitrogen. Matsumoto and
216 colleagues⁶⁹ combined microelectrodes to measure O₂, NH₄⁺, NO₂⁻ and NO₃⁻ profiles and confocal microscopy to
217 determine the distribution of functional groups through these nitrifying granules (FIG. 3b). Considering that the
218 model was not fitted to the data makes the match of microelectrode-measured with IBM-calculated profiles sur-
219 prisingly close; the match of observed biomass distributions with predictions was quite good, especially for the
220 better-known autotrophic groups (FIG. 3b). This provided insight into the self-organization of radial stratification
221 in these granules: oxygen becomes depleted with depth and the aerobic AOB dominate the surface layer because
222 they are first in the food chain, while the also aerobic NOB reside underneath as they are dependent on the AOB's
223 activity and therefore less competitive for oxygen. Reliable predictions are also of great value, as they can be used
224 by engineers to estimate and optimize whole reactor performance.

225

226 **Predicting evolution**

227 IBMs, being based on specifying traits of individuals, can easily incorporate heritable mutations of these traits.
228 Combined with simulating birth, death and competition, natural selection of the mutants is simply an emergent
229 process in IBMs. FIG. 4a shows such an example where the phenotypic traits are digitally encoded, so a change in
230 the digital genome is mapped to a change in phenotype, e.g., a change in the rate of a particular pathway⁷⁰. In
231 spatially structured environments, migration of the evolving microbes can generate spatial patterns and feed-
232 backs between the emerging spatial structure and natural selection can arise (FIG. 4a). This technique has been
233 applied to studies of cyanobacteria dynamics in the ocean⁷¹ where a complex set of interacting biotic and abiotic
234 forces shape the physiology, ecology, and evolutionary dynamics of these microorganisms. In this work, an evo-
235 lutionary IBM was coupled to a hydrodynamic model that resolves vertical gradients in light intensity, tempera-
236 ture, and the amount of dissolved inorganic and organic nutrients, and how these change in time, e.g., due to
237 changes in surface wind speeds, irradiance, or the uptake of nutrients by the cyanobacteria. The model predicted
238 spatial and temporal trends in the physiology, size and abundance of *Synechococcus* and *Prochlorococcus* eco-
239 types. During the summer months when the water column is well stratified, small, high-light adapted cyanobacte-
240 ria dominated in well-lit but nutrient-starved surface waters, and larger, low-light adapted cells dominated at
241 depth (Fig. 4b). These predicted trends were then found to be consistent with observations at the seasonally
242 stratified Bermuda Atlantic Time Series site in the North Atlantic Ocean⁷¹. The IBM helped to identify a three-way

243 trade-off between cyanobacteria cell size, light/nutrient affinity, and growth rate that can explain the observed
244 trends. In the IBM, these different strategies emerged as a result of natural selection – i.e., they were not im-
245 posed. Thus, the extent of microbial diversity within an evolutionary IBM is an emergent property. Further, IBMs
246 can readily simulate both acclimation strategies and adaptive processes, or resolve variations experienced over
247 single division cycles, e.g., in light received by a cell as it is mixed up and down the water column.

248 Another IBM was used to test the hypothesis that dispersal limitation of ocean bacteria is sufficient for
249 the formation of biogeographical patterns⁷². This IBM simulated ~100,000 individuals within a global ocean circu-
250 lation model. The cells grew and divided and their 1 Mbp genomes were subject to neutral evolution, i.e., the
251 mutations were assumed to have no fitness effect. The model showed that biogeographic provinces dominated
252 by different species could be produced from ocean currents and dispersal limitation alone, without any environ-
253 mental selection⁷² (FIG. 4c). IBMs can simulate discrete individuals with their own genome sequence and account
254 for dispersal limitation.

255 IBMs are naturally also well suited to study co-evolution, and have been used to shed light on the co-
256 evolution of host immunity and phage. Given that immunity against phage infection should be of tremendous
257 advantage in a world where bacteria are outnumbered by phages tenfold, it is surprising that less than half of the
258 sequenced prokaryotic genomes from mesophiles contain an adaptive immunity system known as CRISPR (Clus-
259 tered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated genes)⁷³. In contrast, almost all
260 genomes of hyperthermophiles (mostly archaea) code for CRISPR-Cas⁷³. Koonin and coworkers developed an
261 IBM that predicts loss of efficacy for CRISPR at larger population sizes, which are presumably not reached by hy-
262 perthermophiles in their environment, providing a plausible explanation for this puzzling observation²⁷. Their
263 IBM enables host and phage co-evolution by including density dependent encounter of lytic phages and hosts,
264 innate immunity independent of CRISPR (e.g., due to receptor mutation or restriction-modification systems), loss
265 of entire CRISPR cassettes upon cell division and re-acquisition by HGT, loss and addition of single spacers, and
266 mutation of viral proto-spacers (FIG. 4d)²⁷. Counterintuitively, CRISPR immunity is good for phages: the increased
267 host population sustains an increased phage population and concomitant phage diversity (FIG. 4e). Above a phage
268 diversity threshold, CRISPR becomes ineffective and is lost due to its fitness cost²⁷. Importantly, this IBM only
269 provided these insights because (i) all known key processes were included and (ii) population sizes were not fixed,
270 as in previous models, but allowed to emerge through feedbacks between host and phage abundance, diversity
271 and the co-evolving immunity.

272

273 **Beyond cell and population scales**

274 Microbial IBMs are multiscale models by nature, i.e., they link cell and population scales; this is rather useful in
275 itself, but their multiscale nature can be further expanded to include lower and higher levels of organization. In-
276 cluding increasingly more intracellular states and behaviours leads to more mechanistic models of the behaviour
277 of individual cells replacing the empirical descriptions traditionally used. This has already been successful with
278 IBMs incorporating signal transduction mechanisms in chemotaxis⁷⁴ and quorum sensing²⁶. IBMs also advance
279 the new field of synthetic biology because they allow simulation and optimization of how synthetic organisms

280 interact with each other and the environment before actually constructing them^{26,75,76} – the ultimate rational de-
281 sign. Apart from providing stronger mechanistic foundations for individual behaviour and individuality, integra-
282 tion of intracellular mechanisms also enables exploitation of the rapidly growing metagenomic data. After recon-
283 structing genomes from shotgun metagenomics⁷⁷ one can reconstruct whole-genome metabolic models from the
284 stoichiometry of all enzyme reactions coded for by the genome. By assuming that the fluxes through reactions are
285 constrained and optimized such that growth of the cell is maximal, one can predict metabolic fluxes and growth
286 without having to know the kinetics of any enzymes^{78,79}. Such constraint-based reconstruction and analysis (CO-
287 BRA) models have already been scaled up to the population level^{33,80}. Likewise, incorporating COBRA models into
288 IBMs (FIG. 1) has great potential for the future and has already been successful using a model based on a spatial
289 grid, an approach similar to IBMs but with a coarser resolution⁶³.

290 In the other direction, microbial IBMs can be expanded to full-scale ecological or biogeochemical systems,
291 which we have already illustrated with examples (FIG. 4a-c). This will improve the predictive power of ecosystem
292 models. Moreover, subcellular dynamics could be included in full-scale ecological IBMs, which will facilitate the
293 use of omics data sets that measure the composition, acclimation state, activity or genetic make-up of individuals
294 and thus help to bridge the growing gap between omics data and biogeochemical models⁸¹.

295

296 **Conclusions and future directions**

297 Physicists and engineers have long used models to predict dynamics or optimize processes but, since life is far
298 more complex, it has taken considerably longer for models to advance to a stage where they can deal with com-
299 plex biological systems in a realistic way. For this reason, the tradition in most areas of biology has been to view
300 mathematical modelling as too unrealistic to be useful. This, however, has started to change as experimentalists
301 realize that the data being generated have become too complex to handle without models. We have explained
302 how individual-based modelling allows researchers to integrate diverse types of information gathered from stud-
303 ies of molecular mechanisms, single cell observations, community dynamics and spatial patterns, thereby making
304 best use of small and big data.

305 The key general insight from IBMs is that population dynamics and structure emerge from the interac-
306 tions of individuals with each other and the environment. This extends to the community and ecosystem level.
307 Biofilm spatial structure and other self-organized patterns are a prime example of emergence and because of this,
308 complex macroscopic patterns can be predicted from simple microscopic mechanisms. In evolutionary IBMs, di-
309 versity and spatial distribution of species can emerge; the inclusive fitness of individuals is also emergent. Since
310 IBMs link cell and population scales they demonstrate how individual heterogeneity affects population and eco-
311 system function or how new phenomena can arise at the population level, e.g., in populations of signal transduc-
312 ing cells or hosts co-evolving with phages. Moreover, linking individual and population scales can identify mecha-
313 nisms that can or cannot explain observed population behaviour. These insights could not have been obtained
314 with classical PLMs as they do not account for individual heterogeneity, local interactions, or adaptive behaviour.
315 Since IBMs do, they can advance our understanding and ability to predict microbial systems beyond what can be
316 achieved without them. PLMs are also important tools, and are useful when studying general principles in simple

317 systems and for comparing with IBMs, and ideally a variety of modelling approaches, briefly summarized in Sup-
318 plementary Information S1 (text), should be used in conjunction.

319 The key shortcoming of IBMs is the tendency to become too complex and difficult to analyse mathemati-
320 cally. Overly complex models are difficult to understand and communicate, but standardizing the description of
321 IBMs has helped (Box 1). Still, more efforts to standardize IBM descriptions would be just as beneficial as they
322 were for systems biology. If models are too complex to understand, we no longer gain insight, although the pre-
323 dictive power of such models can still be valuable. On the other hand, overly simple models cannot generate the
324 variety of dynamics and patterns the real systems are capable of under different conditions and therefore lack
325 relevance for natural systems, yet they are useful to distil principles. Due to these trade-offs, intermediate model
326 complexity is optimal¹⁹; exactly where the optimum is depends on the purpose of the model and on the data
327 available.

328 Bridging several levels of organization and time scales from molecular dynamics to evolution, multiscale
329 IBMs have the potential to become generic mechanistic models able to predict dynamics in novel conditions or
330 changing environments. Ultimately, sufficient understanding of a system can only be demonstrated when one
331 can write a complete and consistent description of the biology in the formal language of a computer program,
332 which when executed recreates observed system behaviour and generates correct predictions from correct
333 mechanisms. What is required for this revolution in microbial sciences to succeed is the tight integration of ex-
334 perimental and computational research. Community development of computational tools for IBM that enable
335 biologists to explain their system to the computer in their own, biological language will facilitate this. Communica-
336 tion between experimentalists and modellers will be crucial, requiring mutual education and building a communi-
337 ty around the goal of microbial Individual-Based Ecology (μ IBE).

338

339 **Acknowledgements**

340 We thank our colleagues Cristian Picioreanu, Joao Xavier, Barth F Smets, Volker Grimm, Thomas Banitz, Isaac
341 Klapper, Tom Curtis, Helen Kettle, Rosalind Allen, Orkun Soyer, Tobias Grosskopf and many other fellow partici-
342 pants of two workshops for many stimulating discussions: a National Institute for Mathematical and Biological
343 Synthesis (NIMBioS) workshop in June 2011 in the USA (NSF Award #EF-0832858) and the “Understanding Micro-
344 bial Communities” workshop funded by the Isaac Newton Institute (INI), in Cambridge, UK, held 08-12/2014. We
345 thank Babak Momeni and Søren Molin for their unpublished images, and Shinya Matsumoto and Cristian Pi-
346 coreanu for sharing data. We are grateful to the National Centre for the Replacement, Refinement & Reduction
347 of Animals in Research, UK (NC3Rs) for funding our development of IBMs for the gut environment (eGUT grant
348 NC/K000683/1), and to the National Science Foundation for funding development and application of IBMs for
349 phytoplankton.

350 Display items

352 Box 1 | Summary of characteristics, benefits, good practice and pitfalls of individual-based modelling

353 **Characteristics of mathematical models generally.** The purpose of a model is to simplify reality. Since “a mathe-
354 matical model is a logical machine for converting assumptions into conclusions”¹³, it enforces complete and un-
355 ambiguous specification of assumptions, which is essential for rigorous testing of hypotheses. An example of a
356 typical simplifying assumption in IBMs is that cells divide instantaneously when they reach a threshold volume;
357 for studying e.g., lag phase a more mechanistic cell division model would be more appropriate⁸²

358 **When to use IBMs specifically.** IBMs are particularly useful when (i) individual heterogeneity, (ii) local interac-
359 tions, or (iii) adaptive behaviour are potentially important. Nonlinear feedback loops between the variable activi-
360 ties of individuals and environmental resources often make IBMs more appropriate than models using population
361 averages. These models are also useful in situations where sudden, discrete events occur in the lifecycle of the
362 microorganism (e.g., attachment to a surface) or when rare phenotypic variants (e.g., bet hedging) or mutants
363 (e.g., evolutionary IBMs) arise.

364 **Benefits of IBMs.** IBMs act as a bridge between individual and population/community level behaviour, allowing
365 the consistency of assumed individual behaviour and population data to be assessed. IBMs of microbial systems
366 (e.g., biofilms) excel at reconstructing/predicting (i) solute concentration gradients, (ii) effects of spatial structure,
367 (iii) behaviour in more complex environments, and (iv) emergent interactions in complex communities.

368 **Good modelling practice in IBMs.** Adopting the ‘ODD’ (Overview, Design concepts, and Details) protocol as
369 standard for systematic, complete description of IBMs has already facilitated comparison and peer review of
370 IBMs⁸³. ODD is similar in purpose to MIRIAM⁸⁴ (Minimum Information Requested In the Annotation of biochemi-
371 cal Models). Better-developed standards for model exchange and description have been highly beneficial for sys-
372 tems biology. For example, the Systems Biology Ontology (SBO⁸⁵) provides an unambiguous vocabulary for model
373 description and the Systems Biology Markup Language (SBML⁸⁶) enables the exchange of completely and une-
374 quivocally specified models.

375 The most successful models are *structurally realistic*, i.e., the entities and processes in the model correspond to
376 those in the real world (e.g., a chemotaxis model where cells carry out runs and tumbles and responding to
377 changes in attractant concentration⁷⁴ rather than cells taking the steepest ascent towards the concentration
378 peak). Demonstrate robustness of model predictions: parameter sensitivity analysis to identify important pa-
379 rameters (e.g., an IBM of plasmid transfer evaluating how strongly the rate of plasmid transfer changes when
380 pilus reach, lag times between transfers, and other conjugation parameters are varied⁸⁷); structural sensitivity
381 analysis to identify important processes (e.g., an IBM of aging systematically including/excluding processes such
382 as segregation, repair and/or toxicity of damage and growth of the cells to test whether they interact and/or
383 change simulation results qualitatively or quantitatively²³).

384 **Potential pitfalls of IBMs.** Avoid using global environmental or population states in deciding the activities of an
385 individual, since no individual has global knowledge (e.g., assuming that growth of cells depends on population
386 density as in logistic growth⁸⁸). Ignoring processes that are in fact important. Probably the most common mistake

387 is that of imposing behaviour of individuals that one wishes to study as emergent (e.g., using a biofilm model to
388 predict biofilm structure formation but assuming that cells inside the biofilm cannot divide which will affect the
389 structure⁸⁹).

390

391 **Figure 1 | Simplified overview of approaches useful for modelling communities and/or single cells.** ODEs and
392 PDEs describe rates of change of populations (X) and/or resources (R) directly, i.e. the level of individuals is ab-
393 sent. Comparing the non-spatial ODEs with the spatially explicit PDEs illuminates the effect of spatial structure.
394 IBMs describe activities of individuals. Changes on the population level are not directly described because they
395 emerge from individual behaviour. Hence, IBMs can make use of data on both levels: individual-level data as input
396 and population-level data to compare with simulation output. Comparing PDEs with IBMs elucidates the effect of
397 individuality and adaptive behaviour. Combining all three approaches is therefore best practice. Most IBMs to
398 date include only simple kinetic models of growth and rules for cell division, but since IBMs treat individuals as
399 discrete agents, they enable incorporation of intracellular dynamics as modelled in systems biology – bridging the
400 scales from intracellular reactions to ecosystem function. Only two major types of models for intracellular dynam-
401 ics are shown: dynamic kinetic models use full kinetic equations only known for a select number of enzyme reac-
402 tions while flux balance models only need the stoichiometry of the reactions and constraints, which enables a
403 genome-wide prediction of metabolic fluxes in steady state⁷⁹. Image of mouse gut mucosa courtesy of Kristen
404 Earle of Justin Sonnenburg’s lab. Dynamic kinetic model modified with permission from REF. ⁹⁰.

405

406 **Figure 2 | Using IBMs to predict population dynamics, substrate and growth rate gradients and the effect of**
407 **spatial structure.** **a** | Observed cellular phosphate content (quota q) in the marine cyanobacterium *Synechococ-*
408 *cus* WH8103 and photosynthesis rate (μ) calculated using the non-linear Droop kinetics shown as a black line⁴⁰.
409 Calculating the rate for each individual based on its quota (full circles) and averaging over individuals (red line:
410 $\text{ave}[\mu(q)]$) gives a lower population productivity than averaging the quotas and calculating the rate based on that
411 average quota (blue line: $\mu(\text{ave}[q])$). **b** | Predicting unobserved concentration gradients. From a confocal biofilm
412 image (courtesy of Søren Molin) we can estimate biomass distribution for PDE or cell positions for IBM and calcu-
413 late the concentration field based on growth kinetics and laws of diffusion. This may then be verified by microe-
414 lectrode transects where possible. **c** | Growth kinetics parameterised from chemostat experiments can be used
415 in PDEs or IBMs to predict biofilm structure, growth rate and concentration gradients. IBMs could include adapta-
416 tions in kinetics or heterogeneity in the population, e.g. persisters. **d** | Predicting feedbacks between emerging
417 biofilm structure and metabolic interactions. Once clusters of red cells that consume resources economically have
418 formed by chance, they grow faster than clusters of the blue fast growing cells because their economy sustains
419 locally higher substrate concentrations⁶¹.

420

421 **Figure 3 | Inter-species interactions lead to spatial patterns that may be predicted and explained using IBMs.** **a**
422 | Members of a microbial community may be engineered to depend on one another for growth, referred to as
423 synthetic obligate cross-feeding. A cheating strain that receives secreted nutrients but does not produce any is

424 excluded by spatial self-organisation of the co-operators; this is shown both experimentally, where strains are
425 fluorescently tagged, and in an IBM of the system⁶². Images courtesy of Babak Momeni. **b** | Verifying IBM predic-
426 tions for a nitrifying food-web in a lab scale aerobic upflow fluidized bed reactor with microelectrodes and mi-
427 croscopy. Based on standard literature parameters for nitrifiers rather than fitting the model to data, an IBM can
428 predict the measured solute profiles and biomass distributions of the autotrophic AOB and NOB, and of EPS, quite
429 well (simulated profiles were averaged over concentric layers). Predicted distributions of the less understood
430 groups of heterotrophic bacteria (for this figure lumped together as 'Het') are roughly correct. Colour coding of
431 solutes and biomass indicated next to the respective graphs. Modified with permission from REF.⁶⁹.

432

433 **Figure 4 | Combining ecology and evolution is facilitated by IBMs.** **a** | Many of the processes observed in real-
434 life microbial ecology and evolution can be mapped directly to those modelled in IBMs. **b** | Emergent spatial and
435 temporal patterns in cyanobacterial biomass and cell size distribution in an evolutionary IBM based upon a gener-
436 ic, cell-based model for cyanobacteria and coupled to a hydrodynamic model of vertical transport (modified with
437 permission from REF.⁷¹). **c** | Biogeographical provinces emerge from the interaction of dispersal limitation and
438 neutral evolution of genomes in a global surface ocean IBM. Colors demarcate regions where different OTUs
439 dominate (reproduced with permission from REF.⁷²). **d** | The processes included in an IBM of phage-host co-
440 evolution where phages mutate and hosts have innate and adaptive immunity based on CRISPR-Cas. The host can
441 acquire and lose single spacers and the entire cassette. **e** | This IBM predicts that increased immune evasion by
442 mutant phage will, counter-intuitively, reduce overall phage population size and diversity despite an increased
443 number of phages per host cell as the host population declines (data from REF.²⁷)

- 445 1. Brehm-Stecher, B. F. & Johnson, E. A. Single-cell microbiology: tools, technologies, and applications. *Microbiol. Mol.*
446 *Biol. Rev.* **68**, 538–559 (2004).
- 447 2. Balagaddé, F. K., You, L., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-term monitoring of bacteria undergoing pro-
448 grammed population control in a microchemostat. *Science* **309**, 137–140 (2005).
- 449 3. Groisman, A. *et al.* A microfluidic chemostat for experiments with bacterial and yeast cells. *Nat. Methods* **2**, 685–689
450 (2005).
- 451 4. Keymer, J. E., Galajda, P., Muldoon, C., Park, S. & Austin, R. H. Bacterial metapopulations in nanofabricated landscapes.
452 *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17290–17295 (2006).
- 453 5. Wagner, M. Single-Cell Ecophysiology of Microbes as Revealed by Raman Microspectroscopy or Secondary Ion Mass
454 Spectrometry Imaging. *Annu. Rev. Microbiol.* **63**, 411–429 (2009).
- 455 6. Zengler, K. Central role of the cell in microbial ecology. *Microbiol. Mol. Biol. Rev.* **73**, 712–729 (2009).
- 456 7. Tadmor, A. D., Ottesen, E. A., Leadbetter, J. R. & Phillips, R. Probing individual environmental bacteria for viruses by
457 using microfluidic digital PCR. *Science* **333**, 58–62 (2011).
- 458 8. Musat, N., Foster, R., Vagner, T., Adam, B. & Kuypers, M. M. M. Detecting metabolic activities in single cells, with em-
459 phasis on nanoSIMS. *FEMS Microbiol. Rev.* **36**, 486–511 (2012).
- 460 9. Dénervaud, N. *et al.* A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proc. Natl. Acad.*
461 *Sci. U. S. A.* **110**, 15842–15847 (2013).
- 462 10. Wessel, A. K., Hmelo, L., Parsek, M. R. & Whiteley, M. Going local: technologies for exploring bacterial microenviron-
463 ments. *Nat. Rev. Microbiol.* **11**, 337–348 (2013).
- 464 11. Hol, F. J. H. & Dekker, C. Zooming in to see the bigger picture: Microfluidic and nanofabrication tools to study bacteria.
465 *Science* **346**, 1251821 (2014).
- 466 12. Zhang, Q. *et al.* Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science*
467 **333**, 1764–1767 (2011).
- 468 13. Gunawardena, J. Models in biology: ‘accurate descriptions of our pathetic thinking’. *BMC Biol.* **12**, 29 (2014).
- 469 14. Evans, M. R. *et al.* Do simple models lead to generality in ecology? *Trends Ecol. Evol.* **28**, 578–583 (2013).
- 470 15. Klapper, I. & Dockery, J. Mathematical description of microbial biofilms. *SIAM Rev.* **52**, 221–265 (2010).
- 471 16. Hellweger, F. L. & Bucci, V. A bunch of tiny individuals - individual-based modeling for microbes. *Ecol. Model.* **220**, 8–22
472 (2009).
- 473 17. Kreft, J.-U. *et al.* Mighty small: Observing and modeling individual microbes becomes big science. *Proc. Natl. Acad. Sci.*
474 *U. S. A.* **110**, 18027–18028 (2013).
- 475 18. DeAngelis, D. L. & Mooij, W. M. Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol.*
476 *Evol. Syst.* **36**, 147–168 (2005).
- 477 19. Grimm, V. *et al.* Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* **310**, 987–
478 991 (2005).
- 479 20. Railsback, S. F. & Grimm, V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. (Princeton University
480 Press, 2012).
- 481 21. Sher, D., Thompson, J. W., Kashtan, N., Croal, L. & Chisholm, S. W. Response of *Prochlorococcus* ecotypes to co-culture
482 with diverse marine bacteria. *ISME J.* **5**, 1125–1132 (2011).
- 483 22. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**,
484 667–684 (2012).
- 485 23. Clegg, R. J., Dyson, R. J. & Kreft, J.-U. Repair rather than segregation of damage is the optimal unicellular aging strategy.
486 *BMC Biol.* **12**, 52 (2014).
- 487 24. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*
488 **522**, 98–101 (2015).
- 489 25. Lardon, L. A. *et al.* iDynoMiCS: next-generation individual-based modelling of biofilms. *Environ. Microbiol.* **13**, 2416–
490 2434 (2011).
- 491 26. Mina, P., di Bernardo, M., Savery, N. J. & Tsaneva-Atanasova, K. Modelling emergence of oscillations in communicating
492 bacteria: a structured approach from one to many cells. *J. R. Soc. Interface* **10**, 20120612–20120612 (2013).
- 493 27. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary dynamics of the prokaryotic adaptive immunity sys-
494 tem CRISPR-Cas in an explicit ecological context. *J. Bacteriol.* **195**, 3834–3844 (2013).
- 495 28. Davey, H. M. & Kell, D. B. Flow cytometry and cell sorting of heterogeneous microbial populations: The importance of
496 single-cell analyses. *Microbiol. Rev.* **60**, 641–696 (1996).
- 497 29. Musat, N. *et al.* A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc. Natl. Acad. Sci. U. S.*
498 *A.* **105**, 17861–17866 (2008).
- 499 30. Dusny, C., Fritzsche, F. S. O., Frick, O. & Schmid, A. Isolated Microbial Single Cells and Resulting Micropopulations Grow
500 Faster in Controlled Environments. *Appl. Environ. Microbiol.* **78**, 7132–7136 (2012).
- 501 31. Avery, S. V. Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.* **4**, 577–587
502 (2006).
- 503 32. Ackermann, M. Microbial individuality in the natural environment. *ISME J.* **7**, 465–467 (2013).

- 504 33. Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D. & Luthey-Schulten, Z. A. Heterogeneity in protein expression induces
505 metabolic variability in a modeled *Escherichia coli* population. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14006–14011 (2013).
- 506 34. Mitchell, J. G., Okubo, A. & Fuhrman, J. A. Microzones Surrounding Phytoplankton Form the Basis for A Stratified Ma-
507 rine Microbial Ecosystem. *Nature* **316**, 58–59 (1985).
- 508 35. Blackburn, N., Fenchel, T. & Mitchell, J. Microscale nutrient patches in planktonic habitats shown by chemotactic bac-
509 teria. *Science* **282**, 2254–2256 (1998).
- 510 36. Capra, E. J. & Laub, M. T. Evolution of Two-Component Signal Transduction Systems. *Annu. Rev. Microbiol.* **66**, 325–347
511 (2012).
- 512 37. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* (2014).
513 doi:10.1038/nature13582
- 514 38. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43**, 95–100
515 (2011).
- 516 39. Korobkova, E., Emonet, T., Vilar, J. M., Shimizu, T. S. & Cluzel, P. From molecular noise to behavioural variability in a
517 single bacterium. *Nature* **428**, 574–578 (2004).
- 518 40. Bucci, V., Nunez-Milland, D., Twining, B. S. & Hellweger, F. L. Microscale patchiness leads to large and important intra-
519 specific internal nutrient heterogeneity in phytoplankton. *Aquat. Ecol.* **46**, 101–118 (2012).
- 520 41. Klausen, M. *et al.* Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Mol. Mi-
521 crobiol.* **48**, 1511–1524 (2003).
- 522 42. Klausen, M., Aaes-Jørgensen, A., Molin, S. & Tolker-Nielsen, T. Involvement of bacterial migration in the development
523 of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Mol. Microbiol.* **50**, 61–68 (2003).
- 524 43. Piciooreanu, C. *et al.* Microbial motility involvement in biofilm structure formation – a 3D modelling study. *Water Sci.
525 Technol.* **55**, 337–343 (2007).
- 526 44. Barken, K. B. *et al.* Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature
527 multicellular structures in *Pseudomonas aeruginosa* biofilms. *Environ. Microbiol.* **10**, 2331–2343 (2008).
- 528 45. Silander, O. K. *et al.* A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in *Escherichia coli*. *PLoS Genet*
529 **8**, e1002443 (2012).
- 530 46. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol.
531 Syst. Biol.* **4**, n/a–n/a (2008).
- 532 47. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc.
533 Natl. Acad. Sci. U. S. A.* **108**, E67–E76 (2011).
- 534 48. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*
535 **309**, 2075–2078 (2005).
- 536 49. Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival
537 games. *J. Theor. Biol.* **234**, 227–253 (2005).
- 538 50. Hellweger, F. L., Kravchuk, E. S., Novotny, V. & Gladyshev, M. I. Agent-based modeling of the complex life cycle of a
539 cyanobacterium (*Anabaena*) in a shallow reservoir. *Limnol. Oceanogr.* **53**, 1227–1241 (2008).
- 540 51. Whiteley, M. *et al.* Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413**, 860–864 (2001).
- 541 52. Sauer, K., Camper, A. K., Ehrlich, G. D., Costerton, J. W. & Davies, D. G. *Pseudomonas aeruginosa* displays multiple phe-
542 notypes during development as a biofilm. *J. Bacteriol.* **184**, 1140–1154 (2002).
- 543 53. Hentzer, M., Eberl, L. & Givskov, M. Transcriptome analysis of *Pseudomonas aeruginosa* biofilm development: anaero-
544 bic respiration and iron limitation. *Biofilms* **2**, 37–61 (2005).
- 545 54. Dötsch, A. *et al.* The *Pseudomonas aeruginosa* Transcriptome in Planktonic Cultures and Static Biofilms Using RNA
546 Sequencing. *PLoS ONE* **7**, e31092 (2012).
- 547 55. Kreft, J.-U. in *Food-Borne Microbes: Shaping the Host Ecosystem* (eds. Jaykus, L. A., Wang, H. H. & Schlesinger, L. S.)
548 347–377 (ASM Press, 2009).
- 549 56. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**,
550 1622–1625 (2004).
- 551 57. Maisonneuve, E. & Gerdes, K. Molecular mechanisms underlying bacterial persisters. *Cell* **157**, 539–548 (2014).
- 552 58. O’Connor, J. R., Kuwada, N. J., Huangyutham, V., Wiggins, P. A. & Harwood, C. S. Surface sensing and lateral subcellu-
553 lar localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol. Microbiol.*
554 **86**, 720–729 (2012).
- 555 59. Siryaporn, A., Kuchma, S. L., O’Toole, G. A. & Gitai, Z. Surface attachment induces *Pseudomonas aeruginosa* virulence.
556 *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16860–16865 (2014).
- 557 60. Rice, A. R., Hamilton, M. A. & Camper, A. K. Apparent surface associated lag time in growth of primary biofilm cells.
558 *Microb. Ecol.* **40**, 8–15 (2000).
- 559 61. Kreft, J.-U. Biofilms promote altruism. *Microbiology* **150**, 2751–2760 (2004).
- 560 62. Momeni, B., Waite, A. J. & Shou, W. Spatial self-organization favors heterotypic cooperation over cheating. *eLife* **2**,
561 e00960 (2013).
- 562 63. Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and
563 spatial dynamics. *Cell Rep.* **7**, 1104–1115 (2014).

- 564 64. Estrela, S. & Brown, S. P. Metabolic and demographic feedbacks shape the emergent spatial structure and function of
565 microbial communities. *PLoS Comput. Biol.* **9**, e1003398 (2013).
- 566 65. Batstone, D. J., Picioreanu, C. & van Loosdrecht, M. C. M. Multidimensional modelling to investigate interspecies hy-
567 drogen transfer in anaerobic biofilms. *Water Res.* **40**, 3099–3108 (2006).
- 568 66. Momeni, B., Brileya, K. A., Fields, M. W. & Shou, W. Strong inter-population cooperation leads to partner intermixing in
569 microbial communities. *eLife* **2**, (2013).
- 570 67. Wang, G. & Or, D. Trophic interactions induce spatial self-organization of microbial consortia on rough surfaces. *Sci.*
571 *Rep.* **4**, (2014).
- 572 68. Matsumoto, S. *et al.* Experimental and simulation analysis of community structure of nitrifying bacteria in a mem-
573 brane-aerated biofilm. *Water Sci. Technol.* **55**, 283–290 (2007).
- 574 69. Matsumoto, S. *et al.* Microbial community structure in autotrophic nitrifying granules characterized by experimental
575 and simulation analyses. *Environ. Microbiol.* **12**, 192–206 (2010).
- 576 70. Clark, J. R., Daines, S. J., Lenton, T. M., Watson, A. J. & Williams, H. T. P. Individual-based modelling of adaptation in
577 marine microbial populations using genetically defined physiological parameters. *Ecol. Model.* **222**, 3823–3837 (2011).
- 578 71. Clark, J. R., Lenton, T. M., Williams, H. T. P. & Daines, S. J. Environmental selection and resource allocation determine
579 spatial patterns in picophytoplankton cell size. *Limnol. Oceanogr.* **58**, 1008–1022 (2013).
- 580 72. Hellweger, F. L., van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-
581 based model. *Science* **345**, 1346–1349 (2014).
- 582 73. Koonin, E. V. & Wolf, Y. I. Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observa-
583 tions on virus–host coevolution. *Mol. Biosyst.* **11**, 20–27 (2014).
- 584 74. Emonet, T. & Cluzel, P. Relationship between cellular response and behavioral variability in bacterial chemotaxis. *Proc.*
585 *Natl. Acad. Sci. U. S. A.* **105**, 3304–3309 (2008).
- 586 75. Rudge, T. J., Steiner, P. J., Phillips, A. & Haseloff, J. Computational modeling of synthetic microbial biofilms. *Acs Synth.*
587 *Biol.* **1**, 345–352 (2012).
- 588 76. Stevens, J. T. & Myers, C. J. Dynamic Modeling of Cellular Populations within iBioSim. *ACS Synth. Biol.* **2**, 223–229
589 (2013).
- 590 77. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples
591 without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
- 592 78. Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L. & Palsson, B. O. Reconstruction of biochemical networks in microor-
593 ganisms. *Nat. Rev. Microbiol.* **7**, 129–143 (2009).
- 594 79. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype–phenotype relationship using a phy-
595 logeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
- 596 80. Klitgord, N. & Segrè, D. Environments that induce synthetic microbial ecosystems. *PLoS Comput. Biol.* **6**, e1001002
597 (2010).
- 598 81. Fuhrman, J., Follows, M. & Forde, S. Applying ‘-omics’ Data in Marine Microbial Oceanography. *Eos Trans. Am. Ge-*
599 *ophys. Union* **94**, 241–241 (2013).
- 600 82. Dens, E. J., Bernaerts, K., Standaert, A. R., Kreft, J.-U. & Van Impe, J. F. Cell division theory and individual-based model-
601 ing of microbial lag: Part II. Modeling lag phenomena induced by temperature shifts. *Int. J. Food Microbiol.* **101**, 319–
602 332 (2005).
- 603 83. Grimm, V. *et al.* The ODD protocol: a review and first update. *Ecol. Model.* **221**, 2760–2768 (2010).
- 604 84. Novere, N. L. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotech*
605 **23**, 1509–1515 (2005).
- 606 85. Courtot, M. *et al.* Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, (2011).
- 607 86. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochem-
608 ical network models. *Bioinformatics* **19**, 524–531 (2003).
- 609 87. Merkey, B. V., Lardon, L. A., Seoane, J. M., Kreft, J.-U. & Smets, B. F. Growth dependence of conjugation explains lim-
610 ited plasmid invasion in biofilms: an individual-based modelling study. *Environ. Microbiol.* **13**, 2435–2452 (2011).
- 611 88. Gharasoo, M., Centler, F., Fetzner, I. & Thullner, M. How the chemotactic characteristics of bacteria can determine their
612 population patterns. *Soil Biol. Biochem.* **69**, 346–358 (2014).
- 613 89. Wimpenny, J. W. T. & Colasanti, R. A unifying hypothesis for the structure of microbial biofilms based on cellular au-
614 tomaton models. *FEMS Microbiol. Ecol.* **22**, 1–16 (1997).
- 615 90. Hellweger, F. L. Resonating circadian clocks enhance fitness in cyanobacteria *in silico*. *Ecol. Model.* **221**, 1620–1629
616 (2010).

617 **FURTHER INFORMATION**

618 Software pages:

619 CellModeller: www.cellmodeller.org

620 CHASTE (Cancer, Heart And Soft Tissue Environment): www.cs.ox.ac.uk/chaste/

621 CompuCell3D: www.compuCell3d.org

622 FLAME (Flexible Large-scale Agent Modelling Environment): www.flame.ac.uk

623 iDynoMiCS (individual-based Dynamics of Microbial Communities Simulator): www.idynamics.org

624 NetLogo: ccl.northwestern.edu/netlogo/

625 Laboratory home pages:

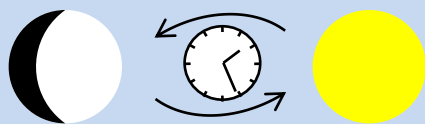
626 Robert J Clegg's and Jan-Ulrich Kreft's lab home page:

627 <http://www.biosciences-labs.bham.ac.uk/kreftlab/>

628 Ferdi L. Hellweger's home page:

629 www.systemsbioecology.org

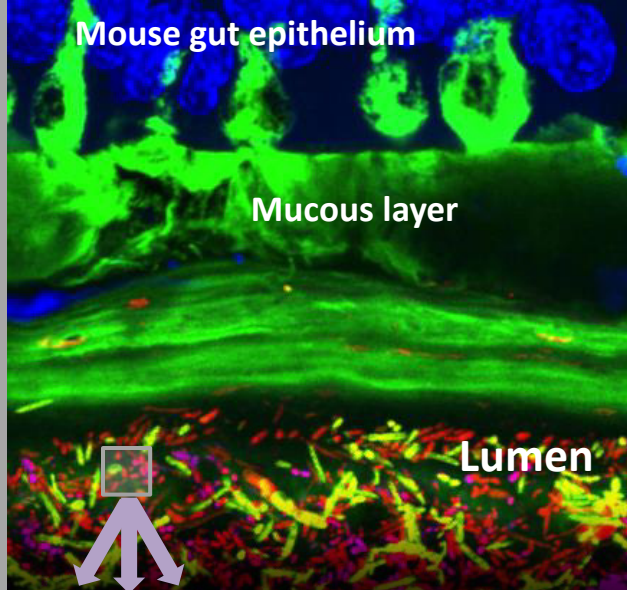
Population/Community



Mouse gut epithelium

Mucous layer

Lumen



Modelling approaches used for a wide range of environments:

Non-spatial

Spatial

Continuous

Discrete

ODE

PDE

IBM

$$\dot{R} = g(X,R)$$

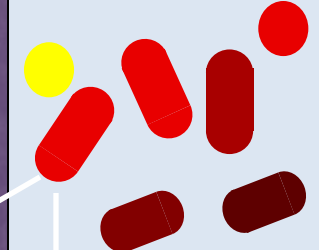
$$\dot{X} = f(X,R)$$

$$\dot{R} = g(X,R) + \text{flux}(R)$$

$$\dot{X} = f(X,R) + \text{flux}(X)$$

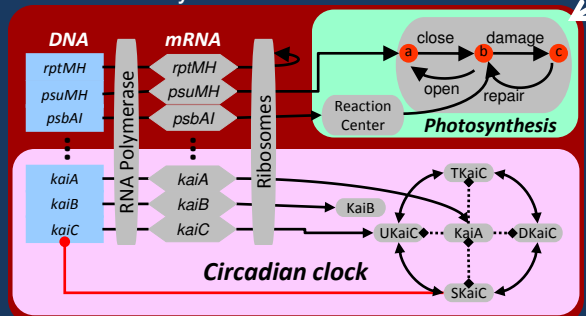
$$\dot{R} = \sum g_i(R)I_i + \text{flux}(R)$$

$$\dot{I}_i = f_i(R)I_i + \text{move}(\Sigma I_j)$$

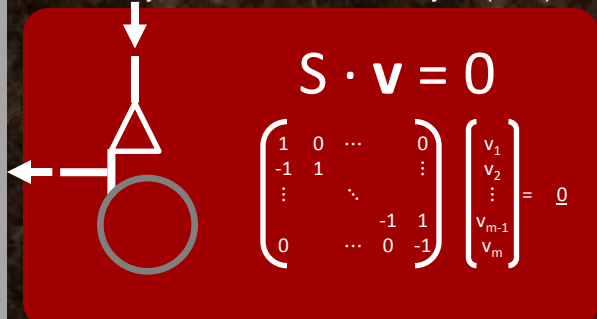


Intracellular

Dynamic kinetic model

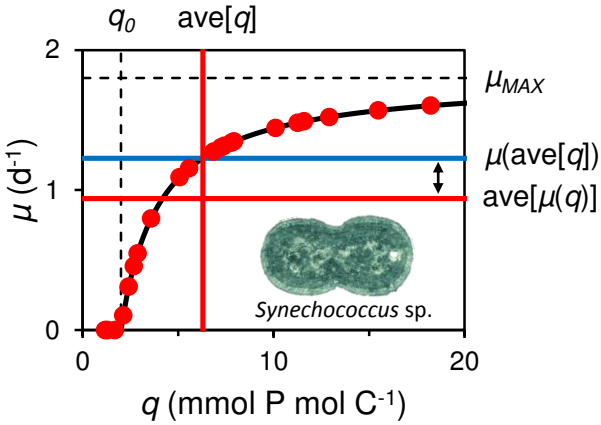


Steady-state flux balance analysis (FBA)

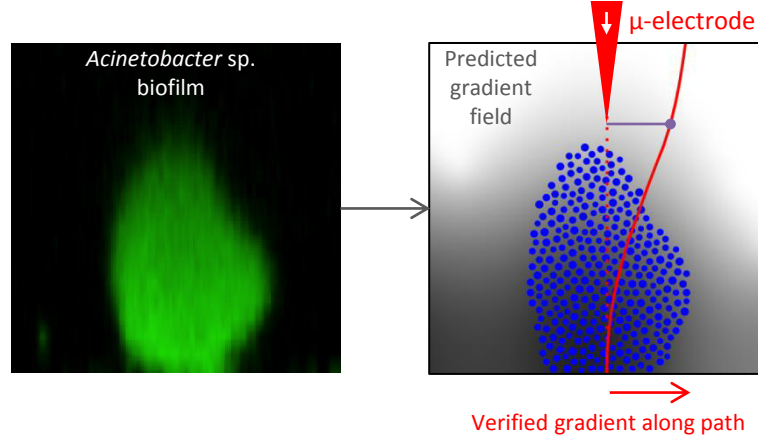


IBMs bridge scales from intracellular to community

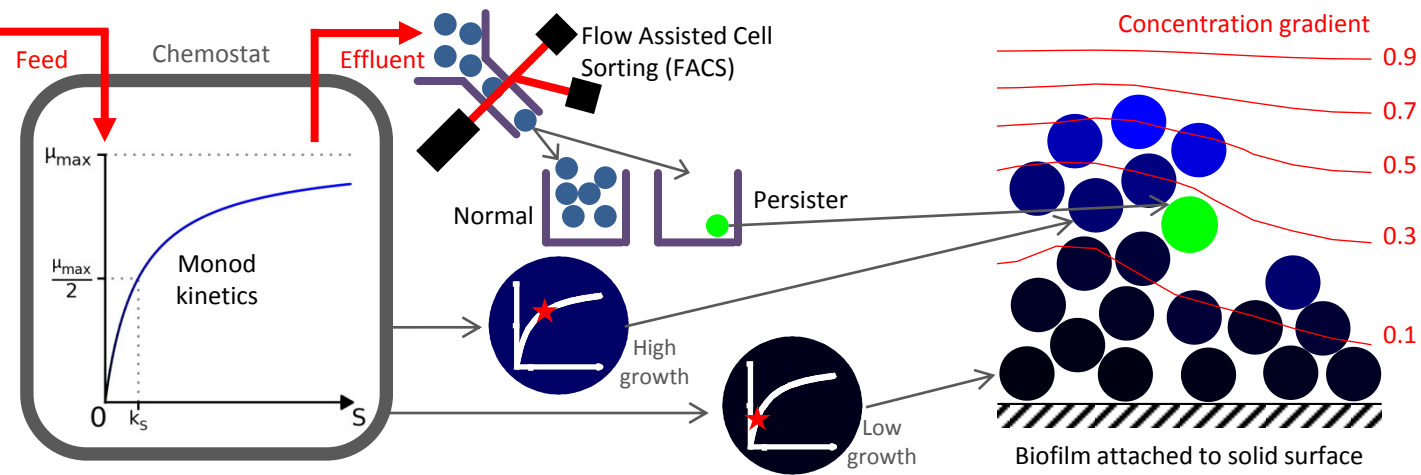
a Non-linear Droop kinetics and the averaging fallacy



b Predicting unobserved diffusion gradients from observed biomass (PDE) or cell (IBM) distribution



c Predicting biofilm growth and structure from chemostat kinetics (PDE or IBM if, e.g., persisters)

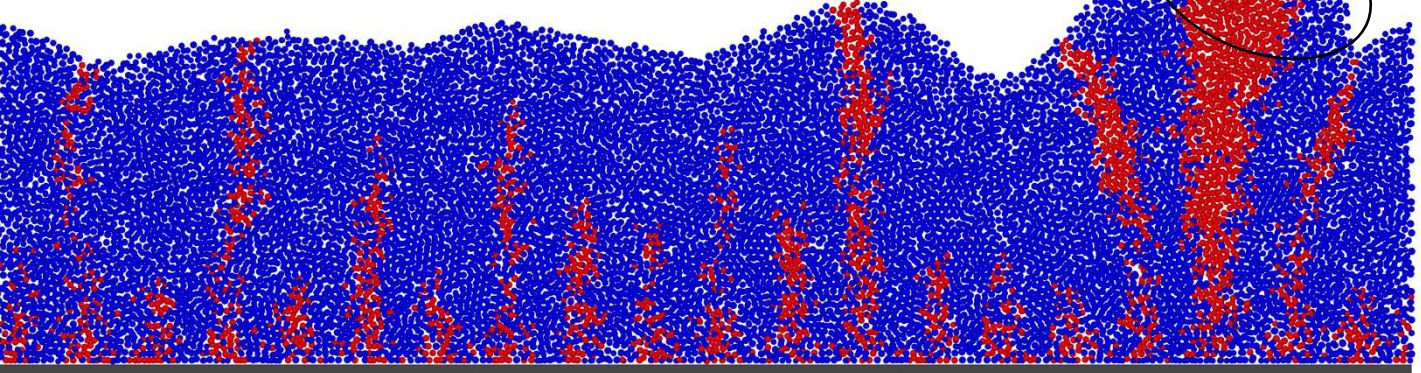


d Predicting emerging interactions and structure feedbacks

Red cells economically use shared resources (benefit neighbours: altruists)
 Blue cells consume more resources to grow faster (benefit self: cheaters)

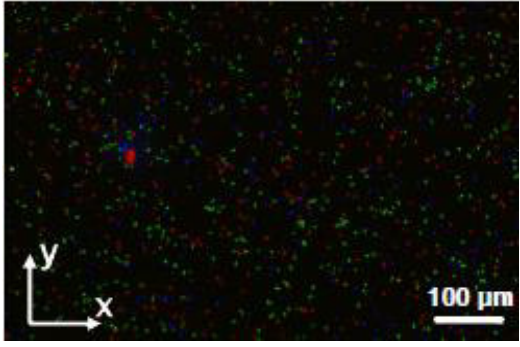
Spontaneously emerging clusters of altruists benefit cluster neighbours and outgrow cheaters

If altruists are scattered they benefit nearby cheaters

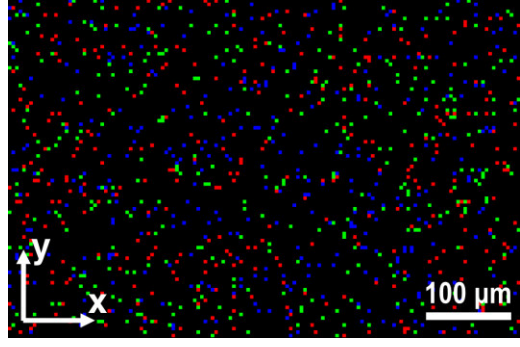


a Locally cross-feeding G and R strains benefit each other such that cheater C becomes spatially excluded from the cross-feeding over time

Experiment with engineered yeast strains

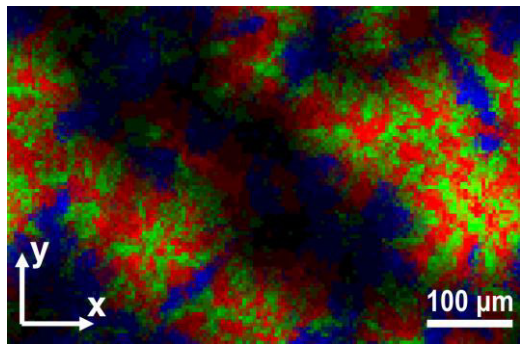
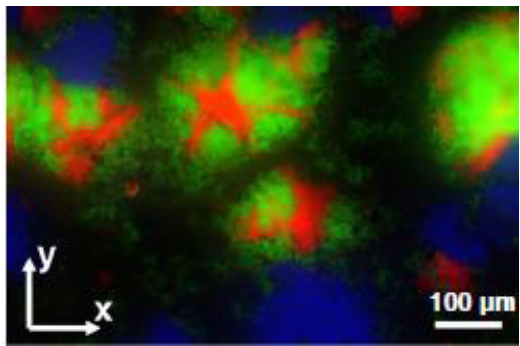
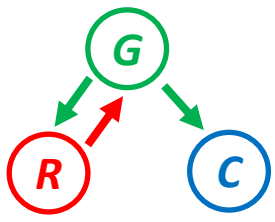


Simulation

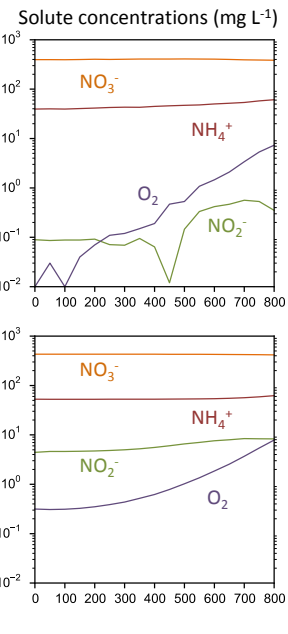
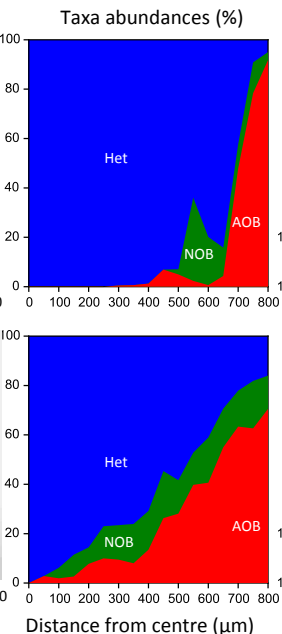
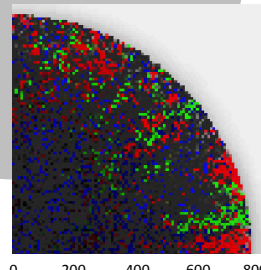
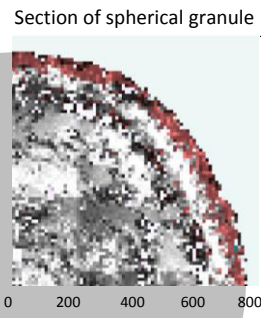
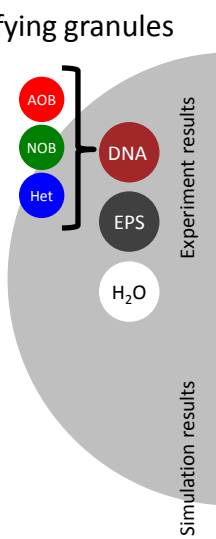
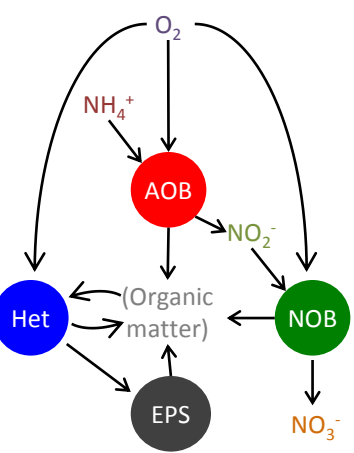


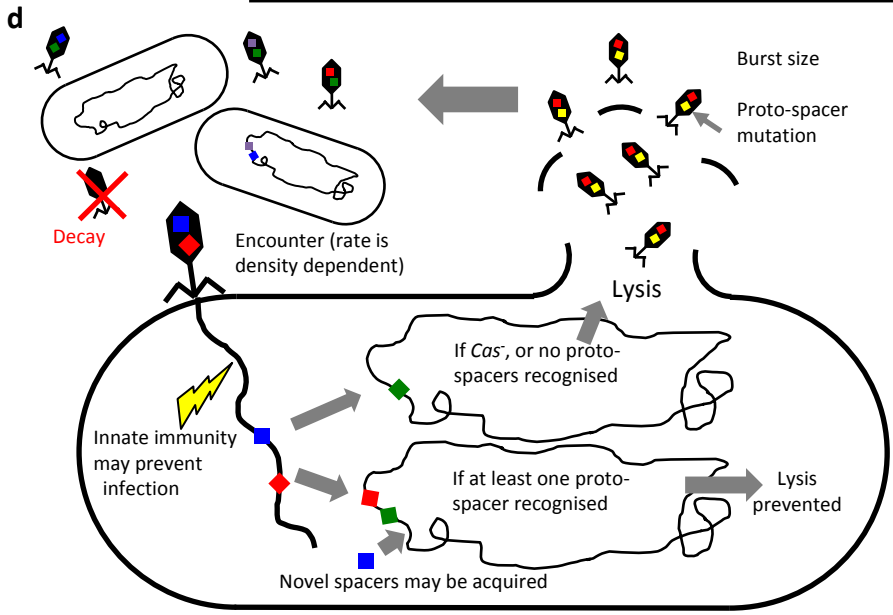
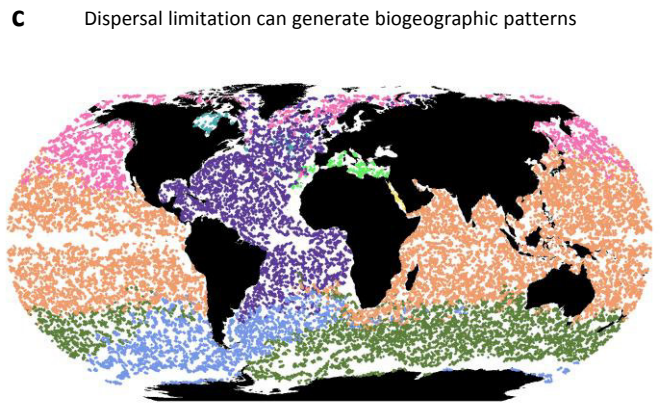
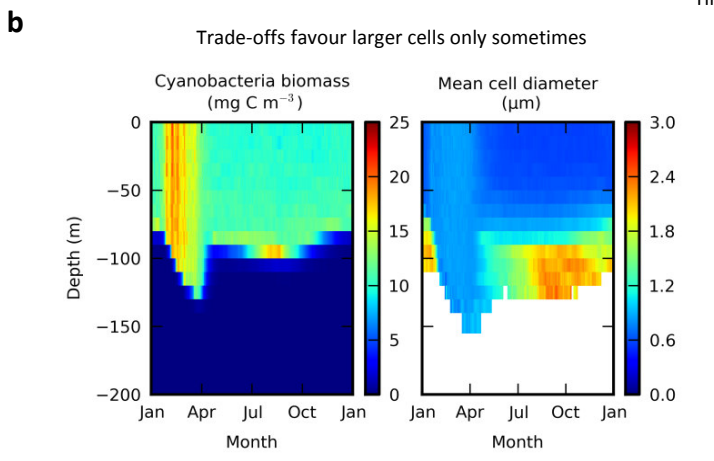
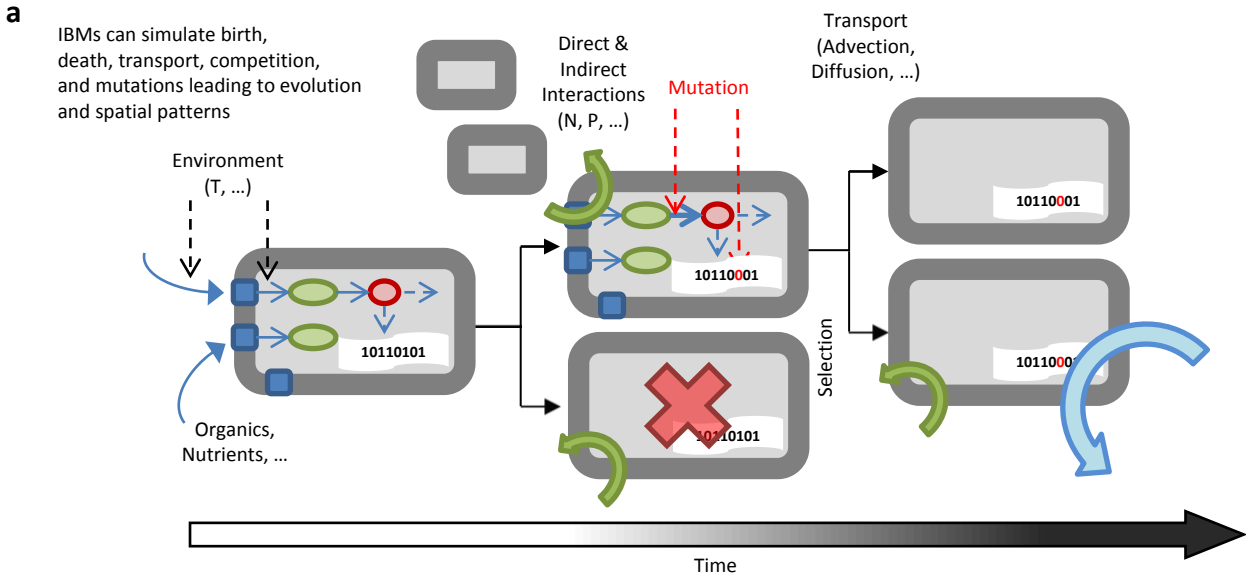
Growth ↓

Growth ↓



b Validation of IBM-predicted solute and biomass profiles in nitrifying granules

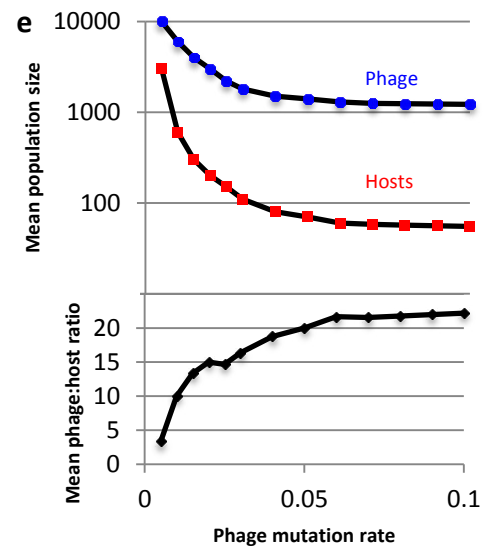


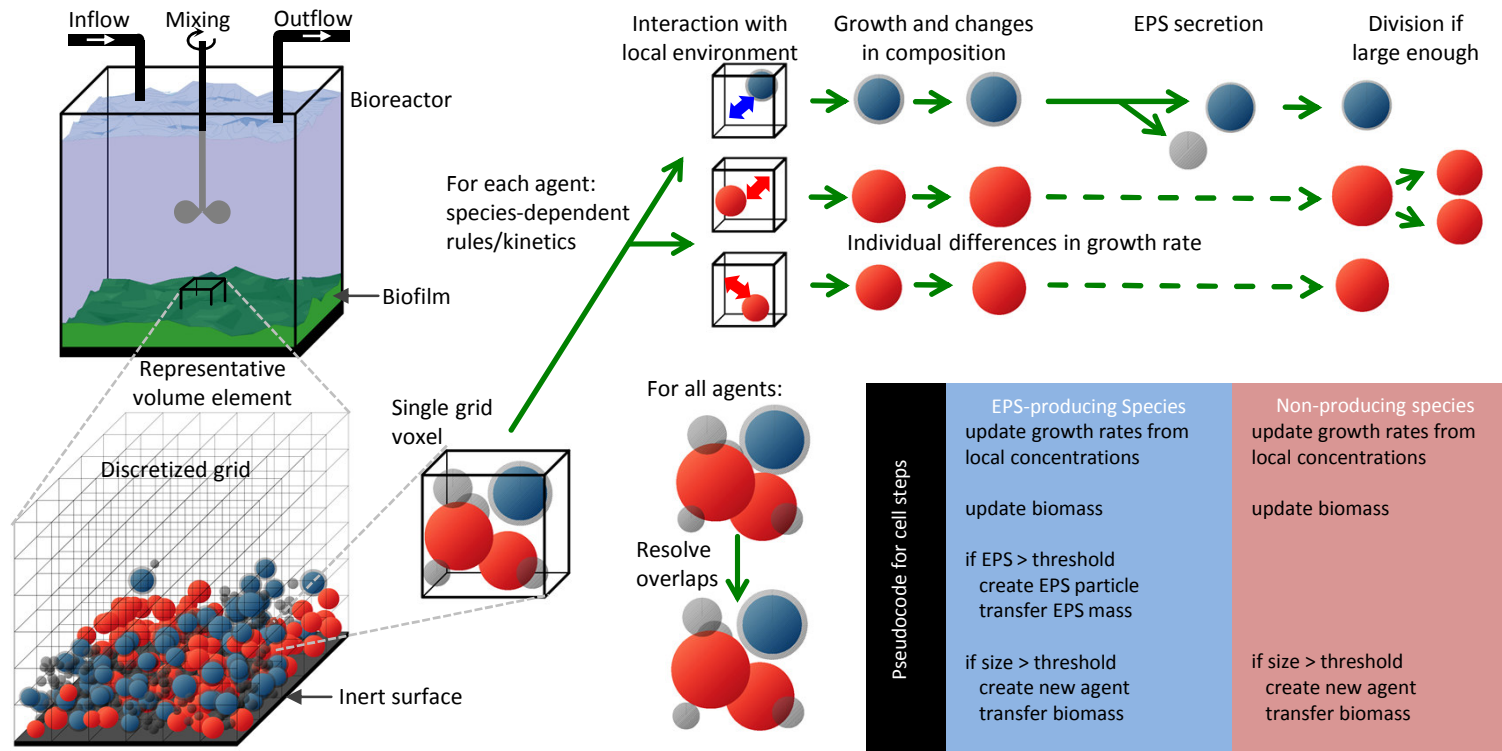


- Also modelled:
- Loss of CRISPR-Cas cassette or individual spacers during host division
 - Horizontal gene transfer of CRISPR-Cas cassette between hosts
 - Growth cost for *Cas*⁺ hosts

Higher mutation rate or lower spacer-acquisition rate
 → More phage per host, but fewer hosts
 → Fewer phage overall and less diversity

Higher immunity
 → Higher host population
 → Higher phage population
 → Higher phage diversity





Primer of some modelling approaches used in microbial ecology

Note that this characterisation of various approaches simplifies in order to emphasize what is typical; in fact, there is a range of complexities for each type of model. For example, IBMs do not need to have spatial structure and can be quite simple to reveal general principles, while PLMs can be applied to particular systems and then become quite complex.

Population-level models (PLMs) directly describe the changes of the populations^{1,2}. This can be achieved with differential equations if time is considered to be continuous or difference equations if time is considered to be discrete, e.g., if the population is stepped from generation to generation, or from year to year. PLMs are typically simple models that use a mass action approach for modelling interactions between different species. The mass action approach was adopted from kinetics of chemical reactions, where the probability of two molecules colliding and interacting is proportional to the concentration of each molecular species³. PLMs may consider the dynamics of resources explicitly, or assume a density dependence of growth rate such as logistic growth which implicitly considers resource depletion at higher population density^{1,2}. The main advantages of PLMs are that they are relatively easy to describe and analyse, and they require less knowledge and data as they have fewer parameters. Their main purpose is to discover ‘general principles’ or concepts, as details are avoided. They have therefore been classified as ‘strategic’, ‘demonstration’ or ‘conceptual’ models⁴. Due to their general nature, they are less suitable to predict specific populations in specific ecosystems as would be desired for ecosystem management^{4,5}.

PLMs are often based on ordinary differential equations (ODEs) or partial differential equations (PDEs); these are closely related, facilitating the exploration of corresponding ODE and PDE models^{1,2}. ODEs assume homogenous space and are therefore appropriate for well-mixed systems such as chemostats or batch cultures. However, a non-uniform system may be represented reasonably well as consisting of different compartments. Then, a different set of ODEs for each compartment, and exchange between compartments, can approximate spatial structure. For example, a predator-prey model could have two types of habitat: one with food and predator, and one that is a refuge^{1,2}.

PDEs are mostly used for fully spatially explicit models^{1,2}. The effects of spatial structure can be inferred from comparing corresponding ODEs and PDEs. These are both continuum models, where time, space and species densities are continuous, rather than discrete, variables. As a result, populations may become infinitesimally small without becoming extinct. The similar difference equations are discrete in time and either discrete in population size or continuous in population density. If they are discrete in population size, extinction occurs more readily. As a consequence of using discrete time, population responses have a built-in delay, i.e. poor weather in one year affects the population size only in the next year. This delay renders dynamics less stable^{1,2}.

Whilst PLMs usually neglect population heterogeneity, they can incorporate population structure in two ways². One is to separate the population into multiple age classes or life-cycle stages and describe the rate of change for each class by a separate ODE, e.g., ‘graduation’ from one age class to the next would be based on the growth rate. Another is to use a PDE to represent the population structure, e.g., age or size structure, as a continuum². Unavoidably, the more complex a PLM becomes, the more it loses its advantage over a corresponding IBM of being simpler and more tractable mathematically⁴.

Individual-based models (IBMs), in contrast to PLMs, do not describe changes on the population level at all: they only describe the activities and properties of individuals, how they change the environment, and how they respond to the environment⁶⁻¹⁰. Changes on the population level emerge automatically from all the interactions between individuals and the environment. Therefore, IBMs are classified as bottom-up models, describing the lower level to predict the higher level. For the same reason, PLMs are classified as top-down models. While PLMs can be made more and more complex to include population and spatial structure, thereby coming closer to IBMs, they remain top-down, describing the changes on the population level, rather than directly describing individuals like IBMs.

Because IBMs map individual behaviour to population dynamics, they can bridge these two scales and use data on both levels: observations of individual behaviour can be used as input into the model and observations of population dynamics can be compared with model output. Alternatively, individual behaviours can be inferred from comparing the kinds of population dynamics and patterns an assumed individual behaviour would produce with those dynamics and patterns observed, over a range of conditions. This has been called pattern-oriented modelling¹¹.

Individuals in IBMs are always discrete, but they may be either cells in a spatial grid (lattice) or particles in continuous space. In the first case, the IBM can also be called a cellular automaton (CA) where updating of lattice elements, diffusion of metabolites, cell division and movements are all specified by rules^{12,13}. Individuals typically occupy a single grid element and can move to a neighbouring grid element only in certain directions (at angles that are multiples of 45° or 90°), which can lead to spatial artefacts¹⁴. A common rule for cell division is this: if a threshold mass is reached, divide into equal daughter cells. One daughter cell picks one of the free neighbouring cells at random, or if there are none free, pushes a random neighbouring cell away, which then moves according to the same rules¹⁵⁻¹⁷. To avoid artefacts, lattice elements should be updated in random order¹⁸.

Modelling individuals as particles with real size in a continuous space facilitates physically correct modelling of mechanical interactions between cells; these may be collisions or pushing away of cells that have encroached on one another due to motility or expansion of cellular volumes¹⁹⁻²¹. Such models can also be called particle-based models, but note that IBMs are only that subset of particle-based models where the particles may differ and have adaptive behaviour. Whether based on CAs or particles, a useful feature of IBMs is that behaviour can be described using simple rules and "if statements" that are not easily captured with differential equations⁶⁻¹⁰. For example, cell division is commonly based on a cell size threshold²².

Individuals in IBMs are autonomous agents that have their own state and carry out activities according to their state and in response to the environment. Hence, individual-based models are often called agent-based models. However, the term agent is more general as an agent does not have to be an individual. Agents can cover many scales, from molecular entities, cells, individual organisms, to social groups of organisms such as families, or larger social or economic organizations¹⁰.

Since IBMs explicitly simulate individuals, they can simulate population heterogeneity in a straightforward manner. However, it should be noted that IBMs of systems with a very high number of individuals generally do not explicitly simulate all microbial cells, but representative ones called "super-individuals"^{23,24}. So even in IBMs there may be some lumping. Therefore, in terms of population heterogeneity, there is no hard distinction between PLMs and IBMs: The resolution increases smoothly from PLMs to super-individual IBMs to true IBMs. However, the two approaches are still fundamentally different in that the PLM describes the behaviour of the population and that the IBM describes the behaviour of an individual^{23,24}.

This difference between PLMs and IBMs can be illustrated with the example of microbial growth, which is fundamental for any modelling of population dynamics. Growth kinetics are non-linear, and this has important consequences. In Droop kinetics, a commonly-used model for growth of phytoplankton, the specific growth rate depends on the cell's internal content of the limiting nutrient²⁵. If internal nutrient contents vary between individuals, as observed in samples from the environment, the sum of the growth rates of all individuals will be different from the growth rate of a population with an average nutrient content²⁵. This is an example of a well-known mathematical theorem, known as Jensen's Inequality, that the average of a non-linear response to some heterogeneous input is different from the response to the average input²⁶.

In Monod kinetics, the standard model for growth of heterotrophic bacteria, the specific growth rate does not depend on the internal nutrient content of the individual, but on the substrate concentration in the environment²⁷. Thus, it could be modelled with PLMs or IBMs, depending on the purpose of the model, e.g., whether other effects of individuality are to be considered or not. IBMs would be more appropriate for the purpose of modelling growth if the Monod kinetic parameters (maximum specific growth rate and substrate affinity²⁷) would differ between individuals. Such individual differences could be due to variation in expression of genes for uptake and metabolic enzymes between cells²⁸. Variation in maximal specific growth rates have been observed^{29,30}, most notably in populations with non-growing persister cells^{31,32}, but variation in substrate affinity between different cells has, to our knowledge, not been investigated. This could be studied in microfluidic single-cell chemostats³³. If individuals had different Monod kinetics, the kinetics of the population, which could be inferred with an IBM summing the rates of the individuals, would deviate from Monod kinetics. However, this would be difficult to observe in large populations, especially as individual growth rates fluctuate over time³⁴ and faster growing lineages would become more frequent in the population over time and so come to dominate the population kinetics.

Models of intracellular dynamics, such as metabolism or gene regulation, can be integrated into IBMs, since IBMs have that flexibility of describing the activities of individuals by any means available to a programmer: from simple rules to complex, computationally expensive submodels. Focussing here on metabolism, there are two main ways in which metabolic fluxes can be predicted: dynamic kinetic models and steady state flux-balance models³⁵. Ideally, one would like to be able to use a dynamic kinetic model and write down the kinetic equations for all enzymes in a cell and then simulate the resulting fluxes through the metabolic network, from which growth rates could be predicted. The advantage of such dynamic kinetic models is that they can simulate the effect of changes in metabolite or enzyme concentrations, or in regulation of enzyme activity. However, this is not feasible for a genome-wide metabolic network, as the kinetics are only known for a limited number of enzymes from a limited number of species and often not under physiological conditions. Therefore, one usually either neglects large parts of the metabolic network or represents them as a stoichiometric model, and focusses instead on energy metabolism, where more is known³⁶.

For most species, even the kinetics of catabolic enzymes are not known sufficiently to use dynamic kinetic models or one wants to include less well studied enzymes. Then, genome-wide flux-balance models, also known as constraint-based models, can be used instead because they only require knowledge of the list of enzyme reactions coded for by the genome and the stoichiometries of these reactions^{37,38}. However, the reaction rates can only be calculated when the equations are simplified by assuming that the system is in steady state, i.e., that the concentrations of the metabolites do not change with time. This is a reasonable assumption during exponential growth. Then, the distribution of fluxes (reaction rates) through the metabolic network that fulfil the stoichiometry can be calculated. To narrow down the space of possible solutions for these flux distributions, one uses

constraints, the more the better. For example, using experimentally measured fluxes, placing upper bounds on reaction rates, using thermodynamics, or using gene expression data to remove reactions catalysed by those enzymes that are not produced under given conditions. To obtain a unique solution for the flux distribution within the narrowed down solution space, one assumes that the flux distribution is optimal for the growth of the cell^{37,38}. Commonly, the objective function for optimization is to maximize biomass production (growth yield), although the choice of objective function can be debated³⁹.

References

1. Edelstein-Keshet, L. *Mathematical models in biology*. (McGraw-Hill, 1988).
2. Gurney, W. & Nisbet, R. M. *Ecological dynamics*. (Oxford University Press, 1998).
3. Voit, E. O., Martens, H. A. & Omholt, S. W. 150 Years of the Mass Action Law. *PLoS Comput. Biol.* **11**, e1004012 (2015).
4. Evans, M. R. *et al.* Do simple models lead to generality in ecology? *Trends Ecol. Evol.* **28**, 578–583 (2013).
5. Topping, C. J., Hoyer, T. T., Odderskaer, P. & Aebischer, N. J. A pattern-oriented modelling approach to simulating populations of grey partridge. *Ecol. Model.* **221**, 729–737 (2010).
6. Grimm, V. & Railsback, S. F. *Individual-based modeling and ecology*. (Princeton University Press, 2005).
7. Railsback, S. F. & Grimm, V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. (Princeton University Press, 2012).
8. *Individual-based models and approaches in ecology: populations, communities, and ecosystems*. (Chapman and Hall, 1992).
9. DeAngelis, D. L. & Mooij, W. M. Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol. Syst.* **36**, 147–168 (2005).
10. Wilensky, U. & Rand, W. *An introduction to agent-based modeling: modeling natural, social and engineered complex systems with NetLogo*. (MIT Press, 2015).
11. Grimm, V. *et al.* Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* **310**, 987–991 (2005).
12. Hogeweg, P. Cellular Automata as a paradigm for ecological modeling. *Appl. Math. Comput.* **27**, 81–100 (1988).
13. Ermentrout, G. B. & Edelstein-Keshet, L. Cellular Automata approaches to biological modeling. *J. Theor. Biol.* **160**, 97–133 (1993).
14. Schönfisch, B. Anisotropy in cellular automata. *Biosystems* **41**, 29–41 (1997).
15. Picioreanu, C., van Loosdrecht, M. C. M. & Heijnen, J. J. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotechnol. Bioeng.* **58**, 101–116 (1998).
16. Kreft, J.-U., Picioreanu, C., Wimpenny, J. W. T. & van Loosdrecht, M. C. M. Individual-based modelling of biofilms. *Microbiology* **147**, 2897–2912 (2001).
17. Tang, Y. & Valocchi, A. J. An improved cellular automaton method to model multispecies biofilms. *Water Res.* **47**, 5729–5742 (2013).
18. Schönfisch, B. & de Roos, A. Synchronous and asynchronous updating in cellular automata. *Biosystems* **51**, 123–143 (1999).
19. Farrell, F. D. C., Hallatschek, O., Marenduzzo, D. & Waclaw, B. Mechanically driven growth of quasi-two-dimensional microbial colonies. *Phys. Rev. Lett.* **111**, 168101 (2013).

20. Rudge, T. J., Federici, F., Steiner, P. J., Kan, A. & Haseloff, J. Cell Polarity-Driven Instability Generates Self-Organized, Fractal Patterning of Cell Layers. *ACS Synth. Biol.* **2**, 705–714 (2013).
21. Storck, T., Picioareanu, C., Viridis, B. & Batstone, D. J. Variable cell morphology approach for Individual-based Modeling of microbial communities. *Biophys. J.* **106**, 2037–2048 (2014).
22. Kreft, J.-U., Booth, G. & Wimpenny, J. W. T. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* **144**, 3275–3287 (1998).
23. Scheffer, M., Baveco, J. M., DeAngelis, D. L., Rose, K. A. & Van Nes, E. H. Super-Individuals: a simple solution for modeling large populations on an individual basis. *Ecol. Model.* **80**, 161–170 (1995).
24. Hellweger, F. L. & Bucci, V. A bunch of tiny individuals - individual-based modeling for microbes. *Ecol. Model.* **220**, 8–22 (2009).
25. Bucci, V., Nunez-Milland, D., Twining, B. S. & Hellweger, F. L. Microscale patchiness leads to large and important intraspecific internal nutrient heterogeneity in phytoplankton. *Aquat. Ecol.* **46**, 101–118 (2012).
26. Jensen, J. L. W. V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **30**, 175–193 (1906).
27. Kreft, J.-U. in *Food-Borne Microbes: Shaping the Host Ecosystem* (eds. Jaykus, L. A., Wang, H. H. & Schlesinger, L. S.) 347–377 (ASM Press, 2009).
28. Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D. & Luthey-Schulten, Z. A. Heterogeneity in protein expression induces metabolic variability in a modeled *Escherichia coli* population. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14006–14011 (2013).
29. Dusny, C., Fritsch, F. S. O., Frick, O. & Schmid, A. Isolated Microbial Single Cells and Resulting Micropopulations Grow Faster in Controlled Environments. *Appl. Environ. Microbiol.* **78**, 7132–7136 (2012).
30. Iyer-Biswas, S. *et al.* Scaling laws governing stochastic growth and division of single bacterial cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15912–15917 (2014).
31. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
32. Maisonneuve, E. & Gerdes, K. Molecular mechanisms underlying bacterial persisters. *Cell* **157**, 539–548 (2014).
33. Hol, F. J. H. & Dekker, C. Zooming in to see the bigger picture: Microfluidic and nanofabrication tools to study bacteria. *Science* **346**, 1251821 (2014).
34. Wang, P. *et al.* Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
35. Klipp, E., Herwig, R., Kowald, A., Wierling, C. & Lehrach, H. *Systems biology in practice*. (Wiley-VCH, 2005).
36. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K. & Reuss, M. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* **79**, 53–73 (2002).
37. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
38. Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L. & Palsson, B. O. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **7**, 129–143 (2009).
39. Schuster, S., Pfeiffer, T. & Fell, D. A. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.* **252**, 497–504 (2008).

SI for Hellweger FL, Clegg RJ, Clark J, Plugge CM, Kreft JU (2015). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*

Box S2 | Software for individual-based modelling in microbial ecology

Overview of available platforms. The open-source platforms characterized here are fairly generic platforms we regard as particularly suited for microbial ecology: NetLogo^{1,2} and FLAME³ are the most generic, and often used for non-microbial IBMs. CellModeller^{4,5}, CHASTE⁶ and CompuCell3D⁷ were designed for modelling tissues but can easily model biofilms as they are similar. CompuCell3D and iDynoMiCS⁸ require the least programming skills. See Supplementary Information S3 (table) for more details. Apart from these platforms, generic agent-based modelling software libraries such as Repast⁹ can be used by programmers to rapidly build custom models (see Supplementary Information S1 (text) for an explanation of agent-based modelling). Nevertheless, more specialized software may be better suited for a specific application: iAlgae for photosynthetic microbes¹⁰, Virtual Ecology Workbench (VEW)/Planktonica for plankton models¹¹, INDISIM for carbon and nitrogen cycling in soil¹² or lag phase in liquid media¹³, Framework for biofilm models¹⁴, AgentCell for chemotaxis signalling¹⁵, COMETS for metabolite exchange¹⁶ and BSim for gene regulation¹⁷.

CellModeller^{4,5} focusses on mechanical forces between cells as it has been developed to model the growth of plant tissues and bacterial colonies. Applied to explaining formation of fractal boundaries between fluorescently-tagged, rod-shaped *E. coli* in expanding colonies. Also models signalling between cells, but lacks substrate diffusion.

*CHASTE*⁶ (Cancer, Heart And Soft Tissue Environment) is a generic simulator for animal tissues. Since biofilms are tissue-like, the excellent capabilities of CHASTE (cellular behaviour, mechanical forces, metabolite and signal transport) could be adapted with some programming effort for biofilms; this would be particularly suited for biofilms associated with tissues.

*CompuCell3D*⁷ is also a tissue simulator, but has already been used for biofilm structure formation. Cells have variable shape as they are made-up of several grid elements; their interactions are specified by 'contact energies', which is natural for the mechanical forces between growing or motile, differentially adherent cells, but can also specify, e.g., signalling.

*FLAME*³ (Flexible Large-scale Agent Modelling Environment) is very general and suited for large scale simulations as agents interact by broadcasting messages. This enables automatic parallel execution on compute clusters. Biological applications include various tissue models and *E. coli* interacting with oxygen.

*iDynoMiCS*⁸ (individual-based Dynamics of Microbial Communities Simulator) was developed to enable biofilm researchers without programming skills to run biofilm simulations. Users are guided through the specification of species, reaction kinetics, substrates etc. by a web tool. The number of environments iDynoMiCS can simulate is increasing but there are still important gaps.

NetLogo^{1,2} is an easy-to-use platform for IBMs with a large user community. It requires some programming aptitude, but its own high-level language makes NetLogo programs very concise and a de-facto standard for communicating IBMs. It has been used for simulating microbial dynamics but lacks a powerful physics engine for simulating solute diffusion.

See also Table S3 for a feature and characteristics matrix of the above platforms.

SI for Hellweger FL, Clegg RJ, Clark J, Plugge CM, Kreft JU (2015). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*

References

1. Wilensky, U. NetLogo [online]. (1999). at <<http://ccl.northwestern.edu/netlogo>>
2. Railsback, S. F. & Grimm, V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. (Princeton University Press, 2012).
3. Holcombe, M. *et al.* Modelling complex biological systems using an agent-based approach. *Integr. Biol.* **4**, 53–64 (2012).
4. Rudge, T. J., Steiner, P. J., Phillips, A. & Haseloff, J. Computational modeling of synthetic microbial biofilms. *Acs Synth. Biol.* **1**, 345–352 (2012).
5. Rudge, T. J., Federici, F., Steiner, P. J., Kan, A. & Haseloff, J. Cell polarity-driven instability generates self-organized, fractal patterning of cell layers. *ACS Synth. Biol.* (2013). doi:10.1021/sb400030p
6. Mirams, G. R. *et al.* Chaste: an open source C++ library for computational physiology and biology. *PLoS Comput. Biol.* **9**, e1002970 (2013).
7. Poplawski, N. J., Shirinifard, A., Swat, M. & Glazier, J. A. Simulation of single-species bacterial-biofilm growth using the Glazier-Graner-Hogeweg model and the CompuCell3D modeling environment. *Math. Biosci. Eng.* **5**, 355–388 (2008).
8. Lardon, L. A. *et al.* iDynaMiCS: next-generation individual-based modelling of biofilms. *Environ. Microbiol.* **13**, 2416–2434 (2011).
9. Macal, C. M. & North, M. J. Tutorial on agent-based modelling and simulation. *J. Simul.* **4**, 151–162 (2010).
10. Hellweger, F. L. & Bucci, V. A bunch of tiny individuals - individual-based modeling for microbes. *Ecol. Model.* **220**, 8–22 (2009).
11. Hinsley, W., Field, T. & Woods, J. Creating individual based models of the plankton ecosystem. *Comput. Sci. - ICCS 2007 Pt 1 Proc.* **4487**, 111–118 (2007).
12. Gras, A., Ginovart, M., Valls, J. & Baveye, P. C. Individual-based modelling of carbon and nitrogen dynamics in soils: Parameterization and sensitivity analysis of microbial components. *Ecol. Model.* **222**, 1998–2010 (2011).
13. Ginovart, M., Prats, C., Portell, X. & Silbert, M. Exploring the lag phase and growth initiation of a yeast culture by means of an individual-based model. *Food Microbiol.* **28**, 810–817 (2011).
14. Xavier, J. B., Picioreanu, C. & van Loosdrecht, M. C. M. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environ. Microbiol.* **7**, 1085–1103 (2005).
15. Emonet, T. & Cluzel, P. Relationship between cellular response and behavioral variability in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3304–3309 (2008).
16. Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* **7**, 1104–1115 (2014).
17. Gorochofski, T. E. *et al.* BSim: an agent-based tool for modeling bacterial populations in systems and synthetic biology. *PLoS ONE* **7**, e42790 (2012).

Supplementary Table S3 | Generic open source platforms for individual-based modelling in microbial ecology

	CellModeller	CHASTE (Cancer, Heart And Soft Tissue Environment)	CompuCell3D	FLAME (Flexible Large-scale Agent Modelling Environment)	iDynoMiCS (individual-based Dynamics of Microbial Communities Simulator)	NetLogo
Availability	www.cellmodeller.org	www.cs.ox.ac.uk/chaste/	www.compuCell3d.org	www.flame.ac.uk	www.idynamics.org	ccl.northwestern.edu/netlogo
Primary domains	Plant tissue growth, Bacterial colonies	Tissue development, Cardiac electrophysiology	Tissue development, Biofilms	Tissue development, Economics	Microbes, Biofilms, Chemostats	Ecology, Sociology, Economics, Teaching, others
Characteristics	Cells with various shapes interact mechanically and by signaling; gene regulation. Growth constrained by forces not substrate	IBM of cells with cell cycles, mechanical interactions; Metabolite and signal transport by PDEs, electrophysiology by PDEs	Contact energies between cells are minimized; Environment is described by PDEs	Agents move from state to state with transition functions and interact by broadcasting messages, which enables automatic parallel execution. Environment is also an Agent	Agents are discrete objects in continuous 3D space with processes in continuous time. Environment is described by PDEs	Agents are 'turtles' that live on a grid of 'patches', models are specified in the programming language Scala. Extensions link NetLogo to Matlab, R, GIS
Users can	Specify model and control simulations by programming. Initialize simulations with location data from microscopic images	Build code from source; specify cell-based models and control simulations by programming; use CHASTE as library for own development	Specify model with tool-specific XML file or Python script, helped by Wizard, GUI to run simulation	Specify model with tool-specific XML file combined with specifying agent transition functions by programming in C; all inputs are converted into C source code which the user compiles and runs	Specify model with tool-specific XML file, which requires no programming skills; GUI helps user generate XML file; further customization or extension does require programming skills	Rapidly develop model by programming agent behaviour and using a GUI to control the simulation and view output
Input and output formats	Models loaded/saved as Python scripts; simulation states loaded/saved as Pickles (Python object serialization)	Non cardiac models are specified by C++ code input; many standard text file formats for cell, mesh and other data output, suitable for VTK and meshing software	Models specified as tool-specific XML or Python scripts and lattice and concentration field text files; data output as VTK and other text files	Output of simulation data in tool-specific XML format	Models specified and simulation data saved/loaded as tool-specific XML files; can also read in previous simulation state and random number generator state to continue simulation with same or altered conditions/agents	Models are specified in Scala, simulation state or time series data can be saved/loaded as CSV files
Documentation and user support in addition to website	Publications, Library of demos	Publications, Tutorials, Library of demos, Code documentation, Mailing list, Bug tracking, Wiki, Workshops	Publications, Tutorials, Library of demos, Manuals, Workshops	Publications, Tutorials, Manual, Code documentation	Publication, Tutorial, Library of demos, Code documentation, Mailing list, Bug tracking, Wiki, Workshops	Publications, Books, Tutorials, Large library of demos and user contributed models, Comprehensive online manual, Mailing lists, User groups, Wiki, Chat channel, Twitter, Workshops
Appeared in	2005	2008	2004	2006	2011	1999
Stable release	4.2.1 (07/2015)	3.3 (01/2015)	3.7.4 (08/2015)	0.17.0 (07/2012)	1.3 (06/2015)	5.2 (04/2015)
Programming language	Python	C++	C++, user specifies models in XML or Python	C, model specified with XML files and C functions	Java (simulation output analysis in Matlab, R, Python)	Scala (compiles to Java byte code so can be run on any Java virtual machine)
Influenced by	Engineering tissue shapes, synthetic biology	Systems biology, software engineering	Cellular Potts Model, Complexity science	State machines, Parallel computing	Swarm, Gecko, BacSim, Framework, Biofilm models, Complexity science	Logo, Teaching emergence by creating ABMs, Complexity science

	CellModeller	CHASTE (Cancer, Heart And Soft Tissue Environment)	CompuCell3D	FLAME (Flexible Large-scale Agent Modelling Environment)	iDynoMiCS (individual-based Dynamics of Microbial Communities Simulator)	NetLogo
OS	Any with Python	Linux, OS X, (Win)	Windows, OS X, Linux	Any with C compiler	Any with Java	Any with Java
Example applications	Plant meristem growth, Rod-shaped bacteria generating fractal patterns	Intestinal crypts/colorectal cancer, Heart electromechanics	Tissue morphogenesis (limb and somite formation, tumour growth, angiogenesis), <i>Dictyostelium</i> fruiting body development	Skin, Signalling cascade, Neo-angiogenesis in cancer, <i>E. coli</i> interacting with oxygen, Market economy	Metabolic switching aerobic/anaerobic Plasmid transfer in biofilms Metabolic cooperation Aging in chemostats	Land use, Crowd dynamics, Traffic, Stock market, Cooperation, Peer review Foraging ants, Mice in agriculture, Daphnia, Plant facilitation, Bacterial colonies on leaves

Abbreviations:

CA: Cellular Automaton

CSV: Comma Separated Value text file

GIS: Geographical Information System (for spatial or geographical data)

GUI: Graphical User Interface

OS: Operating System

PDE: Partial Differential Equations

VTK: Visualization Tool Kit

XML: Extensible Markup Language