Advancing microbial sciences by individual-based modelling

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

DOI:
10.1038/nrmicro.2016.62

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

Advancing microbial sciences by individual-based modelling

Ferdi L. Hellweger[1], Robert J. Clegg[2], James Clark[3], Caroline M. Plugge[4], Jan-Ulrich Kreft[2]

[1]Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA.

[2]Centre for Systems Biology & Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK.

[3]Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK.

[4]Laboratory of Microbiology, Wageningen University, 6703 HB Wageningen, The Netherlands.

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

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

Recent technological advances[1–12], e.g., in microscopy, flow cytometry, microfluidics, spectroscopy, isotope and molecular probes, have brought us much closer to the holy grail of microbial ecology: observing and understanding who does what, when, where, and next to whom. We need no longer envy plant and animal ecologists who have studied individual organisms for over a century. In fact, we are in a better situation now as it is much easier to manipulate the environment of microbes in the laboratory and mix species together into synthetic communities of defined composition. What is more, the rich data from these technologies facilitate a mechanistic description of the behaviour of microbial individuals not yet feasible for larger organisms.

Complementing the experimental approach with mathematical modelling has, in all areas of science, provided valuable insights that would be difficult to obtain through experimentation alone[13]. A model consolidates our knowledge of the system gathered from a variety of experiments, tests the consequences of our assumptions, and exposes gaps and inconsistencies in our knowledge and understanding. Once validated, a mathematical model can be used to make predictions; for example, of system dynamics under conditions not yet investigated in the laboratory or field, or of system properties that may be difficult to observe directly.

Traditionally, models of microbial systems have been constructed at the population level (FIG. 1). Population-level models (PLMs) are typically "strategic models made to be as simple as possible to reveal general explanations"[14], and have proven to be of immense value, both in microbial ecology and ecology in general. PLMs are good choices when the goal is to find general principles that apply across a broad range of organisms, such as the tendency of predator-prey systems to generate oscillatory population dynamics; or as a first step to studying a particular, complex system in detail. For an environment that is assumed to be spatially homogenous, PLMs can be formulated in terms of difference equations if time is treated as being discrete or in terms of ordinary differential equations (ODEs) if time is treated as being continuous (FIG. 1). When considering spatially structured environments, it is traditional to model the time-evolution of population densities (biomass per unit space) using partial differential equations (PDEs) (FIG. 1). For an executive summary of different modelling approaches see Supplementary Information S1 (text).

Spatially explicit PLMs that simulate the temporal evolution of density distributions, e.g., in biofilms[15], are a valuable resource. However, they make several assumptions that are increasingly at odds with our growing knowledge of microbial systems. Firstly, PLMs ignore the ever more apparent phenotypic heterogeneity between individuals within a population, and the role these differences play in determining system level properties (e.g., population growth rates). At the same time, they can make little direct use of the information contained within data describing the state and behaviour of individual microbes. Secondly, PLMs do not resolve the broad range of interactions between individual organisms and their local biotic or abiotic environment. Thirdly, PLMs do not resolve adaptive processes at the level of individuals. Thus, while PLMs can simulate the dynamics of a system (e.g., changes in the spatial distribution of a microbial population), it is impossible to trace such changes back to the behaviour of individual organisms.

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

of the individual; activities may include the uptake of substrates from the environment, or the synthesis of new biomass; and interactions may include competitive, synergistic or parasitic interactions between individuals within a population or community[21–24]. Such processes may be described as continuous and equation-based (e.g., growth) or discrete and rule-based (e.g., division). In IBMs, the collective action of each individual determines population or community level properties. Feedbacks between the behaviour of individuals and the population as a whole emerge automatically; as does fitness, since it depends on what other individuals do, and how this changes the environment. Furthermore, IBMs can make direct use of single-cell data during their construction, and of both individual and population level data during validation. An IBM therefore mimics the natural system it models (FIG. 1). However, care must be taken to avoid the model becoming too complex to be useful. Precisely because various approaches have their advantages and disadvantages (for a deeper discussion see REF [14]), it is particularly beneficial to use them in conjunction[25–27] (FIG. 1). For example, by comparing ODEs to PDEs of the same system, the effects of local interactions in a spatially structured environment can be revealed. Likewise, comparing PDEs to IBMs reveals the effects of individuality and adaptive behaviour.

In parallel to the technological advances for single-cell observations mentioned at the outset, the technology for developing, running and analysing individual-based models has also progressed significantly, making it easier for scientists to use IBMs to help decipher and understand patterns within experimental data[25]. For most potential users, generic open-source platforms for individual-based modelling will be the best choice. A generic platform is a software tool that allows the user to create models of a range of systems. This is done by selecting the physical processes (e.g., diffusion, convection and mechanical interactions) together with the environment (e.g., liquid culture, agar plate, biofilm flow cell or gut) and the set of species and their biological processes (e.g., growth, cell-cell communication or motility). Being able to select from a range of processes, one can readily identify those processes that affect the behaviour under investigation. Current platforms are approaching this goal, and their further community-based development is the most effective way to implement an ever-wider range of processes. However, specific applications may be better served by software that is more specialized. For recommendations of software, see Supplementary Information S2 (box) and S3 (table).

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

Characteristics of individual-based modelling

The IBM approach has advantages over PLMs when simulating (i) individual heterogeneity (ii), local interactions and (iii) adaptive behaviour[14,18–20] (BOX 1), features increasingly recognized as important in microbial ecology. Mounting evidence from flow cytometry[28] and single cell observations[1,5,6,29,30] demonstrates the existence of individual heterogeneity, even in clonal laboratory cultures[12,31–33]. Local interactions are important because most ecosystems are spatially structured and individuals only interact with neighbours. For example, even in well-mixed marine or fresh water environments, mixing at the microbial scale is limited enough for hot spots of nutrients excreted by phytoplankton to persist long enough to attract and nourish chemotactic bacteria[34,35]. Adaptive behaviour is prevalent in microbes since practically everything they do is in response to their environment, e.g., their genomes typically contain between 50 and 200 genes for two-component systems for sensing and responding to conditions[36]. While rarely considered by modellers, adaptive behaviour is common in nature, of great fitness benefit, and easy to implement in IBMs[14,18–20]. IBMs are flexible, enabling behaviour to be specified in a variety of ways. In the simplest cases, rules like "if oxygen concentration below threshold, switch from aerobic to anaerobic metabolism" can be implemented; such rules can be made stochastic. Kinetic equations with some oxygen inhibition term for the rates of aerobic and anaerobic metabolism would lead to a smoother transition. Incorporating a gene regulatory network submodel for oxygen sensing would replace the phenomenological with a mechanistic description, see FIG. 1. Individual differences, local interactions, and adaptive behaviour may interact in ways that are difficult to foresee without using an IBM to include and exclude these effects systematically (BOX 1).

Microbial individuality and its consequences. Phenotypic differences between individuals have consequences for both the population and the ecosystem. Many factors contribute to phenotypic heterogeneity; these include stochastic gene expression[31], stochastic metabolism and growth[37], epigenetically-regulated modifications such as phase variation[31], phase in the cell cycle or biological clock[31], asymmetric cell division[38], and differences in the environment or stochastic sensing of the environment[39]. Finally yet importantly, the interactions between the above factors in a particular chronological order can affect the current state of an individual; this is simulated in an IBM tracking changes of the local environment and cellular state. For example, some cells may have chanced upon a nutrient-rich patch in the past and therefore downregulated their high-affinity transporters, later making them less acclimated to nutrient-poor environments.

An important and surprising consequence of individual variation is that population averages can be misleading if the functional relationship between an explanatory variable, e.g., substrate concentration, and a response variable, e.g., specific growth rate, is non-linear. Non-linearities are prevalent in biology, e.g. they are found in Monod kinetics, Droop kinetics and most other observed relationships between 'dose' and 'response'. Thus, awareness of the averaging fallacy is important[17]. FIG. 2a shows an example of Droop kinetics, here the non-linear increase of specific growth rate of the marine cyanobacterium *Synechococcus* WH8103 with its intracellular phosphorus content; this is typical for phototrophic microbes[40]. As can be seen from the figure, the growth rate that a population of cells with average P content would have is higher than the average growth rate of individual

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

IBMs are also ideal for incorporating rare events, like mutations or phenotypic switches, since they suddenly change the behaviour of one particular individual (BOX 1). For example, *Pseudomonas aeruginosa* cells in laboratory flow-cell biofilms form 'mushroom' structures under specific conditions because a subpopulation of immotile cells forms mushroom 'stalks' through growth and division, while the motile subpopulation forms 'caps' on top of these stalks[41,42]. These subpopulations are reminiscent of castes in social insects. Contrary to expectations , an IBM based on these experimental observations showed that surface-bound twitching motility cannot explain the formation of mushroom caps[43]; it was later shown experimentally that cap formation requires flagella-driven swarming and chemotaxis[44].

Much of the large phenotypic heterogeneity observed in nature is likely due to environmental differences: the expression of most genes in *Escherichia coli* does not show any bursts[45], selection tends to minimize noise in gene expression for most functions[46,47], and individual cells in a microfluidic device maintaining a strictly constant environment showed reduced variation of specific growth rates[30]. Nevertheless, non-heritable phenotypic differences that are intrinsically generated, and therefore independent of environmental conditions, are important as the basis for bet-hedging and division-of-labour strategies[32,48,49]. A division-of-labour example is the segregation of the population into motile and immotile cells discussed earlier. A typical example of bet-hedging strategies is the differentiation of some cells into non-growing resting stages. Less fit under benign conditions, these have increased chances of survival under stress. For example, some cells of the filamentous cyanobacterium *Anabaena flos-aquae* differentiate into akinetes, which sink to the sediment bed to germinate and re-emerge under favourable conditions. It is clear that these dormant cells help the population survive adverse conditions, but is this trait required to survive annual challenges or more irregular and extreme events? Using an IBM, a heterogeneous population of cells was simulated in a reservoir model that also tracked environmental conditions (light, temperature, nutrients)[50]. When the akinete differentiation trait was "knocked out" in the model, the knockout population did not survive the first winter. As an unexpected insight from the simulations, akinetes provided an additional benefit by taking up nutrients while in the sediment bed[50]. Although experiments with mutants are straightforward in the laboratory, we cannot introduce genetically engineered mutants into the field for ethical reasons so will have to rely on modelling approaches.

Predicting complex spatial patterns

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

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

Predicting complex spatial patterns. A microbe, whether growing in a well-mixed liquid or a spatially structured biofilm, has the same genome and therefore the same potential to sense and react to the environment. Indeed, microbes respond to nutrient-poor conditions within biofilms in much the same way as they would to similar conditions in a chemostat[51,52] or stationary-phase batch culture[53,54]. It should therefore be possible to predict the behaviour of a microbe anywhere if we fully understand how it influences, and is influenced by, its environment[55]. FIG. 2c explains the approach of using a chemostat to measure the dependence of population growth rate on substrate concentration. PDEs or IBMs including such growth kinetics can then calculate the consumption and diffusion of substrate in order to predict the resulting substrate concentration and growth rate gradients in a biofilm. This then tracks how, over time, the changing growth rates of the cells give rise to the emerging spatial structure of the biofilm. This approach has been surprisingly successful given the simplifications involved; see the example of nitrifying granules below. However, if such predictions fail, we can conclude that the simplifications are inappropriate or that other factors may play a role, e.g., individual variation in kinetics of growing cells[33], the presence of persisters[56,57] or changes in gene expression upon attachment[58,59] that affect growth kinetics[60] or induce virulence[59]. Refining the model by including such effects may than better match experimental results. Such refinements would be straightforward to implement in IBMs but difficult to include in PDEs.

An example where feedbacks between substrate concentration gradients, growth rates and biofilm structure can lead to spontaneous formation of clusters from initially small stochastic perturbations is shown in Fig. 2d. Once spontaneously arisen, clusters of slowly growing cells that utilize resources more economically grow faster than clusters of cells that grow faster at any given substrate concentration[61]. This counterintuitive result is due to the locally higher substrate concentration in the clusters of economical cells. Thus, complex spatial patterns can emerge from stochastic positioning of cells.

Predicting interactions in mixed cultures. Microbial species are often studied in isolation, yet in their natural environment they interact with many other species in a variety of ways; most interactions are indirect, mediated by diffusible compounds such as metabolites, autoinducers or toxins. Using modelling to predict interactions is extremely useful, as the number of potential interactions increases exponentially with the number of species in a habitat, e.g., at 5 or 6 species there are already 31 or 63 potential interactions[17], making it difficult to study all possible interactions experimentally.

In mixed species biofilms, additional mechanisms and phenomena of pattern formation come into play. Strains of *Saccharomyces cerevisiae* can be engineered to depend on each other for growth by producing a metabolite that the other requires[62]. Engineering another strain requiring one metabolite without reciprocating adds a cheater to the cooperating pair of obligate cross-feeding strains. When randomly placed cells start to grow into colonies, mutually cross-feeding colonies that happen to be close by will grow well towards each other, forming

large areas of contact. In contrast, the cheater strain becomes squeezed out, as it does not facilitate the growth of the strain that it depends on (FIG. 3a)[62]. The general insight from this and other IBMs is that spatial patterns in microbial communities are to some extent self-organized as they emerge from different types of metabolic interactions: cooperation consisting in restraint from competition[61], cross-feeding[63,64], interspecies hydrogen transfer[65,66] and the combination of particular trophic interactions with motility of cells on hydrated rough surfaces[67].

The most rigorous demonstration to date of the ability of IBMs to predict solute gradients and spatial distributions of interacting metabolic/functional groups from kinetics measured in batch and chemostat cultures has been carried out for nitrifying biofilms[68] and granules[69]. These are assemblages of a few types of autotrophic ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB), plus a few types of heterotrophic denitrifiers, forming a food-web that is utilized in wastewater treatment to convert ammonia to dinitrogen. Matsumoto and colleagues[69] combined microelectrodes to measure O_2, NH_4^+, NO_2^- and NO_3^- profiles and confocal microscopy to determine the distribution of functional groups through these nitrifying granules (FIG. 3b). Considering that the model was not fitted to the data makes the match of microelectrode-measured with IBM-calculated profiles surprisingly close; the match of observed biomass distributions with predictions was quite good, especially for the better-known autotrophic groups (FIG. 3b). This provided insight into the self-organization of radial stratification in these granules: oxygen becomes depleted with depth and the aerobic AOB dominate the surface layer because they are first in the food chain, while the also aerobic NOB reside underneath as they are dependent on the AOB's activity and therefore less competitive for oxygen. Reliable predictions are also of great value, as they can be used by engineers to estimate and optimize whole reactor performance.

Predicting evolution

IBMs, being based on specifying traits of individuals, can easily incorporate heritable mutations of these traits. Combined with simulating birth, death and competition, natural selection of the mutants is simply an emergent process in IBMs. FIG. 4a shows such an example where the phenotypic traits are digitally encoded, so a change in the digital genome is mapped to a change in phenotype, e.g., a change in the rate of a particular pathway[70]. In spatially structured environments, migration of the evolving microbes can generate spatial patterns and feedbacks between the emerging spatial structure and natural selection can arise (FIG. 4a). This technique has been applied to studies of cyanobacteria dynamics in the ocean[71] where a complex set of interacting biotic and abiotic forces shape the physiology, ecology, and evolutionary dynamics of these microorganisms. In this work, an evolutionary IBM was coupled to a hydrodynamic model that resolves vertical gradients in light intensity, temperature, and the amount of dissolved inorganic and organic nutrients, and how these change in time, e.g., due to changes in surface wind speeds, irradiance, or the uptake of nutrients by the cyanobacteria. The model predicted spatial and temporal trends in the physiology, size and abundance of *Synechococcus* and *Prochlorococcus* ecotypes. During the summer months when the water column is well stratified, small, high-light adapted cyanobacteria dominated in well-lit but nutrient-starved surface waters, and larger, low-light adapted cells dominated at depth (Fig. 4b). These predicted trends were then found to be consistent with observations at the seasonally stratified Bermuda Atlantic Time Series site in the North Atlantic Ocean[71]. The IBM helped to identify a three-way

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

Another IBM was used to test the hypothesis that dispersal limitation of ocean bacteria is sufficient for the formation of biogeographical patterns[72]. This IBM simulated ~100,000 individuals within a global ocean circulation model. The cells grew and divided and their 1 Mbp genomes were subject to neutral evolution, i.e., the mutations were assumed to have no fitness effect. The model showed that biogeographic provinces dominated by different species could be produced from ocean currents and dispersal limitation alone, without any environmental selection[72] (FIG. 4c). IBMs can simulate discrete individuals with their own genome sequence and account for dispersal limitation.

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

Beyond cell and population scales

Microbial IBMs are multiscale models by nature, i.e., they link cell and population scales; this is rather useful in itself, but their multiscale nature can be further expanded to include lower and higher levels of organization. Including increasingly more intracellular states and behaviours leads to more mechanistic models of the behaviour of individual cells replacing the empirical descriptions traditionally used. This has already been successful with IBMs incorporating signal transduction mechanisms in chemotaxis[74] and quorum sensing[26]. IBMs also advance the new field of synthetic biology because they allow simulation and optimization of how synthetic organisms

interact with each other and the environment before actually constructing them[26,75,76] – the ultimate rational design. Apart from providing stronger mechanistic foundations for individual behaviour and individuality, integration of intracellular mechanisms also enables exploitation of the rapidly growing metagenomic data. After reconstructing genomes from shotgun metagenomics[77] one can reconstruct whole-genome metabolic models from the stoichiometry of all enzyme reactions coded for by the genome. By assuming that the fluxes through reactions are constrained and optimized such that growth of the cell is maximal, one can predict metabolic fluxes and growth without having to know the kinetics of any enzymes[78,79]. Such constraint-based reconstruction and analysis (COBRA) models have already been scaled up to the population level[33,80]. Likewise, incorporating COBRA models into IBMs (FIG. 1) has great potential for the future and has already been successful using a model based on a spatial grid, an approach similar to IBMs but with a coarser resolution[63].

In the other direction, microbial IBMs can be expanded to full-scale ecological or biogeochemical systems, which we have already illustrated with examples (FIG. 4a-c). This will improve the predictive power of ecosystem models. Moreover, subcellular dynamics could be included in full-scale ecological IBMs, which will facilitate the use of omics data sets that measure the composition, acclimation state, activity or genetic make-up of individuals and thus help to bridge the growing gap between omics data and biogeochemical models[81].

Conclusions and future directions

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

The key general insight from IBMs is that population dynamics and structure emerge from the interactions of individuals with each other and the environment. This extends to the community and ecosystem level. Biofilm spatial structure and other self-organized patterns are a prime example of emergence and because of this, complex macroscopic patterns can be predicted from simple microscopic mechanisms. In evolutionary IBMs, diversity and spatial distribution of species can emerge; the inclusive fitness of individuals is also emergent. Since IBMs link cell and population scales they demonstrate how individual heterogeneity affects population and ecosystem function or how new phenomena can arise at the population level, e.g., in populations of signal transducing cells or hosts co-evolving with phages. Moreover, linking individual and population scales can identify mechanisms that can or cannot explain observed population behaviour. These insights could not have been obtained with classical PLMs as they do not account for individual heterogeneity, local interactions, or adaptive behaviour. Since IBMs do, they can advance our understanding and ability to predict microbial systems beyond what can be 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 optimal[19]; 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 (µIBE).

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 cioreanu 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.

Display items

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

Characteristics of mathematical models generally. The purpose of a model is to simplify reality. Since "a mathematical model is a logical machine for converting assumptions into conclusions"[13], it enforces complete and unambiguous specification of assumptions, which is essential for rigorous testing of hypotheses. An example of a typical simplifying assumption in IBMs is that cells divide instantaneously when they reach a threshold volume; for studying e.g., lag phase a more mechanistic cell division model would be more appropriate[82]

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

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

Good modelling practice in IBMs. Adopting the 'ODD' (Overview, Design concepts, and Details) protocol as standard for systematic, complete description of IBMs has already facilitated comparison and peer review of IBMs[83]. ODD is similar in purpose to MIRIAM[84] (Minimum Information Requested In the Annotation of biochemical Models). Better-developed standards for model exchange and description have been highly beneficial for systems biology. For example, the Systems Biology Ontology (SBO[85]) provides an unambiguous vocabulary for model description and the Systems Biology Markup Language (SBML[86]) enables the exchange of completely and unequivocally specified models.

The most successful models are *structurally realistic*, i.e., the entities and processes in the model correspond to those in the real world (e.g., a chemotaxis model where cells carry out runs and tumbles and responding to changes in attractant concentration[74] rather than cells taking the steepest ascent towards the concentration peak). Demonstrate robustness of model predictions: parameter sensitivity analysis to identify important parameters (e.g., an IBM of plasmid transfer evaluating how strongly the rate of plasmid transfer changes when pilus reach, lag times between transfers, and other conjugation parameters are varied[87]); structural sensitivity analysis to identify important processes (e.g., an IBM of aging systematically including/excluding processes such as segregation, repair and/or toxicity of damage and growth of the cells to test whether they interact and/or change simulation results qualitatively or quantitatively[23]).

Potential pitfalls of IBMs. Avoid using global environmental or population states in deciding the activities of an individual, since no individual has global knowledge (e.g., assuming that growth of cells depends on population density as in logistic growth[88]). Ignoring processes that are in fact important. Probably the most common mistake

is that of imposing behaviour of individuals that one wishes to study as emergent (e.g., using a biofilm model to predict biofilm structure formation but assuming that cells inside the biofilm cannot divide which will affect the structure[89]).

Figure 1 | **Simplified overview of approaches useful for modelling communities and/or single cells.** ODEs and PDEs describe rates of change of populations (X) and/or resources (R) directly, i.e. the level of individuals is absent. Comparing the non-spatial ODEs with the spatially explicit PDEs illuminates the effect of spatial structure. IBMs describe activities of individuals. Changes on the population level are not directly described because they emerge from individual behaviour. Hence, IBMs can make use of data on both levels: individual-level data as input and population-level data to compare with simulation output. Comparing PDEs with IBMs elucidates the effect of individuality and adaptive behaviour. Combining all three approaches is therefore best practice. Most IBMs to date include only simple kinetic models of growth and rules for cell division, but since IBMs treat individuals as discrete agents, they enable incorporation of intracellular dynamics as modelled in systems biology – bridging the scales from intracellular reactions to ecosystem function. Only two major types of models for intracellular dynamics are shown: dynamic kinetic models use full kinetic equations only known for a select number of enzyme reactions while flux balance models only need the stoichiometry of the reactions and constraints, which enables a genome-wide prediction of metabolic fluxes in steady state[79]. Image of mouse gut mucosa courtesy of Kristen Earle of Justin Sonnenburg's lab. Dynamic kinetic model modified with permission from REF. [90].

Figure 2 | **Using IBMs to predict population dynamics, substrate and growth rate gradients and the effect of spatial structure.** **a** | Observed cellular phosphate content (quota q) in the marine cyanobacterium *Synechococcus* WH8103 and photosynthesis rate (μ) calculated using the non-linear Droop kinetics shown as a black line[40]. Calculating the rate for each individual based on its quota (full circles) and averaging over individuals (red line: ave[$\mu(q)$]) gives a lower population productivity than averaging the quotas and calculating the rate based on that average quota (blue line: μ(ave[q])). **b** | Predicting unobserved concentration gradients. From a confocal biofilm image (courtesy of Søren Molin) we can estimate biomass distribution for PDE or cell positions for IBM and calculate the concentration field based on growth kinetics and laws of diffusion. This may then be verified by microelectrode transects where possible. **c** | Growth kinetics parameterised from chemostat experiments can be used in PDEs or IBMs to predict biofilm structure, growth rate and concentration gradients. IBMs could include adaptations in kinetics or heterogeneity in the population, e.g. persisters. **d** | Predicting feedbacks between emerging biofilm structure and metabolic interactions. Once clusters of red cells that consume resources economically have formed by chance, they grow faster than clusters of the blue fast growing cells because their economy sustains locally higher substrate concentrations[61].

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

excluded by spatial self-organisation of the co-operators; this is shown both experimentally, where strains are fluorescently tagged, and in an IBM of the system[62]. Images courtesy of Babak Momeni. **b** | Verifying IBM predictions for a nitrifying food-web in a lab scale aerobic upflow fluidized bed reactor with microelectrodes and microscopy. Based on standard literature parameters for nitrifiers rather than fitting the model to data, an IBM can predict the measured solute profiles and biomass distributions of the autotrophic AOB and NOB, and of EPS, quite well (simulated profiles were averaged over concentric layers). Predicted distributions of the less understood groups of heterotrophic bacteria (for this figure lumped together as 'Het') are roughly correct. Colour coding of solutes and biomass indicated next to the respective graphs. Modified with permission from REF. [69].

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

References

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

33. 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).
34. Mitchell, J. G., Okubo, A. & Fuhrman, J. A. Microzones Surrounding Phytoplankton Form the Basis for A Stratified Marine Microbial Ecosystem. *Nature* **316,** 58–59 (1985).
35. Blackburn, N., Fenchel, T. & Mitchell, J. Microscale nutrient patches in planktonic habitats shown by chemotactic bacteria. *Science* **282,** 2254–2256 (1998).
36. Capra, E. J. & Laub, M. T. Evolution of Two-Component Signal Transduction Systems. *Annu. Rev. Microbiol.* **66,** 325–347 (2012).
37. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* (2014). doi:10.1038/nature13582
38. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43,** 95–100 (2011).
39. Korobkova, E., Emonet, T., Vilar, J. M., Shimizu, T. S. & Cluzel, P. From molecular noise to behavioural variability in a single bacterium. *Nature* **428,** 574–578 (2004).
40. 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).
41. Klausen, M. *et al.* Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Mol. Microbiol.* **48,** 1511–1524 (2003).
42. Klausen, M., Aaes-Jørgensen, A., Molin, S. & Tolker-Nielsen, T. Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Mol. Microbiol.* **50,** 61–68 (2003).
43. Picioreanu, C. *et al.* Microbial motility involvement in biofilm structure formation – a 3D modelling study. *Water Sci. Technol.* **55,** 337–343 (2007).
44. Barken, K. B. *et al.* Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms. *Environ. Microbiol.* **10,** 2331–2343 (2008).
45. Silander, O. K. *et al.* A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. *PLoS Genet* **8,** e1002443 (2012).
46. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4,** n/a–n/a (2008).
47. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl. Acad. Sci. U. S. A.* **108,** E67–E76 (2011).
48. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309,** 2075 –2078 (2005).
49. Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *J. Theor. Biol.* **234,** 227–253 (2005).
50. Hellweger, F. L., Kravchuk, E. S., Novotny, V. & Gladyshev, M. I. Agent-based modeling of the complex life cycle of a cyanobacterium (Anabaena) in a shallow reservoir. *Limnol. Oceanogr.* **53,** 1227–1241 (2008).
51. Whiteley, M. *et al.* Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413,** 860–864 (2001).
52. Sauer, K., Camper, A. K., Ehrlich, G. D., Costerton, J. W. & Davies, D. G. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *J. Bacteriol.* **184,** 1140–1154 (2002).
53. Hentzer, M., Eberl, L. & Givskov, M. Transcriptome analysis of *Pseudomonas aeruginosa* biofilm development: anaerobic respiration and iron limitation. *Biofilms* **2,** 37–61 (2005).
54. Dötsch, A. *et al.* The Pseudomonas aeruginosa Transcriptome in Planktonic Cultures and Static Biofilms Using RNA Sequencing. *PLoS ONE* **7,** e31092 (2012).
55. 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).
56. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305,** 1622 –1625 (2004).
57. Maisonneuve, E. & Gerdes, K. Molecular mechanisms underlying bacterial persisters. *Cell* **157,** 539–548 (2014).
58. O'Connor, J. R., Kuwada, N. J., Huangyutitham, V., Wiggins, P. A. & Harwood, C. S. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol. Microbiol.* **86,** 720–729 (2012).
59. Siryaporn, A., Kuchma, S. L., O'Toole, G. A. & Gitai, Z. Surface attachment induces *Pseudomonas aeruginosa* virulence. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 16860–16865 (2014).
60. Rice, A. R., Hamilton, M. A. & Camper, A. K. Apparent surface associated lag time in growth of primary biofilm cells. *Microb. Ecol.* **40,** 8–15 (2000).
61. Kreft, J.-U. Biofilms promote altruism. *Microbiology* **150,** 2751 –2760 (2004).
62. Momeni, B., Waite, A. J. & Shou, W. Spatial self-organization favors heterotypic cooperation over cheating. *eLife* **2,** e00960 (2013).
63. Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* **7,** 1104–1115 (2014).

64. Estrela, S. & Brown, S. P. Metabolic and demographic feedbacks shape the emergent spatial structure and function of microbial communities. *PLoS Comput. Biol.* **9,** e1003398 (2013).
65. Batstone, D. J., Picioreanu, C. & van Loosdrecht, M. C. M. Multidimensional modelling to investigate interspecies hydrogen transfer in anaerobic biofilms. *Water Res.* **40,** 3099–3108 (2006).
66. Momeni, B., Brileya, K. A., Fields, M. W. & Shou, W. Strong inter-population cooperation leads to partner intermixing in microbial communities. *eLife* **2,** (2013).
67. Wang, G. & Or, D. Trophic interactions induce spatial self-organization of microbial consortia on rough surfaces. *Sci. Rep.* **4,** (2014).
68. Matsumoto, S. *et al.* Experimental and simulation analysis of community structure of nitrifying bacteria in a membrane-aerated biofilm. *Water Sci. Technol.* **55,** 283–290 (2007).
69. Matsumoto, S. *et al.* Microbial community structure in autotrophic nitrifying granules characterized by experimental and simulation analyses. *Environ. Microbiol.* **12,** 192–206 (2010).
70. Clark, J. R., Daines, S. J., Lenton, T. M., Watson, A. J. & Williams, H. T. P. Individual-based modelling of adaptation in marine microbial populations using genetically defined physiological parameters. *Ecol. Model.* **222,** 3823–3837 (2011).
71. Clark, J. R., Lenton, T. M., Williams, H. T. P. & Daines, S. J. Environmental selection and resource allocation determine spatial patterns in picophytoplankton cell size. *Limnol. Oceanogr.* **58,** 1008–1022 (2013).
72. Hellweger, F. L., van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* **345,** 1346–1349 (2014).
73. Koonin, E. V. & Wolf, Y. I. Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus–host coevolution. *Mol. Biosyst.* **11,** 20–27 (2014).
74. 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).
75. Rudge, T. J., Steiner, P. J., Phillips, A. & Haseloff, J. Computational modeling of synthetic microbial biofilms. *Acs Synth. Biol.* **1,** 345–352 (2012).
76. Stevens, J. T. & Myers, C. J. Dynamic Modeling of Cellular Populations within iBioSim. *ACS Synth. Biol.* **2,** 223–229 (2013).
77. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32,** 822–828 (2014).
78. 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).
79. 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).
80. Klitgord, N. & Segrè, D. Environments that induce synthetic microbial ecosystems. *PLoS Comput. Biol.* **6,** e1001002 (2010).
81. Fuhrman, J., Follows, M. & Forde, S. Applying '-omics' Data in Marine Microbial Oceanography. *Eos Trans. Am. Geophys. Union* **94,** 241–241 (2013).
82. Dens, E. J., Bernaerts, K., Standaert, A. R., Kreft, J.-U. & Van Impe, J. F. Cell division theory and individual-based modeling of microbial lag: Part II. Modeling lag phenomena induced by temperature shifts. *Int. J. Food Microbiol.* **101,** 319–332 (2005).
83. Grimm, V. *et al.* The ODD protocol: a review and first update. *Ecol. Model.* **221,** 2760–2768 (2010).
84. Novere, N. L. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotech* **23,** 1509–1515 (2005).
85. Courtot, M. *et al.* Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7,** (2011).
86. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19,** 524–531 (2003).
87. Merkey, B. V., Lardon, L. A., Seoane, J. M., Kreft, J.-U. & Smets, B. F. Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. *Environ. Microbiol.* **13,** 2435–2452 (2011).
88. Gharasoo, M., Centler, F., Fetzer, I. & Thullner, M. How the chemotactic characteristics of bacteria can determine their population patterns. *Soil Biol. Biochem.* **69,** 346–358 (2014).
89. Wimpenny, J. W. T. & Colasanti, R. A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. *FEMS Microbiol. Ecol.* **22,** 1–16 (1997).
90. Hellweger, F. L. Resonating circadian clocks enhance fitness in cyanobacteria *in silico*. *Ecol. Model.* **221,** 1620–1629 (2010).

FURTHER INFORMATION

Software pages:

CellModeller: www.cellmodeller.org

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

CompuCell3D: www.compucell3d.org

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

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

NetLogo: ccl.northwestern.edu/netlogo/

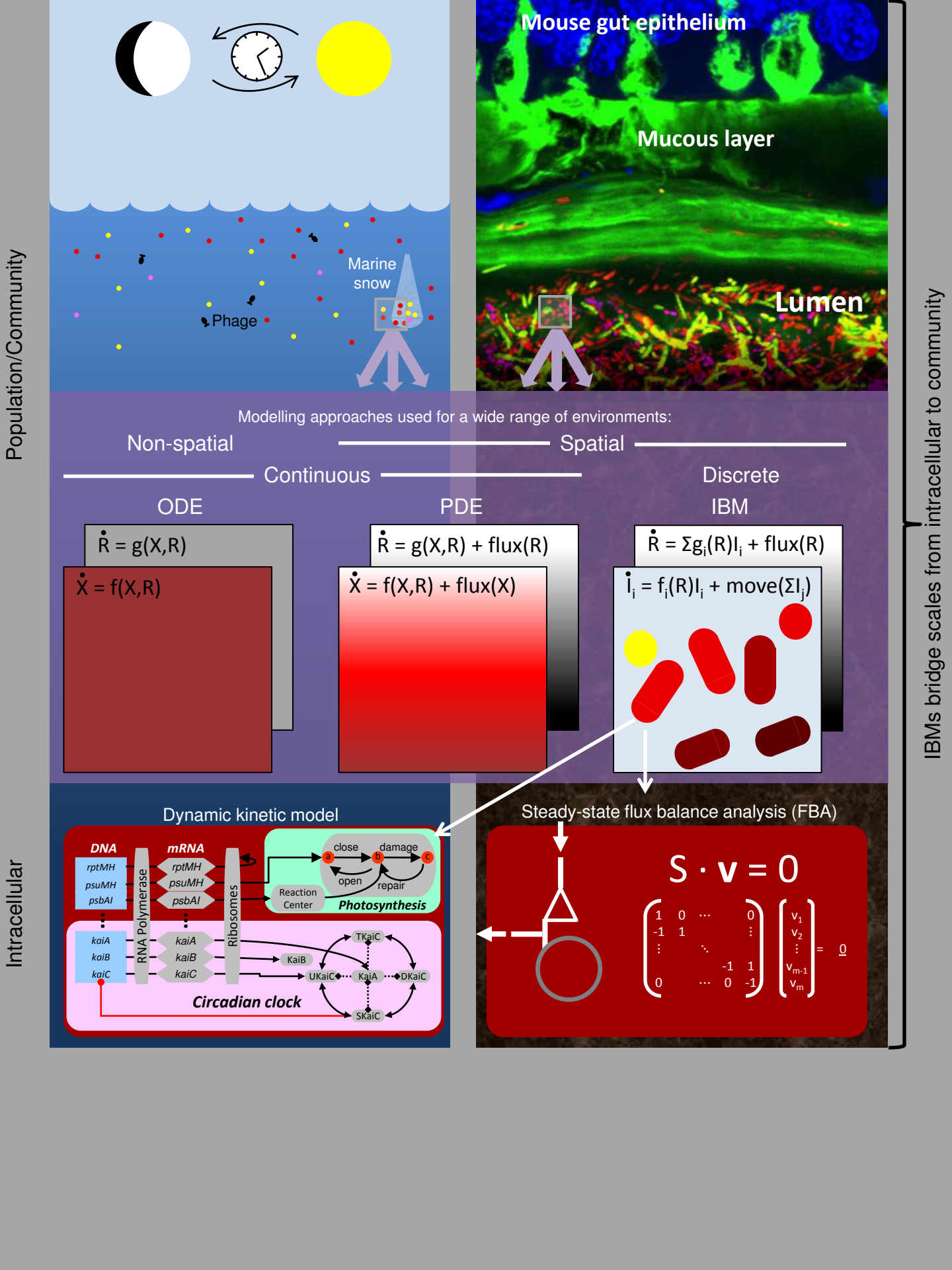
Laboratory home pages:

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

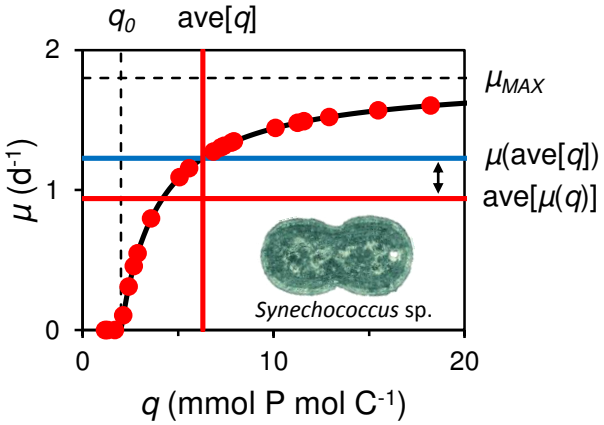
http://www.biosciences-labs.bham.ac.uk/kreftlab/

Ferdi L. Hellweger's home page:

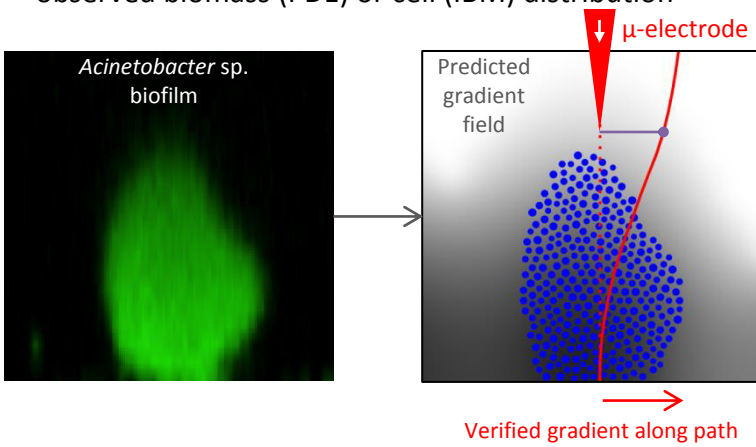
www.systemsbioecology.org


Population/Community
Intracellular
IBMs bridge scales from intracellular to community
Mouse gut epithelium
Mucous layer
Lumen
Marine snow
Phage
Modelling approaches used for a wide range of environments:
Non-spatial
Spatial
Continuous
Discrete
ODE
PDE
IBM
Ṙ = g(X,R)
Ẋ = f(X,R)
Ṙ = g(X,R) + flux(R)
Ẋ = f(X,R) + flux(X)
Ṙ = Σg_i(R)I_i + flux(R)
İ_i = f_i(R)I_i + move(ΣI_j)
Dynamic kinetic model
DNA
mRNA
rptMH
psuMH
psbAI
kaiA
kaiB
kaiC
RNA Polymerase
Ribosomes
close
damage
open
repair
Reaction Center
Photosynthesis
KaiB
TKaiC
UKaiC
KaiA
DKaiC
SKaiC
Circadian clock
Steady-state flux balance analysis (FBA)
S · v = 0

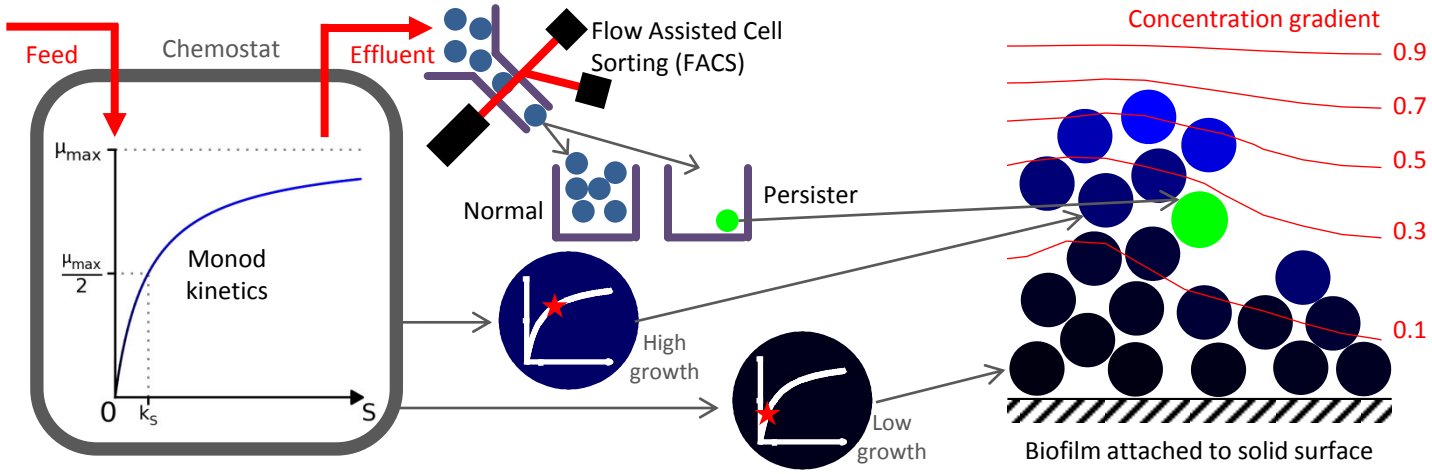
a Non-linear Droop kinetics and the averaging fallacy



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)

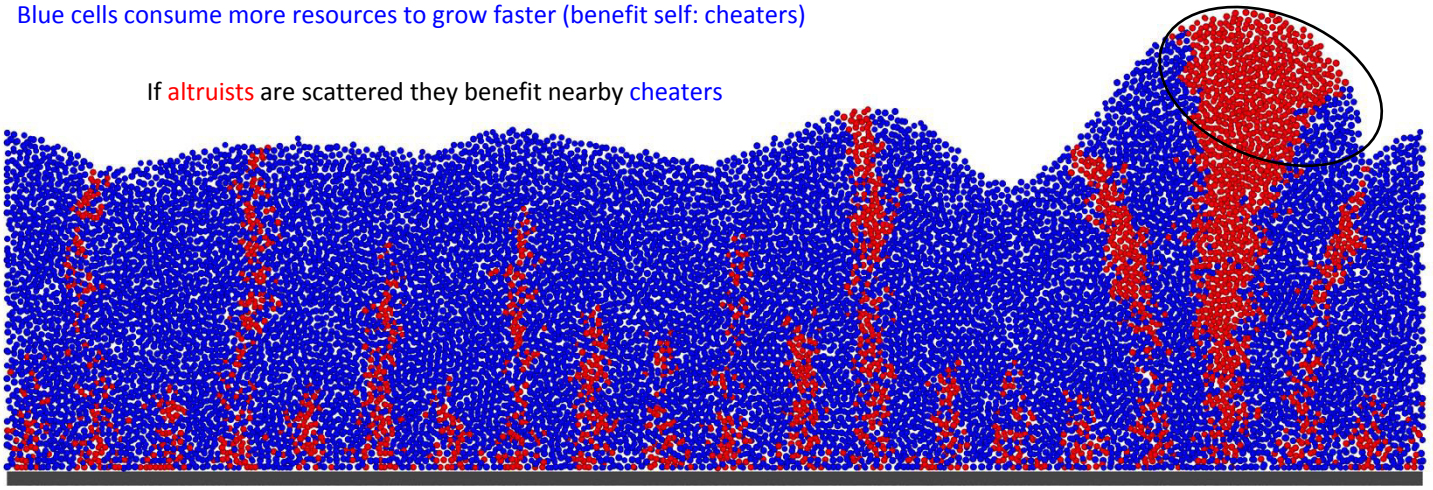


d Predicting emerging interactions and structure feedbacks

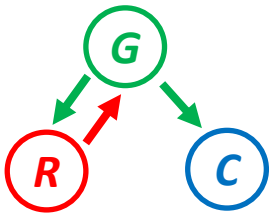
Red cells economically use shared resources (benefit neighbours: altruists)
Blue cells consume more resources to grow faster (benefit self: cheaters)

Spontaneously emerging clusters of altruists benefit cluster neighbours and outgrow cheaters

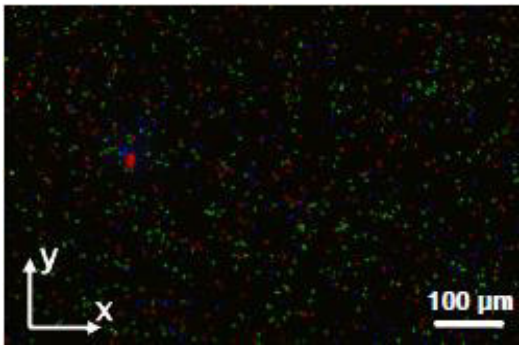
If altruists are scattered they benefit nearby cheaters



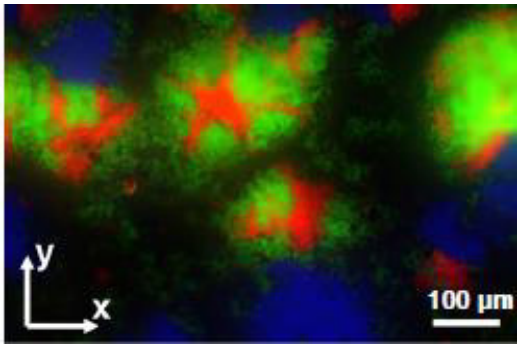
a Locally cross-feeding G and R strains benefit each other such that cheater C becomes spatially excluded from the cross-feeding over time



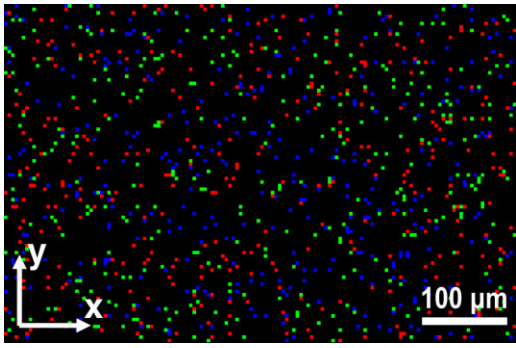
Experiment with engineered yeast strains



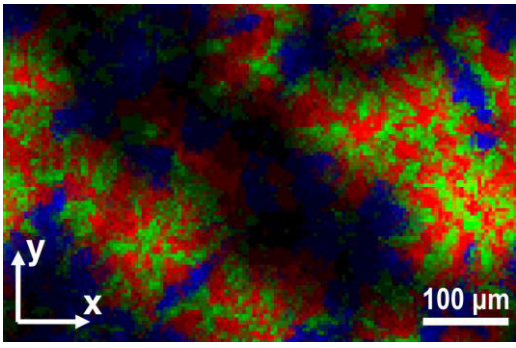
Growth



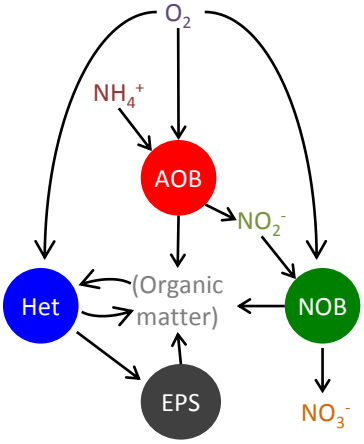
Simulation

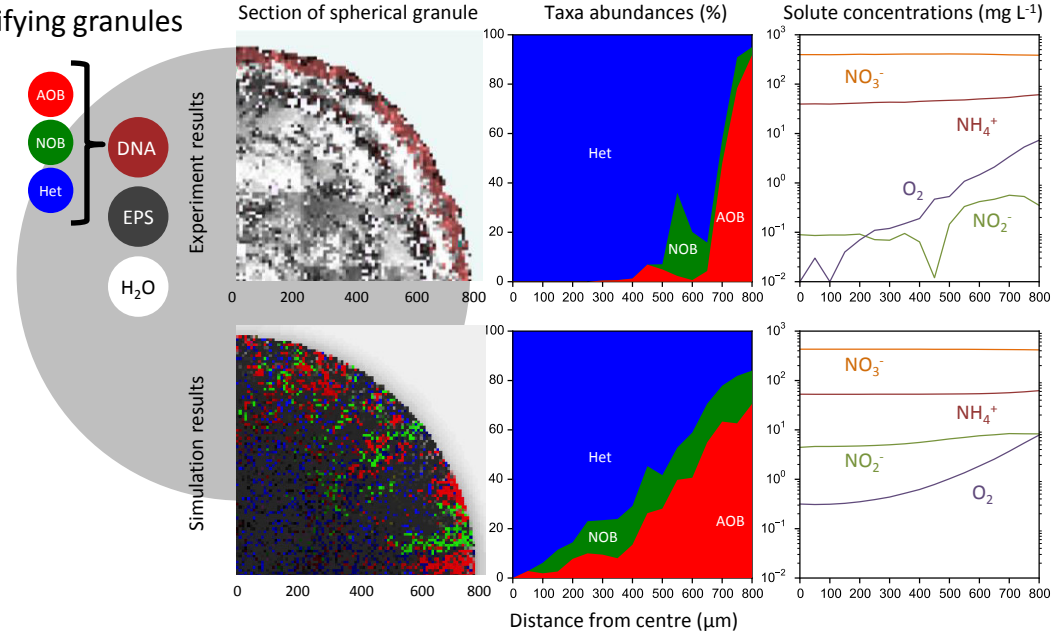


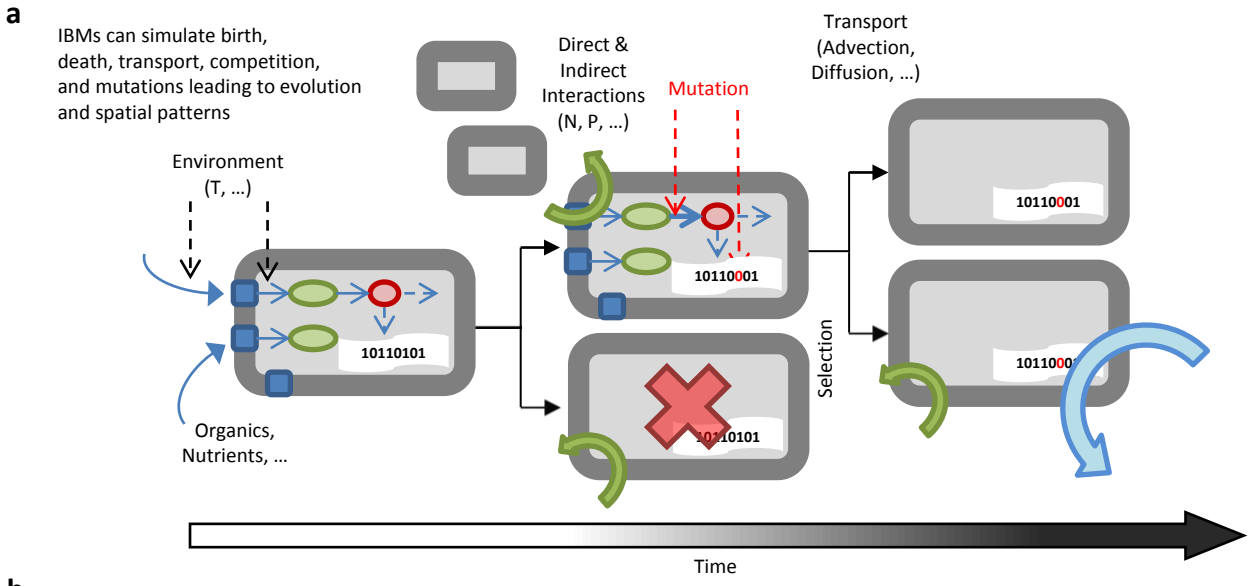
Growth

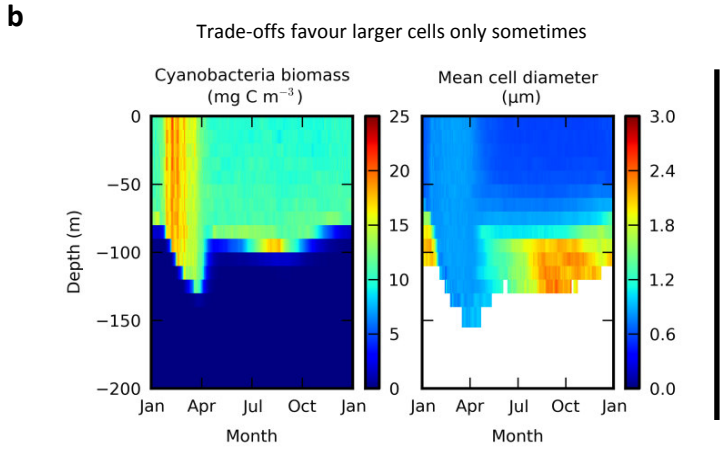


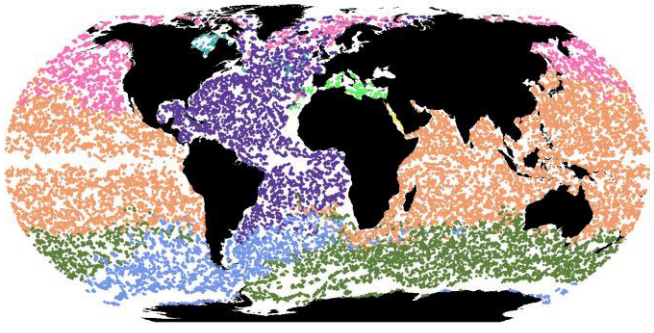
b Validation of IBM-predicted solute and biomass profiles in nitrifying granules

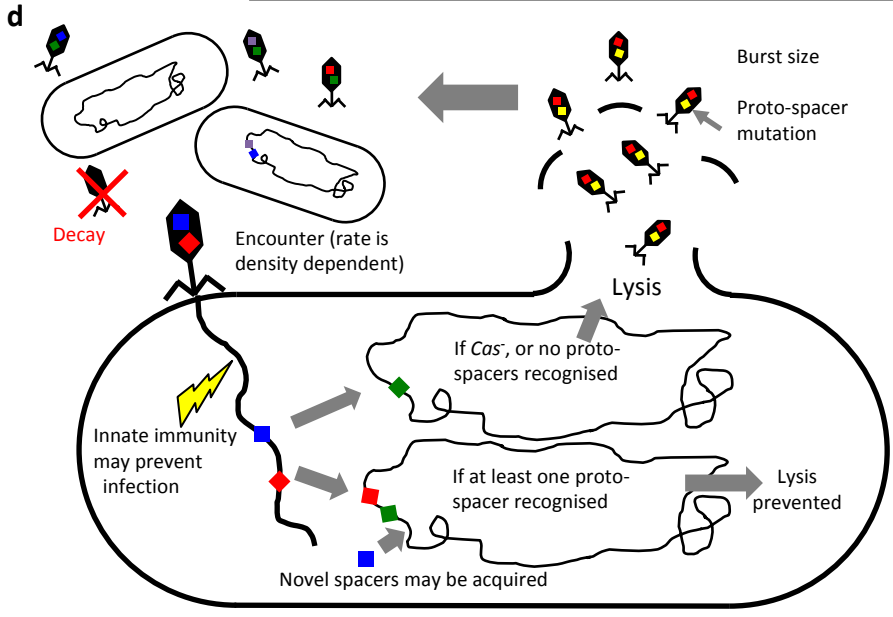




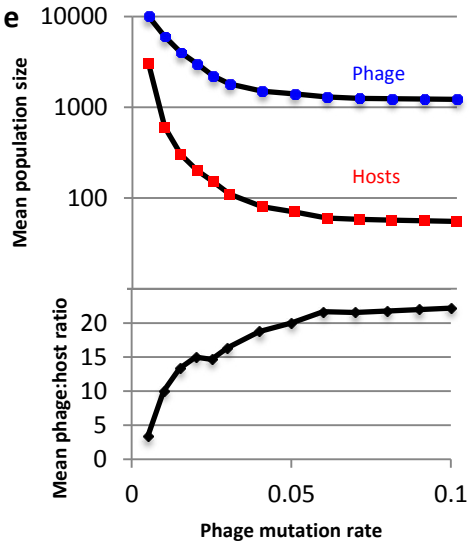






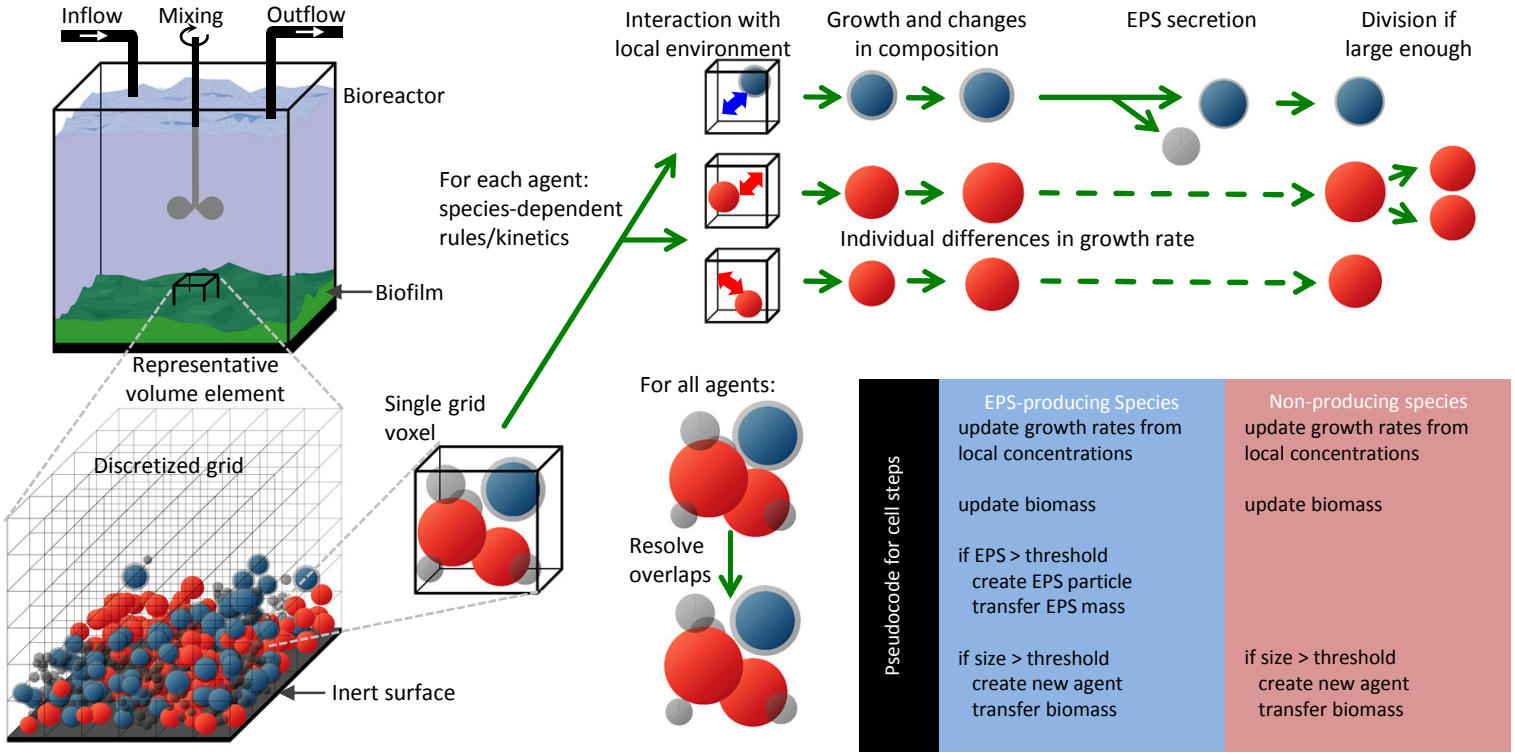


Also modelled:
- Loss of CRISPR-Cas cassette or individual spacers during host division
- Horizontal gene transfer of CRISPR-Cas cassette between hosts
- Growth cost for *Cas*⁺ hosts



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


Inflow
Mixing
Outflow
Bioreactor
Biofilm
Representative volume element
Discretized grid
Inert surface
Single grid voxel
For each agent: species-dependent rules/kinetics
Interaction with local environment
Growth and changes in composition
EPS secretion
Division if large enough
Individual differences in growth rate
For all agents:
Resolve overlaps
Pseudocode for cell steps
EPS-producing Species
update growth rates from local concentrations
update biomass
if EPS > threshold
create EPS particle
transfer EPS mass
if size > threshold
create new agent
transfer biomass
Non-producing species
update growth rates from local concentrations
update biomass
if size > threshold
create new agent
transfer biomass

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 species[3]. 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' models[4]. 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 ways[2]. 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 continuum[2]. Unavoidably, the more complex a PLM becomes, the more it loses its advantage over a corresponding IBM of being simpler and more tractable mathematically[4].

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 environment[6–10]. 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 modelling[11].

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 artefacts[14]. 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 rules[15–17]. To avoid artefacts, lattice elements should be updated in random order[18].

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 volumes[19–21]. 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 equations[6–10]. For example, cell division is commonly based on a cell size threshold[22].

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 organizations[10].

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 nutrient[25]. 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 content[25]. 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 input [26].

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 environment[27]. 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 affinity[27]) would differ between individuals. Such individual differences could be due to variation in expression of genes for uptake and metabolic enzymes between cells[28]. 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 chemostats[33]. 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 time[34] 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 models[35]. 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 known[36].

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 debated[39].

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., Hoye, 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. Evol. 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., Picioreanu, C., Virdis, 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., Fritzsch, 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).

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 FLAME[3] are the most generic, and often used for non-microbial IBMs. CellModeller[4,5], CHASTE[6] and CompuCell3D[7] were designed for modelling tissues but can easily model biofilms as they are similar. CompuCell3D and iDynoMiCS[8] 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 Repast[9] 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 microbes[10], Virtual Ecology Workbench (VEW)/Planktonica for plankton models[11], INDISIM for carbon and nitrogen cycling in soil[12] or lag phase in liquid media[13], Framework for biofilm models[14], AgentCell for chemotaxis signalling[15], COMETS for metabolite exchange[16] and BSim for gene regulation[17].

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.

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.* iDynoMiCS: 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. Gorochowski, 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.idynomics.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 signalling; 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), *Dictyostelium* fruiting body development	Skin, Signalling cascade, Neo-angiogenesis in cancer, *E. coli* 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