

# BacArena: Simulation of Interactions in Microbial Communities using Genome-wide Metabolic Reconstructions

Johannes Zimmermann

Eugen Bauer

October 20, 2013

## **Abstract**

Microbial communities are essential for global ecosystems and human health.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Microbial metabolic ecology . . . . .	3
1.2	Constrained based modeling . . . . .	3
1.3	Agent based modeling . . . . .	4
1.4	Aim of the project . . . . .	4
<b>2</b>	<b>Methods</b>	<b>5</b>
2.1	Model overview . . . . .	5
2.2	Representation . . . . .	6
2.2.1	Environment & Grid . . . . .	6
2.2.2	Bacteria . . . . .	6
2.2.3	Substrate . . . . .	6
2.3	Interactions as rules . . . . .	7
2.3.1	Movement . . . . .	7
2.3.2	Diffusion . . . . .	7
2.3.3	Flux balance analysis . . . . .	7
2.3.4	Growth . . . . .	8
<b>3</b>	<b>Results</b>	<b>9</b>
3.1	Movement and Diffusion . . . . .	9
3.2	Flux Balance Analysis . . . . .	9
3.3	Growth models . . . . .	9
3.3.1	<i>Escherichia coli</i> core . . . . .	10
3.3.2	<i>Escherichia coli</i> . . . . .	10
3.3.3	<i>Methanosarcina barkeri</i> . . . . .	10
3.3.4	<i>Clostridium beijerinckii</i> . . . . .	10
3.4	Mixed communities . . . . .	10
3.4.1	<i>Escherichia coli</i> & <i>Methanosarcina barkeri</i> . . . . .	10
3.4.2	<i>Escherichia coli</i> & <i>Clostridium beijerinckii</i> . . . . .	10
3.4.3	<i>Clostridium beijerinckii</i> & <i>Methanosarcina barkeri</i> . . . . .	10
<b>4</b>	<b>Discussion</b>	<b>10</b>
4.1	Population models of single organisms . . . . .	10
4.2	Synthropy in mixed communities . . . . .	10
4.3	From heterogeneity to biofilms . . . . .	10
4.4	Conclusions & outlook . . . . .	10

# 1 Introduction

## 1.1 Microbial metabolic ecology

Microbial communities pose important roles in the cycle of matter (\*) as well as human health and disease (\*). The comprehensive understanding of the interaction between microbes is thus a major goal in microbial ecology and systems biology.

Microbial consortia often consist of a diverse composition, which degrade complex compounds in multiple steps by different microbes (\*). This division of labour can be realized by the aggregation to biofilms in which multiple layers of microbes are associated with each other and interact by exchanging various metabolites (\*). An example for the cross-feeding between microbes is the interspecies hydrogen transfer. Here, methanogenic archaea are associated with bacteria or protists, which produce hydrogen after anaerobic degradation (\*). The hydrogen can be taken up as an essential substrate by the methanogens, which profits the producer by the removal of the products and thus the thermodynamic limitation (\*). Therefore both partners benefit from this interaction. A close spatial aggregation of the partners can further optimize their individual benefits, since the hydrogen can be exchanged faster. Recent advances in systems biology made it possible to study the metabolic interactions of multiple species on the systems level (\*). In particular constrained based modeling can be applied to model interspecies metabolic exchanges (\*).

## 1.2 Constrained based modeling

Constrained based modeling is a successfully applied method in systems biology ([1], [6] p. 353), where the metabolism of single species is considered as a network of biochemical reaction. The reaction network itself can be represented more formally with differential equations using mass action kinetics. Because of the high numbers of reactions and metabolites the resulting system of equation and the solution space is high dimensional. Therefore, linear algebra is used for a simplified description:

$$\frac{dx}{dt} = S \cdot v$$

where  $x \in \mathbb{R}^m$  is a vector consisting of concentrations of all  $m$  metabolites,  $S \in \mathbb{Z}^{m \times r}$  is the stoichiometric matrix, which includes the net consumption/production of all  $r$  biochemical reactions and  $v \in \mathbb{R}^r$  is the flux vector which contains in general nonlinear kinetic relationships.

Now several *constraints* could be applied to solve the problem more easily. The most prominent constraint is the equilibrium or steady state  $dx/dt \stackrel{!}{=} 0$ . It is a reasonable assumption for a metabolic model, because there is evidence for a metabolic steady state in general (i.e. no net change for every metabolites at each time point) ([3] p. 10-11).

One important constrained based modeling approach is flux balance analysis (fba) ([10], [7]), which will be of utmost importance for this project. Here, the former nonlinear problem  $dx/dt = S \cdot v$  diminishes to  $dx/dt = S \cdot v \stackrel{!}{=} 0$ , which constitutes a normal linear equation systems. Nevertheless, there are far more reactions than metabolites ( $r > m$ ), so that this linear reaction system is underdetermined. That is why other constraints like flux limits are added. Flux limits are reaction limits, which narrow down each reaction to some interval (e.g. irreversible reaction number  $i$  has a flux  $v_i > 0$ ). By this the solution space is shrunked. If the biomass composition, non and growth associated maintenance

(ngam/gam) is known, it is possible to formulate an optimization problem:

$$\begin{aligned} & \underset{v}{\text{maximize}} \quad b(v) \\ & \text{subject to} \quad S \cdot v = 0 \\ & \quad l_i < v_i < u_i \end{aligned}$$

where  $l_i$  and  $u_i$  are the lower and upper limits for reaction  $i$  and  $b(v)$  is biomass function, which is going to be maximized with respect to a certain flux  $v$ . Thus, we search a vector  $v$  carrying quantitative values for all fluxes in the whole reaction system, so that a certain function (here biomass function) is optimal. Although constrained based modeling frameworks for studying species interactions exist (\*), the complexity of microbial communities is still difficult to asses with this approaches.

### 1.3 Agent based modeling

*Complexity theory* is a part of system science since 1970s, where order is not longer considered as something given but made by itself. Moreover, order is producible as a surface phenomenon by a complex process, which is i) self organizing, ii) secures its autonomy and iii) proceeds far from an equilibrium ([4] p. 8-10).

According to John Holland, who introduced the important notion of an *agent*, a complex adaptive system (CAS) is defined as follows:

„We will view CAS as systems composed of interacting agents described in terms of rules. These agents adapt by changing their rules as experience accumulates.” ([5] p. 10)

In this modeling paradigm no general differential equation governs the macro behaviour. The parts of the system called agents are explicitly described by *rules* instead of a theory. This enables the possibility to model „microscopic” phenomena, which give individual properties and defined information to the agents. „If-then rules” are heuristic and could depict relationships, where no mathematical description exists. Therefore, agents have individuality, live in a surrounding area (grid) with limited radius, so that only local interactions in the neighbourhood governed by rules are relevant to produce a global phenomena.

From this microscopic actions the global organization is produced. New properties and behaviour could occur and this is noted by the slightly magical term *emergence* ([12] p. 36-39).

Agent based modeling (abm) has been successfully applied in ecological studies to model the complexity of global behaviours by simple local interactions between species (\*).

### 1.4 Aim of the project

The aim of this project is to combine for the first time constrained based modeling in an ecological context with abm to model interactions in microbial communities. In particular microbes are represented as agents which interact with their surrounding substrate concentrations stored in a grid. These interactions are realized with fba, which is used as a rule.

This framework will be used to represent common observations in microbial ecology such as syntropy.

## 2 Methods

### 2.1 Model overview

To model the metabolic interaction of multiple species populations, each individual was represented as an agent on a grid environment (Figure 1). The grid environment was composed of different

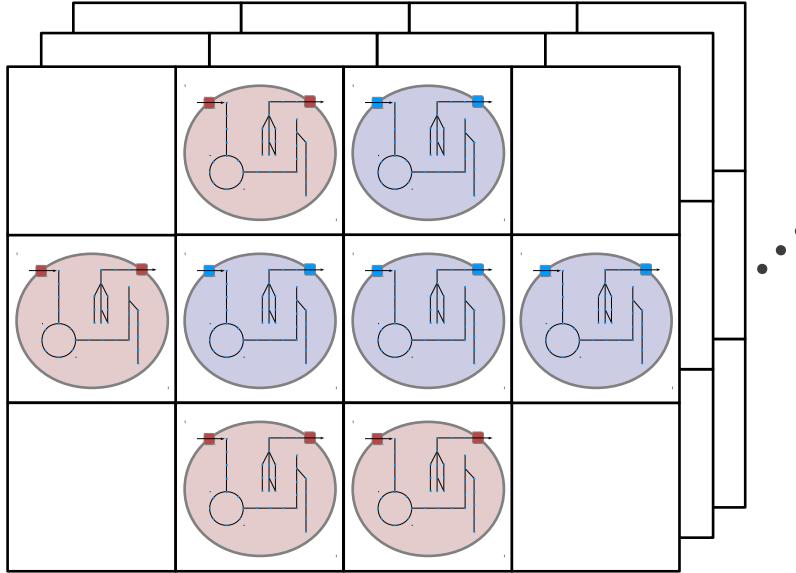


Figure 1: Overview of the grid environment with two different bacterial agent types (species) in red and blue. Each substrates has its own grid representation.

metabolite concentrations. Furthermore, the species type was recorded for each agent to call the respective genome-wide metabolic reconstruction in every iteration (Algorithm 1). The constraints for the subsequent fba were set according to the current metabolite concentrations of the grid cell, where the respective agent was located. The solution of the fba was used to adjust the biomass

---

**Algorithm 1:** Main model iterations called by `diffbac.R` with the different functions applied to bacterial agents and metabolite concentrations.

---

```
for number of iterations do
    diffusion();
    for number of bacteria do
        fba();
        movement();
        growth();
    end
end
```

---

of each agent and to modify the metabolite concentrations according to the produced products and consumed substrates. This uptake and output of metabolites constitute the exchange with the virtual environment. A diffusion model was applied to spread the metabolite concentrations over the grid environment and a movement function allowed the random dispersal of each bacterial agent.

A central theme of `BacArena` was the modularity of each funcsubtion and metabolic model to allow the extension and replacement of certain parts of the framework. Furthermore, the modularity also

enables the inclusion of any desired amount of microbial species. In the following sections the modules of **BacArena** will be regarded more closely.

## 2.2 Representation

The main parts of the framework were implemented in the programming language R (\*). Additionally, certain parts were implemented in C++ and integrated in R with the package **Rcpp** (\*).

### 2.2.1 Environment & Grid

Agents in **BacArena** were assigned to specific two-dimensional  $n \times m$  grid positions  $i, j \in \mathbb{N}$  and a type variable indicating the species. The grid is a discretization of space and could be imagined as a chess board, where the agents can move like chess pieces. One single part of the grid is called *cell* with no biological meaning. For each grid positions certain metabolite concentrations were also recorded and stored in a separate matrix. For each metabolite a own matrix was constructed (Figure 1). The bacterial agents could interact with this environment by the consumption and production of metabolite concentrations. Attention is necessary for the boundaries of the grid. Here continuous boundary conditions were choosen, i.e. rectangle grid is forming a surface of a torus/donut (horn-torus in square case).

### 2.2.2 Bacteria

Bacteria were represented as rows in a matrix, which had four columns and rows according to the current number of agents on the grid. The first two columns contained the discrete positions of the agents on the grid. The third column indicated the species type of the respective agent. The fourth column stored the current biomass value of the agents.

For certain bacterial species published genome-wide metabolic reconstruction were used as a representation of the individual metabolism and to perform the flux balance analysis. In the present study the *Escherichia coli* (\*), *Methanosaerina barkeri* (\*) and *Clostridium beijerinckii* (\*) SBML model were used. For each model the metabolites of the published artificial minimal medium (except the metabolites mentioned below) were set to concentrations, which were always present.

In each iteration in **diffbac.R** (Algorithm 1), rows were deleted and included according to the growth function (see below). The grid positions were updated according to the movement function and the biomass increase as a solution of the fba was added to the current biomass value for each agent.

### 2.2.3 Substrate

Metabolite concentrations were represented as matrices  $m_s \in \mathbb{R}^{n \times m}$  for all substrates  $s$ . According to the requirements of the used genome-wide metabolic models certain metabolites were selected. In the present study these metabolites were: acetate, aketoglutarate, carbon dioxide, ethanol, formiate, fumarate, glucose, water, proton, lactate, oxygen, phosphate, pyruvate, succinate, hydrogen, methanol, methane, acetone, butyrate and butanol. This could be extended easily by just adding another substrate matrix. Nonetheless it's necessary to define the corresponding exchange reaction for all organisms for this new added metabolite because there exists no common nomenclature.

The entries of the metabolite matrices were initialized with a concentration of 70 mmol per grid cell. Additionally, in certain simulations the concentration was of selected metabolites set to 0 mmol to generate diverse metabolic phenotypes, such as anaerobic metabolism.

### 2.3 Interactions as rules

The different representations were closely integrated with rules, which acted on each individual level.

#### 2.3.1 Movement

---

**Algorithm 2:** Movement of bacterial agents in the von Neumann neighbourhood with  $i$  and  $j$  as the current positions on the grid

---

```

 $a \leftarrow i + \text{random}(1, -1);$ 
 $b \leftarrow j + \text{random}(1, -1);$ 
for number of bacteria do
    if  $(a, b) \in \text{bacteria positions}$  then
        | do not move
    else
        |  $i \leftarrow a;$ 
        |  $j \leftarrow b;$ 
    end
end
```

---

#### 2.3.2 Diffusion

In agent based modeling there are two different rules for updating. A Synchronous mode updates all cells simultaneously, i.e. local changes are stored in a temporary copy and only after computation of all cells this changes will be efficacious. Contrary to this, in an asynchronous mode changes will be made immediately ([Matthies2002] p. 92).

Implementing a simple, naive diffusion model, which interchanges states between neighboured cells, asynchronous updating with randomly chosen cells is preferred because i) synchronous updating violates conservation laws and ii) nonrandomly asynchronous is causing a biased diffusion direction [Bandman1999]. Further work could be done by implementing more sophisticated diffusion models like block-rotation [Bandman1999] or the discrete diffusion model by Grajdeanu [Grajdeanu2007].

#### 2.3.3 Flux balance analysis

As described in chapter 1.2 flux balance analysis (fba) determines a flux vector for all reactions by maximizing (in our case) biomass. To solve the optimization problem we use *lpsolve* [**lpsolve**], which has also a *R* package [**Rlpsolve**]. *lpsolve* is a mixed integer linear programming solver, „based on the revised simplex method and the branch-and-bound method for integers“.**[lpsolvedocu]**

For each organism, who is located somewhere on the grid, the corresponding *sbml* model is loaded. In the next step, several constraints could be applied:

1. Lower bound of irreversible reaction fluxes is set to 0
2. All internal metabolites are assumed to be in flux equilibrium, i.e.  $S \cdot v = 0$ .

---

**Algorithm 3:** Diffusion is implemented in c++ using Rcpp [**dirk**]

---

```
for all substrates  $s$  do
     $l \leftarrow 0$ ;
     $n \leftarrow \text{rows}$ ;
     $m \leftarrow \text{columns}$ ;
    while  $l + + < n \cdot m$  do
         $i \leftarrow \text{random}(1, n)$ ;
         $j \leftarrow \text{random}(1, m)$ ;
        min  $\leftarrow$  random local minimum in neighbourhood of cell  $i, j$ ;
         $m \leftarrow (s(\min_i, \min_j) + s(i, j))/2$ ;
         $s(i, j) \leftarrow m$ ;
         $s(\min_i, \min_j) \leftarrow m$ ;
    end
end
```

---

3. Available substrates control the flux of all exchange reactions (lower bound is set to substrate concentration)
4. Certain empirically validated flux boundaries (e.g. maximal uptake rate) limits exchange reactions even further.

Linear optimization delivers a steady state flux vector  $v$ , which maximizes the biomass reaction. All fluxes are multiplied according to previous growth, because the unit of  $v$  is  $[\text{mmol}/(g_{\text{dryweight}} \cdot h)]$  (higher biomass/dryweight  $\sim$  higher fluxes). The solution of the optimization is taken to redefine substrate concentrations. All uptaking exchange reactions lower the available substrates and the providing exchange reactions rise it. The flux of the biomass reaction defines the growth rate, which is added to the previous growth variable.

### 2.3.4 Growth

Growth is considered as output of the fba-optimized biomass function. Bacterial uptake of substrate leads to a growth rate  $[\text{mmol}/(g_{\text{dryweight}} \cdot h)]$ , which is the flux through the biomass reaction. Two different processes are part of the model: First, accumulation of biomass allows reproduction of bacteria (duplication). And second, a certain time of starvation consumes available biomass and drives finally to cell death.

**Duplication** For each organism it's obligatory to define maintenance. Always some non growth associated maintenance ( $\text{ngam}$ ) is needed, which accounts normal energy consumption for upkeeping the cell. During growth there is additionally some growth associated maintenance ( $\text{gam}$ ) to describe energy consumption necessary to replicate the cell. The unit for  $\text{ngam}$  and  $\text{gam}$  is  $\text{mmol}/(g_{\text{dryweight}} \cdot h)$ , again [9].

In **BacArena** all cells start with a biomass of 1 symbolizing an intact cell. Flux balance analysis (fba) estimates fluxes and returns especially a maximized flux through the biomass function. This biomass flux is added to the initial biomass value. A series of such steps symbolizes growth. If accumulated biomass reaches twice the initial biomass value, duplication is possible

**Death** Flux balance analysis considers a fixed ngam flux and growth depending gam part of the biomass function. Under starving conditions, where no fluxes according to fba calculations can be found, we consider the current biomass as reservoir, from which the cell can feed on.

$$\text{biomass}_{t+1} = -\frac{\text{ngam}}{\text{gam}} \cdot \text{biomass}_t + \text{biomass}_t$$

## 3 Results

### 3.1 Movement and Diffusion

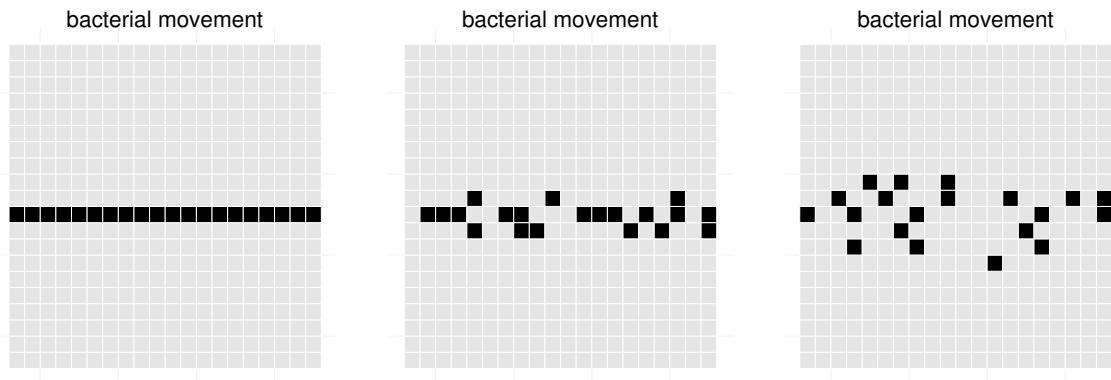


Figure 2: Bacterial movement starting with a line of bacteria in the middle (timestep 1, 2, 5)

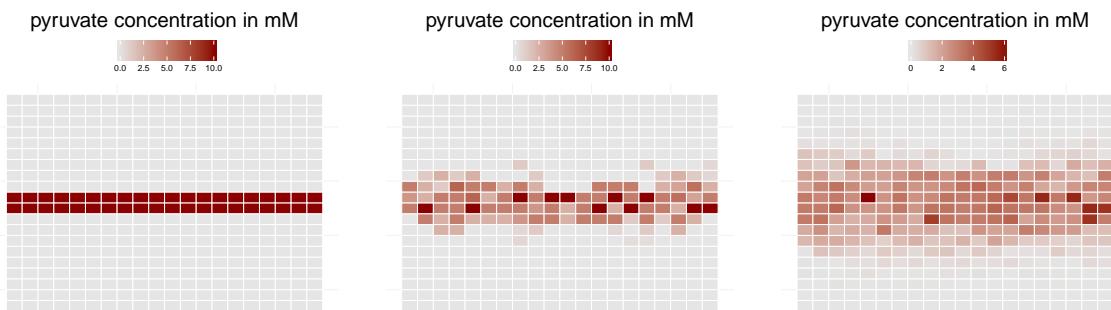


Figure 3: Diffusion on a  $20 \times 20$  grid starting with 10 mmol pyruvate in the middle (timestep 1, 2, 5)

### 3.2 Flux Balance Analysis

### 3.3 Growth models

### **3.3.1 Escherichia coli core**

### **3.3.2 Escherichia coli**

### **3.3.3 Methanosaerina barkeri**

### **3.3.4 Clostridium beijerinckii**

## **3.4 Mixed communities**

### **3.4.1 Escherichia coli & Methanosaerina barkeri**

### **3.4.2 Escherichia coli & Clostridium beijerinckii**

### **3.4.3 Clostridium beijerinckii & Methanosaerina barkeri**

## **4 Discussion**

### **4.1 Population models of single organisms**

### **4.2 Synthropy in mixed communities**

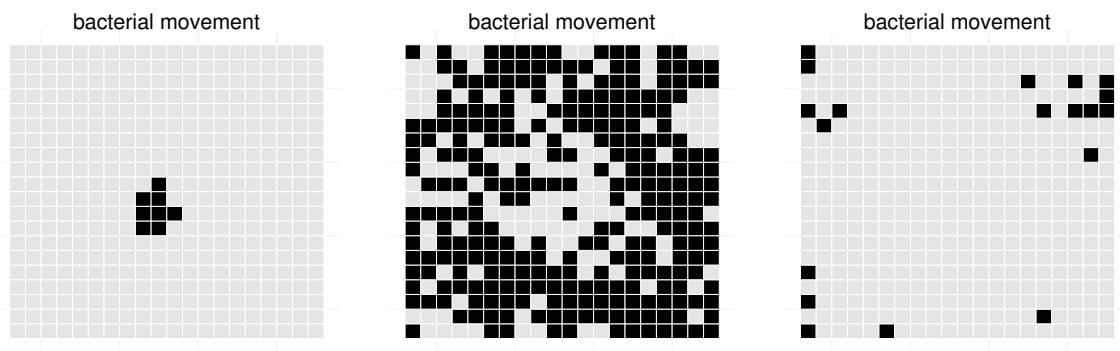
### **4.3 From heterogeneity to biofilms**

### **4.4 Conclusions & outlook**

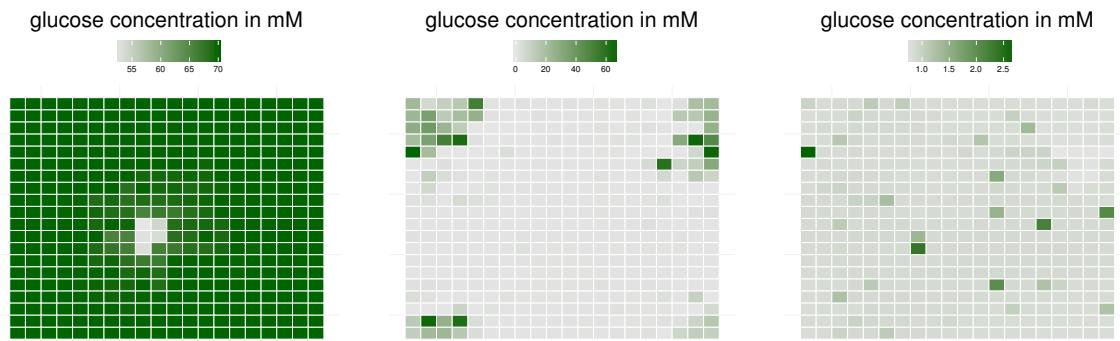
## **References**

- [1] Kevin M Esvelt and Harris H Wang. “Genome-scale engineering for systems and synthetic biology”. In: *Molecular systems biology* 9.1 (2013).
- [2] Adam M. Feist et al. “Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosaerina barkeri*.” In: *Mol Syst Biol* 2 (2006), p. 2006.0004.
- [3] David A. Harris. *Bioenergetics at a glance*. Oxford: Blackwell Science, 1995.
- [4] Francis Heylighen, Paul Cilliers, and Carlos Gershenson. “Philosophy and complexity”. In: *Complexity, science and society* (2007), pp. 117–134.
- [5] John Henry Holland. *Hidden order: How adaptation builds complexity*. Basic Books, 1995.
- [6] Edda Klipp, ed. *Systems biology*. Weinheim: Wiley-VCH, 2010.
- [7] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø. Palsson. “What is flux balance analysis?” In: *Nat Biotechnol* 28.3 (2010), pp. 245–248. DOI: 10.1038/nbt.1614.
- [8] Thomas Petzoldt and Karsten Rinke. “Simecol: an object-oriented framework for ecological modeling in R”. In: *Journal of Statistical Software* 22.9 (2007), pp. 1–31.
- [9] Jan C Thiele and Volker Grimm. “NetLogo meets R: linking agent-based models with a toolbox for their analysis”. In: *Environmental Modelling & Software* 25.8 (2010), pp. 972–974.
- [10] A. Varma and B. O. Palsson. “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.” In: *Appl Environ Microbiol* 60.10 (1994), pp. 3724–3731.

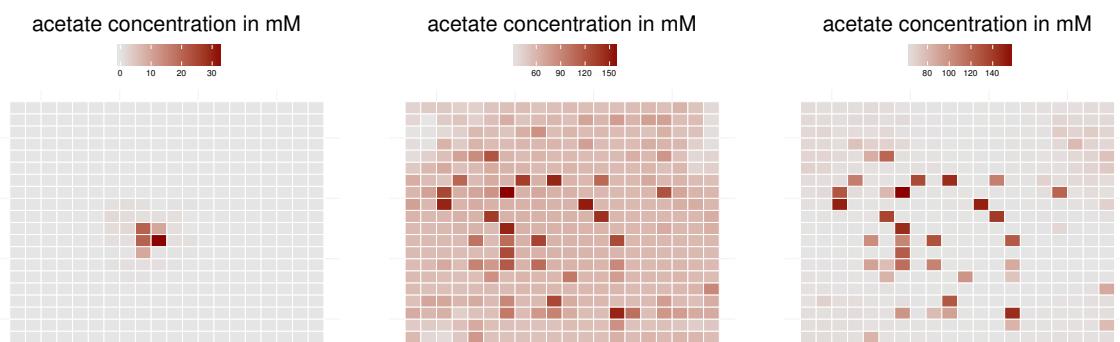
- [11] U Wilensky. “NETLOGO itself: NetLogo”. In: *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston*, (1999).
- [12] J. Zimmermann. “Für eine Geschichte der Systemwissenschaft”. In: ed. by M. Matthies. <http://www.usf.unisnabruueck.de/usf/literatur/beitraege/texte/053-zimmermann.pdf>. Beiträge des Instituts für Umwelt-systemforschung der Universität Osnabrück, 2010.



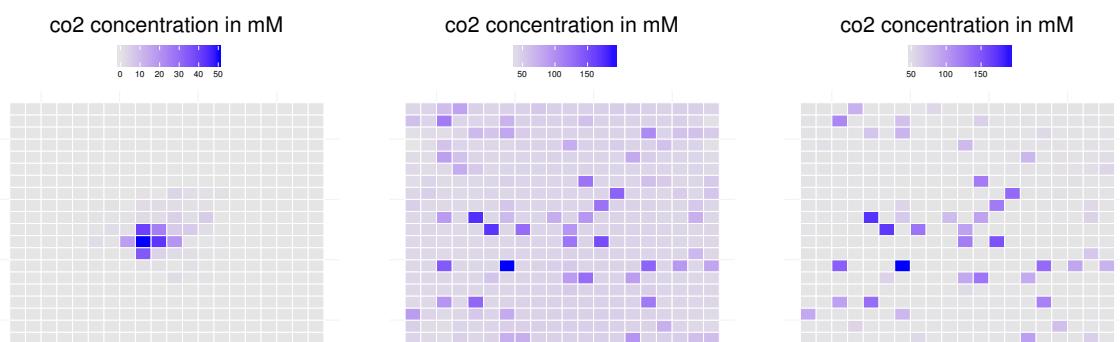
(a)



(b)



(c)



(d)

Figure 4: Aerobic growth of ecoli core model, grid 20x20, seed=55

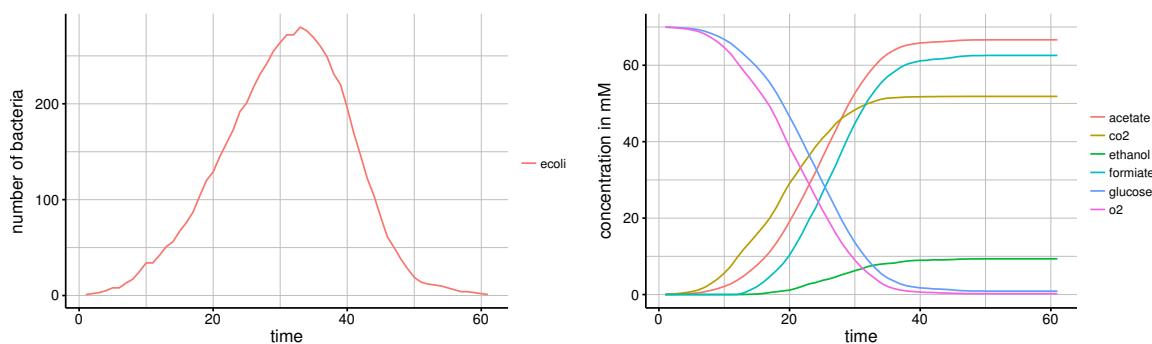
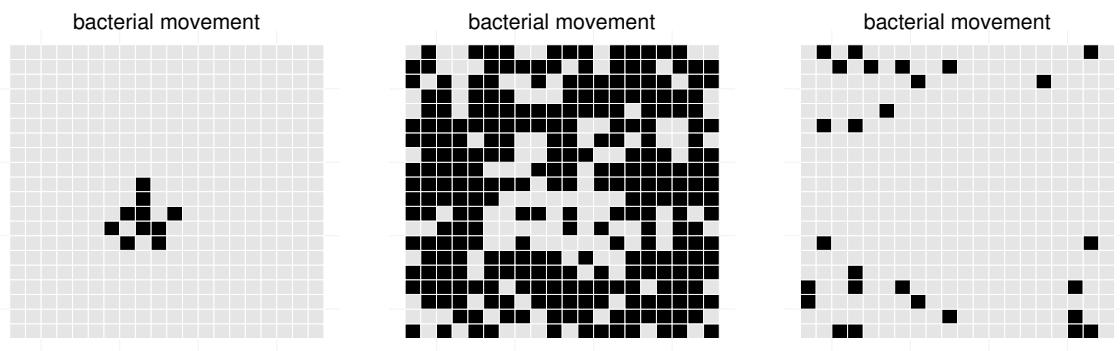
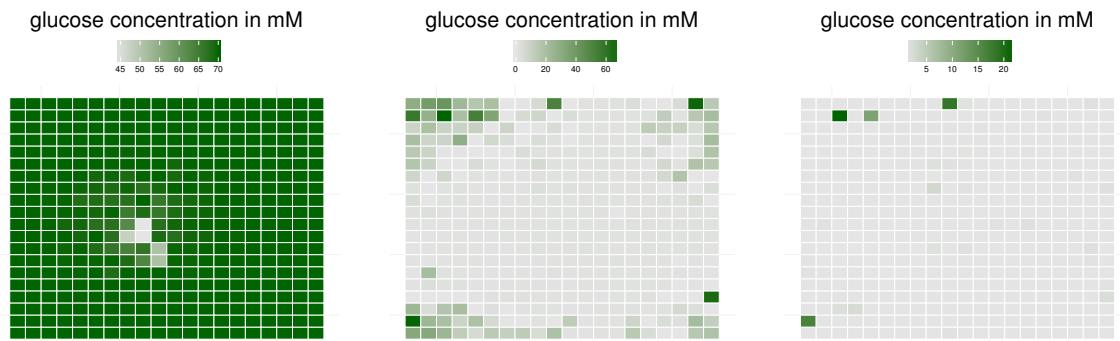


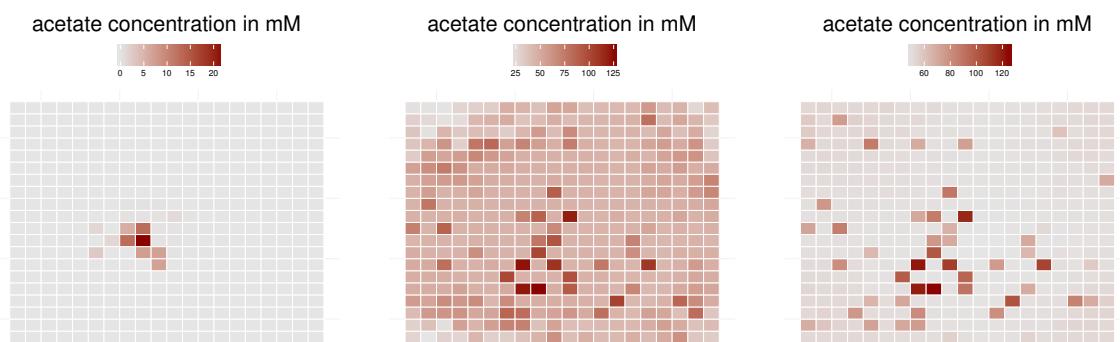
Figure 5: Aerobic growth of *ecoli* core model, grid 20x20, seed=55



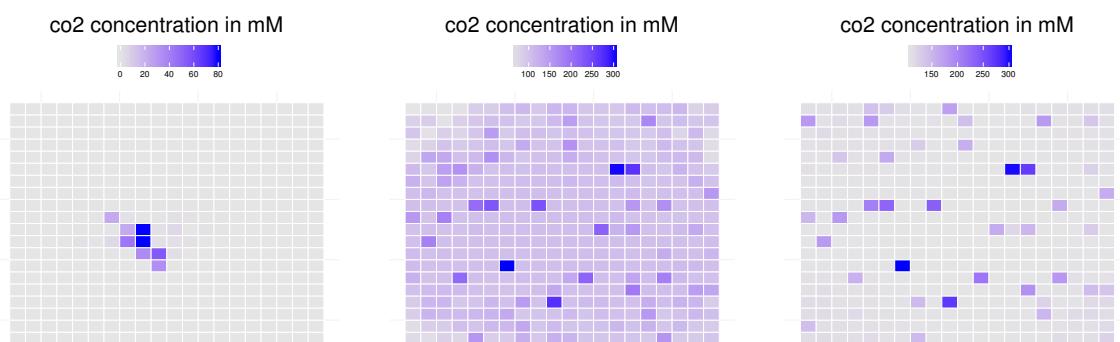
(a)



(b)



(c)



(d)

Figure 6: Aerobic growth of ecoli core model, grid 20x20, seed=55

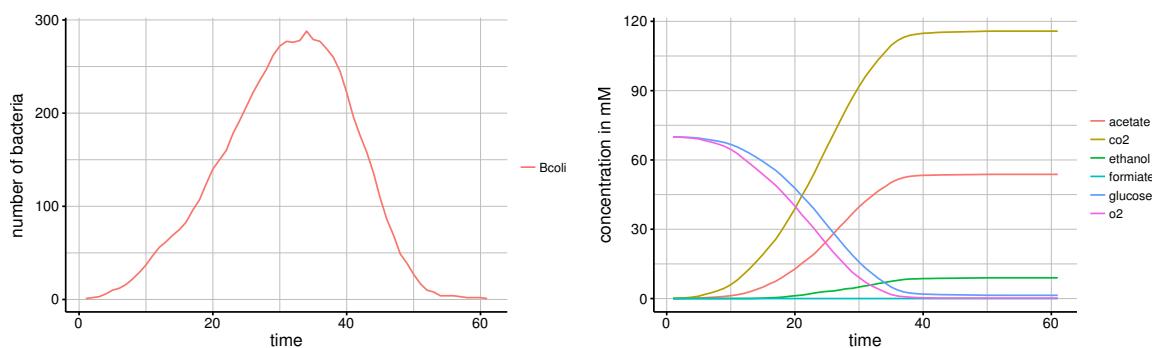
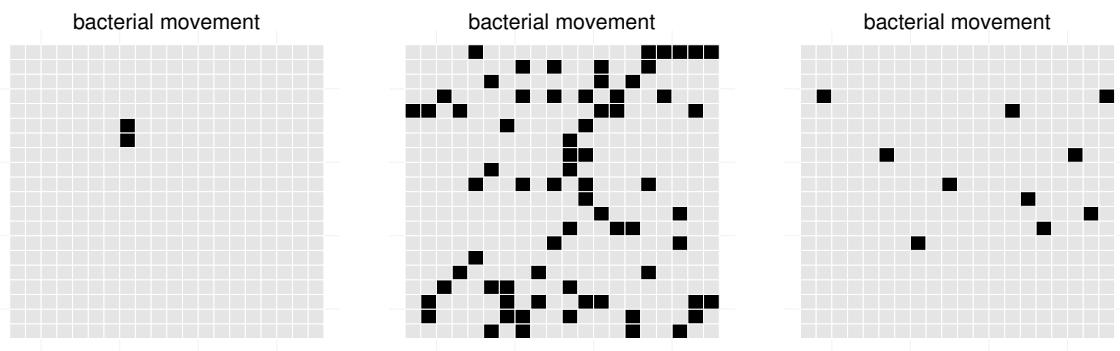
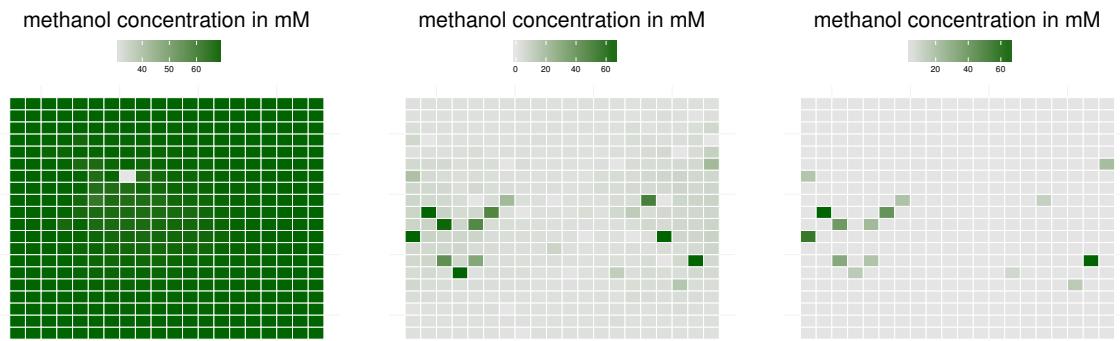


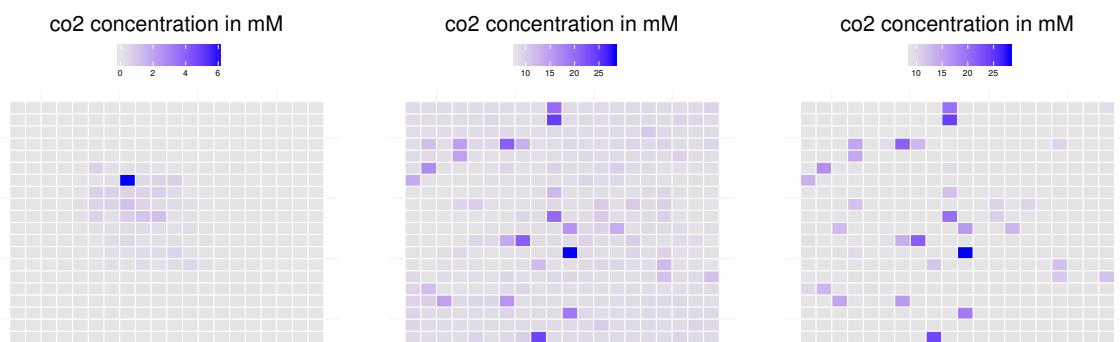
Figure 7: Aerobic growth of ecoli core model, grid 20x20, seed=55



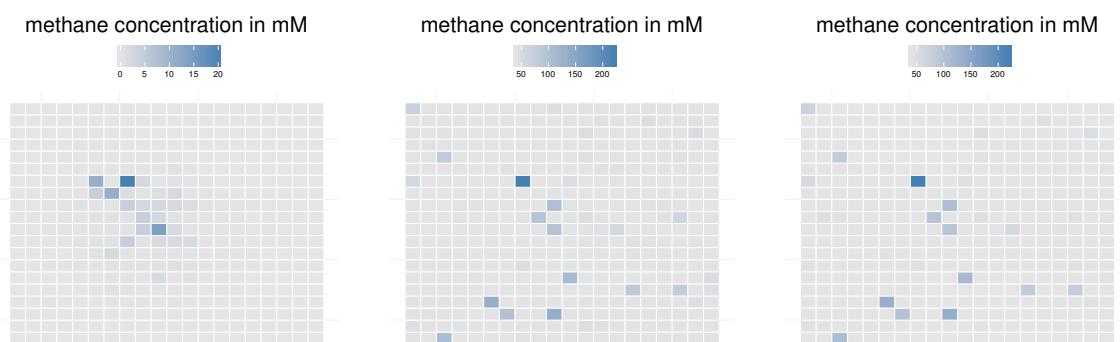
(a)



(b)



(c)



(d)

Figure 8: Aerobic growth of ecoli core model, grid 20x20, seed=55

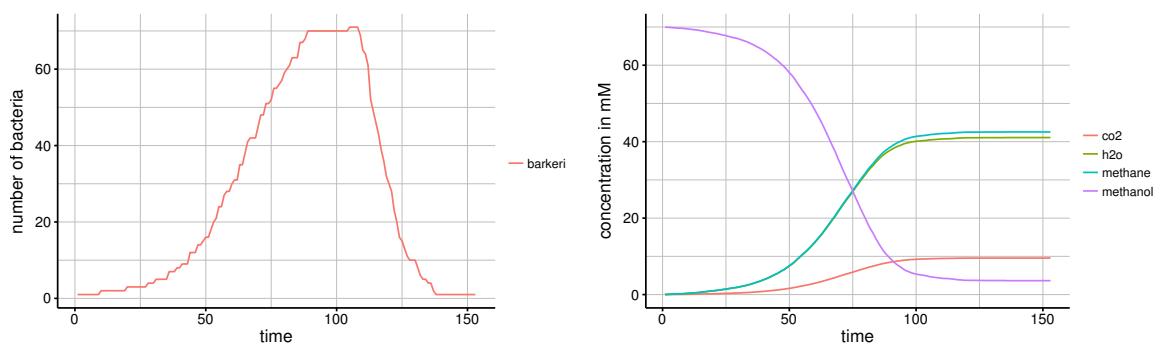


Figure 9: Aerobic growth of ecoli core model, grid 20x20, seed=55

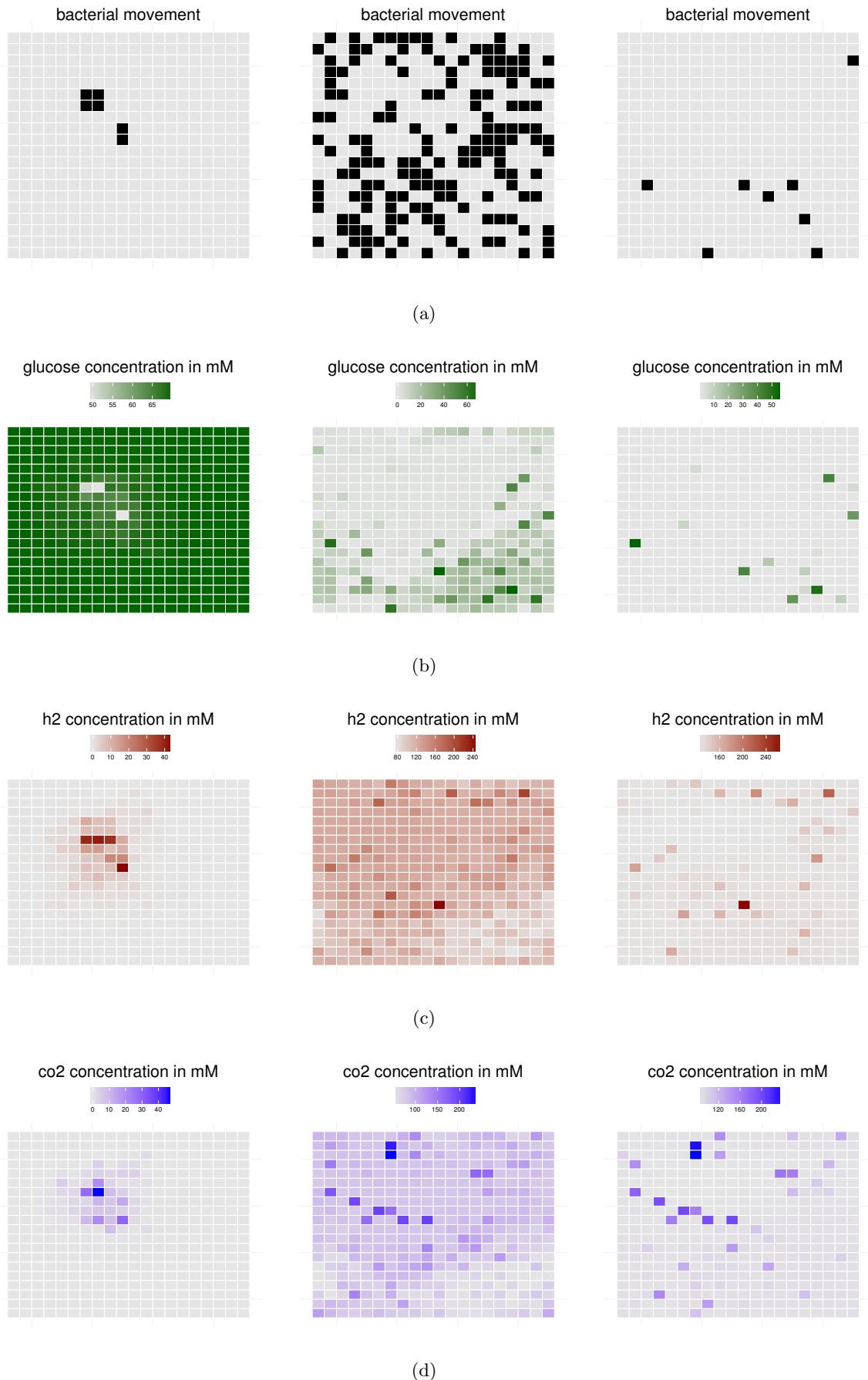


Figure 10: Aerobic growth of ecoli core model, grid 20x20, seed=55

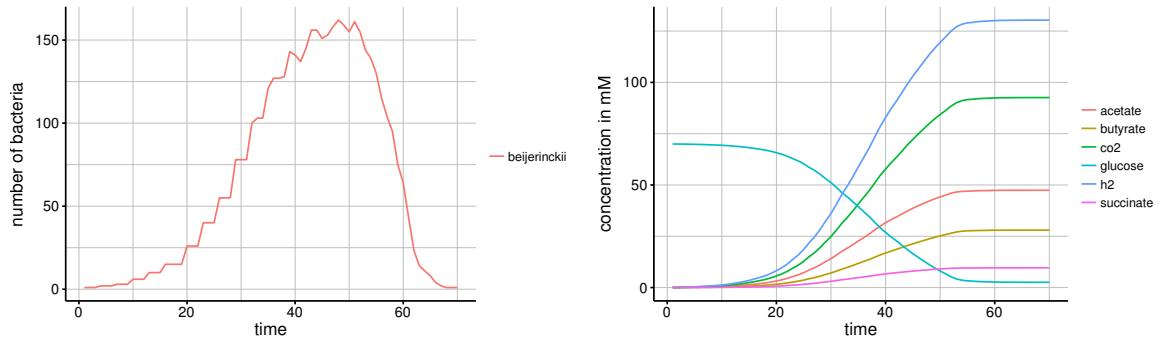


Figure 11: Aerobic growth of ecoli core model, grid 20x20, seed=55

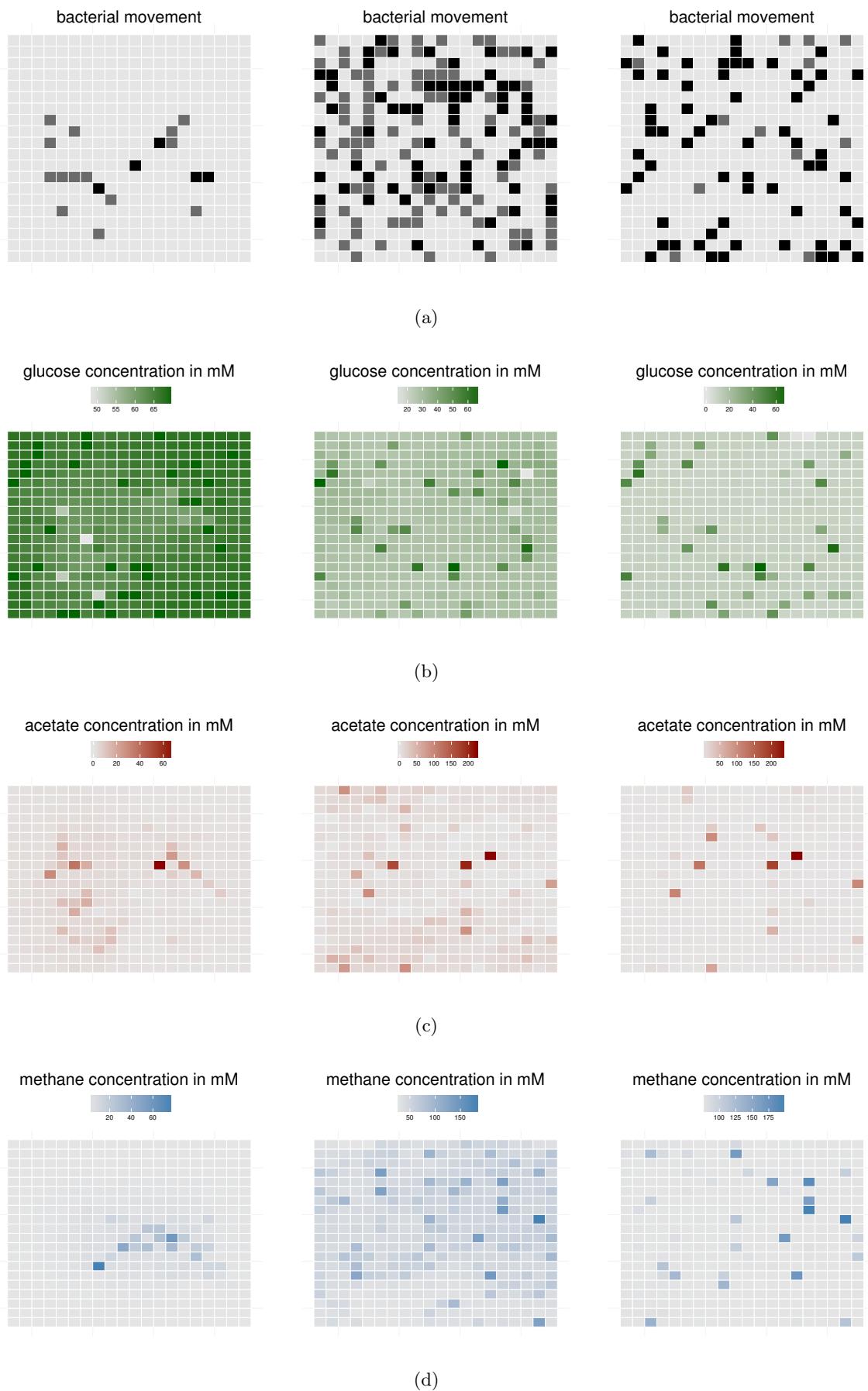


Figure 12: Aerobic growth of ecoli core model, grid 20x20, seed=55

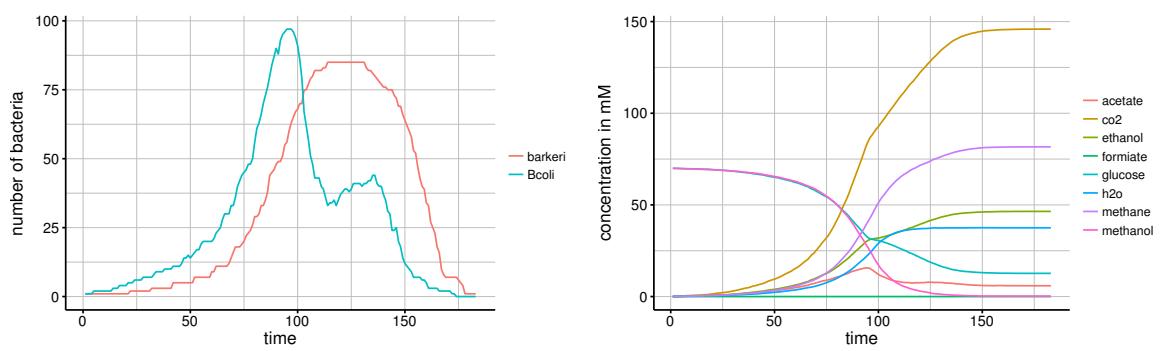
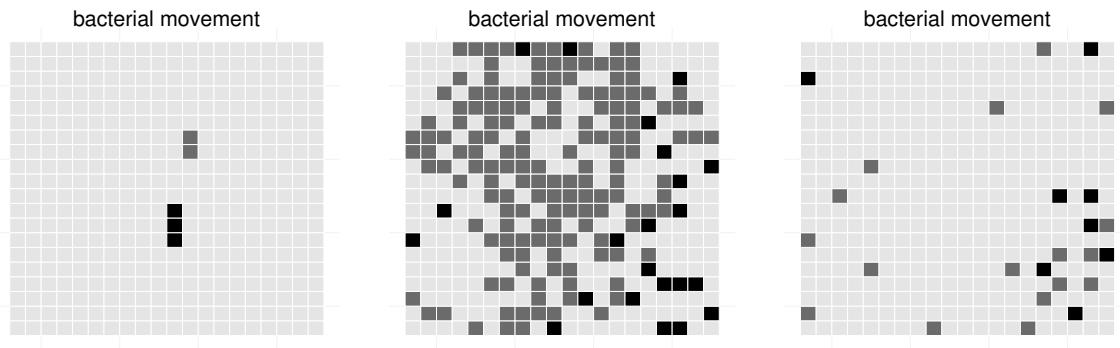
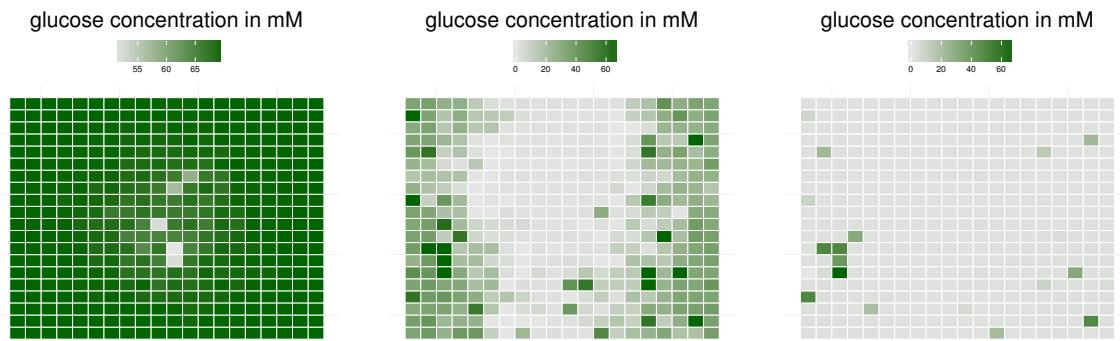


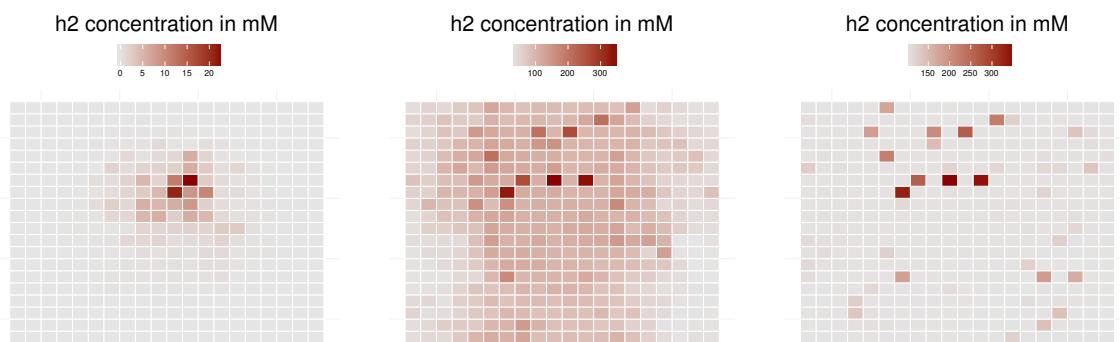
Figure 13: Aerobic growth of *ecoli* core model, grid 20x20, seed=55



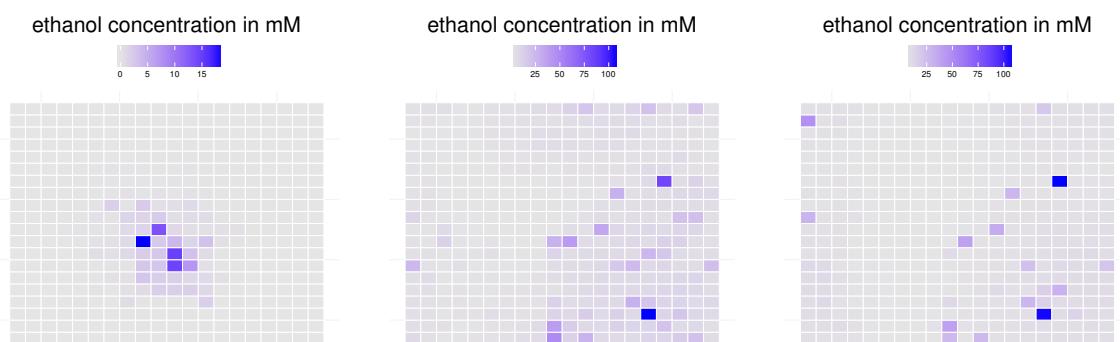
(a)



(b)



(c)



(d)

Figure 14: Aerobic growth of ecoli core model, grid 20x20, seed=55

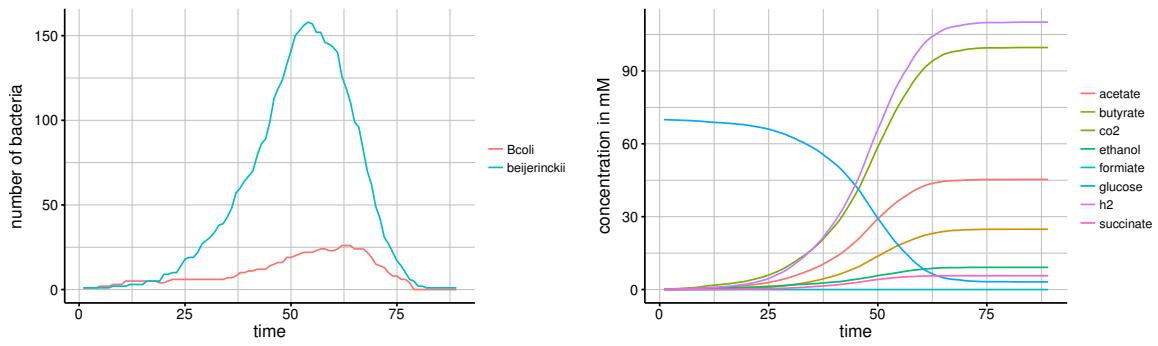


Figure 15: Aerobic growth of ecoli core model, grid 20x20, seed=55

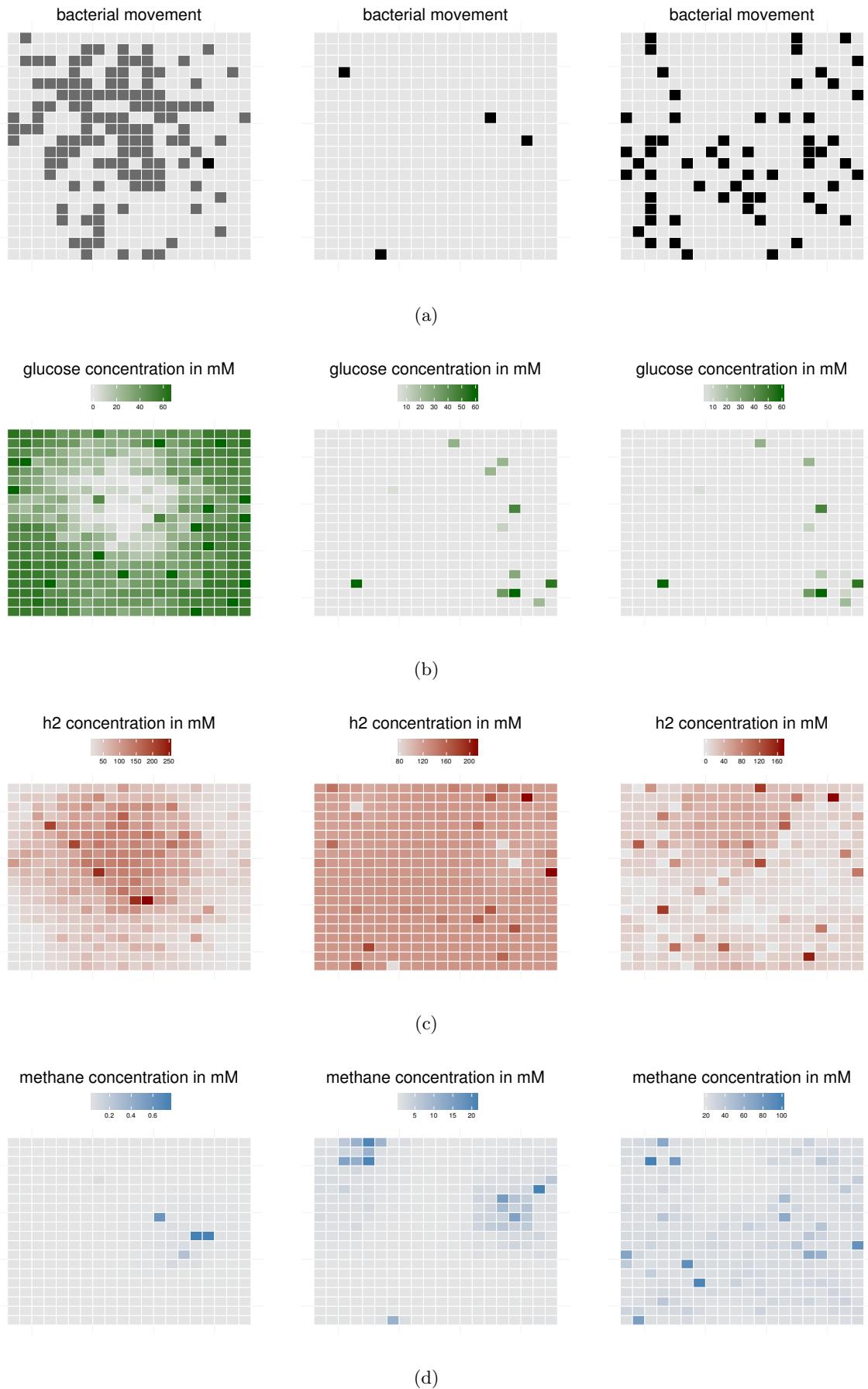


Figure 16: Aerobic growth of ecoli core model, grid 20x20, seed=55

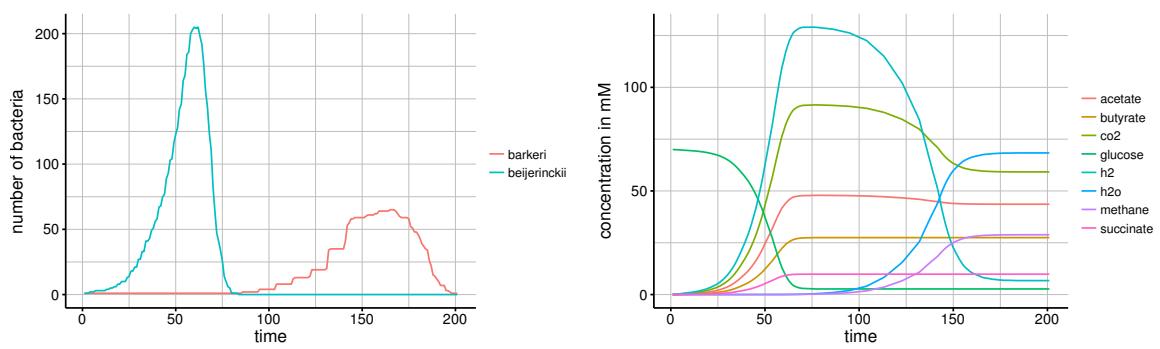


Figure 17: Aerobic growth of ecoli core model, grid 20x20, seed=55