*Category: Bioinformatics*

# Individual-Based Modeling of Bacterial Genetic Elements

**Venetia A. Saunders**
*Liverpool John Moores University, UK*

**Richard Gregory**
*University of Liverpool, UK*

**Jon R. Saunders**
*University of Liverpool, UK*

## INTRODUCTION

Individual-based computational modeling of biological systems is an important complement to experimental research. The individual-based model (IbM) is a bottom-up approach that considers the fate of individuals, their properties and interactions, and the influence of these interactions, holistically, on properties of the system. This contrasts with population-based models dependent on averaged behaviour of the whole system (DeAngelis & Gross, 1992; Huston, DeAngelis, & Post, 1988). IbMs can track individuals in time so that unusual events can be captured. They are particularly suited to biological simulations, where individuals might represent virtual plants, animals, or microorganisms in differing ecosystems. Lower complexity, coupled with the wealth of genetic knowledge about bacteria, allow for more realistic simulations compared with higher organisms. Accordingly, a lineage of IbMs, including Bacteria Simulator (BacSim) (Kreft, Booth, & Wimpenny, 1998; Kreft, Picioreanu, Wimpenny, & van Loosdrecht, 2001), INDividual DIScrete SIMulation (INDISIM) (Ginovart, Lopez, & Gras, 2005; Ginovart, Lopez, & Valls, 2002; Prats, Lopez, Giro, Ferrer, & Valls, 2006), COmputing Systems of Microbial Interactions and Communications (COSMIC) (Gregory, Paton, Saunders, & Wu, 2004; Paton, Gregory, Vlachos, Saunders, & Wu, 2004), RUle-based BActerial Modeling (RUBAM) (Paton, Vlachos, Wu, & Saunders, 2006; Vlachos, Paton, Saunders, & Wu, 2006) and COSMIC-Rules (Gregory, Saunders, & Saunders, 2006, 2008b), based on COSMIC and RUBAM, has been developed for bacterial simulations.

Although all these models are individual-based, underlying simulation mechanisms and aims vary. BacSim was the first to use IbM in a recognizable biological context (Kreft *et al*., 1998, 2001) aiming to model growth and cell division, quantitatively, at the population level, using a pseudocontinuous 2-dimensional world with restricted nutrients. INDISIM is based on stronger mathematical foundations, and is a discrete space and time stochastic simulation of colony growth, largely based on random variables (Ginovart *et al*., 2002). Each cell is a set of parameters existing at a discrete location. COSMIC uses pseudocontinuous space and discrete time to model evolution of cells (Gregory *et al*., 2004). Each cell contains a bit string genome that interacts with itself and the environment. This model is largely deterministic, although random events do have a role. It can run in a parallel machine, though any random effects this creates have been removed. RUBAM is a simplification of COSMIC, with pseudocontinuous space, discrete time, and a much more simplified genome. It aims to model adaptation (Vlachos *et al*., 2006). The simplified genome allows for comparatively rapid simulations that show adaptation and acquired resistance to antibiotics. COSMIC-Rules is a culmination of IbM modeling design, having an effective balance of modeling detail while being computationally tractable (Gregory *et al*., 2006, 2008b). Like COSMIC, it is a parallel simulation with pseudocontinuous space and discrete time. It uses a genome abstraction to represent the conditions and outputs of complex biochemical pathways, while incorporating an element of specificity and means of simulating evolution. Like the other IbMs considered here, each individual has its own parameters and state. Unlike the other IbMs, the scope of COSMIC-Rules covers vertical and horizontal gene transfer using populations of millions of cells.

## BACKGROUND

IbMs describe behaviour in a system, acknowledging the uniqueness of the individual, its characteristics and interactions with other individuals. Individuals are only considered together as a population or community when analysed. The majority of IbM approaches to bacterial simulations have focussed on growth and metabolism from ecological and evolutionary perspectives. However, COSMIC-Rules (Gregory *et al*., 2006, 2008b), the IbM described here, has been designed to simulate genetic interactions in bacteria

within a framework that allows adaptation and evolutionary processes to be observed. An advantage of the IbM in these simulations is that bacterial evolution becomes open-ended: emergence, growth, and death of individual bacteria, their interactions with other bacteria, and any infection events can be monitored over time. Such an approach permits questions about impact of individual variability on adaptive evolution to be addressed. Genetic elements, such as plasmids and viruses, can spread within bacterial populations mediating genetic exchange (Sørensen, Bailey, Hansen, Kroer, & Wuertz, 2005) and, coupled with mutation, provide raw materials for adaptive evolution (Marri, Hao, & Golding, 2007). By acquiring new or mutated genes, bacteria can adapt and survive in changing environments. Some plasmids are conjugative (self-transmissible), transferring by the horizontal gene transfer process of conjugation. This requires cell-to-cell contact involving a conjugation ligand, encoded by the donor, and a receptor on the recipient (Manning & Achtman, 1979).

Bacteriophages (phages) are viruses that infect bacteria (Carter & Saunders, 2007). The infective cycle is initiated by attachment of phages to susceptible bacteria through specificity of phage ligand-host cell receptor interactions. Phages may be temperate or virulent. A temperate phage is capable of operating in lytic (host killing) or lysogenic (without harming the host) modes. Phage replication occurs in the lytic cycle, culminating in cell lysis and release of progeny phages. For lysogeny, an inactive phage genome is stably inherited; there is neither bacterial lysis nor production of progeny phages. Phages that lysogenize the host (lysogen) confer immunity to superinfecting, homologous phages, and may effect changes to host properties by lysogenic conversion (Brussow, Canchaya, & Hardt, 2004). Furthermore, lysogeny promotes adaptation to survival in poor/unstable environments where resources are limited, as well as providing a potential reservoir of progeny phages.

## MODELING GENETIC ELEMENTS IN BACTERIA

### The Model: COSMIC-Rules

COSMIC-Rules (Gregory *et al*., 2006, 2008b) incorporates three levels using IbM philosophy: the genome, the cell, and an environment populated by such cells. Organisms possess individually defined physical locations, size, cell division status, and genomes including extrachromosomal elements (*e.g.,* plasmids and phages). The virtual environment consists of multiple, individual substances (substrates and toxic agents *e.g.,* antibiotics) whose relative nutrient status and/or toxicity is specified by the make-up of particular bacterial genomes. Individuals have specific, mutable genotypes and phenotypes evolving in a medium of initially defined, though changeable, composition. The environment is a 3-dimensional space, with the third dimension being of one cell diameter, so that cells effectively move in two dimensions.

Central to the model are novel features for representing genotypes and phenotypes in a compressed manner (genome compression). Each gene/gene set is represented by a unique tagged bit string that defines coding capacity and mediates specific genotypic and phenotypic interactions through bit string matching (Figure 1). This allows modeling of ligand-receptor interactions required, for example, for bacterial cell contact and conjugation or phage infection, metabolism-substance (substrate) interactions for cell growth, and susceptibility-substance (antibiotic/toxic agent) interactions for antibiotic action and cell death. If there is interaction with a resistance tag, probability of death is reduced to zero. A successful outcome demands that tagged bit strings match, and matching depends on them being no more than two bits different. The degree of similarity for matching varies with events under consideration. There are multiple metabolism, susceptibility, and resistance tags for each individual, together with multiple substances, and the best match is used in each case. "Best" refers to an outcome that produces either highest growth rate or highest probability of cell death. Collectively, bit strings form the genome and only genes directly involved in a particular simulation are considered; other required functions are assumed to be provided by covert gene sets to reduce computational load. However, flexibility of the model allows additional bit strings to be incorporated to increase genome complexity as necessary.

COSMIC-Rules models simplifications of real-world situations, aiming to reproduce biologically realistic conditions, by applying a series of rules, informed by physical laws and principles of bacterial genetics. Simulation parameters reflect biological values, as determined experimentally. Rules govern behavior of individuals in simulations and are varied for different scenarios.

The model is built for parallel execution utilizing a development cluster and scales to large HPC systems. To achieve parallelism, the environment is partitioned into demes, each containing individuals that move and interact with other individuals, and for one time step, the deme is isolated from other demes. Individuals then move and may migrate to another deme. Isolating demes in this way facilitates computability, since cell-to-cell synchronization between demes is not required.

The simulation treats each bacterium or free phage as an object instance, with its own associated parameters and genome. Genomes are also handled as objects. Within each genome are tagged bit strings representing compressed genes that make this approach tractable. Individual bacteria or free phages are subject to mutation, and soon each individual has its own unique bit strings associated with each tag. Once parallel synchronization has been achieved, a typical cycle