Bull Math Biol (2011) 73: 701–705
DOI 10.1007/s11538-011-9643-7

EDITORIAL

Algebraic Methods in Mathematical Biology

Reinhard Laubenbacher

Published online: 12 March 2011

The use of mathematical tools to study biological systems has a long history, with such milestones as the Lotka–Volterra model (Lotka 1910), the mathematics of enzyme kinetics (Michaelis and Menten 1913), the Hodgkin–Huxley model of action potentials in neurons (Hodgkin et al. 1952), and Turing's theory of morphogenesis (Turing 1952). Traditionally, the principal tool of the mathematical biologist was the differential equation, which provides models of a variety of dynamic processes, with many applications, in particular in population dynamics, one of the few areas of biology where some quantitative data were available to calibrate models to reality. This lack of suitable data to build realistic quantitative dynamic models was an important reason why mathematical biology was not able to reach its full potential during most of the 20th century. This was to change dramatically with the advent of the technological revolution in molecular biology that began in the 1970s with DNA sequencing machines and in the 1990s with DNA microarrays. Population biology also benefited from better data collection technologies. Ever-improving in vivo imaging methods promise similar advances at the tissue and organ levels for mammals. In genetics and genomics we are now in a position where data generation capabilities have far outstripped the capabilities of analyzing them, both for lack of appropriate algorithms and lack of suitable hardware. Also, entirely new higher quality data types are becoming available, such as CHIP-seq molecular data or measurements of single-cell protein concentrations. With new data types and larger data quantities, new problems can be approached with mathematical methods. Systems biology applies principles from mathematical systems theory to biological systems and networks whose components we can now measure on a large scale. Evolutionary biology benefits from the

R. Laubenbacher (✉)
Virginia Bioinformatics Institute, Washington St. (0477), Blacksburg, VA 24061, USA
e-mail: reinhard@vbi.vt.edu

availability of many related genomes, and multiple genome alignment has become a pressing unsolved mathematical and statistical problem.

New data types and quantities provide new challenges and opportunities. One consequence is that a larger range of mathematical methods has been applied to biological problems in recent years. At the genomic level, biology is essentially digital, and so it is not surprising that combinatorics has been used very successfully there, most spectacularly in connection with the Human Genome Project, but also in other areas, such as the study of secondary RNA structures (Bakhtin and Heitsch 2009). Searching the PubMed database of literature related to biomedical research, one now finds papers utilizing Lie algebras (Sanchez et al. 2006), the Riemann Mapping Theorem (Hurdal and Stephenson 2009), and methods from algebraic topology (Singh et al. 2008) and algebraic geometry (Wang et al. 2005; Laubenbacher and Stigler 2004), to name a few topics not traditionally found in the applied mathematics repertoire. As the biological challenges evolve, there is every reason to expect that more and more fields of mathematics will be in a position to contribute new points of view and new approaches. One field whose tools and results have found many applications in biological research in the last decade is algebraic geometry, broadly defined. In the last 5 years, a series of workshops, conferences, and programs at the US Mathematics Institutes, supported by the National Science Foundation, have helped to foster research in this field. For instance, a year-long program at the Statistical and Applied Mathematical Sciences Institute (SAMSI) in Research Triangle Park, NC, during 2008–2009 focused on "Algebraic Methods in Systems Biology and Statistics."

This special issue of the Bulletin of Mathematical Biology collects together a sampling of papers each of which uses methods grounded in discrete mathematics, combinatorics, and algebraic geometry to help solve a biological problem. We will refer to these methods summarily as "algebraic." Taken together, the papers cover a broad range of applications and draw on a cross section of algebraic geometric tools and results. The collection is intended to showcase this new tool set available to mathematicians working on biological problems, and can serve as a guide to both mathematicians and biologists who are interested in pursuing research in this new application area. It must be emphasized that, of course, this volume is by no means representative of the breadth of the field, as much exciting work is being done that is not represented here.

The majority of papers included in this volume fall into the category of research that uses methods from algebraic geometry for the solution of biological problems. The key concept here is that of a polynomial, which is the basic mathematical object that underlies many algorithms in algebraic geometry and computer algebra. One of the fundamental contributions of computer algebra is to provide algorithms for the solution of systems of nonlinear polynomial equations. These can be used to solve biological problems that can be translated into the polynomial context. One important problem of this kind is the construction and analysis of models for biological networks. Biochemical networks are often described by a system of ordinary differential equations whose right-hand sides have polynomial form. One can use computer algebra methods to analyze such systems, for instance to compute steady states, which are obtained by setting the left-hand sides equal to 0. Beyond that, one can

study other objects, such as siphons, which are collections of molecular species that have the potential of being absent in a steady state of the system (Shiu and Sturmfels 2010). Here the collection of siphons is encoded as an ideal, a particular kind of algebraic object in a polynomial ring, which can then be studied using theoretical tools from algebraic geometry. Another biological problem that can be translated into a problem in computational algebraic geometry is the inference of gene regulatory networks from experimental data (Laubenbacher and Stigler 2004). Here, the polynomials arise as equations that govern the evolution of a time-discrete dynamical system over a finite state space. Algorithms for polynomials can be used to describe the entire set of such evolution equations that are consistent with the given data and to make an optimal selection. Another source of polynomial equations could be the set of constraints on the geometry of a viral capsid controlling its assembly (Sitharam and Agbandje-McKenna 2006).

Another rich source of polynomials lies in evolutionary biology. The central problem of phylogenetics is the inference of an evolutionary tree from sequence data. Stochastic models of sequence evolution along trees give rise to polynomial expressions for the probabilities of observations. The form of the polynomials reflects both the model and the topology of the tree. Algebraic techniques can therefore be used to study such models and their relationship to sequence data (Allman and Rhodes 2010).

The field is still very young and other applications represented here are still less developed, so it is difficult to group papers into well-delineated research themes. The hope is that this volume will catalyze more research around individual papers and will help discover new applications of algebraic methods. Following are brief descriptions of the individual articles, listed in alphabetical order according to first author.

Blöchl et al. (2011) focus on the combined effect of the kinetic parameters on the dynamics of hierarchical complex systems models of molecular signaling networks. The authors provide algebraic expressions for relevant combinations of these parameters that serve as so-called effective parameters, which they determine recursively from the interaction graph.

Bona et al. (2011) use combinatorics and group theory to answer questions about the assembly of icosahedral viral shells. The result helps clarify the effect of symmetry on the probability and number of their assembly pathways and, more generally, for any finite, symmetric macromolecular assembly.

Malaspinas et al. (2011) develop tools for analyzing the robustness of inferred phylogenies to perturbations in alignment parameters in the Needleman–Wunsch algorithm. The main tool is parametric alignment, with novel improvements that are of general interest in parametric inference.

Hower and Heitsch (2011) extend recent algebraic methods for parametric sequence alignment to the parameter space for scoring RNA folds. Using the so-called RNA polytope, the authors derive combinatorial structures to study the effect of variations of parameters not determined experimentally in the free energy model.

McCaig et al. (2011) study the role of "superspreaders" on the dynamics of disease transmission. The authors investigate stochastic individual-based models through conversion to deterministic, population level mean field equations, using process algebra. The mean emergent population dynamics is shown to be equiva-

lent with and without superspreaders. However, simulations confirm expectations of differences in variability, impacting individual dynamics.

Dickenstein and Perez Milan (2011) focus on mass action kinetics in biochemical systems. The authors clarify the relation between the algebraic conditions that must be satisfied by the reaction constants for the existence of detailed or complex or balancing equilibria.

Sainudiin et al. (2011) use algebraic methods to compute the exact likelihood of the site-frequency spectrum and many classical linear combinations of it at a non-recombining locus are neutrally evolving under the infinitely-many-sites mutation model. This is useful for the evaluation of the likelihood function of parameters in highly-structured population genetics models from extant DNA sequences.

Siebert (2011) studies discrete dynamical systems models of bionetworks. The author develops an analysis method of the dynamics of the entire network by studying subnetworks and extending their limit cycles to the entire network. She also uses these concepts to formulate sufficient conditions for the existence of multiple attractors based on the existence of positive and negative feedback loops.

Finally, Weber et al. (2011) investigate algorithmic methods to analyze parameter ranges in systems of ordinary differential equations that are meaningful from a biological point of view and for which one can find oscillating trajectories. The authors show that one can reduce the problem to quantifier elimination over the real numbers.

Acknowledgements Special thanks are due to Seth Sullivant and Yuriko Yoshida for their contribution to the conception of this special issue.

References

Allman, E., & Rhodes, J. (2010). Evolution. Trees, fast and accurate. *Science*, *327*, 1376–1379.

Bakhtin, Y., & Heitsch, C. (2009). Large deviations for random trees and the branching of RNA secondary structures. *Bull. Math. Biol. 71*(1), 84–106.

Blöchl, F., Wittmann, D. M., & Theis, F. J. (2011). Effective parameters determining the information flow in hierarchical biological systems. *Bull. Math. Biol.* doi:10.1007/s11538-010-9604-6.

Bona, M., Sitharam, M., & Vince, A. (2011). Enumeration of viral capsid assembly pathways: tree orbits under permutation group action. *Bull. Math. Biol.* doi:10.1007/s11538-010-9606-4.

Dickenstein, A., & Perez Milan, M. (2011). How far is complex balancing from detailed balancing. *Bull. Math. Biol.* doi:10.1007/s11538-010-9611-7.

Hodgkin, A. L., Huxley, A. F., & Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *J. Physiol.*, *116*, 424–448.

Hower, V., & Heitsch, C. (2011). Parametric analysis of RNA branching configurations. *Bull. Math. Biol.* doi:10.1007/s11538-010-9607-3.

Hurdal, M. K., & Stephenson, K. (2009). Discrete conformal methods for cortical brain flattening. *Neuroimage*, *45*(1 Suppl), S86–98.

Laubenbacher, R., & Stigler, B. (2004). A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.*, *229*(4), 523–537.

Lotka, A. J. (1910). Contribution to the theory of periodic reaction. *J. Phys. Chem.*, *14*(3), 271–274.

Malaspinas, A., Eriksson, N., & Huggins, P. (2011). Parametric analysis of alignment and phylogenetic uncertainty. *Bull. Math. Biol.* doi:10.1007/s11538-010-9610-8.

McCaig, C., Begon, M., Norman, R. A., & Shankland, C. E. (2011). A symbolic investigation of superspreaders. *Bull. Math. Biol.* doi:10.1007/s11538-010-9603-7.

Michaelis, L., & Menten, M. L. (1913). Kinetik der Invertinwirkung. *Biochem. Z.*, *49*, 333–369.

Sainudiin, R., Thornton, K., Harlow, J., Booth, J., Stillman, M., Yoshida, R., Griffiths, R. C., McVean, G., & Donnelly, P. (2011). Experiments with the site frequency spectrum. *Bull. Math. Biol.* doi:10.1007/s11538-010-9605-5.

Sanchez, R., Grau, R., & Mogardo, E. (2006). A novel Lie algebra of the genetic code over the Galois field of four DNA bases. *Math. Biosci.*, *202*(1), 156–174.

Siebert, H. (2011). Analysis of discrete bioregulatory networks using symbolic steady states. *Bull. Math. Biol.* doi:10.1007/s11538-010-9609-1.

Shiu, A., & Sturmfels, B. (2010). Siphons in chemical reaction networks. *Bull. Math. Biol.*, *72*(6), 1448–1463.

Singh, G., Memoli, F., Ishkhanov, T., Sapiro, G., Carlsson, G., & Ringach, D. L. (2008). Topological analysis of population activity in visual cortex. *J. Vis.*, *8*(8), 1–18.

Sitharam, M., & Agbandje-McKenna, M. (2006). Modeling virus self-assembly pathways: Avoiding dynamics using geometric constraint decomposition. *J. Comput. Biol.*, *13*(6), 1232–1265.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.*, *237*, 37–72.

Wang, L., Meitu, R. R., & Donald, B. R. (2005). An algebraic geometry approach to protein structure determination from NMR data. In *Proc. IEEE comput. syst. bioinform. conf.* (pp. 235–246).

Weber, A., Sturm, T., & Abdel-Rahman, E. O. (2011). Algorithmic global criteria for excluding oscillations. *Bull. Math. Biol.* doi:10.1007/s11538-010-9618-0.