Report of Dagstuhl Seminar 16011

# Evolution and Computing

**Edited by**

**Nick Barton[1], Per Kristian Lehre[2], and Nisheeth K. Vishnoi[3]**

1 **IST Austria – Klosterneuburg, AT,** `nick.barton@ist.ac.at`
2 **University of Nottingham, GB,** `perkristian.lehre@nottingham.ac.uk`
3 **EPFL Lausanne, CH,** `nisheeth.vishnoi@epfl.ch`

## Abstract

This report documents the talks and discussions at the Dagstuhl seminar 16011 "Evolution and Computing". The seminar brought together several research disciplines studying evolution, including population genetics and mathematical biology, theoretical computer science, and evolutionary computation.

**Seminar** January 4–8, 2016 – http://www.dagstuhl.de/16011
**1998 ACM Subject Classification** E.1 Data Structures, F.2 Analysis of Algorithms and Problem Complexity
**Keywords and phrases** Evolution, Evolutionary Computation, Natural Algorithms, Theory of Computation
**Digital Object Identifier** 10.4230/DagRep.6.1.1
**Edited in cooperation with** Jorge Perez Heredia

## 1 Executive Summary

*Nick Barton*
*Per Kristian Lehre*
*Nisheeth K. Vishnoi*

Biological evolution has produced an extraordinary diversity of organisms, even the simplest of which is highly adapted, with multiple complex structures. Dynamic structures at even higher levels emerge from collective and social behaviour. These phenomena have traditionally been studied in population genetics, ecology and related disciplines.

However, theoretical computer scientists, endowed with a wide variety of tools, have recently made progress in describing and characterising the computational capabilities of evolution, analyzing natural algorithms, obtaining quantitative bounds for evolutionary models and understanding the role of sex in evolution. The field of evolutionary computation has found that many innovative solutions to optimisation and design problems can be achieved by simulating living processes, such as evolution via random variation and selection, or social behaviour in insects. Researchers in evolutionary computation have recently started applying techniques from theoretical computer science to analyze the optimization time of natural algorithms.

To further the connections and consolidate this burgeoning new discipline, this Dagstuhl seminar brought together participants from the population genetics, mathematical biology, theoretical computer science, and evolutionary computation communities. The seminar opened with a round of introductions, followed by five introductory talks presenting the perspectives of the disciplines attending. Benjamin Doerr introduced runtime analysis of evolutionary algorithms, Paul Valiant discussed evolution from the perspective of learning, Joachim Krug and Nick Barton introduced population genetics, and Nisheeth Vishnoi discussed evolutionary processes from the perspective of theoretical computer science. In addition to talks contributed by participants, there were several breakout sessions on topics identified during the seminar.

The organisers would like to thank the Dagstuhl team and all the participants for making the seminar a success.

# 2 Table of Contents

# 3 Overview of Talks

## 3.1 Ancestral selection graph meets lookdown construction

*Ellen Baake (Universität Bielefeld, DE)*

Population genetics today relies crucially on mathematical concepts such as ancestral lineages and random genealogies. The talk provided an overview. It started from an interacting particle system that describes the joint action of random reproduction, mutation, and selection; then explained the constructions used to trace back the ancestry of individuals and their genealogy, and presented some recent results. In particular, it presented a novel approach that unifies the two established concepts in the field, namely, the ancestral selection graph of Krone and Neuhauser (1997) and the lookdown construction of Donnelly and Kurtz (1999).

### References

1 P. Donnelly and T. G. Kurtz. Genealogical processes for fleming-viot models with selection and recombination. *Ann. Appl. Probab.*, 9(4):1091–1148, 11 1999.
2 S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51(3):210–237, 1997.
3 U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theoretical Population Biology*, 103:27–37, 2015.

## 3.2 Population genetics and recombination

*Nick Barton (IST Austria – Klosterneuburg, AT)*

There has been a long-standing interest in evolutionary theory in how sex and recombination are maintained, despite their obvious costs. In an infinite population, one can find a general formula for the selection on a gene that slightly modifies recombination; this is derived assuming that interactions between genes (epistasis) are weak. The selection on recombination is expressed in terms of the effect of recombination in reducing mean fitness, and in increasing the additive variance in fitness. These results should be applicable to analogous problems concerning evolutionary algorithms.

## 3.3 The effect of epistasis on the response to selection

*Nick Barton (IST Austria – Klosterneuburg, AT)*

Quantitative genetics describes the evolution of complex traits, which depend on many genes with interacting effects. In 1960, Robertson derived a remarkably simple expression for the total change in mean of a trait, that can be produced by selecting on a population: this equals the population size times the change in the first generation. Robertson's result assumed an additive model, but it can be generalised to arbitrary gene interactions (i.e., epistasis). It is an application of the "infinitesimal model", which approximates the evolution of traits that depend on large numbers of freely recombining genes. I review this and other results that attempt to set general limits on the possible response to selection, and discuss approaches to the same problem from computer science.

## 3.4 Game Dynamics and Population Genetics

*Erick Chastain (Rutgers University – Piscataway, US)*

This is a review of recent work by C, Livnat, Papadimitriou & Vazirani on the connection between Game Theory, Algorithms, and Evolution. We also mention some interesting open problems and the progress we have made on them (including a partial extension of our results to Diploid organisms), indicating promising directions for those interested.

## 3.5 Analysis of Evolutionary Algorithms

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

In this first talk of the Dagstuhl seminar "Evolution and Computation", I will give an easy introduction to the field of analyses of evolutionary algorithms, aimed at an audience with backgrounds in general algorithms or theoretical biology.

To get a quick start into the topic, I will present a particular, but typical result first, namely how simple evolutionary algorithms optimize pseudo-Boolean linear functions, and show-case how narrow occasionally our understanding is, namely by discussing that comparable results for monotonic functions are a famous open problem.

I will then give a broader introduction to this field, discuss the main research goals, the types of results targeted and the methods typically used. I will finally make some language precise that will help the audience to follow the other talks from this field in this seminar.

## 3.6 Crossover as Repair Mechanism and the Usefulness Self-Adjusting Parameter Settings: The $(1 + (\lambda, \lambda))$ GA

*Carola Doerr (UPMC – Paris, FR)*

**Joint work of** Benjamin Doerr, Carola Doerr, Franziska Ebel

We present an evolutionary algorithm using crossover that performs better than any purely mutation-based algorithm on the Hamming distance problem. The main idea behind our approach is a novel use of crossover as repair mechanism. We also discuss that for this algorithm self-adjusting parameter settings are provably superior to any static parameter choices.

## 3.7 Eco-evolutionary dynamics Modeling evolution without defining fitness/pay-off

*Paulien Hogeweg (Utrecht University, NL)*

In this talk I addressed the following questions earlier raised in this workshop:

- Can we model evolution without prior fitness (or pay-off) definition?
- How/when do functionally differentiated ecosystems evolve
- Self-adaptation in evolution: evolution of genotype-phenotype mapping

Using the RNA world as an example, I showed that, if we include structure evolution can be modelled without apriori fitness or payoff definition: the inherent structure in the RNA model provides the substrate for the evolution, in which evolution chooses its own fitness. Multi-species communities can evolve, in which the various lineages have different roles in the ecosystem (niche creation). Mutation rate plays a crucial role in the type of community which evolves, i.e. mutation rate has qualitative effect rather than just influencing speed and degree of adaptation. Important in these results is spatial pattern formation: higher order of selection emerges automatically. Moreover we see in this system the evolution of genotype-phenotype mapping in such a way that the mutant cloud (i.e. the quasispecies) is shaped by evolution such that it contains non-fit mutants which nevertheless play a crucial role in the evolved ecosystem, and can lead to fast adaptation to novel circumstances.

Finally I discussed the question: "can we derive general conclusion by studying specific (structured) examples". My answer is yes we can. Features as discussed reoccur in many very differently structured models, and can be seen as generic properties of evolution *provided enough degrees of freedom are available to the evolutionary process.*

## 3.8 Evolution of mutation rates

*Kavita Jain (JNCASR – Bangalore, IN)*

Because most mutations are deleterious, the mutation rate can not be too high and an upper bound is provided by error threshold. The mutation rate is not zero either. I will describe our results on the fixation probability of a nonmutator using a branching process and arrive at a lower bound on the mutation rate in a finite population.

## 3.9 Basic concepts of population genetics: Adaptation in rugged fitness landscapes

*Joachim Krug (Universität Köln, DE)*

The purpose of the lecture was to introduce some basic concepts of evolutionary adaptation that appear in similar form in biological and computational contexts. Starting from the standard Wright-Fisher model of finite populations, the stochastic dynamics of fixation was introduced and used to identify the evolutionary regimes of periodic selection and clonal interference. The main part of the lecture was devoted to genotypic fitness landscapes, their empirical basis, probabilistic modeling, and exploration by random adaptive walks. Finally, the effects of recombination on adaptation in rugged fitness landscapes was briefly addressed.

## 3.10 Genetic mechanisms for the advantage of recombination

*Joachim Krug (Universität Köln, DE)*

**Joint work of** Joachim Krug, Su-Chan Park

I will describe two results pertaining to the effect of recombination on the efficacy of evolutionary searches. First, I present a solvable model for the Fisher-Muller effect that predicts a twofold speedup of adaptation in a linear fitness gradient. Second, I discuss a minimal deterministic two-locus model which shows a phase transition as a function of recombination rate. Beyond the transition the model displays bistability in the dynamical systems sense and the escape time from a local fitness peak is infinite.

## 3.11 Negative Drift in Populations

*Per Kristian Lehre (University of Nottingham, GB)*

The expected running times of evolutionary algorithms are often analysed using so-called drift analysis where the current state of the algorithm is mapped to a real-valued potential. Bounds on the running time are derived from the expected change in the potential per generation. However, finding an appropriate potential function is non-trivial, particularly for population-based evolutionary algorithms.

In this talk, I presented an alternative drift theorem that provides tail bounds on the running time of population-based evolutionary algorithms given simple conditions on the variation operator and selection mechanism employed by the algorithm. I outlined the ideas behind the proof, which uses a combination of multi-type branching processes and classical drift analysis.

### References

1 Per Kristian Lehre. Negative Drift in Populations. In *Proc. of Parallel Problem Solving from Nature (PPSN XI)*, LNCS, Vol. 6238, pp. 244–253, Springer, 2011.

## 3.12 Mutation as a computational event

*Adi Livnat (University of Haifa, IL)*

In recent years it has become clear that germline mutation is affected by DNA sequence and structure and by complex biological mechanisms. Can this new empirical knowledge tell us something about how evolution works? I argue that opening up the black box of the nature of mutation affects fundamental concepts in our understanding of evolution, including the role of sex in evolution.

## 3.13 The Slime Mold Computer: Physarum can compute shortest paths

*Kurt Mehlhorn (MPI für Informatik – Saarbrücken, DE)*

The slime mold Physarum can apparently compute shortest paths. Nakagaki, Yamada, and Tóth (Nature 2000) performed the following experiment: They prepared a maze, covered it with Physarum, and provided food at two locations. After a few hours, the slime mold had retracted to the shortest path connecting the two food sources. Tero et al (J. Theoretical Biology, 2007) provided a mathematical model for the dynamics of the slime mold and verified in computer simulations that the model converges to the shortest path. In the first part of the talk, I survey the experiments (a video is available at http://people.mpi-inf.mpg.de/~mehlhorn/ftp/SlimeAusschnitt.webm) and introduce the mathematical model. In the second part, I describe the path towards a proof of convergence. In the third part, I look into the future. The slime mold can also build beautiful networks. How can we understand its network building capabilities?

### References

1 T. Nakagaki, H. Yamada, and A. Tóth. Maze-solving by an amoeboid organism. *Nature*, 407(6803):470, oct 2000.
2 A. Tero, R. Kobayashi, and T. Nakagaki. A mathematical model for adaptive transport network in path finding by true slime mold. *Journal of theoretical biology*, 244(4):553–64, feb 2007.

## 3.14 The long-term response to directional selection

*Tiago Paixao (IST Austria – Klosterneuburg, AT)*

The role of gene interactions in the response to selection has long been a controversial subject; while some have dismissed them as an important influence on adaptation, others have argued that their long-term effects are of high significance. Here, we derive simple and general predictions for the effect of gene interactions on the long-term response to selection from standing variation (excluding new mutations). We show that when the dynamics of allele frequencies are dominated by genetic drift, the long-term response is surprisingly simple, depending only on the initial components of the trait variance, regardless of the detailed genetic architecture. Moreover, we show that this result applies when many gene contribute to fitness.

## 3.15 Natural Selection, Game Theory and Genetic Diversity

*Georgios Piliouras (Singapore University of Technology and Design, SG)*

**Joint work of** Ruta Mehta, Ioannis Panageas, Georgios Piliouras

In a recent series of papers a strong connection has been established between standard models of sexual evolution in mathematical biology and Multiplicative Weights Updates Algorithm, a ubiquitous model of online learning and optimization. These papers show that mathematical models of biological evolution are tantamount to applying discrete replicator dynamics, a close variant of MWUA on (asymmetric) partnership games. We show that in the case of partnership games, under minimal genericity assumptions, discrete replicator dynamics converge to pure Nash equilibria for all but a zero measure of initial conditions. This result holds despite the fact that mixed Nash equilibria can be exponentially (or even uncountably) many, completely dominating in number the set of pure Nash equilibria. Thus, in haploid organisms the long term preservation of genetic diversity needs to be safeguarded by other evolutionary mechanisms, such as mutation and speciation.

## 3.16 Benefits of Crossover in Combinatorial Search

*Adam Prugel-Bennett (University of Southampton, GB)*

Crossover often proves to be a powerful research tool in finding good solutions to combinatorial optimisation problems. The landscape of such problems are very complex so that understanding why crossover is beneficial is difficult. To help elucidate possible mechanisms where population-based search, particularly using crossover, is beneficial we consider three 'toy' problems. The first demonstrates how a hybrid algorithm combining crossover with local search can solve a problem by combining building blocks. The second problem looks at how a population can efficiently explore a plateau region, while the third problem looks at how a population can solve a problem despite a very large of noise. In each case these problems require super-polynomial time for local search. The talk describes empirical work which has previously been published in [1]. We also describe a run time analysis carried out by Jonathan Shapiro, Jonathan Rowe and the author. Preliminary results have been published in [2].

### References

1 A. Prügel-Bennett. Benefits of a population: Five mechanisms that advantage population-based algorithms. *Trans. Evol. Comp*, 14(4):500–517, Aug. 2010.
2 A. Prugel-Bennett, J. Rowe, and J. Shapiro. Run-time analysis of population-based evolutionary algorithm in noisy environments. In *Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII* (FOGA'15), pp. 69–75, New York, NY, USA, 2015. ACM.

## 3.17 Mixing time of stochastic evolutionary dynamics

*Piyush Srivastava (California Institute of Technology – Pasadena, US)*

We study the time it takes for stochastic evolutionary dynamics to achieve a stationary steady state (i.e., its mixing time). We prove that the the mixing time of a wide class of such dynamics grows only logarithmically in the size of the state space. The class of dynamics we study includes as a special case a finite population stochastic version of the quasispecies model of molecular evolution.

Such dynamics, in particular the finite population quasispecies model, have been used to study the evolution of viral populations with applications to drug design strategies countering them. Here the time it takes for the population to reach a steady state is important both for the estimation of the steady-state structure of the population as well in the modeling of the treatment strength and duration. Our result, that such populations exhibit rapid mixing, may be seen as a theoretical justification for numerical simulations that use the above approach.

## 3.18 Slime Mold Dynamics for Flows and Linear Programming

*Damian Mateusz Straszak (EPFL Lausanne, CH)*

**Joint work of** Damian Mateusz Straszak, Nisheeth K. Vishnoi

We study dynamics inspired by Physarum polycephalum (a slime mold) for solving network flow problems and linear programs. These dynamics are arrived at by a local and mechanistic interpretation of the inner workings of the slime mold and a global optimization perspective has been lacking even in the simplest of instances. Our first result is an interpretation of the dynamics as an optimization process. We show that Physarum dynamics can be seen as a steepest-descent type algorithm on a certain Riemannian manifold. Moreover, we prove that the trajectories of Physarum are in fact paths of optimizers to a parametrized family of convex programs, in which the objective is a linear cost function regularized by an entropy barrier. Subsequently, we rigorously establish several important properties of solution curves of Physarum. We prove global existence of such solutions and show that they have limits, being optimal solutions of the underlying problems.

## 3.19 The impact of genetic drift on the runtime of simple estimation-of-distribution algorithms

*Dirk Sudholt (University of Sheffield, GB) and Carsten Witt (Technical University of Denmark – Lyngby, DK)*

In this talk, we will consider simple estimation-distribution algorithms, including the so-called compact GA (cGA), on the classical OneMax benchmark problem. The perspective is runtime analysis. We will derive lower bounds on the runtime of the cGA, and discuss how genetic drift affects the optimal parameter setting of the algorithm.

## 3.20 Understanding Diversity and Recombination in Simple Evolutionary Algorithms

*Dirk Sudholt (University of Sheffield, GB)*

I presented an open problem in runtime analysis: trying to understand the benefit of recombination in the context of a simple ($\mu$+1) Genetic Algorithm on the OneMax problem. Experiments suggest that crossover is beneficial, especially for large populations. Yet despite the seeming simplicity of the setting, existing runtime analyses are restricted to small populations and often ignore the initial diversity in the population. Can we use techniques from Population Genetics or other fields to show that Genetic Algorithms exploit this diversity efficiently through recombination?

## 3.21 The speed of adaptation of complex traits

*Barbora Trubenova (IST Austria – Klosterneuburg, AT) and Jorge Perez Heredia (University of Sheffield, GB)*

Many adaptations are complex; they involve large numbers of genes that may interact in non-additive ways. The efficacy of natural selection to produce these adaptations has been a long-standing question. In particular, how long does it take for natural selection to evolve such complex adaptations?

Here, we address this question by making use of tools from computer science to characterize the time it takes for a population to reach a particular genotype sequence. We focus on how this time scales with the complexity of the trait, and find the conditions on selection strength that enable efficient adaptation. We quantify the 'cost of complexity' on several classes of fitness landscapes and show that this cost depends strongly on the details of the genetic architecture. We distinguish between polynomial and exponential scalings as efficient and inefficient adaptation and show that there is sharp threshold between the two for populations in the weak mutation regime. Moreover, we show that even when the loci contributing to the trait interact in an extreme form of epistasis, the time required to reach the fitness peak scales polynomially.

## 3.22 Evolutionary Dynamics

*Nisheeth K. Vishnoi (EPFL Lausanne, CH)*

In this talk we survey two fundamental models in evolutionary dynamics: the infinite population Quasispecies model and the stochastic, finite population Wright-Fisher model. Subsequently, we discuss the notions of error threshold, time to convergence and mixing time rigorously.

## 3.23 Coalescent trees, induced subtrees, their topology and site frequency spectrum

*Thomas Wiehe (Universität Köln, DE)*

Genealogies, when viewed forward-in-time, are equivalent to certain types of branching processes, for instance the Yule process. We are interested in measuring topological properties of binary trees which are generated by a Yule process. An easily accessible topological parameter is tree balance at (upper) internal nodes. Tree balance affects the mutation site frequency spectrum (SFS) and can introduce a bias in typical SFS-based statistics. For practical applications it is of interest to understand the dependence of topological properties – and hence of the SFS – in induced subtrees, when conditioned on the topology of a supertree. Furthermore, it is of interest to understand the impact of recombination in changing tree topologies. Last, I will shortly describe a 'topological' measure of linkage disequilibrium.

## Participants

- Ellen Baake
Universität Bielefeld, DE
- Nick Barton
IST Austria – Klosterneuburg, AT
- Arnab Bhattacharyya
Indian Institute of Science – Bangalore, IN
- Erick Chastain
Rutgers Univ. – Piscataway, US
- Duc-Cuong Dang
University of Nottingham, GB
- Harold P. de Vladar
Parmenides Center for the Study of Thinking, DE
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Carola Doerr
UPMC – Paris, FR
- Tobias Friedrich
Hasso-Plattner-Institut – Potsdam, DE
- Paulien Hogeweg
Utrecht University, NL
- Kavita Jain
JNCASR – Bangalore, IN
- Timo Kötzing
Hasso-Plattner-Institut – Potsdam, DE
- Joachim Krug
Universität Köln, DE
- Per Kristian Lehre
University of Nottingham, GB
- Adi Livnat
University of Haifa, IL
- Kurt Mehlhorn
MPI für Informatik – Saarbrücken, DE
- Tiago Paixao
IST Austria – Klosterneuburg, AT
- Ioannis Panageas
Georgia Institute of Technology – Atlanta, US
- Jorge Perez Heredia
University of Sheffield, GB
- Georgios Piliouras
Singapore University of Technology and Design, SG
- Adam Prugel-Bennett
University of Southampton, GB
- Jonathan L. Shapiro
University of Manchester, GB
- Piyush Srivastava
California Institute of Technology – Pasadena, US
- Damian Mateusz Straszak
EPFL Lausanne, CH
- Dirk Sudholt
University of Sheffield, GB
- Andrew M. Sutton
Hasso-Plattner-Institut – Potsdam, DE
- Barbora Trubenova
IST Austria – Klosterneuburg, AT
- Paul Valiant
Brown Univ. – Providence, US
- Nisheeth K. Vishnoi
EPFL Lausanne, CH
- Thomas Wiehe
Universität Köln, DE
- Carsten Witt
Technical Univ. of Denmark – Lyngby, DK
- Xin Yao
University of Birmingham, GB