Preface

When Charles Darwin published *The Origin of Species* 160 years ago, he initiated evolution as a substantial research direction in the biological sciences. The mathematical side of evolutionary theory was founded in the 1920s by Ronald Fisher, Sewall Wright, and John Haldane, and led to the modern synthesis of Darwinian evolution and Mendelian genetics. Today, understanding the processes of biological evolution in greater depth is a continuing challenge to both biological and mathematical research.

Evolution is a complex phenomenon driven by various processes, such as mutation and recombination of genetic material, reproduction of individuals, and selection of favourable types. The outcome of their interplay is impossible to predict without the substantial use of mathematical models and methods. Over many decades, much of this modelling and analysis took place on a deterministic level, using classical dynamical systems and differential equations, and this has led to an elaborate theory. However, the processes of evolution have intrinsically random elements, which give rise to a wealth of phenomena that cannot be explained by deterministic processes. Examples of such effects are the loss of genetic variability due to randomness in reproduction, and the emergence of random genealogies.

The desire to model such phenomena has inspired the theory of stochastic processes since its very beginning. *Wright–Fisher diffusion*, *Feller's branching diffusion* and *Kingman's coalescent* not only had great success as biological models, but also became prototypes of mathematical objects with a rich and fascinating structure. The fact that these concepts and their refinements turned out to be both natural and convincing from the modelling point of view has greatly motivated researchers to analyse them in depth.

Currently, the area of biological evolution attracts increasing attention worldwide, both from an experimental and a theoretical point of view. Keeping up with the increasing complexity of the models provides great challenges for probability theory. It requires to go beyond the long-standing prototype models of mathematical population genetics and to systematically explore the various elementary processes (also referred to as *evolutionary forces*), alone and in combination.

From a mathematical point of view, the objects of evolution are populations of individuals. The *type* of an individual can be described at varying degrees of complexity. It can be the type at a single gene (or *locus* in biological language), the genetic sequence at a set of multiple loci or an entire genome, or some aspect of the *phenotype* (the collection of traits) of an individual, possibly together with the individual's geographical (or more abstract) location. To a certain degree of abstraction, a population can thus be modelled as (a vector of) type frequencies, or as a measure on type space. Its (random) dynamics is governed by the various evolutionary forces. Individuals *reproduce*, transmitting their genetic material to their offspring. In sexual reproduction, this goes along with *recombination*, which combines a maternal and a paternal (geno)type into the "mixed" type of the offspring. In addition, along the individual lines of descent, types may be changed by *mutation*; and individuals may *migrate* between locations. The individual net offspring production rate (*fitness*) can depend on the type as well as on the environment, for instance on the state of the entire population due to *competition* for resources. If reproduction rates vary across types, fitter individuals flourish at the expense of less fit ones. This is known as *selection*.

These evolutionary forces form the basis for the two main theories in the area: *population genetics* (the theory of evolution at the level of individual genotypes within populations) and *population dynamics* (its counterpart at the level of phenotypes). A conceptual and methodological anchor for both of them is the topic of *random genealogies*; they are typically related to the corresponding population models by way of *duality*.

The aforementioned prototype models have been extended to include effects such as selection and recombination. Here, one of the central issues has been, and still is, to make random genealogies available in these complex situations. Beside their benefits for applications, the richness of the various models at the interface of probability and evolution (such as Fleming–Viot and ancestral processes with high offspring variation, coalescents with spatial and genetic structure, and individual-based ecological models) is leading to the exploration of new mathematical structures, and hence gives new impulses to probability theory and stochastic analysis as well.

This was the motivation for a large group of researchers from all over Germany (and one from the Netherlands) to join their efforts in the Priority Programme *Probabilistic Structures in Evolution* (DFG-SPP-1590) over two funding periods from 2011–2020. The current volume is a collection of survey articles that grew out of this work. They span the following topics.

Evolution in random fitness landscapes. The fitness landscape encodes the mapping of genotypes to fitness. In Chapter 1, Joachim Krug explores the structure of various kinds of random fitness landscapes by investigating the existence of mutational paths with non-decreasing fitness connecting distant genotypes; this translates into a problem of *accessibility percolation*. In Chapter 2, Wolfgang König investigates *branching random walks* on such landscapes, that is, the movement (via mutation) combined with birth and death events at rates that depend on the current position in the landscape.

Stochastic processes in models of population genetics and population dynamics, forward in time. The *individual-based* modelling of the evolutionary forces leads to processes that describe the evolution of type frequencies that may then be analysed via suitable limit theorems for various scales of time and population size. In this vein, Ellen Baake and Anton Wakolbinger describe and analyse two population-genetic models of microorganisms under selection (Chapter 3): *Lenski's long-term evolution experiment*, that is, experimental evolution under directional selection, and *pathogen evolution* under balancing selection. The population genetics of bacteria is also the theme of Rolf Backofen and Peter Pfaffelhuber (Chapter 4) – here a mutation model for the *CRISPR-Cas (clustered regularly interspaced short palindromic repeats) system* is developed, together with a classification of such systems via bioinformatics methods.

In Chapter 5, Martin Hutzenthaler and Dirk Metzler formulate an individual-based, spatially structured *predator-prey model with an altruistic defence trait* in the prey population, and analyse the conditions under which this trait will persist. Wolfgang Stephan and Aurélien Tellier investigate the impact of the randomness in reproduction (known as genetic drift in biology) and mutation on the genetic structure of populations that experience *host-parasite coevolution* (Chapter 6). Chapter 7 by Anton Bovier reviews the stochastic individual-based model of *adaptive dynamics* (a concept for the joint description of population regulation and genetic change) and discusses how different scaling limits can be obtained by taking limits of large populations, small mutation rate, and small effect of single mutations together with appropriate time rescaling.

Stochastic population models and random genealogies. *The* fundamental construct of a random genealogy is Kingman's coalescent (1982), which describes the ancestral relationships between the individuals in a random sample from a population with small family sizes in the absence of selection, recombination, and further evolutionary forces. Substantial complications arise when one goes beyond these "standard" assumptions. Matthias Birkner and Jochen Blath (Chapter 8) review recent progress on *coalescents with multiple and simultaneous multiple mergers*, which emerge in populations with highly-skewed offspring distribution, and they discuss questions of inference from genetic data. Inference for multiple-merger coalescents is also the theme of Chapter 9 by Fabian Freund, who additionally considers effects of *variable population size*. Anja Sturm reviews what changes when the population is *diploid* rather than haploid, that is, when individuals carry two copies of each gene which they inherit from two distinct parent individuals (Chapter 10). Götz Kersting and Anton Wakolbinger (Chapter 11) investigate *functionals of multiple-merger coalescents*, such as the total external branch length and the time to the most recent common ancestor, and discuss *evolving coalescents*.

Multiple-merger coalescents may also appear in populations with dormancy, where individuals switch between an active state, and an inactive one in which they do not reproduce, thus generating a seed bank of dormant individuals. Such models, along with the resulting genealogies, are developed and analysed by Jochen Blath and Noemi Kurt in Chapter 12. The effect of a hierarchical spatial structure on the "volatility" (i.e. the strength of fluctuations) of the type frequencies is investigated by Greven and den Hollander (Chapter 13) and compared with spatial effects in a model with highly-skewed offspring distribution, but without dormancy.

Lines of descent are further shaped by interaction within a population. Matthias Birkner and Nina Gantert investigate *spatial population models with local regulation* and explain how an ancestral lineage can be interpreted as a random walk in a dynamic random environment, and how, in particular in the one-dimensional case, the whole collection of ancestral lineages converges to the Brownian web (Chapter 14). In Chapter 15, Jochen Blath and Marcel Ortgiese study the *interface* (that is, the region of coexistence) in the *symbiotic branching model* with nearest-neighbour migration, with methods of duality that may in some cases be interpreted as tracing back ancestral lines. Anja Sturm and Anita Winter provide a formal framework for a type-dependent *branching model with mutation and competition*, and the corresponding genealogies (Chapter 16).

Powerful concepts to deal with recombination or selection are given by ancestral structures that are not trees, but branching-coalescing graphs, namely the *ancestral recombination graph* and the *ancestral selection graph*, respectively; the "true" genealogical trees (for each individual genetic locus) are embedded into these graphs. Ellen Baake and Michael Baake obtain a transparent solution of the deterministic recombination equation by means of its dual, namely an *ancestral partitioning process* derived from the ancestral recombination graph in a law of large numbers regime (Chapter 17). Julien Dutheil (Chapter 18) reviews methods based on the ancestral recombination graph that allow to *infer recombination parameters* from population-genetic data sets, in particular when the recombination rates vary across the genome. Martin Hutzenthaler and Peter Pfaffelhuber study the *genealogical trees embedded into the ancestral selection graph*, in particular their lengths relative to the case without selection, and their behaviour when selection fluctuates (Chapter 19).

Trees as such. Last not least, the properties of trees as such are crucial, regardless of their meaning as genealogies in population models. Thomas Wiehe reviews *combinatorial and topological properties of binary tree*s, the probability distributions in tree space under different generation mechanisms, as well as the effects of the pruning and grafting operations that result when the tree evolves (Chapter 20). The volume concludes with Chapter 21, in which Anita Winter considers *binary tree topologies as algebraic objects*, equips them with a probability measure that allows to sample leaves from the tree, and describes the topology and statistics of the resulting subtrees.

Each chapter has been refereed by two experts (anonymous for the chapter's authors), one from outside and one from inside the Priority Programme. Our thanks go to all of them.

The Priority Programme would not have been thinkable without our fellow members on the steering committee, namely Anton Bovier, Andreas Greven, and Frank den Hollander, who went with us through the adventure of designing the concept of, applying for, and guiding the Programme over so many years.

It is our pleasure to thank those people who made the Priority Programme run smoothly, including Nicole Althermeler, Mareike Esser, and Sebastian Probst, who took responsibility for organisational matters from running the web page to assisting with meetings; and Jochen Röndigs, who managed the general administration including about two thousand (!) travel reimbursements. Special thanks go to Britta Heitbreder for supporting the coordinator (E. B.) in every possible way, and for being the LaTeX wizard who shouldered the complex endeavour of preparing, correcting, and revising this volume from titlepage to index. Last not least, we thank the German Research Foundation (DFG) for having invested into this project.

Ellen Baake Anton Wakolbinger