Biological evolution across scales: mathematical modelling and statistical inference

BEvAS

Bernoulli Center of fundamental studies, EPFL, Switzerland

April 17th - April 21st, 2023

Organizers

Alia Abbara, Celia García Pareja, Umberto Lupo

Book of abstracts

Frederic Alberti (Universität Bielefeld, Germany)

Labelled partitions in action: recombination, selection, mutation, and more

The investigation of mathematical models that describe the interplay of genetic recombination with other forces of evolution such as selection and/or mutation is among the major challenges of mathematical population genetics. In particular, the *deterministic selection-recombination equation*, a high-dimensional system of nonlinear differential equations, has long defied all attempts at its solution. Recently, we solved this equation in the special case of a single selected locus linked to an arbitrary number of neutral loci, and single-crossover recombination [1]. The key to our solution is a connection (via a duality of Markov processes) between the solution of the selection-recombination equation, forward in time, and a stochastic process that describes the random evolution of the genealogy of a sample, backward in time. This process is derived from the *ancestral selection-recombination graph*, which in turn combines the *ancestral selection graph* introduced by Krone and Neuhauser with the ancestral *recombination* graph. Ultimately, this led to an explicit representation of the solution in terms of iterated integrals. This contribution complements our aforementioned work [1] by exploring the interaction between recombination and additional evolutionary forces in a more general setting. We will see that, under certain natural assumptions, the genealogy of a sample can be described as a partitioning process whose blocks carry independent Markovian labels. Intuitively speaking, the partitioning process describes the fragmentation of the genetic material of an individual across its ancestors, while the Markovian labels encode the evolution under the additional evolutionary forces. On the one hand, this deepens our insight into the mathematical structure underlying our results in [1] and on the other, it allows us to construct explicit solutions for a wider class of models. We will illustrate these ideas with the example of the selection-mutation-recombination equation.

[1] F. Alberti and E. Baake, Solving the selection-recombination equation: Ancestral lines and dual processes, Documenta Mathematica 26 (2021), 743–793.

Erik Aurell (KTH Royal Institute of Technology, Sweden)

Statistical Genetics and Direct Coupling Analysis in and out of Quasi-Linkage Equilibrium

Statistical genetics is an interdisciplinary area between statistical physics and population biology. I will briefly review its main problems and results, and then focus on the phase of quasi-linkage equilibrium (QLE). This phase, which has many similarities to a state of thermal equilibrium in statistical mechanics, was discovered by Kimura in a two-locus two-allele model, and was later extended and generalized to the global genome scale by Neher and Shraiman. The resulting Kimura-Neher-Shraiman (KNS) theory describes a population evolving due to the mutations, recombination, natural selection and possibly genetic drift. A QLE phase exists at sufficiently high recombination rate (r) and/or mutation rates (μ) with respect to selection strength. I will show how in QLE it is possible to infer the epistatic parameters of the fitness function from the knowledge of the (dynamical) distribution of genotypes in a population by applying the technique of Direct Coupling Analysis (DCA). I will further consider the breakdown of the QLE regime at high enough selection strength, and identify and characterize a new phase which of non-random coexistence (NRC) where variability persists in the population without either fixating or disappearing.

Maria Anisimova (ZHAW Zurich University of Applied Sciences, Switzerland)

Uncovering the potential of modeling indel evolution in genomic sequence analyses

Multiple sequence alignment (MSA) and phylogeny inference tasks are crucial in genomics and molecular evolution. Changes between homologous characters are typically modelled by a Markov substitution model. In contrast, the dynamics of indels usually are not modelled explicitly. This is because the computation of the marginal likelihood, even under most simple models, has exponential time complexity in the number of taxa. Yet, the failure to model indel evolution may lead to artificially short MSAs due to the biased indel placement, inconsistencies with phylogenetic relationships. Equally, ignoring or incorrectly treating indels has consequences for the accuracy of phylogeny inference. Luckily, the classical indel model TKF91 can be modified to describe indel evolution on a phylogeny via a Poisson indel process (PIP). This allows us to compute the joint marginal probability of an MSA and a tree in linear time. We developed a 3D dynamic programming algorithm for progressive MSA inference under PIP in polynomial time. Our method is the first polynomial time progressive aligner with a rigorous mathematical formulation of indel evolution. Maximum likelihood phylogeny inference under PIP can contribute to improved accuracy, with a particularly large margin for small datasets. Further, we implemented the PIP model for ancestral sequence reconstruction and demonstrated its advantages compared to other existing methods. Finally, we propose how to resolve the dependency between inferences of MSAs and phylogeny by reconstructing them jointly within the same likelihood framework under the PIP model. To demonstrate the applicability of our methods, we use them to systematically study the indel variation in the HIV-1 env gene that encodes the two glycoproteins, gp120 and gp41. In particular, gp120 is the primary target for neutralizing antibodies generated as a response to HIV infection or vaccination, and its rapid mutation rate contributes to the virus's ability to evade host immune response. Therefore a better understanding of the genetic variability of the region could aid effective vaccine development.

Florence Bansept (Max Planck Institute for Evolutionary Biology, Plön, Germany)

Evolution in biphasic life cycles

Microbial communities extend the host functional repertoire, thus making the host and its associated microbes a functional unit. We are only beginning to decipher how host and microbe fitnesses are intertwined: while it is now clear that the microbiota has a vast potential to affect the host physiology, less focus has been put to the microbial perspective, i.e. to understand what benefit or cost can microbes retrieve from their interaction with their host. In contrast to the common hypotheses of strong and continued coevolution proposed to explain the emergence of such elaborate symbioses, we focus here on the steps that can lead a microbial population to transition from a free-living life-style to an association with a host. In particular, we aim at understanding what selection pressures apply to microbes following a biphasic life cycle, in which they can regularly transit in and outside a host. We study three simple models of such biphasic life cycles. In the first model, we study a homogeneous microbial population transiting between a host and its environment and perform a sensitivity analysis to show the existence of two different regimes: one where the effect of migration from the environment to the host dominates, and a second where the within-host replication rate matters most. The second model is an SI-inspired compartmental model, which accounts for the habitats' dynamics. We show that microbial propagation across habitats depends on the product of the transmissibilities, which we propose as a new holistic measure of microbial fitness - bearing similitudes with the R0 of epidemiology. In the third model, we combine microbial population and habitats dynamics and derive predictions that are consistent with experimental observations of an increased ability to form biofilms in bacteria evolved in biphasic conditions with C. elegans.

Pierre Barrat-Charlaix (University of Basel, Switzerland)

Ecology of host and pathogen and limited predictability of evolution

Seasonal influenza viruses repeatedly infect humans in part because they rapidly change their antigenic properties and evade host immune responses, necessitating frequent updates of the vaccine composition. A better understanding of this evolution could allow for better predictions and would be important for vaccine design. We investigated the predictability of frequency dynamics and fixation of amino acid substitutions and found that the current frequency was the strongest predictor of eventual fixation, as expected in neutral evolution. Simulations of models of adapting populations, in contrast, show clear signals of predictability. This indicates that the evolution of influenza HA and NA, while driven by strong selection pressure to change, is poorly described by common models of directional selection such as travelling fitness waves. In consequence, we propose an alternative way to view influenza's evolution where due to continuous changes in human immune response, adaptive mutations do not fix but rather reach an equilibrium frequency. This leads to think about the viral population in terms of equilibrium between strains rather than of fixed individual fitness values. We explore the rich phenomenology of this model which qualitatively matches observations about influenza, such as ladder-like genealogies and rapid but unpredictable evolution.

John Barton (University of Pittsburgh, USA)

Using time series data to learn about evolution, from the lab to the globe

In recent years, massive sequencing efforts have allowed us to observe evolution in real time. My lab studies how to use sequence data collected over time to understand evolutionary dynamics. In this talk I'll give a few examples of our work studying evolution at wildly divergent scales. First, I'll discuss short-term evolution in deep mutational scanning experiments, where we've used new theoretical approaches to interpret data much more reliably than existing state-of-the-art methods. Second, I'll talk about how we've used epidemiological models to study the evolution of SARS-CoV-2 over the course of the pandemic, uncovering mutational drivers of increased viral transmission.

Barbara Bravi (Imperial College London, UK)

Machine learning models for antigen immunogenicity and T-cell recognition

Antigen immunogenicity and the specificity of binding of T-cell receptors to antigens are key properties underlying effective immune responses. Identifying immunogenic antigens, as well as antigen-specific T-cell receptors, is therefore crucial to vaccine and cancer immunotherapy design. In this talk, I will discuss a set of flexible and easily interpretable methods that we have recently developed based on the machine learning scheme of Restricted Boltzmann Machines (RBM). Such scheme allowed us first to build models of antigen presentation by the human leukocyte antigen class I proteins and antigen immunogenicity, which can be used to reconstruct the underlying molecular motifs and as predictors of viral epitopes and cancer neoantigens. I will next introduce RBM-based models of the complementary process of recognition by T cells of presented antigens, which are able to discriminate responses specific to different epitopes and to detect signatures of response at the T-cell repertoire level.

Xiaowen Chen (Ecole Normale Supérieure de Paris, France)

Generalized Glauber dynamics for inference of collective behavior

Large interacting systems in biology often exhibit emergent dynamics, such as coexistence of multiple time scales, manifested by fat tails in the distribution of waiting times. While existing tools in statistical inference, such as maximum entropy models, reproduce the empirical steady state distributions, it remains challenging to learn dynamical models. We present a novel inference method, called generalized Glauber dynamics (GGD). Constructed through a non-Markovian fluctuation dissipation theorem, generalized Glauber dynamics tunes the dynamics of an interacting system, while keeping the steady state distribution fixed. We motivate the need for the method on real data from Eco-HAB, an automated habitat for testing behavior in groups of mice under semi-naturalistic conditions, and present it on simple Ising spin systems. The dynamical inference can be performed by parameterizing the memory kernel and by an expectation-maximization algorithm. We show the applicability of the generalized Glauber dynamics for experimental data, by inferring dynamical models of social interactions in a group of mice that reproduce both its collective behavior and the long tails observed in individual dynamics.

Giancarlo Croce (UNIL, Switzerland)

Exploiting evolutionary patterns in homologous protein sequences to predict short-term polymorphisms: applications to E. coli and SARS-CoV-2

With the advancement of high-throughput sequencing, public databases now contain hundreds of thousands of genomes. Homologous amino-acid sequences provide valuable insights into how proteins have evolved over millions of years. We asked whether this information on long-term evolution could help us predict which amino acids are more likely to mutate in the short term. To do so, we employed Direct-Coupling Analysis (DCA), a data-driven approach that harnesses evolutionary patterns present in homologous protein sequences, to predict which protein residues are more likely to be polymorphic in a set of recently diverged strains. We tested the effectiveness of our approach using two cases: (1) analyzing E. coli data, we highlighted the crucial role of genetic context in predicting mutation effects and that epistatic interactions strongly constrain tolerance of many amino-acid ($30\% - 50\%$ of all the proteome) sites to polymorphisms; and (2) leveraging the vast amount of COVID-19 data, we used DCA to identify positions in the SARS-CoV-2 viral proteome that are more likely to undergo mutations. Our statistical model, trained on sequence data from all coronaviruses and the first known strain of SARS-CoV-2, accurately predicted polymorphic residues that have mutated during the pandemic. We also combined our mutability predictions with emerging immune response data to identify a restricted set of residues that are likely to mutate in the future and may induce immune escape, potentially being overrepresented in current and future SARS-CoV-2 variants of concern. Overall, our study provides novel insights into the use of homologous amino-acid sequences for predicting short-term evolution, suggesting that DCA can be a useful tool in anticipating the emergence of new strains and in tracking and monitoring protein evolution in species for which sufficient sequence data are available.

1)Vigué L*., Croce G*. et al. "Deciphering polymorphism in 61,157 Escherichia coli genomes via epistatic sequence landscapes." Nature Communications (2022)
2) Rodriguez-Rivas J*, Croce G*. et al. "Epistatic models predict mutable sites in SARS-CoV-2 proteins and epitopes." Proceedings of the National Academy of Sciences (2022)

Pete Czuppon (Westfälische Wilhelms-Universität Münster, Germany)

Within-host dynamics of antibiotic resistance

The use of an antibiotic creates a beneficial environment for the evolution of antibiotic-resistant bacterial strains. The dosage of the antibiotic drug during plays an important role during this process. Experimental and theoretical studies have investigated the drug dose that minimizes the risk of resistance evolution over the course of treatment of an individual, showing that the optimal dose will either be the highest or the lowest drug concentration possible to administer; however, no analytical results exist that help decide between these two extremes. To address this gap, we develop a stochastic mathematical model of bacterial within-host dynamics under antibiotic treatment. We explore various scenarios of density regulation (bacterial density affects cell birth or death rates), and antibiotic modes of action (biostatic or biocidal). In this context, we derive analytical results for the survival probability of the resistant subpopulation until the end of treatment, the size of the resistant subpopulation at the end of treatment, the carriage time of the resistant subpopulation until it is replaced by a sensitive one after treatment. Importantly, we obtain an analytical prediction of the antibiotic concentration that maximizes the survival of resistant cells, which may help to decide which drug dosage (not) to administer.

Christian Dallago (NVIDIA Europe and TU München, Germany)

Protein language models for protein science, design and engineering: an overview and outlook.

Biology has recently seen incredible advances supported by methods from Deep Learning. In particular, protein science and engineering has benefitted both from advances in downstream deep learning models, predicting aspects of protein structure and function, and foundational models often based on the Transformer architecture, providing new ways to encode protein information in computational form. The staggering advances in accurate Deep Learning methods for protein science have led to a sprouting of new ways in which Deep Learning can be used to predict and design proteins. This talk will examine the principles that led to foundational models for protein science, the validation of these models through predictions of function and structure, and venues for their application in protein engineering and design accompanied by practical examples.

Florence Débarre (Sorbonne Université - CNRS, France)

Spatial spread of a gene drive

Understanding the temporal spread of gene drive alleles – alleles that bias their own transmission – through modeling is essential before any field experiments. In this talk, I will present recent results on a deterministic reaction-diffusion model describing the interplay between demographic and allelic dynamics, in a one-dimensional spatial context. This work was led by Léna Kläy and done in collaboration with Léo Girardin and Vincent Calvez. I will present our results on the traveling wave solutions, and more specifically, on the speed of gene drive invasion (if successful), and the effect of demographic dynamics on drive spread.

Chaitanya Gokhale (University of Würzburg and Max Planck Institute for Evolutionary Biology, Plön, Germany)

Ecological determinants of structure and function in living systems

Evolution requires a conducive ecology. From protocellular to societal, networks of living systems are complex and multiscale. Thus they also experience ecologies spanning multiple timescales. Essential resources, such as nutrients, water, and space, often fluctuate in regular cycles linked to rhythmic physical processes like the movement of the earth. Such cyclical variation in ecology exerts selective pressure on the interactions, population dynamics, abundance, and richness in a range of systems. However, an interconnected system must first persist long enough to be a subject of natural selection. I will describe the development of a mathematical (and experimental) system of interconnected replicators (yeast) designed to study persistence in dynamic ecologies. We emulate Darwin's "warm little ponds" with an ecology governed by transient compartmentalisation. Persistence then provides a fertile test bed for developing evolutionary processes, such as multilevel selection. Endogenisation of cyclical ecological processes, such as the development of life cycles and circadian and circalunar cycles, may provide an evolutionary advantage to communities. While endogenisation helps the communities escape the pressures of one ecological niche, it makes them susceptible to others - an essence of multilevel selection. In closing, I will discuss how this research framework expands our understanding of co-option between living and non-living processes and how this understanding helps us design better intervention strategies using synthetic microbial communities.

Anastasia Ignatieva (University of Oxford, UK)

Threading new data into reconstructed genealogies

The evolutionary history of a sample of sequences is fully captured by their genealogy in the form of an ancestral recombination graph (ARG), which contains information on shared ancestors and the evolutionary events that have created the observed present-day genetic diversity. The notoriously difficult problem of reconstructing plausible genealogies for a given input dataset has seen significant recent progress, with several powerful and scalable tools now available. The related problem of adding a new sequence to a fixed reference ARG has also drawn significant interest, in the context of constructing proposals for MCMC-based ARG sampling schemes, or reconstructing ARGs from scratch by adding in data sequentially. We present a new approximate but principled method for accurately and quickly threading a sequence into an ARG, which naturally handles common issues such as the presence of recurrent mutation, missing data and sequencing errors, and allows for types of data which most ARG reconstruction methods do not readily accommodate, such as unphased genotypes and non-contemporary samples. This has myriad applications, for instance in adding low-coverage ancient DNA data into previously inferred large-scale human genealogies, allowing for the application of genealogy-based inference methods. We use the timing of the mutations that are shared by the reference ARG and the new sequence, to construct a reduced state space of possible edges with which the new sequence can coalesce at each genomic location, from which the threading position is chosen. Simulation studies with human-like parameters demonstrate that the method maintains excellent accuracy while scaling well (sub-linearly in the sample size), for both simulated and reconstructed reference ARGs, and running in a fraction of the time compared to de novo genealogy reconstruction.

Laurent Jacob (Université Lyon 1, France)

Phyloformer: towards fast and accurate phylogeny estimation with self-attention networks

An important problem in molecular evolution is that of phylogenetic reconstruction, that is, given a set of sequences descending from a common ancestor, the reconstruction of the binary tree describing their evolution from the latter. State-of-the-art methods for the task, namely Maximum likelihood and Bayesian inference, have a high computational cost, which limits their usability on large datasets. Recently researchers have begun investigating deep learning approaches to the problem but so far these attempts have been limited to the reconstruction of quartet tree topologies, addressing phylogenetic reconstruction as a classification problem. We present here a radically different approach with a transformer-based network architecture that, given a multiple sequence alignment, predicts all the pairwise evolutionary distances between the sequences, which in turn allow us to accurately reconstruct the tree topology with standard distance-based algorithms. The architecture and its high degree of parameter sharing allow us to apply the same network to alignments of arbitrary size, both in the number of sequences and in their length.

Paul Jenkins (University of Warwick, UK)

Some properties of the Wright-Fisher diffusion

Evolution is a change in allele frequencies over time. How should we model this? If genetic drift, which can introduce stochastic variation on short timescales, plays a role then it is appropriate to use a diffusion process. This idea was introduced and studied by Wright, Fisher, and Kimura, among others. In this talk I will give an overview of what is called the Wright-Fisher diffusion and show you some of the kinds of questions we can answer with it. I will focus on some recent work which is important in the problem of statistical inference from time series data. Mathematically the question is: when is one diffusion path measure absolutely continuous with respect to another? - or, in simpler terms: if I observe the path of an allele frequency over time, is there anything I can say with certainty about the underlying parameters? It turns out that this question is very closely related to the behaviour of the diffusion at the time an allele either goes extinct or goes to fixation.

Parul Johri (University of North Carolina, Chapel Hill, USA)

The role of non-adaptive evolutionary processes in shaping genomic variation

The question of the relative evolutionary roles of adaptive and nonadaptive processes has been a central debate in population genetics for nearly a century. While advances have been made in the theoretical development of the underlying evolutionary models, as well as in statistical methods for estimating their parameters from large-scale genomic data, a framework to account for the joint effects of multiple evolutionary processes is still lacking. We show how segregating deleterious mutations affect patterns of variation at "neutral" sites near functional genomic components and how not accounting for linked effects of selection affects inference in population genetics. We demonstrate a potential solution to this fundamental problem– a novel statistical framework for jointly inferring the contribution of the relevant selective and demographic parameters, accounting for effects of selection on linked sites. Our approach represents an appropriate baseline model for inference in population genetics at large which is necessary to accurately assess the role of adaptive processes in shaping genomic variation and is especially relevant for species whose genomes have a high density of selected sites.

Alexander Klug (ETH Zürich, Switzerland)

The effect of multi-nucleotide mutations on adaptive evolution

In the past, molecular studies of population genetics and evolution have employed the simplifying assumption that the differentiation of alleles or species is primarily driven by a succession of individual changes, each reflecting a nucleotide mutation at a single site in a genome. However, other types of mutational changes also contribute significantly to evolution and adaptation, e.g., gene duplications or mobile element insertions. In the past decade, it has become clear that multi-nucleotide mutations (MNMs), defined as multiple closely spaced changes within a generation, occur at rates far higher than expected from the coincidence of multiple single-nucleotide substitution mutations (SNMs). The most common form of an MNM is the tandem double mutation, which may occur at rates 0.4% of the rate of SNMs. The contribution of MNMs and other rare mutations to evolution and adaptation is a question of general theoretical interest. Whereas SNMs may be the most commonly occurring class of mutations in many species, they are not the most numerous class. Likewise, while single-nucleotide changes may show a more modest distribution of effects on fitness or other phenotypes, other classes of mutations may show a more extreme distribution of effects. Our goal in this study is to explore the relative contributions of SNMs and tandem double mutations to protein adaptation, as a case of adaptation subject to two classes of mutations, one of which occurs with higher rates, and the other of which covers more possibilities but occurs at lower rates.

Thibault Latrille (UNIL, Switzerland)

A phylogenetic mutation-selection model predicts fitness effects of mutations in extant mammals

At the phylogenetic scale, sequence variation informs us on the selective effects of mutations. Indeed, mutations can be either beneficial, deleterious or neutral for their bearer, influencing the likelihood for a mutation to reach fixation. In this study, we first estimated the selective effects of mutations inside mammalian protein coding sequences, assuming a nearly-neutral model of evolution at the phylogenetic scale. We then confirmed that mutations predicted to be deleterious are purified away in extant populations. Conversely, mutations that are repairing previous deleterious changes and simply restoring existing functions are indeed beneficial in extant populations. Our work confirms that a nearly-neutral model of evolution at the phylogenetic scale informs us on the effect of mutations for extant populations and individuals. Moreover, we show that we can also use this nearly-neutral model of evolution as a null model above which we can detect adaptation in protein-coding DNA sequences.

Jens Lagergren (KTH Royal Institute of Technology, Sweden)

VaiPhy - Variational Inference for Phylogeny

Phylogenetics is a classical methodology in computational biology that today has become highly relevant for medical investigation of single-cell data, e.g., in the con- text of cancer development. The exponential size of the tree space is, unfortunately, a substantial obstacle for Bayesian phylogenetic inference using Markov chain Monte Carlo based methods since these rely on local operations. Although more recent variational inference (VI) based methods offer speed improvements, they rely on expensive auto-differentiation operations for learning the variational parameters. We propose VaiPhy, a remarkably fast VI based algorithm for approxi mate posterior inference in an augmented tree space. VaiPhy produces marginal log-likelihood estimates on par with the state-of-the-art methods on real data and is considerably faster.

Loïc Marrec (Universität Bern, Switzerland)

Evolutionary rescue in a fluctuating environment

No environment is constant over time, and environmental fluctuations impact the outcome of evolutionary dynamics. Survival of a population not adapted to some environmental conditions is threatened unless a mutation rescues it, an eco-evolutionary process termed evolutionary rescue. I investigate evolutionary rescue in an environment that fluctuates between a favorable state, in which the population grows, and a harsh state, in which the population declines. I quantify the probability of evolutionary rescue using a stochastic framework with numerical and analytical tools, resulting in an exact computation of the population's fate under deterministic versus stochastic environmental fluctuations. I also compare a perfectly harsh environment (i.e., fully birth-preventing) to an imperfectly harsh one (i.e., not fully birth-preventing) and identify which growth parameters promote evolutionary rescue using different population growth types. Because the functional form of density dependence is essential in assessing the fate of a population, I discuss the main current methods for inferring growth parameters from curve data. I identify their weaknesses, explain why fitting growth data from deterministic models can sometimes poorly estimate growth parameters, and finally attempt to develop an accurate inference method.

Andreas Mayer (University College London, UK)

Imprints of antigen-driven selection in immune repertoires

The specificity of adaptive immune responses relies on the binding of hyper-variable receptors to diverse ligands. Advances in the depth at which the hyper-variable receptor loci can be sequenced provide unprecedented resolution into the many-to-many mapping between receptors and ligands. What does this data reveal about the sequence determinants of specific binding? In my talk, I will discuss some of our recent progress in addressing this question: First, I will introduce a population genetics approach to infer correlation functions of specificity on sequence space. Second, I will present our initial attempts at using metric learning to learn which receptors recognize common targets. I will conclude with some thoughts on how an ability to read the T cell receptor code will change the way we track immune responses to infection, vaccinations, and cancer.

Carina Mugal (Université Lyon 1, France)

Applications of stochastic reproduction-selection models for the study of non-equilibrium dynamics in molecular evolution

Models in molecular evolution are typically separated into two groups, microevolutionary models that describe intra-species genetic variation and macroevolutionary models that describe inter-species genetic variation. Common to both settings is that derivation of analytical results and methodological developments frequently rely on the stationarity assumption, which corresponds to a stationary allele frequency distribution at the microevolutionary scale and a stationary sequence content at the macroevolutionary scale. Yet, most natural populations are not in equilibrium, since many evolutionary processes fluctuate over time and give rise to non-equilibrium conditions. To study the resulting non-equilibrium dynamics, we formulate a stochastic reproduction-selection model, which in the scaling limit converges to a Poisson random field model. This allows us to obtain exact analytical expressions of relevant nonstationary allele frequency spectra, and to formulate micro- and macroevolutionary measures of key evolutionary processes, such as natural selection and the effective population size. Moreover, our setting provides a step towards bridging the gap between micro- and macroevolution.

Charles Mullon (UNIL, Switzerland)

The different paths to adaptive polymorphism in heterogeneous populations

Research of adaptive dynamics via invasion analysis has helped understand how gradual evolution leads to adaptive polymorphism when individuals interact. In my talk, I review some of the basic tools that have come out of this field to model the evolution of quantitative traits in complex populations. I present mathematical expressions that describe directional and disruptive selection in class- and group-structured populations in terms of individual fitness, with the aims of bridging different models and interpreting natural selection. In particular, my review of disruptive selection suggests there are two main paths that can lead to diversity: (i) when individual fitness increases more than linearly with trait expression; (ii) when trait expression simultaneously increases the probability that an individual is in a certain context (e.g. a given age, sex, habitat, size or social environment) and individual fitness in that context. I provide various examples of these and more broadly argue that population structure lays the ground for the emergence of polymorphism with unique characteristics.

Stephan Peischl (Universität Bern, Switzerland)

TBA

Abstract

Cornelia Pokalyuk (Goethe Universität Frankfurt, Germany)

Fixation of slightly beneficial alleles from a backward and a forward perspective

Haldane's formula is a rule of thumb for the probability of fixation for a slightly beneficial mutant. If the mean offspring number of a slightly beneficial mutant is by a factor 1+s larger than that of the wild type individuals, the probability of being established in a large population is approximately 2s/v, where v is the offspring variance. This approximation has been derived (for v=1) by Haldane (1927) based on a branching process approximation by Fisher (1922) and has also been used as an approximation for the probability of fixation.

In my talk I will argue that this formula gives the correct asymptotics within a class of Cannings models with selection. It turns out that for "moderately strong" selection the forward perspective of the Galton-Watson approximation helps to prove the Haldane asymptotics in the large population limit, whereas for "moderately weak" selection a backward perspective is appropriate, which relies on an ancestral selection graph in discrete time. These parts of the presentation are based on F. Boenkost, A. González Casanova, C.P., A.W., Haldane's formula in Cannings models:

- The case of moderately weak selection, Electron. J. Probab. 26(4) (2021)
- The case of moderately strong selection, Stoch. Proc. Appl. 153 (2022)

In the last part of the talk I will give an outlook on work in progress on analoga of Haldane's formula for a class of Cannings models with asymptotically infinite offspring variance.

Xiang-Yi Li Richter (Université de Neuchâtel, Switzerland)

The evolutionary impact of sex-specific spatial scales of competition

The spatial structure and subdivision of a population can strongly impact its evolutionary trajectory through the interactions between individuals (e.g., resource competition, mate choice), and between individuals and the environment (e.g., resource depletion, genotype-by-environment interactions). In species of separated sexes, the spatial scales of competition (i.e., the softness of selection) can often differ between males and females. I will provide examples and modelling case studies to show why it is important to consider sex-specific softness of selection, and why considering this factor can help us solve several evolutionary puzzles: (1) why does experimental evolution under polyandry sometimes produce stronger genome feminization than under monogamy? (2) why do males and females often differ more in their appearance than what they eat? And (3) why does female preference for highly competitive males sometimes lead to decreased population growth rate or even evolutionary suicide?

Merlijn Staps (Princeton University, USA)

The evolution of flexibility in task allocation systems

Living collective systems, such as multicellular organisms and social insect colonies, have evolved diverse strategies to dynamically allocate individuals to different tasks. These strategies range from rigid, inflexible task allocation that is not adjusted to changing circumstances to more fluid, flexible task allocation that is rapidly adjusted to the external environment. It remains poorly understood to what extent such differences in the flexibility of task allocation can be viewed as adaptive responses to different ecological contexts—for example, different degrees of temporal variability. Motivated by this question, we have developed a general mathematical framework to study the evolution of task allocation in dynamic environments. In my presentation, I will show how this framework can be used to identify general rules for the evolution of collective flexibility and propose potential adaptive explanations for some puzzling empirical observations, such as seemingly unnecessary task switching under constant environmental conditions, apparent task specialization without efficiency benefits, and high levels of individual inactivity.

Jérôme Tubiana (Tel Aviv University, Israel)

Harnessing sequence generative models for inhibitory peptide design: a case study

Peptides that efficiently bind a target protein and interfere with its native protein-protein interactions are attractive tools for basic research and therapeutic applications. However, the vast search space and the physio-chemical properties of protein-peptide interactions make rational design challenging. I will present an integrative peptide design protocol based on a generative model trained on native protein interaction partners of the target. We tested our protocol on Calcineurin, a serine/threonine phosphatase involved in multiple cellular pathways such as T-cell activation. We showed that the generative model i) inferred sequence motifs related to binding, ii) predicted binding affinity changes upon mutation, and iii) generated diverse candidate sequences. After filtering via molecular docking and high-throughput binding assays, we found that 70% of the designed peptides successfully interfered with Cn-substrate interactions. Altogether, our work suggests that generative modeling is a promising strategy for discovering peptide inhibitors.

Sedigheh Yagoobi (Max Planck Institute for Evolutionary Biology, Plön, Germany)

Fixation probability in network-structured metapopulations

To study the effect of underlying structure on the evolutionary fate of a population, a common way is Evolutionary Graph Theory(EGT), in which the nodes of the graph indicate the individuals, and the links indicate the interaction or the replacement pattern in the population. However, one needs to modify this model to include migration or dispersal between subpopulations. To do so, one can generalize the graph of individuals in EGT to the graph of subpopulations. An important question is: to what extent do the characteristics of the graph of individuals carry over to the graphs of subpopulations? Another question is how robust is the dynamics of the network-structured metapopulations to the choice of update mechanism?
