

Probabilistic models in biology
Playa del Carmen, Mexico. October 26-30, 2016

Abstracts

- Eneas Aguirre-Von-Wobeser (Instituto de Ecología, Mexico)
Title: Modeling chemically-mediated antagonistic interactions between bacterial cells at two different scales

Abstract: The microbial world is rich in complex interactions between cells, populations, and even communities and ecosystems, which take place at scales ranging from molecular to planet-wide. Thus, determining appropriate scales for modelling those interactions is critical to obtain meaningful results. We were interested in modelling the production of antimicrobial substances by bacterial cells, which is a common type of interaction in almost any microbial community, in which a strain gains competitive advantage by inhibiting or killing other cells. Specifically, we give estimates of the order of magnitude of the costs of those interactions, in terms of resources invested in producing the molecules involved. Even though we are dealing with a single interaction mechanism, chemically-mediated antagonistic interactions, large variations in the numbers of molecules involved call for different mathematical treatment at different scales. At the very small scale, we considered the random diffusion of single molecules away from the producing cell. For this scenario, the time a molecule stays close to a cell was simulated considering random steps, and could be adjusted to an Inverse Gaussian Distribution. Using this approach we could estimate that a cell would need to invest 3.92×10^{22} kg s^{-1} of organic matter, which is 9 orders of magnitude lower than the estimated expense for growth. On the other hand, for many cells it would be useful to maintain a certain concentration of antimicrobial compounds around itself. In this scenario, the large number of molecules prevented modelling their trajectories individually. Thus, we used differential equations and modelled them as continuous concentrations. These results indicated that, in this scenario, cells would need to invest 1.20×10^{-11} kg s^{-1} , which is 2 orders of magnitude higher than the cost of growth. Since this figure is not likely to be sustainable, the real expense is probably intermediate between these two extremes. (Joint work with Luis Enrique Eguiarte, Valeria Souza and Gloria Soberón-Chávez).

- Vincent Bansaye (École Polytechnique, France)
Title: On the regulation of the size of large populations in some stochastic competitive models

Abstract: We will study the decrease of a large initial population which is due to the competition between individuals of the same specie and/or

between two species. We are focusing on logistic birth and death processes and Lotka-Volterra diffusions. We could complete these results by some approximations results for the process scaled in large populations and end up with more general results or persistence criteria.

- Sylvain Billiard (University of Lille, France)
Title: How fast are the interactions? Deriving stochastic functional responses from the interactions at the individual level.

Abstract: Interactions between individuals, (e.g. mating, competition, predation, cooperation), are at the core of ecology and evolutionary biology. How individuals interact, how fast, with whom, etc. and how it affects ecological and evolutionary dynamics is a long-standing question. Most models are based on simplifying assumptions: e.g. no interactions in birth-death processes, competition for 'space' in population genetics, or functional responses in ecology. Functional responses are a characterization at the macroscopic level of how fast individuals interact with each other. It is well-known that functional responses can dramatically affect the dynamics of species in interaction, especially their coexistence. Dozens, maybe hundreds, different functional responses have been proposed in the literature since the seminal work of Holling in the late 1950's. However, very few functional responses are justified at the microscopic level, and even fewer in a stochastic framework. This has three consequent drawbacks. First, the interpretation and the measure of the parameters are difficult. Second, the inference of functional responses from data has revealed almost impossible. Third, a quantification of the variation of functional responses is lacking. Here, we propose a general stochastic framework to derive and justify functional responses from the microscopic level. We first show how it is possible to recover and extend classical functional responses. Second, we show how to derive any functional response from even complicated ecological scenario. Third, we show how to approximate the variance of any functional responses. (Joint work with Vincent Bansaye, Jean-René Chazottes, Clotilde Lepers)

- Jochen Blath (TU, Berlin, Germany)
Title: Seed banks in population genetics

Abstract: We analyse patterns of genetic variability of populations in the presence of a large seed bank with the help of a new coalescent structure called seed bank coalescent. This ancestral process appears naturally as scaling limit of the genealogy of large populations that sustain seed banks, if the seed bank size and individual dormancy times are of the same order as the active population. Mutations appear as Poisson process on the active lineages, and potentially at reduced rate also on the dormant lineages. The presence of such dormant lineages leads to qualitatively altered times to the most recent common ancestor and non-classical patterns of genetic diversity. To illustrate this we provide a Wright-Fisher model with seed bank component and mutation, motivated from recent models of microbial dormancy, whose genealogy can be described by the seed bank coalescent. Interestingly, the scaling limit of this model (forward in time) can be in-

terpreted in terms of a stochastic delay differential equation. Based on our coalescent model, we derive recursions for the expectation and variance of the time to most recent common ancestor, number of segregating sites, pairwise differences, and singletons. The effect of a seed bank on the expected site-frequency spectrum is also investigated. Our results indicate that the presence of a large seed bank considerably alters the distribution of some distance statistics, as well as the site-frequency spectrum. Thus, one should be able to detect the presence of a large seed bank in genetic data. (Joint work with Bjarki Eldon, Adrián González Casanova, Noemi Kurt, Maite Wilke-Berenguer)

- Nicolas Champagnat (INRIA, France)
Title: Uniform exponential convergence to quasi-stationary distributions in population dynamics.

Abstract: We consider general Markov processes of population dynamics, absorbed when one or several (sub-)populations go extinct and seek for criteria ensuring the exponential convergence of the distribution of sizes conditionally on non absorption to a quasi-stationary distribution, uniformly with respect to the initial distribution. This last point is important in practice because the initial distribution of a population is usually not precisely known. We prove that this uniform convergence is equivalent to two conditions, one saying that that the process goes down fast from infinity and goes away from places with high absorption rate when it is not absorbed, and the other one that the process cannot survive much better than when started from a compact set. We then provide explicit criteria ensuring these conditions for multi-dimensional birth and death processes and diffusion processes. (Joint work with Denis Villemonais.)

- Alison Etheridge (University of Oxford, UK)
Title: How does spatial structure interact with natural selection?

Abstract: Since the pioneering work of Fisher, mathematicians and biologists have been trying to understand the way in which selectively advantageous genetic types spread through a population and to recognise the patterns of genetic variation that will be left behind by such ‘selective sweeps’. Much of this work either ignores the effects of genetic drift (due to locally finite population size) or is restricted to one spatial dimension. Here we describe an approach to modelling the effects of selection in spatially structured populations which allows us to take into account genetic drift in arbitrary dimensions (the most interesting being two) and we survey some recent results.

- Régis Ferrière (University of Arizona, USA & Ecole Normale Supérieure, France)
Title: Insights into eco-evolutionary dynamics from simulations of individual stochastic processes

Abstract: A central question in biology is to understand how macroscopic ecological and evolutionary patterns emerge from stochastic processes op-

erating at the microscopic level of individuals and their interactions. Here i will present several examples in which individuals are characterized by 'traits' that can evolve, thus causing changes in the ecology of the system. I will highlight biologically meaningful predictions from these models and discuss challenges and opportunities that these models raise to bridge the gap between eco-evolutionary theory and empirical data.

- Fabian Freund (University of Hohenheim, Germany)
Title: Multiple-merger n -coalescents vs. standard n -coalescents - models, biology & inference

Abstract: Kingman's n -coalescent is widely used as a standard model for the genealogy of n DNA sequences sampled from a single non-recombining, selectively neutral locus. Two conditions ensuring that it is a suitable model are that the population sampled from is panmictic with a fixed (and high) size over time and that the offspring distributions are not too variable. Kingman's n -coalescent is a bifurcating random tree with n leaves.

Kingman's n -coalescent can be easily adjusted to account for selfing, moderate population size fluctuations and population structure while still being a bifurcating random tree. Other changes, e.g. highly variable offspring distributions, can lead to the genealogy being more suitably modelled by a multiple-merger n -coalescent. Multiple-merger n -coalescents are multifurcating random trees with n leaves.

In this talk, the different n -coalescent models are overviewed and compared, both from a mathematical and a biological perspective. Emphasis will be put on the description of statistical approaches to distinguish the different n -coalescents based on their site-frequency spectra.

- Philip Gerrish (Los Alamos National Lab / NASA, USA)
Title: Characterizing rates and effects of fitness mutations - the elusive raw material of natural selection - from samples taken at the population level in real time.

Abstract: The distribution of fitness effects (DFE) of newly-arising mutations characterizes the mutational "input" to evolution, while natural selection largely determines how this input gets sorted into an evolutionary "output". The output cannot contain novel genetic material that is not present in the input and, as such, understanding the DFE and its dynamics is crucial to understanding evolution generally. Despite its essential role in evolution, however, we know almost nothing about the general nature of the DFE, or even if it has a general nature. We derive a novel theoretical framework for estimating the distribution of fitness effects (DFE) of newly-arising mutations in real time from samples taken at the population level – a challenging inverse problem – as well as how this distribution evolves as a population adapts. Our work promises to provide an unprecedented window into this hitherto elusive "microscopic" realm of evolution. Tests of our methods against simulations are very encouraging, accurately reconstructing the DFE (when our "guess" of distribution family is correct) and accurately estimating the genomic mutation rate. We

apply our methods to data taken from evolving *E. coli* populations, giving a first look at the DFE and its evolution in real populations, and providing a surprising estimate of genomic mutation rate. Pairwise comparisons of cumulant-correlation matrices reveals structural changes in the DFE that roughly coincided with apparent selective sweeps; this observation is consistent with the notion that substitutions can open new evolutionary pathways that change the DFE (i.e., that the DFE is contingent on genetic background). (Joint work with Tanya Singh (PhD student, Univ of Pennsylvania), Guillaume Martin (Univ de Montpellier), Paul Sniegowski (Univ of Pennsylvania)).

- Robert Griffiths (University of Oxford, UK)
Title: Genealogy of Wright-Fisher bridges.

Abstract: The path of the frequency of a gene, $X(t)$ at time t , which enters a population and is then lost at time T later is modeled as a bridge in a Wright-Fisher diffusion process beginning at $X(0) = 0$ when the gene enters and ending at $X(T) = 0$ when the gene is lost at time T . The model can be neutral with no selection or mutation; or there can be selection; or there can be mutation away from the genes' type as new types arise. Theoretical results that will be discussed in the talk are the probability distribution of $X(t)$, $0 < t < T$ in the bridge; the coalescent genealogy in the bridge and the maximum attained in a bridge.

A reference is: Joshua Schraiber, Robert Griffiths, Steven Evans, (2013), Analysis and rejection sampling of Wright-Fisher diffusion bridges, *Theoretical Population Biology* **89** 64–74, however there is more emphasis on genealogy in this talk.

- Götz Kersting (Goethe University of Frankfurt, Germany)
Title: Beta coalescents and their shape

Abstract: Coalescents are mathematical models for the genealogical tree of (a sample of) individuals from some population. Beta coalescents make a family of coalescents which borders on the celebrated Kingman coalescent and includes the Bolthausen-Sznitman coalescent. In the talk we describe the construction of this family and touch on possible applications. Then we discuss geometric properties, such as total lengths of all branches or all external branches of the coalescent. A main tool is a random asymptotic expansion for a class of functionals of Beta coalescents, due to Kersting, Schweinsberg and Wakolbinger (EJP 2014).

- Noemi Kurt and Adrian González Casanova (TU, Berlin, Germany)
Title: An individual-based model for the Lenski experiment, and the deceleration of the relative fitness

Abstract: The Lenski experiment investigates the long-term evolution of bacterial populations. Its design allows the direct comparison of the reproductive fitness of an evolved strain with its founder ancestor. It was observed by Wiser et al. (2013) that the relative fitness over time increases sublinearly, a behaviour which is commonly attributed to effects like clonal

interference or epistasis. In this talk, we present an individual-based probabilistic model that captures essential features of the design of the Lenski experiment. We assume that each beneficial mutation increases the individual reproduction rate by a fixed amount, which corresponds to the absence of epistasis in the continuous-time (intraday) part of the model, but leads to an epistatic effect in the discrete-time (interday) part of the model. Using an approximation by near-critical Galton-Watson processes, we prove that under some assumptions on the model parameters which exclude clonal interference, the relative fitness process converges, after suitable rescaling, in the large population limit to a power law function. This is joint work with Adrian Gonzalez Casanova, Anton Wakolbinger and Linglong Yuan.

- Guillaume Martin (University of Montpellier, France)
Title: Probability of fixation and distribution of extinction/fixation times in polymorphic populations

Abstract: What happens to an allele entering a population where several other alleles are already segregating at the same locus. Theory is surprisingly scarce when it comes to the interaction of selection and drift with more than two alleles. Here I will present a general framework, based on a inhomogeneous Feller diffusion [1], to analytically derive the fixation probability of an allele entering a population with arbitrary standing distribution of fitness. I will describe some subcases and check the results against simulations. As an aside, the use of Feller diffusion [2] also provides the distribution of the time to fixation or extinction of the allele, at least in simple subcases (e.g. just two alleles with arbitrary dominance). These results may have applications for models of stochastic adaptation (mutation-selection-drift) in polymorphic populations or possibly in molecular evolution studies. (joint work with Amaury Lambert)

(1) Martin, G. and A. Lambert, fixation probability in polymorphic populations. in prep.

(2) Martin, G. and A. Lambert, A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theor Popul Biol*, 2015. 101: p. 40-6.

- Sylvie Méléard (École Polytechnique, France)
Title: Competitive populations with vertical and horizontal transmissions.

Abstract: Horizontal transfer of information is recognized as a major process in the evolution and adaptation of population, especially microorganisms. There is a large literature but the previous models are either based on epidemiological models or population genetics stochastic models with constant population size. We propose a general stochastic eco-evolutionary model of population dynamics with horizontal and vertical transfers, inspired by the transfer of plasmids in bacteria. The transfer rates are either density-dependent (DD) or frequency-dependent (FD) or of Bedington-deAngelis form (BDA). Our model allows eco-evolutionary feedbacks.

In the first part we present a two-traits (alleles or kinds of plasmids, etc.) model with horizontal transfer without mutation and study a large population limit. It's a ODEs system. We show that the phase diagrams are different in the (DD), (FD) and (BDA) cases.

We interpret the results for the impact of horizontal transfer on the maintenance of polymorphism and the invasion or elimination of pathogens strains. We also propose a diffusive approximation of adaptation with transfer. In a second part, we study the impact of the horizontal transfer on the evolution. We explain why it can drastically affect the evolutionary outcomes. (Joint work with S. Billiard, P. Collet, R. Ferrière, C.V. Tran)

- Arno Siri-Jegousse (Universidad de Guanajuato, Mexico)
Title: The Bolthausen-Sznitman coalescent and its applications in population genetics.

Genealogical models are of major interest to infer parameters of the evolution. In this talk we will describe a forward in time model of a population with selective mutations, studied recently by Schweinsberg and motivated by biological arguments, for which the limit genealogies are given by the Bolthausen-Sznitman coalescent. We will then define more carefully this stochastic process and consider it as a special example of the class of exchangeable coalescents. Finally we will list various results on functionals of the coalescent tree useful to estimate parameters as the rate of mutation or to build some hypothesis tests on the genealogical mode: namely the sites and allele frequency spectra and the minimal clade size. Their asymptotic behaviour will be compared with other models as the Kingman coalescent.

- Dario Spanò (Warwick University, UK)
Title: Bayesian non-parametric inference for Λ -coalescents.

Abstract. The Λ -coalescent family is a class of stochastic processes parametrised by probability measures (Λ -measures) on $[0, 1]$. This family extends the celebrated Kingman's coalescent (recovered when Λ consists of a single atom at $\{0\}$) and provides viable alternative models to interpret the genealogy of several populations (particularly among some marine species) where, with non-negligible probability, a large proportion of the population can be replaced by the offspring of a single individual from the previous generation. Such large reproductive events correspond to jumps in the allele frequency dynamics forward in time and to multiple mergers in the dual genealogy backward in time. In contrast, Kingman's model exhibits no jumps in the allele frequency dynamics and only binary mergers in the genealogy. Inferring the Λ measure governing the evolution of a population is an important challenging problem. Several techniques have recently been proposed when the Λ -measure is assumed known up to the specification of a finite-dimensional parameter. In many contexts however it is reasonable to consider Λ as entirely unknown, i.e. the object of inference is infinite-dimensional: the problem then becomes a non-parametric one. In this talk I will describe a possible Bayesian approach to non-parametric inference on the Λ -measure underlying a population's

evolution. In particular I will provide criteria for posterior consistency in presence of time-series genetic data, and illustrate feasible computational strategies for sampling the posterior. I will discuss current limitations and future challenges for a Bayesian approach.

- Viet Chi Tran (University of Lille, France)
Title: Phylogenies of a neutral marker linked to an adaptive trait under selection in an eco-evolutionary model

Abstract: We consider a population structured by a neutral marker and an adaptive trait, that are supposed to be completely linked (hence neglecting recombination). The population evolves through births and deaths, and individuals compete together, in a logistic way and depending on their adaptive traits. The marker is neutral with respect to competition and selection.

In a large population with rare mutations, we show that the stochastic individual-based model can be approximated by the following process. Restricting to the trait distribution, we recover the Trait Substitution Sequence process (Metz et al. 96, Champagnat 06) or the Polymorphic Evolution Sequence (Champagnat Méléard 11), depending on whether the invasion of an advantageous mutant implies or not its fixation. During the invasion, the neutral marker associated with the original mutant hitchhikes, thus creating a bottleneck. Then, genetic diversity is rebuilt following a Fleming-Viot or Wright-Fisher (in bi-allelic cases) process.

The phylogenies in this limit are coalescent processes with bottlenecks and interaction between branches.

The first part is a work with S. Billiard, R. Ferrière and S. Méléard. The second part is a work in progress with J.M. Marin, S. Méléard and M. Porte.

Contributed talks

- Airam Aseret Blancas Benitez (CIMAT, Mexico)
Title: Simple nested coalescent process

Abstract: In this talk we define a family of Markov processes with values in the space of nested bivariate partitions of \mathbb{N} , called simple nested exchangeable coalescent, for short snec. Our motivation to introduce this class of processes arises from molecular biology where genes trees are nested in the species phylogeny, that is, genes lineages are allowed to coalesce while they are in the same branch of the species trees.

We present a Poissonian construction for a particular snec and we also establish a necessary and sufficient condition for this nested coalescent to come down from infinity, which means that almost surely the partitions associates with species and genes consists of only finitely many blocks for all $t > 0$.

- Matti Leimbach (TU, Berlin, Germany)
Title: A model for cancer growth: propagation and proliferation

Abstract: Motivated by the propagation and reproduction of cancer cells in an infected tissue, we investigate a combination of two PDEs, the viscous porous medium equation and the Fisher-Kolmogorov-Petrovskii-Piskunov model. Both of them can be attained as the limit of the empirical measure of certain interacting particle systems. By combining these two models, one obtains a model, which qualitatively describes cancer growth. We justify this by proving that the combination can be interpreted as the limit of an interacting particle system with a certain branching mechanism. This is based on ongoing work with Franco Flandoli.

- H el ene Leman (Ecole Polytechnique, France)
Title: Large population and rare mutations asymptotics for a spatially structured population

Abstract: The aim is to study the influence of a spatial structure on Darwinian evolution. The interplay between space and evolution and the heterogeneity of the environment are particularly crucial in the emergence of polymorphism and spatial patterns.

The model considered here is an individual-based model in which any individual birth and death events are explicitly described [2]. Each individual is characterized by its phenotypic trait and its position. Migration is modeled by a Brownian motion reflected at the boundaries. At each birth event, a mutation may occur making the phenotypic trait of the offspring different. We study the convergence of this microscopic process when the population size grows to $+\infty$ and the mutation probability decreases to 0. We prove a convergence towards a jump process that jumps in the infinite dimensional space of the stable spatial distributions [3]. This result is correlated with several works on adaptive dynamics and in particular with the model of Trait Substitution Sequence (TSS) [1,4]. It describes the evolution of phenotypic traits and spatial distributions if the ecological time scale, related to the dynamics of the population, and the evolutionary time scale, related to the mutations, are separated. The proof requires specific studies of the microscopic model.

[1] N. CHAMPAGNAT, *A microscopic interpretation for adaptive dynamics trait substitution sequence models*, Stochastic processes and their applications, 116(8):1127-1160, 2006.

[2] N. CHAMPAGNAT AND S. M EL EARD, *Invasion and adaptive evolution for individual-based spatially structured populations*, J. Math. Biol., 55(2):147-188, 2007.

[3] H. LEMAN, *Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence*, Arxiv:1509.02022, 2015.

[4] JAJ. METZ AND S. GERITZ AND G. MESZ ENA AND F. JACOBS AND JS. VAN HEERWAARDEN, *Adaptive dynamics, a geometrical study of the*

consequences of nearly faithful reproduction, Stochastic and spatial structures of dynamical systems, 45:183-231, 1996.

- Sandra Palau (CIMAT, Mexico)
Title: Stochastic differential equations with jumps applied to branching processes.

Abstract: Branching processes are the simplest probabilistic models that describe the dynamics of a given population. In this talk, we first survey some well-known examples of branching processes in discrete time and describe their continuous analogues, in time and space, using stochastic differential equations. Basically we will focus on Galton-Watson processes, Galton-Watson processes in random environment and logistic branching processes. During the second part of the talk, we will introduce a branching process, in continuous time and space, via a stochastic differential equation with jumps that can be thought of as a generalization of the previous branching processes included at the beginning of this talk.

- Charline Smadi (University Oxford, UK & IRSTEA, France)
Title: The effect of recurrent mutations on genetic diversity in a large population of varying size.

Abstract: Recurrent mutations are a common phenomenon in population genetics. They may cause diseases, or reduce the fitness of an individual in a given environment. They may also be at the origin of the fixation of a new genotype, if they give a phenotypic advantage to the carriers of the new mutation.

In this talk, we are interested in the genetic signature left by a selective sweep induced by recurrent mutations in a given locus from an allele A to an allele a , depending on the mutation frequency. We distinguish three possible scales for the mutation probability per reproductive event, which entail distinct genetic signatures. Besides, we study the hydrodynamic limit of the A - and a -population sizes when mutations are likely to happen (mutation probabilities of order 1), and find non trivial equilibria leading to several possible patterns of polymorphisms.