Stochastic and Deterministic Models for Evolutionary Biology

Vincent Calvez (École normale suprieure de Lyon),
Mark Kirkpatrick (University of Texas),
Sylvie Méléard (École Polytechnique),
Sarah Otto (University of British Columbia),
Gaël Raoul (École polytechnique)

July 31–August 5 2016

1   Overview of the Field

Mathematical approaches have always played a major role in evolutionary biology. The theory of evolution developped by C. Darwin (*The origin of Life*, 1859) was based on simple ideas, and opened a way to develop mathematical models. This was actually started by his half-cousin F. Galton in the 1860's, leading to the celebrated Galton-Watson model. The second major imput in the field came from G. Mendel, who discovered some surprising mathematical laws governing the generation-to-generation dynamics of some species. His work lead to the development of genetics. A third important step was the combination of those two approaches by R. Fisher (we also mention S. Wright, B.S. Haldane), in the first half of the 20th century. This was possible thanks to the development of new mathematical methods.

It is stricking to see how mathematical development has been central in the field since its begining, culminating with the descovery of DNA, which provided an amazing chemical basis for genetics. The mathematical developments made by population geneticists have indeed participated to the emergence of modern probability theory and statistical theory.

Recently, the importance of evolutionary biology has been renewed by a number of societal issues: impact of global changes, emerging epidemics, antibiotic resistant bacteria... These new questions have opened a wide range of new problematics in the field, outside of the traditional framework (the most developed of which being population genetics, that is bounded by strong biological assumptions: population with a constant size, without a spatial structure). Another major change in the field was the arrival of new technologies, such as next generation sequencing or GPS tracking, that provide large amount of data. These technological inovation had implied deep changes in the field, from field data collection to experimental evolution. Many theoretical biologists, hand in hand with field biologists, are working to expend the available mathematical methods to those new challenges. Just as it was the case in the early 20's century, developing such a theoretical framework will require new mathematical methods, and we believe this is a very exciting challenge, that will involve biologists, physiscists and mathematicians.

The goal of this meeting was to gather people working in this direction. The biologists that have participated to the meeting are very knowlegeable in traditional mathematical methods for evolutionary biology
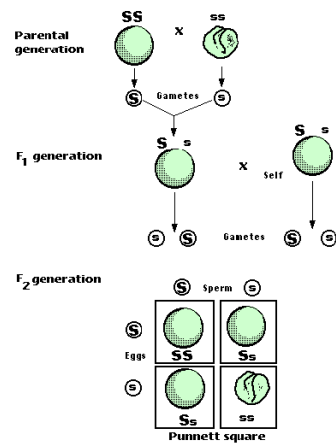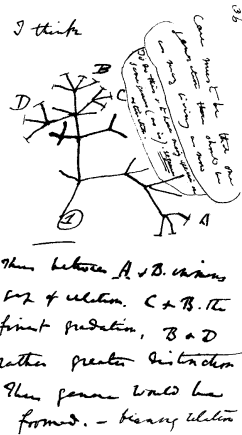
Figure 1: Left: first schetch by C. Darwin describing his point of view on evolution. Right: description of G. Mendel experiment on the heritability on a species of peas.

(population genetics, adaptive dynamics, etc), but also have a good understanding of modern mathematical methods: the interfaces between disciplines have already started to melt. Conversely, a growing number of mathematicians, although working on mathematical problems that are interesting by themselves (and usually publish in generalist mathematical journals), are aware of the theoretical biology litterature, and often associate biologists to their research projects. The presentations and discussions throughout the meeting have thus been very exciting, and exciting questions appealing to both mathematicians and biologists have been numerous.

2 Recent Developments and Open Problems

For this workshop, we have tried to invite the most active researchers of the field worldwide. Given the participants that were able to come to the workshop, we have divided the session according to biological topics. Each thematic consisted in a number of talks, by both biologists and mathematicians, and discussions were encouraged, during questions sessions or during the breaks.

2.1 Evolutionary epidemiology

Evolutionary biology is a very exciting subfield of evolutionary biology: it is directly connected to contemporary health issues (emerging diseases, antibiotic resistence, long term infections...), and a situation where experimental evolution is possible. It is then a very good starting point to develop quantitative mathematical models.

Public health issues

Influenza is a disease that infects millions of people every year (36 000 deaths each year in the United States). This virus evolves continuously, an evolution that is directly linked to the recurrence of the epidemics, and which can lead to dramatic situations (the 1918 flu pandemic was responsible for at least 50 million deaths). This pathogen and its evolution is carefully scrutinised by the World Health Organisation, and the molecular mechanisms for its evolution are well understood.

Julia Gog has presented a state of the art in mathematical models to describe this evolution. She highlighted the importance of an accurate description of the *cross-immunity*, that is the ability of an antibody developed during an infection by a given strain to protect the person from another strain. Doing so in a tractable model is a challenge, and involves a combination of determinist and stochastic approaches.

Jorge Velzquez Castro considered an even more applied question: motivated by current Denge and Zika epidemics, he considered the impact of the geography of cities on the propagation of pathogens. Many pathogens depend on mosquitoes to propagate, which typically stay in a small area around their birth place. This opens the way to targeted anti-mosquito strategies towards specific areas. For instance, it is probably more efficient to target mosquitoes in transportation nodes, rather than vast residential areas. Those ideas were discussed in view of actual data collected in Mexico city.

Drug resistance

Resistance of pathogenic bacteria to antibiotic is a natural phenomena inherent to antibiotic use: microbial are rapidly evolving organisms, and the use of antibiotic implies a selection for resistance. This phenomena is responsible for millions of deaths every year worldwide, and the situation could potentially become even more dramatic, since very few new antibiotic are being discovered. Providing a precise description of resistance dynamics is an important factor to develop public policies to prevent the further spread of those deadly superbugs.

A very important experiment on drug resistance was done by S. Luria and M. Delbrück in 1943, to describe the emergence of a single locus resistance mutation in a bacteria population. This experiment did show that simple mathematical models were indeed able to represent this dynamics. Oskar Hallacheck presented his recent study, where he did re-visit this famous experiment, with the additional help of high speed sequencing. He was able to confirm the results obtained by S. Luria and M. Delbrück, but was also able to uncover more detailed features, that he also described mathematically. Finally, he investigated the impact of a spatial structure on this dynamics, relating it to *gene surfing* problems.
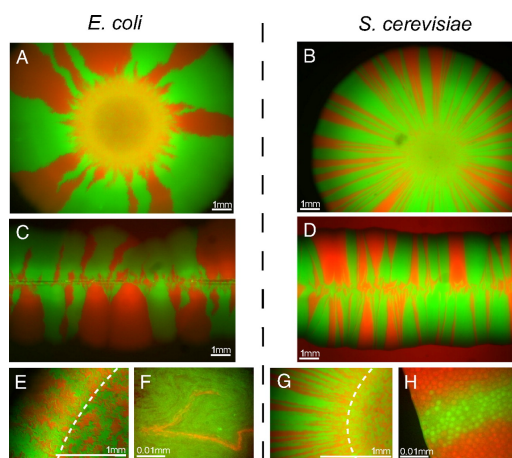

Figure 2: segregation patterns of two different bacterial strains.

In the experiment by S. Luria and M. Delbrück mentionned above, an important assumption was that the resistance mutation involved an only locus. the description of the population is then very simple: a bacteria is resistant or not. In practice however, resistance often involves a combination of loci, and acquiring a beneficial resistance then require the accumulation of several mutation. Matthew Osmond presented a theoretical study of such population, and explained how the faith of the population was affected by the fitness of intermediate strains (where only a part of the necessary mutations occured). He also discussed the impact of a sexual reproduction on this dynamics.

Developing a resistance mutation is actually not the only way for a bacteria to resist an antibiotic treatment: many antibiotic act on bacteria during its replication phase, so that dormant cells remain unaffected. Experimental results show that this phenomena may have a significant impact on the dynamics of a bacterial population. Troy Day developed a mathematical model to describe accurately this dormancy phenomena, which is a plastic (non-heritable) trait. His model is able to capture key experimental data, and he succeeded to describe the dyamics of solutions of this model (although some difficult questions remain open). Another

talk, by Marco Arieli Herrera-Valdez, was devoted to plastic behavior in bacteria. Cells can enter a *competence state*, where they are able to take up extra-cellular DNA from their environment. This particular state is by itself a plastic behaviour, and Marco Arieli Herrera-Valdez was able to validate his theoretical predictions (based on dynamical system arguments) to his experimental results.
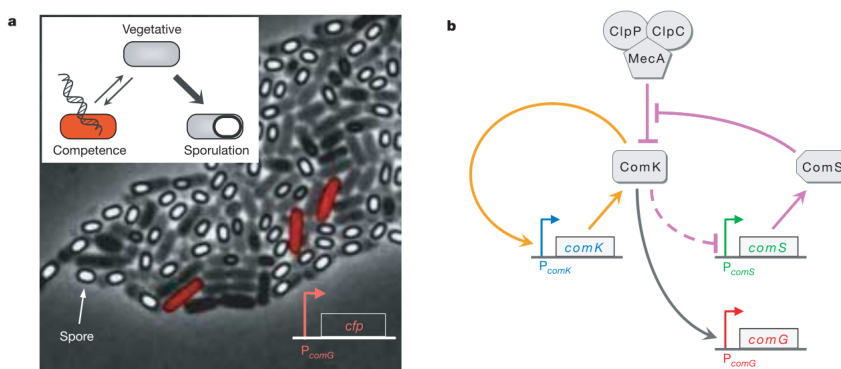

Figure 3: Competence state in *Bacillus subtilis*: an example of a plastic phenomenon.

Yet another way bacteria can survive is through the help of plasmids. Plasmids are strains of DNA that can live inside a bacteria, replicate, and migrate from one bacterium to another. Rafael Pea-Miller has set up simple experiments to assert the impact of the number of plasmids carried by a bacterium on its resistance to anti-biotic. He succeeded to obtain a very agreement of the experimental results to the mathematical models he has developed.

2.2 Spatially structured populations

Spatial structure is often present in natural populations, and is known to have a deep impact on evolutionary dynamics. Most problematics coming from ecological or helath issues involves a spatial structure. Classical evolutionary biology neglect this structure, and developing tractable mathematical models able to take it into account is an important problem today.
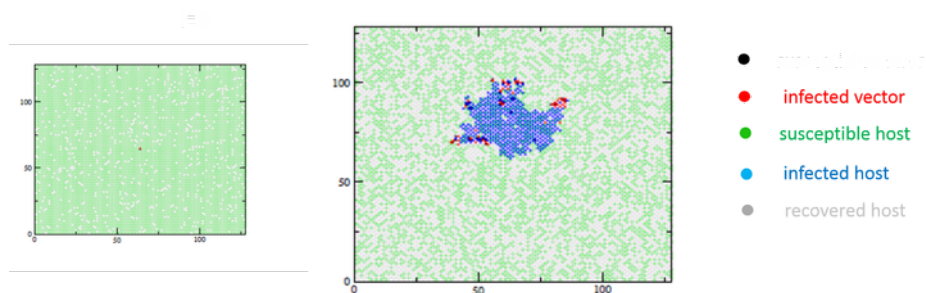
Lattice models

Figure 4: Epidemics in a lattice model.

A first way to model spatial structure is to consider a finite population living on a lattice. It is then possible to define notions such as the neighbours of a given individual. These models are interesting, because it is easy to run simulation, and they allow the emergence of phenomena such as cooperation, that do not appear in non-spatial situations (see the so-called notion of *kin-selection*).

If one is interested in the impact of spatial structure on the dynamics of a population (and not in the distribution pattern of the population, for instance), a natural strategy is to consider equations on *moments*, that is the density of individuals, the density of a couple of individuals that are neighbours, the density of triplets... One then typically obtains an infinite sequence of equations describing those densities. Sebastien Lion presented how it is possible to use those moment equations without the help of (poorly justified) moment closures. He showed how this approach can give useful insight into the short and long term dynamics of populations.

PDE models

Another way to model spatial structure is to consider a continuous spatial variable, and an infinite population size, leading to partial differential equations.

Evolution of Dispersal: advective environments

A first question was introduced by Yuan Lou and Adrian Lam: they consider a population living on a segment, and submitted to a constant speed drift (representing for instance the drift of the water in a river). Then, individuals that do not disperse will eventually be carried out of the interval, and die. Conversely, if they disperse too much, they will randomly escape the domain, and disapear. The optimal strategy should then lie in between those two extrema. It is however not clear that such an optimal strategy exists (because of the spatial structure: a phenotype can be better in one location, and another further downstream...). Yuan Lou and Adrian Lam discussed those situations, depending on the model parameters: when one strain is selected, and when a disruptive selection leads to the emergence of two types.

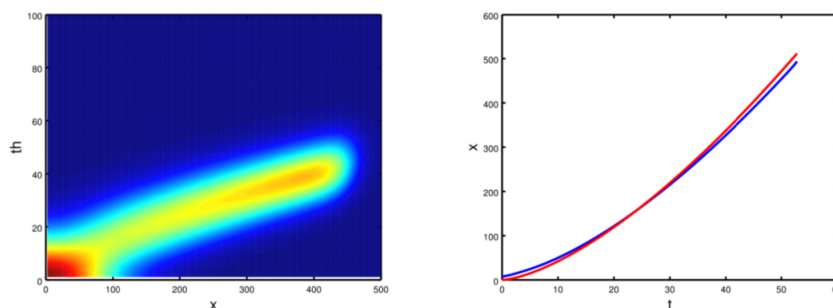**Evolution of Dispersal: invasion of new environments**

Figure 5: Acceleration of the invasion of an empty environment by a population of toads. Left: repartition of the dispersion rate across space in an expending population. Right: position of the edge of the invasion as a function of time.

Mathematical research on those models has been very active recently, especially on the question of the evolution of dispersal: in an expending population, the individuals that acquire a greater dispersion ability are likely to colonize new environments, and thus produce more offspring.

A leap on those questions was made by Olga Turanova, who succeeded to establish an L^∞ estimate on the solutions of a model for a population structured by both a spatial variable and a dispersion trait. She gave a talk introducing the difficulties of the analysis of this equation, and described the Hamilton-Jacobi approach that is made possible by this estimate.

When the phenotypic trait representing the dispersion of the population is unbounded, the dispersion can potentially increase indefinitely. A group of researchers have carefully tracked an invasion of Australia by Cane Toads, that seems to have this type of behaviour. Describing mathematically this dynamics has been a challenging problem for several years, and has been tackled recently by two groups, one of them being represented at the workshop by Emeric Bouin and Christopher Henderson, who presented their results on this fascinating problem.

These theoretical talks where completed by the great talk of the biologist Maxime Deforet, who presented a set of experiments on bacterial populations. He considered several strains of a bacteria, that he obtained by experimental evolution, and that differed from the wild strain by well identified mutations. He set up some competition experiments between those types, and compared the experimental results to asymptotics he obtained through the analysis of partial differential equations.

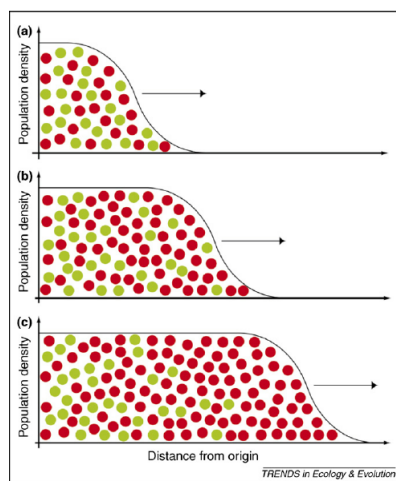**Impact of stochasticity on an expending population**

Figure 6: Impact of the spatial propagation on the diversity of a population.

When a population expends, the individuals on the fore front of the invasion grow without any competition. This allow them to produce a large number of offsprings, and leads to a loss of diversity within the population. This loss of diversity of in the population can lead to the accumulation of deleterious mutations, which can have a real impact on populations.

These accumulations of deleterious mutations can indeed be seen in human populations, as Stephan Peischl showed through an analyis of data from Quebec, where a few number of original europeans colonized a large area over 17 generations. Thanks to this genealogies (with geographic data), he showed that the individuals with many ancestors living on the front of the colonization have a low diversity, and a higher rate of genetic diseases.

It is possible to relate gene surfing phenomena to some properties of Fisher-KPP type equations. An overview of those results was given by Jerome Coville, who, starting from a specific problem (yearly epidemics of pathogens mushrooms along a river), introduced a non-local F-KPP model, and discussed the evolution of neutral diversity for such models.

Impact of a heterogeneous environment

Another interesting topic around spatial structure and evolution is to consider the case of a heterogeneous environment. Typicall questions include: will the population adapt to the specific environment where it leaves (*Local Adaptation*), or adopt traits that are the best on the "averaged environment" ?

A classical setting to consider this question are meta-populations : the population leaves on a finite number of islands, that are connected by some migration events. Sepideh Mirrahimi has presented her ongoing work on this topic: the population is structured by a discrete location variable, and by a continuous trait. She then takes advantage of the Hamilton-Jacobi limit equation (obtained for small mutation rates) to obtained some refined and rigorous result on the phenotypic distribution of the population in both patches.

Matthieu Alfaro describes the situation of a continuous environment, and the situation where the heterogeneity is implied by the climate. In a simple situation, he can describe the dynamics of the population, and consider the impact of a changing climate: will the population succeed to survive, through either evolution or migration, or will it go extinct ?

2.3 Stochasticity in adaptive evolution

Stochasticity is known to play a very important role on evolutionary dynamics. An example was already given with the topic of gene surfing. There are many other challenges, and one highly discussed problem is to extend determinist models (such as the Theory of Adaptive Dynamics) to include the effect of stochasticity.

A basic and relevant setting to consider this question is provided evolutionary epidemiology : how fast will a (relatively small) population evolve ? will it evolve fast enough to avoid extinction when a vaccine is used in the population of hosts ? Sylvain Gandon considered this question, and suggests a way to go beyond the notions of ESS-CSS. This can be especially interesting to describe the impact of transitory dynamics, during the critical moments following the vaccination of the population.

A major gap in the Theory of Adaptive Dynamics is to describe the branching events occuring in a population when an ESS is not a CSS. The descriptions provided by PDE models seem to be of a different nature than the ones observed in Individual Based Models. Joe Yuichiro Wakano advocated for some moment closure formula that would allow a reasonable description of numerical simulations. These formula are not derived through rigorous mathematical arguments, but seem to provide valuable information.
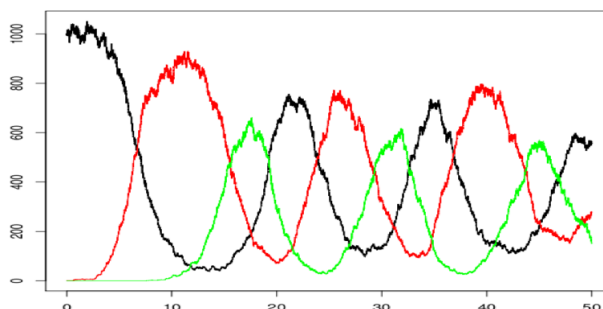

Figure 7: Oscillating population sizes for three competing populations.

Charline Smadi has described her recent work on the impact of stochasticity on population that have periodically a small population size. She mentions for instance the case of a lezard, living is California, where three strains of males exist, leading to a periodic dynamics where one of the strains is dominant for a while, before being replaced by another one... She explained how it is possible to combine an ODE model to describe the abundant population to stochastic methods to model the striving small populations.

2.4 Evolution models and the renewal evolution

In the models we have described so far, individuals did not have any variable descring their age. This is of course a very rough simplification: the key traits of individuals (reproduction rate, death rate, etc) do indeed depend on individual age.

Developing more precise model has been one of the concerns of Odo Diekmann, who has worked on the development of a powerful mathematical setting to study age-structured populations (through so-called *Renewal equations*). Being able to use mathematical arguments such as bifurcation analysis indeed rapidly becomes difficult when an age structure is present.

The problem of obtain renewal equations starting from stochastic Individual Based Model has been considered by Nicolas Champagnat. More precisely, he considered a population of bacteria in a chemostat, where the bacteria are characterized by their size x, and a phenotypic trait. The individuals typically divide more when they have a large size. He is able to extend the notion of fixation probability to this context, and thus to extend the notions of ESS, CSS, etc to size-structured populations.

If we do not assume that mutations are very rare, then the population satisfies a renewal equation in the age variable, with a mutation operator (typically a diffusion) in the trait variable. Jimmy Garnier and
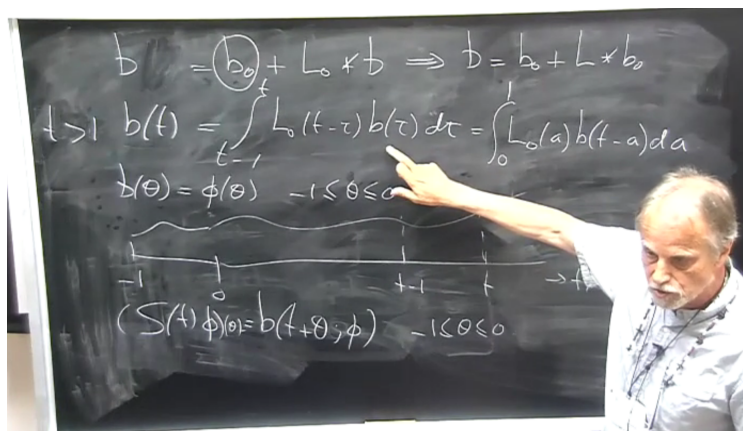
Figure 8: Odo Diekman presenting an analysis framework for renewal equations.

Thibault Bourgeron have considered the case where the individuals where best fit to the environment if their trait coincides with an optimal trait $x(t)$. The impact of a mal-adaptation is here to increase the death rate for individuals of age larger than a given parameter. They are able to discuss the dynamics of the population (survival or death, maladaptation...) depending on the parameters of the model. The main idea being that it is difficult to adapt when a trait affects individuals at an old age only. This analysis, using a Hamolting-Jacobi approach, was able to discuss both asexual and sexual reproduction cases, which produce quite different dynamics.

2.5   Branching process models and genealogies

Branching process models with mutations

Galton-Watson processes are a natural way to model the reproduction of asexual individuals, and has been widely used in evolution theory, since its introduction by the geneticist F. Galton. This model is very useful, because many quantities caracterizing the population are tractable. Extending this simple model to include mutations while remaining tractability rapidly leads to difficulties.

Sandra Palau Caldern presented her work on the subject, where she did add a type to each individual in a Galton-Watson process. The extinction probability of the population can then be related to spectral properties of the matrix describing the various types survival probability. She presents several cases where some quantities, such as the total population size, follow a simple branching process rule.

If we assume that mutations are neutral and rare, it is possible to push the analysis further. Airam Aseret Blancas Bentez presented how we can then characterize the distribution law of the population, which opens the way for statistical inferences of key properties of the population.

The most famous evolutionary evolution experiment was set up by R. Lenski in 1988, and is still continuing today. In consists in twelve populations of E-coli bacteria, a subset of them being introduce in a fresh medium every day. Sample of those populations have been collected over the years, which provides exceptional data, with genealogical data over a large number of generation, in a controlled experiment. Anton Wakolbinger decided to re-visit the mathematical models existing to describe this experiment. He was able to obtain a good description of the increase of the mean fitness of the population throughout the experiment, although many open questions remain.

Reconstructing Phylogenies, and the importance of selection in species evolution

The relative importance of selection (understood as an *almost* deterministic evolution of species toward optimal traits) and mutation (implying random evolutionary trajectories) has been a highly debated topic during the second half of the 20th century, with biological evidence that either of those two possibility exist in

Figure 9: The populations of bacteria in the long-terml Lenski experiment.

nature. Rapid progress of sequencing technologies have made possible the use of massive genetic information to reconstruct the history of species. These statistical methods rely on probability models to reconstruct phylogenies, and the development of those models is a very challenging question.
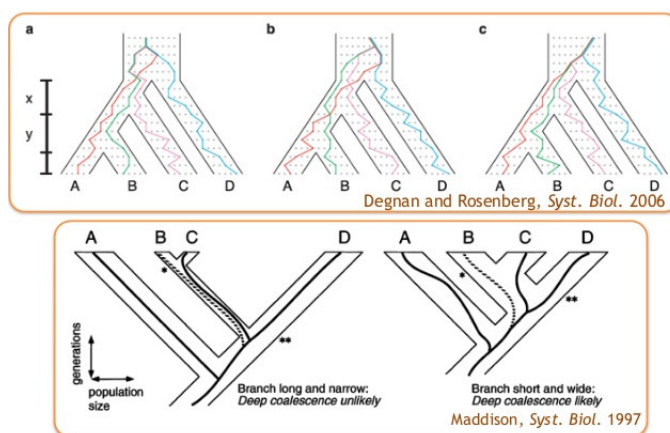

Figure 10: reconstruction of individuals and species genealogies.

If one wants to estimate properties of species evolution based on individuals genetic information, one faces a difficulty: individuals have their own genealogies, that are different from the genealogies of species. The first one is however submitted to the second, since individuals are part of their species at all time. The challenge is then to estimate simultaneously the species genealogies, and genealogies at the level of individuals. This complex coagulation question has been investigated by Arno Siri-Jgousse, who presented a theoretical setting that succeeds to include all necessary features.

A first way to investigate the relation between directed and directed evolution is to consider two types of mutations. This is the idea that was developped by Viet Chi Tran, who constructed phylogenies based on the assumption that individuals are submitted to frequent mutations with neutral effect, and rare mutations with significant effect. Those rare mutations with effect on the fitness drive the species phylogenic tree, while the neutral mutations are dominant within each species. He can then use this model to explain markers distributions in biological populations.

Amaury Lambert undertook a ambitious research project to improve those methods. Stating the problem is already complicated: finding the maximum likelyhood tree, one needs to define distances between trees, which brings in a strong bias, since distances and not equivalent on the infinite dimensional spaces containing all possible trees. Amaury Lamber succeeded to include some inter-specific competition, felt by young

species only, and thus to go beyond the classical *null-model* assumption (where selection is completely neglected). The trees he obtains families of trees depending on a parameter describing the strength of selection, and that interpolate between the completely asymmetric tree and the Kingman coalescent.

3   Workshop Highlights

Over the past decade, several groups of mathematicians have undertook a serious re-visit of mathematical models existing in evolutionary biology. This has lead to the development of inovative mathematical approach. These development have now reached a very stimulating pace, and are done in close collaboration with leading theoretical biologists. Here is a short list of striking recent developments:

- **Asexual populations structured by a phenotypic trait and a spatial variable**: These models lead to monostable reaction-diffusion equations models, with non-local competition terms. The dynamics of several such models are now well understood, based on either Hamilton-Jacobi methods, or travelling wave ideas.

- **Stochasticity in adaptive evolution**: A important mathematical work has been devoted to the development of a robust mathematical framework for the Theory of Adaptive Dynamics. This construction is now mature enough to consider extensions of this theory, in particular to capture more subtle effects of the stoichasticity.

- **Reconstruction of genealogies**: The come of age of rapid sequencing has open incredible new opportunities for evolutionary biology, from the investigation of the early stages of life to experimental evolution. The mathematical side of this problem is very challenging, the choice of a particular model having a big impact on the reconstructed philogenies. There have however been significant progress to include more realistic assumptions in those models.

4   Outcome of the Meeting

The meeting has been a great opportunity to gather this new community, which is a very stimulating composite of mathematicians, physiscists and biologists. Many challenging open questions have been discuss, helping the mathematicians to get in line with current questions of evolutionary biology. Conversely, the mathematical results obtained by mathematician have really shade a new light on several biological questions.

During this workshop, a number of new collaboration have been set up, and should lead to exciting joint research publication. A group has teamed up to produce a review article on the highly discussed problem of mathematical models to describe sexual reproduction. Finally, several long-term invitations have been issued, to deepen discussions initiated during this meeting.

Several topics such as evolutionary epidemiology, sexual reproduction, horizontal transfers or plasticity, that play an increasingly important role in today's evolutionary biology yet haven't received the attention they deserve from the mathematical community, and open very exciting perspectives for future work, taking advantage of the ties we have now built across communities. Reconvening for a second workshop centred on those new questions in a few years' time seems to be a very attractive perspective, and we project to apply for a similar gathering in 2019 or 2020.