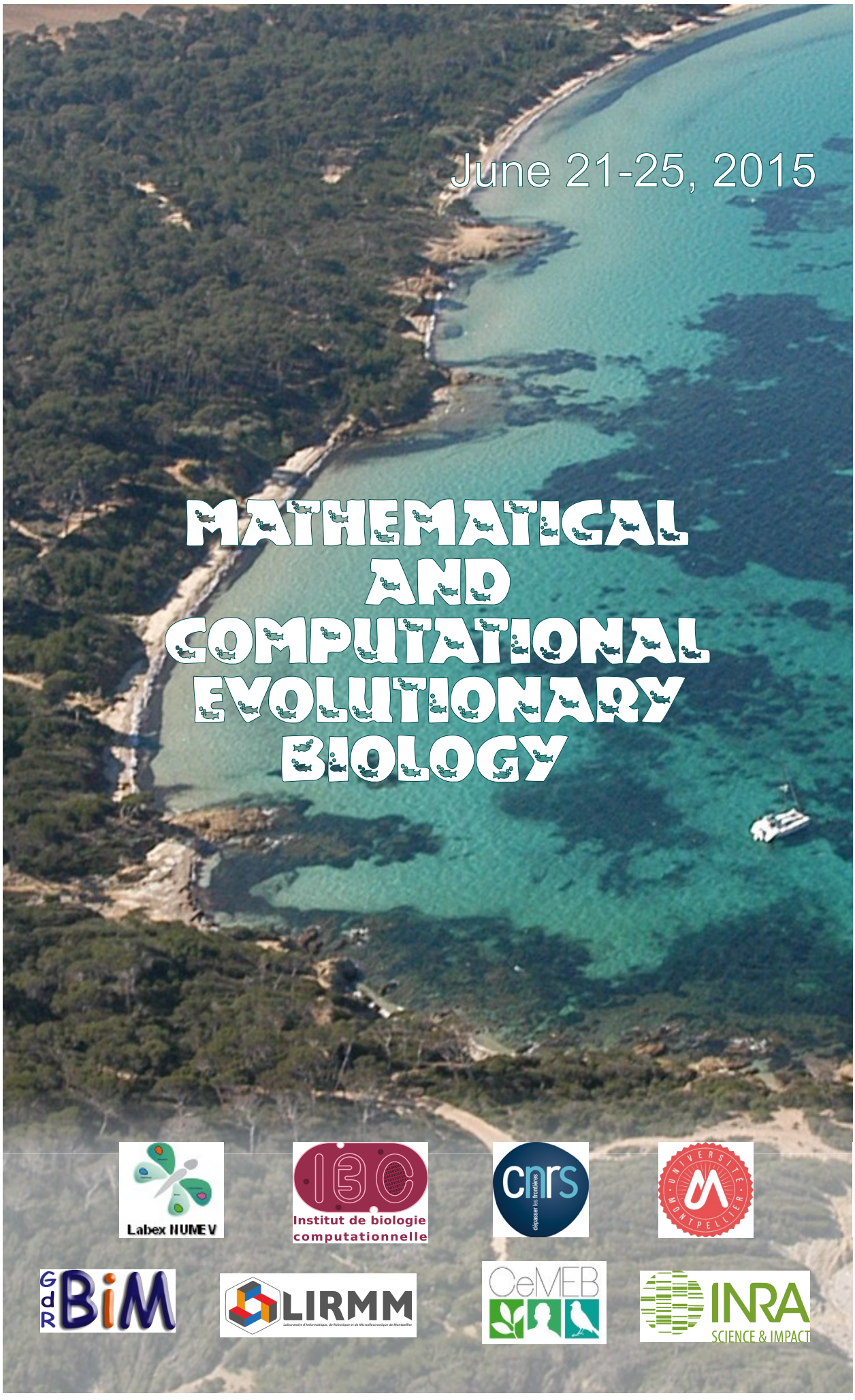

June 21-25, 2015
MATHEMATICAL AND COMPUTATIONAL EVOLUTIONARY BIOLOGY
Labex NUMEV
Institut de biologie computationnelle
cnrs
GdR BiM
LIRMM
CeMEB
INRA
SCIENCE & IMPACT

INFORMATION

Meeting Point

Directly at IGESA Center for the drinks at « le théâtre de verdure » (next to reception) on Sunday 7 :30 pm.
If you stay at IGESA, you can get the key from 4pm.
If you stay at Les MEDES, you can get the key from 3pm

Aim to get the 6.30pm ferry (or earlier) at La Tour Fondue.

In case of problems :

Olivier Gascuel : +33 (0)6 17 80 37 57

Fabio Pardi : +33 (0)6 83 23 20 14

Krister Swenson : +33 (0)7 81 71 60 90

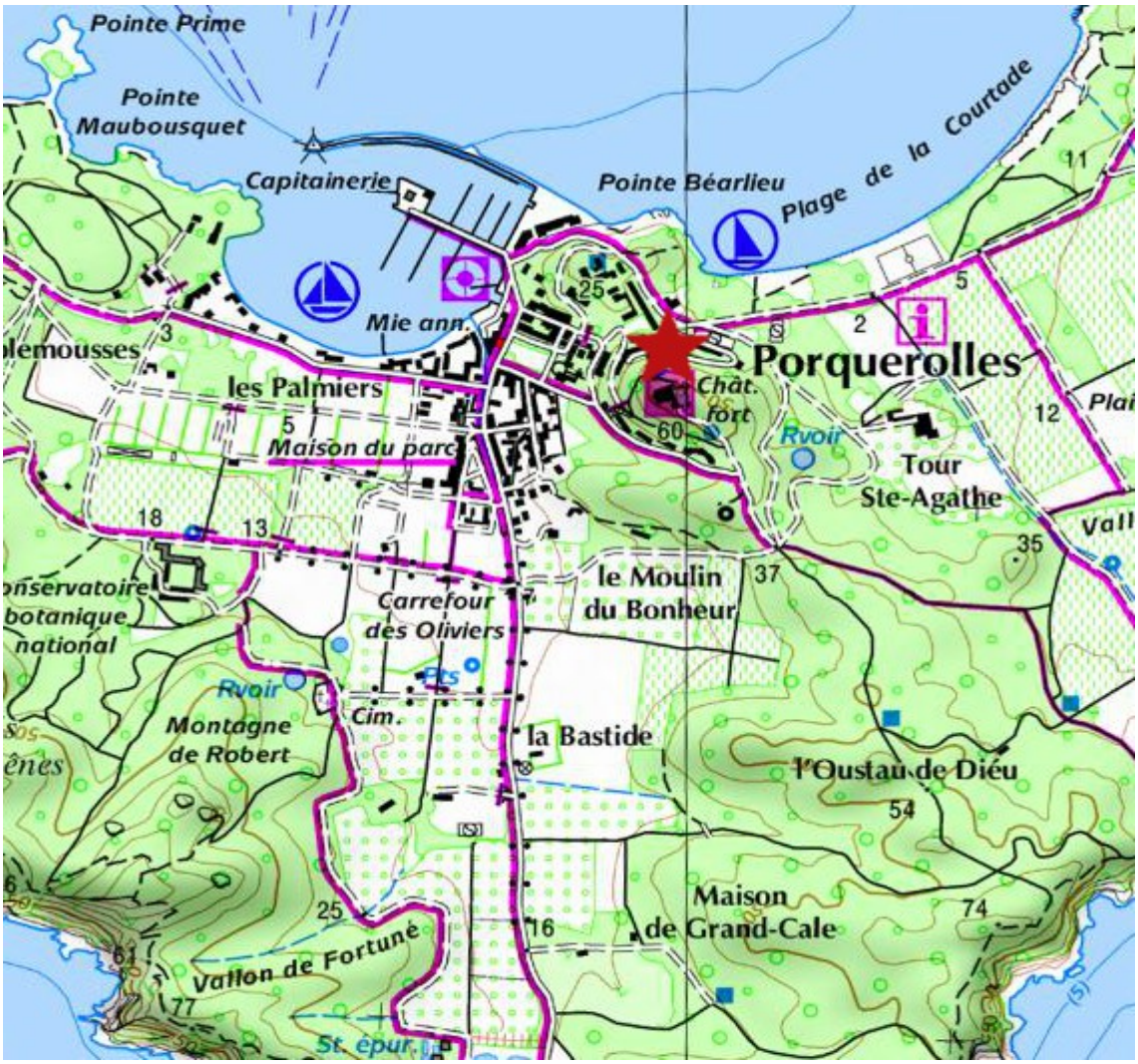
Raphael Leblois : +33 (0)7 81 38 35 03

Céline Scornavacca : +33 (0)7 70 11 94 45

=> inside France use 0 to start (06 17...), outside France use +33 without the 0 (336 17)

Location

The conference will be held at the ***IGESA*** centre, in the village of Porquerolles, five minutes from the harbour and the beach. It is located in a nature reserve, the Port-Cros National Park, one of the most intact coastal areas in the Mediterranean.



Practical information

Accommodations :

Hôtel Club IGESA
Rue de la Douane
Ile de Porquerolles
83 400 HYERES
Tel : 04 94 12 31 80

Coordonnées GPS : 43.000402,6.205076

WI-FI : available but very limited speed

Hôtel Résidence Les Mèdes
Rue de la Douane
Ile de Porquerolles
83 400 HYERES
Tel : 04 94 124 124

free WI-FI

Ferry :

TLV-TVM : 04 94 58 21 81
Return ticket: 19,50€
La Tour Fondue -> Porquerolles :
7:30, 9:00, 9:30, 10:00, 10:30, 11:00, 11:30, 12:00, 12:30, 13:30, 14:30, 15:30, 16:30, 17:30, 18:30
Porquerolles -> La Tour Fondue :
7:00, 8:30, 9:30, 10:00, 10:30, 11:00, 11:30, 14:00, 15:00, 16:00, 17:00, 18:00, 19:00

After hours shuttle :
One way ticket : 18,50€ (buy your ticket directly on board)
La Tour Fondue -> Porquerolles : 19 :45, 23 :15
Porquerolles -> La Tour Fondue : 20:00, 23:30

Taxi-Boat:

Taxi Boat « Le Pelican » : 06 09 52 31 19
The taxi-boat costs 16.50€ per person, when at least 6 people are booked on it.

Car Park on Tour Fondue :

Cars are not allowed on the island, so if you come by car you'll have to leave it in a car park at La Tour Fondue. You can book a place.
Car park Coulomb (watchman 24/24) : 04 94 58 14 67 (24h : 14€)
Car park des Iles (videosurveillance) : 04 94 58 90 78 (24h : 15€)
Car park Vinci (watchman 24/24) : 04 94 01 99 28 (24h : 16€)

Bus Hyères-La Tour Fondue :

If you come by train or plane to Hyeres, you can take the bus line 67 on « Reseau Mistral »
More information is available with your mobile : http://m.reseaumistral.com/
Or call a taxi : Taxi Hyères 04 94 00 60 00

PROGRAM

Sunday, June 21

> 19:30 : Drinks

> 20h00 : Dinner

Monday, June 22

> 09h00 : Welcome

> 09h15 – 10h30 : *KEYNOTE* *p.7*

Asger Hobolth (Bioinformatics Research Center - BiRC, Aarhus University, DK)
« Modelling DNA sequence evolution within and between species »

> 10h30 : Coffee break

> 11h00 – 12h20 : 4x 20 min *TALKS* (including questions)

Guillaume ACHAZ (Université Pierre and Marie Curie, FR) *p.10*
« How models of population growth differ in the light of genetical data? »

Min YE (EPFL, CH) *p.14*
« A generative model for protein protein interaction network evolution based on modularity »

Véronique LADRET (Institut de Mathématiques et de Modélisation de Montpellier - I3M, FR) *p.12*
« Evolutionary game dynamics in a finite continental-island model and emergence of cooperation »

Giulio DALLA RIVA (Biomathematics Research Centre-University of Canterbury, NZ) *p.10*
« Evolving in a tangled world »

> 12h20 : Lunch

> 14h00 – 15h00 : 3x20 min *TALKS* (including questions)

Claudia SOLIS-LEMUS (University of Wisconsin-Madison, USA) *p.14*
« Inferring phylogenetic networks from quartets with maximum pseudolikelihood estimation »

Patrick GEMMELL (Department of Zoology, The University of Oxford, UK) *p.11*
« The evolutionary dynamics of endogenous retroviruses »

Etienne GJ DANCHIN (Institut Sophia Agrobiotech, Université de Nice Sophia Antipolis, FR) *p.11*
« Evolutionary successful and asexual: the paradox of the root-knot nematodes? »

> 15h15 – 16h30 : *KEYNOTE* *p.7*

Philippe LEMEY (Rega Institute, Clinical and Epidemiological Virology, BE)
« Data integrating in viral evolutionary inference: from spatial dynamics to trait evolution »

> 16h30 - 18h30 : Freetime, beach...

> 18h30-20h00 : *POSTERS* *p.16-20*
Wine and discussion (posters 1-10)

> 20h00 : Dinner

Tuesday, June 23

> 09h15 – 10h30 : *KEYNOTE* *p.7*

Jukka CORANDER (University of Helsinki, FI)
« ABC meets machine learning - fitting intractable models to genome data »

> 10h30 : Coffee break

> 11h00 – 12h15 : *KEYNOTE* *p.8*

Bernard MORET (Laboratory for Computational Biology and Bioinformatics, EPFL, CH)
« Phylogenetic Transfer of Knowledge »

> 12h15 : Lunch

> 14h00 - 20h00: Free afternoon (beach, hiking, theorems, etc.)

> 20h00 : Dinner

Wenesday, June 24

> 09h15 – 10h30 : *KEYNOTE* *p.7*

David BRYANT (University of Otago, NZ)
« Recovering phylogeny and demographics from SNPs: prospects and limitations »

> 10h30 : Coffee break

> 11h00 – 12h20 : 4x 20 min *TALKS* (including questions)

Murray PATTERSON (Universite Claude Bernard Lyon 1, FR) *p.12*
« Correlated evolutionary scenarios of metabolic functions »

Fanny POUYET (Lab. de Biométrie et Biologie Evolutive - LBBE, Villeurbanne, FR) *p.12*
« A new codon substitution model to better estimate evolutionary processes »

François ROUSSET (Institut des Sciences de l'Evolution, Université de Montpellier, FR) *p.13*
« A calibrated method for simulation-based inference »

Peter ARNDT (Max Planck Institute for Molecular Genetics, Berlin, DE) *p.10*
« How Evolution of Genomes Is Reflected in Exact DNA Sequence Match Statistics »

> 12h20 : Lunch

> 14h00 – 15h00 : 3x20 min *TALKS* (including questions)

Konrad SCHEFFLER (University of California, San Diego, USA) *p.13*
« RELAX: Detecting Relaxed Selection in a Phylogenetic Framework »

Magali SEMERIA (Lab. de Biométrie et Biologie Evolutive - LBBE, Villeurbanne, FR) *p.14*
« Probabilistic reconstruction of ancestral gene orders using reconciled gene trees »

Chris ILLINGWORTH (University of Cambridge, UK) *p.11*
« Evaluating the role of selection in the within-host evolution of the influenza virus »

> 15h00 - 15h15 : Break

> 15h15 – 16h30 : *KEYNOTE* *p.9*

Katherine ST JOHN (City University of New York, USA)
« Towards Improving the Search for Optimal Phylogenetic Trees »

> 16h30 - 18h30 : Freetime, beach...

> 18h30 - 20h00 : *POSTERS* *p.21-25*
Wine and discussion (posters 11-19)

> 20h00 : Dinner

Thursday, June 25

> 09h15 – 10h30 : *KEYNOTE* *p.9*

Molly PRZEWORSKI (Columbia University, New york, USA)
« An evolutionary perspective on human germline mutation »

> 10h30 : Coffee break

> 11h00 – 12h15 : *KEYNOTE* *p.8*

Nicolas ORLANDO (Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen)
« Ancient DNA: from very old molecules to genomes and epigenomes »

> 12h15 : Lunch and then farewell session

> 14h00 : First ferry to « La Tour Fondue »

KEYNOTE SPEAKERS

> ***David BRYANT***
University of Otago, NZ

Recovering phylogeny and demographics from SNPs: prospects and limitations
Single Nucleotide Polymorphisms (SNPs) are fundamental to applications of genetics in medicine, pharmacology, and breeding. As such, there has been massive investment in technology for obtaining and analysing SNP data, technology which now routinely permits genotyping of hundreds of thousands of SNPs. These SNPs should be a rich source of information about species structure and history.In this talk I will examine how SNP data can aid inference of species trees and ancestral demographics. I will review how SNP data is obtained and then talk about the various evolutionary models which have been developed. I discuss model-based tools recently developed by various research groups for analysing these data. I conclude by talking about limitations of these data and some ideas for future directions.

> **Jukka CORANDER**
University of Helsinki, FI

ABC meets machine learning - fitting intractable models to genome data
For a statistical model specified through a data generating process with an intractable likelihood, simulation-based inference methods have been proposed where the model parameters are inferred by finding values which yield simulated data that resemble the observed data. Two major dificulties in this approach are the choice of the discrepancy measure, and computationally eficient identification of regions in the parameter space where the discrepancy is low. We give an introduction to our recent work which tackles the two dificulties through classification and Bayesian optimization. Applications to modeling transmission dynamics and evolution of bacterial populations are used to illustrate the advantages brought by our method which can speed up likelihood-free inference by three orders of magnitude compared with the state of the art adaptive sequential Monte Carlo algorithms and MCMC.

> **Asger HOBOLTH**
Bioinformatics Research Center (BiRC), Aarhus University, DK

Modelling DNA sequence evolution within and between species
We consider the diffusion approximation of the multivariate Wright-Fisher model with mutation. Analytically tractable formulaefor the first- and second-order moments of the model are derived, and the moments are subsequently used to better understand key population genetics parameters and modelling frameworks. In particular we investigate the behaviour of the expected homozygosity (the probability that two randomly sampled genes are identical) in the transient and stationary phases, and how appropriate the Dirichlet model is for modelling allele frequency changes at different evolutionary time scales. We find that the Dirichlet model is adequate for the pure drift model (no mutation), but the model is not sufficiently flexible for more general mutation structures. We suggest a new hierarchical Beta approximation for the Wright-Fisher model with a mutation model on the nucleotide level that distinguishes between transitions and transversions.

> **Philippe LEMEY**
Rega Institute, Clinical and Epidemiological Virology, BE

Data integrating in viral evolutionary inference: from spatial dynamics to trait evolution
Combating pathogen spread and their associated disease burden is a tremendous challenge requiring sustained research effort and decided public health measures, and the availability of genomic data provides a major asset in characterizing these pathogens. Recent developments in pathogen phylodynamics aim at a marriage of statistical thinking and evolutionary biology to integrate these data through phylogenetic reconstructions with host, phenotypic and geographic sampling information. Here, we will highlight recent advances in Bayesian evolutionary inference methodology that focus on data integration in pathogen phylodynamics. These approaches include connecting sequence to trait evolution, but also the incorporation of covariates of the evolutionary process in the reconstruction procedures. Specifically, these developments rely on phylogenetic implementations of Bayesian hierarchical and mixed-effects modeling, and Bayesian multidimensional scaling. We will discuss how these approaches allow testing the ecological impact on viral evolution in wildlife hosts, identifying the drivers of spatial dispersal and the determinants of cross-species transmission, and quantifying evolutionary patterns of pathogen phenotypes. Applications will focus on influenza, HIV and rabies viruses.

> **Bernard MORET**
Laboratory for Computational Biology and Bioinformatics, EPFL, CH

Phylogenetic Transfer of Knowledge
Advances in biotechnology have enabled researchers to study molecular biology from the point of view of systems, from focused efforts at functional annotation to the study of pathways, regulatory networks, protein-protein interaction networks, etc. However, direct observation of these systems has proved difficult, time-consuming, and often unreliable. Thus computational methods have been developed to infer such systems from high-throughput data, such as sequences, gene expression levels, ChIP-Seq signals, etc. For the most part, such methods have not yet proved accurate and reliable enough to be used in automated analysis pipelines.Most methods used to infer biological networks rely on data for a single organism; a few attempt to leverage existing knowledge about some related organisms. Today, however, we often have data about a large variety of organisms as well as good consensus about the evolutionary relationships among these organisms, so that the latter can be used to integrate the former in a well founded manner, thereby gaining significant power in the analysis. We have coined the term Phylogenetic Transfer of Knowledge (PTK) for this approach to inference and analysis. A PTK analysis considers a family of organisms with known evolutionary relationships and transfers biological knowledge among the organisms in accordance with these relationships. The output of a PTK analysisthus includes both predicted (or refined) target data for the extant organisms and inferred details about their evolutionary history.While a few ad hoc inference methods used a PTK approach almost a dozen years ago, we first provided a global perspective on such methods just 6 years ago. The last few years have seen a significant increase in research in this area, as well as new applications. We will thus review the principles behind the PTK approach, present some graphical models for specific applications, discuss the range of possible applications, and review some open problems.

> **Ludovic ORLANDO**
Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK

Ancient DNA: from very old molecules to genomes and epigenomes
DNA molecules can survive in fossil specimens and archaeological material over hundreds of thousands of years. The recent development of high-throughput sequencing and molecular advances tailored to the recovery of ultra-short and extremely damage DNA fragments have enabled the reconstruction of whole genomes from ancient individuals and extinct species. The sensitivity of current methodologies increased the time window for genome sequencing by an order of magnitude, to at least one million years in permafrozen regions and half-a-million years in temperate caves. Beyond genomes, the profiling of the epigenetic landscape has also become feasible and genome-wide nucleosome and methylation maps from past organisms have now been reconstructed, paving the way for evaluating the evolutionary role of epigenetic reprogramming. The taxonomic composition of ancient microbiomes can also be reconstructed, providing a unique opportunity to follow the changes possibly introduced by recent cultural changes in our life-style, following the Neolithic revolution from hunter-gathering to farming or post the industrial revolution and the modern antibiotic era. Despite such fantastic advances, paleogenomics also faces its own challenges. In particular, demonstrating that the vast majority of the sequence data underpinning ancient genomes does not derive from fresh DNA contaminants is key to data authentication. Additionally, as damage affecting DNA templates after death can lead to nucleotide mis-incorporations during sequencing, analytical procedures, including read alignments and variant calling, must account for inflated sequencing error rates. We will present key technological developments underpinning this recent paleogenomic revolution, and show that ancient sequence information can illuminate our evolutionary past at unprecedented levels. We will then present our recent work focusing on the reconstruction of ancient horse genomes spanning different stages of the domestication process and illustrate how these can help better understand the evolutionary origins of the modern horse, as well as identify the genes that have been selected during the domestication process. More specifically, we will compare the genomes of modern domesticated horses to the genomes from horses that lived in the Late Pleistocene, prior to their domestication and we will track the recent evolutionary past of Yakutian horses, which represent the domesticated breed that populates the coldest country from the Northern hemisphere.

> **Molly PRZEWORSKI**
Columbia University, New York, USA

An evolutionary perspective on human germline mutation

Germline mutation is the ultimate source of genetic differences among individuals and species. Yet our understanding of this fundamental process is limited by its extremely infrequent occurrence. The cumulative effects of mutations over many thousands of meioses, however, are captured by polymorphism and divergence data. Moreover, the revolution in sequencing technologies has made it feasible to directly detect de novo mutations that arise in transmissions from parents to offspring. Together, analyses of these three levels of variation data - pedigrees, populations, and species - are beginning to reveal how mutations arise and how they evolve. I will provide an overview of the methods by which mutation processes are studied and what we have learned, focusing on three questions: (i) On a mechanistic level, what proportion of mutations is introduced through mistakes in the replication process versus non-replicative, « spontaneous » errors? (ii) In terms of variation among individuals, why do mutation rates depend so strongly on sex and age, and are there other sources of variation? (iii) From an evolutionary perspective, how do mating systems and life history traits shape the mutation rate of a species?

> **Katherine ST. JOHN**
City University of New York, USA

Towards Improving the Search for Optimal Phylogenetic Trees

Finding the optimal evolutionary history for a set of taxa is a challenging computational problem, even when restricting possible solutions to be "tree-like" and focusing on the maximum-parsimony optimality criterion. This has led to much work on using heuristic tree searches to find approximate solutions. We present an approach for finding exact optimal solutions that employs and complements the current heuristic methods for finding optimal trees. Given a set of taxa and a set of aligned sequences of characters, there may be subsets of characters that are compatible, and for each such subset there is an associated (possibly partially resolved) phylogeny with edges corresponding to each character state change. These perfect phylogenies serve as anchor trees for our constrained search space and delimit a region in the space of trees within which the best tree is guaranteed to be found. Limiting the search for the optimal tree to this region can significantly reduce the number of trees that must be examined in a search of the space of trees. We analyze this method empirically using four different biological data sets as well as surveying 400 data sets from the TreeBASE repository, demonstrating the effectiveness of our technique in reducing the number of steps in exact heuristic searches for trees under the maximum-parsimony optimality criterion.

TALKS

> Marguerite Lapierre[1,2,3] ; Cécile Delaporte[3,4] ; Amaury Lambert[3,4] ; **Guillaume Achaz**[1,2,3]
[1] Evolution Paris-Seine, UMR7138, UPMC, Paris, France
[2] Atelier de Bioinformatique, UPMC, Paris, France
[3] SMILE, CIRB, UMR7241 [4] Laboratoire des Probabilités et Modèles Aléatoires (LPMA, UMR 7599), UPMC, Paris, France

How models of population growth differ in the light of genetical data?

Interestingly, since the advent of the neutral theory of molecular evolution, we have been assuming a single reference model in population genetics and even more generally in molecular evolution. This model, coined as “The Standard Neutral Model” assumes a finite panmictic population of constant size and no selection. This model has taken the position of a statistical reference model (H0) so that one needs to reject it to propose an alternative model.Classical violations of this model include demography (no constant size) or selection. Demographical inferences are often based on large number of loci, even complete genomes. Using the data from the 1000 human genomes project, one can easily fit a scenario of continuous expansion to explain the pattern of observed diversity in the African Human populations.We will show that, even in the case of extremely simple models with 1 parameter, several different models of expansion (e.g. linear expansion, exponential expansion, sudden expansion or even growing branching processes) can account for the observed pattern. Indeed, we will show that for some parameter range, all models seem almost confounded. This work shows (1) that some care is needed when we interpret genetic data in the light of a model and (2) that different kinds of growth for a population can result in extremely similar coalescent trees. Ultimately, it shed light on the fundamental behaviors of the models themselves.

> Florian Massip[1,2] ; Michael Sheinman[1] ; Sophie Schbath [2] ; **Peter F. Arndt** [1]
[1] Department for Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 63–73, 14195 Berlin, Germany
[2] UR1077, Unite Mathematiques Informatique et Genome, INRA, domaine de Vilvert, Jouy-en-Josas, FR

How Evolution of Genomes Is Reflected in Exact DNA Sequence Match Statistics

Genome evolution is shaped by a multitude of mutational processes, including point mutations, insertions, and deletions of DNA sequences, as well as segmental duplications. These mutational processes can leave distinctive qualitative marks in the statistical features of genomic DNA sequences. One such feature is the match length distribution (MLD) of exactly matching sequence segments within an individual genome or between the genomes of related species. These have been observed to exhibit characteristic power law decays in many species. Here, we show that simple dynamical models consisting solely of duplication and mutation processes can already explain the characteristic features of MLDs observed in genomic sequences. Surprisingly, we find that these features are largely insensitive to details of the underlying mutational processes and do not necessarily rely on the action of natural selection. Our results demonstrate how analyzing statistical features of DNA sequences can help us reveal and quantify the different mutational processes that underlie genome evolution.

> **Giulio Valentino Dalla Riva**
Biomathematics Research Centre (UC), Christchurch, New Zealand.

Evolving in a tangled world

Are phylogenies relevant in understanding the assembly of ecological networks? Yes, they are. However, the research has been restrained by the fact we classically considered food webs as deterministic binary relational structures. In this talk we will show that one can interpret empirically observed food webs as realisations of a family of stochastic processes, namely Random Dot Product Graph models, providing an ideal extension of food-web models beyond the limitations of current deterministic or partially probabilistic models and into a space where the relevant information is given by a metric structure. Interestingly, our results indicate that the evolutionary signature in food webs is present. Indeed, this is not a completely surprising result, as we know that, generally speaking, species in the kingdom Plantae tend not to eat species in the kingdom Animalia. However, we will show that we can be more detailed than that: for example, we will argue that the evolutionary signal in food webs' backbones is stronger than it is in food webs' fine-scaled architecture.

> Romain Blanc-Mathieu[1] ; Laetitia Perfus-Babeoch[1] ; Martine Da Rocha[1] ; Jérôme Gouzy[2] ; Erika Sallet[2] ; Philippe Castagnone-Sereno[1] ; Jean-Marc Aury[3] ; Arnaud Couloux[3] ; Jean-François Flot[4] ; Corinne Da Silva[3] ; Thomas Schiex[2] ; Pierre Abad[1] ; **Etienne G.J. Danchin**[1]
[1] Institut Sophia Agrobiotech, UMR 1355, INRA, Université de Nice Sophia Antipolis, CNRS, Sophia-Antipolis, France.
[2] LIPM, UMR 441, INRA, CNRS, Castanet-Tolosan, France.
[3] Genoscope, CEA, Institut de Génomique, Evry, France. [4] University College London, Department of Genetics, Evolution and Environment, London, UK.

Evolutionary successful and asexual: the paradox of the root-knot nematodes?
Root-knot nematodes show an intriguing diversity of modes of reproduction from obligatory sexual reproduction to fully asexual reproduction. Surprisingly, the most damaging species to the world agriculture are those that reproduce without meiosis and without sex. This observation seems to contradict the evolutionary advantages of sex and genetic exchanges. To disentangle this parasitic success despite asexuality, we have compared the genomes of root-knot nematodes with different modes of reproduction. We have sequenced and assembled the genomes of 3 obligatory asexual species, Meloidogyne incognita, M. arenaria and M. javanica, unable to do meiosis. We have compared these genomes to those of the facultative asexual M. hapla and the obligatory asexual M. floridensis, both able to do meiosis. Our comparative genomics analysis shows that the genomes of the ameiotic asexual root-knot nematodes are composed of duplicated regions that show a high average nucleotide divergence of 10% within a species. Phylogenomic analysis of the genes present on duplicated blocks suggest that these regions result from multiple independent hybridization events. Nucleotide divergence level between duplicated genes (4%) may allow functional divergence and provide plasticity despite the absence of sexual recombination. In contrast, mitochondrial genome divergence between the three ameiotic asexuals is very low (~2%) and suggests that these hybrids share a common or closely related maternal donor lineage. We also observed high variations in TE contents. Two of the obligatory asexuals, M. arenaria and M. javanica show a high proportion of TE (~50%) while the third asexual, M. incognita has a lower TE content (~25%), closer to the proportion of the facultative sexual M. hapla (~20%). Proliferation of TE following hybridization might have taken place in these asexuals. The intriguing and paradoxical parasitic success of root-knot nematodes without sex could be partly explained by transgressive phenotypes resulting from the hybridization events.

> **Patrick Gemmell** [1] ; Jotun Hein [2] ; Aris Katzourakis [1]
[1] Department of Zoology, The University of Oxford, UK;
[2] Department of Statistics, The University of Oxford, UK

The evolutionary dynamics of endogenous retroviruses
Animal genomes, including our own, contain genes left behind by the viral pathogens that infected our ancestors. Over tens of millions of years, many of these endogenous retroviruses (ERVs) continued to replicate as selfish DNA so that today they comprise a large fraction of the genetic material that we pass on to our offspring. Surprisingly, ERVs have been shown to be involved in many important host processes including placentation in mammals and the maintenance of stem cell identity in humans. In this presentation we describe how we use recently released whole genome alignments to answer quantitative questions about the evolutionary dynamics of ERVs across the primate lineage. These questions are addressed by processing alignments using non-deterministic finite automata and then applying a phylogenetic maximum likelihood method. Our method recovers know qualitative facts about the behaviour of different ERV families but, for the first time, places them in direct quantitative comparison. Preliminary results suggest that our method also reveals that a biologically important family of ERVs displays distinct dynamics when placed in comparison with other families.This work is supported by the EPSRC.

> **Chris Illingworth** [1]
[1] University of Cambridge

Evaluating the role of selection in the within-host evolution of the influenza virus
The evolution of the influenza virus is often considered on a global scale, yet proceeds via successive individual infection and transmission events. I aim to understand this latter, small-scale evolution of the virus. Quantifying the role of selection acting upon the virus is made challenging by the presence of interference between allele frequencies caused by hitchhiking and clonal interference, such that observed changes in allele frequencies cannot be simply identified with direct selection. I discuss mathematical and statistical techniques to untangle this interference and quantify the role of within-host selection based upon both long- and short-read sequence data. By means of illustration, I present results inferred from data collected from evolutionary transmission experiments in swine and ferret populations.

> **Véronique Ladret**
Institut de Mathématiques et de Modélisation de Montpellier (I3M), Université de Montpellier, France

Evolutionary game dynamics in a finite continental-island model and emergence of cooperation

We consider the continental island model for a finite haploid population with a total number of n demes consisting in one continent and n-1 islands. We assume viability differences in the population captured by a linear game within each deme as a result of pairwise interactions. Under the structured coalescent hypothesis, assuming weak selection and conservative migration, we derive the first order approximation for the fixation probability of a single mutant, initially introduced in the continent, with respect to the intensity of selection. This result is applied to the case of the iterated Prisoner's Dilemma, when the resident strategy is always defect (AllD) and the mutant cooperative strategy is tit-for-tat (TFT). In this case, we investigate conditions under which selection favors the emergence of cooperation and we derive an extension of the "one-third law" of evolution. We find that the population subdivision of the continental island model weakens the one-third law that holds for a panmictic population. When the model is symmetric, i.e. when the continent and the islands are the same size, we compare this condition to the one obtained when the population structure is replaced by a Wright's finite island model with the same number of demes, the same deme sizes and the same expected total number of migrants per generation after population regulation as in the continental island model. We investigate under what conditions the continental island structure facilitates the emergence of the cooperative tit-for-tat strategy in comparison with its Wright's island model counterpart. And when the deme sizes differ, we investigate how the asymmetry in the deme sizes of the continental island model can better promote the evolution of tit-for-tat compared to its equal deme sizes model counterpart.

> **Murray Patterson** [1]; Thomas Bernard [1]; Daniel Kahn [1,2]
[1] Laboratoire de Biometrie et Biologie Evolutive, UCBL, Villeurbanne;
[2] Departement MIA, INRA

Correlated evolutionary scenarios of metabolic functions

Metabolism refers to the whole set of biochemical reactions that occurs within the cell. A pathway usually designates an arbitrary sub-graph of the whole metabolic network, for instance leading to the biosynthesis of some particular compound. Although numerous studies have been undertaken on metabolic networks, we still understand little about their evolution. Up to now, most of the evolutionary aspects have focused on the evolutionary history of enzymes that catalyze the reactions and not on the evolution of the structural properties of the network itself. The understanding of how a rather complicated and robust network of reactions may have emerged and evolved over time is still an important and open question. We study metabolic functions in over a thousand species for which the complete genome sequence is available, in which several hundred thousand families of homologous proteins have been identified (HOGENOM v.6 [1]). Given a protein sequence, PRIAM search [2] -- using a set of profiles trained on the SWISS-PROT knowledge-base -- is able to deliver probabilities of the presence of enzymes (ECs), assigning possible enzymatic function(s) to thousands of these protein families. By applying EC-specific rules to the set of proteins encoded in a genome, PRIAM also allows to estimate probabilities for each EC in an organism. In each of the species, we are thus able to detect on average several hundred ECs with PRIAM, from which we can generate a draft metabolic network and the corresponding stoichiometry matrix. We then propagate these ECs to the ancestral nodes of the species tree using maximum likelihood methods [3], resulting in draft ancestral metabolic networks. These evolutionary scenarios are systematically compared using pairwise mutual information between ECs. We identify co-evolving enzyme sets from the graph of these relationships using community detection algorithms [4,5], which sheds light on the structure of the metabolic networks in terms of co-evolving metabolic modules. We also study these modules from a functional perspective using stoichiometric models of metabolic networks.

[1] Penel et al., BMC Bioinformatics, 10(6):S3, 2009
[2] Claudel-Renard et al., Nucleic Acids Research, 31(22):6633-6639, 2003
[3] Dutheil and Boussau, BMC Evolutionary Biology, 2(255), 2008
[4] Ahn et al., Nature, 446:761-764, 2010 [5] Blondel et al., Journal of Statistical Mechanics, 2008(10):P10008, 2008

> **Fanny Pouyet** [1], Marc Bailly-Bechet [1], Laurent Guéguen [1]
[1] Laboratoire de Biologie et Biométrie Evolutive, UMR5558, UCB Lyon1, Villeurbanne, France

A new codon substitution model to better estimate evolutionary processes

The genetic code is redundant, and some codons, called synonymous, are translated in the same amino acid (AA). Degeneracy of the code does not however lead to a uniform or a random usage of those synonymous codons : it is called codon usage bias (CUB). It may vary between species and between genes. These differences are easily observable and measurable on extant sequences. However, to fully understand how this bias arose and is maintained in a set of genes, we need a model that studies codon

usage in an evolutionary approach. We develop a codon substitution model, inspired from Yang and Nielsen [Yang, Z. and Nielsen, R. Mol. Biol. Evol, (2008)] that explicitely considers selection on CUB. It is available in Bio++ suite [Guéguen L., et al. Mol. Biol. Evol, (2013)].Our model distinguishes evolution of coding sequences at nucleotidic, codons and AA layer. It explicitely describes separately mutational bias and BGC which applies on all nucleotides independently of their position in the codon, the selection between synonymous codons and the preferences among AA. We distinguish selection on codon usage from observed CUB. We are able to measure the importance of each of these three layers in sequences evolution from simple estimators such as partial effective number of codons (ENC) statistics. We apply our model in an homogeneous and non-stationary context, in a maximum likelihood framework. We study the forces that drive the genomic content in the core genome of twenty species of pathogens bacterias [Lassalle F. et al., Plos Genet, (2015)]. On the one hand, we compare our model to classical codon models [Yang, Z. and Nielsen, R. J Mol Evol., (1998)]. We observe a global AT enrichment at equilibrium in every genomes which agrees with Hershberg and Petrov [Hershberg R. and Petrov D. A., Plos Genet, (2010)]. On the other hand, we can focus on specific species. We contrast codon specific predictions (codons layer) and prediction at the full scale (nucleotides layer). For some AA the codons layer dominates CUB, while for some others nucleotidic layer is stronger. We confirm the existence of both a universal mutation bias towards AT and selection pressure on codon usage, and are able to measure the relative importance of each one in the evolutionary process.

> **François Rousset** [1]
[1] Institut de Biologie Computationnelle (IBC), Montpellier, France and Institut des Sciences de l'Evolution CNRS UMR 5554

A calibrated method for simulation-based inference

I will present the current state of development of a new approach for simulation-based inference, alternative to the Approximate Bayesian Computation (ABC) framework. In particular, this project aims to provide better calibrated confidence intervals, i.e. with better controlled coverage than those ABC may provide. As ABC, it uses simulations of summary statistics under the process which parameters are to be inferred, but it analyzes simulations results differently, essentially providing a likelihood-based analysis of the information retained in the summary statistics.

> **Konrad Scheffler** [1]; Joel Wertheim [1]; Ben Murrell [1]; Martin Smith [1]; Sergei Kosakovsky Pond [1]
[1] University of California, San Diego, USA

RELAX: Detecting Relaxed Selection in a Phylogenetic Framework

Relaxation of selective strength, manifested as a reduction in the efficiency or intensity of natural selection, can drive evolutionary innovation and presage lineage extinction or loss of function. Mechanisms through which selection can be relaxed range from the removal of an existing selective constraint to a reduction in effective population size. Standard methods for estimating the strength and extent of purifying or positive selection from molecular sequence data are not suitable for detecting relaxed selection, because they lack power and can mistake an increase in the intensity of positive selection for relaxation of both purifying and positive selection. Here, we present a general hypothesis testing framework (RELAX) for detecting relaxed selection in a codon-based phylogenetic framework. Given two subsets of branches in a phylogeny, RELAX can determine whether selective strength was relaxed or intensified in one of these subsets relative to the other. We establish the validity of our test via simulations and show that it can distinguish between increased positive selection and a relaxation of selective strength. We also demonstrate the power of RELAX in a variety of biological scenarios where relaxation of selection has been hypothesized or demonstrated previously. We find that obligate and facultative c-proteobacteria endosymbionts of insects are under relaxed selection compared with their free-living rela- tives and obligate endosymbionts are under relaxed selection compared with facultative endosymbionts. Selective strength is also relaxed in asexual Daphnia pulex lineages, compared with sexual lineages. Endogenous, nonfunctional, bornavirus-like elements are found to be under relaxed selection compared with exogenous Borna viruses. Finally, selection on the short-wavelength sensitive, SWS1, opsin genes in echolocating and nonecholocating bats is relaxed only in lineages in which this gene underwent pseudogenization; however, selection on the functional medium/long- wavelength sensitive opsin, M/LWS1, is found to be relaxed in all echolocating bats compared with nonecholocating bats.

> **Magali Semeria**[1], Laurent Guéguen[1], Eric Tannier[1,2]
[1]Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université Claude Bernard Lyon 1, Villeurbanne, France;
[2] INRIA Grenoble Rhône-Alpes, Montbonnot, France;

Probabilistic reconstruction of ancestral gene orders using reconciled gene trees

Phylogenetic reconstruction methods are traditionally based on models of nucleotide or amino-acid sequence evolution. But the evolution of genomes can be studied at different scales, such as the gene level, accounting for gains and losses, and the genome level, accounting for rearrangements of chromosome organization.Models that integrate several of these levels of information reconstruct more plausible evolutionary histories than traditional sequence-based models. For this reason, the sequence and gene levels are now often combined in a single model through reconciled gene trees. However, very few attempts have been made to integrate the genome level as well and reconstruct gene trees and gene orders simultaneously.We propose a model that takes reconciled gene trees and gene orders in extant genomes and computes the probabilities of conservation, origination and loss of gene orders along the branches of the gene trees. We use a framework similar to [1] in which we consider that two consecutive genes on a chromosome share a relationship that we call an adjacency. As two adjacent genes undergo speciation, duplication and loss, the adjacency can be conserved, duplicated or lost. The adjacency can then be considered as an evolutionary object itself and we can represent the evolution of gene orders with a set of phylogenetic trees of adjacencies. We can then compute a pseudo-likelihood of observed adjacencies using the usual dynamic algorithm proposed by Felsenstein.Our model yields the most likely gene orders in ancestral species given reconciled gene trees and extant gene orders. We assess the performance of our model by comparing the ancestral gene orders we obtain with those obtained with [1]. We then show that the probabilistic reconstruction of ancestral genomes can be used as a criterion to evaluate the quality of the reconciled trees given as input [3].We argue that our model is a first step towards the integration of gene order data in phylogeny :[2] already uses a probabilistic framework to reconstruct evolutionary histories by maximizing the combined likelihoods of the alignment and of the reconciliation. Future work will explore the possibility of maximizing alignment, reconciliation and ancestral gene order in a single model.

[1] S. Bérard, C. Gallien, B. Boussau, G. J. Szöllősi, V. Daubin, and E. Tannier, “Evolution of gene neighborhoods within reconciled phylogenies.,” Bioinformatics, vol. 28, no. 18, pp. i382–i388, Sep. 2012.
[2] B. Boussau, G. J. Szöllosi, L. Duret, M. Gouy, E. Tannier, and V. Daubin, “Genome-scale coestimation of species and gene trees.,” Genome Res., vol. 23, no. 2, pp. 323–30, Feb. 2013.
[3] C. Chauve, N. El-Mabrouk, L. Guéguen, M. Semeria, and E. Tannier, “Duplication, Rearrangement and Reconciliation: A Follow-Up 13 Years Later,” in Models and Algorithms for Genome Evolution, C. Chauve, N. El-Mabrouk, and E. Tannier, Eds. Springer, 2013, pp. 47–62.

> **Claudia Solis-Lemus**[1]; Cecile Ane[2]
[1] Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, United States;
[2] Department of Statistics and Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin, United States;

Inferring phylogenetic networks from quartets with maximum pseudolikelihood estimation

Inferring phylogenetic networks to represent the tree of life expanded by hybridization edges is a necessary endeavor because not all species follow the paradigm of vertical inheritance of their genetic material. A great deal of research has flourished for the inference of phylogenetic trees, but methods to infer phylogenetic networks are still limited and under development. Furthermore, testing whether gene tree discordance can be explained entirely by incomplete lineage sorting or we need to invoke reticulation events is still an difficult question. Here, we respond to both problems by developing (i) a statistical method to infer phylogenetic networks from sequence data in a pseudolikelihood framework and (ii) a theoretical test to determine whether a species tree (or a species network) is a good fit for the gene tree discordance. Our methodology has two main advantages. First, the computation of the pseudolikelihood is fast and simple. It avoids the burdensome calculation of the full likelihood, which is untractable on many species and many reticulations. Second, the estimation at the quartet-level is easily parallelizable. Our techniques to learn phylogenetic networks will enable scientists to incorporate organisms to the "tree of life" in parts that are more net-like than tree-like, and thus, complete a broader picture of evolution.

> **Min Ye**; Bernard Moret
EPFL, 1015 Lausanne, Switzerland

A generative model for protein protein interaction network evolution based on modularity

The study of the evolution of biological networks remains in its infancy. Most work to date has been on protein-protein-interaction (PPI) networks and gene regulatory networks, but here we will focus on PPI networks. Several types of events are commonly accepted to play a role in the evolution of PPI networks. They are duplication and loss of proteins (vertices) or interactions (edges). Models have been proposed including one or more of them [1, 2, 3]. Mentioned as a generative model by Middendorf et al. [4], the duplication-mutation with complementarity (DMC) model follows the duplication-divergence principle that allows the modification of either the inherited edge of the newly entered vertex or of its compliment edge. Navlakha and Kingsford have concluded the DMC model to be a suitable network growth model, after analyzing and comparing it against two others [1]. This model has been extended by integrating the age information of proteins [2]. Another parsimonious framework by Patro and Kingsford is based on a directed hypergraph formulation [3]. Reasonable models for PPI network evolution are also crucial for network comparisons or alignments. Alignment algorithms have been proposed, based on vertices but not on edges [5]. Local alignment approaches have been developed to identify functionally conserved modules in multiple networks [6]. We describe a new model that is based explicitly on modularity and thus goes beyond the existing models of adding or deleting vertices or edges. Our model includes the development and disappearance of modules, as driven by the inheritance and evolution of their internal connectivity as well as by their external connections to other modules. Such a model is a first step toward phylogenetic analysis of PPI networks and inference of ancestral PPI networks. We present comparisons between our model and existing ones in terms of their generative characteristics.

[1] Saket Navlakha and Carl Kingsford. Network archaeology: Uncovering ancient networks from present-day interactions. PLoS Comput Biol, 7(4):e1001119, 04 2011.

[2] Jiawei Luo, Dan Song, Cheng Liang, and Guanghui Li. Model the evolution of protein interaction network assisted with protein age. Journal of Theoretical Biology,333(0):10 – 17, 2013.

[3] Rob Patro and Carl Kingsford. Predicting protein interactions via parsimonious network history inference. Bioinformatics, 29(13):i237–i246, 2013.

[4] Middendorf M, Ziv E, Wiggins CH (2005) Inferring network mechanisms: the Drosophila melanogaster protein interaction network. Proc Natl Acad Sci USA 102:3192–3197.

[5] Leonid Chindelevitch, Cheng-Yu Ma, Chung-Shou Liao, and Bonnie Berger. Optimizing a global alignment of protein interaction networks. Bioinformatics, 29(21):2765–2773, 2013.

[6] Jialu Hu and Knut Reinert. LocalAli: An Evolutionary-based Local Alignment Approach to Identify Functionally Conserved Modules in Multiple Networks. Bioinformatics, 31(3):363-372, 2014.

POSTERS

Poster 1

> **Bassam AlKindy**[1] ; Christophe Guyeux[1] ; Jean-François Couchot[1] ; Michel Salomon[1] ; Jacques M. Bahi[1]
[1] Institute de FEMTO-ST/DISC, UMR 6174 CNRS, University of Franche-Comté, Besançon, France

Using Genetic Algorithm for Optimizing Phylogenetic Tree Inference in Plant Species

Discovering homoplasy in phylogenetic reconstruction is considered as a difficult computational problem, because the number of situations to investigate dramatically increases with the number of core genes and taxa. We describe here an optimization pipeline of three stages using genetic algorithm that can efficiently optimize the searching space for well supported phylogenetic inference. Genetic algorithms are simulations of natural selection in which selected individuals encode solutions to the problem of interest. The objective is to obtain a well-supported phylogenetic tree (using RAxML) by using the largest possible subset of obtained core genes with preferred outgroup. If this goal cannot be reached by taking all n core genes, the first thing to investigate is to test whether one particular gene is not responsible of this problem. We then systematically compute all trees we can obtain by removing exactly one gene in the core genome, leading to n new phylogenetic trees, where n is the core size. If, during this systematic stage, one well-supported tree is obtained, then it is returned as the phylogeny of the species under consideration. Conversely, if all obtained trees have at least one problematic branch, then deeper investigations are required. Random mode is applied by investigating all phylogenetic trees that can be obtained by removing 2 genes among core genome of size n. Obviously, the number of cases explodes, and hoping to investigate all reachable trees by discarding 10% of a core genome having 100 genes is illusory. This is why genetic algorithm has been proposed for optimization. The optimization stage of genetic algorithm is then employed by selecting the best scoring trees produced from early stages for crossover and mutation operations, to predict new solutions that are computed and added to the solution space for tree reconstruction. The well supported phylogenetic tree is produced and considered as the solution of the given family. We applied this pipeline to various families of plant species, more than 65% of phylogenetic trees produced from this pipeline presents well supported bootstrap value greater than 95.

Poster 2

> **Yoann Anselmetti**[1] ; Sèverine Bérard[1] ; Vincent Berry[2][3] ; Annie Chateau[2][3] ; Éric Tannier[4][5]
[1] Institut des Sciences de l'Évolution de Montpellier (ISE-M), Montpellier, France
[2] Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Montpellier, France
[3] Institut de Biologie Computationnelle (IBC), Montpellier, France;
[4] Laboratoire de Biométrie et Biologie Évolutive (LBBE), Villeurbanne, France
[5] Institut National de Recherche en Informatique et en Automatique (INRIA) Rhône-Alpes, Montbonnot, France

Ancestral genome reconstruction with synteny helps to improve scaffolding of extant species.

Currently, it is difficult to conduct studies on the structural evolution of genomes because for many vertebrates, the complete assembly of genomes is not resolved.For example, in Ensembl release 78 (December 2014), on average, genomes consist of 2885 contigs and for 57 of the 69 species present in the Ensembl database, the genome is composed of over 100 contigs.DeCo (Detection of Coevolution) [1] is an algorithm to reconstruct ancestral gene adjacencies from extant gene adjacencies and reconciled gene trees. We chose to adapt this existing algorithm to consider incomplete genome assembly. In a first step, DeCo reconciles gene trees with species tree by a parsimony approach using LCA (Last Common Ancestor) reconciliation within a model based on gene duplications and losses [2]. Then gene adjacencies observed in current genomes are split into different classes, adjacencies within a class being susceptible of deriving from a common ancestral adjacency. To reconstruct the evolution of such an ancestral adjacency (depicted in an adjacency tree), the algorithm considers a couple of gene families (those involved in the observed adjacency). After assigning a cost to adjacency breaks and creations, a

parsimony principle is applied, preferring an history of the adjacency that is of lower cost (e.g., necessitating the smallest number of hypotheses for explaining current adjacencies).Such an history of minimum cost is inferred by a dynamic programming principle [3] based on a joint exploration of the reconciled trees of the gene families involved in the adjacency.In its first version, DeCo excludes extant genes without adjacency from the analysis. In order to account for these genes in ancestral genome reconstruction, we define the probability for two genes to be adjacent, for each pair of genes. Then, we modified Deco's algorithm to include these probabilities during the inference of the adjacenciy tree of minimum cost.We tested the modified algorithm on a data set covering 11 species, 10,426 gene trees and 224,975 genes generated from Ensembl release 78 with a pipeline that we have developed downstream of the DeCo algorithm. The algorithm works in less than 10 minutes on a computer with 6GB RAM and proposes 779 new adjacencies between extant genes providing a new alternative for the scaffolding of extant genomes. The first analysis of results are very encouraging, for example, the two first adjacencies analyzed on horse's genome propose the fusion between contigs and chromosomes that are highly probable.

[1] Bérard, S. et al. (2012) Evolution of gene neighborhoods within reconciled phylogenies. Bioinformatics, 28, 382-388.
[2] Goodman, M. et al. (1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool., 28, 132–163.
[3] Sankoff, D. (1975) Minimal mutation trees of sequences. SIAM J. Appl. Math., 28, 35-42.

Poster 3

> **Paul Bastide**[1,2] ; Stéphane Robin[1] ; Mahendra Mariadassou[2]
[1] UMR 518 Mathématiques et Informatique Appliquées (MIA) - AgroParisTech/INRA, Paris, France;
[2] UR Mathématiques et Informatique Appliquées du Génome à l'Environnement (MaIAGE) - INRA, Jouy-en-Josas, France

Shifted stochastic processes evolving on trees: application to models of adaptive evolution on phylogenies

In comparative and evolutive ecology, quantitative traits measured on related species can be seen as the outcome of a stochastic process running on a phylogenetic tree. This modeling accounts for correlations between species that share a common evolutionary history [2]. The stochastic process must capture the mechanisms of a trait evolution. Ecologists hence prefer the Orstein-Uhlenbeck (OU) process to the simpler but less realistic Brownian Motion (BM) [3]. The OU process has a tendency to revert to a central value, interpreted in ecology as the optimal value of a trait in a given environment. A shift in this central value therefore models an abrupt change of evolutionary niche. This model has been studied under the hypothesis that the position of the shifts on the tree are known [1]. In this study, we relax this hypothesis and develop an efficient method that allows us to simultaneously infer the (continuous) parameters of the OU process and the position of the shifts. We do so using an incomplete-data formulation of the problem and an Expectation Maximization algorithm. We leverage the tree structure of the data to achieve an efficient implementation (https://github.com/pbastide/Phylogenetic-EM). Evolutionary niches are inherited from ancestral to descendant species throughout the tree until superseded by a change of niche. Our trait-evolution model therefore naturally induces a clustering of the species into unobserved evolutionary niches. Niches are phylogenetically coherent by construction and defined by their optimal trait value. We show that only this clustering, rather than the complete history of shifts, is identifiable. Indeed, we exhibit equivalent shift allocations, i.e. different allocations that lead to the exact same trait distribution at the leaves and therefore can not be distinguished. A recursive algorithm allows us to efficiently count and enumerate all allocations in an equivalence class. This has important implications for biological interpretations: additional information is required to choose between scenarii in an equivalence class.Finally we consider the choice of the number of shifts as a model selection problem and adopt a penalized likelihood criteria. The penalty depends on the number of equivalence classes. We prove a recursive formula to count this number. The formula depends on the tree topology in general except in the binary case where a closed formula is available.

[1] Butler and King (2004). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist, 164(6):683-695.
[2] Felsenstein (1985). Phylogenies and the Comparative Method. The American Naturalist, 125(1):1-15.
[3] Hansen (1997). Stabilizing selection and the comparative analysis of adaptation. Evolution, 51(5):1341-1351

Poster 4

> **Priscila Biller**[1] ; Eric Tannier[2]
[1] Institute of Computing, University of Campinas, Brazil
[2] Laboratoire de Biométrie et Biologie Evolutive (LBBE), UCB Lyon 1, France

Genome rearrangement with a dynamic intergene size distribution

Genome rearrangements as inversions usually break intergenic sequences, with a probability proportional to their sizes. Traditional models of evolution by inversions assume an uniform distribution of intergene sizes, whereas it is neither realistic nor an equilibrium distribution under a process of inversions. We show that the equilibrium distribution is closer to real ones, and that with this equilibrium distribution, distance estimations assuming uniform distribution have a low performance. We propose a new distance estimator inspired from the evolution of random graphs, which can be computed on any distribution. We test it on several simulated and real datasets.

Poster 5

> **Manuel Binet**[1,2,3], Olivier Gascuel[1,2], Celine Scornavacca[3], Emmanuel J.P. Douzery[3] and Fabio Pardi[1,2]
[1] Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier, UMR5506 CNRS, Université de Montpellier, France
[2] Institut de Biologie Computationnelle, UMR5506 CNRS, Université Montpellier, France
[3] Institut des Sciences de l'Evolution de Montpellier, ISEM, UMR5554 CNRS, Université de Montpellier, France

Fast and accurate branch length estimation for phylogenomic trees.

Branch lengths are an important attribute of phylogenetic trees, as they provide essential information for many questions in evolutionary biology. Yet, much of the current methodology in phylogenomics --- the study of evolution at a genomic level --- principally focuses on the topology or structure of a phylogeny, rather than the evolutionary divergences associated to it. Moreover, those methods that do estimate branch lengths are limited by large demands in memory and computing time, and may become impractical when the data sets are too large.Here, we present a novel phylogenomic distance-based method, named ERaBLE (Evolutionary Rates and Branch Length Estimation), to estimate the branch lengths of a given reference topology, and to estimate the relative evolutionary rates of the genes employed in the analysis. ERaBLE uses as input data a potentially very large collection of distancematrices, where each matrix is obtained from a different genomic region --- either directly from its sequence alignment, or indirectly from a gene tree inferred from the alignment.Our experiments show that our method is fast, and relatively accurate when compared to other possible approaches for the same tasks. Specifically, ERaBLE can provide an efficient alternative to the probabilistic analysis of a concatenated alignment for large data sets, such as the OrthoMaM database, composed of 6953 alignments of exonic DNA sequences from up to 40 mammals.

Poster 6

> **Lauriane Cacheux**[1] ; Loïc Ponger[1] ; Florence Richard[2] ; Christophe Escudé[1]
[1] Muséum national d'Histoire naturelle de Paris (MNHN), Paris, France
[2] Université de Versailles-Saint Quentin en Yvelines (UVSQ), Versailles, France

Evolution and diversity of alpha-satellite DNA in cercopithecines

Tandem repeated DNA is highly abundant in the centromeric regions of eukaryotic chromosomes. In primates, the major centromeric DNA repeats are AT-rich and made of monomeric units of 171 or 172bp in length; they are called alpha-satellite DNA [1]. Very few studies focused on the diversity of these repeats, and their evolutionary features remain poorly characterized even if they appear peculiar and unusual compared to classic genomes evolution pathways [2,3]. Cercopithecines are a clade of Old world monkeys composed of thirty five species that have diverged ten million years ago. They evolve through specific processes involving numerous chromosomal fissions, events that systematically require the appearance of so-called evolutionary new centromeres. Therefore, these species represent a relevant model to

study the diversity and evolution of alpha-satellite DNA, and the emergence of centromeres during evolution. With an innovative approach combining enzymatic isolation and NGS of alpha-satellite DNA, we obtained several hundreds of thousands of alpha-satellite sequences from two species of cercopithecines. A bioinformatic analysis (pca, clustering and phylogeny) showed that these sequences can be classified in several families, revealing an unknown genomic and interspecific diversity for alpha-satellite DNA in cercopithecines. Using a cytogenomic approach [4], we showed that the distribution of these families differ between chromosomes and species. For the first time, promising new methods will allow us to infer the way alpha-satellite DNA diversify through a clade of primates. They will reveal some evolutionary features of centromeric DNA which could probably be extended far beyond this group. Furthermore, we will be able to determine if the original diversity and evolution of alpha-satellite sequences can make this DNA a pioneering marker for phylogenetic studies in primates.

[1] Ugarković DI. Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol, 2009; 48:53-76.
[2] Shepelev VA, Alexandrov AA, Yurov YB, Alexandrov IA. The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet, 2009; 5(9):e1000641.
[3] Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. Science, 2001; 293:1098-1102.
[4] Moulin S, Gerbault-Seureau M, Dutrillaux B, Richard F. Phylogenomics of African guenons. Chromosome Res, 2008; 16:783-99.

Poster 7

> **Daniel Doerr** [1] ; Pedro Feijao [2,3] ; Cedric Chauve [4]
[1] Laboratory for Computational Biology and Bioinformatics, School of Computer and Communication Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland;
[2] Genome Informatics, Faculty of Technology
[3] Institute for Bioinformatics, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany; [4] Department of Mathematics, Simon Fraser University, Burnaby, Canada

On the Family-free Breakpoint Median of Three

Recently, a new line of research emerged in the field of computational comparative genomics, whose goal is to devise new methods for gene order analysis that do not require prior gene family assignments [1-3]. Following this idea, we present a model for constructing a breakpoint median of three genomes in a family-free analysis. Our model is a generalization of the well-studied mixed multichromosomal breakpoint median of three genomes [5-6]. We show that the computational complexity of our family-free model is NP-hard by reduction from the weighted independent set problem. We then discuss algorithms for its solution and study their performance on simulated datasets.

[1] Doerr, D., Thévenin, A. & Stoye, J. Gene family assignment-free comparative genomics. BMC Bioinformatics 13, S3 (2012).
[2] Braga, M. D. V., Chauve, C., Doerr, D., Jahn, K., Stoye, S., Thévenin, A., & Wittler, R. The potential of family-free genome comparison. In Models and Algorithms for Genome Evolution, Chr 13 (2013).
[3] Martinez, F. V., Feijao, P., Braga, M. D. V. & Stoye, J. On the family-free DCJ distance. Algorithms in Bioinformatics (2014).
[5] Sankoff, D., Sundaram, G. & Kececioglu, J. Steiner points in the space of genome rearrangements. International Journal of Foundations of Computer Science 7, (1996).
[6] Tannier, E., Zheng, C. & Sankoff, D. Multichromosomal median and halving problems under different genomic distances. BMC Bioinformatics 10, 120 (2009).

Poster 8

> **Janina Dordel**[1] ; Matt W Mitchell[1] ; Peter H Sudmant [2] ; Mary Katherine Gonder [1]
[1] Drexel University, Philadelphia, USA
[2] Massachusetts Institute of Technology (MIT), Cambridge, USA

Towards a high resolution understanding of the evolutionary forces shaping the population structure of common chimpanzees

Despite our close evolutionary relationship and our deep curiosity for the common chimpanzee we are only starting to understand the rich genetic diversity within this species. Recent advantages in utilizing genetic markers have revealed that Pan troglodytes can be divided into two major geographically and genetically distinct lineages consisting of two subspecies each: a western African group that includes P. t. verus and P. t. ellioti and a central/eastern African group including P. t. troglodytes and P. t. schweinfurthii. The natural division of the two lineages is the

Sanaga River in Cameroon. However, we have shown that this barrier and therewith neutral evolutionary processes alone can not explain the differentiation between P. t. ellioti and P. t. troglodytes. Caveats of previous studies aimed to understand the population structure of chimpanzees and the forces that drive diversification in these populations include the use of evolutionary neutral markers, small numbers of genetic markers and individuals, and the use of samples from captive chimpanzees in contrast to wild living individuals. To tackle these problems we present a high resolution SNP typing scheme based on the whole genome sequences of 31 individuals from all four subspecies. Whole genome sequences were mapped against the P. t. verus reference genome and high quality variants were identified using a haplotype-based variant caller and subsequent filtering. From the resulting variants ~4,000 neutral SNPs spaced across the genome and defining and reflecting the diversity of the four subspecies were selected. Furthermore, around 2,000 SNPs under selection or of general interest (e.g. in known pathogen-related loci) were determined. All SNPs were then used to genotype our extensive collection of geo-referenced non-invasive DNA samples. Screening large numbers of geo-referenced samples using a large number of SNPs will help to understand the population structure of the four chimpanzee subspecies in a spatial context. The inclusion of markers under selection will furthermore facilitate the identification of the forces that shape those populations and will shed light on the distribution of genetic loci that are involved in e.g. host-pathogen interactions, pathogen resistance, and adaptation. Chimpanzees are an endangered species and threatened by climate change. Understanding diversity and the processes leading to and maintaining it are not of scientific interest only – more importantly this knowledge can help to inform conservation strategies and save the common chimpanzee.

Poster 9

> **Wandrille Duchemin** [1] ; Eric Tannier [1][2]
[1] Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne F-69622, France;
[2] INRIA Rhône-Alpes, Montbonnot F-38322, France

Reconstruction and comparison of an ancestral genome with an ancient genome of Yersinia pestis

Ancient genomes are sequenced from archeological remains, whileancestral genomes are computed using the genomes of their extantdescendants. Ancient genomes allow the study of extinct organisms withoutdescendants, while ancestral reconstruction methods enable the potentialreconstruction of much older genomes. These two complementary means ofexploring the past forms of life on earth can benefit from the comparativestudy of one another. For instance, ancestral genomes can be used as a template for ancientgenome assembly [1], or for the design of better baits for ancient geneticmaterial. Conversely, ancient genomes are good tools for the validation ofancestral genomes (usually restricted to robustness analysis). We propose a method to reconstruct an ancestral genome in order to compareit to an extinct descendant. We illustrate our method with the example ofthe bacteria Yersinia pestis and its late medieval ancient genome [2].Because of their internal position in the phylogenetic tree, ancient andancestral genomes can be expected to show greater similarity than with anyof the other extant species. We compare the Yersinia pestis ancestral andancient sequences both in terms of local mutations (SNPs, small insertionsand deletions), but also and particularly in terms of structuralrearrangements (losses and duplications, inversions,rearrangements), in order to have insights on their adaptive potential.

[1] Rajaraman et al., 2013, Bioinformatics, doi:10.1093/bioinformatics/btt527
[2] Bos et al., 2013, Nature, doi : 10.1038/nature10549.A

Poster 10

> **Katharina Jahn** ; Jack Kuipers ; Niko Beerenwinkel
D-BSSE, ETH Zurich, Basel, Switzerland

Probabilistic approaches to estimate the mutation history of a tumor

The recent development of single-cell sequencing techniques revealed that the genetic make-up of tumors is better described as a heterogeneous cell population than a monoclonal cell mass. Sequencing data of a large number of single cells can be used to reconstruct the evolutionary

history of a tumor population. A key task in this procedure is to populate the ancestral states in the phylogenetic tree with mutations that split the cell population into subclones. Recurrent mutation orders observed in multiple tumor instances may lead to a better understanding of the mutational patterns associated with a specific tumor type, and help with the identification of tumor subtypes. Currently one of the major challenges in analyzing single cell sequencing data is its low quality due to the limited amount of DNA obtainable from a single cell. The main sources of error are a high allelic dropout rate and an increased false discovery rate compared to bulk sequencing.We introduce a probabilistic approach to estimate the mutation history of a heterogeneous tumor based on single-cell sequencing data that can deal with various error-types in the data. The method is evaluated in a simulation study and used to reconstruct mutation histories of different tumor types.

Poster 11

> **Anne Kupczok** ; Giddy Landan ; Tal Dagan
Institute for general microbiology, University of Kiel, Germany

The contribution of genetic recombination to CRISPR array evolution

CRISPR (clustered regularly interspaced short palindromic repeats) is a microbial immune system against foreign DNA. Recognition sequences (spacers) encoded within the CRISPR array mediate immunity in a sequence-specific manner. Mechanisms described for CRISPR array evolution are spacer acquisition at the 5' end and deletion of successive spacers throughout arrays. Here we study the contribution of genetic recombination between homologous CRISPR arrays to the evolution of spacer repertoire. Acquisition of spacers from exogenic arrays may confer the recipient with immunity against unencountered antagonists.To detect recombination events, we compare the spacer order in CRISPR arrays from multiple strains of a single species. In the absence of recombination, we expect the ordering to be conserved on the 3' end of the CRISPR array and diversified on the 5' end. Lateral spacer transfer can introduce a different pattern of spacer content similarity, termed order divergence event, that is composed of a shared segment of spacers followed towards the 3' end by diverse spacers. To this end, we construct a spacer graph, where nodes designate unique spacers and directed edges connect spacers that are consecutive in the 5' to 3' direction. We note, however, that order divergence events can be generated by two additional scenarios. Independent acquisitions of the same spacer sequence due to biased sampling of protospacers from invasive genomes may lead to a shared segment of a single spacer. In addition, pervasive deletions in CRISPR arrays may create proximal deletions resulting in order divergence events. To estimate the frequency of order divergence events created by deletions, the ratio of order divergence to deletion events in the data set is compared to the corresponding ratio in simulation replicates with simulated deletions but absence of recombination.We analyze spacer content and order in four bacterial species. We find that CRISPR array evolution in E. coli and S. agalactiae can be explained solely by vertical inheritance and proximal spacer deletion. In P. aeruginosa, we find an excess of single spacers potentially incorporated into the CRISPR loci during independent acquisition events. S. thermophilus shows the highest ratio of order divergence to deletion events among the data sets analyzed here. This high ratio cannot be explained by proximal deletions or independent acquisitions and is thus attributed to recombination.Genetic recombination has been proposed to accelerate adaptation by combining beneficial mutations that arose in independent lineages. However, for most species under study, we find that CRISPR evolution is shaped by spacer acquisition and loss rather than recombination. Since the evolution of spacer content is characterized by a rapid turnover, it is likely that recombination is not beneficial for improving phage resistance, or that it cannot be detected in the resolution of intra-species comparisons.

Poster 12

> **Laurent Guéguen**
Laboratoire de Biométrie et Biologie Évolutive (LBBE), Lyon, France

Computing Models of Sequence Evolution in a Coalescent Context

Usual phylogenetic studies consider gene trees directly as they are computed from sequences. But for a better understanding of evolution, these gene trees should be considered in the context

of species trees, where features and mechanisms typical of population genetics occur (such as polymorphism, recombination, hybridization, ...). More and more methods are built to handle jointly phylogenetic and population genetic evolutionary processes. Ancestral recombination graphs [6] and phylogenetic networks [5] are tools to get beyond the concept of tree to represent evolution processes.As for lineages and polymorphism, coalescence theory provides efficient concepts and for-mula to consider sequence evolution at the scale of lineages within species [4, 8, 7, 1] and jointly several evolution modelings have been built to handle explicitly polymorphism [3, 2]. I present a method to integrate models of sequence evolution inside a coalescent process, and to compute expected -- among lineage histories --transition probabilities between sequences. This method can be built upon any kind of sequence evolution model (nucleotidic, codon or proteic). Once the transition probabilities are built, it is very rapid to perform phylogenetic studies (in the same manner as in "usual" phylogenetic) that consider explicitly population genetics features, such as incomplete lineage sorting or ancestral polymorphism.

[1] D. Bryant, R. Bouckaert, J. Felsenstein, N.A. Rosenberg, and A. RoyChoudhury. Infering species trees directly from biallelic genetic markers: Bypassing gene tree in a full coalescent analysis. Mol. Biol. Evol., 29:1917--1932, 2012.
[2] N. De Maio, C. Schlötterer, and C. Kosiol. Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. Mol. Biol.Evol., 30(10):2249--2262, 2014.
[3] I. Gronau, L. Arbiza, J. Mohammed, and A. Siepel. Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. Mol. Biol. Evol., 30:1159--1171, 2013.
[4] A. Hobolth, O.F. Christensen, T. Mailund, and M.H. Schierup. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLOS Genetics, 3(2), 2007.
[5] K.J. Liu, J. Dai, K. Truong, Y. Song, M.H. Kohn, and L. Nakhleh. An HMM-based comparative genomic framework for detecting introgression in eukariotes. PLOS Comp. Biol.,10(6), 2014.
[6] M.D. Rasmussen, M.J. Hubisz, I. Gronau, and A. Siepel. Genome-wide inference of ancestral recombination graphs. PLOS Genetics, 10(5), 2014.
[7] Y. Wu. Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. Evolution, 66(3):763--775, 2011.[8] Z. Yang and B. Rannala. Bayesian species delimitation using multilocus sequence data. PNAS, 107:9264--9269, 2010.

Poster 13

> **Pierpaolo Maisano Delser** [1] ; Shannon Corrigan [2] ; Gavin Naylor [2] ; Stefano Mona [1]
[1] Muséum national d'Histoire naturelle, Paris, France
[2] College of Charleston, Charleston, USA

A shark tale: population genomics of Carcharhinus melanopterus.

Population genomics of non-model organisms is generally challenging to perform due to the lack of a reference genome. Here, we applied a recently developed re-capturing approach to deep sequence ~1000 independent autosomal regions (Li et al. 2013) of the blacktip reef shark (Carcharhinus melanopterus). To investigate its demographic history and population structure, we designed a spatial sampling approach studying individuals from a single deme (Australia) and a scatter sample (from the whole Indian Ocean) sensu Wakeley (Wakeley 1998, 1999), for a total of 18 sharks. We developed an Approximate Bayesian Computation (ABC) algorithm with recombination using the Site Frequency Spectrum computed on unphased data as summary statistics. We contrasted the genealogical signature of population dynamics detected from both samples (Australia vs. scatter) finding strong signals of demographic expansion in the scatter sample while Australia suggested a pattern compatible with a constant population-size model. Theoretical predictions of meta-population models suggest that this is compatible with a low Nm value. This is consistent with the migratory pattern of this philopatric species and the moderately high Fst found using microsatellite markers (Vignaud et al. 2013). To further corroborate our hypothesis, we tested a non-equilibrium finite island model on both samples. ABC model selection confirmed that a finite island is the most supported model on the two sampling schemes. Moreover, both of them provided an estimate of Nm≈40, consistent with the lack of signature of population expansion from Australia. In summary, we showed how with new NGS approaches and an ad hoc sampling scheme even few individuals can provide a wealth of information on the spatial structure of a species.Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJ (2013) Capturing protein-coding genes across highly divergent species. Biotechniques 54: 321-6Vignaud T, Clua E, Mourier J, Maynard J, Planes S (2013) Microsatellite analyses of blacktip reef sharks (Carcharhinus melanopterus) in a fragmented environment show structured clusters. PLoS One 8: e61067Wakeley J (1998) Segregating sites in Wright's island model. Theor Popul Biol 53: 166-74Wakeley J (1999) Nonequilibrium migration in human history. Genetics 153: 1863-71

Poster 14

> **Coralie Merle** [1,2,3] ; Raphaël Leblois [2,3] ; Pierre Pudlo [1,3]
[1] Institut de Mathématiques et de Modélisation de Montpellier (I3M), Universitée de Montpellier, France
[2] Centre de Biologie pour la Gestion des Populations (CBGP) INRA, Montferrier-sur-Lez, France
[3] Institut de Biologie Computationnelle (IBC), Montpellier, France.

Resampling : an improvement of Importance Sampling in vaying population size models

Importance Sampling (IS) distributions on coalescent histories become inefficient when the population size varies in time, making the inference in many demo-graphical situations impossible. Construct a new IS distribution, efficient for varying population size models, is a problem that is not solved yet. The question is : is it possible to improve the inference in this framework without a new IS proposal ? In this work, we aim to improve the efficiency of the inference based on IS, still using an inefficient IS distribution adapted from the constant population size IS distribution. To this end, we applied a resampling technic, for the first time in the Sequential Importance Sam-pling (SIS) algorithm in order to reduce the computational cost for the same accuracy of inference. We tested our new algorithm, called Sequential Importance Sampling with Resampling (SISR) on simulated data sets under different demographical cases. In most demographic cases, we divided the computational cost by 2 for the same accuracy of in-ference, in some cases even by 10. These results reveal that resampling technics improve the efficiency of the inference with IS on coalescent histories for demographic models with population size varying in time. The possible implications of this study are to make the inference possible for new demographical situations ; moreover to bring out the advantage of the resampling technic.

Poster 15

> **Venelin Mitov**[1]; Tanja Stadler[1];
[1] Department of Biosystems Science and Engineering (D-BSSE), Basel, Switzerland;

Importance of trait selection in estimating the viral contribution to virulence of an HIV infection

It has been observed that HIV-1 infected patients have a largely variable time to AIDS. The inverse of this time span is called virulence. It is controversially discussed to what extent the viral genotype and to what extent the patient's immune system and other environmental factors influence virulence. Proving the existence of viral genetic traits with significant effect on virulence has at least two important implications: 1) From a clinical perspective, it would motivate the search for viral targets of a vaccine or a treatment; 2) From an evolutionary perspective, if such viral traits tend to be preserved during the transmission and the timecourse of an infection, these would provide the missing part in support of the hypothesis that, at host population level, stabilizing selection favors the proliferation of viral strains coding for an optimal trade-off between infectiousness and longevity of the carriers.Thanks to treatment, we do not have direct information on HIV virulence, and thus, consider set point viral load (spVL) as a proxy for the inverse of virulence. In order to quantify the contribution of the virus to spVL, the so-called heritability of spVL is estimated. Borrowed from quantitative genetics, the term heritability is defined in two senses: The broad sense heritability, H^2, is the fraction of phenotypic variance attributable to transmittable and non-transmittable viral genetic factors and is a good descriptor of the viral contribution to spVL with respect to its clinical implication; The narrow sense heritability, h^2, is the fraction of phenotypic variance attributable to those viral genetic factors that are preserved during transmission and is a good descriptor of the viral contribution to spVL with respect to its evolutionary implication. I'll begin the talk with a broad/narrow sense interpretation of the current estimates of heritability of spVL that were obtained from traditional donor-recipient regression and more recent phylogenetic methods. Then, I'll use simulation results to show that the currently widely used phylogenetic methods based on a Brownian Motion (BM) trait evolution model significantly underestimate heritability in the case of stabilizing selection. Next, I'll present a new phylogenetic tool to estimate heritability of spVL that explicitly takes selection into account. Finally, I'll show results of applying this tool to HIV genetic sequencing and spVL data from Switzerland and the UK proving that the HIV genotype has a significant impact on spVL, with estimated narrow sense heritability 0.2.

Poster 16

> **Amanda Navas**[1]; Gudrun Brockmann[2]; Martin Schlather[3]; Steffen Weigend[4]; Henner Simianer[1]
[1] Georg-August University, Goettingen, Germany; [2]Humboldt University, Berlin, Germany; [3]Friedrich-Loeffler Institute, Neustadt, Germany

A spatial scan statistic as a tool for uncovering heterozygosity clusters in highly inbred lines

Inbreeding is a process that drastically reduces genetic diversity in a population if it persists for many generations, by which time genome-wide heterozygosity levels should be almost zero. However, in several highly inbred chicken lines heterozygosity is still found to be remarkably high for certain genomic regions. The main objective in this study is to determine if there are compact regions of preserved heterozygosity in the chromosomes of inbred chicken lines, and to determine what process could be generating said regions. The data was obtained by genotyping four inbred chicken lines from three different breeds with the 600k SNP-chip from Affymetrix. The lines differed in breeding history, selection type and inbreeding level reflected by the overall proportion of heterozygous SNPs, spanning from 1.18% to 13.84%. To evaluate if a region of a chromosome had larger heterozygosity than expected by chance we designed a spatial scan statistic that scans throughout a chromosome with windows of variable sizes, and compares such windows with permutations of the data to detect if the cluster is significantly different from the background levels of heterozygosity. We found several clusters of heterozygosity in many of the chromosomes in all lines, differing in length and location between chromosomes and lines. Some clusters coincide with gene positions. One of the most intriguing findings was a region of chromosome 20 in the C-line (a line that is 99.9% homozygous), which showed the formation of a heterozygous haplotype in one sub-line, while the other sub-lines remained fully homozygous. We have yet to determine if these highly different patterns accross lines, sub-lines and chromosomes are due to selective pressures, genetic drift or processes such as inversions. Further steps will involve analyzing inbred lines from other organisms, to evaluate the generality of our findings, as well as working with sequences, and evaluating the validity of our methodological approach using a spatial scan statistic to detect such clusters.

Poster 17

> David Salthouse[1]; Bernard Cazelles[1]
[1] IBENS CNRS, Paris, France

Phylodynamics framework for describing dengue epidemics in the French Polynesian

As opposed to some continental countries where the dengue virus is endemic and multiple serotypes co-circulate, French Polynesian outbreaks of the disease are characterised by the absence of co-circulation and a rapid propagation of the virus between the different peninsular islands. The Institut Malardé has recorded the weekly number of reported dengue cases, together with their associated genetic sequences.Various models have been postulated to describe the stochastic non-linear epidemiological dynamics and recent advances in Bayesian inference are particularly promising to validate these models. Using the algorithms of Stadler [1] and Rasmussen [2] to combine birth-death or coalescent based phylogenetic trees in conjunction with the time series of cases we attempt to better infer model parameters of dengue dynamical models.Here we present the early results of working with the Institute Malardé's data.

[1] Stadler T., Bonhoeffer S., 2013, "Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods." Phil. Trans. R. Soc. B
[2] Rasmussen D. A., Volz E. M. and Koelle K., 2014, "Phylodynamic inference for structured epidemiological models." PLoS comp. Biol.

Poster 18

> **Guillaume Scholz**[1]; Katharina. T. Huber[1]
[1] University of East Anglia (UEA), Norwich, UK

Bridging the gap between rooted and unrooted networks

Phylogenetic trees have been introduced to investigate evolutionary relationships between genes or species. However, they are unable to capture efficiently some evolutionary events, such as hybridizations. This is why people turned their interest to rooted phylogenetic networks.

The simplest class of such networks which encompasses the class of phylogenetic trees is the class of 1-nested (phylogenetic) networks. An important problem concerning those structures is to find good ways to construct them. We present here concepts from split networks, which are a certain type of unrooted phylogenetic network, to shed light into this question.

Poster 19

> **Mingfu Shao**[1]; Bernard M.E. Moret
[1] Laboratory for Computational Biology and Bioinformatics, EPFL, Lausanne, Switzerland

Comparing Genomes with Rearrangements and Segmental Duplications

Large-scale evolutionary events such as genomic rearrangements and segmental duplications form an important part of the evolution of genomes and are widely studied from both biological and computational perspectives. A basic computational problem is to infer these events in the evolutionary history for given modern genomes, a task for which many algorithms have been proposed under various constraints. Algorithms that can handle both rearrangements and content-modifying events such as duplications and losses remain few and limited in their applicability. We study the comparison of two genomes under a model including general rearrangements (through DCJ) and segmental duplications. We formulate the comparison as an optimization problem, and describe an exact algorithm to solve it by using an integer linear program. We also devise a sufficient condition and an efficient algorithm to identify optimal substructures, which can simplify the problem while preserving optimality. Using the optimal substructures with the ILP formulation yields an exact, yet practical, algorithm -- the first practical method to provide exact solutions to the problem of comparing two arbitrary genomes under rearrangements and duplications. We then apply our algorithm to assign in-paralogs and orthologs (a necessary step in handling duplications), and compare its performance with that of the state-of-the-art method MSOAR (an approximation method), using both simulations and real data. On simulated datasets our method outperforms MSOAR by a significant margin, and on 5 well-annotated species, MSOAR achieves high accuracy, yet our method performs slightly better on each of the 10 pairwise comparisons.

ATTENDEES

ACHAZ	Guillaume	Université Pierre and Marie Curie, FR
ALKINDY	Bassam	University of Franche-Comte, FR
ALTHERMELER	Nicole	Bielefeld University, DE
ANSELMETTI	Yoann	Institut des Sciences de l'Évolution de Montpellier (ISE-M), FR
ARNDT	Peter	Max Planck Institute for Molecular Genetics, DE
BASTIDE	Paul	AgroParisTech - Université Paris Sud, FR
BERRY	Vincent	IBC, Lirmm, FR
BILLER	Priscila	University of Campinas, BR
BINET	Manuel	LIRMM-ISEM-IBC, FR
BRYANT	David	University of Otago, NZ, NZ
CACHEUX	Lauriane	Muséum national d'Histoire naturelle de paris, FR
CASSAN	Elodie	Université de Montpellier, FR
CASTEL	Guillaume	INRA, FR
CORANDER	Jukka	University of Helsinki, FI
DALLA RIVA	Giulio Valentino	Biomathematics Research Centre - University of Canterbury, NZ
DANCHIN	Etienne G. J.	Institut Sophia Agrobiotech, UMR 1355, INRA, CNRS, Université de Nice Sophia Antipolis, FR
DOERR	Daniel	Laboratory for Computational Biology and Bioinformatics, School of Computer and Communication Sciences, École polytechnique fédérale de Lausanne, CH
DORDEL	Janina	Drexel University, US
DUCHEMIN	Wandrille	UMR CNRS 5558 - LBBE "Biométrie et Biologie évolutive", FR
GASCUEL	Olivier	Institut de Biologie Computationnelle - Montpellier, FR
GEMMELL	Patrick	University of Oxford, GB
HOBOLTH	Asger	Bioinformatics Research Center (BiRC), Aarhus University, DK
ILLINGWORTH	Chris	University of Cambridge, GB
JAHN	Katharina	ETH Zurich, CH
KUPCZOK	Anne	University of Kiel, DE
LADRET	Véronique	Institut de Mathématiques et de Modélisation de Montpellier (I3M), Université de Montpellier, FR
LARTILLOT	Nicolas	LBBE, FR
LAURENT	Guéguen	LBBE - Université Lyon 1, FR
LEBLOIS	Raphael	INRA, FR
LEMEY	Philippe	Rega Institute, Clinical and Epidemiological Virology, BE
LONGUEVILLE	Jean-emmanuel	CNRS, FR
MAISANO DELSER	Pierpaolo	Muséum national d'Histoire naturelle, FR
MARIADASSOU	Mahendra	INRA - MaIAGE, FR
MARIN	Jean-Michel	Universite de Montpellier, FR
MATIAS	Catherine	CNRS - LPMA, FR
MERLE	Coralie	Université de Montpellier, FR
MITOV	Venelin	ETH Zurich, CH
MORET	Bernard	Laboratory for Computational Biology and Bioinformatics, EPFL, CH
MOURAD	Raphaël	Université Paul Sabatier, FR
NAVAS	Amanda	Georg-August University, Goettingen, DE
ORLANDO	Ludovic	Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, FR
PARDI	Fabio	CNRS - LIRMM, IBC, FR
PATTERSON	Murray	Universite Claude Bernard Lyon 1, FR
POUYET	Fanny	LBBE, FR
PRZEWORSKI	Molly	Columbia University, New york, USA
ROUSSET	François	Institut des Sciences de l'Evolution CNRS UMR 5554, FR
SALTHOUSE	David	CNRS, FR
SAULNIER	Emma	MIVEGEC, IBC, FR
SCHEFFLER	Konrad	University of California San Diego, US
SCHOLZ	Guillaume	University of East Anglia, GB
SCORNAVACCA	Céline	I-SEM, CNRS UM2 IRD, FR
SEMERIA	Magali	Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université Claude Bernard Lyon 1, FR
SHAO	Mingfu	EPFL, CH
SOLIS-LEMUS	Claudia	University of Wisconsin-Madison, US
ST. JOHN	Katherine	City University of New York, US
SWENSON	Krister	MAB - LIRMM - Universite de Montpellier, FR
TRUSZKOWSKY	Jakub	EBI Goldman Group and Cancer Research UK, University of Cambridge
VILLABONA	Ch. Julian	Institut de Biologie Computationnelle - IRD, FR
YE	Min	EPFL, CH