June 12-16, 2016

Hameau de l'étoile, Montpellier, France

MATHEMATICAL AND COMPUTATIONAL EVOLUTIONARY BIOLOGY

INFORMATION

Meeting Point

Bus station, Parking du grand Saint-Jean
(Close to the train station and served by the tram)
A « Bancarel » bus will leave Montpellier on Sunday at **17H00**.

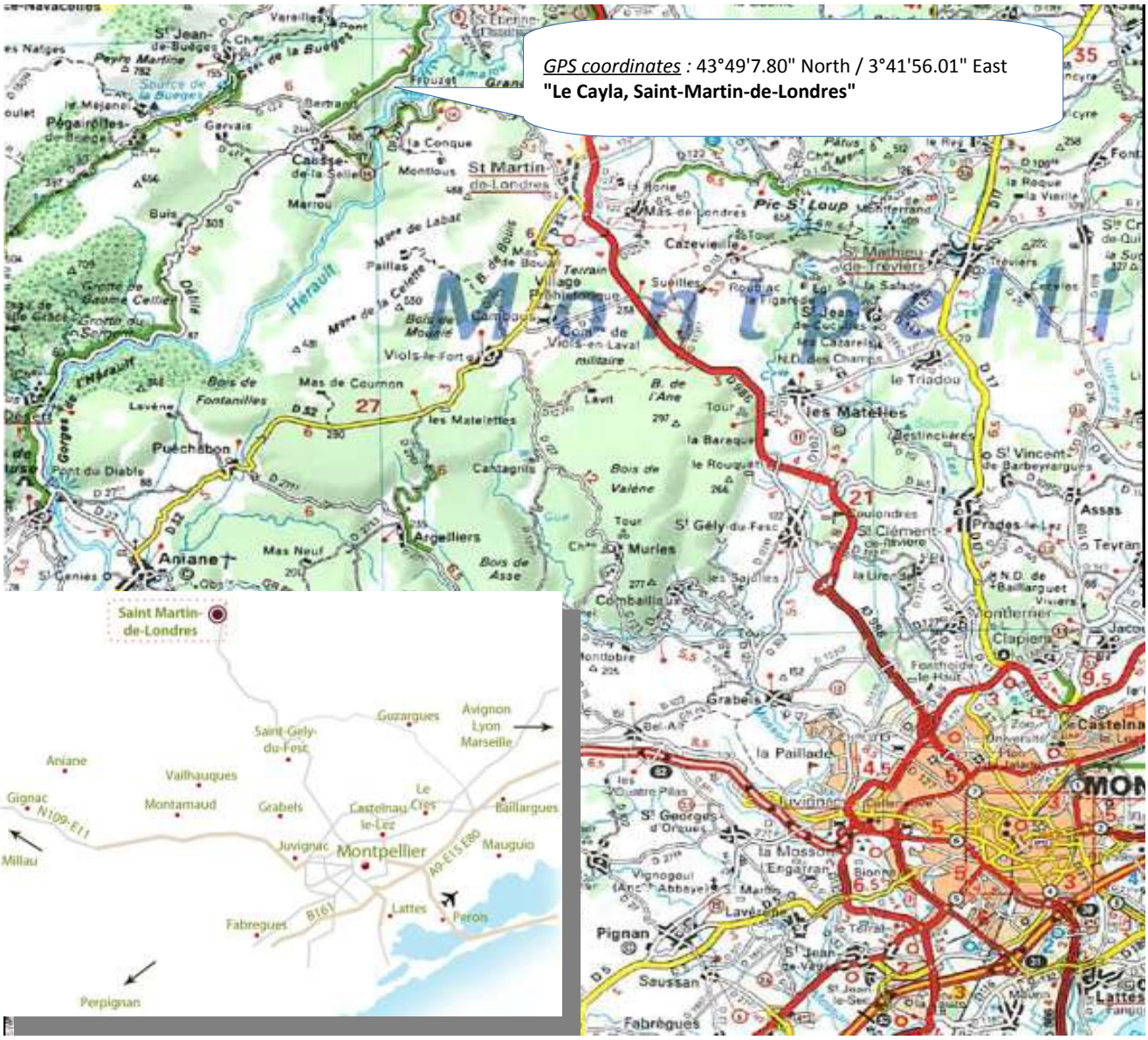
In case of problems :

Olivier Gascuel : 33 (0) 06 48 12 14 82

Sylvain Milanesi : 33 (0) 06 74 13 05 58

Location

The conference will be held at the ***Hameau de l'Etoile***, a hamlet dedicated to seminars and conferences, located at St Martin de Londres, about 25 km north of Montpellier (south of France).



Practical informations

Domaine Le Hameau de l'Etoile
Route de Frouzet

34380 ST-MARTIN-DE-LONDRES

Tél (+33) **04 67 55 75 73**

Fax (+33) 04 67 55 09 10

Taxi :

Taxi de St Martin de Londres
Call first « Florent » at **06 69 34 03 03**
Ex : rates, in week, tram station Saint-Roch Train Station = 59 € / Airport=74 €
more details : http://www.hameaudeletoile.com/en/access-plan.html

Hotels in Montpellier :

Hôtel d'Aragon *** 10, rue Baudin 34000 MONTPELLIER Tél : 33 (0)4 67 10 70 00 fax : 33 (0)4 67 10 70 01	Hôtel d'Angleterre ** 7, rue Maguelone 34000 MONTPELLIER Tél : 33 (0)4 67 58 59 50 fax : 33 (0)4 67 58 29 52	Hôtel le Mistral ** 25, rue Boussairolles 34000 MONTPELLIER Tél : 33 (0)4 67 58 45 25 / 33 (0)6 60 53 73 40 fax : 33 (0)4 67 58 23 95
Hôtel Le Guilhem *** 18, rue Jean Jacques Rousseau 34000 MONTPELLIER Tél : 33 (0)4 67 52 90 90 fax : 33 (0)4 67 60 67 67	Hôtel des Arceaux ** 33/35, boulevard des Arceaux 34000 MONTPELLIER Tél : 33 (0)4 67 92 03 03 fax : 33 (0)4 67 92 05 09	Hôtel Nova ** 8, rue Richelieu 34000 MONTPELLIER Tél : 33 (0)4 67 60 79 85 fax : 33 (0)4 67 60 89 06
Newhotel du Midi *** 22, boulevard Victor Hugo 34000 MONTPELLIER Tél : 33 (0)4 67 92 69 61 fax : 33 (0)4 67 92 73 63	Hôtel des Arts ** 6, boulevard Victor Hugo 34000 MONTPELLIER Tél : 33 (0)4 67 58 69 20 fax : 33 (0)4 67 58 85 82	Hôtel du Palais ** 3, rue du Palais 34000 MONTPELLIER Tél : 33 (0)4 67 60 47 38 fax : 33 (0)4 67 60 40 23
Royal Hotel *** 8, rue Maguelone 34000 MONTPELLIER Tél : 33 (0)4 67 92 13 36 fax : 33 (0)4 67 92 59 80	Hôtel les Fauvettes * 8, rue Bonnard 34000 MONTPELLIER Tél : 33 (0)4 67 63 17 60 / 33 (0)6 89 26 63 58	Hôtel du Parc ** 8, rue Achille Bégé 34000 MONTPELLIER Tél : 33 (0)4 67 41 16 49 fax : 33 (0)4 67 54 10 05
Hôtel Acapulco ** 445, rue Auguste Broussonnet 34090 MONTPELLIER Tél : 33 (0)4 67 54 12 21 fax : 33 (0)4 67 52 26 10	Hôtel François de Lapeyronie ** 80, rue des Pétètes 34090 MONTPELLIER Tél : 33 (0)4 67 52 52 20 fax : 33 (0)4 67 63 56 65	Hôtel Les Troenes ** 17, avenue Emile Bertin Sans 34040 MONTPELLIER Tél : 33 (0)4 67 04 07 76 / 33 (0)6 29 02 31 17 fax : 33 (0)4 67 61 04 43
Hôtel Les Alizés ** 14, rue Jules Ferry 34000 MONTPELLIER Tél : 33 (0)4 67 12 85 35 fax : 33 (0)4 67 12 85 30	Hôtel Littoral ** 3, Impasse Saint Sauveur 34000 MONTPELLIER Tél : 33 (0)4 67 92 28 10 fax : 33 (0)4 67 92 72 20	

PROGRAM

Sunday, June 12

> 19:30 : Drinks

> 20h30 : Dinner

Monday, June 13

> 09h15 : Opening

> 09h30 – 10h45 : *KEYNOTE* *p.7*

Laura Kubatko (Dpt of Statistics and Evolution, Ecology, and Organismal Biology - The Ohio State University, USA)
« Coalescent-based Species-level Phylogenetic Inference Using Site Pattern Frequencies »

> 10h45 : Coffee break

> 11h15 – 12h35 : 4x 20 min *TALKS* (including questions)

Andreas FUTSCHIK (Johannes Kepler University Linz, AT) *p.13*
« On the inadmissibility of common population genetic estimates and their improvement by using shrinkage »

Alice LEDDA (Imperial College London, UK) *p.14*
« A formal model of clonal expansion in bacterial population genetics »

Daniel DOER (Ecole Polytechnique Fédérale de Lausanne, CH) *p.12*
« Principled synteny analysis »

Nicolas GALTIER (University Montpellier - CNRS, FR) *p.13*
« Incomplete lineage sorting in mammalian phylogenomics »

> 12h45 : Lunch and swimming pool

> 14h15 – 15h15 : 3x20 min *TALKS* (including questions)

Olga CHERNOMOR (CIBIV, Max F. Perutz Laboratories, Medical University of Vienna, AT) *p.11*
« Phylogenomics: Efficient analysis by accounting for phylogenetic terraces »

Sébastien LION (Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, FR) *p.15*
« Parasite evolution in heterogeneous and spatially structured host populations »

Chris WYMANT (Dpt of Infectious Disease Epidemiology, Imperial College London, UK) *p.17*
« Phylogenetics Between and Within: Seeing Transmissions and Dual Infections in HIV Deep Sequence Data »

> 15h15 - 15h30 : Break

> 15h30 – 16h45 : *KEYNOTE* *p.7*

Jeffrey JENSEN (Ecole Polytechnique Fédérale de Lausanne, CH)
« Progress, prospects, and practical suggestions for jointly inferring selection and demography: an application to human cytomegalovirus (HCMV) »

> 16h45 - 19h15 : Freetime, swimming pool...

> 19h15-20h30 : *POSTERS* *p.18 - 23*
Wine and discussions (posters 1 to 12)

> 20h30 : Dinner

Tuesday, June 14

> 09h30 – 10h45 : *KEYNOTE* *p.8*

Bruce RANNALA (UC Davis - University of California, USA)
« Species delimitation »

> 10h45 : Coffee break

> 11h15 – 12h35 : *4x 20 min TALKS* *(including questions)*

Inês FRAGATA (Instituto Gulbenkian de Ciência, PT) *p.12*
« Predictability of long-term, but not short-term phenotypic evolution of Drosophila »

Benjamin LINARD (LIRMM, Montpellier, FR) *p.14*
« Metagenome skimming of species-rich lineages: an alternative to eDNA barcoding »

Adrian GONZALEZ CASANOVA (Weierstraß-Institut Berlin, DE) *p.14*
« Modeling the Lenski experiment »

Stéphane GUINDON (CNRS, Montpellier, FR) *p.14*
« Demographic inference under the coalescent in a spatial continuum »

> 12h45 : Lunch

> 14h00 - 20h30: Free afternoon (canoe, hiking, theorems, etc.)

> 20h30 : Dinner

Wednesday, June 15

> 09h30 – 10h45 : *KEYNOTE* *p.7*

Amaury LAMBERT (UPMC - Université Paris 06, FR)
« Random tree shapes and the future loss of phylogenetic diversity »

> 10h45 : Coffee break

> 11h15 – 12h35 : 4x 20 min *TALKS* (including questions)

Veronika BOSKOVA (ETH Zurich, CH) *p.10*
« Phylogenetic reconstruction of viral quasispecies dynamics »

Luca FERRETTI (Pirbright Institute, UK) *p.12*
« Partial selective sweeps and genetic signatures of ongoing adaptation »

Martin GUILLAUME (Institut des Sciences de l'Evolution, Montpellier, FR) *p.16*
« Predicting adaptive fitness trajectories in experimental evolution from deleterious mutation data »

Chieh-Hsi WU (University of Oxford, UK) *p.16*
« Co-estimation of migration patterns and demographic history in structured populations using a structured coalescent approach »

> 12h45 : Lunch and swimming pool

> 14h15 – 15h15 : 3x20 min *TALKS* (including questions)

Champak BEERAVOLU REDDY (The City College of New York, USA) *p.9*
« Approximate likelihood inference of arbitrary population histories and recombination from multiple incomplete or whole genome sequences »

Odile MALIET (ENS, FR) *p.15*
« The effects of phylogenetic tree shape and non random extinctions on the expected loss of phylogenetic diversity »

Nicola DE MAIO (University of Oxford, UK) *p.11*
« SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent »

> 15h15 - 15h30 : Break

> 15h30 – 16h45 : *KEYNOTE* *p.7*

Richard NEHER (Max Planck Institute for Developmental Biology, DE)
« Rapid adaptation and the predictability of evolution »

> 16h45 - 19h15 : Freetime, swimming pool...

> 19h15 - 20h30 : *POSTERS* *p.23-27*
Wine and discussions (posters 13 to 24)

> 20h30 : Dinner

Thursday, June 16

> 09h30 – 10h45 : *KEYNOTE* *p.8*

Isabel SANMARTIN BASTIDA (Real Jardín Botánico (RJB-CSIC), Madrid, SP)
« Bayesian Markov Chain Monte Carlo models and their application in biogeography »

> 10h45 : Coffee break

> 11h15 – 12h15 : *3x 20 min TALKS* *(including questions)*

Marta CASANELLAS (Universitat Politècnica de Catalunya, SP) *p.10*
« Linear invariants and the space of phylogenetic mixtures for Felsenstein'81 and other models »

Claudia BANK (Instituto Gulbenkian de Ciência, PT) *p.9*
« On the (un-)predictability of a large intragenic fitness landscape »

Guillaume ACHAZ (MNHN, FR) *p.9*
« Deeper into the duality between coalescent and drift »

> 12h30 : Lunch, swimming pool, goodbyes etc

> 14h00 : Bus back to Montpellier (train station ~15:30, airport ~16:00)

KEYNOTE SPEAKERS

> **Jeffrey JENSEN**
Ecole Polytechnique Fédérale de Lausanne

Progress, prospects, and practical suggestions for jointly inferring selection and demography: an application to human cytomegalovirus (HCMV)
I will begin with a practical overview of the current best practices in the field with regards to inferring the effects of selection and demography from genomic variation. I will further outline challenges which remain outstanding, and suggest some potential routes forward. Finally, I will discuss an application of such inference to the HCMV system, as well as highlight the important role that population genetics may play in governing clinical outcomes and treatment strategies.

> **Laura S. KUBATKO**
Departments of Statistics and Evolution, Ecology, and Organismal Biology - The Ohio State University

Coalescent-based Species-level Phylogenetic Inference Using Site Pattern Frequencies
The advent of rapid and inexpensive sequencing technologies has necessitated the development of computationally efficient methods for analyzing sequence data for many genes simultaneously in a phylogenetic framework. The coalescent process is the most commonly used model for linking the underlying genealogies of individual genes with the global species-level phylogeny, but inference under the coalescent model is computationally daunting in the typical inference frameworks (e.g., the likelihood and Bayesian frameworks) due to the dimensionality of the space of both gene trees and species trees. In this talk, I will review recent progress on this problem in which the data are viewed as a collection of site patterns, and the algebraic structure associated with the probability distribution on the site patterns under the coalescent model is utilized for inference. In particular, identifiability results for four-taxon species trees based on site pattern frequencies can be used to build a quartet-based inference algorithm for trees of arbitrary size. Furthermore, species tree branch lengths are also identifiable for four-taxon trees, and simple estimators can be derived for branch lengths based on observed site pattern frequencies. Because the estimators for both the species tree and the branch lengths are derived in a fully model-based framework (i.e., the coalescent process is used to model the relationship between gene trees and the species tree, and standard nucleotide substitution models (GTR+I+G and all submodels) are used for sequence-level evolution), these methods are promising approaches for computationally efficient, model-based inference for the large-scale sequence data available today.

> **Amaury LAMBERT**
Stochastics & Biology Group - Laboratoire de Probabilités & Modèles Aléatoires - UPMC - Université Paris 06

Random tree shapes and the future loss of phylogenetic diversity
We think of a phylogeny of contemporary species as an ultrametric binary tree generated by some macroevolutionary process. The sum of branch lengths of this tree is called phylogenetic diversity and is taken as a measure of the evolutionary heritage carried by the clade. First, we seek to characterize mathematically the footprint that the macroevolutionary process leaves on the shape of the tree. To do this, we review families of random tree shapes, including Aldous' beta family, and use these to characterize the trees produced by some well-known stochastic processes of macro-evolution. Second, we study the effect on phylogenetic diversity of random extinctions of contemporary species. We show how the future of phylogenetic diversity loss can be predicted by the distribution of the tree shape.

> **Richard NEHER**
Max Planck Institute for Developmental Biology

Rapid adaptation and the predictability of evolution
Evolution is simple if adaptive mutations appear one at a time. However, in large microbial populations many mutations arise simultaneously resulting in a complex dynamics of competing variants. I will discuss recent insight into universal properties of such rapidly adapting populations and compare model predictions to whole genome deep sequencing data of HIV-1 populations at many consecutive time points. Genetic diversity data can further be used to infer fitness of individuals in a population sample and predict successful genotypes. We validate these prediction using historical influenza virus sequence data. Successful predictions of the composition of future influenza virus population could guide strain selection for seasonal influenza vaccines.

> **Bruce RANNALA**
UC Davis - University of California

Species delimitation
DNA-based approaches to systematics have changed dramatically during the last two decades with the rise of DNA barcoding methods and newer multi-locus methods for species delimitation. During the last half-decade, partly driven by the new sequencing technologies, the focus has shifted to multi-locus sequence data and the identification of species within the framework of the multi-species coalescent (MSC). I discuss model-based Bayesian methods for species delimitation that have been developed in recent years using the MSC. Several approximate methods for species delimitation (and their limitations) are also discussed. Explicit species delimitation models have the advantage of clarifying more precisely what is being delimited and what assumptions are invoked in doing so. Moreover, the methods can be very powerful when applied to large multi-locus datasets and thus take full advantage of data generated using today's technologies. A simple example is presented for which DNA fingerprinting based on a a sequence distance threshold is certain to fail whereas multi-locus species delimitation methods yield reliable inferences.

> **Isabel SANMARTÍN BASTIDA**
Real Jardín Botánico (RJB-CSIC), Madrid

Bayesian Markov Chain Monte Carlo models and their application in biogeography
Bayesian island biogeographic (BIB) models allow joint estimation of the posterior distribution of phylogenetic relationships, divergence times, and ancestral ranges given molecular data and the geographic location of sequences. Because of their simple underlying biogeographic model, based on continuous-time discrete state Markov chain processes (CTMC), and the use of Bayesian MCMC to ease computational tractability, these models have become very popular in phylogeography to answer a wide range of questions, from routes of viral spread to historical patterns of gene flow across populations. We were first to propose the use of these models in biogeography to estimate dispersal rates and area carrying capacities (equilibrium frequencies) from DNA sequences and species distribution data. Here, I focus on the use of these models to test hypotheses in macroevolution and ecology using datasets of multiple clades inhabiting the same region. For example, we explored the interplay between abiotic versus ecological factors in the long-term carrying capacities of islands or the prevalence of phylogenetic niche conservatism in plant versus animal lineages. The original BIB model was a time-homogenous CTMC process, assuming constancy of rates over time. Extensions of the BIB model have gone in the direction of relaxing the time homogeneity of the Markov process by allowing dispersal rates to vary across time intervals, or the use of non-stationary models for detecting the signal of catastrophic mass extinctions events, evidenced as a sudden decrease in area carrying capacities. By allowing the intensity of these events to depend on clade-specific traits, these models might be used to "predict" the fate of lineages, for example, did some island guilds react more severely to environmental perturbations than others? I illustrate this work with empirical studies on the Canary Islands and the African flora.

TALKS

> Emmanuel Schertzer[1,2]; Amaury Lambert[1,2]; **Guillaume Achaz**[1,3]
[1] SMILE, CIRB UMR7241, Collège de France, Paris, France
[2] LPMA UMR7599, UPMC, Paris, France
[3] Atelier de Bioinformatique, ISyEB UMR7205, MNHN, Paris, France

Deeper into the duality between coalescent and drift

Since the advent of the "neutral theory of molecular evolution", the reference model (H0) that we assume, sometimes blindly, when we analyze population genetics datasets is a mutation-drift equilibrium. In practice, H0 is often the backward coalescent process that arises under the same assumptions. Although, we keep exchanging drift and coalescent, as if they were identical, a closer look at the two processes (forward drift and backward coalescent) is extremely informative on their duality. Does the forward fixation time have the same distribution than the backward coalescent time? Why? To what corresponds the MRCA (defined backward) in the forward fixation drift? How the fixation trajectory relates to the coalescent tree? Can we say something about the allele frequency when we know the coalescent times (and vice-versa)? In this work, we will try to address some fundamental questions on the most popular population genetics model.

> **Claudia Bank**[1,2]; Sebastian Matuszewski[2]; Pamela Cote[3]; Ryan T. Hietpas[3]; Daniel N.A. Bolon[3]; Jeffrey D. Jensen[2]
[1] Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
[3] University of Massachusetts Medical School, Worcester, MA, USA

On the (un-)predictability of a large intragenic fitness landscape

The study of fitness landscapes, which aims at mapping genotypes to fitness, is receiving ever-increasing attention (1). Novel experimental approaches combined with next-generation sequencing methods enable accurate and extensive studies of the fitness effects of mutations – allowing us to validate theoretical predictions and improve our understanding of the shape of the true underlying fitness landscape, and its implications for the predictability and repeatability of evolution. A novel biochemical approach termed EMPIRIC (2) allows us to assess the growth rate of large systematic sets of mutants in a chosen region of a microbial genome. We have previously described the single-step distribution of fitness effects and the cost of adaptation across environments in both yeast and influenza virus (3-5), and the amount of epistasis between pairs of mutations under standard laboratory environmental conditions in yeast (6). Here, we present a uniquely large multi-allelic fitness landscape comprised of 640 engineered mutants that represent all possible combinations of 13 amino-acid changing mutations at 6 sites in the heat-shock protein Hsp90 in S. cerevisiae under elevated salinity. Despite a prevalent pattern of negative epistasis in the landscape, we identify a surprisingly high global fitness peak that is reached via four synergistically interacting mutations. We extend a simulation framework proposed by Draghi and Plotkin (7) to study the potential for, and predictability of, adaptive walks, and apply various statistics to characterize the topology of the landscape (8, 9). Using subsets of mutations, we demonstrate that extrapolation beyond a known part of the landscape is feasible, but that it is confounded by the biased selection of assessed mutations, and by specific mutations acting as epistatic hotspots.

1. J. A. G. M. de Visser, J. Krug, Nat Rev Genet. 15, 480–490 (2014).
2. R. T. Hietpas, J. D. Jensen, D. N. A. Bolon, Proc Natl Acad Sci USA. 108, 7896–7901 (2011).
3. R. T. Hietpas, C. Bank, J. D. Jensen, D. N. A. Bolon, Evolution. 67, 3512–3522 (2013).
4. C. Bank, R. T. Hietpas, A. Wong, D. N. Bolon, J. D. Jensen, Genetics. 196, 841–852 (2014).
5. L. Jiang et al., J. Mol. Biol. (2015), doi:10.1016/j.jmb.2015.11.027.
6. C. Bank, R. T. Hietpas, J. D. Jensen, D. N. A. Bolon, Molecular Biology and Evolution. 32, 229–238 (2015).
7. J. A. Draghi, J. B. Plotkin, Evolution, 3120–3131 (2013).
8. I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, J. Stat. Mech., P01005 (2013).
9. L. Ferretti et al., Journal of Theoretical Biology (2016), doi:10.1016/j.jtbi.2016.01.037.

> **Champak Beeravolu Reddy**[1]; Michael J. Hickerson[1;2;3]; Lynsey Bunnefeld[4]; Laurent A.F. Frantz[5]; Konrad Lohse[4]
[1] Biology Department, The City College of New York, New York, New York 10031, USA
[2] The Graduate Center, The City University of New York, New York, New York 10016, USA
[3] Division of Invertebrate Zoology, American Museum of Natural History, New York, New York 10024, USA
[4] Institute of Evolutionary Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3FL, UK
[5] Paleogenomics and Bio-Archaeology Research Network, Research Laboratory for Archeology and History of Art, University of Oxford, Oxford OX1 3QY, UK

Approximate likelihood inference of arbitrary population histories and recombination from multiple incomplete or whole genome sequences

Inferring the demographic history of populations from whole genome data is an area of active research and the increasing volume of genome data presents an incredible opportunity for inference of population histories with unprecedented resolution. Though, in the process, it also adds significant computational burden which calls for modeling only the more salient aspects of population histories. Current approaches to demographic inference can be broadly divided based on how they deal with linkage between adjacent nucleotides in the genome. Methods based on the Site Frequency Spectrum (SFS) are easy to implement but throw away a lot of information by assuming free recombination. The other class of methods approximate the coalescent process with recombination by considering the histories between two adjacent nucleotides to be sequentially Markovian. Between these two extremes, we build upon an existing analytical approach accounting for linkage and recombination by dividing the genome into blocks of arbitrary size. We make use of the distribution of blockwise SFS (bSFS) patterns, which retains topological and recombinational information, to generate approximate likelihoods via Monte Carlo simulations. We illustrate our Approximate Blockwise Likelihood Estimation (ABLE) approach by reconstructing the demographic history of orangutans from Sumatra and Borneo from whole genome data and compare our results to previous analyses based on the SFS. Since each simulated genealogy contributes to estimating the probability of all bSFS patterns compatible with it, our Monte Carlo approximation of the likelihood is extremely efficient and converges rapidly to analytical expectations. Our method also overcomes the limitations of a previously developed analytic approach which assumes no recombination within sequence blocks and is limited to very small samples of genomes and simple demographic models.

> **Veronika Boskova**[1,2]; Tanja Stadler[1,2]
[1] Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH) Zürich, Basel, Switzerland
[2] SIB Swiss Institute of Bioinformatics, Basel, Switzerland

Phylogenetic reconstruction of viral quasispecies dynamics

Especially in fast evolving and reproducing populations such as RNA viruses, the population of sequences present in one host at a time is often very diverse but also very repetitive. Deep-sequencing approaches allow for quantification of sequences and their diversity. The amount of sequences from sequencing efforts represents a computational overload for current phylogenetic and phylodynamic model implementations in a full Bayesian framework. Heuristic approaches aim at reducing the computational burden by applying the inference models only to a subset of the sequences. One can only use the unique sequences, i.e. ignoring frequencies of the different sequences and instead assuming each one occurs only once [see e.g. Redd et al., 2012 and Wertheim et al., 2013]. Alternatively, only a random subsample of the full dataset can be used to reconstruct the phylogeny and the corresponding population dynamics [see e.g. Poon et al., 2012]. We investigated these heuristics in terms of how much loss of information on dynamic properties of the process occurs. We found that in both cases the computational time is drastically reduced, however, the parameter estimates are less exact, and/or less precise. Based on the identified drawbacks of the heuristics, we propose a new tool for efficient reconstruction of viral epidemiological and evolutionary dynamics from full quasispecies datasets and implement it in BEAST v2.1 [Bouckaert et al., 2014]. The framework takes as input viral haplotypes and their frequencies within the quasispecies reconstructed from the raw reads. Phylogenetic analyses are then performed on the haplotype alignment and the frequency information to obtain exact and precise parameter estimates of the epidemiological and evolutionary process. Enabling the use of complete quasispecies datasets should lead to a more complete picture of pathogen dynamics, insight into transmission bottlenecks, and in more reliable parameter estimation.

References:
Bouckaert, Remco, et al. "BEAST 2: a software platform for Bayesian evolutionary analysis." PLoS Comput Biol 10.4 (2014): e1003537.
Poon, Art FY, et al. "Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data." PLoS Comput Biol 8.11 (2012): e1002753.
Redd, Andrew D., et al. "Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions." Journal of Infectious Diseases 206.9 (2012): 1433-1442.
Wertheim, Joel O., et al. "Phylogenetic relatedness of HIV-1 donor and recipient populations." Journal of Infectious Diseases 207.7 (2013): 1181-1182.

> **Marta Casanellas**[1]; Mike Steel[2]
[1] Universitat Politècnica de Catalunya, Barcelona, Spain
[2] Biomathematics Research Centre, Christchurch, New Zealand

Linear invariants and the space of phylogenetic mixtures for Felsenstein'81 and other models

The reconstruction of phylogenetic trees from molecular sequence data relies on modelling site substitutions by a Markov process, or a mixture of such processes. In general, allowing mixed processes can result in different tree topologies becoming indistinguishable from the data, even for

infinitely long sequences. However, when the underlying Markov process supports linear phylogenetic invariants, then provided these are sufficiently informative, the identifiability of the tree topology can be restored [SV]. We present a class of processes that support linear invariants once the stationary distribution is fixed, the `equal input model'. This model generalizes the `Felsenstein 1981' model from four states to an arbitrary number of states (finite or infinite), and it can also be described by a `random cluster' process. We describe the structure and dimension of the vector space of phylogenetic mixtures (and the complementary space of linear invariants) for any fixed phylogenetic tree (and for all trees -- the so called `model invariants'), on any number n of leaves. Our results are based on [CS], where we combine techniques from discrete random processes and (multi-) linear algebra and give a generalization of Lake's invariants [L].

[CS] M. Casanellas and M. Steel, Phylogenetic mixtures and linear invariants for equal input models, http://arxiv.org/abs/1602.04671

[L] J. Lake, A rate-independent technique for analyais of nucleic acid sequences: evolutionary parsimony, Molec. Biol. Evol., 4 (1987), pp. 167--191.

[SV] D. Stefakovic and E. Vigoda, Phylogeny of mixture models: robustness of maximum likelihood and non-identifiable distributions, J. Comput. Biol., 14 (2007), pp. 156-189.

> **Olga Chernomor**[1]; Bui Quang Minh[1]; Arndt Von Haeseler[1]
[1] Center for Integrative Bioinformatics Vienna (CIBIV), Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria

Phylogenomics: Efficient analysis by accounting for phylogenetic terraces

In phylogenomics the analysis of concatenated gene alignments, the so-called supermatrix, is commonly accompanied by the assumption of partition models. Under such models each gene, or more generally partition, is allowed to evolve under its own evolutionary model. Though partition models provide a more comprehensive analysis of supermatrices, missing data may hamper the tree search algorithms due to the existence of phylogenetic terraces - collections of trees with identical score (maximum likelihood or parsimony score). For sparse supermatrices, the number of terraces and the number of trees on terraces can be very large. If terraces are not taken into account, a lot of computation time might be unnecessarily spent to evaluate many trees that in fact have identical score. To save computation time during the tree search, it is worthwhile to quickly identify such cases. The score of a species tree is the sum of scores for the so-called induced partition trees. Therefore, if the topological rearrangement applied to a species tree does not change the induced partition trees, the score of these partition trees is unchanged. Here, we provide the conditions under which the three most widely used topological rearrangements (nearest neighbor interchange, subtree pruning and regrafting, and tree bisection and reconnection) change the topologies of induced partition trees. During the tree search, these conditions allow us to quickly identify whether we can save computation time on the evaluation of newly encountered trees. We also generalize the phylogenetic terrace concept to partial terraces, which occur more frequently than the original ‘‘full’’ terrace and provide additional timesaving possibilities. We developed and implemented a phylogenetic terrace aware (PTA) data structure in IQ-TREE which provides an efficient detection of partial terraces during the tree search and facilitates synchronisation between species tree and induced partition trees. As we show on 12 real alignments, the identification of (partial) terraces speeded up the tree search with IQ-TREE for up to 5 and 8 times compared to the standard implementation (terrace-unaware) and RAxML, respectively.

(i) O. Chernomor, B.Q. Minh, and A. von Haeseler (2015) Consequences of Common Topological Rearrangements for Partition Trees in Phylogenomic Inference. Journal of Computational Biology, Dec; 22(12):1129-42. (DOI:10.1089/cmb.2015.0146)

(ii) O. Chernomor, A. von Haeseler, and B.Q. Minh. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Under review in Systematic Biology.

> **Nicola De Maio**[1]; Chieh-Hsi Wu[1]; Daniel J Wilson[1]
[1] Nuffield Department of Medicine, University of Oxford, UK

SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent

Exploiting pathogen genomes to reconstruct transmission represents a powerful tool in the fight against infectious disease. However, their interpretation rests on a number of simplifying assumptions that regularly ignore important complexities of real data, in particular within-host evolution and non-sampled patients. Here we propose a new approach to transmission inference called SCOTTI (Structured COalescent Transmission Tree Inference). This method is based on a statistical framework that models each host as a distinct population, and transmissions between hosts as migration events. Our computationally efficient implementation of this model enables the inference of host-to-host transmission while accommodating within-host evolution and non-sampled hosts. SCOTTI is distributed as an open source package for the phylogenetic software BEAST2. We show that SCOTTI can generally infer transmission events even in the presence of considerable within-host variation, can account for the uncertainty associated with the possible presence of non-sampled hosts, and can efficiently use data from multiple samples of the same host, but can be inaccurate when samples are collected close to the infection time. We illustrate the features of our approach by investigating transmission from genetic and epidemiological data in a Foot and Mouth Disease Virus (FMDV)

veterinary outbreak in England and a Klebsiella pneumoniae outbreak in a Nepali neonatal unit. Transmission histories inferred with SCOTTI will be important in devising effective measures to prevent and halt transmission.

> **Daniel Doerr**[1]; Bernard Moret[1]
[1] School of Computer and Communication Sciences, EPFL, 1015 Lausanne, Switzerland

Principled synteny analysis

Many methods in comparative genomics require prior knowledge of syntenic blocks, making synteny deduction a fundamental task in genomic analyses. Yet, a recent study suggests that different synteny tools yield very different collections of syntenic blocks on the same data, thus indirectly throwing some doubt on results in comparative genomics obtained with different synteny tools. That is because most current tools for synteny analysis are ad hoc methods specified solely by their implementation and characterized by their computational merits and limits rather than the result of a principled methodology with clear definitions and objective criteria. We present a principled approach to synteny with measurable objectives that support analyses in comparative genomics. Our new workflow facilitates the identification of syntenic blocks from raw genomic sequences of multiple species, requiring no assumptions on genes and gene families (although its pipeline can also be entered at a later stage with predefined markers and homology assignments). Our method is robust against segmental duplications, insertions and deletions of one or few markers, as well as large insertions into otherwise conserved homologous syntenic blocks. Rather than constructing a single set, our method builds a hierarchy of syntenic blocks. The hierarchy not only captures mutational changes between homologous syntenic blocks, but also allows users to influence granularity and number of syntenic blocks suitable for any dedicated subsequent comparative genomics analyses. In evaluating our method, we study a genomic dataset composed of four Drosophila species. Our results support the prevailing understanding of Drosophilian synteny, but also pinpoints fine-grained differences between the genomic sequences that to the best of our knowledge were previously unreported.

> **Luca Ferretti**[1,2]; Guillaume Achaz[2]
[1] The Pirbright Institute, Woking, UK
[2] UMR 7138, UPMC and Collège de France, Paris, France

Partial selective sweeps and genetic signatures of ongoing adaptation

Selective sweeps from standing variation and incomplete sweeps represent common modes of recent and ongoing adaptation to changing environments. Their typical genomic signature is a partial sweep, i.e. a local decrease in nucleotide variability that however does not sweep out all the diversity in the population. We provide a simple guide to the effects of partial sweeps on the decrease in nucleotide variability and on the behaviour of common neutrality tests. We use exact coalescent results on the spectrum of linked sites to derive the analytic form of the site frequency spectrum for soft, incomplete selective sweeps without recombination and we obtain the values of statistics such as Watterson's theta, Tajima's Pi, Tajima's D and Fay and Wu's H tests in terms of the initial and final frequency of the selected allele. These results provide a simple way to understand the nature of partial sweeps found in sequences from natural populations or evolution experiments under recent selective pressure. These methods can also be used to discriminate loci that experienced rapid sequence evolution from loci that are still under recent selective pressure.

> **Inês Fragata**[1,2]; Pedro Simões[2]; Claudia Bank[1]; Margarida Matos[2]
[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal
[2] Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

Predictability of long-term, but not short-term phenotypic evolution of Drosophila

Combining real-time evolution experiments with a theoretical framework is a powerful approach to create and test expectations on the repeatability and predictability of evolution. Several studies have focused on this issue1,2 and have shown that repeatability of phenotypic evolution is generally pervasive3. However, an important question that remains unanswered is how much short-term evolution can tell us about evolution in the long-term, and which one of these is easier to predict. We founded 5 new laboratory populations of Drosophila subobscura from two natural populations in Sintra and Arrábida, Portugal, in 3 different years (1998, 2001, 2005)4. We then followed the real-time evolution of life history and physiological traits in the 15 replicated populations throughout adaptation to the laboratory. Here, we compare the short and long-term (20 and 60 generations, respectively) evolutionary dynamics of these populations. We have previously shown4 that evolutionary contingencies during short-term adaptation were most prevalent for traits more loosely related with fitness, consistent with less selective pressure on these. Curiously, when performing the long-term analysis, we found differences in evolutionary rate for several traits, independently of their relationship to fitness. When analyzing the dependence of the overall evolutionary rate on the initial differentiation

to the control populations, we found that differentiation is a good predictor for both short and long-term evolution. However, differences between populations seem to be more common in the initial period of adaptation. We also found that the short-term evolutionary rate (combining all traits) is a good predictor of long-term patterns in most cases. These results suggest that stochastic events play a role on both short and long-term evolution of each trait. The presence of these contingencies indicates that, despite a similar evolutionary trend between populations, the repeatability of evolution is not pervasive. Combining all analyzed traits, the strong dependency of short and long-term evolution on early differentiation and the later convergence supports the presence of a single phenotypic optimum, common to all populations. This suggests that adaptation to this environment is somewhat predictable. However, the specific evolutionary dynamics seem to be different between populations, consistent with a role of stochastic events in the path taken towards this optimum. In light of these findings, we discuss a theoretical approach to combine information across traits in order to achieve a better characterization of the evolutionary dynamics, and address the issue of how predictable phenotypic evolution is.

References :
1 – de Visser JAGM, Krug J. 2014. Nat. Rev. Genet. 15:480–490
2 – Lind PA, Farr AD, Rainey PB. 2015. eLife Sci. 4:e07074
3 – Orgogozo V. (2015) Interface Focus 5:20150057.
4 – Simões P, Santos J, Fragata I, Mueller LD, Rose MR, Matos M. 2008. Evolution 62:1817–1829.

> **Andreas Futschik**[1]; Kerstin Gärtner[2]
[1]Johannes Kepler University Linz,
[2] University of Veterinary Medicine Vienna, Austria

On the inadmissibility of common population genetic estimates and their improvement by using shrinkage

Two key parameters in population genetics are the scaled mutation parameter theta and the scaled recombination rate rho. Estimates of these parameters have been obtained either from suitable summary statistics, or via more sophisticated methods such as maximum (composite) likelihood. We show that there is often room for uniform improvement of these estimates. We first explain why the popular Ewens-Watterson estimate of theta is inadmissible in terms of the mean squared error under the classical Wright-Fisher model and how it can be uniformly improved. Also other estimates of theta such as the MLE can be improved, although to a smaller extent. Then we look at estimating the scaled recombination rate rho, and explore possible gains that can be obtained with the two popular (approximate) Bayesian estimates LDhat and LDhelmet. It turns out that there is often still room for uniform improvement. This may be surprising, as it is well known that Bayesian estimates are admissible under weak assumptions. Both for Wattersons estimate and when estimating recombination, the improvement can be achieved without a lot of computational effort. However, as no explicit formulas for bias and variance are available when estimating rho, the coefficients turning up in the improvement need to be estimated from simulations.

> Céline Scornavacca[1]; **Nicolas Galtier**[1]
[1] Institut des Sciences de l'Evolution, Montpellier, France

Incomplete lineage sorting in mammalian phylogenomics

The impact of incomplete lineage sorting (ILS) on among genes phylogenetic conflicts and the need to account for ILS in species tree reconstruction are a matter of intense controversy (e.g. Gatesy & Springer 2014, Liu et al 2015, Springer & Gatesy 2016, Edwards et al. 2016). Here, focusing on full-genome data in placental mammals, we empirically test two assumptions underlying current usage of tree-building methods that account for ILS. We show that in mammals (i) distinct exons from a common gene do not share a common genealogy, (ii) ILS is only a minor determinant of the existing phylogenetic conflict between genes, and (iii) classical supertree methods are faster and perform equally well as methods recently designed to correct for ILS. These results shed new light on the relevance and conditions of applicability of ILS-aware methods in phylogenomic analyses of protein coding sequences.

References :
Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR, Leaché AD, Liu L, Davis CC. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol Phylogenet Evol. 2016 Jan;94(Pt A):447-62.
Gatesy J, Springer MS. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80:231-266.
Liu L, Xi Z, Wu S, Davis CC, Edwards SV. 2015. Estimating phylogenetic trees from genome-scale data. Ann. N. Y. Acad. Sci. 1360:36-53.
Springer MS, Gatesy J. 2016. The gene tree delusion. Mol Phylogenet Evol 94:1-33.

> **Adrián González Casanova**[1]; Noemi Kurt[2]; Anton Wakolbinger[3]; Linglong Yuan[4]
[1] WIAS Berlin
[2] TU Berlin
[3] Uni Frankfurt
[4] Uni Mainz

Modeling the Lenski experiment

The Lenski experiment investigates the long-term evolution of bacterial populations. Its design allows the direct comparison of the reproductive fitness of an evolved strain with its founder ancestor. It was observed by Wiser et al. (2013) that the mean fitness over time increases sublinearly, a behaviour which is commonly attributed to effects like clonal interference or epistasis. In this talk we present an individual-based probabilistic model that captures essential features of the design of the Lenski experiment. We assume that each beneficial mutation increases the individual reproduction rate by a fixed amount, which corresponds to the absence of epistasis in the continuous-time (intraday) part of the model, but leads to an epistatic effect in the discrete-time (interday) part of the model. Using an approximation by near-critical Galton-Watson processes, we prove that under some assumptions on the model parameters which exclude clonal interference, the relative fitness process converges, after suitable rescaling, in the large population limit to a power law function.

> **Stéphane Guindon**[1,2]; Hongbin Guo[1]; David Welch[1]
[1] The University of Auckland, New Zealand
[2] LIRMM, CNRS Montpellier, France

Demographic inference under the coalescent in a spatial continuum

Characterizing population dynamics from the analysis of molecular and spatial data relies on sound statistical modelling of the underlying processes. Current approaches rely on the assumption that a population is naturally partitioned into discrete demes, which is not always biologically realistic. Other models predict the formation of clusters of individuals in space, which, again, is not always relevant. Building on recent theoretical work, we introduce a new genealogy-based inference framework that alleviates these important issues. Simulations indicate that estimates of local population density obtained with this approach are more accurate and precise than those derived from standard techniques. Useful information about rates of dispersal can also be recovered, which is not possible with the traditional approach based on the analysis of pairs of sequences. Describing populations and/or closely related species distributed on a spatial continuum using this new approach therefore opens new perspectives in population genetics and ecology. The validity of this inference framework is further confirmed with the analysis of influenza sequences collected over five seasons in the USA.

> **Alice Ledda**[1]; Xavier Didelot[1]
[1] Department for Infectious Disease Epidemiology, Imperial College, London

A formal model of clonal expansion in bacterial population genetics

With the recent advances in sequences techniques, the availability of bacterial sequences in the last few years has dramatically increased. Given the relative small cost of sequencing, it is now possible to sequence many very closely related bacteria and thus to study their evolution at short time scales in great detail. One phenomenon that is particularly evident in the short timescale of bacterial evolution is the so called "Clonal expansion". Clonal expansion occurs when a certain lineage in a bacterial population has a selective advantage and thus expands, creating a clonal subpopulation. When talking about clonal expansion, the interest is not in the time-scales at which this clone gets to fixation (if it ever does), but only in the transient, that is the small time-scale at which the expansion begins. The concept of clonal expansion is quite straightforward and widely used in literature, yet has not been formally conceptualised. In this talk we will present a formal model of clonal expansion. We will show how this model can be used to simulate pattern resembling those observed in natural bacterial populations. We will also present a method to perform Bayesian inference under the clonal expansion model, thus allowing to estimate key parameters of our model such as effective populations size, time scales and expansion rates. In an ongoing epidemics, these quantities can be used to foresee the strength of the expansion and to tune the strength of public health responses to be put in place to limit the epidemics.

> **Benjamin Linard**[1,2]; Paula Arribas[1,3]; Carmelo Andujar[1,3]; Alfried Vogler[1,4]
[1] Natural History Museum of London, London, UK
[2] LIRMM, Montpellier, France
[3] University of La Laguna, Tenerife, Spain
[4] Imperial College, London, UK

Metagenome skimming of species-rich lineages: an alternative to eDNA barcoding

Shotgun sequencing is finally coming to the exploration of animal biodiversity, overcoming past

paradigms of biology. Complex eukaryotic communities can now be explored through shallow sequencing of bulk samples. In particular, the “Metagenome Skimming” of plants [1] or arthropod genomes [2] showed its potential for rapidly extracting organelle genomes [3] or the most abundant nuclear genome motifs [4], appearing today as a flexible alternative to classic eDNA barcoding [5]. Exploiting only the mitochondrial fraction already showed great potential for rapidly building large-scale phylogenies [6], resolving patterns of biodiversity [7] and exploring biological frontiers such as soil arthropod biodiversity [8]. These approaches come with their own challenges, mainly related to DNA reads prioritization and de-novo assembly. I will describe cases related to the metagenome skimming of insect pools [2,12]. These complex species assemblages can be targeted for their organelles genomes (identification, phylogenetics), recurrent gene families (genetic variation), and associated gut remnants/bacterial symbionts (ecology). New pipelines are needed to rapidly target these components and produce reliable contigs, but the presence of many undescribed clades makes subsequent phylogenetics unattainable for many laboratories. The accumulation of such data also requires the development of dedicated DNA reads mining and new visualization tools.

1. Straub SCK, et al.: Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot 2012, 99:349–364.
2. Linard B, et al.: Metagenome skimming of insect specimen pools: potential for comparative genomics. Genome Biol Evol 2015.
3. Malé PJG, et al.: Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol Ecol Resour 2014, 14:966–975.
4. Dodsworth S: Genome skimming for next-generation biodiversity analysis. Trends Plant Sci 2015.
5. Coissac E, et al.: From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol 2016.
6. Crampton-Platt A, et al.: Soup to Tree: The Phylogeny of Beetles Inferred by Mitochondrial Metagenomics of a Bornean Rainforest Sample. Mol Biol Evol 2015:msv111–.
7. Tang M, et al.: High-throughput monitoring of wild bee diversity and abundance via mitogenomics. Methods Ecol Evol 2015:n/a–n/a.
8. Andújar C, et al.: Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. Mol Ecol 2015:n/a–n/a.
12. Linard B, et al.: Lessons from metagenome skimming of arthropod-preserving ethanol. (in press)

> **Sébastien Lion**
Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) -- CNRS UMR 5175, Montpellier, France

Parasite evolution in heterogeneous and spatially structured host populations

Evolutionary epidemiology aims at understanding how host-parasite interactions evolve in response to various ecological factors. However, theoretical studies often assume that the host population is well-mixed, thereby neglecting potential selective pressures caused by genetic and epidemiological spatial structuring. I will present some theoretical and experimental results in order to elucidate the evolutionary impacts of parasite and host dispersal patterns. I will first focus on a population where all hosts have the same quality, and show that the predictions of non-spatial theory are altered by kin competition for susceptible hosts. I will then examine what happens in a heterogeneous population in which a fraction of the hosts are vaccinated. I show that different types of vaccines may lead to different evolutionary outcomes, which depend on the interplay between vaccine efficacy, vaccination coverage, and spatial structure. Kin selection is shown to be a useful conceptual tool to understand the ecological feedbacks on parasite traits and to generate predictions for the management of infectious diseases.

> **Odile Maliet**[1,2]; Fanny Gascuel[1,2,3]; Amaury Lambert[2,4]
[1] Institut de Biologie de l’Ecole Normale Superieure, CNRS UMR 8197, Ecole Normale Superieure, Paris, France
[2] Center for Interdisciplinary Research in Biology, CNRS UMR 7241, College de France, Paris, France
[3] Sorbonne Universites, UPMC Univ Paris 06, CNRS UMR 7625, Laboratoire Ecologie et Evolution, Paris, France
[4] Sorbonne Universites, UPMC Univ Paris 06, CNRS UMR 7599, Laboratoire Probabilites et Modeles Aleatoires, Paris, France

The effects of phylogenetic tree shape and non random extinctions on the expected loss of phylogenetic diversity

Phylogenetic diversity is a measure of the evolutionary legacy of a group of species, defined as the sum of branch lengths of the phylogeny spanned by these species. It can be used to define conservation priorities. It has been shown that an important loss of species diversity can lead to a much less important loss of phylogenetic diversity (Nee & May 1997). This result depends strongly on the topology of the tree, and particularly on its balance and the age of its nodes (Nee & May 1997), but the effect of the correlation between clade age and its species richness has not been studied so far. It should also be sensitive to the order in which species are lost, depending on how species abundances covary with the phylogeny. I present a three-parameter model generating trees with random topology and random node ranking, also endowed with random numbers summing to 1 at the tips (interpreted as abundances of contemporary species). This model can be seen as an extension to Aldous’ beta-splitting model (Aldous 1996; Aldous 2001) with two additional parameters, a new parameter alpha

quantifying the correlation between ages of subtrees and their richness and another parameter eta quantifying the correlation between richness of subtrees and their frequency in the community. We develop an MCMC algorithm to infer the first two parameters from real data and show that the combined effect of the parameters can make phylogenetic diversity loss much more sensitive to species loss.

Aldous, D.J., 1996. Probability Distributions on Cladograms. In D. J. Aldous & R. Pemantle, eds. Random Discrete Structures. Springers (IMA Volumes Math. Appl. 76), pp. 1–18.
Aldous, D.J., 2001. Stochastic Models and Descriptive Statistics for Phylogenetic Trees , from Yule to Today. , 16(1), pp.23–34.
Nee, S. & May, R.M., 1997. Extinction and the loss of evolutionary history. Science (New York, N.Y.), 278(5338), pp.692–694.

> **Guillaume Martin**
Institut des Sciences de l'Evolution (ISEM), Montpellier, France

Predicting adaptive fitness trajectories in experimental evolution from deleterious mutation data

Various models describe asexual evolution undergoing mutation, selection and drift. Some focus directly on fitness, explicitly modelling drift but simplifying mutation effects distributions and epistasis (Mutation load, Muller's ratchet, clonal interference models). Others follow the dynamics of quantitative traits determining fitness (quantitative genetics, Fisher's geometrical model), imposing a complex but fixed form of mutation effects and epistasis, and often ignoring drift. In all cases, predictions are typically obtained in high or low mutation rate limits and for long-term stationary regimes, thus loosing information on transient behaviors and the effect of initial conditions. Here, we connect fitness-based and trait-based models in a single framework, and seek explicit solutions even away from stationarity. The expected fitness distribution is followed over time via its cumulant generating function, using a deterministic approximation that ignores the bias induced by drift. In several cases, the resulting partial differential equation can be solved, yielding predictable trajectories of the full fitness distribution, for arbitrary mutation rates and standing variance. For non-epistatic models, this deterministic approximation fails over the long term, but captures the nonlinear transient dynamics, thus complementing stochastic fitness-based models. For diminishing returns epistasis models (e.g. with a phenotypic optimum), the approximation is accurate and analytically tractable at and away from equilibrium. Such epistatic models display a sort of 'phase transition', where the equilibrium fitness distribution changes qualitatively beyond some threshold mutation rate. Both fitness and trait distribution trajectories over single peak landscapes take simple analytic form beyond this threshold, independently of the underlying mutant phenotype distribution. Finally, an application to trajectory predictions from fisher's geometrical model, compared to observed trajectories from various evolution experiments will be discussed.

> **Chieh-Hsi Wu**[1]; Nicola de Maio[1]; Daniel Wilson[1]
[1] University of Oxford, Oxford, United Kingdom

Co-estimation of migration patterns and demographic history in structured populations using a structured coalescent approach

Inferring the historic trends in the effective population size is of substantial interest in various fields of biology. For example, information on the demographic history provides insights into the epidemiological dynamics of pathogen populations and enables hypothesis testing of the impact of the change in environment (e.g. climate) on past ecosystems. Trends in the past could potentially shed light on the future dynamics of the populations of interest. The skyline-plot methods (Pybus et al., 2000; Drummond et al., 2005; Minin et al., 2008) aim to recover the temporal variation in the effective population size from molecular sequence data. However, these methods assume that the samples are all collected from a single panmictic population. Many studies have shown that if a population is structured, ignoring the population structure can result in misleading inference of the population trends (Taedler et al., 2009; Heller et al., 2013). Therefore, in order to avoid model misspecification, estimating demographic history of structured populations requires a method that accounts for the effect of the population trend and the migration process on the shape of the genealogy. On the other hand, the structured coalescent extends the Kingman coalescent to geographically structured populations, while modelling the migration process among the subpopulations and incorporating its effect on the tree shape. Until recently, performing Bayesian inference under the structured coalescent model was hampered by its computational burden. Recently, more efficient structured coalescent-based methods have been developed (Vaughan et al., 2014; de Maio et al., 2015) which can handle datasets with much larger samples sizes and number of demes than previous methods. However, these methods do not accommodate changes in population size through time. Here, we present a new method extending a recently proposed structured coalescent approximation (de Maio et al., 2015) to allow efficient joint inference of the demographic history and the migration process within structured populations. Our method employs a Bayesian nonparametric smoothing approach in the reconstruction of the demographic history of the subpopulations. In a simulation study, we demonstrate how the inferred population trend is affected when population structure is ignored under various sampling and demographic scenarios, including complex situations where the population trend varies

among subpopulations. Furthermore, we apply our method to a number of empirical virus datasets and compare the reconstructed demographic history with that recovered from methods that ignore population structure.

> **Chris Wymant**[1]; Marion Cornelissen[2]; Astrid Gall[3]; Matthew Hall[1]; François Blanquart[1]; Mariska Hillebregt[4]; Margreet Bakker[2] ; Jan Albert[5]; Daniela Bezemer[4]; Anne-Marte Bakken Kran[6]; Anne Margarita Dyrhol Riise[6]; Katrien Fransen[7]; M. Kate Grabowski[8]; Barbara Gunsenheimer-Bartmeyer[9]; Huldrych Günthard[10]; Roger Kouyos[10]; Claudia Kücherer[9]; Oliver Laeyendecker[11]; Kirsi Liitsola[12]; Laurence Meyer[13]; Kholoud Porter[14]; Matti Ristola[12]; Anders Sönnerborg[5]; David van de Vijver[15]; Guido Vanham[7]; Ard van Sighem[4]; Swee Hoe Ong[3]; Martin Hunt[3]; Dan Frampton[14]; Sima Zaheri[4]; Frank de Wolf[1]; Ben Berkhout[2]; Peter Reiss[2]; Paul Kellam[3]; Christophe Fraser[1]
[1] Imperial College London, London, UK
[2] Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands
[3] Wellcome Trust Sanger Institute, Hinxton, UK
[4] HIV Monitoring Foundation, Amsterdam, The Netherlands
[5] Karolinska University Hospital, Stockholm, Sweden
[6] Oslo University Hospital, Oslo, Norway
[7] Institute of Tropical Medicine, Antwerp, Belgium
[8] John Hopkins University, Baltimore, USA
[9] Robert Koch-Institute, Berlin, Germany
[10] University Hospital Zurich, Zurich, Switzerland
[11] Laboratory of Immunoregulation NIAID NIH, Baltimore, USA
[12] Helsinki University Hospital, Helsinki, Finland
[13] Université Paris Sud, Paris, France
[14] University College London, London, UK
[15] Erasmus Medical Centre, Rotterdam, The Netherlands

Phylogenetics Between and Within: Seeing Transmissions and Dual Infections in HIV Deep Sequence Data

The large-scale sequencing of pathogens facilitated by next-generation sequencing (NGS) methods provides rich data that informs us about evolutionary history and epidemiology of a disease, and the genetic determinants of virulence. However interpreting NGS data for pathogens that diversify within-host, such as HIV, has proven challenging due to the difficulty in assembling short reads from a quasispecies. The usual approach involves collecting together all reads from across the genome for each patient. Here, we collect together the reads at fixed positions in the genome from all patients in a sample, for a full picture of viral evolution within and between hosts simultaneously. Analysis of such phylogenies, when constructed in sliding windows spanning the whole genome, reveals 1) dually-infected patients, namely those with two phylogenetically distinct groups of reads; 2) contamination, as recurrent exact duplication of reads; 3) transmission, as sub-trees nested or entangled between patients; and 4) points of recombination, directly visible in the aligned reads of the covering window. I will present a powerful and user-friendly tool for generating such phylogenies, and show the results of its application to 68 European seroconverters from the BEEHIVE study into the viral-molecular basis of HIV virulence. Identifying cases of dual infection is critical in large-scale analysis of sequence data, as such cases are not only special clinically but require a different bioinformatic approach to avoid artefactually recombinant results. Identifying transmission allows associated epidemiological (risk) factors to be found, informing targetted public health interventions; it also quantifies the heritability of infection properties. Both points are important for attempts to forecast pathogen evolution. Our tool and analyses of the results should be relevant to sequence data for other pathogens showing appreciable within-host diversity.

POSTERS

Poster 1

> **Yoann Anciaux**[1]; Martin Guillaume[1]; Lionel Roques[2]; Amaury Lambert[3]; Ophélie Ronce[1];
[1] Institut des Sciences de l'Evolution (ISEM), Montpellier, France
[2] Biostatistiques and Spatial Process (BioSP), INRA, Avignon, France
[3] Center for Interdisciplinary Research in Biology (CIRB), Collège de France, Paris, France

Evolutionary Rescue In large asexual Populations

Evolutionary rescue occurs when a population genetically adapts to a new stressful environment that would otherwise cause its extinction. Forecasting the probability of persistence under stress, such as emergence of drug resistance or invasion of a new environment is a major concern in ecology and evolution. Models which are describing the eco-evolutionary dynamics of evolutionary rescue consider the stochastic apparition of new alleles and their fixation (origin-fixation models) or the creation of genetic variance in a polymorphic population by recombination or mutation. However none of them allow us to forecast the dynamic of the whole distribution of fitness of the population and mutation are often independent of the genetic background where they appear. Here, we explore the use of an analytical tool to model the dynamics of the distribution of fitness in large asexual populations, through their generating function. This allows following the effect of selection among many co-segregating types and of background-dependent mutation effects (epistasis) under Fisher's geometric model. Analytical results based on diffusion approximations for the probability of evolutionary rescue are confronted with individual based simulation in different scenario such as adaptation from standing variance and of de novo mutation.

Poster 2

> **Florian Clemente**[1]; Mathieu Gautier[2]; Renaud Vitalis[2]
[1] Institut de Biologie Computationnelle (IBC), Montpellier, France
[2] CBGP, Montpellier, France

KimTree: dealing with ascertainment bias and selection

The revolution in sequence data generation for model and non-model species has led to various approaches trying to make this ever-increasing amount of information accessible. Whereas early methods were forced to rely on mathematically convenient approximations to the underlying evolutionary processes, advances in computing power now enable complex likelihood-based inference. Here, we present an extension of KimTree, a previously developed method for estimating divergence times among populations based on the Kimura diffusion approximation for the evolution of neutral alleles. We explore different ways to correct for ascertainment bias created due to analysing SNP data and apply the model to simulated data of autosomes and sex chromosomes to jointly infer divergence times, which are informative about the effective sex-ratio in the studied populations. Moreover, we extend the method to identify loci under positive selection, using information about the expected allele frequency distribution after a selective sweep. The performance of the model is evaluated under various demographic scenarios. We find considerable improvement in the accuracy and robustness of parameter estimation compared to the original version of KimTree.

Poster 3

> **Ramsès Djidjou-Demasse**[1,2]; Frédéric Fabre[1]; Arnaud Ducrot[2]
[1] Institut National de la Recherche Agronomique (INRA), Bordeaux, France
[2] Institut Mathématique de Bordeaux (IMB), Bordeaux, France

Eco-evolutionary dynamics of life-history traits of fungal plant pathogens: an analysis of evolutionarily stable phenotype(s).

Modeling the epidemiology and the evolution of pathogens has long been addressed through analysis of invasibility assuming that epidemiological and evolutionary time scales are distinct. These analysis ignore short-term evolutionary and epidemiological dynamics despite the major interests of these dynamics in agroecosytems. Deriving sustainable management strategies of resistance gene to plant pathogens is one of the many examples where one want to make quantitative predictions far away from the epidemiological equilibrium. Up to date, most of the modelling approaches devoted to design sustainable strategies of resistance management tackle the case of qualitative resistance based on gene for gene interactions that triggers host immunity to the disease. Few works consider adaptation to quantitative resistances that reduce pathogen aggressiveness instead of conferring immunity. Pathogen aggressiveness is determined experimentally by plant pathologists by measuring life history traits which described the disease life cycle: the infection efficiency, the latency time, the sporulation

period and the rate of spore production. In this work, we used integro-differential equations to model the evolutionary epidemiology of spore-producing pathogens in a homogeneous host population. The host population is subdivided into four compartments (Susceptible or healthy host tissue, Latently infected tissue, Infectious tissue and Airborne spores). The model also consider the evolutionary dynamics within the pathogen population. To do so, we consider genetically diverse pathogen population characterize by a continuous distribution of phenotypic trait values. Each strain is represented by phenotypic trait value x. Mutations, causing a transition from one strain to another, are modelled by an integral operator. In this framework, the life history traits describing the disease life cycle are function of phenotypic trait value x. We then studied analytically the stationary values of the evolutionary process, the so-called Evolutionarily Stable Strategies (ESS). By defining a specific function depending on the pathogens life history traits, we first characterized the set (called EES-pathogen set) of pathogens strains which can defined an ESS. In the literature, there is some well-known results when the EES-pathogen set is reduced to a single pathogen strain. But, little is done when the EES-pathogen set is composed by at least two pathogens strains. Here we give some results for this latter case. We also investigated:(i) the density of infected host with respect to phenotypic trait values x, (ii)the transient evolutionary dynamics of pathogens strains before they reach an equilibrium and (iii)whether strains can co-exit and why. With the view of tackling the eco-evolutionary dynamics of spore producing pathogen in agro-ecosystems, we are now working to extend these results to the case of heterogeneous host metapopulations bearing distinct quantitative resistance genes

Poster 4

> **Jesús Fernández-Sánchez**[1]; Marta Casanellas[1]
[1] Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Reconstruction of phylogenetic trees with Erik+2

I present a reconstruction method for quartets, called Erik+2 (Fernández-Sánchez and Casanellas, 2016), consistent with the most general Markov model of nucleotide substitution, which can also deal with data coming from mixtures on the same topology. We compare Erik+2 to popular methods as neighbor-joining (with paralinear distance), maximum likelihood (with different underlying models), and maximum parsimony. The results show that this method is accurate and robust, has a similar performance to ML when data satisfies the assumptions of both methods, and outperforms the other methods when these are based on inappropriate substitution models. If alignments are long enough, then it also outperforms other methods when some of its assumptions are violated. Erik+2 also provides a system of weights that can be used as input for quartet-based methods. We analyze the performance of Erik+2 with some quartet-based methods in phylogenetic reconstruction, namely Quartet Puzzling (Strimmer and von Haeseler, 1996), Weight Optimization (Ranwez and Gascuel, 2001) and Willson's method (Willson, 1999). We believe the results obtained somehow revalidate quartet methods to reconstruct phylogenetic trees. References :

J. Fernández-Sánchez and M. Casanellas, Invariant versus classical quartet inference when evolution is heterogeneous across sites and lineages, Syst Biol (2016) 65(2): 280-291
V. Ranwez and O. Gascuel. Phylogenetic reconstruction algorithms based on weighted 4-trees. In Computational biology, volume 2066 of Lecture Notes in Comput. Sci., pages 84-98. Springer, Berlin,2001.
K Strimmer and A von Haeseler. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Molecular Biology and Evolution, 13(7):964, 1996.
SJ Willson. Building phylogenetic trees from quartets by using local inconsistency measures. Molecular Biology and Evolution, 16(5):685, 1999.

Poster 5

> **Alexandre Gouy**[1,2]; Laurent Excoffier[1,2]
[1] Institute of Ecology and Evolution (IEE), Bern, Switzerland
[2] Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

Detecting and measuring polygenic selection in gene networks

To better predict how species will adapt to changing environments, it is of central importance to understand the genetic basis of adaptation. This process is typically viewed as involving selective sweeps that drive beneficial alleles from low to high frequency in a population. By contrast, classical models of artificial and natural selection in the quantitative genetics literature emphasize the importance of modest changes in allele frequencies at many loci [1]. Indeed, adaptive events in natural populations could occur by the evolution of polygenic traits rather than via the fixation of single beneficial mutations [2]. Recent studies in various model organisms have confirmed that variation at many important traits is controlled by a large number of loci dispersed throughout the genome (e.g. [3]). Numerous statistical tests have been developed to detect selection from genomic data based on a simple selective sweep model [1]. Small allele frequency changes due to polygenic selection may however remain below the detection limit of most of these outlier detection methods [4]. Moreover, the conclusions drawn from significant tests can be seriously challenged because phenotypic traits exhibiting clear-cut molecular signatures of selection may represent a biased subset of all adaptive

traits [5]. In order to better understand to what extent species adapt by polygenic selection, we have developed a method to detect and measure this type of selection in natural populations. The general idea is to search for subsets of interacting genes within gene networks (e.g. biological pathways) that present unusual features. This search is a typical combinatorial optimization problem that can be solved using a heuristic approach like simulated annealing. We implemented such an algorithm to search for high-scoring subnetworks of genes in biological pathways data. We also developed a significance test procedure that explicitly takes into account the optimization process involved in this search. We applied our methodology to find evidence of polygenic adaptation in various human populations.

[1] Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Molecular ecology, 25(1), 79-88.
[2] Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation–not by sweeps alone. Nature Reviews Genetics, 11(10), 665-667.
[3] Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... & Ferreira, T. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature, 467(7317), 832-838.
[4] Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. Molecular ecology, 21(7), 1548-1566.
[5] Chevin, L. M., & Hospital, F. (2008). Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics, 180(3), 1645-1660.

Poster 6

> **Valentin Hivert**[1]; Mathieu Gautier[1]; Renaud Vitalis[1]
[1] Centre de Biologie pour la Gestion des Populations (CBGP), Montpelliers, France

A hierarchical bayesian model for measuring the extent of local adaptation from haplotype data.

The recent advent of high throughput sequencing and genotyping technologies (Next Generation Sequencing, NGS) enables the comparison of patterns of polymorphisms at a very large number of markers, which makes it possible to characterize genomic regions involved in the adaptation of organisms to their environment. Here, we present some recent developments to SelEstim (Vitalis & al., 2014), a hierarchical bayesian model that identifies and measures genomic signatures of selection from gene frequency data. In particular, we extend the model to analyse multi-allelic markers. This allows to use haplotype data, defined by means of unsupervised classification methods, (Scheet & Stephens 2006), and considering haplotype blocks as multi-allelic markers. We expect that such approach makes more use of linkage disequilibrium information across individual markers , as compared to SNP data . We will show some analyses conducted on simulated data, comparing the information brought by haplotype data relatively to analyses using SNP data alone. Finally we apply SelEstim on human data, showing evidence of selection acting on the lactase gene. We will also discuss some potential extensions of this method, that would allow to test for correlations between haplotype frequencies and environmental variables, which could ultimately be used to predict the potential of evolution for some particular populations (e.g., in the context of invasion dynamics).

References :
Vitalis, R., M. Gautier, K.J. Dawson, and M.A. Beaumont, 2014 Detecting and Measuring Selection from Gene Frequency Data. Genetics 196 : 799-817.
Scheet, P., and M. Stephens 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American journal of human genetics 78 : 629-644

Poster 7

> **K K Irwin**[1,2]; S Laurent[1,2]; S Matuszewski[1,2]; S Vuilleumier[1,2]; L Ormond[1,2]; H Shim[1,2]; J D Jensen[1,2]
[1] School of Life Sciences, EPFL, Lausanne, Switzerland
[2] Swiss Institute of Bioinformatics (SIB), Switzerland

Improving the inference of viral population histories by accounting for multiple merger events and background selection

Many features of viral populations make them excellent candidates for population genetic study. Examples include a very high rate of mutation (especially in RNA viruses), high levels of nucleotide diversity (such that some viruses exist as quasispecies), exceptional effective population sizes (Ne up to 105), and frequent positive selection (in response to host immunity or antiviral drugs). Though such attributes make them ideal subjects for models requiring extreme values (e.g., high Ne required for selection on standing variation, Pennings 2012), everyday population genetic models must be used with caution. Confidently inferring demographic or selective histories of viral populations in coalescent models can prove challenging, and special attention must be paid to model assumptions. For example, the skewed offspring distributions and frequent bottleneck events in viruses render them susceptible to multiple-merger coalescent events (Tellier & Lemaire 2014), and extraordinary rates of mutation lead to frequent removal of deleterious variants and hence frequent background selection (Renzette et al. 2015), a phenomenon not often accounted for in population models. Given these distinctive features and the increasing popularity of population genetic inference for elucidating viral evolutionary patterns, it is crucial to account for processes that, though they may be less common in other organisms, occur

regularly in viruses. In particular, we draw attention to multiple-merger coalescent events and background selection, two processes that may greatly affect viral evolution and hence its inference. We review each one, its particular application to viruses, and the repercussions of its omission from population genetic models.

Pennings, P. 2012. Standing genetic variation and the evolution of drug resistance in HIV. PLoS Computational Biology 8:31002527

Renzette, N., Kowalik, T. F., and J. D. Jensen. 2016. On the relative roles of background selection and genetic hitchhiking in shaping human cytomegalovirus genetic diversity. Molecular Ecology 25: 403-413

Tellier, A. and C. Lemaire. 2014. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Molecular Ecology 23: 2637-2652

Poster 8

> **Sohta Ishikawa**[1]; Yuji Inagaki[2]; Tetsuo Hashimoto[2]
[1] University of Tokyo, Tokyo, Japan
[2] University of Tsukuba, Tsukuba, Japan

A multi-grained MPI/OpenMP parallelization of the maximum-likelihood phylogenetic inference with the non-homogeneous model

Recent advances in DNA sequencing techniques enable us to phylogenetically analyze large matrices including sequences from diverse species, (and genes). At present, the vast majority of maximum-likelihood (ML) phylogenetic programs only implement 'homogeneous' substitution models enforcing an uniform evolutionary process to across the tree. Because the nucleotide and amino acid sequences in distantly related species certainly evolve under different evolutionary processes (non-homogeneous evolution), the assumption of homogeneous models is often violated, and such model violation may result in various forms of phylogenetic artifacts. For accurate phylogenetic inferences from real-world sequence data, non-homogeneous (NH) substitution models, which allow model parameters to vary across the tree, are more realistic than homogeneous models. However, the analyses with NH models, in which an enormous amount of model parameters need to be optimized, can be computationally intense. Therefore, efficient parallelization for the phylogenetic programs is critical to analyze real-world sequence datasets based on NH models within a reasonable computational time. We here present our computational challenges to parallelize a ML method-based phylogenetic program, 'NHML', which implements a NH model that allows the AT content to vary across lineages (Galtier N. & Gouy M. 1998, Mol. Biol. Evol., 15:871-879). We applied three parallel computing methods to multiple levels of the algorithm for the phylogenetic inferences with NHML; (i) a fine-grained parallelization by OpenMP was applied to the site-wise log-likelihood calculation for a given tree, and (ii) medium- and iii) coarse-grained parallelization by MPI were applied respectively to the optimization of model parameters for a given tree, and to the computation of multiple tree topologies (e.g., those proposed by SPR method during the ML tree search). The performance of the 'multi-grained' parallelization was benchmarked by analyzing simulated nucleotide sequence datasets, on up to 32 computational nodes and 512 CPU cores of a current supercomputer system, COMA (PACS-IX), in Center for Computational Sciences, University of Tsukuba. We analyzed several datasets with different number of taxa and positions to assess the power and limit of each parallel region of the present code against, i) different number of trees to be simultaneously computed during ML tree search, ii) different number of parameters to be optimized for a single tree, and iii) different number of site-wise likelihood calculations for a given tree. Consequently, we archived ~180 times speedup from the serial code on the use of 512 cores with the datasets composed of 17 taxa and 10,000 bp. From the results, we here propose a guideline to efficiently distribute parallel computing resources (OpenMP threads and MPI processes) on multiple computational nodes, to obtain the best performance of the present multi-grained parallel scheme with NH models.

Poster 9

> N. Vakirlis[1]; A.Herbert[2]; D.A.Opulente[3]; C.T.Hittinger[3]; G.Fischer[1]; J.J. Coon[2]; **I. Lafontaine**[1]
[1] Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Institut de Biologie Paris-Seine, UPMC, Paris
[2] Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
[3] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706, USA

Evolution of gene repertoire in yeast genomes : the emergence of novel genes

We recently reconstructed the genome history of the entire Lachancea yeast genus covering a continuous evolutionary range from closely related to more diverged species, unifying the evolution of the genome architecture and gene repertoire [1]. To the end we generated a high quality genome dataset, developed a tool for the reconstruction of ancestral genome architectures, and performed a comprehensive analysis of gene repertoire evolution. We showed that the vast majority of extant Lachancea genes were already present in the Lachancea common ancestor. We observed that gene content variations between extant species are mainly driven by differential gene losses, while gene duplications remained globally constant in all lineages. We stated that a significant part of gene losses

were due to breakpoints generated by chromosomal rearrangements. However, a significant fraction of the genes [ca. 2%] has been gained since divergence from the Lachancea common ancestor, with several dozens of horizontally transferred genes and hundreds of potential gene creations [corresponding to annotated ORF with no detectable similarity outside the genus, also called Taxonomically Restricted Genes [TRG] or ORfans] [1]. We subsequently performed a comprehensive in silico analysis of the TRG in the Lachancea genus and in the Saccharomyces genus. We devised a novel classification method to estimate a coding probability for each TRG and propose the possible scenario for the emergence of new genes in Lachancea and Saccharomyces. Comparisons of our in silico predictions to transcriptomic and proteomic data, revealed that transcriptomic data are not always a good indicator of « geneness ».

1. Vakirlis N, Sarilar V, Drillon G, Agier N, Meyniel JP, Blanpain L, Carbone A, Devillers H, Dubois K, Fleiss A, Gillet Markowska A, Graziani S, Nguyen HV, Poirel M, Reisser C, Schott J, Schacherer J, Lafontaine I, Llorente B, Neuvéglise C, Fischer G. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Research, in press.

Poster 10

> **Jason Lapeyronnie**[1]; Claude Chevalet[1]; Bertrand Servin[1]
Institut National de la Recherche Agronomique (INRA), Toulouse Midi-Pyrénées, France

Statistical methods for the functionnal characterisation of selection signatures

Statistical methods for the functionnal characterisation of selection signatures Understanding processes that shaped population diversity in the past can help in predicting evolution. For example, inferring the evolution of genetic polymorphism in response to past adaptive constraints can help anticipating genetic response to future changes in the environment. The detection of genomic regions that responded to selection in the past, sometimes called "selection signatures", can be based on genetic data alone, or exploit additional information on individual or populations (eg. Coop et al. [1]). Here I will present the extensions of two models based on genetic data alone (FLK [1] and hapFLK [3]) to include covariates and exploit phenotypic or environmental information on populations. Based on simulations, I will show to what extent including covariates while exploiting linkage disequilibrium information can improve detection power compared to other approaches. Finally I will show how our simulations can allow to predict the efficacy of different sampling strategies to study genome response to adaptation. Keywords : Population genetic, selection, covariates, statistics

[1] Coop, Graham et al. “Using Environmental Correlations to Identify Loci Underlying Local Adaptation.” Genetics 185.4 (2010): 1411–1423.
[2] Bonhomme, Maxime et al. “Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended.” Genetics 186.1 (2010): 241–262.
[3] Fariello, María Inés et al. “Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations.”Genetics 193.3 (2013): 929–941.

Poster 11

> Damien Correia[1]; Olivia Doppelt-Azeroual[2]; **Vincent Lefort**[3]; Fabien Mareuil[2]; Sarah Cohen-Boulakia[4]; Olivier Gascuel[2][3]
[1] Centre national de la recherche scientifique (CNRS), Paris, France
[2] Institut Pasteur, Paris, France
[3] Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Montpellier, France
[4] Laboratoire de Recherche en Informatique (LRI), Orsay, France

New Generation Phylogeny.fr: Refactoring Phylogeny.fr for innovative phylogenetic services

With 50,000 data analysis per month and more than 1,500 citations (google scholar), the phylogenetic analysis pipeline Phylogeny.fr [1] is one of the most visible French IT resources both at the national and international levels. Phylogenetic analysis is performed by chaining (selected) programs together. Today, users’ needs have evolved; they can use Phylogeny.fr for teaching, inducing possibly hundreds of users at the same time, or employ it in batch mode leading to the submission of large amount of requests to the same server. Those practices have led to several engorgements of our servers. In this project, we thus plan to increase the robustness of Phylogeny.fr. The originality of the new version of Phylogeny.fr lies in considering a scientific workflow environment (Galaxy) coupled with a web interface allowing visualization and interaction with phylogenetic objects. More precisely, this project will provide (i) a large set of phylogenetic analysis bricks and for each brick, access to diverse programs, all encapsulated into Galaxy thus making the system able to deal with large groups of users and/or large sets of data, (ii) a set of optimized, robust and expressive workflows extending the basic phylogenetic workflow to various and rich contexts of phylogenetic analyses, (iii) an easy-to-install environment equipped with a new visualization layer, on top of the Galaxy system, and dedicated to phylogenetic analyses.

[1] Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Nucleic Acids Res. 2008, 36(Web Server issue):W465-9
[2] Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Goecks J, Nekrutenko A, Taylor J; Galaxy Team. Genome Biol. 2010;11(8):R86.

Poster 12

> **Florian Massip**[1]; Marie-Noelle Prioleau[2]; Laurent Duret[1]; Franck Picard [1]
[1] Laboratoire Biométrie et Biologie Evolutive, UMR CNRS 5558 Univ. Lyon 1, Villeurbanne, France
[2] Institut Jacques Monod, CNRS, University Paris 7, Équipe Labellisé Ligue Contre le Cancer, Paris, France

Characterizing Functional Sequence Motifs in Vertebrates DNA Replication Origins.

The faithful replication of DNA at the beginning of each cell division is a complex process involving the combination of genetic and epigenetic factors. For this reason, the molecular mechanisms that trigger the initiation of this process in precise genomic regions, called replication origins, are not well understood. The recent availabity of the human, mouse and chicken replication origin maps at the kb resolution grants a unique opportunity to adopt an evolutionary perspective of replication origins. Understanding the evolutionary dynamics that govern these regions, through specific conservation or mutational patterns may be a powerful strategy to better characterize the molecular functionning of replication origins. In particular, we will focus on evolutionary interplay between origins and specific motifs called G-quadruplexes, known to be involved in the replication process. Finally, we will discuss the relationships between replication origins and genomic repetitions.

Poster 13

> **Sebastian Matuszewski**[1]; Claudia Bank[2]; Nicholas Renzette[3]; Ping Lui[3]; Hyunjin Shim[1]; Mathieu Foll[4]; Daniel N.A. Bolon[3]; Konstantin B. Zeldovich[3]; Timothy Kowalik[3]; Robert W. Finberg[3]; Jennifer Wang[3]; Jeffrey D. Jensen[1]
[1] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2] Instituto Gulbenkian de Ciência, Oeiras, Portugal
[3] University of Massachusetts Medical School, Worcester, USA
[4] International Agency for Research on Cancer, Lyon, France

Favipiravir to the (evolutionary) rescue: First evidence of resistance evolution against drug-induced lethal mutagenesis in influenza A virus.

The rapid evolution of drug resistance remains a critical public health concern. The treatment of influenza A virus (IAV) has proven particularly challenging, due to the ability of the virus to develop resistance against most current antivirals (1). Here we evaluate the novel antiviral drug therapy favipiravir; this mutagenic drug's mechanism of action in IAV involves an interaction with the viral RNA polymerase, resulting in an effective increase in the viral mutation rate and, in principle, eventual population extinction (i.e., lethal mutagenesis (2, 3)). Using time-sampled data from an in vitro selection experiment where populations of IAV were evolved either in the presence or absence of favipiravir, and comparing these results to a previous, similar experiment using oseltamivir (4, 5), we indeed demonstrate that favipiravir induces lethal mutagenesis. In fact, all experimental populations exposed to escalating drug concentrations eventually became extinct. This result starkly contrasts those seen in the oseltamivir experiment: there, resistance mutations arose and fixed quickly after the introduction of drug pressure. We continued by evaluating different concentrations of favipiravir treatment, quantifying the dose necessary to induce this lethal effect, and here provide the first evidence for potential adaptation to favipiravir under a constant drug concentration. Furthermore, using population genetic modeling within an approximate Bayesian computation framework (4), we were able to identify positively selected candidate mutations underlying favipiravir resistance from whole-genome time-sampled SNP data. Our results demonstrate the promise of drug-induced lethal mutagenesis as a means for combating virus infections, and particularly for favipiravir as an effective strategy against IAV. Importantly, our findings also highlight experimental scenarios of clinical relevance, emphasizing that proper drug dosage is essential for effective treatment, and demonstrate the utility of evolutionary theory in informing drug development (6).

1. Ghedin E., E. C. Holmes, J. V. DePasse, L. T. Pinilla, A. Fitch, et al. J. Infect. Dis. 206, 1504–1511 (2012).
2. Baranovich, T., S. S. Wong, J. Armstrong, H. Marjuki, R. J. Webby, R. G. Webster et al., J. Virol. 87, 3741–3751 (2013).
3. Furuta Y., K. Takahashi, M. Kuno-Maekawa, H. Sangawa, et al. Antimicrob. Agents. Chemother. 49, 981–986 (2005).
4. Foll M., Y. P. Poh, N. Renzette, A. Ferrer-Admetlla, C. Bank, et al. PLoS Genet. 10, e1004185 (2014).
5. Renzette N., D. R. Caffrey, K. B. Zeldovich, P. Liu, G. R. Gallagher, et al. J. Virol. 88, 272–281 (2014).
6. Alexander H.K., G. Martin, O. Y. Martin, and S. Bonhoeffer, Evol. Appl. 7, 1161– 1179 (2014)

Poster 14

> **Coralie Merle**[1,2,3]; Jean-Michel Marin[1,3]; François Rousset[3,4]
[1] Institut Montpelliérain Alexander Grothendieck (IMAG), Université 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
[4] Institut des Sciences de l'Evolution CNRS UMR 5554, Montpellier, France

Detecting past contraction in population size using haplotype homozygosity

The recent development of high throughput sequencing technologies have revolutionized the data generation for many organisms. Therefore, we want to adapt and test new inference methods of the demographic history of a population from next generation genetic data. Classical methods (Importance Sampling or ABC), suitable for polymorphism data sets consisting in some loci, exploit the genetic recombination assuming the genealogies of different loci are independent. To take advantage of wider data consisting of entire genomes or genome fragments, we need to consider the dependence of genealogies of adjacent positions in the genome. Thus, we developed a statistical method appropriate to the inference of demographic parameters modeling the evolution of the population size over time. Assuming that we observe entire genomes, we chose to summarize the polymorphism information between two (haploïds) genomes or within a diploïd genome by the conserved sequence lengths in this pairwise alignment. Our inference approach of the demographic history is based on the comparison of this summary of the data to a theoretical predictor. In many cases, the biological question of interest takes the form of a model choice question. In particular, has the population undergone a size change in the past ? Has the population size undergone a past change? To achieve this, we used a model choice procedure between a simple model of constant population size and a more complex model with a past change in the population size. We highlight the results obtained under different simulated demographic scenarios and on a cattle data set.

Poster 15

> Lam-Tung Nguyen, Arndt von Haeseler, **Bui Quang Minh**
Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, Austria

Complex models of sequence evolution require accurate estimators as exemplified with the invariable site plus Gamma model

We show that frequently used phylogenetic inference programs cannot reliably estimate the shape parameter, the fraction of invariable sites and the tree length for the invariable site plus discrete Γ model. The inability to infer the true parameters is caused by inaccurate numerical optimization routines implemented in these programs. Here, we propose a simple optimization strategy to improve accuracy for maximum-likelihood methods and we recommend exercising care when implementing estimation routines for complex evolutionary models.

Poster 16

> **Veronica Miro Pina**[1,2]; Guillaume Achaz[1,3]; Emmanuel Schertzer[1,2]; Amaury Lambert[1,2]; Nicolas Lartillot[4]
[1] Centre Interdisciplinaire de Recherche en Biologie (CIRB), UMR 7241, Collège de France, Paris, France
[2] Laboratoire de Probabilités et Modèles Aléatoires (LPMA), UMR 7599, UPMC & CNRS, Paris, France
[3] Atelier de BioInformatique (ABI), UMR 7205, Institut de Systématique, Évolution, Biodiversité (ISyEB), Paris, France
[4] Laboratoire de Biométrie et Biologie Évolutive, (LBBE), UMR CNRS 5558, UCB Lyon 1, Lyon, France

Inference under a non-exchangeable model of diversification

A popular line of research in evolutionary biology is to use time-calibrated phylogenies to infer the diversification processes having generated them. Most models of diversification assume that species are exchangeable, i.e. play the same role (e.g. the neutral theory of biodiversity (1) and all lineage-based models (2). Our goal here is to develop an individual-based model of species diversification where species play different roles, depending on their order of appearance (3,4), in the sense that competition is only felt by individuals of older species from individuals of younger species. This asymmetric competition can result from various ecological mechanisms collectively gathered under the name of the Red Queen hypothesis. Achaz et al. (3) have shown that in a model of speciation by mutation, the dynamics of the metacommunity converge, in the limit of rare mutations, to a lineage-based Markov process of diversification where the speciation rate of each species is a deterministic function of its rank in the order of species appearances. Backwards in time, we obtain a coalescent process that depends on two parameters, the ratio of intraspecific to interspecific competition -a-, and the foundation time of the metacommunity -T0. We have shown that this model interpolates between the caterpillar tree (a close to 0), the Yule tree (T0 finite, a close to 1), and the Kingman coalescent (T0 large, a close to 1). We developed an MCMC algorithm for the estimation of the parameters of the model. In particular, we estimated a value of a quite close to one for the phylogeny of cetaceans, but significantly different from one, which rejects the Kingman coalescent hypothesis.

References:
1) S.P. Hubbell. The unified neutral theory of biodiversity and biogeography. Princeton Univ. Press, 2001.
2) T. Stadler. Recovering speciation and extinction dynamics based on phylogenies. J Evol Biol.,2013.
3) A. Lambert, T. Parsons, G. Achaz, N. Lartillot. A non-exchangeable coalescent arising in phylogenetics, in prep.
4) J.P. O'Dwyer, R. Chrisholm. A mean-field model for competition: from neutral ecology to the Red Queen. Ecology Letters, 2014.

Poster 17

> **Robert Noble**[1]; Michael E Hochberg[1,2]
[1] Institut des Sciences de l'Evolution (ISEM), Montpellier, France
[2] Santa Fe Institute, Santa Fe, USA.

Forecasting tumour progression

The interacting ecological and evolutionary processes that shape tumour progression and therapeutic outcomes remain poorly understood. I will show how simple mathematical models, based on empirical data, can be used to characterize processes such as competitive and cooperative cell interactions, and adaptation to changing, heterogeneous microenvironments. I will further illustrate how different sampling regimes compare in terms of their abilities to predict future population states. My findings add to understanding of the dynamics of tumour heterogeneity and the emergence of drug resistance.

Poster 18

> **Susanne P. Pfeifer**[1,2]; Stefan Laurent[1,2]; Ricardo Mallarino[3]; Matthieu Foll[1,2]; Catherine R. Linnen[4]; Jeffrey D. Jensen[1,2]; Rowan D. H. Barrett[5]; Hopi E. Hoekstra[3]
[1] School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[2] Swiss Institute of Bioinformatics, Switzerland
[3] Department of Organismic and Evolution Biology, Harvard University, USA
[4] Department of Biology, University of Kentucky, USA
[5] Department of Biology, McGill University, Canada

Genomic and functional basis of adaptive change: The selective history of camouflaged deer mouse populations

Global warming (and climate change more generally), along with anthropogenic interference in wild habitats, forces many natural populations to adapt to novel ecological conditions. However, both the causes as well as the magnitude of the genomic changes underlying these adaptations often remain poorly understood. Despite the fact that recent progress in genomics has enhanced our understanding of adaptive trait evolution at the molecular level over the past years, relatively few studies have so far been published that link beneficial mutations (and thus the molecular mechanism) to specific selective pressures in natural populations, thereby providing a complete picture of the source of evolutionary change. In addition, the genomic basis of adaptive traits and the ecological causes of selection are often very complex, generally limiting the predictability of local adaptation. In this study, we are investigating the genomic and functional basis of adaptive coat coloration of deer mouse populations to the ecologically distinct, recently formed Sand Hills in Nebraska. Previous research has shown that this coat color change, caused by allelic variation at the Agouti pigmentation locus, is an adaptation for crypsis (helping mice to avoid predation from visually hunting avian predators, which preferentially predate on poorly background-matched prey) [Linnen et al. 2013]. We take advantage of this established relationship between the camouflaging coat-color and the underlying beneficial mutations, and combine ecological sampling and population sequencing with novel statistical and computational methods in order to address questions surrounding the mode and tempo of adaptive evolution as well as the predictability of phenotypic evolution. Beginning with a transgenic study confirming that a previously identified deletion within the Agouti gene results in the lighter coat color of deer mice inhabiting this region, we then performed a large-scale field-based selection experiment that involved the introduction of 480 wild deer mice (with known genotypes and phenotypes) into replicated field enclosures, representing the two extremes in substrate color found in their natural habitat, enabling us to characterize selection on both the genotype and phenotype.

Linnen, CR, Poh, YP, Peterson, BK, Barrett, RD, Larson, JG, Jensen, JD, Hoekstra, HE (2013). Adaptive evolution of multiple traits through multiple mutations at a single gene. Science, 339, 6125:1312-6.

Poster 19

> **Pierre Pudlo**[1,2]; Arnaud Estoup[4,2]; Jean-Michel Marin[3,2]; Louis Raynal[3]; Mathieu Ribatet[3]; Christian P. Robert[5]
[1] Aix-Marseille University (AMU), Marseille, France
[2] Institut de Biologie Computationnelle (IBC), Montpellier, France
[3] Université de Montpellier, Montpellier, France
[4] Centre de Biologie pour la Gestion des Populations INRA, Montpellier, France
[5] Université Paris-Dauphine, France

Sharp approximate Bayesian computation with random forest

Approximate Bayesian computation (ABC, see Marin et al., 2012) provides an elaborate approach to Bayesian inference on complex models, including model choice. It has been mainly developed to answer inferential issues in neutral population genetics (Tavaré et al., 1997; Pritchard et al., 1999; Beaumont, 2010). Both theoretical arguments and simulation experiments indicate, however, that model posterior probabilities, as well as parameter posterior distributions, may be poorly evaluated by

standard ABC techniques (Robert et al., 2011; Blum et al., 2013; Biau et al., 2015). We propose a novel approach based on a machine learning tool named Random Forest (Breiman, 2001) to conduct both model selection among possible evolutionary scenarios and parameter inference such as population sizes, dates of divergence or migration rates. The main step of the method is to train various Random Forests on the many simulations produced by the ABC algorithm in order to predict either the scenario or the parameter based on a set of summary statistics. We will illustrate the power of this novel methodology by analyzing population genetic datasets.

Beaumont, M. (2010). Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics, 41, 379–406.
Biau, G., Cérou, F., and Guyader, A. (2015). New insights into Approximate Bayesian Computation. Annales de l'Institut Henri Poincaré B, Probability and Statistics, 51(1), 376–403.
Blum, M., Nunes, M., Prangle, D., and Sisson, S. (2013). A comparative review of dimension reduction methods in Approximate Bayesian Computation. Statistical Science, 28(2), 189–208.
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
Marin, J., Pudlo, P., Robert, C., and Ryder, R. (2012). Approximate Bayesian computational methods. Statistics and Computing, 22(6), 1167–1180.
Pritchard, J., Seielstad, M., Perez-Lezaun, A., and Feldman, M. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Molecular Biology and Evolution, 16, 1791–1798.
Robert, C., Cornuet, J.-M., Marin, J.-M., and Pillai, N. (2011). Lack of confidence in ABC model choice. Proceedings of the National Academy of Sciences, USA, 108(37), 15112–15117.
Tavaré, S., Balding, D., Griffith, R., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. Genetics, 145, 505–518.

Poster 20

> **David Salthouse**[1]; Clara Champagne[1]; Maite Aubry[2]; Yoann Teissier[2]; Elodie Descloux[2]; Mai Cao-Lormeau[2]; Bernard Cazelles[1]
[1] IBENS CNRS, Paris, France
[2] Institut Louis Malardé, Papeete, Polynésie Française

French Polynesian Dengue epidemics : The coupling of phylogeny and mathematical models.

As more and more virus sequences become available, phylogenetics is increasingly used to study infection history (who infected whom) through the reconstruction of phylogenetic trees. If enough mutations accumulate at the time scale of the population dynamics then the structure of the phylogenetic tree may be modified by the past dynamics. The field of phylodynamics is concerned with recovering information on the tree-shaping dynamics through statistical analysis. Luckily for epidemiologists, viral genetics have such high mutation rates that some information on the dynamics of viral epidemics can be retrieved through its phylogeny. Starting with coalescence theory [1], new phylodynamic methods [2,3,4] have been developed to estimate the parameters of increasingly complex stochastic non-linear dynamical models. Here we present the results of our study of the dengue epidemics in French Polynesia. Characterised by the absence of co-circulating dengue serotypes, these epidemics rapidly propagate between the different peninsular islands. The surveillance program was conducted by the Institute Louis Malardé who recorded the historic incidence and collected more than 500 genetic sequences of dengue strains emerging in French Polynesia since 1982. Using methods derived from [2,4], an estimation of epidemiological parameters is obtained via Bayesian inference coupling simultaneously genetic sequence data and historic incidence data.

[1] Kingman J.F.C., 1982, "The coalescent", Stoch. Proc. Appl.
[2] Volz E.M., 2012, "Complex population dynamics and the coalescent under neutrality", Genetics.
[3] Stadler T., Bonhoeffer S., 2013, "Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods." Phil. Trans. R. Soc. B
[4] Rasmussen D. A., Volz E. M. and Koelle K., 2014, "Phylodynamic inference for structured epidemiological models." Plos Comp. Biol.

Poster 21

> **Guillaume Scholz**[1]
[1] University of East Anglia, Norwich, UK.

Representing orthology relations beyond trees

Reconstructing the evolutionary past of a gene family G is an important aspect of many genomic studies. To help with this, simple operations on a set of sequences, called orthology relations, have been introduced. An attractive aspect of such relations is that a characterization for when they can be represented in term of a phylogenetic tree is known. For many reasons, it is however too much to hope for that real biological orthology relations satisfy this characterization. Rather than trying to correct the data, we propose representing orthology relations in terms of a structurally very simple phylogenetic network called a level-1 network, which generalizes phylogenetic trees. We first provide a mathematical formalization of orthology relations and their link with phylogenetic trees. Then, we review relevant existing results, and present our most recent ones.

Poster 22

> **Mircea T. Sofonea**[1,3]; Samuel Alizon[2,3]; Yannis Michalakis[2,3]
[1] Université de Montpellier (UM), Montpellier, France
[2] Laboratoire Maladies Infectieuses et Vecteurs : Écologie, Génétique, Évolution et Contrôle (MIVEGEC), Montpellier, France
[3] Centre National de la Recherche Scientifique (CNRS), France

Modelling multiple infections: from within-host interactions to parasite evolution

Multiple infections are a major concern in public health, human and veterinary medicine and phytopathology but they are also an interesting subject for ecology and evolution. In a coinfected host, the different parasite genotypes can interact in various ways, for instance via the production of public goods molecules. Within-host interactions thus create diverse parasite load dynamics that makes it difficult to predict the epidemiology at the between-host level. In particular, the best genotype at one level may not be the best at the other level. Most of epidemiological models avoid this complexity by considering only one parasite genotype or by arbitrarily assuming either a superinfection or a coinfection pattern. Here we develop a deterministic model that explicitly nests within-host interactions into a generic epidemiological model. Owing to feasibility and stability arguments, we manage to map a diversity of epidemiological scenarios, way beyond the co/superinfection dichotomy. We then integrate analytical and numerical results from both within and between-host levels into the adaptive dynamics framework to investigate the evolution of parasite traits. We show in particular that epidemiological feedbacks select for intermediate levels of within-host cooperation. As a conclusion, we provide tools to open the within-host black box which is a necessity to predict parasite evolution.
Ref: Sofonea MT, Alizon S, Michalakis Y. 2015 From within-host interactions to epidemiological competition: a general model for multiple infections. Phil. Trans. R. Soc. B. 370: 20150303.

Poster 23

> **Nihan Tokac**[1]; Magnus Bordewich[1]
[1] School of Engineering and Computing Sciences, Durham University, Durham, United Kingdom

An algorithm for reconstructing ultrametric tree-child networks from intertaxa distances

Traditional “distance based methods” reconstruct a phylogenetic tree from a matrix of pair-wise distances between taxa. A phylogenetic network is a gen- eralization of a phylogenetic tree that can describe evolutionary events such as reticulation and hybridization that are not tree-like. Although evolution has been known to be more accurately modelled by a network than a tree for some time, only recently have efforts been made to directly reconstruct a phylogenetic network from sequence data, as opposed to reconstructing several trees first and then trying to combine them into a single coherent network. In this work we present a generalisation of the UPGMA algorithm for ultrametric tree reconstruction which can accurately reconstruct ultra- metric tree-child networks from the set of distinct distances between each pair of taxa.

Poster 24

> **Matteo Tomasini**[1][2]; Stephan Peischl[1][2]
[1] Interfaculty Bioinformatics Unit, University of Bern, Switzerland; [2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Evolutionary rescue in spatially structured populations

Evolutionary rescue (ER) describes the process of adaptive evolutionary change that restores positive growth to declining populations and hence prevents extinction. Theoretical and experimental studies of ER identified the demographic, genetic, and extrinsic factors that affect the probability of ER. Identifying factors that affect the probability of ER is important in the face of rapid climatic and environmental change and in the context of evolution of drug resistance during treatment. It has been argued theoretically and experimentally that population structure can strongly impact the probability of rescue. In particular, it has been suggested that intermediate migration rates are optimal for evolutionary rescue and that global migration (e.g., as in the island model) generally lead to a higher probability of rescue as compared to local migration (e.g., as in a stepping stone model). Here we build on these results and investigate the effects of a gradually changing environment in a spatially structured population with a variety of migration kernels, including the island model, the stepping stone model, and a mixture of both. We present a mathematical analysis using multi-type branching process, complemented by extensive computer simulations. We derive simple analytical approximations that allow us to better understand the process of evolutionary rescue in spatially structured populations and help us to guide our intuition. Our key findings are that (i) the probability of rescue is larger for local migration as compared to global migration for a wide range of parameters, (ii) habitat fragmentation has a strong impact on ER in the stepping stone model but not the island model, (iii) the range of parameters for which we observe ER is generally larger under the stepping stone model.

ATTENDEES

ACHAZ Guillaume MNHN
ANCIAUX Yoann Institut des Sciences de L'Evolution de Montpellier
BALDISSERRI Agnese Department of Mathematics and Computer Science, University of Florence
BANK Claudia Instituto Gulbenkian de Ciência
BEERAVOLU REDDY Champak The City College of New York
BOSKOVA Veronika ETH Zurich
BRETON Catherine ISEM
CASANELLAS Marta Universitat Politècnica de Catalunya
CASSAN Elodie Université de Montpellier / CNRS
CHERNOMOR Olga Center for Integrative Bioinformatics Vienna (CIBIV), Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna
CHIFOLLEAU Anne-muriel LIRMM
CLEMENTE Florian University of Montpellier
DE MAIO Nicola University of Oxford
DJIDJOU-DEMASSE Ramses INRA-Bordeaux & Mathematics Institute of Bordeaux (IMB)
DOERR Daniel Ecole Polytechnique Fédérale de Lausanne (EPFL)
FERNÁNDEZ SÁNCHEZ Jesús Universitat Politècnica de Catalunya
FERRETTI Luca Pirbright Institute
FRAGATA Inês Instituto Gulbenkian de Ciência
FUTSCHIK Andreas Johannes Kepler University Linz
GALTIER Nicolas University Montpellier - CNRS
GANDON Sylvain CNRS
GASCUEL Olivier CNRS and Institut Pasteur
GONZALEZ CASANOVA Adrian Weierstraß-Institut Berlin
GOUY Alexandre Institute of Ecology and Evolution - University of Berne
GUGLIELMINI Julien Institut Pasteur
GUINDON Stephane CNRS
HIVERT Valentin INRA CBGP
IRWIN Kristen Ecole Polytechnique Fédérale de Lausanne (EPFL)
ISHIKAWA Sohta University of Tokyo
JENSEN Jeyffrey Ecole Polytechnique Fédérale de Lausanne (EPFL)
KUBATKO Laura S. Departments of Statistics and Evolution, Ecology, and Organismal Biology - The Ohio State University
LADRET Véronique Université de Montpellier
LAFONTAINE Ingrid Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Institut de Biologie Paris-Seine, UPMC, Paris
LAMBERT Amaury Stochastics & Biology Group - Laboratoire de Probabilités & Modèles Aléatoires - UPMC - Université Paris 06
LAPEYRONNIE Jason INRA
LEBLOIS Raphael INRA
LEDDA Alice Imperial College London

LEFORT Vincent	LIRMM - CNRS
LEMOINE Frederic	Institut Pasteur
LINARD Benjamin	LIRMM
LION Sébastien	CNRS
MALIET Odile	ENS
MARTIN Guillaume	CNRS
MASSIP Florian	Laboratoire de Biométrie et de Biologie Evolutive
MATUSZEWSKI Sebastian	Ecole Polytechnique Fédérale de Lausanne (EPFL)
MERLE Coralie	Institut Montpelliérain Alexander Grothendieck (IMAG)
MINH Bui Quang	Center for Integrative Bioinformatics Vienna
MIRO PINA Veronica	Université Pierre et Marie Curie
NEHER Richard	Max Planck Institute for developmental biology
NOBLE Robert	Institut des Sciences de l'Evolution, University of Montpellier
PARDI Fabio	CNRS - LIRMM
PERRET Cédric	UM (student) MIVEGEC (current internship)
PFEIFER Susanne	Ecole Polytechnique Fédérale de Lausanne (EPFL)
PUDLO Pierre	Aix-Marseille University / CNRS / IBC
PULICANI Sylvain	LIRMM
RANNALA Bruce	UC Davis - University of California
ROUSSET Francois	CNRS
SAGULENKO Pavel	Max Planck Institute for developmental biology
SALTHOUSE David	IBENS CNRS
SANMARTÍN BASTIDA Isabel	Real Jardín Botánico (RJB-CSIC)
SAULNIER Emma	Laboratoire MIVEGEC, LIRMM, IBC, UM
SCHOLZ Guillaume	Unversity of East Anglia
SOFONEA Mircea T	Université de Montpellier
SWENSON Krister	LIRMM, IBC
TOKAC Nihan	Durham University
TOMASINI Matteo	University of Bern
TRUSZKOWSKI Jakub	EBI Goldman Group And Cancer Research UK (S.Tavare)
WU Chieh-hsi	University of Oxford
WYMANT Chris	Department of Infectious Disease Epidemiology, Imperial College London
ZHUKOVA Anna	Institut Pasteur, Paris

Thanks to the Agropolis Fondation, PhD students and postdocs have benefited of financial support for all conference fees.