



Roscoff (France), 31 March – 5 April 2012

Theoretical and empirical advances in evolutionary genomics

Développements théoriques et empiriques en génomique évolutive

PRESIDENT: Juliette DE MEAUX

Institute for Evolution and Biodiversity, Münster University, Germany

VICE-PRESIDENT: Xavier VEKEMANS

Génétique et Evolution des Populations Végétales, CNRS – Université Lille 1, France

With collaboration from:

Guillaume ACHAZ, UPMC Paris 6

Laurent DURET, CNRS Lyon

Oscar GAGGIOTTI, Univ. Joseph Fourier Grenoble

Nicolas GALTIER, CNRS Montpellier

Lluis QUINTANA-MURCI, Institut Pasteur Paris

Olivier TENAILLON, INSERM Paris

Renaud VITALIS, INRA Montpellier

Frédéric AUSTERLITZ, CNRS Paris

Conference Report

Rapport sur la conférence

PART I PROGRAM OVERVIEW

The conference took place in Roscoff, France, from March 31st to April 4th

The program included an opening session with 2 keynote lectures, 6 oral communication sessions with a total of 42 talks, and 3 poster sessions with a total of 61 posters.

Session I: Theoretical Population Genomics

Session II: Towards a fine-grained map of genetic variation in model organisms

Session III: Evolutionary Genomics

Session IV: Evolutionary Systems Biology

Session V: The ecological genetics of molecular systems controlling complex phenotypes

Session VI: Open session

PART II PARTICIPANTS



Invited speakers

The 27 invited speakers gave each a 30 minute talk about their recent work.

BALDING David - London, UNITED KINGDOM

BLUM Michael - Grenoble FRANCE

CARBONE Alessandra - Paris, FRANCE

DE MEAUX Juliette – Munster, GERMANY

DURET Laurent - Lyon, FRANCE

DUTHEIL Julien - Montpellier, FRANCE

EXCOFFIER Laurent - Bern, SWITZERLAND

GLÉMIN Sylvain - Montpellier, FRANCE

GORDO Isabel - Lisbon, PORTUGAL

HEYER Evelyne - Paris, FRANCE

KAESSMANN Henrik - Lausanne, SWITZERLAND

LANDRY Christian - Québec, CANADA

LASSIG Michael - Köln, GERMANY
LOUDET Olivier - Versailles, FRANCE
McLYSAGHT Aoife - Dublin, IRELAND
NORDBORG Magnus - Vienna, AUSTRIA
PATTERSON Nick – Boston, USA
PETROV Dmitri - Stanford, USA
QUINTANA-MURCI Lluís - Paris, FRANCE
RATTEI Thomas Rattei - Vienna, AUSTRIA
ROCHA Eduardo - Paris, FRANCE
SCHMID Karl - Stuttgart, GERMANY
SCHMITT Johanna - Providence, USA
TENAILLON Olivier - Paris, FRANCE
VEKEMANS Xavier – Lille, FRANCE
WHEAT Christopher W. - Helsinki, FINLAND
WITTKOPP Trisha - Michigan, USA

From these 27 invited speakers, 7 were females, and 10 did attend the previous Jacques Monod conference on evolutionary genomics (May 2007). Eleven invited speakers were from France, 11 were from Europe (not including France), and 5 were from outside Europe (we had several cancellations from invited speakers from outside Europe, including one due to a Visa problem).

A total of 79 non-invited participants attended the conference. We selected 17 contributed talks of 15 minute from these participants.

FRANCOIS Olivier – Grenoble, FRANCE
FERRETTI Luca –Barcelona, SPAIN
VAN HEERWAARDEN Joost – Wageningen, THE NETHERLANDS
SEGUREL Laure – Chicago, USA
TENAILLON Maud – Paris, FRANCE
BATAILLON Thomas – Aarhus, DENMARK
ROMIGUIER Jonathan – Montpellier, FRANCE
SZOLLOSI Gergely – Lyon, FRANCE
PERFEITO Lilia – Köln, GERMANY
DILLMANN Christine – Paris, FRANCE
ACHAZ Guillaume – Paris, FRANCE
CHAN Frank – Plön, GERMANY
AARTS Mark – Wageningen, THE NETHERLANDS
JORON Mathieu – Paris, FRANCE
BOITARD Simon – Toulouse, FRANCE
CASTRIC Vincent – Lille, France

From these 79 other participants, 52 were from France, 24 from Europe (not including France) and 3 from outside Europe.

PART III SCIENTIFIC PROGRAM

Opening session

The conference was launched with two plenary lectures held on Saturday evening. The first was held by Michael Lässig, professor in Theoretical Physics at the University of Cologne in Germany, who brilliantly illustrated how concepts and methods of theoretical and statistical physics help our understanding of genome evolution. M. Lässig presented his latest findings on the evolution of the influenza virus, showing that, although influenza viruses are subject to very strong natural selection, variation remains at all time. The comparative analysis of different portions of the genomes showed that most evolution occurs in gene regions detected by the human immune system. More importantly, he showed that evolution of influenza presented both strong selection and high diversity as a result of evolutionary competition between different gene functions, and not between equally fit clones as expected under classical scenarios of clonal interference.

Evelyne Heyer, Professor at the Natural History Museum in Paris, held the second plenary lecture. She presented an impressive summary of 10 years of research on Human population genetics in Central Asia. She began her lectures with a brief description of population history in Central Asia, with times to population expansion being older in the Eastern area. Genetic differences among human populations in Central Asia correlated better with linguistic rather than with geographical distances, showing the importance of culture in shaping genetic diversity in humans. E. Heyer further presented evidence that ecological differences have shaped the human genome. Her lab conducted an extensive genotype-phenotype association study in this region, associating lactase persistence with life-style and dating the expansion of the persistence allele to 6000-12000 years ago. This seminar set a valuable framework for studies of human genetics that were to be presented on the next day, it further eloquently outlined how population genomics can help understand the history and dynamics of human populations.

Session I: Theoretical Population Genomics

This session presented some of the new developments in the statistical framework used to analyze genome-wide polymorphism datasets. Nick Patterson, senior scientist at the Harvard's Broad Institute, presented some developments in analysis of human datasets. In particular he showed that large polymorphism datasets cannot be explained by a single phylogeny of human populations, and he introduced a "test of treeness" in order to verify this property. He also introduced models of evolution comprising instantaneous admixture, instead of continuous migration, in order to better mimic the genetic effects of the Neanderthal introgression. He also presented a new database of human polymorphism data that comprises carefully controlled ascertainment in order to avoid biases of the previous databases. Michael Blum, CNRS research associate at Université Joseph Fourier (Grenoble) presented new approaches to detect anisotropic isolation by distance in human populations, and to detect rapid spatial range expansion processes. Olivier François, professor at the same institution, presented new methods, entitled "Latent Factor Mixed Models", to simultaneously detect population genetic structure and gene-environment associations, with the aim of detecting environmental adaptation. Isabel Gordo, from the Instituto Gulbenkian de Ciencia in Portugal, presented empirical tests in populations of *Escherichia coli* of two important models of adaptation, the clonal interference model and the moving optimum fitness landscape model. Most results obtained previously in *in vitro* conditions were found to be preserved in *in vivo* conditions within mice hosts. Luca Ferreti, from the Centre for

Research in Agricultural Genomics in Barcelona, presented new tests of neutrality that are integrating data on the frequency distribution of single nucleotide polymorphisms with patterns of linkage disequilibrium among sites. He showed the power of these methods in studies aiming at inferring evolutionary scenarios using Approximate Bayesian Computation approaches. Julien Dutheil, a CNRS Research Associate that is currently working in the Max-Planck Institute of Marburg, presented methods that are exploiting the incomplete lineage sorting observed in gene genealogies of large primates. In particular it was shown that higher correlations in gene expression between humans and chimpanzees were observed for genes lying in regions where humans and chimpanzees were the closest relatives in gene genealogies. Also, it was shown that current genomic patterns of recombination in humans were similar to those of the direct ancestor of humans and chimpanzees. These results suggest that the genomic environment of a gene needs to be taken into account when testing for adaptation in large primates. Joost van Heerwaarden, from Wageningen University (The Netherlands), presented results from simulations of genomic and phenotypic evolution under divergent selection, in order to test different approaches to perform Marker-Assisted selection in domesticated organisms. Finally, David Balding, senior scientist at the Genetics Institute from the University College London, presented new approaches to estimate relatedness among individuals, using large genomic datasets, which allow better estimates of important quantities such as heritability.

Session II: Towards a fine-grained map of genetic variation in model organisms

Since the previous evolutionary genomics conference, whole genome sequence datasets obtained by re-sequencing approaches in model organisms have been accumulated at high rate. This session was dedicated to studies exploiting these new datasets. Laurent Excoffier, Professor at the University of Bern, documented the use of such datasets to understand the consequences of range expansion in ancient human history. Mutation surfing is an important process associated to range expansion, and is due to the fact that genetic drift is an important evolutionary force at the migration front. Using demographic models, he showed that the relatively high level of introgressed genes from Neanderthals found today in the modern humans gene pool could be due to much lower rates of interbreeding, associated to a mutation surfing process. He also presented results illustrating the effect of mutation surfing at a much more recent time, in association with the colonization of Quebec by European immigrants. Guillaume Laval, from Institut Pasteur, Paris, presented results aiming to find genomic regions subject to positive selection based on re-sequenced whole genome datasets. He introduced a new test, called the intra-allelic nucleotide diversity test (DIND) and compared it to other tests based on haplotype length or frequency-spectra. He showed that this test is highly powerful and allows to detect positive selection even in datasets characterized by low sequence coverage such those from the "1000 genomes project". Laure Segurel, postdoc at the Department of Human Genetics in University of Chicago, focused on a single genomic region, the ABO blood group system locus, and compared polymorphisms from different primates. Two hypotheses could account for the observation of shared polymorphisms between great Apes species, either convergent evolution, or trans-species polymorphism. Based on allelic genealogies of A and B alleles she concluded very convincingly that the ABO system showed trans-species polymorphism due to balancing selection. Magnus Nordborg, Director of the Gregor Mendel Institute in Vienna, presented results on 230 genomes of *Arabidopsis thaliana*, as part of the 1001 Arabidopsis Genomes initiative. The most striking result was the observation of an important variation in genome size, with up to 20 Mb differences among accessions, in

an overall genome of only about 135 Mb. This variation was geographically structured, and was mainly accounted for by differences in regions encompassing ribosomal repeats and centromeric repeats. Some other regions that do not map to the reference Col-0 genome were found to map to the close relative *A. lyrata* genome, or were found to be shared with other thaliana accessions, suggesting a substantial flexibility in genome composition among accessions. Maud Tenaillon, from Paris (Le Moulon) presented the result of an investigation combining mapping approaches and experimental evolution in maize to reveal the determinants of a model complex trait in plants, i.e. flowering time. This study revealed residual heterozygosity in the parental lines at a locus affecting flowering time. Polymorphism at that locus was subsequently found in a panel of accessions, and the association with flowering time variation and other morphological traits suggested pleiotropic effects, possibly due in part to the occurrence of a cluster of genes linked through epistasis. The last contribution was due to Lluís Quintana-Murci from Institut Pasteur (Paris), who illustrated how population genomic data in humans could be used to investigate the evolutionary consequences of cultural changes in mode of subsistence. He compared whole genome polymorphism data between Bantu farmers and Pygmy hunter-gatherer from central Africa. Higher values of linkage disequilibrium were found in pygmies, as compared to Bantu populations, suggesting bottlenecks or founder effects in the former. Genomic signatures of population admixture, in the form of large introgressed genome chunks, were found in some pygmy populations. Moreover a signal of sex-biased admixture was found, with more introgressed regions across the X chromosome than autosomes. Finally, differences in selection signatures at many genes were found between the two groups of populations, suggesting indeed that the differences in modes of subsistence have impacted human genetic diversity. Overall this session illustrated a wide diversity of studies on model organisms that used whole-genome or region-specific re-sequencing approaches to highlight evolutionary processes within species.

Session III: Evolutionary Genomics

This session aimed at illustrating the insight gained from recent comparative genomic approaches in deciphering the processes of molecular evolution, and discriminating between the many evolutionary forces (mutation, selection, genetic drift, recombination) that shape genomes in the long run. The first talk, by Dmitri Petrov, Professor at Stanford University, focused on the process of adaptation in *Drosophila* populations and specifically aimed at distinguishing between hard versus soft sweep processes as primary sources of adaptation. Using pesticide resistance as a model phenotype, he showed evidence for the occurrence of soft sweeps in *Drosophila*, suggesting that effective population sizes are several orders of magnitude higher than previously acknowledged. Using a combination of genome-wide re-sequencing data, sliding window and ABC approaches, he showed that indeed effective population sizes are very high in populations of *Drosophila melanogaster*. Aoife Mc Lysaght, from the University of Dublin, investigated the process of paralogous gene loss after whole genome duplication. She showed that for pairs of gene products involved in protein-protein interaction, there is a general bias in gene loss leading to the co-occurrence of such pairs of genes within the same genome block, rather than in paralogous genome blocks, which suggests that physical clustering of genes in eukaryotic genomes is biologically significant overall. Thomas Bataillon, from Aarhus University (Denmark), presented estimates of adaptive evolution as well as purifying selection in chimpanzees for the X chromosome as compared to autosomes. In both cases, evidence for selection acting on nucleotide sequences is much higher for the X chromosome as compared to

autosomes. It is suggested that most adaptive mutations are recessive and therefore positive selection is more efficient on the X chromosome as recessive mutations will be expressed in males. Eduardo Rocha, from Institut Pasteur, Paris, compared patterns of genome organization and evolution in fast-growing versus slow-growing Prokaryotes. He found substantial differences in terms of gene distribution, gene content, and codon usage bias. Hence, retrospectively, quantifying such patterns from metagenomic data allows identification of the amount of fast-growing versus slow-growing organisms, and this allows further ecological inference. Patterns of genomic evolution were also used by Jonathan Romiguier, from the University of Montpellier 2 (ISEM lab), to make inference on the life-history properties of mammalian ancestors. Their analysis suggests that such ancestors would have been rather large and long-lived, in striking contrast to most previous scenarios based on fossil evidence. Karl Schmid from the University of Hohenheim (Germany) presented results from genome-wide analysis of polymorphisms in *Arabidopsis thaliana*. Large differences in effective population sizes were inferred from different geographic regions, and these correlated with the strength of selection inferred from molecular evolution features. The dataset was also used to test a recent suggestion that selection could act on the efficiency of nitrogen use. Gergely Szollosi, from Université Claude Bernard in Lyon (LBBE lab), suggested that documented patterns of horizontal gene transfers in Prokaryotes could be used to improve molecular dating of phylogenies, as they impose the simultaneous occurrence of the two organisms exchanging genetic information. He explored models aiming at incorporating gene transfer processes, in addition to the speciation, duplication and gene loss processes of molecular evolution. Sylvain Glémin, from the University of Montpellier 2 (ISEM lab), focused on the process of GC-biased gene conversion in plants. By analyzing datasets of EST data from over 200 plant species, he showed that GC-rich and GC-heterogeneous genomes occur in several groups outside of the monocot clade, and also that this is not the rule for all monocots. The respective roles of mutation bias, selection for codon usage, and biased gene conversion on the relative occurrence of GCs at first, second and third codon positions were discussed.

Session IV: Evolutionary Systems Biology

Combining sensitive and accurate functional methods with genomics analysis promises to unravel the laws determining the evolution of molecular networks. The goal of Session IV was to highlight the bearings of such approaches on the reconciliation between modern molecular biology and evolutionary theory. Christian Landry, Professor at University of Laval, Quebec, inaugurated this session with a report on a gigantic experiment performed in his lab: the analysis of interspecific variation in protein-protein interaction systems in yeast. He presented a convincing methodology applied to two large protein complexes and concluded that protein-protein interactions are generally well conserved between species. This approach will be applied to the extent of protein interactions across yeast species.

After him, Thomas Rattei, from University of Vienna, Austria, documented the dynamics of protein-protein interaction in this genus and the progressive gain (and supposedly parallel loss) of protein interaction networks. This was followed by a short presentation by Lilia Perfeito, who presented the outcome of an elegant experiment designed to describe the fitness landscape of an iconic molecular pathway: the lactose operon.

The session went on after a short break. Dr. Olivier Tenaille, from the Pasteur Institute in Paris, presented ground-breaking results: the whole genome re-sequencing of experimental bacterial populations. This experiment demonstrates the existence of adaptive convergence. Specific gene complexes are commonly recruited to respond to

selection, although mutations themselves rarely occur at the same site. This was followed by a short presentation by Christine Dillmann showing how nutrient resources shape metabolic diversity and plasticity in *E. coli*. Henrik Kaessmann, Professor at University of Lausanne, Switzerland, presented an exhaustive study of tissue-specific transcriptomic variation across mammals. Among other things, his data showed that species with a small population size diverge more. It further analyzed the dynamics of dosage compensation on sex chromosome. His dataset was particularly informative in this respect since some lineages present different sex chromosomes. The session was followed by a short presentation by Guillaume Achaz analyzing the order of mutations in real and model fitness landscape and closed by a presentation of recent advances in domain annotation by Alessandra Carbone, Professor at University Paris VI.

Session V: The ecological genetics of molecular systems controlling complex phenotypes

Illustrating how molecular components underpinning the adaptive phenotypes are involved in the response to selection was the goal of this session, to help establish an overarching link between specific mutations and fitness.

The session was opened by a presentation by Trisha Wittkopp, Professor at Ann Arbor University, Michigan. She described the results of an experiment designed to highlight the molecular and evolutionary properties of regulatory mutations. Studying a population of *Saccharomyces cerevisiae* exposed to weak EMS doses, she identified 231 regulatory mutants. By analyzing their expression in haploid and diploid states, she could show that the various classes of regulatory mutations differ markedly in both their effect and their dominance properties. This experiment is the first to demonstrate that cis- and trans- regulatory mutations can show different evolutionary trajectories since they will respond differently to selection and occur at different rates. Olivier Loudet, from INRA, Versailles, followed with a seminar presenting the careful dissection of molecular variants controlling phenotypic variation in plant growth, in the model plant species *Arabidopsis thaliana*. Among other things, his findings show that gene duplications can generate Dobzhansky-Müller incompatibilities, via the dys-regulation of epigenetic silencing mediated by small RNAs. Such findings were in fact thought provoking for genome scientists. In a short talk, Frank Chan, from Max Planck Institute for Evolutionary Biology in Plön, Germany, exposed a brief theory of how adaptation shapes genome.

The session was continued in the late afternoon, after a long break, with a presentation of the first association genetics study of local fitness and genome variation by Johanna Schmitt, professor at Brown University, Rhode Island, and member of the National Academy of Science. This research, at the forefront of ecological genetics, uses *Arabidopsis thaliana* to establish innovative connections between gene networks and fitness expressed in natural conditions. Among other things, she showed how quantitative trait loci (QTLs) maximize fitness in a manner dependent on the environment in which these QTLs are expressed. Subsequently, Mark Aarts, Professor at the University of Wageningen, The Netherlands, presented evidence showing the importance of another trait of major ecological importance, heavy metal accumulation, for fighting off herbivores. His result suggests that, in *Nocceae caerulescens*, heavy metal accumulation may stand in a trade-off between growth and defense. Christopher Wheat, from University of Helsinki, Finland, presented a fascinating example of adaptation in a metapopulation. He showed how, in butterflies, flight performance is associated with specific gene variants, an advancement that takes full advantage of the new sequencing technologies. Right afterwards, Mathieu Joron, from the Museum of Natural History in

Paris, dissected the molecular basis of mimicry, the unique ability of non-toxic species of butterflies to develop wing patternings closely matching unrelated toxic genera. Juliette de Meaux closed the session by presenting her recent research aiming at characterizing the adaptive properties of microRNAs.

Session VI: Open session

The last session was planned to remain broad in topic, to present aspects that had not been covered throughout the other sessions. It was initiated by a fascinating talk by Laurent Duret, analyzing the evolution of recombination hotspots in humans and their closest relatives. By using accurate population genetics model, he clarified an observation that was so far paradoxical, namely that mutated versions of the motif recognized by PRDM9, the major determinant of recombination hotspots, are enriched in regions of high recombination. Simon Boitard followed to present a new analytical method for detecting signatures of selective sweeps drawn from pooled samples subjected to deep sequencing. The session was then closed by two coordinated talks by Xavier Vekemans and Vincent Castric from the University of Lille. They presented recent advances in the molecular evolution of alleles responsible for self-incompatibility in outcrossing Brassicaceae species. Their recent advances highlight that dominance relationships among alleles become crucial when strong negative frequency-dependent selection is at play. These relationships seem to be determined by the diversity of epigenetic signals.

Final Comments

The goal of this conference was to present current advances in both theoretical and empirical approaches in Evolutionary Genomics. This discipline is currently experiencing an unprecedented acceleration since novel sequencing technologies have i) dramatically increased the size and accuracy of datasets describing genomes and their variation and ii) made access to genetic information possible for virtually any species, thereby breaking the exclusivity of model systems in the field of genomics. We believe the conference met its goal. Invited speakers gave cutting-edge science talks, using the best communication standards (clear slides, clear speeches). Research presented was often recently published in major scientific journal such as Nature, Science, Plos Biology or Nature Genetics (or about to be published there). The speakers were accessible, present at all sessions and interested in each other's research. The audience was so lively that questions sometimes fused during the talks, smoothing barriers between the speakers and the audience and creating an informal atmosphere most favorable to scientific exchanges. In addition, we felt it was crucial for the success of the conference to give posters the attention and credit they deserved. For this, we spared three time-slots for visiting each time 20 to 30 posters. Poster presenters really appreciated this opportunity.

Perspective

The participants felt there is a need to organize another meeting on this topic in 2015. Xavier Vekemans agreed to serve as president and Christian Landry accepted to take over the charge of vice-president for the next meeting. We believe that Christian Landry, a Harvard-trained scientist now established in Quebec, is best suited to promote the international visibility of the genomic and population genetics excellence at work in the French scientific landscape.