

**Mind the gap:
joining theoretical and empirical
population genetics**

Freiburg, 2nd and 3rd October 2009.

funded by the



and supported by the BMBF through



Table of contents

Program	3
Abstracts	5
List of Participants	27

Program

	October 2nd
8:30-9:00	Registration
9:00-9:05	Welcome
9:05-9:45	AMKE CALIEBE A Test for Gene Conversion and Results in the Human Growth Hormone (<i>GHI</i>) Gene Promoter
9:45-10:25	JOACHIM HERMISSON The role of ecology in the adaptive process
10:25-10:40	JAREK BRYK Copy number variants, gene expression and positive selection in wild mice
10:40-11:00	Break
11:00-11:40	ARNE NOLTE Analyses of genetic introgression as a means to identify genetic differentiation among populations of sculpins
11:40-12:30	PETER PFAFFELHUBER Primer for Discussion on <i>Speciation models</i>
12:30-14:30	Lunch break
14:30-15:10	SYLVAIN MOUSSET Coalescent simulations: how simulation methods shape the trees
15:10-15:50	FLOYD REED Underdominance and Population Transformations
15:50-16:05	FABIAN STAUBACH The emergence of a new gene from an intergenic region in the house mouse lineage is associated with a selective sweep
16:05-16:30	Break
16:30-17:10	AURÉLIEN TELLIER Coalescence methods to estimate ecological parameters of seed bank and metapopulation structure
17:10-17:50	CHRISTIAN SCHLÖTTERER Massively parallel sequencing of pooled DNA samples-the next generation of molecular markers
17:50-18:05	FRANZ BAUMDICKER A mathematical model for the genomic diversity of bacterial populations

	October 3rd
9:00-9:40	TIM SHARBEL Sex, evolution and agriculture
9:40-10:20	LESLIE TURNER Associations between testis gene expression variation and male sterility in a house mouse hybrid zone
10:20-10:35	EMANUEL HEITLINGER Divergence between European and Asian populations of the swimbladder nematode <i>Anguillicoloides crassus</i>
10:35-11:00	Break
11:00-11:40	INES HELLMANN Positive selection has a significant impact on genome-wide patterns of variation in humans
11:40-12:30	BERNHARD HAUBOLD Primer for Discussion on <i>Next Generation Sequencing Data</i>
12:30-14:30	Lunch break
14:30-15:10	JOCHEN WOLF Non-linear dynamics of non-synonymous (dN) and synonymous (dS) substitution rates affects inference of selection
15:10-15:50	THOMAS STÄDLER The Impact of Sampling Schemes on the Site Frequency Spectrum in Non-equilibrium Subdivided Populations
15:50-16:05	TOBIAS LENZ Evolution of adaptive immune genes in two sympatric stickleback species (<i>G. aculeatus</i> & <i>P. pungitius</i>)
16:05-16:30	Break
16:30-17:10	Carolin Kosiol The influence of population size on patterns of natural selection in mammals
17:10-17:25	EMILIE HARDOUIN Mice genetic diversity and history on Kerguelen
	October 4th
9:30	Meeting point: Hotel am Rathaus, Rathausgasse 4, for a small hike to the Freiburger Schlossberg

Abstracts

A mathematical model for the genomic diversity of bacterial populations

FRANZ BAUMDICKER

Zentrum für Biosystemanalyse, Universität Freiburg

The genomes of bacteria are less conserved than the genomes of eukaryotes. In particular, as predicted by the distributed genome hypothesis, bacteria from the same population/ecotype may carry different genes. Datasets describing such genomic diversity currently become available by massive next-generation sequencing and require mathematical models for the explanation and prediction of the number, distribution and transfer of genes in populations of closely-related microorganisms.

Existing approaches describing the genomic diversity of bacteria are based on the total number of genes found in a distributed genome (pangenome) and the frequency distribution of genes (supragenome). We extend these ideas by explicitly taking the sample genealogy into account. We introduce a new mathematical model based on the assumption that an individual needs a minimal number of genes to be viable (core genome). In addition, new genes are taken up from the environment and may be lost again along ancestral lines. A specific feature of the model is that genes which are not in the core genome evolve neutrally.

In our presentation, we are concentrating on a dataset of marine cyanobacteria of the *Prochlorococcus/Synechococcus* clade. These distant relatives of plant chloroplasts are the most abundant phototrophic organisms in the oceans (ca 10^5 - 10^6 cells per ml of sea water). Applying our quantitative model we are able to estimate the size of the core genome of the cyanobacteria as well as the rates of gene gain and loss along ancestral lines. Using these parameters we see a good fit of the observations to our mathematical model for *Prochlorococcus*. The fit is less good for *Synechococcus* showing a potential expansion of the population or an increased amount of lateral gene transfer.

This model should also be applicable to other microbial populations, forming microbial mats or sharing other ecological niches, like gut or soil bacteria. Including elements like selection, horizontal gene transfer and structured populations into this mathematical model might elucidate how genetic diversity evolves and which mechanisms are most important in shaping metagenomic diversity.

Copy number variants, gene expression and positive selection in wild mice

JAREK BRYK

(joint with Diethard Tautz)

Max-Planck-Institute for Evolutionary Biology, Plön

Recent years brought many examples of the role of copy number variants (CNVs) - large scale deletions and insertions in the genome - in organisms' physiology. Apart from CNVs shaping genomic variation and their association with human disorders, they can influence genome-wide transcriptional activity and facilitate adaptations. I will describe our project on searching for CNVs in recently-diverged populations of *Mus musculus domesticus* and establishing their possible role in adaptations using whole-genome approaches (gene expression, comparative genome hybridization and microsatellite scans) and functional analyses. I am going to present initial data and analyses regarding whole-genome gene expression patterns and their correlation with copy number variants and discuss next steps that involve microsatellite scans to identify CNVs that underwent selective sweep.

A Test for Gene Conversion and Results in the Human Growth Hormone (*GH1*) Gene Promoter

AMKE CALIEBE

(joint with Andreas Wolf, David S. Millar, Michael Krawczak, David N. Cooper)

Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität Kiel

Gene conversion is an important mechanism of mutation which has not only served to fashion the structure of extant human genes but has also played an important role in pathology. We developed a coalescence-based test for gene conversion which employs the similarity of putative donor and acceptor sequences. The proximal promoter region of the human growth hormone (*GH1*) gene is highly polymorphic, an observation which has been attributed to gene conversion. For 14 SNPs located in the 535 bp human *GH1* promoter, a total of 60 different haplotypes were observed in a total of 577 individuals from different ethnic backgrounds (156 Britons, 116 Spaniards, 163 West Africans and 142 Asians). When all four population groups were tested separately, evidence was found in the British, Spanish and African groups for the action of *GH1* as an acceptor of gene conversion, with at least one of the four paralogous *GH* gene promoters serving as donor. A putative gene conversion hotspot spanning the transcriptional initiation site (position -6 to +25) of the *GH1* gene was found to contain several DNA sequence motifs previously shown to be associated with gene conversion. Of the *GH1* paralogues, the *GH2* gene promoter appears in particular to have acted as a donor in both Britons and Spaniards. The occurrence of gene conversion during the evolution of the human *GH* locus therefore has been established.

Reference:

Wolf A, Millar DS, Caliebe A, Horan M, Newsway V, Kumpf D, Steinmann K, Chee IS, Lee YH, Mutirangura A, Pepe G, Rickards O, Schmidtke J, Schempp W, Chuzhanova N, Kehrer-Sawatzki H, Krawczak M, Cooper DN. A gene conversion hotspot in the human growth hormone (*GH1*) gene promoter. *Hum Mutat*, 30 (2009), 239-247

Mice genetic diversity and history on Kerguelen

EMILIE HARDOUIN

(joint with Diethard Tautz)

Max-Planck-Institute for Evolutionary Biology, Plön

The house mouse (*Mus musculus domesticus*) colonized the entire world, even various islands such as Kerguelen (48°25'-50°S; 68°27'-70°35'E). This archipelago is located in the Indian Ocean more than 4000 km off the coast of Africa and Australia. Mice on Kerguelen have experienced a variety of novel selective pressures, as this environment differs from typical mouse habitat in several ways: cold temperatures, no human population establish and new types of food due to sub-Antarctic fauna and flora. To investigate the genetic diversity and the genetic history of Kerguelen, the D-loop and 18 microsatellites was typed. As expected the mice show a very low genetic diversity due to founder effect because of the geographical isolation and the fact that these are new colonizes sites. But interestingly the islands of the archipelago show different genetic pattern suggesting different sources of colonization populations and the excess of homozygous genotypes that are typical for land based populations is not evident on island populations. Here I will give an overview of the mice genetic diversity and present the future work on mice adaptation to sub-Antarctic climate.

Primer for Discussion on *Next Generation Sequencing Data*

BERNHARD HAUBOLD

Max Planck Institute for Evolutionary Biology, Plön

Thanks to next generation sequencing technology, genomics-scale data collection is becoming routine. Analysis of this data, however, is still far from routine and the purpose of our discussion is to focus on the current discrepancy between the data gathering and the data analysis capabilities of evolution laboratories new to genomics.

Analysis of genome-scale sequence data can usefully be divided in two parts, data preparation and data interpretation. Data preparation concerns such aspects as the creation of error scores, the mapping of reads and the assembly of long sequences. In all three cases quality assurance is a major concern and participants are encouraged to report on successes and failures in these areas. One possible outcome of this part of the discussion would be the identification of a set of reference procedures for reliable mapping and assembly of genomics sequencing data.

The challenge in data interpretation remains—as always—to infer parameters. In the context of genomics data, it may be particularly important to incorporate detailed error models in the estimation procedures. Moreover, in this field the implementation of new theory into robust computer programs is challenging in its own right. Any implementation needs to strike the familiar balance between ease of use and ease of programming. A further requirement that is often overlooked is long-term availability and stability of genomics software tools. In this part of the discussion we should therefore concentrate on how in evolutionary genomics the relationship between mathematicians, programmers and users can be made as fruitful as possible.

Divergence between European and Asian populations of the swimbladder nematode *Anguillicoloides crassus*

EMANUEL HEITLINGER

Universität Karlsruhe and University of Edinburgh

Anguillicoloides crassus, a nematode parasite of freshwater eels, was introduced from Asia to Europe early in the 1980s where it rapidly spread. As this introduction was accompanied by a host switch from *Anguilla japonica* to *Anguilla anguilla*, the parasite has been confronted (60-70 generations) with a new environment and host species. Cross infection experiments using both host-species and parasite populations suggest a divergence of life history traits in European populations compared to those from Asia. This divergence is being investigated using high throughput sequencing (454-pyrosequencing) of the nematode's transcriptome. Furthermore, gene expression will be analysed in *A. crassus* from cross-infection experiments using digital transcriptomics (tag-sequencing) or mRNA-seq (direct sequencing of cDNA) on an Illumina Solexa platform. Based on the divergence of both sequence and gene expression levels combined with inferences drawn from functional annotation, possible adaptation to the new host should be tested. The relative extent of direct cDNA sequencing, producing additional sequence information vs. tag sequencing, producing gene-expression information only, has yet to be determined. This will depend on both the amount of sequence divergence found in functionally interesting genes and the methods developed by theoreticians to analyse such data under strong demographic influence.

Positive selection has a significant impact on genome-wide patterns of variation in humans

INES HELLMANN

(joint with Anders Albrechtsen, Yingrui Li, Jun Wang, Rasmus Nielsen)

Max F. Perutz Laboratories, University of Vienna

Population genetic evidence for positive selection in the human genome is accumulating. However, most methods used to detect selection are designed to identify outliers and it remains unclear how many outliers truly are selected. This is because we do not know with what frequency advantageous mutations occur and get fixed, i.e. the frequency of selective sweeps. A couple of recent papers have attempted to get such estimates from the relationship of the nucleotide diversity and recombination rates. However, those estimates are confounded by the effects of purifying selection on linked neutral variants (background selection). Here, we devise the allele frequency spectrum (AFS) as a means to distinguish between the effect of selective sweeps and background selection. Here, we investigate the relationship between AFS, recombination rates and the density of putatively functional sites in two genome-wide data-sets. We find that the AFS in regions of low recombination and with a higher density of functional sites are more skewed towards rare variants than regions with high recombination rates and few functional sites. The observed variation in the AFS with recombination rates is consistent with selective sweeps. Neither background selection nor selectively neutral phenomena such as changes in demography or biased gene conversion can explain the data fully. Hence, positive selection had a non-negligible impact on genome-wide patterns of variation, but it remains a challenge to quantify the effect.

The role of ecology in the adaptive process

JOACHIM HERMISSON

(joint with Michael Kopp)

Max F. Perutz Laboratories, University of Vienna

My talk treats a classic problem of evolutionary biology. Assume that a population faces a new environmental challenge. What is then the expected distribution of step-sizes (measured as mutational effects on a phenotype)? And what is the order of adaptive steps; i.e. should we expect large steps before small steps or vice-versa? In the past, this topic has usually been studied from a purely genetic point of view, either in the context of clonal competition with fitness as a trait or on the phenotypic level using Fisher's geometrical model. Results for these models, in particular by H. A. Orr, conclude that the distribution of adaptive substitutions should be approximately exponential under very general conditions. It appears, however, that all previous approaches focus on a single ecological scenario, where the population responds to a single sudden change of the environment. Since in nature, environmental change is often gradual, my recent work with Michael Kopp considers the influence of the ecological dynamics on the problem. It turns out that this influence is profound: In particular, we do not usually expect an exponential distribution of adaptive steps.

I will start my talk with an overview of the state of the art and the most important concepts that have emerged in the research since Darwin and Fisher. I will then describe the "moving-optimum" model that introduces an ecological component to the adaptation problem. For the methods, we proceed from biology to mathematics, using ideas from theoretical physics to bridge the gap. Finally, I will briefly discuss our results.

The influence of population size on patterns of natural selection in mammals

CAROLIN KOSIOL

University of Veterinary Medicine, Vienna

There has been some discussion on the effect of population size on selection scans. It has been argued that positive selection as well as purifying selection is less efficient in humans and chimps than rodents because of their smaller population sizes than murids (Keightley et al., 2005; Rhesus Macaque Sequence and Analysis Consortium, 2007).

I have performed a positive selection scan on six mammalian genomes: human, chimpanzee, mouse, rat and dog. We find fewer positive selected genes on the primate clade than on the rodent clade. We try to quantify this effect by estimating ratios of population sizes of the species from the genomic data. The estimation is based on an extension of population genetic interpretation of a branch model of codon substitutions by Halpern and Bruno 1998 (see also Nielsen and Yang, 2003). The estimates we obtain for the ratio of macaque and human effective population sizes are in good agreement with the ratio calculated from polymorphism data (Wall et al., 2003, Hernandez et al, 2007). I will also discuss limitations of the estimation of population ratios from genomes of single individuals.

Evolution of adaptive immune genes in two sympatric stickleback species (*G. aculeatus* & *P. pungitius*)

TOBIAS LENZ

(joint with Christophe Eizaguirre, Martin Kalbe and Manfred Milinski)

Max Planck Institute for Evolutionary Biology, Plön

The Major Histocompatibility Complex (MHC) is part of the adaptive immune system in vertebrates and most of its genes show outstanding polymorphism in terms of allelic and genetic diversity, which has evolved to confer adaptive resistance against pathogens. Interactions between distinct MHC alleles and distinct pathogen taxa have been reported and we therefore assume host populations to differ in their respective MHC allele pool if they are facing different pathogen communities. Three-spined and Nine-spined sticklebacks occur in sympatry in different ecological habitats, which are known to harbour distinct parasite communities. We show that overall parasite communities differ among habitats, but detect a substantial overlap of parasites between host species in the same habitat. With our study we are testing whether local immunogenetic adaptation to different parasite communities leads to convergent evolution in these two sympatric fish species. We will compare phylogenetic relationships between habitats and species on the adaptive level (MHC exons), the neutral level (MHC introns, D-loop) and mitochondrial genes (Cytb, 12S, 16S). We expect the mitochondrial markers to reflect clear species separation, whereas neutral markers are probably more informative on the population level. Immunologically adaptive markers (MHC exons) in turn are expected to contrast with these results and show closer relationship between species than between habitats, reflecting local convergent evolution.

Coalescent simulations: how simulation methods shape the trees

SYLVAIN MOUSSET

CNRS, Laboratoire de Biométrie et Biologie Évolutive, Lyon

(joint with Sabas Ramos-Onsins)

Coalescent simulation is a widely-used tool in population studies of molecular polymorphism. In such simulations a coalescent tree is first produced conditional on a particular demographic model (usually a panmictic constant-size Wright-Fisher population); mutations are then independently spread on this tree conditional on some mutational process, and simulated samples are compared to observed data.

Several simulation methods have been proposed for this independent mutational process, the two most popular being either using a fixed mutation rate or using a fixed number of mutations. With a Bayesian approach, we show that these and other methods can lead to statistical errors by radically changing the shape of the simulated trees, especially in non-recombining regions.

Analyses of genetic introgression as a means to identify genetic differentiation among populations of sculpins

ARNE W. NOLTE

Max-Planck-Institute for Evolutionary Biology, Plön

Our understanding of the genomic basis underlying early evolutionary divergence is still very limited. However, current technological advances will provide a surge of data that may help to understand evolutionary processes in great detail provided that a strong interpretative framework exists. Genes that affect divergence between nascent species can be expected to be subject to natural selection and hence affect genomic isolation between divergent populations. This may be due to intrinsic genetic incompatibilities or extrinsic ecological factors. Therefore, an understanding of functional variation may require analyses of evolutionary processes *in situ* to include environmental factors. Here, the utility of secondary contact zones to identify functional differentiation in nature is assessed. Genetic introgression between different species of sculpins (*Cottus*) was analyzed using microsatellites and SNPs. The analysis tests for deviations from neutral patterns of introgression at individual loci based on expectations given genome-wide levels of admixture. Current analyses reveal conspicuous patterns of introgression that suggest that widespread selection affects different genomic fragments in different ways. Deviations suggest that different loci may be subject to positive selection. In the case of heterotypic genotypes, signs for overdominance or underdominance were detected. Selection is so common in the study system that it is difficult to define an appropriate null model. Although the results provide a good basis to identify loci that cause evolutionary divergence, it is also clear that the analysis could be improved by refining theoretical expectations and analytical approaches.

Primer for Discussion on *Speciation models*

PETER PFAFFELHUBER

Zentrum für Biosystemanalyse, Universität Freiburg

Traditionally, microevolution modelled through population genetics is a field separated from phylogenetics. It is common to both fields to think about relationships of individuals, best describes by genealogical trees. Since a few years the two fields become closer thanks to models (e.g. introduced by Wakeley and Hey (1997)) giving a good connection between phylogenetic and coalescent trees. The punchline of these models is that the species coalescent tree is embedded into the phylogenetic tree. Such scenarios are most often studied with only two recent species which speciated a short time ago, such that lineages in the coalescent tree might not follow the species tree.

By their nature, such models are not capable of explaining how speciation occurs, but they allow for a fine description of patterns of sequence variation. However, the theoretical analysis of these models has just started, e.g. only neutral evolution since the time of speciation is considered so far. We will discuss possible theoretical results which can help to analyse sequence variation data.

Underdominance and Population Transformations

FLOYD REED

Max Planck Institute for Evolutionary Biology, Plön

There is growing interest in methods to “drive” genes to high frequency in wild populations (i.e. to genetically transform entire populations), particularly with regard to human “pest” species (e.g. transforming wild mosquitoes to be resistant to *Plasmodium* and thus prevent malaria in humans). Genetic systems that can drive to a high frequency in populations are also of central interest for their possible natural evolutionary roles. I will describe part of our research program where we are focused on a unique type of gene drive, underdominance, which has many useful properties. In our program we are also attempting to “mind the gap” and simultaneously develop a theoretical understanding of the predictions of underdominance and empirically screen for naturally occurring underdominance alleles, but the latter is very much a work in progress. Here I will focus on our theoretical results; particularly the conditions for stable population transformations in the presence of wildtype immigrants. Then I will briefly describe our plan to detect loci affected by underdominant selection in wild populations of *Drosophila melanogaster*.

Massively parallel sequencing of pooled DNA samples-the next generation of molecular markers

CHRISTIAN SCHLÖTTERER

(joint with A. Futschik)

University of Veterinary Medicine, Vienna

Next generation sequencing (NGS) is about to revolutionize genetic analysis. NGS is currently mainly used to sequence individual genomes. Due to the high sequence coverage required, the costs for population scale analyses are still too high to allow an extension to non-model organisms. Here, we show that NGS of pools of individuals is more effective in SNP discovery and provides more accurate allele frequency estimates. Using simple adjustments it is possible to account for sequencing errors, without severely compromising the advantage of the pooling approach. We conclude that NGS of pools of individuals is a cost effective approach to estimate allele frequencies on a genome-wide scale. It is anticipated that NGS will replace classic surveys of molecular markers.

Sex, evolution and agriculture

TIM SHARBEL

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung, IPK, Gatersleben

Apomixis, or asexual reproduction through seed, is a naturally occurring reproductive form which has been observed in more than 400 plant species (not to mention a large number of animals). Naturally occurring apomictic taxa provide important tools to study the evolution, molecular genetic mechanisms, environmental constraints and consequences, and population-level variability of this form of reproduction. Furthermore, it represents a potentially important agricultural tool, since introduction of apomixis into crops could be an effective way to fix and propagate a given genotype. We are studying naturally-occurring variation in male and female gamete formation in wild populations of the *Boechera holboellii* complex, a wild relative of *Arabidopsis* which includes both sexual and apomictic reproductive pathways.

In short, our research aims at identifying genes associated with apomixis by performing high-throughput “omics” analyses on sexual and apomictic ovules. Comparisons of such datasets have demonstrated thousands of differentially-expressed genes, and thus while we are closer to understanding the mechanisms behind asexual seed formation, we are nonetheless at a crossroad whereby evolutionary theory is required to differentiate signal (candidate genes or mechanisms) from noise (technical and biological). My presentation will thus outline the high levels of variability exhibited by asexual organisms, from the population level through to that of the pollen and egg cell, with a focus on how hypotheses on the evolution of sex can help us make sense of it all.

The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations

THOMAS STÄDLER

(joint with Bernhard Haubold, Wolfgang Stephan and Peter Pfaffelhuber)

Institute of Integrative Biology, ETH Zurich

Our contribution resulted from a successful cross-talk (or so we prefer to think) between a theory-conscious empiricist and several theoreticians, where the initial impetus arose as the empiricist struggled to explain his data in light of the prevalent theoretical expectations. In this presentation, I shall first recount the main features of our multi-locus sequence data on wild tomatoes (that also happen to characterize SNP data sets in humans, *Drosophila*, and plants) that provided the main motivation for our study implementing coalescent simulations of subdivided populations: pooled and geographically scattered samples show an excess of low-frequency polymorphisms (as summarized by Tajima's D_T and/or Fu and Li's D_{FL} statistics), compared to geographically local samples. We studied the impact of three different sampling schemes on patterns of neutral diversity: local (all sequences are from one deme), pooled (samples from 4 demes treated as a single entity), and scattered (single sequences from each of many demes). Specifically, we were interested in the two D statistics based on the site frequency spectrum as a function of migration rate, demographic history of the entire population (including timing and magnitude of species-wide expansions), and the sampling scheme. Using simulations implementing both finite-island and two-dimensional stepping-stone spatial structure, we found strong effects of the sampling scheme on D_T and D_{FL} , especially under species-wide (range) expansions. Pooled samples yield average D_T and D_{FL} values that are generally intermediate between those of local and scattered samples. Under species-wide expansion scenarios, these effects of spatial sampling may persist up to very high levels of gene flow ($Nm > 25$), implying that local samples cannot be regarded as being drawn from a panmictic population. Our simulation results capture the genealogical features and effects of sampling scheme seen in many empirical data sets. Complementary to our findings, Jost (2008, 2009) has convincingly shown that F_{ST} does not measure differentiation and that both the number of demes and the mutation rate, jointly with migration rate, determine levels of differentiation in the finite-island model. These results have broad implications for empirical studies aimed at distinguishing effects of demographic history from those of natural selection. They also suggest that hardly any species can be regarded as being panmictic and that a new generation of inference schemes needs to be developed to do justice to these inconvenient facts.

The emergence of a new gene from an intergenic region in the house mouse lineage is associated with a selective sweep

FABIAN STAUBACH

Max-Planck-Institute for Evolutionary Biology, Plön

It has long been thought that new genes arise only by gene duplication mechanisms, since the signals for regulation and transcript processing would be unlikely to evolve in parallel with a new gene function. In *Drosophila*, such newly evolved genes are often associated with signs of positive selection. We have identified here a transcript (*Poldi*) in the house mouse (*M. musculus*) that has arisen within a large intergenic region. Extensive population genetic sequence analysis of the surrounding region in natural populations of the house mouse provides evidence that *Poldi* has experienced a recent selective sweep in *M. m. musculus*. This region is conserved in many mammals, including humans, i.e. transposons or genome-rearrangements were not involved in generating this novel transcript. *Poldi* contains three exons, shows alternative splicing and is specifically expressed in postmeiotic cells of the testis (shown by in situ hybridization). A specific knock-out of the entire gene region in laboratory mice results in reduced testis weight. Our results suggest that preexisting signals for transcript regulation and processing are present in intergenic regions and can become the basis for the evolution of a new gene, which as soon as it is expressed, can become a target of natural selection and gain functional importance as presented here.

Coalescence methods to estimate ecological parameters of seed bank and metapopulation structure

AURÉLIEN TELLIER

(joint with Stefan Laurent and Wolfgang Stephan)

Biocenter, University of Munich

Wild tomato species, which originated in western South America and the Galapagos Islands, are found in a wide range of habitats, and have thus to cope with various abiotic (e.g. temperature fluctuations, drought) and biotic stresses (e.g. attack by pathogens and herbivores). In order to study molecular signature of adaptation, we develop here first a demographic model taking into account the life history traits of such species, namely seed bank, spatial structure of populations and range expansion. Our aim is to explain the discrepancies between the high effective population size inferred from genetic data and the very small census population sizes observed in nature. Metapopulation structure with restricted migration among demes and seed banks are two well known mechanisms increasing effective population size. Here we estimate parameters of seed banks (germination rate and maximum life expectancy of seeds) and of metapopulation (migration rates) using an Approximate Bayesian Computation framework. We also investigate how bias in the unknown total number of demes in the metapopulation affects these estimates. Finally, we show how different models of metapopulation structure can affect the estimates of the seed bank parameters.

Associations between testis gene expression variation and male sterility in a house mouse hybrid zone

LESLIE M. TURNER

(joint with Bettina Harr)

Max Planck Institute for Evolutionary Biology, Plön

Reduced hybrid fertility is a common feature of reproductive barriers separating recently diverged species, thus identifying the genetic mechanisms underlying hybrid sterility is key to understanding the molecular basis of speciation. Recent studies demonstrate that misexpression of genes is common in hybrids of several taxa and is associated with hybrid male sterility in *Drosophila*, suggesting divergence in gene regulation plays an important role in reproductive isolation. House mice (*Mus musculus*) are an excellent system for investigating genetic causes of reproductive isolation because of the wealth of genetic tools available for this model organism and the existence of a natural hybrid zone between two subspecies. In laboratory crosses of these subspecies, F1 males are sometimes sterile, suggesting reduced hybrid fertility is a major reproductive barrier maintaining subspecies distinctness. Yet little is known about the specific phenotypes, prevalence, or genetic basis of male sterility in the hybrid zone. We investigate the contribution of divergence in testis gene expression to hybrid male sterility in natural populations of house mice. First, using microarrays, we characterized genome-wide expression patterns in testis of 26 males captured in the *M. m. musculus*-*M. m. domesticus* hybrid zone in southern Germany. We identify several hundred candidate genes which are either 1) differentially expressed between the subspecies and show a sharp cline in expression level across the hybrid zone or 2) are misexpressed in hybrids. Second, we measured phenotypes associated with fertility (testis weight, sperm count, sperm motility, sperm morphology) in 100 first generation lab-born offspring of wild captured individuals from the hybrid zone. We report evidence for transgression (larger variance in hybrids) in these fertility traits and identify multiple distinct phenotypes likely indicative of sterility. Finally, we find associations between these phenotypes and expression levels of several candidate genes identified from the microarray data. Together, these results make an important contribution to characterizing both the phenotypic and genotypic basis of a reproductive isolating barrier acting in nature.

Non-linear dynamics of non-synonymous (dN) and synonymous (dS) substitution rates affects inference of selection

JOCHEN B. W. WOLF

(joint with Axel Künstner, Kiwoong Nam, Mattias Jakobsson, Hans Ellegren)

Department of Evolutionary Biology, Uppsala University

Selection modulates gene-sequence-evolution in different ways, by constraining potential changes of amino acid sequences (purifying selection) or by favouring new and adaptive genetic variants (positive selection). The number of non-synonymous differences in a pair of protein-coding sequences can be used to quantify the mode and strength of selection. To control for regional variation in substitution rates the proportionate number of non-synonymous differences (dN) is divided by the proportionate number of synonymous differences (dS). The resulting ratio (dN/dS) is a widely used indicator for functional divergence to identify particular genes that underwent positive selection. With the ever growing amount of genome data, summary statistics like mean dN/dS allow gathering information on the mode of evolution for entire species. Both applications hinge on the assumption that dS and mean dS (branch length) are neutral and adequately control for variation in substitution rates across genes and across organisms respectively. We here explore the validity of this assumption using empirical data based on whole genome protein sequence alignments between human and 15 other vertebrate species and several simulation approaches. We find that dN/dS does not appropriately reflect the action of selection, as it is strongly influenced by its denominator (dS). Particularly for closely related taxa, such as human and chimpanzee, dN/dS can be misleading and is not an unadulterated indicator of selection. Instead, we suggest that inconsistencies in the behaviour of dN/dS are to be expected and highlight the idea that this behaviour may be inherent to taking the ratio of two randomly distributed variables that are non-linearly correlated. New null hypotheses will be needed to adequately handle these non-linear dynamics.

List of participants

Name	Address	Email
Baake, Ellen	University of Bielefeld	ebaake@techfak.uni-bielefeld.de
Baumdicker, Franz	Zentrum für Biosystemanalyse, University of Freiburg	franz.baumdicker@gmx.de
Birkner, Matthias	University of Mainz	birkner@biologie.uni-muenchen.de
Borck, Cornelia	Zentrum für Biosystemanalyse, University of Freiburg	Cornelia.Borck@gmx.de
Bryk, Jarek	Max-Planck-Institute for Evolutionary Biology, Plön	bryk@evolbio.mpg.de
Caliebe, Amke	Institut für Medizinische Infor- matik und Statistik, University of Kiel	caliebe@medinfo.uni-kiel.de
Depperschmidt, Andrej	Zentrum für Biosystemanalyse, University of Freiburg	depperschmidt@stochastik.uni-freiburg.de
Ewing, Greg	Max F. Perutz Laboratories, Vienna	gregory.ewing@univie.ac.at
Hardouin, Emilie	Max-Planck-Institute for Evolutionary Biology, Plön	hardouin@evolbio.mpg.de
Haubold, Bernhard	Max-Planck-Institute for Evolutionary Biology, Plön	haubold@evolbio.mpg.de
Heitlinger, Emanuel	Universities of Karlsruhe und Edinburgh	emanuelheitlinger@googlemail.com
Hellmann, Ines	Max F. Perutz Laboratories, Vienna	ines.hellmann@univie.ac.at
Hermisson, Joachim	Max F. Perutz Laboratories, Vienna	joachim.hermisson@univie.ac.at

Name	Address	Email
Hustedt, Thiemo	University of Bielefeld	thustedt@gmx.de
Hutzenthaler, Martin	Biocenter, University of Munich	hutzenth@math.uni-frankfurt.de
Klassmann, Alexander	Institut für Genetik, University of Cologne	alex.klassmann@koeln.de
Klauke, Nadine	Faculty of Biology, University of Freiburg	nadine.klauke@merkur.uni-freiburg.de
Kosiol, Carolin	University of Veterinary Medicine, Vienna	ckosiol@gmail.com
Lang, Daniel	Faculty of Biology, University of Freiburg	daniel.lang@biologie.uni-freiburg.de
Lenz, Tobias	Max-Planck-Institute for Evolutionary Biology, Plön	lenz@evolbio.mpg.de
Metzler, Dirk	Biocenter, University of Munich	metzler@biologie.uni-muenchen.de
Mousset, Sylvain	CNRS, Laboratoire de Biométrie et Biologie Évolutive, Lyon	mousset@biomserv.univ-lyon1.fr
Nadu, Lisha	Biocenter, University of Munich	LishaNadu@biologie.uni-muenchen.de
Nolte, Arne	Max-Planck-Institute for Evolutionary Biology, Plön	nolte@evolbio.mpg.de
Pardoux, Etienne	Centre de Mathématiques et d'Informatique, Université de Provence, Marseille	pardoux@cmi.univ-mrs.fr
Pennings, Pleuni	Biocenter, University of Munich	pennings@zi.biologie.uni-muenchen.de

Name	Address	Email
Pfaffelhuber, Peter	Zentrum für Biosystemanalyse, University of Freiburg	peter.pfaffelhuber@stochastik.uni-freiburg.de
Reed, Floyd	Max-Planck-Institute for Evolutionary Biology, Plön	reed@evolbio.mpg.de
Rensing, Stefan	Faculty of Biology, University of Freiburg	stefan.rensing@biologie.uni-freiburg.de
Rudolf, Johannes	Abteilung Mathematische Stochastik, University of Freiburg	rudolf.johannes@googlemail.com
Rüschendorf, Ludger	Abteilung Mathematische Stochastik, University of Freiburg	ruschen@stochastik.uni-freiburg.de
Schlötterer, Christian	University of Veterinary Medicine, Vienna	christian.schloetterer@vu-wien.ac.at
Sharbel, Tim	Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben	sharbel@ipk-gatersleben.de
Staab, Paul	Abteilung Mathematische Stochastik, University of Freiburg	paul.staab@googlemail.com
Städler, Thomas	Institute of Integrative Biology, ETH Zurich	thomas.staedler@env.ethz.ch
Stanciu, Julia	Fachbereich Informatik und Mathematik, University of Frankfurt	
Staubach, Fabian	Max-Planck-Institute for Evolutionary Biology, Plön	staubach@evolbio.mpg.de
Stephan, Wolfgang	Biocenter, University of Munich	stephan@zi.biologie.uni-muenchen.de
Surovcik, Katharina	Zentrum für Biosystemanalyse, University of Freiburg	katharina.surovcik@stochastik.uni-freiburg.de

Name	Address	Email
Tellier, Aurelien	Biocenter, University of Munich	tellier@zi.biologie.uni-muenchen.de
Turner, Leslie	Max-Planck-Institute for Evolutionary Biology, Plön	turner@evolbio.mpg.de
Ücker, Hildegard	Max F. Perutz Laboratories, Vienna	hildegard.uecker@univie.ac.at
Vogl, Claus	University of Veterinary Medicine, Vienna	claus.vogl@vu-wien.ac.at
Wakolbinger, Anton	Fachbereich Informatik und Mathematik University of Frankfurt	wakolbin@math.uni-frankfurt.de
von Wangenheim, Ute	University of Bielefeld	bandwurm_ute@gmx.net
Weisshaupt, Heinz	Zentrum für Biosystemanalyse, University of Freiburg	heinz.weisshaupt@univie.ac.at
Wiehe, Thomas	Institut für Genetik, University of Cologne	twiehe@uni-koeln.de
Wolf, Jochen	Department of Evolutionary Biology, Uppsala University	wolf@evolbio.mpg.de
Wollstein, Andreas	Institut für Genetik, University of Cologne	andreas.wollstein@uni-koeln.de
Zivkovic, Daniel	Biocenter, University of Munich	daniel.zivkovic@gmx.at