Science Summit 2023 September 20

Genomics of Biodiversity and Evolution
Aula Magna, Stockholm

SciLifeLab

Contents

Modelling insect diversity across Sweden with environmental DNA and deep learning
biodiversity, modelling, deep learning, GIS, remote sensing

Adrian Baggström, Biodiversity Data Lab (Uppsala University) Sweden. Tobias Andermann Biodiversity Data Lab (Uppsala University) Sweden.

Despite being in the middle of a global biodiversity crisis, we still have comparably little knowledge of the spatial distribution of biodiversity for most organism groups. Such knowledge is crucial in making informed conservation priority decisions. Here we present a project where we develop deep learning biodiversity modelling tools that can in theory predict the expected species diversity of any organism group, given a set of publicly available geospatial data-products. We train the model on biodiversity data of insects derived from a Sweden-wide environmental DNA (eDNA) inventory, produced by the Insect Biome Atlas. We utilize unique DNA barcode sequences derived from over 5000 eDNA samples collected from 200 sites distributed across Sweden. By combining this data with spatial information such as temperature, precipitation, elevation, ground cover classification, NDVI, human impact indices etc., we can train a convolutional neural network to predict the expected number of insects at any given location.

A generalized benchmark for all types of enrichment analysis methods
systems biology, pathway enrichment analysis, functional enrichment, protein network, functional association network

Davide Buzzao, Department of Biochemistry and Biophysics, Stockholm University, Sweden. Miguel Castresana-Aguirre, K7 Department of Oncology-Pathology, Karolinska Institutet, Sweden. Dimitri Guala, Department of Biochemistry and Biophysics, Stockholm University, Sweden. Erik L.L. Sonnhammer, Department of Biochemistry and Biophysics, Stockholm University, Sweden.

Enrichment analysis (EA) is the standard approach to functionally characterize omics results. As a consequence, a large number of EA methods have been developed, yet it is unclear which method is the best for a given dataset. The main issues with previous benchmarks include the complexity of correctly assigning true pathways to a test dataset, and lack of generality of the evaluation metrics, for which the rank of a single target pathway is commonly used.

We here present a generalized EA benchmark of all four categories of currently existing approaches. In order to address the shortcomings of the single target pathway approach and to enhance the sensitivity evaluation, we introduced a Disease Pathway Network based on functional associations between KEGG pathways. By making the assumption that no pathway should be enriched in a randomized gene expression dataset, we measured the specificity of the methods in an independent manner, and determined whether any methods might be biased towards certain pathways.

Genetic parallelism and adaptation to brackish water bodies in sprat (Sprattus sprattus) and Atlantic herring (Clupea harengus)
genetic parallelism, salinity adaptation

Qiaoling Deng, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden.

Mats E. Pettersson, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden. María Quintela, Dept of Population Genetics, Institute of Marine Research, Norway. François Besnier, Dept of Population Genetics, Institute of Marine Research, Norway. Andreas Wallberg, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden. Kevin Glover, Dept of Population Genetics, Institute of Marine Research, Norway. Leif Andersson, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden.

The similar evolutionary pressures faced by closely related species that colonize a new environment provides an opportunity to explore the mechanisms underlying genetic parallelism and adaptation, an essential knowledge for understanding evolution and the maintenance of biodiversity. The European sprat (Sprattus sprattus) and Atlantic herring (Clupea harengus) are two clupeids that have both, independently, colonized and adapted to brackish waters, in particular the Baltic Sea, providing an ideal situation for studying genetic parallelism in the context of salinity-related adaptation.

Here, we have utilized a draft assembly of the sprat genome and whole-genome pool-seq data from 19 population samples of sprat, spanning from the Baltic to the Black Sea, to detect strong signals of genetic adaptation to brackish waters, including six putative inversions with fixed, or nearly fixed, differences between brackish and oceanic populations. We have intersected these signals with previously identified signals of adaptation in C. harengus, and are able to identify three narrow signals, covering single genes, or in one case a small gene cluster, that overlap between the two species. These genes – PRLRA, THRB and TNNI2 – constitute strong candidates for parallelism, with regard to genetic adaptation to low salinity, between the two species. Our results reinforce the importance of the genes in question, and thus shed light on the genomic basis of salinity adaptation.

Inferring gene regulatory networks - exploration of bioregulation upon perturbation

GRN, RNA-seq, transcritomics, single-cell

Mateusz Garbulowski, Department of Biochemistry and Biophysics, Stockholm University, Sweden. Thomas Hillerton, Department of Biochemistry and Biophysics, Stockholm University, Sweden. Claudia Kutter, Department of Microbiology, Tumor, and Cell Biology, Karolinska Institutet, Sweden. Erik L. L. Sonnhammer, Department of Biochemistry and Biophysics, Stockholm University, Sweden.

Genes tend to interact with each other to initiate or stop cell functions. To find regulator-target relationships, perturbation screening of a transcriptome can be performed. However, as the real data lack the ground truth gene regulatory network (GRN), synthetic-based modeling is highly beneficial for benchmarking inferred GRNs. To tackle this issue, we developed tools such as GeneSNAKE and GeneSPIDER that use state-of-the-art methods for generation and simulation of bulk-like perturbations. Recently we implemented GeneSPIDER2 that can simulate single-cell-like perturbations for large GRNs.

Apart from simulations, we show that our tools can be applied for GRN inference from real data using two examples. First, we explored the regulome of hepatocellular carcinoma (HepG2) using shRNA-seq experiments. Second, we inferred cluster-specific GRNs from single-cell Perturb-seq experiments of SARS-CoV-2 (Calu-3). Overall, our research aims at constructing bulk- and single-cell-based GRNs that are supported by simulations and benchmarking.

Which field work strategy is more efficient on capturing fungal biodiversity?
eDNA, fungi, field work

Monica Guilera Recoder, Uppsala University, Sweden.

The current biodiversity crisis brings us the necessity to increase our knowledge on species biodiversity of those groups which still remain mostly undescribed. High-throughput sequencing of enviornmental DNA (eDNA) has shown promising results for the detection of the hidden diversity of undescribed cryptic species. In this study we investigate how different eDNA sampling techniques impact species detection. We focus on targeting fungi using soil eDNA, deadwood and insect malaise traps. We are testing different variables such as how deep to sample the soil, which wood decay stages to target, whether or not to pool multiple samples, and how frequently we need to sample to effectively capture the species community. We will focus on a forest site near Uppsala with a rich species inventory record, allowing benchmarking of the methods against known species records. The sampling is performed twice in the year to capture seasonal variations of diversity.

We're all holobiointes
microbiome, host-microbe interaction, gut-brain axis, sex hormones

Luisa W. Hugerth, Imbim, Uppsala University, Sweden.

Humans harbour complex microbial communities, with whom we share a long evolutionary history. Microbiome acquisition begins at birth, and is strongly shaped by the mother, but also by other people, animals and the surrounding environment. Much of the ancestral diversity of the microbiome has been lost with urbanization, industrialization and medicalization. In parallel with this, many non-infectious diseases have gained prevalence. Previous research lifts the microbiome as modifiable risk factor for these; however, neither the causal route between microbiome and disease, nor the intrinsic dynamics of the microbiome are sufficiently known. Host factors, such as sex hormones, shape the environment the microbiome live in, e.g. by modulating mucosal structure and composition. This makes the microbiome of women specially challenging to study. Still, women are key to shaping the microbiome of coming generations. Therefore, we aim to study the interplay between sex hormones, microbiome and health in pregnant and non-pregnant women.

Reconstructing the last meal of an extinct steppe bison living 50,000 years ago
metagenomic, aDNA, gut, steppe bison

Chenyu Jin, Adrian Forsythe, Miklós Bálint, Frank Kienast, Katerina Guschanski, Tom van der Valk.

During the last cold stage containing MIS 3, Arctic Eurasia exhibited a distinctive plant composition consisting of steppe and xerophilous arctic plants. This combination formed the Pleistocene tundra-steppe, which provided a habitat for large herbivores like mammoths and bison. In this study, we capitalize on a unique opportunity to analyze a plant paleo-community of the Arctic mammoth steppe through the well-preserved gut content of a steppe bison (Bison priscus) discovered in the Anyuy River, Russia.

Radiocarbon datings of its tissues and the sedimentary context indicate that this steppe bison lived in the range of 48000 and 50000 14C YBP, during the last cold stage (MIS 3). Macrofossil analysis of this sample reveals a floral composition resembling that of a

saline meadow, likely resulting from increased aridity due to a more northern coastal line compared to the present. Additionally, the presence of larch (Larix gmelinii) suggests the existence of a hidden tree refugium in Northern Siberia, which may have facilitated the rapid recolonization. By employing ancient DNA, we anticipate not only to validate and enhance taxonomic identification with greater resolution than macrofossils, but also to enable population genetics research of plants (e.g., larch) that could help evaluate the hypothesis of founder effects originating from a refugium.

Fish recombination landscapes
recombination, linkage maps, comparative genomics, fish

Kivioja Teemu, University of Helsinki, Finland. Rastas Pasi, University of Helsinki, Finland.

Meiotic recombination is a common aspect of eukaryotic life. Even though each pair of homologous chromosomes typically has one or few recombination events per meiosis, their numbers and locations vary both within and between species. Fish are particularly interesting because females have a higher overall recombination rate than males in many but not all species. Numerous mechanistic and evolutionary hypotheses have been suggested to explain the observed variation. Comparing different explanations requires high-quality estimates of local recombination rates along the chromosomes across many species but such resource of comparable recombination landscapes has not been available. Thus, we have generated high-resolution sex-specific linkage maps for 35 fish species by processing publicly available pedigree data from different studies through the same computational pipeline. The set allows for the first time quantitative comparison of recombination landscapes across both closely related and distant species.

Exploring the possibilities of population metagenomics on Swedish insects
metagenomics, population genetics, bioinformatics, biodiversity, arthropoda

Samantha López Clinton, Swedish Museum of Natural History, Centre for Palaeogenetics, Stockholm University, Sweden. Chenyu Jin, Swedish Museum of Natural History, Centre for Palaeogenetics, Stockholm University, Sweden. Andreia Miraldo, Swedish Museum of Natural History, Sweden. Elzbieta Iwaszkiewicz-Eggebrecht, Swedish Museum of Natural History, Sweden. Fredrik Ronquist, Swedish Museum of Natural History, Sweden. Tom van der Valk, Swedish Museum of Natural History, Centre for Palaeogenetics, Sweden.

Biological and genetic diversity monitoring, including population genetics measures, are crucial for conservation as they provide information on the status of and threats to species and ecosystems. Here, we explore the use of terabyte-scale databases, taxonomic classification software and subsequent sequence alignments on genome-wide shotgun sequencing data from mixed arthropod samples collected across Sweden by the Insect Biome Atlas Project. We evaluate the strengths and limitations of genome-wide approach to assess, not just species composition, but also gene flow, genetic diversity, phylogeography and additional population genomics measures for populations sampled across different locations and gradients. By exploring the use of innovative metagenomics bioinformatic methods applied on mixed and environmental DNA samples, we contribute to the accessibility and effectiveness of genomic based conservation strategies, and set an important precedent for metagenomic-scale population genomics.

The SciLifeLab Ancient DNA unit - a community resource
ancient DNA, genomics, DNA sequencing, archaeology, biodiversity

Magnus Lundgren, SciLifeLab Ancient DNA unit, Sweden. E-Jean Tan, SciLifeLab Ancient DNA unit, Sweden. Thijessen Naidoo, SciLifeLab Ancient DNA unit, Sweden. Feizabadi Farahani, Motahare, SciLifeLab Ancient DNA unit, Sweden

Ancient DNA provides a wealth of information on the origins, development and history of humans and other organisms. SciLifeLab Ancient DNA unit is a core facility based at Stockholm University and Uppsala University. The unit has state-of-the art clean-room laboratories for ancient DNA processing, and staff with expertise in laboratory work and computational analysis. We analyze a range of different type of samples and perform extraction of human, animal and sediment DNA for both screening and high coverage DNA data production. As a service to researchers in Sweden and abroad, we perform standardized and custom data analysis with regard to sex, mitochondrial and Y-chromosome haplogroup, kinship, pathogens and demography. You are welcome to contact us for further information and project discussions.

Mapping Moose Short-Read Sequences on the Bovine Genome – a Tool to Investigate the White Moose
moose, genome sequencing, coat color, biodiversity

Iryna Shutava, SLU. Theodor Kalbfleisch, University of Kentucky, Lexington. Dipanwita Paul, SLU. Maya Baghdy Sar, Uppsala University. Baptiste Herlemont, SLU. Tomas Bergström, SLU. Sofia Mikko SLU.

Scandinavian Moose (Alces alces alces) has a wildtype agouti coat color. White individuals are seen in some areas of Scandinavia and Canada, but the genetic basis is unknown. Two moose genomes are publicly available on scaffold level, none of them yet annotated. One is derived from a Scandinavian moose, and one is a compilation of three North American sub-species (A.a.shirasi, A.a.americana and A.a.gigas). We sequenced the genome of a white male moose and mapped this to the Bovine genome. Raw moose data mapped at ~82% to the Bovine genome. We filtered for an autosomal recessive inheritance of white coat color in the moose, and annotated variants using the bovine genome as a reference. We detected 2,914 missense, 100 nonsense, and 313 frame-shift putative variants. Further filtering detected 39 missense variants within known pigmentation genes. Using the same filtrations, we detected 1456 putative SVs, of which 29 within known pigmentation genes.

Discovery of a Mega-NUMT in Drosophila paulistorum
NUMT, mtDNA, genomics, long read sequencing, speciation

Merce Montoliu Nerin, Dept of Cell and Molecular Biology, Uppsala University, Sweden. Anton Strunov, Dept of Cell and Developmental Biology, Medical University of Vienna, Austria. Wolfgang Miller, Dept of Cell and Developmental Biology, Medical University of Vienna, Austria. Lisa Klasson, Dept of Cell and Molecular Biology, Uppsala University, Sweden.

Mitochondrial DNA (mtDNA) has been widely used as a molecular marker in animals, essential to resolve taxonomic uncertainties. The development of genomic methods has revealed the existence of mitochondrial genome insertions in the nucleus (NUMTs). Differences in copy number of NUMTs vs mtDNA would suffice to resolve the correct

identification of mtDNA into NUMT, but a NUMT in high copy number (Mega-NUMT), could lead to incorrect inferences of heteroplasmy.
The neotropical fly Drosophila paulistorum is a superspecies undergoing sympatric speciation. Studies have shown incongruencies between nuclear and mtDNA phylogenies, suggesting introgression events. Moreover, mitochondrial genomes split in two polyphyletic clades, alpha and beta. We analysed the genome of one semispecies containing both mitotypes. Using long read data we were able to confirm that the alpha mitotype is a Mega-NUMT of around 60 copies, with a possible structural function as telomere.

Genomic insights into the interconnected Demographic History between Sheep and Humans in Iberia
ancient DNA, sheep, demographic history, bell beaker, iberia

Pedro Morell Miranda, M. Ángeles Galindo-Pellicena. Paolo Mereu. Juan Antonio Antolinos. José-Miguel Carretero. Juan Luis Arsuaga. Ignacio de Gaspar. Colin Smith. Mario Barbato. Eneko Iriarte. Cristina Valdiosera. Torsten Günther.

Animal domestication was crucial in prehistory, with sheep as the oldest livestock species. As Neolithic humans spread, sheep were brought along, and their remains can be used to trace them. However, the impact of later migrations on sheep populations is less well understood. We sequenced 11 sheep genomes from El Portalón, a cave in Northern Iberia with a rich record of remains from the Neolithic to the Roman period. We found significant changes in sheep's gene-pool that corresponded with cultural changes and migrations in humans. We observed an increase in Eastern ancestry in Iberian sheep around the appearance of the Bell Beaker phenomenon, which coincided with a shift towards milk and wool production. Further shifts were observed during the later periods, and we also found an increase in genetic diversity over time matching human populations. These findings demonstrate the complex and interconnected nature of the demographic history of sheep and humans.

Scoring genomes with phylogenetically informative microRNAs
microRNA, ancient sRNA, genome assembly

Vanessa M. Paynter, UiT The Arctic University of Norway, Tromsø, Norway. Sinan Ugur Umu, Department of Pathology, Institute of Clinical Medicine, University of Oslo, Norway. Kevin J. Peterson, Department of Biological Sciences, Dartmouth College, Hanover, USA. Bastian Fromm, UiT The Arctic University of Norway, Tromsø, Norway.

MicroRNAs are highly conserved, evolutionarily hierarchical short non-coding RNAs. They function as regulators of gene expression influencing development and disease and are valuable taxonomical and phylogenetic markers. We overcame the disparity between increasing availability of genomes and their respective microRNA complement by developing the tool MirMachine, enabling accurate annotation of conserved microRNAs from genomes directly. It has been successfully applied to various animal species, including extinct species where small RNA sequencing is challenging. Here we present an expansion to MirMachine: the microRNA score that quantifies conserved microRNAs in relation to those expected based on the species' phylogenetic position. We analyzed over 1000 animal genomes from Ensembl and found good correlation with popular genome statistics like N50 and BUSCO, while being on average 6 times faster and computationally efficient. It is highly relevant in the field of genomics, facilitating genome reconstruction by indicating coverage which is especially useful for extinct genomes.

Evolutionary constraint identifies candidate non-coding drivers in glioblastoma
cancer, evolutionary constraint, regulatory mutations, glioblastoma

Pensch Raphaela, Uppsala University, Sweden; Roy Ananya, Uppsala University, Sweden; Sakthikumar Sharadha, Uppsala University, Sweden; Kozyrev Sergey, Uppsala University, Sweden; Mäkeläinen Suvi, Uppsala University, Sweden; Jansson Ylva, Uppsala University, Sweden; Lindblad-Toh Kerstin, Uppsala University, Sweden; Forsberg-Nilsson Karin, Uppsala University, Sweden

Glioblastoma (GBM) is the most common malignant brain tumor in adults and is characterized by poor prognosis and a lack of effective treatment options. Non-coding mutations drive cancer by targeting regulatory elements, however, their role in GBM remains poorly understood. We apply the Zoonomia evolutionary constraint scores from 240 mammals to whole-genome sequencing data of 136 GBM patients to predict the functional impact of non-coding mutations. We observe an enrichment of non-coding constraint mutations (NCCMs) in the regulatory regions of both novel and cancer genes. The genes with the strongest enrichment are DLX5 and DLX6 and NCCMs around them as well as SHFM1 and SSTR1 are proposed as novel candidate non-coding drivers. We conclude that evolutionary constraint is a powerful tool to identify non-coding mutations that have the potential to drive GBM. Their discovery could lead to more robust patient stratification and better treatment in the future.

Structural variation evolution during hybridisation in wheatears
transposable elements, structural variants, hybridisation, birds, genomics

Valentina Peona, Swedish Museum of Natural History, Stockholm, Sweden. Dave Lutgen, Vogelwarte, Switzerland. Reto Burri, Vogelwarte, Switzerland.

Structural variants (SVs) are getting attention as facilitators of the evolution of genetic incompatibilities and eventually of reproductive barriers. Here, we aim to investigate how hybridisation itself can trigger the generation of new SVs by causing a partially uncontrolled TE activity in hybrids. We hypothesize that the uncoupling of active TEs and their repressors caused by hybridisation and recombination can escalate genetic conflicts in two ways: 1) increased TE activity can cause more insertions and the origin of new incompatibility loci; 2) new insertions can act as substrate for further SVs through non-allelic homologous recombination between these repeats. This hypothesis predicts SVs and TEs to be a source of incompatibilities, then we should see that regions resilient to gene flow are denser in SVs and that SVs are preferentially induced by TEs. We test this hypothesis using a phased genomic dataset of 300 wheatears that hybridise in multiple contact zones.

Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample
genomics

Marcela Sandoval-Velasco, Center for Evolutionary Hologenomics, University of Copenhagen, Denmark. Olga Dudchenko, The Center for Genome Architecture, Baylor College of Medicine, USA. Juan Antonio Rodríguez, Center for Evolutionary Hologenomics, University of Copenhagen, Denmark. Cynthia Pérez Estrada, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Marianne Dehasque, Centre for Palaeogenetics, Stockholm, Sweden. Claudia Fontsere, Center for Evolutionary Hologenomics, University of Copenhagen, Denmark. Sarah S.T. Mak, Center for Evolutionary Hologenomics, University of Copenhagen, Denmark. Valerii

Plotnikov, Academy of Sciences of Sakha Republic, Yakutsk, Russia. Ruqayya Khan, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. David Weisz, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Vinícius G. Contessoto, Center for Theoretical Biological Physics, Rice University, Houston, USA. Antonio B. Oliveira Junior, Center for Theoretical Biological Physics, Rice University, Houston, USA. Achyuth Kalluchi, Department of Genetics, Cell Biology and Anatomy, Nebraska Medical Center, Omaha, USA. Arina D. Omer, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Sanjit S. Batra, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Muhammad S. Shamim, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Neva C. Durand, The Center for Genome Architecture, Baylor College of Medicine, Houston, USA. Brendan O'Connell, Department of Biomolecular Engineering, University of California, USA. Alfred L. Roca, University of Illinois at Urbana-Champaign, USA. Andreas Gnirke, Broad Institute of MIT and Harvard, USA. Isabel Garcia-Treviño, San Antonio Zoo, USA.Rob Coke, San Antonio Zoo, USA. Joseph P. Flanagan, Houston Zoo, USA. Kelcie Pletch, Houston Zoo, USA. Aurora Ruiz-Herrera, Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Spain. Eric S. Lander, Broad Institute of MIT and Harvard, USA. M. Jordan Rowley, Department of Genetics, Cell Biology and Anatomy, Nebraska Medical Center, USA. José N. Onuchic, Center for Theoretical Biological Physics, Rice University, USA. Love Dalén, Centre for Palaeogenetics, Stockholm, Sweden. Marc A. Marti-Renom, Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain. M. Thomas P. Gilbert, Center for Evolutionary Hologenomics, University of Copenhagen, Denmark. Erez Lieberman Aiden, The Center for Genome Architecture, Baylor College of Medicine, USA.

Ancient DNA (aDNA) sequencing analysis typically involves alignment to a modern reference genome assembly from a related species. Since aDNA molecules are fragmentary, these alignments yield information about small-scale differences, but provide no information about larger features such as the chromosome structure of ancient species. We report the genome assembly of a female Late Pleistocene woolly mammoth (Mammuthus primigenius) with twenty-eight chromosome-length scaffolds, generated using mammoth skin preserved in permafrost for roughly 52,000 years. We began by creating a modified Hi-C protocol, dubbed PaleoHi-C, optimized for ancient samples, and using it to map chromatin contacts in a woolly mammoth. Next, we developed "reference-assisted 3D genome assembly," which begins with a reference genome assembly from a related species, and uses Hi-C and DNA-Seq data from a target species to split, order, orient, and correct sequences on the basis of their 3D proximity, yielding accurate chromosome-length scaffolds for the target species. By means of this reference-assisted 3D genome assembly, PaleoHi-C data reveals the 3D architecture of a woolly mammoth genome, including chromosome territories, compartments, domains, and loops. The active (A) and inactive (B) genome compartments in mammoth skin more closely resemble those observed in Asian elephant skin than the compartmentalization patterns seen in other Asian elephant tissues. Differences in compartmentalization between these skin samples reveal sequences whose transcription was potentially altered in mammoth. We observe a tetradic structure for the inactive X chromosome in mammoth, distinct from the bipartite architecture seen in human and mouse. Generating chromosome-length genome assemblies for two other elephantids (Asian and African elephant), we find that the overall karyotype, and this tetradic Xi structure, are conserved throughout the clade. These results illustrate that cell-type specific epigenetic information can be preserved in ancient samples, in the form of DNA geometry, and that it may be feasible to perform de novo genome assembly of some extinct species.

FunCoup 5: Functional Association Networks in All Domains of Life, Supporting Directed Links and Tissue-Specificity
functional association network, tissue-specific network, protein network

Emma Persson, Miguel Castresana-Aguirre, Davide Buzzao, Dimitri Guala, Erik L.L. Sonnhammer. Department of Biochemistry and Biophysics, Stockholm University, Sweden.

FunCoup (https://funcoup.sbc.su.se) is one of the most comprehensive functional association networks of genes/proteins available. Functional associations are inferred by integrating evidence using a redundancy-weighted naïve Bayesian approach. FunCoup's high coverage comes from using eleven different types of evidence, and extensive transfer of information between species. Since the latest update of the database, the availability of source data has improved, and user expectations have grown. To meet these requirements, we have made a new release of FunCoup with updated source data and improved functionality, now including 22 species from all domains of life. In this release, directed regulatory links can be visualized for the human interactome, and subnetworks can be filtered for genes expressed in specific tissues. FunCoup 5 includes the SARS-CoV-2 proteome, allowing users to visualize and analyze interactions between SARS-CoV-2 and human proteins. This new release of FunCoup constitutes a major advance for the users, with updated sources, new species and improved functionality for analysis of the networks.

Seascape genomics of Scandinavian eelgrass (Zostera marina)
zostera marina, baltic sea, seascape genomics, population structure

Stefanie R. Ries, University of Gothenburg, Sweden. Per R. Jonsson, University of Gothenburg, Sweden. Per-Olav Moksnes, University of Gothenburg, Sweden. Katharina Kesy, University of Greifswald, Germany. Camilla Gustafsson, University of Helsinki, Finland. Tiia Möller-Raid, University of Tartu, Estonia. Marlene Jahnke, University of Gothenburg, Sweden.

The Baltic Sea can serve as a model system for future coastal oceans because of the strong influence of climate change and its extreme environment with a steep salinity gradient and varying temperatures. There is generally a low species diversity and high clonality in the Baltic Sea. In this study, we add available genetic data from 15 eelgrass meadows sampled along the Swedish coast and analyze a more extensive dataset with 24 additional meadows from the Baltic Sea and the Skagerrak-Kattegat area. Genetic data will be combined with environmental information to relate Scandinavian eelgrass population structure and connectivity to the seascape. We try to understand how oceanography, geography distance, and environment affect connectivity, adaptation, and future distribution of Scandinavian eelgrass population in a changing environment.

Arbuscular mycorrhizal fungi and their signature phosphorus transporters
mycorrhiza, soil, uptake

Lovisa Lundberg, Maliheh Mehrshad, Anna Rosling

Arbuscular mycorrhizal (AM) fungi have been instrumental to plant growth since the emergence of terrestrial plants and enhanced phosphorus (P) uptake is one of the key plant benefits of this association. Upon mycorrhizal colonization, the plant pathway for P uptake is downregulated in favor of the mycorrhizal pathway and a number of different P transporters orchestrate in the fungal uptake of P from soil to plants. Motivated by the central role that

P uptake plays in our understanding of the AM symbiosis, we set out to explore whether the emergence of these obligate plant symbionts is associated with a diversification of P transporter genes.

The Role of Population History in Shaping the Mutational Load of Structural Variants Relative to SNPs, in Distinct Island versus Continental Lagopus Lineages
demographic history, deleterious mutations, structural variatio, distribution of fitness effects

Rödin-Mörch, Patrik; David, Gabriel; Squires, Theodore; Magnusson, P, Kristinn; Höglund Jacob

While the importance of specific demographic histories in shaping patterns of mutational load conferred by deleterious single nucleotide polymorphisms (SNPs) has received considerable attention in the recent past, few studies have investigated the corresponding fitness consequences of structural variation in distinct evolutionary lineages. We performed heuristic-based filtering and rapid automated curation of short-read-discovered SVs callsets from 102 re-sequenced individuals across two recently (~2 million years) diverged ptarmigan (Lagopus) species. Population genetic analyses of the resulting high-confidence SV callsets reveal that the relative proportion of deleterious structural variants is consistently greater in small effective population sizes, but that the relative frequency of deleterious variants differs between population having experienced recent bottlenecks versus longer-term low Ne. Similar to SNPs, ratios of non-synonymous to synonymous polymorphisms in SVs are higher for historically small versus large populations, suggesting that many SVs may largely conform to nearly-neutral expectations.

Spatial metaTranscriptomics resolves host-bacteria-fungi interactomes
spatial transcriptomics, microbial diversity, host response

Sami Saarenpää, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. Or Shalev, Max Planck Institute for Biology Tübingen, Germany. Haim Ashkenazy, Max Planck Institute for Biology Tübingen, Germany. Vanessa Carlos, Max Planck Institute for Biology Tübingen, Germany. Derek Severi Lundberg, Max Planck Institute for Biology Tübingen, Germany. Detlef Weigel, Max Planck Institute for Biology Tübingen, Germany. Stefania Giacomello, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden.

Microbial diversity forms complex interaction networks with the host. Despite the recent advances in infection biology to study the host response and quantify microbes, we lack a comprehensive understanding of the microbial diversity at the spatial scale and the impact of these microbial taxa on host gene expression. To fill this gap, we present Spatial metaTranscriptomics (SmT), a sequencing-based technology that leverages 16S-18S/ITS-poly-d(T) multimodal arrays for simultaneous host transcriptome- and microbiome-wide characterization of tissues at 55-µm resolution. We showcase SmT on Arabidopsis thaliana leaves to study the spatial microbial distributions and the associated host response identifying 1,376 and 1,159 unique bacterial and fungal taxa, respectively. We unveiled leaf-scale spatial microbial hotspots and uncovered that the spatial distribution of microbes consistently drives their interactions.

SmT demonstrates the feasibility of studying spatially resolved host-pathogen interactions and microbial diversity elucidating complex infection processes where the spatial context is key for understanding infection processes.

Haplotype blocks, ancestral recombination graphs and the future of population genomics
haplotypes, ARG, selection, inference

Shipilina Daria, Uppsala University, Sweden. Pal Arka, IST Austria. Stankowski Sean, IST Austria. Chan Yingguang Frank, Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany. Barton Nicholas, IST Austria.

Population genetics is undergoing a transformative shift, replacing summary statistics with methods that leverage ancestral recombination graphs (ARGs) to synthesize genetic data inferences about molecular evolution. We propose defining "haplotype block" as fundamental unit for such analysis: based on ARG structure and containing complete information on the ancestry of a sample. We use simulated examples to demonstrate key features of the relation between haplotype blocks and ancestral structure, emphasising the stochasticity of the processes that generate them. In Heliconius butterflies and Littorina snails, we demonstrate using haplotype blocks for adaptive introgression and selective sweeps.

Understanding and applying the concept of the haplotype block will be essential to fully exploit frontier methods in modern data driven life science such as long and linked-read sequencing technologies.

Deleterious mutations and genetic load in inbred Scandinavian wolves
conservation genomics; deleterious mutations; genetic load

Linnéa Smeds, Uppsala University, Sweden. Hans Ellegren, Uppsala University, Sweden.

The Scandinavian wolf population was founded 40 years ago by three individuals, a recovery occurring after the original population had become extinct by persecution. The small number of founders in combination with geographic isolation from other wolf populations have led to extreme inbreeding. Still, the population has grown to over 500 individuals and wolf management is subject to an extensive political debate in both Sweden and Norway.
We have used functional annotation and evolutionary conservation scores on 209 whole genome wolf sequences from Scandinavia and neighbouring populations in Finland and Russia to study deleterious variation and how inbreeding and the founding bottleneck have affected the genetic load.

Our observations provide genome-wide insight into the character of genetic load and genetic rescue at the molecular level, and in relation to population history. Our results emphasize the importance of securing gene flow in the management of endangered populations.

Signatures of Selection in Rock Ptarmigan
reference genome, PacBio, vertebrate biology, conservation biology

Theodore Squires, University of Akureyri & Uppsala University, Iceland/Sweden. Patrik Rödin Mörch, Uppsala University, Sweden. Jacob Höglund, Uppsala University, Sweden. Kristin P. Magnusson, University of Akureyri & Icelandic Institute of Natural History, Iceland.

Rock Ptarmigan are largely sedentary birds with distribution across high-latitudes and alpine regions of the global north. Using whole genome data from individuals collected across Europe and North America, we sought to investigate the drivers of selection in

different populations to review the differences between them. Using active searches for selective sweeps, traditional SNP association, and redundancy analysis, in conjunction with environmental data, we were able to identify many candidate SNPs of evolutionary interest. Additionally, within the well represented Icelandic population, classical cycling has been observed at 10- and 5-year intervals. By using data from individuals collected leading into peak and trough years we aimed to identify if there were strong genetic signals predicting increasing or declining numbers. Using the same loci, we explored the relationship between different polymorphisms and individual fitness among different temporal cohorts of Icelandic birds.

Dimensionality Reduction Using Contrastive Learning
deep learning, AI, dimensionality reduction, PCA

Filip Thor, Carl Nettelblad

Interpreting the genetic variation in a dataset through visualization techniques is a powerful first step when studying population structure. This requires a dimensionality reduction method, where PCA is the most common. Frequently, pairwise distances in PCA embeddings correlate with distances in actual geographic origin for natural populations.
Other methods for dimensionality reduction include UMAP and t-SNE, which are non-linear, in contrast to PCA. While these methods are often superior to PCA for local population structure, they can fail to reconstruct global patterns.

We develop dimensionality reduction techniques using deep learning methods, tailored towards applications on SNP data. Specifically, we use contrastive learning, where the aim is to learn an embedding mapping similar samples to similar coordinates, without the need for labeled data.

In human and dog data, we produce rich embeddings capturing non-linear relationships better than PCA, resulting in a more pronounced clustering of the data, while retaining global structure.

Genetic basis of a regionally isolated sexual dimorphism involves cortex
genome assembly, cortex, melanin, sexual dimorphism, sexual selection

Kalle Tunström, Stockholm University, Sweden. Ramprasad Neethiraj, Stockholm, Sweden. Rachel Steward, Stockholm University, Stockholm Sweden, Lund University, Lund, Sweden. Christopher W. Wheat, Stockholm University, Stockholm Sweden.

In the Pieris napi butterfly, females in Northern Scandinavia uniquely exhibit a melanized form, adalvinda, absent in males. This intriguing sexual dimorphism prompted our study: is it due to local adaptation or intrapopulation sexual selection? Using extensive field trials and genetic data, we found males universally preferred unmelanized females. Further, through a nuanced comparative genomics approach, bulk segregant analysis, and population genetic comparisons, we discovered the adalvinda morph's association with the cortex gene, previously linked to color variations in other Lepidopterans. Our study underlines the role of cortex as a central evolutionary locus for Lepidopteran color variation.

Identifying novel non-coding driver mutations in canine and human diffuse large B-cell lymphoma using evolutionary constraint
non-coding mutations, evolutionary constraint, lymphoma

Anna Darlene, van der Heiden, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden. Suvi Mäkeläinen, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden. Sophie Agger, Department of Veterinary Clinical Sciences, University of Copenhagen, Denmark. Cheryl London, Cummings School of Veterinary Medicine, Tufts University, USA. F. Jamie Modiano, Department of Veterinary Clinical Sciences, College of Veterinary Medicine, University of Minnesota, USA. Karin Forsberg Nilsson, Department of Immunology, Genetics and Pathology, Uppsala University, Sweden. Kerstin Lindblad-Toh, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden. Maja Louise Arendt, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden.

Diffuse large B-cell lymphoma (DLBCL) is a prevalent lymphoma subtype in dogs and humans. In this study we focus on the non-coding genome, utilising an evolutionary constraint approach to identify conserved sites with functional potential, where modifications could contribute to DLBCL development. We conducted whole-genome sequencing on 73 canine and 41 human patients. Our analysis of non-coding constraint mutations (NCCMs) identified 108 and 81 NCCM-enriched genes within the canine and human cohorts, respectively, of which 10 were shared. An NCCM hotspot proximal to the proto-oncogene BCL6 was identified in both species, coinciding with a previously proposed enhancer region that regulates BCL6. Additional shared top-genes include known cancer genes RTP2, SST, BACH2, BCL7A, MLXIP, ZFP36L1, ZCCHC7, ACTN1, and PAX5, suggesting that regulation abnormalities in these genes might be a feature of DLBCL. These preliminary findings support the value of evolutionary constraint for detecting regulatory mutations in non-coding regions within cancer research.

New insights into the evolutionary history and adaptive potential of World Ocean krill using comparative population transcriptomics
krill; climate change; comparative population genomics; genetic adaptation

Marvin Choquet, Uppsala University, Sweden. Felix Lenner, Uppsala University, Sweden. Arianna Cocco, Uppsala University, Sweden. Gaëlle Toullec, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Erwan Corre, Sorbonne Université, France. Jean-Yves Toullec, Sorbonne Université, France. Andreas Wallberg, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden.

Genetic adaptation could be crucial to sustain populations under climate change but is poorly understood in pelagic species like zooplankton. We applied comparative genomics to a dataset of 20 krill species spanning tropical to polar environments to elucidate their evolutionary history and genomic bases of adaptation. Southern Ocean krill, such as the keystone species the Antarctic krill Euphausia superba, showed low levels of genetic variation and evolutionary rates, suggesting their potential to adapt to rapid climate change may be low. Nevertheless, we uncovered hundreds of candidate genes for long-term cold-adaptation among Antarctic Euphausia species. These include genes that govern ion transmembrane homeostasis and thermosensation, which have also been detected in Antarctic fish. These results suggest parallel evolutionary responses to similar selection pressures across groups of Antarctic fauna and provide new insights into the adaptive potential of zooplankton with critical roles in ocean ecosystems.