Norwegian Biodiviersity and genomics conference

11th and 12th of April, 2024.
Forskningskparken 21,
Auditorium: Forum (0. etg)

Thursday 11th of April

08:30	Registration with coffee and tea
09:00	Opening remarks by Kjetill S. Jakobsen
	Chair: Simen Rød Sandve
09:15	**Keynote lecture: Aoife McLysaght** Patterns and consequences of genome duplication in animals
09:55	Contributed talk **Benedicte Garmann-Aarhus** New insights into the Antarctic icefish radiation: Promises of the new congolli (*Pseudaphritis urvillii*) genome
10:10	Break
10:30	Contributed talks **Laura Valencia** De novo assembly of the crucian carp genome for the study of anoxia tolerance **Andrii Tarieiev** Complete plastid genome assembly for the critically endangered endemic birch species *Betula klokovii* Zaverucha and plastome-based phylogeny for *Betula* L.genus **Bram Danneels** Gene loss in the chemical defensome of Cetaccea and other marine mammals

	Lars Martin Jakt How changing the meiotic recombination landscape may have resulted in a teleost specific intron size distribution **Monica Hongrø Solbakken** The genome of Glacier Lantern fish reveals ongoing degradation of adaptive immunity
12:00	Lunch
	Chair: Torsten Hugo Struck
13:00	**Keynote Lecture: Alexander Suh** A bird's-eye view on why we need to sequence germline genomes
13:40	Contributed talk **Andrew Foote** Evolutionary insight from a cromosomal genome assembly of a Norwegian killer whale
14:00	Break
14:20	Contributed talks **Helle Tessand Baalsrud** An unusual mode of chromosomal evolution in Mucoromycota fungi **Pia Merete Eriksen** Another step up the phylogenetic hill: Assessing the viability of macrosynteny as a phylogenetic trait in Lophotrochozoa **Jule Drewalowski** Investigating the genomic basis of diversity in Syngnathiformes
15:10	Break

15:20	**Keynote Lecture: Lene Lange** From genome sequence to Biology and Applied-relevance
16:40	**Ivar Grytten** KAGE 2: Fast and accurate genotyping of structural variation using pangenomes **Jean-François Flot** GraphUnzip, HairSplitter and genomeTailor: a general framework for generating reference-quality genome assemblies of non-model organisms
17:00	Apéritif at eventlokalet: Toppsenteret
18:00	Dinner at eventlokalet: Toppsenteret

Friday 12th of April

09:00	Opening remarks
	Chair: Helle T. Baalsrud
09:10	**Dr. Olga Vinnere Petterson** Democratisation of genomics and hidden figures of reference genome sequence projects
09:50	Contributed talk **Torsten Struck** Species selection for large-scale genome projects – an automated process based on explicityly bottom-up defined criteria
10:10	Break
10:30	Contributed talks **Charlotte Benedict** Anonymous anemones: using target-enrichment to identify cryptic diversity in understudied taxa **Veronica O. T. Phan** Dietary adaptations along the northern limit of distribution: what does smooth snake (*Coronella austriaca*) eat in Norway? **Marvin Choquet** Unlocking the first large genome of the key zooplankton genus Calanus: challenges and insight **Anna Rix** High quality petrel reference genome from the snow petrel, *Padodroma nivea*
11:30	Break **Dabao Sun Lu** Investigating the history and consequences of secondary contact between divergent populations of *Trichaptum abietinum* in Europe

	James F. Flemming BOSTIn – User friendly software to assess potential problems in genomic datasets **Alberto Valero-Garcia** Homeotic genes and 3D genomics in Nemertea
12:30	Lunch
	Chair: Kjetill S. Jakobsen
13:30	Societal Impacts **Michael Dondrup** ELIXIR3: Empowering Biodiveristy Research through Collaboration and Innovation **Olav Lanes** ArcticZymes Technologies – Enzyme innovations from the Arctic **Hogne Bleie** Ervik Marine Research: Marine bioscience laboratories with an excellent access to the oceans
14:15	Break
14:30	**Hanne Mette D. Kristensen** The Life Science Cluster's reflections on EBP-Nor: Earth Biogenome Project in a societal context – from new knowledge to solutions for society **Tonje Heggeset** SINTEF: Harvesting the riches of earth through bioprospecting and biotechnological exploitation of genomic data
15:00	Closing remarks
15:15	End of conference

Abstracts

Keynote speakers

Professor Aoife McLysaght

Molecular Evolution Lab, Smurfit Institute of Genetics, Trinity College Dublin

Patterns and consequences of genome duplication in animals

Whole genome duplication (WGD) has long been recognised as a feature in the evolution of various lineages, especially plants, but questions persist regarding the timing and relevance of WGD events. This can partly be attributed to the complex and asynchronous resolution of the polyploid genome that generates misleading phylogenetic signals. One such example is found in the acipenseriform lineage where paddlefish and sturgeon in fact share an ancestral WGD event despite a plurality of phylogenetic trees showing independent duplication events. This is due to asynchronous rediploidisation of loci across the genome.

Taking this into consideration reveals a more ancient date for the WGD event, which occurs close to the Permian-Triassic mass extinction event. This timing is interesting because other work, primarily in plants, has suggested a link between polyploidy and survival through stressful times. In parallel we test these ideas for the first time in animals where we have developed a C elegans model for examining the short-term effects of polyploidy. We find that neotetraploid C elegans have improved fitness under stress, most notably cold shock. We are still trying to understand why and how.

Professor Alexander Suh

Head of Centre for Molecular Biodiversity Research (zmb)
Head of Section Molecular Biodiversity, Leibniz Institute for the Analyses of Biodiversity Change

A bird's-eye view on why we need to sequence germline genomes

The Earth BioGenome Project and related genome sequencing initiatives are in the ambitious process of generating one reference genome per species of eukaryotes. For some organisms, however, a single reference genome may not even be representative for an individual organism. A diversity of unrelated animals, ranging from some nematodes over some dipteran insects to lampreys and songbirds, undergo a phenomenon called programmed DNA elimination during germline/soma differentiation of their development. This can render genomes of somatic vs. germline cells from the same individual animal vastly different in terms of chromosome number, genome size, repeat content and/or gene content. Here I will discuss lessons learned from studying germline genomes of over 30 songbird species, together representing 2/3 of all 10,500 extant bird species, with regards to approaches for sampling, sequencing, and analyzing such fascinating cases of germline/soma genome differences.

Dr. Lene Lange

Company founder and owner of LL-Bioeconomy

From genome sequence to Biology and Applied-relevance

The fungal enzymatic invasive power is a prominent Fungal Kingdom characteristic. It is also the key asset for improved upgrading of biological resources to value-added product; valorizing the bio-resources now wasted (34% food waste) or downgraded (additional 20% food processing side-streams). However, annotation of genome sequence to only protein family of the fungal enzymes cannot be used for, neither biological understanding nor applied potential as each enzyme protein family holds several different functions. This presentation will introduce a series of new methods for peptide-based functional annotation of digestive enzymes directly from genome sequence; hereby opening for improved insight in fungal biodiversity, and for identifying the enzyme Hotspots of biomass degrading enzymes, found among all the more than 2.000 genome sequenced fungal species. The Hotspot fungal species of function diversity and the Hotspot fungal species for total degrading capacity will be exemplified. Furthermore, a case study of “learning from nature-guidance” will be described, to illustrate how to find the most interesting, unique, and efficient enzyme blends. Such blends of fungal enzymes have been refined during millions of years of evolution, degrading different types of substrates, cellulose, hemicellulose, pectin, and lignin.

Dr. Olga Vinnere Pettersson

Department of Immunology, Genetics and Pathology. Facilities:
Uppsala Genome Center, Uppsala University

Democratisation of Genomics and Hidden Figures of Reference Genome Sequencing Projects

Since the discovery of DNA, scientists wanted to unveil its mysteries. In 1970s, sequencing of nucleic acids became possible. As technologies and methods of analysis have evolved, generation of a reference genome sequence does not require huge international consortia of hundreds of labs and thousands of trained molecular biologists, as it was at the time of sequencing baker's yeast or the first draft of the Human Genome Project. Sequencing became available to any biologist, for virtually any organism on the planet. But what has changed in the field of the reference genome sequencing and how do these changes impact our view on the genome sequencing process?

Abstracts

Contributed talks

New insights into the Antarctic icefish radiation: Promises of the new congolli (Pseudaphritis urvillii) genome

Benedicte Garmann-Aarhus (Natural History Museum, University of Oslo) Michael Matschiner (Natural History Museum, University of Oslo) Arild Johnsen (Natural History Museum, University of Oslo) Torsten Hugo Struck (Natural History Museum, University of Oslo) Chiara Papetti (University of Padova) Bruce Deagle (National Research Collections Australia, CSIRO) Floriaan Devloo-Delva (National Research Collections Australia, CSIRO) Thomas Desvignes (Institute of Neuroscience, University of Oregon)

In the past 10 million years, Antarctic icefishes, a group comprising five families of the suborder Notothenioidei, have radiated into more than 100 species. During their radiation, they have developed key innovations, such as antifreeze glycoproteins, which has helped them dominate the below freezing water temperatures of the Southern Ocean. Whether this radiation was facilitated by gene flow between the different notothenioid species is not yet analyzed. By creating high quality genomic resources, analyses like these are now possible. In this talk, I will present the first chromosome level whole-genome assembly of the congolli (Pseudaphritis urvillii). The congolli is an early notothenioid lineage, and by using a reference genome created with HiFi and Hi-C data, assembled to EBP-Nor standards, we can aid in the understanding of how the icefish adaptive radiation occurred. The preliminary results show an estimated haploid genome size of 682 Mbp, with a contig N50 of 23.8 Mbp. With this reference genome, we hope to add to our main project of creating a complete and dated species tree to reconstruct the icefish radiation based on SNP data acquired through Illumina sequencing. This SNP data will also be used to assess the degree of gene flow between the icefish species. The congolli reference genome will also be used to study the timing of the diversification of early notothenioid lineages, with respect to paleoclimactic and geological changes.

De novo assembly of the crucian carp genome for the study of anoxia tolerance

Laura Valencia (Section for Physiology and Cell Biology, Department of Biosciences, University of Oslo), Sjannie Lefevre (Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo), Sissel Jentoft (Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo)

The crucian carp (*Carassius carassius*) is a fish with a remarkable anoxia tolerance, enabling it to survive the Norwegian winters in the wild, where seasonal anoxia is common in small lakes and ponds. While the physiological mechanisms are well documented, there is still a need for better understanding regulation at molecular and genetic level, and a fully sequenced, assembled and annotated genome will facilitate such studies. To obtain a high quality genome assembly, we sequenced crucian carp genomic DNA using chromosome conformation capture (HiC), PacBio long-read sequencing, and short-read sequencing for error correction. The structural and functional annotation were produced using RNA-seq data, teleost ortholog proteins, and UTRs were annotated with the PASA-pipeline and long-read RNA sequencing (Iso-seq data). Functional annotation of proteins predicted by the structural gene models were obtained by performing protein motif searches using Diamond blast, InterProScan and comparing to a well curated protein database from UniProt. With this new genome assembly with structural and functional annotation of a wild specimen of Norwegian crucian carp, we can begin to formulate more hypotheses about mechanisms that support the physiological adaptations that the crucian carp employs in order to tolerate survival in anoxic environments. Furthermore, the crucian carp has a recent whole-genome duplication (shared with some other cyprinids like common carp and goldfish), which provides the opportunity for diversified functions of the multiple gene paralogs; diversions that can be further studied with a high-quality reference genome.

Complete plastid genome assembly for the critically endangered endemic birch species Betula klokovii Zaverucha and plastome-based phylogeny for Betula L. genus

Andrii Tarieiev (Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, Göttingen, Germany), Oliver Gailing (Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, Göttingen, Germany)

In the current study, we assembled complete plastid genomes for the rare Ukrainian endemic Betula klokovii Zaverucha, related species (Betula pubescens Ehrh. and Betula pendula Roth), and the potential hybrid B. klokovii × pendula from short-read low-coverage whole-genome sequencing data. Assembled plastomes were annotated, aligned together with 50 birch plastome assemblies from NCBI GenBank, and used for phylogenetic reconstruction (Maximum Likelihood in RAxML-NG and IQ-tree, and Bayesian Inference in MrBayes) in two variants: 1) with partitioning according to annotation and assigning a separate model to each part and 2) without partitioning with a single model. Optimal models were identified by ModelTest-NG. All phylogenetic reconstructions revealed a clear separation between B. klokovii and B. pubescens, and therefore provided additional evidence that B. klokovii is a separate taxon. Overall, plastome-based phylogeny in the case of Betula L. genus provided better resolution in comparison to previously used markers. However, some taxa remain unresolved. Partitioning with assignment of different models has minor or no impact on topology and support values in comparison to reconstructions obtained on non-partitioned data with a single model. To resolve the status of B. klokovii and to estimate the level of its hybridization better there is a need to sequence much more samples of this birch (ideally – the whole population), related species, and prospective hybrids from close locations. In addition, performing reference-grade WGS and genome assembly would be very beneficial not only to address taxonomic problems with B. klokovii and related taxa but understand processes of speciation in birch better.

Gene loss in the chemical defensome of Cetacea and other marine mammals

Bram Danneels (Department of Informatics, University of Bergen, NO) Diogo Oliveira (CIIMAR, University of Porto, PT) Raquel Ruivo (CIIMAR, University of Porto, PT) Filipe Castro (CIIMAR, University of Porto, PT) Odd-André Karlsen (Department of Biological Sciences, University of Bergen, NO) Anders Goksøyr (Department of Biological Sciences, University of Bergen, NO) Inge Jonassen (Department of Informatics, University of Bergen, NO)

The marine ecosystem is under pressure from anthropogenic activity, for example from chemical pollution. Many marine mammals are top predators in the marine food web and thus especially at risk for accumulation of chemical pollutants. Accordingly, marine mammals are among the species with the highest measured levels of chemical contaminants in their tissues. Most animals have evolved a system of genes and cellular signalling pathways that help them detect, convert, and excrete foreign chemicals (xenobiotics), commonly referred to as the chemical defensome. Previous research has found that the nuclear receptors NR1I2 (pregnane X receptor; PXR) and NR1I3 (constitutive androstane receptor; CAR) are pseudogenized and non-functional in cetaceans (whales & dolphins). These nuclear receptors play a key role in detecting xenobiotics and activating the xenobiotic metabolism in mammals. In this work we developed a pipeline that uses reference genomes to detect chemical defensome genes (genes related to xenobiotic metabolism) in target genomes and assesses whether they are pseudogenes or not. We used this pipeline to identify and analyse the chemical defensome in a range of marine mammals and their close relatives. We detected cetacean gene losses in many defensome-related pathways and some of these losses are shared in other marine mammal lineages, such as pinnipeds. This work further demonstrates that gene detection and pseudogene assessment are crucial tools in functional analysis of the many reference genomes generated by large-scale genome projects such as the Earth Biogenome Project.

How changing the meiotic recombination landscape may have resulted in a teleost specific intron size distribution.

Ann-Christin Zinner (Nord University), Lars Martin Jakt (Nord University)

In mammals, and vertebrates generally, the sites of meiotic recombination are thought to be defined by the sequence specific binding of the PRDM9 protein. This leads to double stranded breaks that are either repaired by homologous recombination or, in the minority of cases, lead to crossover and meiotic recombination. In teleosts, most PRDM9 orthologues do not contain the aKRAB domain that is necessary for its function in mice. Our observations indicate that aKRAB has been lost multiple times during the teleost evolution and that this has resulted in the evolution of a strongly bimodal intron size distribution with an antimode at around 256 bp. I will describe the evolution of PRDM9 orthologues across the teleosts and demonstrate, by simulation, how this may have resulted in an apparently convergent evolution towards a common distribution in the absence of selection.

The genome of Glacier lantern fish reveals ongoing degradation of adaptive immunity

Monica Hongrø Solbakken[1,2*], Ole Kristian Tørresen[1], Torkild Bakken[3], Kjetill S. Jakobsen[1]

[1]University of Oslo. [2]Norwegian University of Life Sciences, current affiliation. [3]Norwegian University of Science and Technology

Lantern fishes (Myctophiformes) are centrally placed in the ocean food web and contributes significantly to midwater biomass. They are potential targets for human exploitation due to their biomass but concerns regarding the robustness of all mid/deep-water ecosystems have been raised. Genomic resources are valuable tools for e.g. stock management, and here we present a draft genome of the glacier lantern fish (*Benthosema glaciale*) using PacBio HiFi and Hi-C sequencing. Both haplotype assemblies are ~1.3Gb in size with 22 chromosomes, scaffold N50 ~44Mb and BUSCO scores indicating ~96 % complete genes. We characterized central adaptive immune genes including three additional Myctophiformes genomes.

The adaptive immune system is organized around Major Histocompatibility Complex (MHC), T-cell receptors (TCR) and B-cell receptors (BCR, antibodies). Collectively, they establish tailored responses to infection. However, in glacier lantern fish we observe loss of classical MHCI as well as destroyed functionality of MHCII and TCRs (via the loss of crucial co-factors). Remaining is an intact BCR system likely producing generic germline encoded antibodies. Interestingly, similar findings are observed in *Electrona antarctica*, but not in the two *Gymnoscopelus* investigated. The underlying reason for eliminating adaptive immunity in some Myctophiformes, and how it potentially affects these species' robustness to e.g. environmental changes remain unknown. Thus, the genome of *Benthosema glaciale* is a great example of why genomic resources are crucial when evaluating species in the context of resource exploitation or conservation initiatives.

Sequencing: Norwegian Sequencing Centre. Funding: S. G. Sønneland Foundation. Specimen collected by trawl in Stjørnfjorden (Trondheimsfjorden) Norway.

Evolutionary insights from a chromosomal genome assembly of a Norwegian killer whale

Andrew Foote, Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, 0316 Oslo, Norway

Here I provide background on how the search for the natal population of a Norwegian killer whale which stranded on the Netherlands coast, led to the first high throughput sequencing genome of a marine mammal species. That first assembly provided insights into the signatures of convergent evolution along the branches to different marine mammal lineages. In addition it provided the reference for mapping short-read sequence data to, providing population genomic insights into the evolutionary histories of distinct killer whale ecotypes. The assembly was updated by the Darwin Tree of Life team using the original DNA extract for PacBio long-read sequencing and combined with Hi-C data from an Icelandic killer whale generated by the DNAzoo team, providing a high quality chromosomal assembly. This has provided the contiguity necessary for accurate inference of runs of homozygosity and structural genetic variants among populations and ecotypes. I will conclude this presentation with an outline of a new project using the killer whale as a comparative model to study the genetic basis of the evolution of a mid-life female menopause, which is common to both humans and killer whales. In both species the evolutionary drivers of a female menopause are thought to be selection against cross-generational reproductive conflict and kin selection for investment in existing offspring. I will make the case that the X-linked grandmother hypothesis is a good fit for killer whale sociality and that the X-chromosome may hold the key to the genetic underpinning of this rare trait.

An unusual mode of chromosomal evolution in Mucoromycota fungi

Helle Tessand Baalsrud (Norwegian University of Life Sciences), Thu-Hien To (Norwegian University of Life Sciences), Dana Byrtusova (Norwegian University of Life Sciences), Mariann Arnyasi (Norwegian University of Life Sciences), Lars Grønvold (Norwegian University of Life Sciences), Juan Fu (Norwegian University of Life Sciences), Volha Shapaval (Norwegian University of Life Sciences), Simen Rød Sandve (Norwegian University of Life Sciences)

Decades of comparative genomics studies have revealed that gene content and order on chromosomes, known as synteny, can be conserved over hundreds of million years of evolution. Synteny will degrade with increasing evolutionary distance between species, however, this process is poorly studied in most lineages as it requires dense sampling of high-quality genomes. Using chromosome-scale assemblies from 25 species in the understudied Mucoromycota fungi we show that one lineage displays an unusual mode of chromosome evolution resulting in conserved mesosynteny. Mesosynteny entails conservation of gene content within homologous chromosomes, but with complete reshuffling of gene order. We also found several polyploidization events, with one species having mesosynteny between its subgenomes. Mesosynteny has previously been reported in a distantly related group of Ascomycota fungi, but its evolutionary significance is currently unknown. We found a unique link in Mucoromycota between mesosynteny and the loss of centromere proteins (CENP-A and C), which are conserved across all other Eukaryotes. This could have affected how new chromosomal rearrangements navigate meiosis as heterozygotes. The mesosyntenic species also have a significantly higher repeat content. The cause-and-effect relationship between repeat evolution, meiosis and chromosomal rearrangements is still an open question in biology. Our study sheds new light on constraints affecting synteny evolution by demonstrating that the mode of synteny evolution varies between lineages, possibly due to differences in the meiotic machinery and/or repeat landscape. Mucoromycota is thus an excellent new system to explore the evolutionary drivers and consequences of breaking up gene linkages.

Another step up the phylogenetic hill: Assessing the viability of macrosynteny as a phylogenetic trait in Lophotrochozoa

Pia Merete Eriksen (University of Oslo), James F Fleming (University of Oslo), Torsten H Struck (University of Oslo)

Finding the "true" phylogenetic tree of life is a Sisyphean task that taxonomists and phylogeneticists have worked on since the realisation that the vast diversity of life on our planet is interconnected. Before the sophisticated molecular methodologies used in the modern day, evolutionary relationships were exclusively inferred through macroscopic morphological analyses such as the conservation of limb structure, developmental patterns, organ structure, and similar phenotypic traits. In our modern era, morphological analysis has fallen out of fashion amongst extant taxa. However, with increasing accessibility of genomic data comes opportunities to examine morphology on a new scale, by treating chromosomal macrosynteny and arrangement as phenotypic traits: karyo-morphology. In this presentation, I will discuss the viability of macrosynteny as a trait for phylogenetic reconstructions. Synteny, briefly explained, is the linkage of genes across species (Renwick, 1971), whilst macrosynteny is this phenomenon on a larger, chromosomal scale, observing multiple genes or genomic regions. By following macrosyntenic patterns across species, I will evaluate the efficacy of macrosynteny as a phylogenetic trait. I will constrain my discussion to Lophotrochozoa, a diverse clade containing a plethora of familiar invertebrates such as the annelids and molluscs. Previous research has been conducted into the role of synteny for other groups, such as Ctenophora and Porifera (Schultz et al., 2023), and I intend to build on this foundation and extend the exploration of the potential of synteny into this, at time of writing, underexplored clade.

Investigating the Genomic Basis of Diversity in Syngnathiformes

Jule Drewalowski (University of Copenhagen), Josefin Stiller (University of Copenhagen)

In the realm of biodiversity genomics, the utilization of genome resources is pivotal for advancing our understanding of biodiversity, evolution, and conservation. Despite a substantial increase in the availability of genomes, a significant number of marine vertebrates lack genome references. Syngnathiformes, a captivating group of ray-finned fishes, exhibit vast morphological and ecological diversity, including seahorses, pipefish and seadragons. Yet, the genomic basis of these differences remains largely unexplored due to missing availability of densely sampled genomes. Addressing this gap, we created a whole-genome dataset encompassing over 250 syngnathiform fishes, sampled across all major lineages. This dataset comprises short-read draft assemblies, assembly statistics, repeat annotations and phylogenetic analyses, allowing for a range of comparative studies that shed light on both lineage-specific and convergent trait evolution. A first investigation has unveiled substantial genome size variation, ranging from 283 Megabases (Mb) to 1,396 Mb. The large majority of this diversity is attributed to repetitive sequences, particularly transposable elements, which are positively correlated with the genome size. Our ongoing research aims to explain the noteworthy differences in speciation across lineages, exploring potential associations with genomic, ecological and morphological traits. This project promises to deepen our understanding of the intricate interplay between genetic factors and the rich biodiversity exhibited by syngnathiform fishes.

KAGE 2: Fast and accurate genotyping of structural variation using pangenomes

Ivar Grytten (University of Oslo, Department of Informatics), Knut Dagestad Rand (University of Oslo, Department of Informatics), Geir Kjetil Sandve (University of Oslo, Department of Informatics)

Structural variation is known to play an important and often overlooked role in regulating disease and traits, but accurately detecting structural variants from sequencing data has traditionally been difficult. However, recent improvements in high-quality genome assembly along with methodological advancements in pangenome creation have opened up the landscape for methods that use such pangenomes for structural variant calling and genotyping. We here present KAGE2, which accurately and efficiently genotypes structural variation by exploiting the availability of pangenomes that represent known variation in a population. Through comprehensive benchmarking, we highlight limitations of existing methodology and show that KAGE2 is more accurate and considerably faster than existing methods.

GraphUnzip, HairSplitter and GenomeTailor: a general framework for generating reference-quality genome assemblies of non-model organisms

Roland Faure[1] Jean-François Flot1[,2]

[1]Evolutionary Biology & Ecology, Université libre de Bruxelles (ULB), Brussels, Belgium

[2]Interuniversity Institute of Bioinformatics in Brussels – (IB)², Brussels, Belgium

Unlike model organisms that have generally very low heterozygosity (because of having been inbred for many generations in laboratory settings and/or because of having experienced population bottlenecks or dramatic range expansions), the genomes of non-model organisms are characterized by extensive polymorphism comprising substitutions, indels and structural rearrangements. This poses specific challenges for assembling them, a challenge that long reads help address. We present here three distinct tools, published and unpublished, that we devised to tackle this challenge. All three tools present the characteristics of taking as input and/or producing GFA assembly graphs rather than FASTA files of contigs. GraphUnzip detects and duplicates regions that were collapsed during assembly of diploid/polyploid organisms, thereby improving the continuity and accuracy of such genome assemblies; HairSplitter detects and recovers variations that were lost during assembly of genome repeats (such as collapsed haplotypes), thereby disentangling them; and GenomeTailor checks the accuracy of an assembly by mapping on it a set of long reads, detecting assembly mistakes and rearranging the assembly to maximize the number of reads that map onto it end-to-end. Using these tools, we have been able to produce reference-quality assemblies of the genomes of a variety of non-model organisms, including several organisms sequenced in the framework of the ERGA (European Reference Genome Atlas) pilot. By producing highly accurate assemblies that are structurally consistent with the set of input reads, our framework paves the way for meaningful comparisons of genome assemblies of closely related organisms.

Species selection for large-scale genome projects– an automated process based on explicitly bottom-up defined criteria

Torsten Struck (University of Oslo) Astrid Böhne (Leibniz Institute for the Analysis of Biodiversity Change) Christian de Guttry (Swiss Institute of Bioinformatics) Rosa Fernandez (Spanish National Research Council) Isabelle Florent (National Museum of Natural History) Carmela Gissi (University of Bari Aldo Moro) Bernhard Hausdorf (Leibniz Institute for the Analysis of Biodiversity Change) Jennifer Leonard (Spanish National Research Council) Thomas Marcussen (University of Oslo) Camila Mazzoni (Leibniz Institute for Zoo and Wildlife Research) Seanna McTaggart (Earlham Institute) Jose Melo Ferreira (University of Porto) Alice Minotto (Earlham Institute) Rita Monteiro (Leibniz Institute for the Analysis of Biodiversity Change) Rebekah Oomen (University of Oslo) Joao Pimenta (University of Porto) Jaakko Pohjoismäki (University of Eastern Finland) Katja Reichel (Free University of Berlin) Ana Riesgo (Spanish National Research Council) Andrii Tarieiev (University of Goettingen) Olga Vinnere Pettersson (University of Uppsala)

As there are usually more species to be sequenced than is possible given the available resources, a crucial part of all large-scale genome projects is the selection of species to be sequenced. Accordingly, some criteria and procedures must be in place to accomplish this. The BGE project employed a bottom-up procedure for the development of the selection process by involving the larger ERGA community. Herein, we present the development of the species selection process at all its stages. It is an automated process based only on clearly outlined objective criteria. For example, subjective motivations are not part of the decision process., which is without human intervention. This allows for a more objective, less biased species selection for large genome projects. The process itself is a four-stage process including (1) an exclusion stage, (2) a prioritization stage employing a decision-tree model and additional ranking for country and researcher representation, (3) a feasibility check with additional adjustment for genera with multiple species suggestions and (4) a final permit check. We also report the effect of the process on the composition of the pool of the selected species in relation to the pool of all suggested species. For example, half of the species were regarded as not feasible and in consequence fewer phyla and countries were represented in the final pool. The main reasons for this are that they are either too small in body size, too large in genome size, in too bad condition of preservation or too challenging to collect.

Anonymous Anemones: using target-enrichment to identify cryptic diversity in understudied taxa

Charlotte Benedict- The Ohio State University Department of Evolution, Ecology, and Organismal Biology, 1315 Kinnear Rd, Columbus, OH, 43212 Alonso Delgado- The Ohio State University Department of Evolution, Ecology, and Organismal Biology, 1315 Kinnear Rd, Columbus, OH, 43212 Isabel Pen- The Ohio State University Department of Evolution, Ecology, and Organismal Biology, 1315 Kinnear Rd, Columbus, OH, 43212 Claudia Vaga- Department of Invertebrate Zoology, Smithsonian Institution's National Museum of Natural History, 10th and Constitution Ave NW, Washington DC 20560 Marymegan Daly- The Ohio State University Department of Evolution, Ecology, and Organismal Biology, 1315 Kinnear Rd, Columbus, OH, 43212 Andrea M. Quattrini- Department of Invertebrate Zoology, Smithsonian Institution's National Museum of Natural History, 10th and Constitution Ave NW, Washington DC 20560

Sea anemones (Order Actiniaria) are a diverse group of marine invertebrates ubiquitous across marine ecosystems. Despite their wide distribution and success, a knowledge gap persists in our comprehensive understanding of their diversity within tropical systems, owed to sampling bias of larger and more charismatic species overshadowing cryptic lineages. This study aims to delineate the sea anemone diversity in Moorea (French Polynesia) with the use of a dataset from the Moorea Biocode's "BioBlitz" initiative, which prioritized the sampling of more cryptic and understudied taxa. Implementing a target enrichment approach, we integrate 71 newly sequenced samples into an expansive phylogenetic framework and contextualize Moorea's diversity within global distribution patterns of sea anemones. Our analysis corroborates the presence of several previously documented sea anemones in French Polynesia and identifies for the first time the occurrence of members of genera Andvakia and Aiptasiomorpha. This research unveils the diverse sea anemone ecosystem in Moorea, spotlighting the area's ecological significance and emphasizing the need for continued exploration. Our methodology, encompassing a broad BLAST search coupled with phylogenetic analysis, proved to be a practical and effective approach for overcoming the limitations posed by the lack of comprehensive sequence data for sea anemones. We discuss the merits and limitations of current molecular methodologies and stress the importance of further research into lesser-studied marine organisms like sea anemones. Our work sets a precedent for future biodiversity studies stemming from "BioBlitz" endeavors, particularly among cryptic taxa with less comprehensive sequence data available

Dietary adaptations along the northern limit of distribution: What does smooth snake (Coronella austriaca) eat in Norway?

Marielle C. Samslått (independent researcher), Veronica Q. T. Phan (Norwegian University of Life Sciences), Alison C. Cleary (British Antarctic Survey), Audun Slettan (University of Agder), Beate S. Johansen (University of Agder)

Application of molecular analysis to investigate diet composition can greatly increase the knowledge about how species survive along their northern limit of distribution. The diet of smooth snakes (*Coronella austriaca*) primarily consists of reptiles in central Europa. However, prey choice numbers decrease towards the Arctic, thus the variety of food available to smooth snakes are limited in Norway compared to central Europe. In this study, we identify the potential prey species of smooth snakes in Southern Norway using Illumina next-generation sequencing (NGS) combined with DNA metabarcoding. Taxonomy for Amplicon Sequence Variants (ASVs) of particular abundance or interest were confirmed or further refined using manual BLASTN searches against the NCBI Genbank database. Stomach content DNA from 17 dead smooth snake individuals was amplified with a universal primer (fwhF2+fwhR2n) for a short cytochrome c oxidase I (COI) fragment to determine dietary items. Eight primary prey species were detected, of which were four reptile species and four mammal species. Short-tailed field vole (*Microtus agrestis*) was the most preyed upon and composed 12.7% of total occurrences. Second most common prey species was Eurasian shrew (*Sorex araneus*), followed by common lizard (Zootoca vivipara), slow worm (*Anguis fragilis*) and wood mouse (*Apodemus sylvaticus*). DNA from grass snake (*Natrix natrix*) was also detected in all but one individual, and bank vole (*Myodes glareolus*) and European adder (*Vipera berus*) were revealed in two separate individuals.

Unlocking the first large genome of the key zooplankton genus Calanus: challenges and insight

Marvin Choquet (Natural History Museum, University of Oslo, Norway), Andreas Wallberg (Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden)

Copepod species of the genus Calanus dominate the zooplankton biomass in the North Atlantic and Arctic oceans, where they play critical roles in the stability of marine ecosystems and fisheries. They are among the first organisms to respond to environmental changes by shifting the distribution of their populations and are thereby widely used as climate change indicators. Despite their ecological importance, very little is known about the evolutionary mechanisms underlying Calanus spp.' biological success and even less on the molecular basis behind their adaptative strategies. To understand the potential of Calanus spp. to adapt to climate change, and consequently be able to implement strategies to anticipate future ecological shifts, we need knowledge about their genome and molecular mechanisms encoded within. However, the combination of small body sizes, large genomes and high inter-individual genetic diversity observed in Calanus spp. have hindered our ability to assemble a reference genome, until now. We devised our strategy combining the largest body-sized specimen of Calanus on Earth with the least DNA-demanding technology of long-read sequencing. Despite challenges linked to limited sequencing coverage, we successfully assembled the genome of the Arctic C. hyperboreus and were able to retrieve most genes. Our analyses revealed a high prevalence of transposable elements, whose proliferation may be responsible for the large genome size. With resequencing experiments, we also discovered a remarkable inter-individual genome size variability, independent of geography. The potentially adaptive role of this biological phenomenon remains to be investigated.

High quality petrel reference genome from the snow petrel, Pagodroma nivea

Anna Rix (Durham University) Rus Hoelzel (Durham University) Erin McClymont (Durham University)

We present the first high-quality genome for petrels constructed from one of the southernmost breeding species of birds, snow petrels, Pagodroma nivea. The genome was constructed using PacBio HiFi reads from the blood of a female bird. The constructed genome has a BUSCO score of 97.5% complete [(s:96.3%, D:1.2%), F:0.5%, M:2.0%, n:8338] when compared to birds. There are 228 contigs covering 1,335,759,149 bases in the haplotype genome with a scaffold N50 of 30,282,880 bases. The nearest relative to snow petrels with a chromosome level assemble is the Humboldt penguin, which has a similar BUSCO score and 389 contigs. We compared the genome to those assembled for other birds from the Austrodyptornithes clade and consider patterns associated with both genome structure and evolutionary history. The historical population dynamics for snow petrels was additionally calculated using PSMC. Snow petrels have had two major expansions and declines over the past 10 million years.

Investigating the history and consequences of secondary contact between divergent populations of *Trichaptum abietinum* in Europe

Dabao Sun Lu1[*], Ine-Susanne Hopland Methlie[1*], Michelle Vera Castellanos[1], Jørn-Henrik Sønstebø[2], Anneli Andersen[1], Kathleen Helleland[1], David Peris[3], Sundy Maurice[1] , Håvard Kauserud[1], Inger Skrede[1]
[1]University of Oslo, [2]University of South-Eastern Norway, [3]Institute of Agrochemistry and Food Biotechnology (IATA) CSIC

The glacial cycles of the Quarternary have been a major factor in shaping the present-day genetic structure of species on the northern hemisphere. During the glacial periods, populations of species were commonly isolated in multiple glacial refugia where they could diverge in allopatry. In the ensuing interglacials these diverged populations expanded their ranges and came into secondary contact. The resulting admixture of genomes in these hybrid zones can be seen as natural experiments for the study of speciation. There is limited information on postglacial migration of fungi, and it is unclear to what extent they have followed the same recolonization route as plants and animals. Here we trace the postglacial history of the widespread wood-decay fungus *Trichaptum abietinum* in Europe using population genomic analyses of 138 genomes. We detected a Mediterranean lineage in Southern Europe and a Boreal lineage in Eastern Europe and Fennoscandia, that we connect to glacial survival in southern and eastern refugia, respectively. These two lineages form an admixture zone in Central Europe, and local ancestry analyses of the admixed individuals reveal that the second half of their largest chromosome is entirely inherited from the Boreal lineage, indicating either strong selection or genomic incompatibilities. To understand this pattern further we have generated F1 lab hybrids between the Boreal and Atlantic lineages to compare to the natural hybrids, which our analyses suggest originated from secondary contact at least 600 generations ago. In addition, a fourth Atlantic lineage was detected in western Scandinavia and Europe, and we connect its origin to a western refugia. Intriguingly, the Atlantic and Boreal populations appear to have limited admixture in Scandinavia despite being in proximity and able to mate in the lab, and we assess the fitness of these hybrids with different growth and decomposition experiments. In conclusion, the postglacial history of *T. abietinum* in Europe resemble what has been found in numerous plants and animals, with survival and divergence in multiple separate glacial refugia and the formation of hybrid zones in Central Europe.

BOSTIn - User-friendly software to assess potential problems in genomic datasets

James F. Fleming (University of Oslo), Torsten H Struck (University of Oslo)

As genomic datasets grow progressively larger, and phylogenetic analysis becomes an expected part of any new manuscript detailing new genomic information, effective use phylogenetic tools becomes a concern of paramount importance. Unfortunately, phylogenetics is a field that can often be seen as impenetrable to non-specialists. BOSTIn (Broad Overview of Sequence Topology and INcongruence) is designed from the ground-up to address this issue. As a "one-stop shop" for phylogenetic artifact identification - a quick and easy to use software that provides users with detailed information about their alignment's vulnerability to artifacts such as compositional heterogeneity, branch length heterogeneity and site saturation. Unique to BOSTIn is its new "narrative" report feature, which conversationally explains what each metric means with respect to the user's dataset, and provides practical next steps considering the context of their data. Our hope is that BOSTIn will be immediately useful to non-specialists whilst also helping to develop their abilities to evaluate common phylogenetic model violation metrics effectively and independently.

Homeotic genes and 3D genomics in Nemertea

Alberto Valero-Gracia (Natural History Museum, University of Oslo, Norway), Michael Matschiner (Natural History Museum, University of Oslo, Norway), Torsten H. Struck (Natural History Museum, University of Oslo, Norway)

Nemertea, a predominately marine phylum composed of over 1,300 species, is a Spiralia/Lophotrochozoa clade that can be subdivided into three major groups: Palaeonemertea, Hoplonemertea, and Pilidiophora. They are ambush hunters that use an eversible proboscis and toxins to capture prey. Moreover, nemerteans develop either directly or indirectly by primary larvae. This makes this clade a promising system for studies of toxin and life-cycle evolution. Despite its potential, nemerteans have so far been severely neglected from genomic studies with only two genomes of pilidiophoran species being published. To fill this gap, here we present six new high-quality genomes containing representatives of all three major groups. These genomes, generated de novo by means of PacBio HiFi sequencing and assembled with Hifiasm, were compared based on a whole-genome alignment together with the two publicly available Nemertea genomes. Our study is the first to support the phylogenetic relationships of the phylum with whole genome data and enables an understanding of the genome evolution dynamics within the clade. In this work we specifically investigate the evolution of selected gene clusters to address whether there is homeotic gene contiguity, and whether there are major structural rearrangements of the genome architecture (i.e., gene flow modifications, transposable element behaviour, and introgression changes). In a broader perspective, our genome-based study aims to understand how conserved and lineage-specific genetic features shape evolution.