ESEB Tandem Repeat Satellite Symposium Program on August 17

Time	Name	Title	Abstract	Country
3:00 - 3:15 pm	Ye Jin	Genome-wide patterns of tandem repeat abundance across species are driven by retrotransposons	Short Tandem Repeats (STRs) are sequential repeat elements in DNA sequences that are abundant in most eukaryotic species. STRs are known to contribute to variation in gene function and gene expression across species. It has been observed that the abundance and characteristics of STRs vary widely across genomes of different species. However, the evolutionary origins of this variation and its implications remain largely unexplored. Previous work suggested that species- or clade-specific retrotransposon elements may represent a major source of STRs in eukaryotic genomes. Here, we explored the composition of STRs across the tree of life with species ranging from yeast to primates and investigated the critical role retrotransposons play in driving species-specific differences in genomic STR content. Our results confirm previous reports that homopolymers are strongly enriched in primate genomes, and that this trend is driven by SINE retrotransposon elements. Further, we identify DNAREP1_DM and TART_DV elements as drivers of dinucleotide abundance in multiple fruit fly species and CR1-Y1_Ave and CR1-Y2_Aves as drivers of pentanucleotide abundance in birds. Finally, we find that dinucleotide repeats are enriched in abundance and tend to be longer in rodent species, but that this enrichment is likely not driven by rodent-specific retrotransposon elements and is therefore likely driven by alternative mechanisms. Overall, our work highlights the major role retrotransposon elements play in shaping the composition of STRs across the genomes of diverse species.	United States
3:15 - 3:30 pm	Oxana Lundstrom	STR Explorer - workflow for genome wide studies on short tandem repeats	Microsatellite sequences are short tandem repeats(STRs) of two to six nucleotides. They are known to be hypervariable due to the accumulation of length mutations by intra-allelic polymerase slippage on microsatellite sequence during replication. STRs have been poorly studied until very recently due to their highly polymorphic nature which complicates their annotation. But more and more evidence suggests that STRs play an important role in cancer and other diseases. In colorectal cancer(CRC) variation in STR regions have been shown to influence protein-expression levels and increase tumor progression (Contente et al. 2002). The instability of microsatellite regions is used by clinicians today to classify tumors into different groups. This marker includes only five microsatellite regions though and doesn't exploit the whole variety of this phenomena. Currently it is unknown whether distinct CMS groups exhibit distinct patterns of STR variation and in what CRC-related functional pathways STRs are involved. To explore this issue we have developed a pipeline to systematically search for STR biomarkers in colorectal cancer that can be easily adapted for similar studies in other medical conditions. Our pipeline is the first building block towards a workflow for personalised medicine approach through STRs. The statistical framework TRAL (Schaper 2015) have been used to find STRs in human reference genome. We share the logic behind creating a genotyping panel relevant for cancer research and how we arrived at the microsatellite panel we use today. These inferred STRs have been further genotyped on more than 400 genomes from patients with colorectal cancer available to us through the TCGA (The Cancer Genome Atlas Network 2012). The results are presented as a relational database with programmatic access through a REST-full API, - STR Explorer. Additional Python modules are available to add more genotyped data to the database . We demonstrate how to combine STR annotations from STR Explorer with gene expression analysis and other clinically relevant data to perform a genome-wide association study (GWAS) on STRs in colorectal cancer as a case study. The inferred STR risk variants can later be validated as novel RNA and protein targets, which will serve as additional information for patient risk stratification and therapy-response prediction.	Sweden
3:30 - 3:45 pm	Max Verbiest	Uncovering the Diverse Roles of Short Tandem Repeat Variation in Colorectal Cancer	Insertions and deletions of repeat motifs are common in short tandem repeat (STR) loci and can affect gene expression levels and protein structures. In cancer, their high mutability can also result in tumor-specific neoantigens, making them promising targets for immune therapies. Although recently developed computational approaches allow for accurate genotyping of STRs from sequencing data, many investigations of STR variation in cancer predate these methods. We therefore suspect that the contribution STRs have to the molecular picture of cancer is currently underestimated. Therefore, we used repeat-specific methods to generate and genotype a panel of over 1.8 million STR loci in colorectal cancer (CRC) patients from The Cancer Genome Atlas. We detected tumor STR variants by comparing repeat lengths between patient-matched healthy and diseased tissue. We then estimated the contribution of these tumor STR variants to gene expression changes in CRC using existing catalogues of STRs known to affect expression. For STR variants in coding regions, we determined the expected changes in protein structure and monitored the generation neoantigens. While a lot of this is still work in progress, we expect our results will provide a better understanding of the diverse roles of STR variation in CRC. Using computational methods specifically designed to analyse STRs, we will demonstrate the importance of this abundant but often bypassed source of variation in cancer. Furthermore, by making our panel of STR loci and their variation in CRC patients available to the community, we hope to stimulate future investigations into this important topic.	Switzerland
3:45 - 4:00 pm	Brankica Mravinac	Tandem repeats constitute the major genomic difference between Tribolium sibling species	The flour beetle Tribolium freemani is a sibling species of the model organism and important pest Tribolium castaneum. The two species can produce hybrid progeny, but F1 hybrids are sterile. To address genomic incompatibilities between T. freemani and T. castaneum, we sequenced the T. freemani genome by PacBio HiFi technology and assembled it de novo using the T. castaneum genome assembly as a reference. Comparison of T. freemani and T. castaneum genome assemblies showed that the genomes are very similar in their coding sequence, and we successfully annotated 96% of the T. castaneum genes in the T. freemani assembly. By contrast, the two genomes differ drastically in repetitive DNA, and the main differences come from tandemly repeated satellite DNA sequences. Each genome is dominated by a species-specific major satellite DNA. The major satellite DNAs comprise 20% and 30% of the T. castaneum and T. freemani genomes, respectively, and are the main constituents responsible for genome size disproportion (200 Mb vs. 320 Mb). In addition to the most abundant tandem repeats, we also analyzed a number of low-copy satellite DNAs that revealed concerted evolution between the two species. In conclusion, tandem repeats make the most substantial quantitative and qualitative difference in genomic sequence between T. freemani and T. castaneum. Since the highly abundant major satellite DNAs predominately build pericentromeric and centromeric heterochromatin of all chromosomes, it is possible that the species-specific tandem repeats play a role in postzygotic reproductive isolation betweenTribolium sibling species.	Croatia
4:00 - 4:15 pm	Inês Borges	Ancient satellite DNA and rapid turnover across passerine birds	Satellite DNA (satDNA) is among the fastest evolving elements in the genome and is highly abundant in some eukaryotic genomes. Its highly repetitive nature means it is challenging to assemble, and thus underrepresented in most assemblies and often understudied as a result. Birds are an ideal model organism for the study of satDNA and its evolution, since the large amount of available sequenced genomes of this clade allows for dense sampling across the evolutionary timescale and the low number of satDNA families within their satellitomes facilitates their study and comparison between species. Here, we characterize satDNA and its evolution across Passeriformes, an avian clade containing two-thirds of all bird species spanning nearly 50 million years of evolution. With this goal we used both short-read data and long-read assemblies of species representative of over 30 families in this clade to shed light on the evolution of its satellitome. We focused on examining the phylogenetic relationship between satellites common to most species as well as characterizing satellite array structure and location in the genome. We found a unique satellitome in each sampled bird family containing satellites from 3 to 18 different families, and identified differences in the satellite content between male and female individuals suggesting occasional satDNA expansion on the female-specific W chromosome.	Sweden
4:15 - 4:30 pm	Abhishek Singh	Ancestral deconvolution confronts socio-cultural structure in western himalayas	The Himalayas, a mountainous habitat of a diverse group of people with distinctive cultures and social structures, is one of the most undisturbed regions due to its difficult geography and climatic conditions. For thousands of years, several populations of western Himalayas followed strict social barriers leading to the present caste endogamy. With this conserved genetic pool based on the caste system, the genetic distinction must follow social division between the Himalayan populations. However, due to earlier migrations and demographic changes, the genetic makeup of the populations might be affected which further needs to be assessed on a larger scale. Here we obtain the genetic data of 1026 unrelated individuals belonging to 13 populations based on 20 STRs from the high elevational sites of western Himalayas, India. Most of these populations follow strict marital practices within their caste and thus lesser chance of shared genetic components leading to the population sub-structure. However, both Bayesian and non-Bayesian analyses revealed a lack of population substructure indicating deep ancestral mixing and panmixia. The majority of the populations showed close genetic affinity irrespective of their caste and social structure. Although, the populations of western Himalayas clustered together when assessed with the other populations of central and eastern Himalayas. These observations suggest that the genetic structure of western Himalayan populations is a panmictic population that defies the ancestral gradient related to the caste system.	India