



Abstract for EIHA conference (June 2020)

Cannabis gene variation - comparison of multiple genome assemblies

By Dr. Oori Weisshaus/NRGene

In recent years the *Cannabis sativa* genome is at the centre of research efforts due to the species' large array of secondary metabolism products and their potential to interact with mammalian systems. Removal of legislative barriers has enabled the publication of multiple genome assemblies using various technologies for sequencing and assembly, mostly using highly heterozygous plants. In this work we demonstrate a pan-genome comparison of multiple cannabis genomes of hemp and drug type cultivars. We have de-novo assembled two new heterozygous elite-line genomes with fully phased high accuracy assemblies and compared them to four public reference-level non-phased assemblies and several commercial clones at WGS-level assembly. This comparison was carried out in a pan-genome structure based on a common coordinate system of the CBDRx reference genome. We have aligned and ordered multiple genome assemblies with phased haplotypes and created uniform chromosome mapping. We have generated a non-redundant dataset of 43,000 transcripts and mapped it to each haplotype. This enabled identification of allelic variation and novel homologues of important genes for cannabinoid biosynthesis, as well as an accurate comparison of copy number, present-absent and structural variations, identification of highly conserved gene region duplications and identification of novel candidate genes. We have identified hyper-variable regions and massive genome rearrangements that may hold great significance to cannabis and hemp research and breeding.