GENOMIC, TERPENE AND CANNABINOID PROFILES OF A PUTATIVELY NOVEL *CANNABIS* SPECIES

Kevin McKernan[1*], Cameron Miller[2*], Brad Douglass[2], Jessica Spangler[1], Braden Doane[2], Vasisht Tadigotla[1], Colin Montgomery[2], Yvonne Helbert[1], Lei Zhang[1], Douglas Smith[1], Jeffrey Raber[2]

[1]Medicinal Genomics Corporation, Woburn MA, 01801, USA
[2]The WercShop, 15405 SE 37th St. Bellevue WA 98006
*These authors contributed equally to this work

An Australian feral male cannabis strain (Australian Bastard Cannabis or ABC) was crossed with *Cannabis Sativa L.* (Purple skunk Oregon) and presented with atypical leaf patterns and stunted but hardy growth (offspring termed ABCh). Microscopy identified small trichomes and encouraged further evaluation with UV-HPLC (UV confirmed high performance liquid chromatography) that identified THCA and CBCA peaks. This cultivar was further tested for terpene profiles with GC-FID (gas chromatography-flame ionization detector) identifying many common terpenes in *Cannabis*. Thin layer chromatography (TLC) was used to confirm the cannabinoid results against Cerilliant standards and the sample was selected for genomic analysis.

Quantitative PCR for a Y chromosome *Cannabis* marker and a cannabinoid synthase gene were both positive, suggesting further analysis with whole genome sequencing. 28 million paired 251bp reads (7.5Gb, ~10X coverage) were mapped against an existing female *Cannabis* reference (CanSat3 or Purple Kush) to assess the genomic similarity to a recently sequenced *Cannabis* strain[1]. As a control, reads from several male and female *Cannabis* strains reported to be *C.sativa* or *C.indica* dominant cultivars were sequenced and mapped to the same CanSat3 reference genome to derive read mapping percentages using a single trimming and mapping algorithm. The ABCh strain had 94.6% of the reads mapping to the CanSat3 reference compared to 99.58% and 99.60% of the control reads from two respective male genomes (WIFI and Grape Stomper). The female C.*indica* cultivar 'LA Confidential' and the female hemp strain Finola mapped at 99.1% and 97.7% respectively. These data suggest ABCh is more distant to Purple Kush than Finola Hemp and a variety of other 'drug type' cultivars. Mapping Bonobo to the human genome produced 95.87% reads mapping underscoring a more distant relationship of ABCh mapped to Purple Kush than speciated mammals mapping to each other. Since mapping percentages can be influenced by microbial contamination we also performed SNP calling and principle component analysis (PCA). PCA also revealed distance albeit less extreme and perhaps more related to commonalities in the hybrid *C.Sativa* genetics.

We then analyzed the ABCh read mappings to THCA synthase reported by Sirikantaramas *et al.*[2] and found 7 SNPs in the 1635bp open reading frame. One amino acid changing variant (Ala5Thr) was in the N-terminal signal peptide but no SNPs were found in the FAD binding domain suggesting a fully functional THCA synthase enzyme with a variant of unknown significance in the N-terminal signal peptide.

In conclusion, the combination of cannabinoid and terpene profiling with genome sequencing can be a powerful toolset for cannabis classification. These methods offer synergistic perspectives to the *Cannabis Sativa L.* taxonomic debate and suggest a possible fourth *Cannabis* species or subspecies to *C.indica*, *C.sativa* and *C.ruderalis.* Sequencing of the original unhybrized genetics is required to further this debate.

1. van Bakel, H. et al. The draft genome and transcriptome of Cannabis sativa. *Genome Biol* **12**, R102 (2011).

2. Sirikantaramas, S. et al. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *The Journal of biological chemistry* **279**, 39767-39774 (2004).