Male associated DNA Cannabis (MADC2) markers have previously been described by Mandolino et al. These markers were reported to target the MADC2 repeat and generate many bands with gel electrophoresis of which the male cannabis plant delivered a unique band. While these markers are reported to accurately detect male plants, the use of gels and visual inspection of banding patterns could be improved with more scalable quantitative PCR (qPCR) methods. Towards this end, we cloned and sequenced these bands to over 1000x coverage using next generation sequencing. We discovered highly variable MADC2 sequences complicating qPCR assay design.

We further aligned many female whole genome sequencing reads to the published MADC2 references and found significant sequence homology and sequence coverage suggesting over 500 copies of this degenerate repeat in female genomes. Several SNPs exist in the reported Mandolino primers and suggest an allele specific amplification of male versions of MADC2.

As a result, we chose to whole genome sequence two male cannabis plants to find markers more compatible with a qPCR assay. 30X whole genome shotgun of male cultivars WIFI and Grape Stomper were aligned to existing female references. Unmapped male reads were then assembled to generate 2523 high quality contigs (1.5Mb of contigs). These contigs were in-silico screened for microbial contamination, extreme coverage and heterozygosity to prioritize 259 hypo-variable contigs >500bp in length. Quantitative PCR assays were designed and initially screened against 15 males and 15 females with no errors. To finalize and confirm our assay design we sequenced a distant male cultivar known as ABCh to 10X coverage and aligned those reads to the 259 hypo-variable male contigs to find adequate coverage but many polymorphisms in the contigs targeted for qPCR. Eleven percent of the male contigs had zero coverage in ABCh underscoring the value in the extensive screening required for assay design in polymorphic plants.