Accumulation of somatic mutations leads to genetic mosaicism in Cannabis

Kristian Adamek[1], Davoud Torkamaneh[1], Andrew Maxwell Phineas Jones[1*]

[1] Department of Plant Agriculture, University of Guelph, Guelph, N1G 2W1, ON, Canada

* Correspondence: amjones@uoguelph.ca

Abstract

Cannabis is typically propagated using stem cuttings taken from mother plants to produce genetically uniform propagules. However, producers anecdotally report that clonal lines deteriorate over time and eventually produce clones with less vigour and lower cannabinoid levels than the original mother plant. While the cause of this deterioration has not been investigated, one potential contributor is the accumulation of somatic mutations within the plant. To test this, we used deep sequencing of whole genomes (>50x depth of coverage) to compare the variability within an individual *Cannabis sativa* cv. "Honey Banana" plant sampled at the bottom, middle and top. Overall, we called over 6 million sequence variants based on a published reference genome (SNPs, MNPs, and indels) and found that that the top had the most by a sizable amount. We compared the variants among the samples and uncovered that nearly 600K (34%) were unique to the top while the bottom only contained 148K (12%) and middle with 77K (9%) unique variants. Bioinformatics tools were used to identify high impact mutations in critical cannabinoid/terpene biosynthesis pathways. While none were identified, some contained more than double the average level of nucleotide diversity (π) in or near the gene, including OLS, CBDAS, HMGR2 and CsTPS9FN. The first two genes code for essential enzymes required for the cannabinoid pathway while the other two are involved in the terpene pathways, demonstrating that mutations were accumulating within these pathways and could influence their function. Overall, these data identified a measurable number of intra-plant genetic diversity that could impact the long-term genetic fidelity of clonal lines and potentially contribute to the observed decline in vigour and cannabinoid content.

Keywords: Cannabis, Whole-genome sequencing, Somatic mutation, Genetic mosaicism

Introduction

Cannabis sativa L. (marijuana, hemp, cannabis; Cannabaceae) is regarded as one of the first crops humans domesticated and is primary a dioecious diploid annual species (2*n* =20) cultivated for fiber, oil, seed and its medicinal and psychoactive properties (Hillig, 2005). The main pharmaceutical and psychoactive compounds are cannabinoids that accumulate in trichomes produced primarily on floral tissues of female plants (van Bakel *et al.*, 2011). To date, 177 cannabinoids have been identified and described, with the two most naturally abundant, well-studied, and sought after being (–)-*trans*-Δ^9–tetrahydrocannabinol (THC) and cannabidiol (CBD) (Hanuš and Hod, 2020; Hurgobin *et al.*, 2020). These compounds have many demonstrated therapeutic properties, including alleviating symptoms in epilepsy (Devinsky *et al.*, 2014) and multiple sclerosis (van Amerongen *et al.*, 2018) and are being investigated for other ailments such as Alzheimer's disease (Watt and Karl, 2017). Additionally, cannabis produces hundreds of different terpenes/terpenoids, which are also known to have therapeutic effects, including antifungal, antiviral, anticancer, anti-inflammatory, antiparasitic, antioxidant, and antimicrobial activities, and are thought to interact with cannabinoids to alter their activities (Hanuš and Hod, 2020). Cannabinoids, terpenes and other secondary metabolites are produced in the capitate stalked glandular trichomes and minor variations in concentrations of each may result in distinct therapeutic effects (Hurgobin *et al.*, 2020).

The polyketide, cannabinoid, and methylerythritol phosphate (MEP) pathways are responsible for the creation of the cannabinoids (i.e. THCA, CBDA, CBCA, CBGA) and the monoterpene,

mevalonate (MEV), sesquiterpene and methylerythritol phosphate (MEP) pathways produce several different terpenes. The medicinal effects of Cannabis plants depend on the relative concentration of these compounds and are often classified into three main categories based on the THC:CBD ratio. Type I plants express a well over 1:1 THC:CBD ratio, Type II plants have moderate amounts with a near equal ratio, and Type III plants contain a less than 1:1 ratio (Hurgobin *et al.*, 2020). However, it should be noted that this is an oversimplification of the chemical diversity found within Cannabis and that each plant has a unique chemical fingerprint that may impact its biological activity.

Clonal propagation is the primary method used when cultivating Type I & II cannabis plants for medicinal or recreational use to ensure genetic and phenotypic uniformity (McKernan *et al.*, 2020). To achieve this, mother plants are established from elite seedlings that have been selected largely based on their specific chemical profile. Mother plants are capable of supplying hundreds or thousands of cuttings, usually taken from the apical region of the plant. The mother plants are maintained in an indefinite vegetative state for many years using a constant long day photoperiod (18:6 hours) and are occasionally replaced using a clonal propagule taken from themselves.

Vegetative propagation is used in many other domesticated plants to maintain valuable genotypes including bananas, potatoes, grapes, hops, and coffee trees (McKey *et al.*, 2010; Carrier *et al.*, 2012). Theoretically, clones produce plants that are genetically identical and phenotypically similar to the parent stock, however, cannabis producers have observed a decline in the quality of clones taken from a mother plant, usually resulting in reduced cannabinoid production and plant vigour (Cannabis growers, personal communication, 2020). While no peer-reviewed study has investigated this in cannabis, it has been discussed widely in the gray literature (Burnstein, 2019)

and this is a well-known phenomenon in other species that demonstrate a decline in plant vigour during extended periods of vegetative propagation (Muller, 1932, 1964).

Muller's Rachet, a term first proposed by Felsenstein (1974), is a theory developed to explain this phenomenon and suggests that in the absence of sexual recombination, species accumulate irreversible somatic mutations (Muller, 1932; Govindaraju *et al.*, 2020). Further, since the majority of random mutations are deleterious, the long-term effect of their accumulation is a decline in plant vigour similar to what has been observed in cannabis. While somatic mutations have also positively impacted agriculture by producing unique clonal varieties of apples, citrus, and wine grapes derived by propagating genetically diverse bud sports, the phenomenon is problematic for long term genetic preservation of elite individuals (Cruzan, Streisfeld and Schwoch, 2018). Whether somatic mutations are beneficial or deleterious, it is well-known that they occur and can lead to genetic diversity even within a single plant.

The source of this diversity lies in the nature of plant growth and development, in which meristematic regions grow and develop independent from one another. As they grow, they each accumulate a unique set of somatic mutations that leads to genetic diversity within a plant. This phenomenon has been under investigation since the 1970s where isozyme analysis was initially used to demonstrate genetic variation within single trees (Marshall and Allard, 1970). More recently, DNA sequencing verified this and identified that genetic distance increases systematically throughout a tree such that the mutation load is greatest at the distal end (Diwan *et al.*, 2014; Orr *et al.*, 2020). This phenomenon is known as the Genetic Mosaicism Hypothesis (GMH) and states that individual plants become genetically diverse due to the accumulation of spontaneous mutations occurring randomly as independent branches grow (Gill *et al.*, 1995). While this phenomenon is often neglected during preservation of clonal lines, it could have a

significant impact on long term genetic fidelity, especially in species with higher than average mutation rates.

In this study, we examined if the GMH applies to cannabis by sequencing the full genome of three samples taken from different locations within a single mother plant. These data were used to identify nucleotide variants within the plant using bioinformatics tools and assess the degree of variation within the plant. We then calculated the unique and shared variants among samples. Lastly, we investigated the potential impact of the nucleotide variants in critical cannabinoid and terpene synthesis genes. Overall, this study confirms there was a significant degree of genetic variability within a single cannabis plant and raises concerns about long term genetic preservation using clonal propagation.

Results

Deep full genome sequencing of bottom, middle and top regions from a mother plant

We performed deep whole-genome sequencing (WGS) on three samples, taken from the stems located at 59 cm, 151 cm, and 226 cm from the top of the pot of a 2.4 m tall, 1.5-year-old mother plant (Figure 1A). This mother plant was a clonally propagated seedling selection from *Cannabis sativa* cv. "Honey Banana" (HB) plant (BrantMed Inc., Brantford, ON), which is a high THC type-I plant. In total, we generated >1 billion 150-bp paired-end reads using an Illumina Novaseq 6000 technology (Table 1). This represents an average 58x depth of coverage across the three samples. These reads were mapped against the public cannabis reference genome (cs10 v. 2.0; GenBank Accession No. GCA_900626175.2) with a mapping success rate of greater than 93% (Table 1), thus covering >97% of the cs10 v. 2.0 genome sequence. Using the Fast-WGS (Torkamaneh *et al.*, 2018) bioinformatics pipeline, we detected 1.3, 0.9, and 1.7 M nucleotide variants (i.e. single- and

multiple nucleotide variants (SNVs and MNVs) and small insertions/deletions (indels)) in the samples derived from the bottom, middle, and top of the mother plant, respectively. Originally, 6.4 M nucleotide variants were detected but due to the high level of heterozygosity discovered, we employed a filter to both the minimum number of reads (minNR) and the minimum number of reads containing variants (minNV) to greater than 10. Altogether, over 3.8 M variants compared to the reference genome were identified with a transition/transversion (Ts/Tv) ratio of 1.9. As detailed in Table 1, the top of the plant had the most variants, followed by the bottom, while the middle had the fewest, demonstrating intra-plant genetic diversity. When comparing the locations there was a difference of ~400K variants between the bottom and middle as well as the bottom and top. However, the difference between the middle and top was much greater with a total of 850K variants.

Table 1. Statistics related to the whole-genome sequencing of Cannabis samples.

Sample	Nb of reads (M)	Depth of coverage (x)	Mapping rate (%)	SNV	MNV	INS	DEL	Total[a]
Bottom	362	62	94.1	966,945	104,519	95,910	98,124	1,265,498
Middle	314	54	93.8	670,974	63,953	64,168	64,988	864,083
Top	336	58	93.9	1,290,497	154,789	133,460	136,061	1,714,807

[a]Number of variants compared to the reference genome with minNR>10 and minNV>10.

Mutational variation detected inside a singular *C. sativa* mother plant

For further analysis, the catalogues of detected variants relative to the cannabis reference genome (cs10 v. 2.0) for each sample (i.e. bottom, middle, and top) were compared to assess the overlapping and unique variants. As can be seen in Figure 1B, more than 600K variants, compared to reference genome, were shared among all samples representing 51%, 75%, and 38% of variants detected in the sample from the bottom, middle, and top section of the mother plant, respectively. Bottom and top samples shared the greatest number of variants (403K), however top and middle shared only 72K variants. The lowest overlap was observed between bottom and middle, 67K. Most interestingly, the top sample contained the most unique variants at 34% (592K) followed by the bottom with 12% (147K) and the middle with 9% (77K). The most intriguing aspect of the result is that the top sample contained the most *de novo* (new) mutations. We document a very high rate of somatic mutations among the bottom, middle and top of a cannabis mother plant during vegetative growth.

According to the π statistic, the nucleotide diversity within three samples from a single plant was $\pi = 6.0 \times 10^{-4}$. To explore the genetic similarity between samples using nucleotide variants, we constructed a phylogenic tree using a genetic distance approach and the neighbour-joining method with repetition 1000x bootstrap test (Figure 1C). This shows two main branches, an individual bottom branch as well as a middle and top branch as the second, since they share a node. Within a single plant, we observed a very quickly declined linkage disequilibrium (LD) (Supplemental Figure 1). The LD was seen to decay to its half in only a few kb.

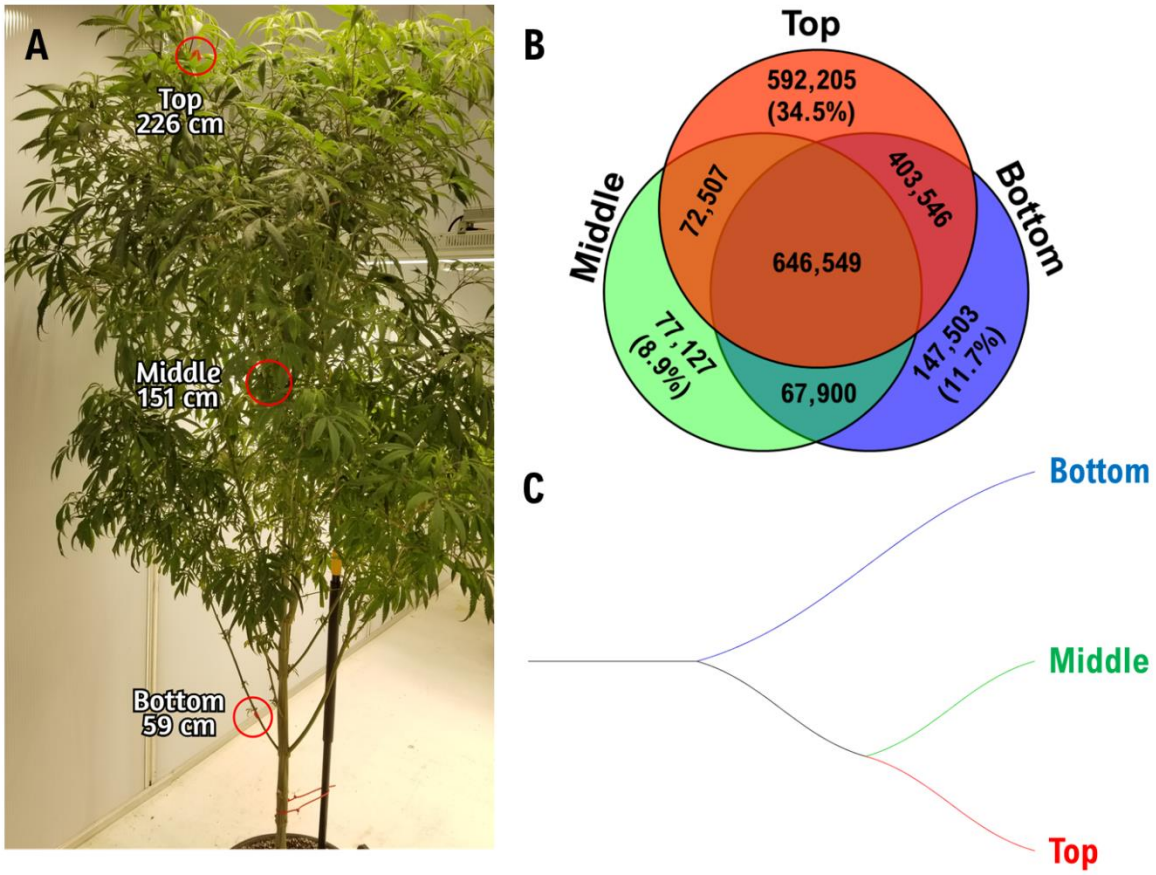

Figure 1. Whole-genome sequencing of a cannabis mother plant. (**A**) Honey Banana, a mother plant (age: ~1.5 years), exhibited incredible growth, vigour and excellent traits which made this an ideal cultivar for this study. Three stem samples were taken from the top (226cm), middle (151cm) and bottom (59cm) and had their genomes fully sequenced. (**B**) A Venn diagram reveals the overlapping and unique nucleotide variants for the top, middle and bottom. (**C**) A phylogenic tree created using a genetic distance approach and the neighbour-joining method demonstrates that the top and middle are more similar by sharing a common node while the bottom has a separate branch.

Finally, based on a visualization approach using IGV (Robinson *et al.*, 2011), we determined if the variants were clustered in specific regions of the genome. As can be seen from Supplementary Figure 2, there were certain areas in the genome that contain fewer mutations and others where elevated levels of mutations emerged. Specifically, one, five, eight and the X chromosomes had an apparent increased count while chromosomes two, three, four, six, seven and nine showed to

have fewer. Furthermore, an intriguing observation was that chromosomes with more mutations also had higher levels of mutations in euchromatin regions (i.e. gene rich). Altogether, clusters of mutations appeared across the genome and as a result could indicate mutational hotspots exist within the cannabis genome.

Functional impact on the genome from mutations are divergent depending on the location on the plant

To explore the potential functional impact of the mutations, we classified sequence variants into five categories based on their localization and identified the putative impact of the mutations. As can be seen in Figure 2A, more than half (51%) of the variants were in up/downstream regions, hence in close proximity of genes (5kb before and after gene) and the other 49% of the variants were located in intergenic regions (35%), genic regions (13%), splice sites (0.4%), and untranslated regions (UTR; 0.4%). Additionally, the genic category consists of exons, introns, and transcriptional variants at 2%, 4.5%, and 6.7%, respectively. From a functional standpoint, we were particularly interested in the subset of mutations predicted to have a large impact. Therefore, we explored the category of the high impact mutations (i.e. variants which are predicted to have a disruptive impact on the protein, probably leading to protein truncation, loss of function or triggering non-sense-mediated decay). Figure 2B presents the unique and shared high impact mutations from the bottom, middle and top samples. This information follows a similar pattern as seen for entire variants (Figure 1B) where the percentage of variants are similar. The number of shared high impact mutations between all samples was 845 which corresponds to 41%, 77%, and 28% of the total high impact mutations for the bottom, middle and top, accordingly. The top sample had the most unique high impact mutations with 1,234 (40%), next the bottom with 247 (12%) and lastly the middle with 63 (6%). The most intriguing was middle and top because there was a large

difference between the unique variants (1.2K) (Figure 2B). Also, they shared 952 with 107 exclusively together, which represents 31% of top and 87% of middle total high impact mutations. To provide a more relevant perspective, we mainly focused on the high impact mutations from the middle and top samples as they are new mutations and ultimately, they might show a different phenotype compared to the original plant (i.e. bottom). In total, 1,333 high impact mutations were divided into four categories and represented frameshift (46%), premature stop codon (32%), splice site (17%), and stop loss (4%) mutations (Figure 2C). These results show that a large number of high impact mutations (>1K) arise as the cannabis mother plant grows and is maintained for a long time and as a consequence potentially severely affects the function of important genes. Owing to the lack of any significant enrichment in terms of gene ontology (GO) annotation, we investigated the functional impact of these mutations in the important cannabinoid and terpene production related genes individually using public databases.

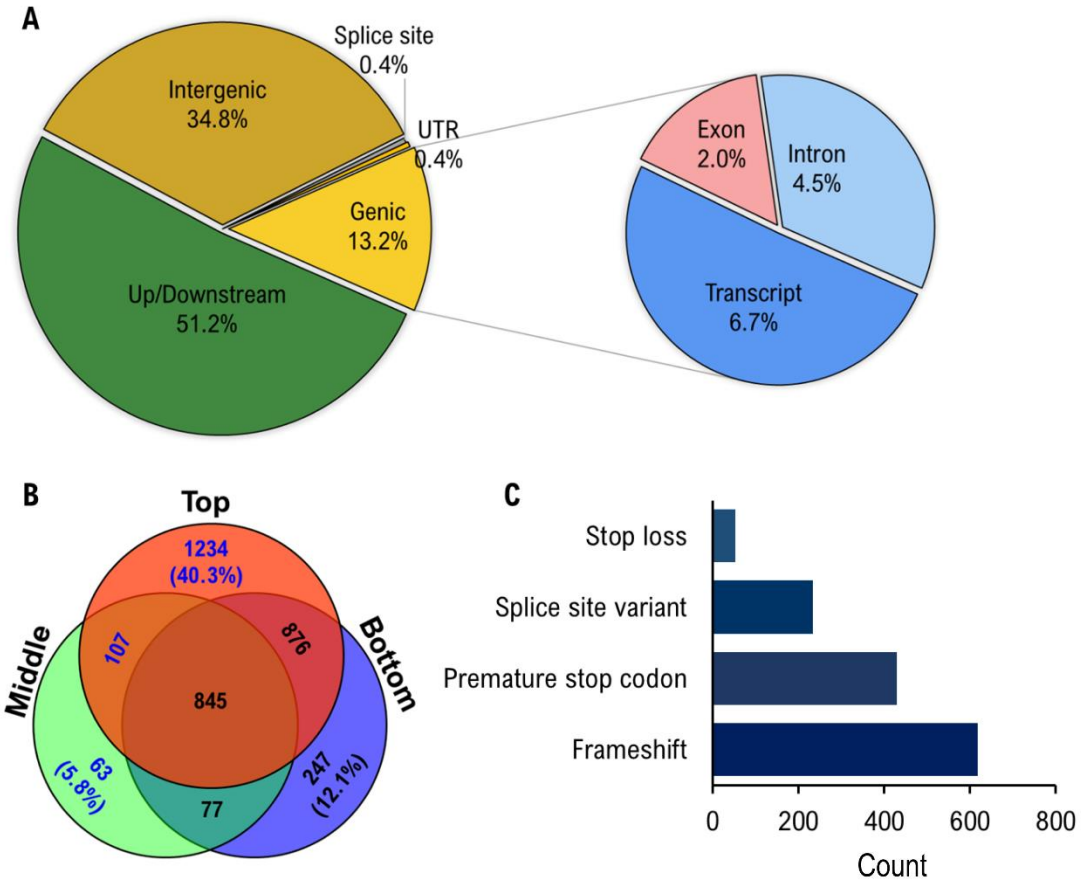

Figure 2. Analysis of the location and impact of nucleotide variants. (**A**) A pie graph visualization that displays the percentages from the five main categories of found variants (intergenic, up/downstream, genic, UTR, and splice site) while a secondary pie graph reveals the breakdown of the genic category (transcript, exon and intron). (**B**) A Venn diagram showing the high impact variants throughout and between the top, middle and bottom with the blue text highlighting *de novo* variants. (**C**) High impact mutations were organized into four categories and represented frameshift (46%), premature stop codon (32%), splice site (17%), and stop loss (4%) mutations.

Properties of novel mutations on cannabinoid and terpene pathways genes

To more specifically focus on the potential impact of these mutations on secondary metabolite production, we studied the functional impact of mutations that occurred in or near the necessary cannabinoid and terpene pathway genes. We determined which enzymes were required to create

essential chemical compounds and which chromosome they can be found on the cs10 v. 2.0 reference genome for both cannabinoid and terpene pathways from public databases (Table 2). As seen from the Table 2, we calculated and predicted the number and type of mutations, nucleotide diversity (π) in 20kb window encompassing these genes and the contrast of the π in these genic regions compared to the average genome wide. Analysis of the prediction of the functional impact of the mutations determined that none of these genes contained a high impact mutation and recorded all observed variants as modifier. The ratio of gene to genome wide π allowed us to provide a sense of potentially conserve or somatic mutation prone genes. The most notable genes that undergo somatic mutation were ones with ratio values greater than 2.0 which includes OLS, CBDAS, HMGR2 and CsTPS9FN, with values of 2.08, 4.35, 2.95, and 2.50, respectively. Of the remaining 40 genes, 32 were under 2.0 and the other eight are missing data due to no data found on the NCBI's cannabis protein table. Overall, four genes are noteworthy because the number of variants emerging, and seem to be more prone to somatic mutations than others, but this will require additional research to determine if this is a common trend.

1 **Table 2.** Variants in cannabinoid and terpene pathway genes

Abbreviation	Name	Chromosome	Reference	Nb of variants[a]	Gene π[b]	Related to Avg[c]
Polyketide pathway						
HCS/AAE1	Hexanoyl-CoA synthetase 1	3	Stout et al. 2012	10	2.86E-04	0.47
OLS	Olivetol synthase	8	Taura et al. 2009	44	1.26E-03	2.08
OAC	Olivetolic acid cyclase	9	Gagne et al. 2012	28	8.14E-04	1.35
Cannabinoid pathway						
CBGAS	Cannabigerolic acid synthase	X	Page and Boubakir 2012	31	8.95E-04	1.48
THCAS	Inactive tetrahydrocannabinolic acid synthase	7	Sirikantaramas et al.2004	2	5.71E-05	0.09
CBDAS	Cannabidiolic acid synthase	7	Taura et al. 2007	92	2.63E-03	4.35
CBCAS	Cannabichromosomeomenic acid synthase	7	Page and Stout 2015	2	5.71E-05	0.09
MEP pathway						
DXS1	DXP synthase	9	Booth et al. 2017	19	5.57E-04	0.92
DXS2	DXP synthase	4	Booth et al. 2017	13	3.71E-04	0.61
DXR	DXP reductoisomerase	3	Booth et al. 2017	0	0.00E+00	0
MCT	MEP cytidylyltransferase	4	Booth et al. 2017	29	8.29E-04	1.37
CMK	CDP-ME kinase	2	Booth et al. 2017	1	2.86E-05	0.05
MDS	MECDP-synthase	5	van Bakel et al. 2011	16	4.71E-04	0.78
HDS	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	2	Booth et al. 2017	1	2.86E-05	0.05
HDR	Hydroxymethylbutenyl diphosphate reductase	X	Booth et al. 2017	11	3.14E-04	0.52
Monoterpene pathway						
GPPS.lsu	Geranyl pyrophosphate synthase large subunit	4	Booth et al. 2017	8	2.43E-04	0.4
GPPS.ssu1	Geranyl pyrophosphate synthase small subunit 1	6	Booth et al. 2017	4	1.29E-04	0.21
GPPS.ssu2	Geranyl pyrophosphate synthase small subunit 2	ND	Booth et al. 2017	-	-	-

CsTPS1SK/CsTPS14CT	(-)-limonene synthase	5	Günnewich et al. 2007	16	4.57E-04	0.76
CsTPS2SK	(+)-α-pinene synthase	ND	Günnewich et al. 2007	-	-	-
CsTPS3FN/CsTPS15CT	β-myrcene synthase	5	Booth et al. 2017	12	3.43E-04	0.57
CsTPS5FN	β-myrcene, (-)-α-pinene synthase	9	Booth et al. 2017	1	2.86E-05	0.05
CsTPS6FN	(E)-β-ocimene synthase	5	Booth et al. 2017	9	2.57E-04	0.43
CsTPS13PK	(Z)-β-ocimene synthase	ND	Booth et al. 2017	-	-	-
CsTPS30PK	β-myrcene synthase	5	Booth et al. 2017	9	2.57E-04	0.43
CsTPS33PK	α-terpinene, γ-terpinene synthase	5	Booth et al. 2017	37	1.06E-03	1.75
CsTPS37FN	Terpinolene synthase	ND	Livingston et al. 2019	-	-	-
CsTPS38FN	(E)-β-ocimene synthase	ND	Livingston et al. 2019	-	-	-
MEV pathway						
HMGS	Hydroxymethylglutaryl-CoA synthase	5	Booth et al. 2017	16	4.57E-04	0.76
HMGR1	Hydroxy-methylglutaryl-CoA reductase 1	X	Booth et al. 2017	5	1.57E-04	0.26
HMGR2	Hydroxy-methylglutaryl-CoA reductase 2	X	Booth et al. 2017	62	1.79E-03	2.95
MK	Mevalonate Kinase	2	Booth et al. 2017	15	4.29E-04	0.71
PMK	Phosphomevalonate Kinase	5	Booth et al. 2017	35	1.00E-03	1.65
MPDC	Mevalonate diphosphate decarboxylase	1	Booth et al. 2017	15	4.29E-04	0.71
IDI	Isopentenyl-diphosphate delta-isomerase	ND	Booth et al. 2017	-	-	-
Sesquiterpene pathway						
FPPS1	Farnesyl diphosphate synthase 1	4	Booth et al. 2017	8	2.43E-04	0.4
FPPS2	Farnesyl diphosphate synthase 2	6	Booth et al. 2017	3	8.57E-05	0.14
CsTPS4FN	Alloaromadendrene synthase	6	Booth et al. 2017	8	2.29E-04	0.38
CsTPS7FN	δ-selinene synthase	ND	Booth et al. 2017	-	-	-
CsTPS8FN	γ-eudesmol, valencene synthase	6	Booth et al. 2017	15	4.38E-04	0.72
CsTPS9FN	β-caryophyllene, α-humulene synthase	6	Booth et al. 2017	53	1.51E-03	2.5
CsTPS16CC	Germacrene B synthase	ND	Zager at al. 2019	-	-	-

CsTPS20CT	Hedycaryol synthase	6	Zager at al. 2019	38	1.10E-03	1.82
CsTPS18VF/CsTPS19BL	Nerolidol/linalool synthase	1	Zager at al. 2019	5	1.43E-04	0.24

[a] Includes 10kb before and after gene

[b] Window range 21k bp

[c] Whole genome nucleotide diversity = 6.0 x 10-4

ND: Not determined in cs10 v 2.0 genome

Discussion

Genomic diversity within cannabis

Numerous studies have investigated intra-plant somatic mutations in long-lived perennials, but few have examined this in annuals and this research is the first to look at intra-plant mutations in cannabis. (Diwan *et al.*, 2014; Plomion *et al.*, 2018; Hanlon, Otto and Aitken, 2019; Orr *et al.*, 2020). Many previous studies used molecular markers to investigate and calculate mutation rates, which are prone to miss more rare mutations that whole-genome sequencing (WGS) will capture. Some research on intra-plant genetic variation were unreliable as they used extrapolations from microsatellites or chlorophyll mutations, however, more recent research has been able uncover a more complete understanding due to advances and technologies becoming more accessible (Schoen and Schultz, 2019). For example, two studies using WGS on a long-lived Oak tree (*Quercus robur*) reported a much lower mutation rate than what was obtained from scaling up estimates from *Arabidopsis thaliana*. Based on this they hypothesized that perennials may have a mechanism to reduce the accumulation of mutations that is lacking in short lives annual species (Schmid-Siegert *et al.*, 2017; Plomion *et al.*, 2018). Likewise in mammals, it has been shown that mice have more mutations per cell division than humans, which is consistent with the hypothesis that short lived species have a higher mutation rates (Milholland *et al.*, 2017). Currently, somatic mutations are believed to be common in plants but the mutation rate, distribution, morphological effects, age or size influence, and the differences between annuals and perennials remains poorly understood (Schoen and Schultz, 2019).

Naturally, cannabis is an annual species where it lives until its flowers are pollinated and seeds are produced. This all occurs during a single season that can range from a few months to closer to a

year, and then it naturally dies. In contrast, cannabis plants maintained as mothers are artificially maintained in a perpetual vegetative state and replaced periodically using clonal propagaules that can extend their life span to several years or even decades. Based on Muller's ratchet, a decline in plant vigour is likely during this period due to the accumulation of deleterious mutations and the absence of recombination. The extended life-span of mother plants significantly increases this concern because each *de novo* deleterious variant bears the potential for both multiplicative and cumulative effects to genome instability, altered gene expression, molecular heterogeneity, tissue disintegration and vulnerability to stress (Govindaraju, Innan and Veitia, 2020). In other species, propagules that inherit deleterious mutation from a mother plant exhibited smaller leaves, reduced nutrient translocation capacity, degraded genetics, less rooting, lower plant vigour, and a decrease in growth (Wendling, Trueman and Xavier, 2014; Singh *et al.*, 2015; Orr *et al.*, 2020). Therefore, our results lead us to hypothesize that prolonging the lifespan of Cannabis plants and constantly pruning clonal cuttings is leading to the observed plant decline.

In this study, we called over 2 million nucleotide variants within a single cannabis plant. In comparison, the long-lived oak study only called 5,330 potential SNVs that had accumulated over many years (Schmid-Siegert *et al.*, 2017). Interestingly, the Napoleon Oak genome was less than 1% heterozygous while the cannabis genome is known to be highly heterozygous (estimated at 12.5 – 40.5%) and contains substantial amounts of repetitive elements (estimated at 70%) (Hurgobin *et al.*, 2020). We identified somatic mutations with deep sequencing (>50x depth of coverage) and used systematic filtering steps to reduce errors that may have occurred from next-generation sequencing (NGS) and incorrect mapping (Ajay *et al.*, 2011). A recent cannabis study sequenced 40 cannabis genomes and they reported an average of 12.8 million SNPs+Indels for dispensary grade cannabis (Type I and Type II plants) which equated to a variant every ~73 bases

while our results equated to a variants every ~425 bases (McKernan *et al.*, 2020). Although our variants were six-fold less than the average found in the other study, partly due to clonal origin of our samples, they still represent a substantial quantity of variants that contain the ability to interfere with the stability and quality of the plant.

One might assume that calculating the ratio of transition to transversion from three samples derived from one individual mother plant is less interesting from a statistical point of view, but the comparison of this ratio with other cannabis research and other plant species was appealing. The Ts/Tv ratio found in this study (1.88) was similar to a recent study (Soorni *et al.*, 2017) where the average Ts/Tv ratio for 69 cannabis individuals was 1.65, showing an intriguing level of intra-plant genomic diversity. As the Ts/Tv ratio is comparable, we can verify that the value is within an expected range, which reduced the chances of high false positives or bias. Compared to other plant species, this ratio is similar to oil palm (1.67) and significantly higher than maize (1.02) (Batley *et al.*, 2003; Pootakham *et al.*, 2015).

Variants across intra-plant genome

We expected to witness a systematically hierarchical nature of mutation accumulation from the bottom to the top, as was seen by Diwan *et al* (2014) in their research on genome differences within Yoshino cherry (*Prunus* × *yedoensis*) and Japanese beech (*Fagus crenata*) trees. In this study, we witnessed a minor drop in total variants between the bottom and middle compared to the middle to top. Although we didn't observe a uniform increase, the uppermost sample was the more genetically distant from the bottom than was the middle, as seen by other previous intra-plant studies (Schmid-Siegert *et al.*, 2017; Plomion *et al.*, 2018; Hanlon, Otto and Aitken, 2019). Additionally, the primary analytical focus in this study was the middle and top variants because

they represent *de novo* mutations while the bottom section corresponds to the oldest growth tissue where, in theory, the mutations counts should be the lowest.

While many uncertainties remain about somatic mutation in plants, our results demonstrating significant genetic variation may be explained by a few possibilities 1) environmental factors 2) long-term pruning from the top 3) difference between perennials and annuals 4) small sample size. Environmental factors have been well studied and are known to impact mutation rates, especially during shock or stress (Gill *et al.*, 1995). Thus, any changes to the environment during the growth and maintenance of the mother plant, such as light or temperature stress, may have altered the mutation rates and contributed to their accumulation over time. Next, due to the nature of cannabis mother plants, they are consistently apically pruned and clonal propagules are taken for propagation. This practice has a direct impact on the balance between auxin and cytokinin levels, promotes rapid growth from originally inhibited axillary bud sites, and as a result could impact the rate of mutations within the plant (Prusinkiewicz *et al.*, 2009; Schaller, Bishopp and Kieber, 2015). Recently, next generation sequencing (NGS) studies identified major differences in the rate of mutations between perennials and annuals species, with perennials having a significantly lower rate than expected, suggesting they may possess a mechanism to suppress the rate (Schoen and Schultz, 2019). These findings are consistent with our results that found a relatively high number of mutations and lead us to believe that annuals contain a more severe and sporadic mutation rate such that prolonging their normal lifespan may lead to a substantial mutation load with potentially deleterious effects. Lastly, due to the expensive costs of full genome sequencing and intention to simply identify intra-plant variations, only three positions (i.e. bottom, middle, and top) were sequenced with the ability to call ultra-rare mutations. As such, it is unknown if the degree of diversity observed in this study is representative of the species. Further work is needed using

larger sample sizes across multiple genotypes in different environments to gain a better understanding of this phenomenon.

Impact on cannabinoid and terpene synthase

We investigated the impact that these mutations may have had on both the cannabinoid and terpene pathways because of their medical and product quality importance (Hanuš and Hod, 2020; Singh, Bilichak and Kovalchuk, 2020). First, we examined mutations that were categorized as high impact from analysis using SnpEff (Cingolani *et al.*, 2012). These mutations are known to have a large impact on protein production, protein truncation, loss of function or trigger nonsense mediated decay (Cingolani *et al.*, 2012). Although we didn't identify any high impact mutations in the cannabinoid and terpene synthase genes, the analysis revealed a similar arrangement of the distribution of mutations from the total number of variants. Interestingly, the top sample contained an even larger total percentage by ~6% while the middle had 3% less and the bottom had a less than 1% difference. This further supported of the notion that apical regions are more genetically distant than the basal regions.

Furthermore, we examined mutations in these pathway genes that fell under the category of *moderate*, *low*, or *modifier* which relates to non-disruptive changes to protein effectiveness, mostly harmless or unlikely to change protein performance, and non-coding variants where predictions are difficult or there is no evidence of impact, respectively. Our analysis revealed that all discovered variants fell under the modifier category and encouraged the assessment of nucleotide diversity (π) with a 20kb scope of critical genes for both pathways. Most of the 44 genes investigated had a similar or lower value than the average nucleotide diversity (π), however, four major genes had values over double. Two of these are a part of terpene production (HMGR2 and CsTPS9FN) and the other two are involved in the cannabinoid pathway (OLS and CBDAS).

HMGR2 role is to convert acetyl-CoA into mevalonic acid which undergoes a few more steps to produce the synthesis of both γ-eudesmol or β-caryophyllene terpenes. CsTPS9FN is the last enzyme necessary to produce the β-caryophyllene terpene. OLS and CBDAS were particularly intriguing because OLS is an essential enzyme for the whole cannabinoid pathway as it converts hexanoyl-CoA into olivetolic acid which then converts into CBGA, a precursor to many well-known cannabinoids (i.e. THCA, CBDA, CBCA). CBDAS is the last enzyme required to convert CBGA into CBDA and this is especially important as its the main cannabinoid used to treat various health concerns (Maroon and Bost, 2018). Both critical enzymes revealed a more than double nucleotide diversity (π) above average and as a result could indicate initial signs of decay in the cannabinoid pathway.

Mother plants are usually selected through large scale, costly, screening programs, and marketed as strains with unique properties. The alteration of genes over time represents a significant challenge in long-term batch to batch consistency. Also, cannabis used for medicinal purposes must ensure a consistent product that provides the appropriate properties and quality that are necessary for treatments. Thus, a greater insight into somatic mutations may enable new or superior procedures to assist the preservation of elite cultivars in clonally propagated plants. Overall, our research highlights an important phenomenon related to maintaining elite genetics and could provide an underlying mechanism for the decay of cannabinoids and plant vigour that has been anecdotally observed (Cannabis growers, personal communication, 2020).

Conclusion

The findings in this study demonstrate that the genetic diversity exists with a single cannabis plant and the genetic mosaicism hypothesis applies to *C. sativa.* This study is the first to investigate the existence of this phenomenon in cannabis plants and the potential consequences from accumulating somatic mutations in an artificially prolonged annual species. As cannabis normally lives for ~3-6 months, this process likely enables an unknown, but manageable, amount of somatic mutations to accumulate. Currently, somatic mutations in plants have many uncertainties remaining, but due to modern genetic technologies and more affordable whole-genome sequencing, there has been more contributions with higher degrees of accuracy and precision on this topic. From a practical standpoint, this significantly benefits the cannabis industry as understanding this phenomenon will help establish best practices for maintaining mother plants to minimize, slow or prevent the accumulation of mutations. Based on these data, we advocate replacing mother plants using cuttings from the basal portion of the plant and discourage excessively extending the life of a mother plant. Additionally, important genetics should be preserved using cryopreservation techniques where the original genetic profile can be maintained and accessed indefinitely (Uchendu *et al.*, 2019). The research here provides a concrete basis for cannabis mutation research. However, the current study lacked different cultivars, generational data, mutation rates and multiple biological replicates. Thus, future research will be necessary to enhance and solidify our understanding of somatic mutations and the mutagenic potential that exists within a cannabis mother plant.

Materials & Methods

Plant material

A high-THCA mother plant of *Cannabis sativa* cv. "Honey Banana" (15-20% THC; <1% CBD), was grown indoors at BrantMed Inc. in Ontario, Canada. The plant was grown in nutrient rich soil with regular feeding of a full nutrient solution developed for vegetative growth adjusted to ~6.5 pH. The proprietary nutrient solution was relatively high in nitrogen, with moderate levels of phosphorous, potassium (NPK) and micronutrients, administered every 3-5 days as needed. The plant was grown from seed and transplanted to larger pots as needed, until reaching the 76L (20gal) pot where it remained until this study was conducted. The environmental conditions were maintained at 20-25°C and 55-65% relative humidity using BrantMed Inc LED lighting (Grow Light E1-300W) under long day photoperiods at a 18:6 hour light:dark cycle to maintain the mother plant in an indefinitely vegetative state. These broad "white" spectrum lights (4.2% red 650-670 nm) with 300-watts provided a photosynthesis photon flux of >440 µmol/s and came with a PAR photon efficacy of 2.2 µmol/J. Samples were removed when the plant had reached an age of approximately 1.5 years. We isolated ~2.5cm of fresh stem tissue from three location at 59cm, 151cm, and 226cm which represented the bottom, middle and top samples, respectively (Figure 1A). The samples were frozen and stored in a freezer until DNA extraction.

DNA extraction and whole genome resequencing

Frozen stem tissues were ground using a Qiagen TissueLyser. DNA was extracted from approximately 100 mg of ground tissue using the Qiagen Plant DNeasy Mini Kit according to the manufacturer's protocol. DNA was quantified on a NanoDrop spectrophotometer and on a Qubit fluorometer. Illumina Paired-End libraries were constructed for three DNA samples using the

Illumina Tru-seq DNA Library Prep Kit (Illumina, San Diego CA, USA) following the manufacturer's instructions. The quality of DNA library was verified on an Agilent Bioanalyzer with a High Sensitivity DNA chip. The sequencing was performed on an Illumina NovaSeq 6000 platform at the McGill University-Génome Québec Innovation Center in Montreal, QC, Canada generating >1 billion 150-bp paired-end reads. to provide >50x death of coverage by sequencing reads against the mapped the public domain *C. sativa* reference genome (cs10 v. 2.0; Hurgobin *et al.*, 2020)

Bioinformatic data analysis

Illumina paired-end reads were processed using Fast-WGS bioinformatics pipeline (Torkamaneh et al., 2018). In summary, the reads were mapped against cannabis reference genome (cs10 v. 2.0; GenBank Accession No. GCA_900626175.2; Hurgobin *et al.*, 2020) using BWA-MEM (Li, 2013). The nucleotide variants were called using Platypus (Rimmer *et al.*, 2014). In general, we removed variants if: 1) they had more than two alleles, 2) an allele was not supported by reads on both strands, 3) the overall quality (QUAL) score was <32, 4) the mapping quality (MQ) score was <20, 5) read depth (minNR) was <10 and 6) the number of reads supporting variant (minNV) was <10. Functional annotation of nucleotide variation was performed by SnpEFF and SnpSift (Cingolani *et al.*, 2012) using a customized reference built using cannabis reference genome annotation file downloaded from NCBI. To obtain the description of genes with large impact, we used the NCBI's protein table for cannabis sativa database. The gene ontology (GO) analysis was done using the singular enrichment analysis (SEA) method implemented in agri-GO (Du *et al.*, 2010).

Diversity, LD and clustering analysis

Nucleotide diversity (π) was calculated using VCFtools (Danecek *et al.*, 2011), with a window of 20K bp on the full dataset. An average π across all windowed calculations was used to obtain a genome-wide average π. A neighbor-joining phylogenetic tree was constructed using full dataset in Tassel 5.0 (Bradbury *et al.*, 2007). The linkage disequilibrium (LD) decay was determined using PopLDdecay version 3.40 Beta (Zhang *et al.*, 2019). IGV 2.8 (Robinson *et al.*, 2011) was utilized to display the distribution of variants within the three samples which was produced with the indexed version of the VCF file.

Acknowledgments

This work was conducted as part of a collaborative research project funded by Brantmed Inc. and NSERC Alliance [#ALLRP555969 -20 to A.M.P.J and D.T.]. The authors wish to thank BrantMed Inc. for supporting this project.

Author contributions

DT, AMPJ, and KA conceived the project. DT carried out the WGS and variant calling. KA performed bioinformatics analysis. KA, DT, and AMPJ contributed to writing the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

Ajay, S. S. *et al.* (2011) 'Accurate and comprehensive sequencing of personal genomes', *Genome Research*, 21(9), pp. 1498–1505. doi: 10.1101/gr.123638.111.

van Amerongen, G. *et al.* (2018) 'Effects on Spasticity and Neuropathic Pain of an Oral Formulation of Δ9-tetrahydrocannabinol in Patients With Progressive Multiple Sclerosis', *Clinical Therapeutics*, 40(9), pp. 1467–1482. doi: 10.1016/j.clinthera.2017.01.016.

van Bakel, H. *et al.* (2011) 'The draft genome and transcriptome of Cannabis sativa', p. 18.

Batley, J. *et al.* (2003) 'Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data1', *Plant Physiology*, 132(1), pp. 84–91.

Booth, J. K., Page, J. E. and Bohlmann, J. (2017) 'Terpene synthases from Cannabis sativa', *PLoS ONE*, 12(3), pp. e0173911–e0173911. doi: 10.1371/journal.pone.0173911.

Bradbury, P. J. *et al.* (2007) 'TASSEL: software for association mapping of complex traits in diverse samples', *Bioinformatics*, 23(19), pp. 2633–2635. doi: 10.1093/bioinformatics/btm308.

Burnstein, J. (2019) 'What is Genetic Drift and How Does it Apply to Cloning and Micro-propagation?', *Cannabis & Tech Today*, 30 September. Available at: https://cannatechtoday.com/genetic-drift/ (Accessed: 16 June 2020).

Carrier, G. *et al.* (2012) 'Transposable Elements Are a Major Cause of Somatic Polymorphism in Vitis vinifera L.', *PLoS ONE*. Edited by C. Schönbach, 7(3), p. e32973. doi: 10.1371/journal.pone.0032973.

Chandra, S., Lata, H. and ElSohly, M. A. (2017) *Cannabis Sativa L. - Botany and Biotechnology*. Cham, SWITZERLAND: Springer International Publishing AG. Available at: http://ebookcentral.proquest.com/lib/uoguelph/detail.action?docID=4865024 (Accessed: 10 February 2021).

Cingolani, P. *et al.* (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ; iso-2; iso-3', *Fly*, 6(2), pp. 80–92. doi: 10.4161/fly.19695.

Cruzan, M. B., Streisfeld, M. A. and Schwoch, J. A. (2018) *Fitness Effects of Somatic Mutations Accumulating during Vegetative Growth*. preprint. Evolutionary Biology. doi: 10.1101/392175.

Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

Devinsky, O. *et al.* (2014) 'Cannabidiol: Pharmacology and potential therapeutic role in epilepsy and other neuropsychiatric disorders', *Epilepsia*, 55(6), pp. 791–802. doi: https://doi.org/10.1111/epi.12631.

Diwan, D., Komazaki, S., Suzuki, M., Nemoto, N., Aita, T., Satake, A., & Nishigaki, K. (2014). Systematic genome sequence differences among leaf cells within individual trees. BMC genomics, 15, 142. https://doi.org/10.1186/1471-2164-15-142

Du, Z. *et al.* (2010) 'agriGO: a GO analysis toolkit for the agricultural community', *Nucleic Acids Research*, 38(suppl_2), pp. W64–W70. doi: 10.1093/nar/gkq310.

Felsenstein, J. (1974) 'The evolutionary advantage of recombination', *Genetics*, 78(2), pp. 737–756. doi: 10.1093/genetics/78.2.737.

Gagne, S. J. *et al.* (2012) 'Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides', *Proceedings of the National Academy of Sciences*, 109(31), pp. 12811–12816. doi: 10.1073/pnas.1200330109.

Gill, D. E. *et al.* (1995) 'Genetic Mosaicism in Plants and Clonal Animals', p. 23.

Govindaraju, D. R., Innan, H. and Veitia, R. A. (2020) 'The Muller's Ratchet and Aging', *Trends in Genetics*, 36(6), pp. 395–402. doi: 10.1016/j.tig.2020.02.004.

Günnewich, N. *et al.* (2007) 'Functional Expression and Characterization of Trichome-Specific (-)-Limonene Synthase and (+)-α-Pinene Synthase from Cannabis sativa', *Natural Product Communications*, 2(3), p. 1934578X0700200301. doi: 10.1177/1934578X0700200301.

Hanlon, V.C.T, Otto, S. P. and Aitken, S. N. (2019) 'Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis', *Evolution Letters*, 3, pp. 348–358. doi: 10.1002/evl3.121a.

Hanlon, Vincent C. T., Otto, S. P. and Aitken, S. N. (2019) 'Somatic mutations substantially increase the per-generation mutation rate in the conifer Picea sitchensis', *Evolution Letters*, 3(4), pp. 348–358. doi: https://doi.org/10.1002/evl3.121.

Hanuš, L. O. and Hod, Y. (2020) 'Terpenes/Terpenoids in Cannabis: Are They Important?', *Medical Cannabis and Cannabinoids*, 3(1), pp. 25–60. doi: 10.1159/000509733.

Hillig, K. W. (2005) 'Genetic evidence for speciation in Cannabis (Cannabaceae)', *Genetic Resources and Crop Evolution*, 52(2), pp. 161–180. doi: 10.1007/s10722-003-4452-y.

Hurgobin, B. *et al.* (2020) 'Recent advances in *Cannabis sativa* genomics research', *New Phytologist*, p. nph.17140. doi: 10.1111/nph.17140.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', 00, pp. 1–3.

Livingston, S. J. *et al.* (2020) 'Cannabis glandular trichomes alter morphology and metabolite content during flower maturation', *The Plant Journal*, 101(1), pp. 37–56. doi: https://doi.org/10.1111/tpj.14516.

Maroon, J. and Bost, J. (2018) 'Review of the neurological benefits of phytocannabinoids', *Surgical Neurology International*, 9. doi: 10.4103/sni.sni_45_18.

Marshall, D. R. and Allard, R. W. (1970) 'Isozyme polymorphisms in natural populations of Avena fatua and A. barbata', *Heredity*, 25(3), pp. 373–382. doi: 10.1038/hdy.1970.38.

McKernan, K. J. *et al.* (2020) 'Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes', *bioRxiv*, p. 2020.01.03.894428. doi: 10.1101/2020.01.03.894428.

McKey, D. *et al.* (2010) 'The evolutionary ecology of clonally propagated domesticated plants: Tansley review', *New Phytologist*, 186(2), pp. 318–332. doi: 10.1111/j.1469-8137.2010.03210.x.

Milholland, B. *et al.* (2017) 'Differences between germline and somatic mutation rates in humans and mice', *Nature Communications*, 8(1), p. 15183. doi: 10.1038/ncomms15183.

Muller, H. J. (1932) 'Some Genetic Aspects of Sex', *The American Naturalist*, 66(703), pp. 118–138. doi: 10.1086/280418.

Muller, H. J. (1964) 'The relation of recombination to mutational advance', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1), pp. 2–9. doi: 10.1016/0027-5107(64)90047-8.

Orr, A. J. *et al.* (2020) 'A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant', *Proceedings of the Royal Society B: Biological Sciences*, 287(1922), p. 20192364. doi: 10.1098/rspb.2019.2364.

Page, J. E. and Stout, J. M. (2019) 'Applicant: NATIONAL RESEARCH COUNCIL', *US10364416B2*, p. 32.

Plomion, C. *et al.* (2018a) 'Oak genome reveals facets of long lifespan', *Nature Plants*, 4(7), pp. 440–452. doi: 10.1038/s41477-018-0172-3.

Plomion, C. *et al.* (2018b) 'Oak genome reveals facets of long lifespan', *Nature Plants*, 4(7), pp. 440–452. doi: 10.1038/s41477-018-0172-3.

Pootakham, W. *et al.* (2015) 'Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS)', *Genomics*, 105(5), pp. 288–295. doi: 10.1016/j.ygeno.2015.02.002.

Prusinkiewicz, P. *et al.* (2009) 'Control of bud activation by an auxin transport switch', *Proceedings of the National Academy of Sciences*, 106(41), pp. 17431–17436. doi: 10.1073/pnas.0906696106.

Rimmer, A. *et al.* (2014) 'Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications', *Nature Genetics*, 46(8), pp. 912–918. doi: 10.1038/ng.3036.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26. doi: 10.1038/nbt.1754.

Schaller, G. E., Bishopp, A. and Kieber, J. J. (2015) 'The Yin-Yang of Hormones: Cytokinin and Auxin Interactions in Plant Development', *The Plant Cell*, 27(1), pp. 44–63. doi: 10.1105/tpc.114.133595.

Schmid-Siegert, E. *et al.* (2017) 'Low number of fixed somatic mutations in a long-lived oak tree', *Nature Plants*, 3(12), pp. 926–929. doi: 10.1038/s41477-017-0066-9.

Schoen, D. J. and Schultz, S. T. (2019) 'Somatic Mutation and Evolution in Plants', *Annual Review of Ecology, Evolution, and Systematics*, 50(1), pp. 49–73. doi: 10.1146/annurev-ecolsys-110218-024955.

Singh, A., Bilichak, A. and Kovalchuk, I. (2020) 'The genetics of Cannabis – genomic variations of key synthases and their effect on cannabinoids content', *Genome*, p. gen-2020-0087. doi: 10.1139/gen-2020-0087.

Singh, A. K. *et al.* (2015) 'Parental Age Affects Somatic Mutation Rates in the Progeny of Flowering Plants', *Plant Physiology*, 168(1), pp. 247–257. doi: 10.1104/pp.15.00291.

Sirikantaramas, S. *et al.* (2004) 'The Gene Controlling Marijuana Psychoactivity: Molecular Cloning and Heterologous Expression of Δ1-Tetrahydracannabinolic Acid Synthase from Cannabis Sativa L.', *Journal of Biological Chemistry*, 279(38), pp. 39767–39774. doi: 10.1074/jbc.M403693200.

Soorni, A. *et al.* (2017) 'Assessment of Genetic Diversity and Population Structure in Iranian Cannabis Germplasm', *Scientific Reports*, 7(1), p. 15668. doi: 10.1038/s41598-017-15816-5.

Stout, J. M. *et al.* (2012) 'The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes', *The Plant Journal*, 71(3), pp. 353–365. doi: https://doi.org/10.1111/j.1365-313X.2012.04949.x.

Taura, F. *et al.* (2007) 'Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa', *FEBS Letters*, 581(16), pp. 2929–2934. doi: https://doi.org/10.1016/j.febslet.2007.05.043.

Taura, F. *et al.* (2009) 'Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway', *FEBS Letters*, 583(12), pp. 2061–2066. doi: https://doi.org/10.1016/j.febslet.2009.05.024.

Torkamaneh, D. *et al.* (2018) 'Comprehensive description of genomewide nucleotide and structural variation in short-season soya bean', *Plant Biotechnology Journal*, 16(3), pp. 749–759. doi: 10.1111/pbi.12825.

Uchendu, E. *et al.* (2019) 'Cryopreservation of Shoot Tips of Elite Cultivars of Cannabis sativa L. by Droplet Vitrification', *Medical Cannabis and Cannabinoids*, 2(1), pp. 29–34. doi: 10.1159/000496869.

Watt, G. and Karl, T. (2017) 'In vivo Evidence for Therapeutic Properties of Cannabidiol (CBD) for Alzheimer's Disease', *Frontiers in Pharmacology*, 8. doi: 10.3389/fphar.2017.00020.

Wendling, I., Trueman, S. J. and Xavier, A. (2014) 'Maturation and related aspects in clonal forestry—part II: reinvigoration, rejuvenation and juvenility maintenance', *New Forests*, 45(4), pp. 473–486. doi: 10.1007/s11056-014-9415-y.

Zager, J. J. *et al.* (2019) 'Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis', *Plant Physiology*, 180(4), pp. 1877–1897. doi: 10.1104/pp.18.01506.

Zhang, C. *et al.* (2019) 'PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files', *Bioinformatics (Oxford, England)*, 35(10), pp. 1786–1788. doi: 10.1093/bioinformatics/bty875.