Plant Direct

WILEY | American Society of Plant Biologists SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

RESEARCH ARTICLE OPEN ACCESS

The Structure of the Chemotype Determining Locus in *Cannabis sativa*

Keith Allen[1] | Anthony Torres[2] | Reginald Gaudino[2]

[1]Front Range Biosciences, Lafayette, Colorado, USA | [2]Front Range Biosciences/Cannabis Research Institute at Discovery Partners Institute, University of Illinois, Chicago, Illinois, USA

Correspondence: Keith Allen (kallen83@gmail.com)

Received: 1 July 2025 | **Revised:** 26 March 2026 | **Accepted:** 31 March 2026

ABSTRACT

The chemical phenotype (chemotype) of *Cannabis sativa* is defined by the ratio of cannabidiolic acid (CBDA) to Δ9-tetrahydrocannabinolic acid (THCA). Although the Mendelian segregation of these traits suggests a single-locus biallelic system, recent sequencing and phylogenetic evidence indicate they are encoded by two distinct, tightly linked genes. The precise genomic architecture of this region, known as the *B* locus, has remained poorly defined. In this study, we analyzed recently released high-quality *Cannabis* reference genomes to resolve the structure of the *B* locus. Our results demonstrate that this region functions as a supergene, characterized by suppressed recombination that facilitates Mendelian-like switching between phenotypic states. Comparative genomic analysis reveals substantial structural polymorphism within the locus, including significant variations in gene copy number and large-scale insertions/deletions (indels). Furthermore, we functionally characterized three previously unstudied members of the cannabinoid oxidocyclase family. We find that these enzymes primarily catalyze the production of cannabichromenic acid (CBCA), reinforcing the model that cannabinoid profile is dictated specifically by the presence and expression of THCAS or CBDAS. Finally, we mapped the expression profile of the entire berberine bridge enzyme (BBE) family, identifying widespread expression across plant tissues, including in glandular trichomes. Collectively, these findings resolve the genomic architecture of the *B* locus, clarify the enzymatic basis of cannabinoid profile determination, and establish a framework for understanding the evolutionary maintenance of chemotype diversity in *C. sativa*.

1 | Introduction

1.1 | Cannabinoid Diversity and Biological Significance

Cannabinoids, defined chemically as isoprenylated resorcinol polyketides, are best known in *Cannabis sativa L.*, where more than 120 have been identified (Wishart et al. 2024). The most abundant cannabinoids are Δ9-tetrahydrocannabinolic acid (THCA), which is primarily responsible for the intoxicating aspects of this plant, and cannabidiolic acid (CBDA). While not intoxicating, CBDA has a range of potential medical applications, including promising anticonvulsant properties (Pauli et al. 2020).

The common precursor to the cannabinoids, cannabigerolic acid (CBGA), is often detectable in flower samples and may reach relatively high concentrations in some cultivars. Of the less well-studied cannabinoids, cannabichromenic acid (CBCA) is the most abundant. While commonly detectable in medicinal or recreational samples, CBCA concentration in flower material is typically less than 0.5% of dry weight (ElSohly et al. 2016).

Abbreviations: BBE, berberine bridge enzyme; CBCA, cannabichromenic acid; CBCAS, cannabichromenic acid synthase; CBDA, cannabidiolic acid; CBDAS, cannabidiolic acid synthase; CBGA, cannabigerolic acid; THCA, Δ9-tetrahydrocannabinolic acid; THCAS, tetrahydrocannabinolic acid synthase.

Keith Allen and Anthony Torres contributed equivalently to this work.

1.2 | Chemotype Classification

The two most abundant cannabinoids, THCA and CBDA, define the chemotype of *Cannabis*. THC-dominant cultivars are classified as Type I, or drug-type. Cultivars with an approximately 1:1 ratio of THCA:CBDA are classified as Type II (intermediate). Finally, CBD-dominant cultivars are classified as Type III (hemp-type) (Toth et al. 2020). In this paper, we refer to these categories as THC-type, intermediate, and CBD-type.

In the United States, the Hemp Farming Act of 2018 defined CBD-type as containing less than 0.3% THC, creating a legal distinction between CBD- and THC-type *C. sativa*, which remains classified as a schedule I narcotic. This allows legal cultivation of CBD-type for fiber, seed protein production, or CBDA production. Preparations of CBDA have been approved by the FDA for the treatment of epileptic seizures (Reddy et al. 2023). There has been growing interest in the potential health benefits of CBDA, leading to increased demand and expanded CBD-type *Cannabis* cultivation.

1.3 | The Berberine Bridge Enzyme (BBE) Family

The cannabinoid oxidocyclases belong to the larger BBE family (Daniel et al. 2017). Phylogenetic analysis shows that cannabinoid oxidocyclase genes are specific to *C. sativa*. This family arose from a single progenitor gene in an ancestral plant rather than from an earlier duplication event prior to speciation (Van Velzen and Schranz 2021). Nearly all of the BBE genes in *C. sativa*, including the cannabinoid oxidocyclases, are found on chromosome 7, with the cannabinoid oxidocyclases in two clusters (Lynch et al. 2025).

Copy number variation is high in this family, though THCAS and CBDAS have not been observed in more than single copy (Van Velzen and Schranz 2021; Lynch et al. 2025). While many cannabinoids have been detected in *C. sativa*, the characterized enzymes account only for the most abundant molecules, along with a few cannabinoids for which no standards were available (Hanuš et al. 2016). Given the medicinal potential, understanding which enzymes in this family are active and which products they produce remains important.

In this study, we characterize three cannabinoid oxidocyclase enzymes from the poorly explored Clade C (Van Velzen and Schranz 2021). This work helps clarify the enzymes responsible for THCA production in CBD-type cultivars containing only an inactive THCAS, which is important for CBD-type farmers and breeders maintaining regulatory compliance.

1.4 | The *B* Locus and Chemotype Inheritance

The genes encoding CBDAS and THCAS are tightly linked genetically. Segregation of chemical phenotype is consistent with a simple Mendelian model involving codominant alleles at a single locus, which has been designated as the *B* locus (De Meijer et al. 2003). Intermediate chemotypes are interpreted as due to heterozygosity at this locus.

Initially, THCAS and CBDAS were thought to be alleles of the same gene. However, with more gene sequences and improved genome assemblies, it became clear that these are separate genes. They behave as a single locus because they are so tightly linked (van Bakel et al. 2011). Current understanding suggests that chemotype is determined by two tightly linked genes in repulsion (also referred to as a *trans* configuration). This means that in a heterozygote, one chromosome carries an active THCAS gene and an inactive CBDAS. The other chromosome has the opposite arrangement with an intact CBDAS gene and an inactive THCAS (Grassa et al. 2021; Weiblen et al. 2015). This arrangement would explain why assemblies from intermediate-type cultivars (Cannbio2, GCA_016165845.1 and Jamaican Lion, GCA_003660325.2) failed to assemble intact THCAS and CBDAS genes on the same contig.

The overall layout of the *B* locus has been described through a phylogenetic analysis (Van Velzen and Schranz 2021), and we use their terminology for the clades in this family throughout this paper. The genomic assemblies available at the time of that study were quite fragmented, with several splitting the *B* locus. The availability of fully phased diploid (haplotype-resolved) chromosome-scale "anchor" genomes (Lynch et al. 2025) now provides an opportunity to examine this locus in greater detail.

1.5 | The Supergene Framework

We find that the *B* locus has characteristics of a supergene. Supergenes are defined as genetic architectures where "alternative phenotypes in a balanced polymorphism segregate as if controlled by a single genetic locus, resulting from tight genetic linkage between multiple functional loci" (Thompson and Jiggins 2014). This definition encompasses supergenes with even a single coding sequence, focusing on the functional outcome rather than the number of genes involved. The key feature is suppression of recombination within the region, leading to inheritance of multiple genetic elements as a single unit.

2 | Results

2.1 | Arrangement of Cannabinoid Oxidocyclase Genes

Nearly all BBE gene family members are on chromosome 7. The cannabinoid oxidocyclase genes are found in two clusters. One cluster contains either THCAS or CBDAS plus uncharacterized genes designated as Clade C (Van Velzen and Schranz 2021). The second cluster is an array of cannabichromene synthase (CBCAS) genes. Both clusters also contain various pseudogenes and are interspersed with more distantly related BBE family members.

Previous studies showed that THCAS and CBDAS reside in a nonsyntenic region within a larger syntenic area (Weiblen et al. 2015; Van Velzen and Schranz 2021; Lynch et al. 2025). However, the exact size, structure, and gene content of this region remained unclear. To better characterize this region, we

used the genome comparison tool lastz (Harris 2007) to compare the region containing either THCAS or CBDAS between phased, chromosome-level genomes.

We analyzed five CBD-type cultivars: Boone County (BCM), BoAx (BOAX), Golden Redwood (GRM), Kompolti (KOMP), and YunMa (YMv2), and four THC-type cultivars: Ace High (AH3M), Northern Lights (NLv1), White Widow (WHW), and Sour Diesel (SODL).

2.2 | Definition and Boundaries of the *B* Locus

Figure 1A shows a dotplot comparison between the *B* locus region from the THC-type Sour Diesel (SODLa) and the CBD-type Boone County (BCMa). Note that this is just an example comparison, and there is considerable variability at this locus. The genomes of these cultivars are highly syntenic across the length of chromosome 7, but the region immediately surrounding either THCAS or CBDAS shows almost no similarity. We define the nonsyntenic region containing either the THCAS gene or the CBDAS gene as the *B* locus.

We mark the left side of the nonsyntenic region just to the right of a cationic amino acid transporter homolog (marked in red in Figure 1). This gene is present in one or two copies depending on cultivar. To the left of this point, there is still substantial synteny. We mark the right boundary near a BBE gene with approximately 45% identity to THCAS at the protein level (marked in green in the Figure 1). To the right of this boundary, there is substantial synteny, but between these two boundaries, the sequences have very little similarity to each other. These boundaries are shown as vertical dotted lines in Figure 1A.

2.3 | Gene Content and Size Variation

In BCMa, the only gene within the *B* locus is CBDAS. In contrast, SODLa contains THCAS, a Clade A3 pseudogene, a cluster of Clade B2 pseudogenes (discussed below), and a hypothetical CDS with no informative GenBank similarities. There is a substantial size difference between THC- and CBD-type B loci: SODLa spans 1.5 Mb, while BCMa is only 0.66 Mb.

Interestingly, the Kompolti (KOMPa) CBD-type genome shows a different pattern. The first half of the THC-type *B* locus is inserted at the beginning of the CBD-type *B* locus, including a copy of the Clade A3 pseudogene (Figure 1B). The similarity of this segment to THC-type sequence is evident in Figure 1B comparing KOMPa to SODLa. This segment is found in all the THC-type assemblies and in the CBD-type KOMP and GRMb genomes.

2.4 | Comparison Across Multiple Genomes

To assess the consistency of these patterns, we compared all available THC- and CBD-type cultivars. Figure 2 shows a comparison of the *B* locus region of THC-type cultivars compared to Sour Diesel, while Figure 3 shows CBD-type cultivars compared to BCMa. Despite substantial small-scale variability, the basic structure is consistent: larger regions in THC-type cultivars and two versions in CBD-type, one including a THC-type *B* locus segment with the Clade A3 pseudogene, the other without.

A diagrammatic representation of the *B* locus is shown in Figure 4.

2.5 | Repetitive Element Content

The genome data sets we analyzed included masking of repetitive elements (Lynch et al. 2025). Analysis reveals that 79.5% of SODLa chromosome 7 is covered by annotated repeat sequences, while the *B* locus itself is 92.8% covered. The median size for unmasked regions in the *B* locus is just 105 bp. Of 11 unique unmasked stretches longer than 1500 bp, five contain cannabinoid oxidocyclase genes (one being THCAS), and none of these 11 segments have matches in the corresponding BCMa genome.

2.6 | Heterozygosity at the *B* Locus

We assessed heterozygosity by comparing the two haplotype assemblies for each genome using lastz (Harris 2007). Results for THC- and CBD-type genomes are shown in Figures S1 and S2, respectively, with a graphical summary in Figure 5.

Heterozygosity varies substantially across cultivars (Figure 5). Some cultivars, like WHW and KOMP, appear nearly inbred with highly similar haplotypes. In contrast, others like GRM and SODL show highly divergent chromosomes. Variation in the number of genes in the Clade C cluster is quite evident in these plots, particularly in WHW and AH3M.

Kompolti is homozygous for the variant *B* locus containing the THC-type segment, while GRM is heterozygous for this same variant. Figure S3 shows GRM is also heterozygous for a second insertion (0.64 Mb in length) located to the left of the *B* locus. This insertion is shared with YMv2a (Figure 3B,C; in Figure 3B, this insertion is between 7.5MBase and 8.2MBase on the Y-axis). Notably, this insertion is also found in SODLb in the same position (Figure S1A), indicating it occurs in both THC- and CBD-type cultivars.

2.7 | The BBE Gene Family in *C. sativa*

We identified 20 intact genes from the Jamaican Lion genome using BLAST queries with the THCAS protein sequence as query. Where necessary to retrieve complete reading frames, we used the exonerate alignment tool with a protein2genome model (Slater and Birney 2005). All of these are single-exon genes encoding proteins of approximately 545 amino acids.

The first panel of Figure 6 shows a phylogenetic tree of the *C. sativa* BBE family. We define the cannabinoid oxidocyclases as sequences with greater than 80% identity at the protein level to THCAS, and these are marked in blue in Figure 6. The cannabinoid oxidocyclase genes cluster tightly together, indicative of their relatively recent descent from a single progenitor gene

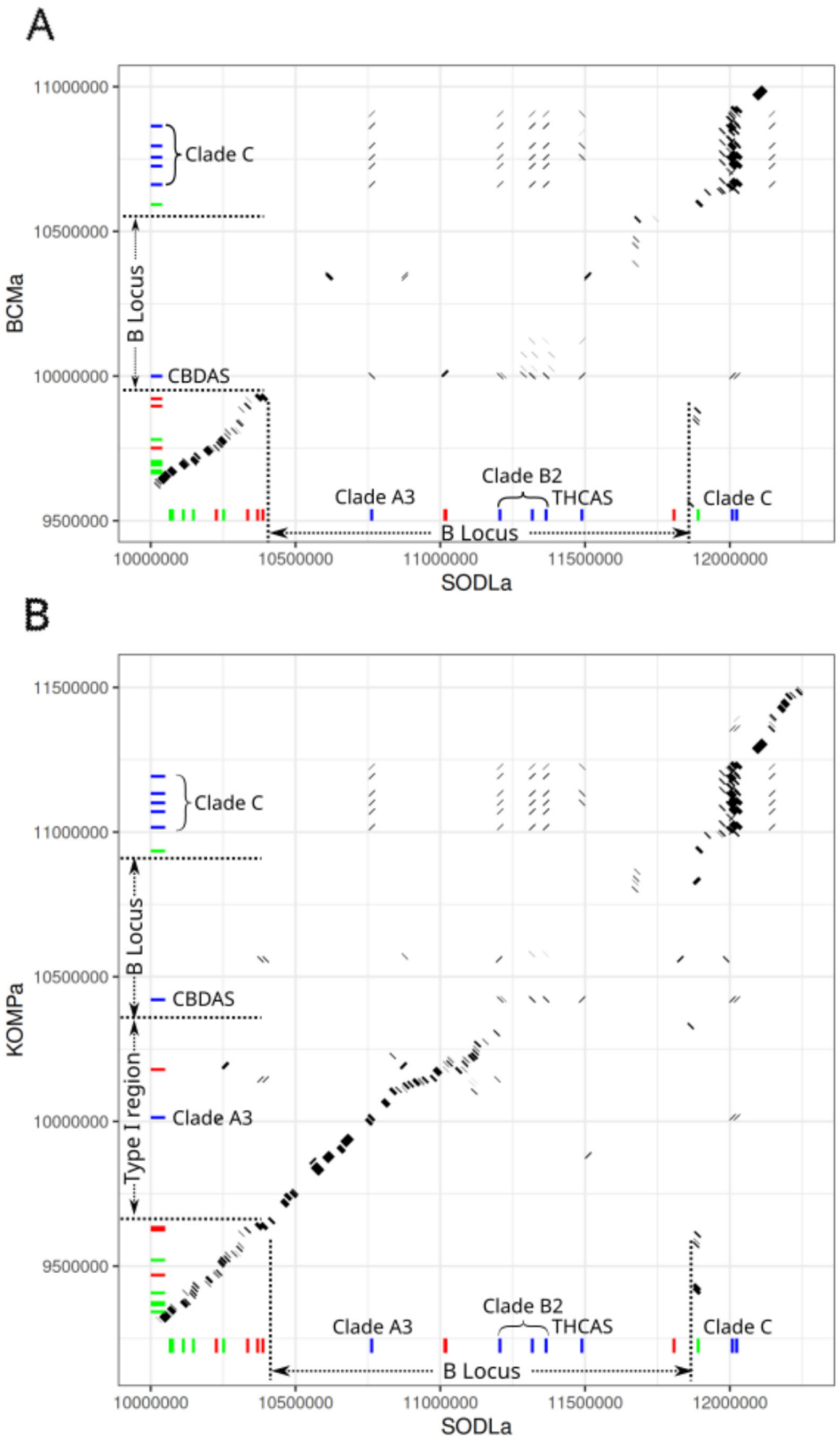

FIGURE 1 | (A) Dotplot comparison of THC-type Sour Diesel (SODLa) and CBD-type Boone County (BCMa) B locus regions. (B) Dotplot comparison of THC-type Sour Diesel (SODLa) and CBD-type Kompolti (KOMPa) B locus regions. Cannabinoid oxidocyclase genes are shown in blue; other members of the BBE family are shown in green while unrelated genes are shown in red.

(Van Velzen and Schranz 2021). The remainder of the BBE family (marked in red in Figure 6) is designated by percentage protein-level identity to THCAS. One gene shows 65% identity to THCAS, while most of the family clusters around 45% identity. For context, this is comparable to the distance seen to BBE family members in *Arabidopsis*.

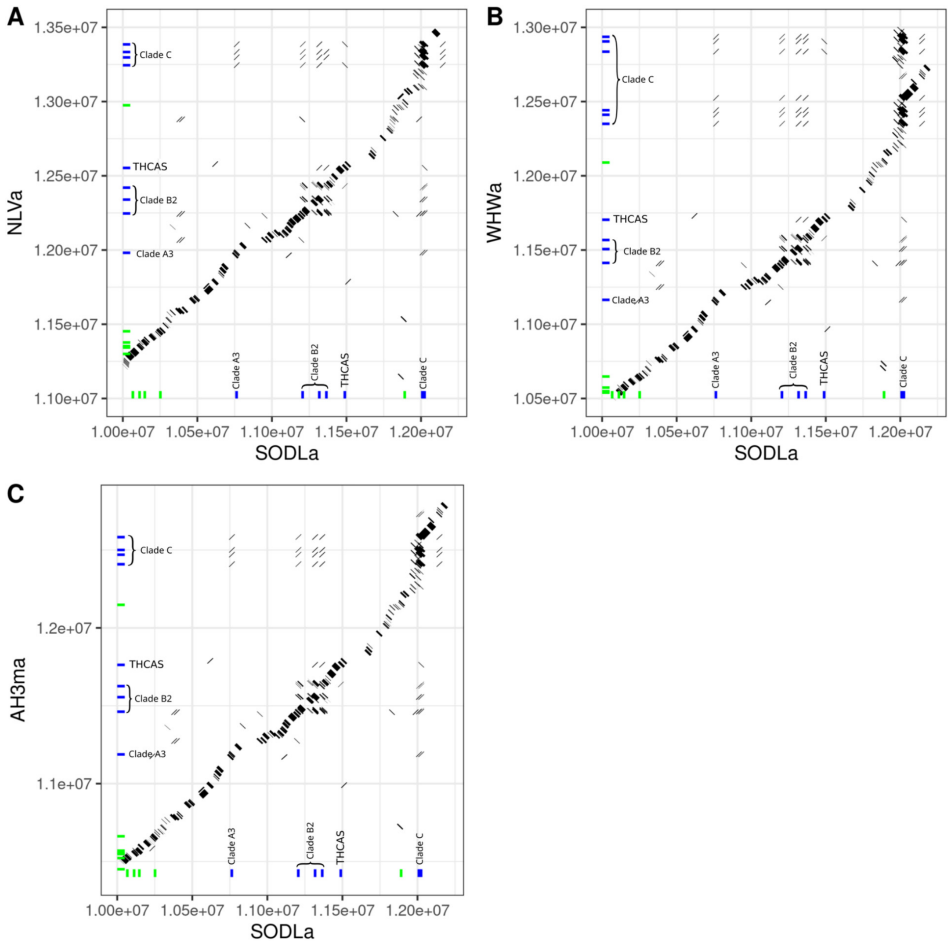

FIGURE 2 | Comparison of THC-type cultivars to Sour Diesel (SODLa). (A) Northern Lights (NLVa), (B) White Widow (WHWa), (C) Ace High (AH3ma). Cannabinoid oxidocyclase are shown as blue ticks, while other BBE family members are shown in green.

2.8 | Expression Patterns in Different Tissues

We assessed BBE family gene expression using publicly available RNA-Seq data from the intermediate-type cultivar Cannbio2 (Braich et al. 2019). Reads were mapped to the Jamaican Lion genome because at the time of analysis, this was the only assembly including intact copies of both CBDAS and THCAS genes. Results are shown in Figure 6.

THCAS and CBDAS dominate gene expression in trichomes and show much lower expression in other tissues, though expression is never zero even in roots. In female leaves, CBDAS and THCAS account for most of the BBE family gene expression. In male leaves, a single clade of three unknown BBE genes accounts for most gene expression (Figure 6).

Several uncharacterized BBE genes are also highly expressed in trichomes, specifically genes designated 43.503, 45.149, and 46.916. The most highly expressed BBE genes are all proximally close to the locus containing either THCAS or CBDAS. While the cannabinoid oxidocyclase genes show the lowest expression in roots, many other BBE genes show their highest expression in roots. Several BBE genes show dramatically different expression in male versus female leaves, though functional characterization would be required to understand potential roles.

2.9 | Clade A3 Gene Variants

Clade A3 was previously described (Van Velzen and Schranz 2021) with a full-length coding sequence from cultivar Finola and two nonfunctional pseudo-gene copies from THC-type cultivars: Purple Kush and Jamaican Lion. Clade A3 sequences are characterized by a duplication of the third codon. With improved genomes and published RNA-Seq data, we can provide further detail on three distinct variants.

2.10 | Clade A3.1: Expressed but Nonfunctional

Clade A3.1 is found in most THC-type cultivars we examined. In addition to the duplication of the second exon, this gene carries a nine-base deletion (relative to the THCAS sequence) as well as two in-frame stop codons. These changes remove many active site residues, making it unlikely that this encodes a functional enzyme.

Despite these mutations, we found ample evidence for expression. Figure 6 shows expression of this gene in Cannbio2 flowers. Figure 7 shows expression of the minor cannabinoid oxidocyclase genes in 14 published trichome expression libraries (Booth et al. 2020; Zager et al. 2019). Based on normalized read counts, the CladeA3.1 gene is expressed at least an order of

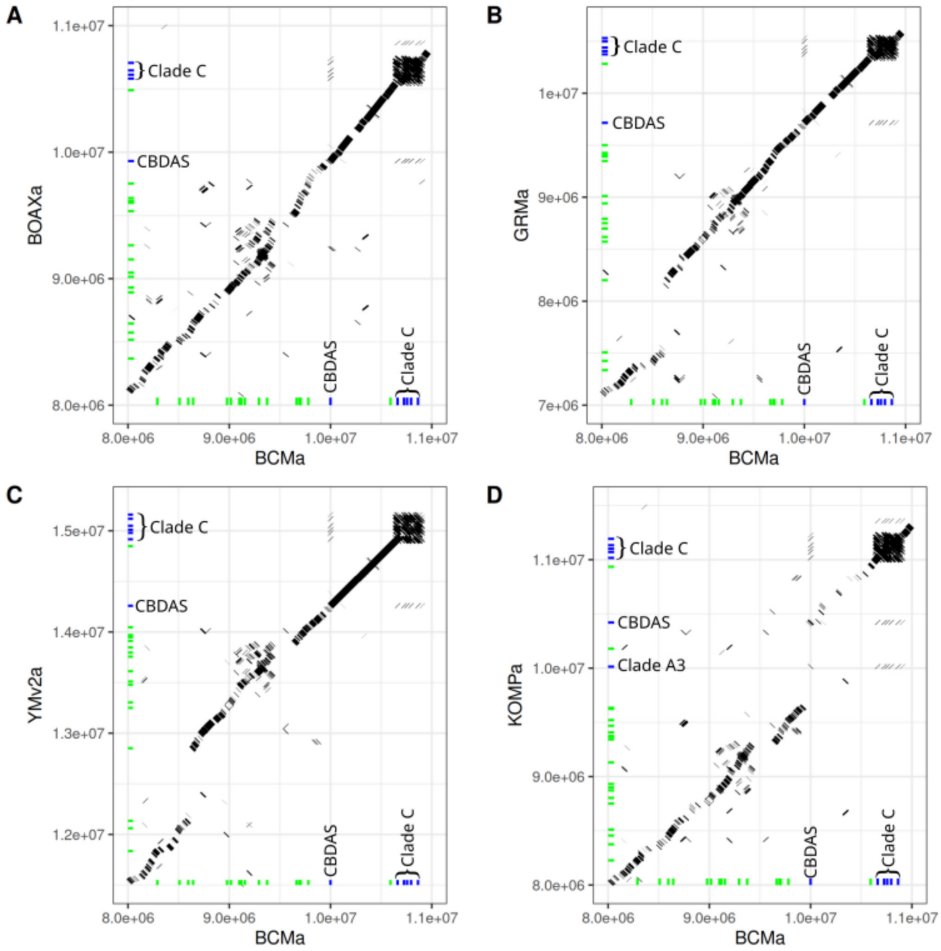

FIGURE 3 | Comparison of CBD-type cultivars to Boone County (BCM). Varieties shown are (A) BoAx (BOAX), (B) Golden Redwood (GRM), (C) Kompolti (KOMP), and (D) YunMa (YMv1).

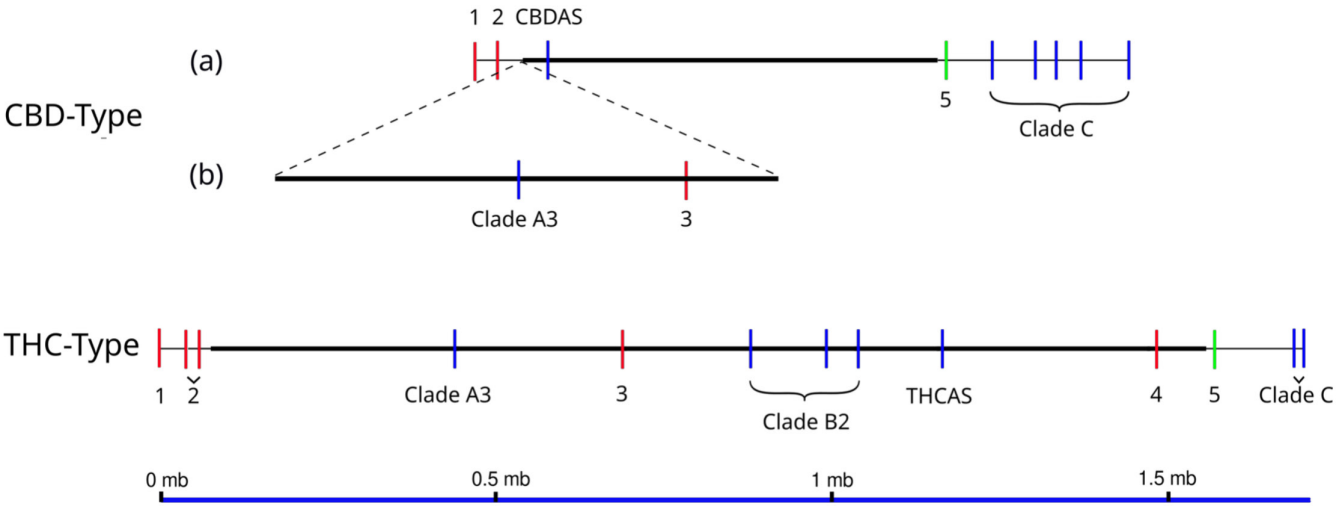

FIGURE 4 | Diagrammatic representation of the *B* locus. *B* locus with immediate surrounding genes. The B locus itself is shown as a thicker black line, bounded on the left by the unrelated genes MBS1-like (1) and CAAT3-like (2), and on the right-hand side by an uncharacterized BBE gene 45.122 (5). Genes designated 3 and 4 are FAR1-related and an unknown gene. The THC-type genomes all displayed the configuration shown here, but the CBD-type sequence were found in two arrangements, with the most common designated as (a) and the arrangement found in KOMP and Tibet shown as (b), where this appears to be an insertion of THC-type sequence into the start of the B locus, including the Clade A3 pseudogene and the FAR1-related gene.

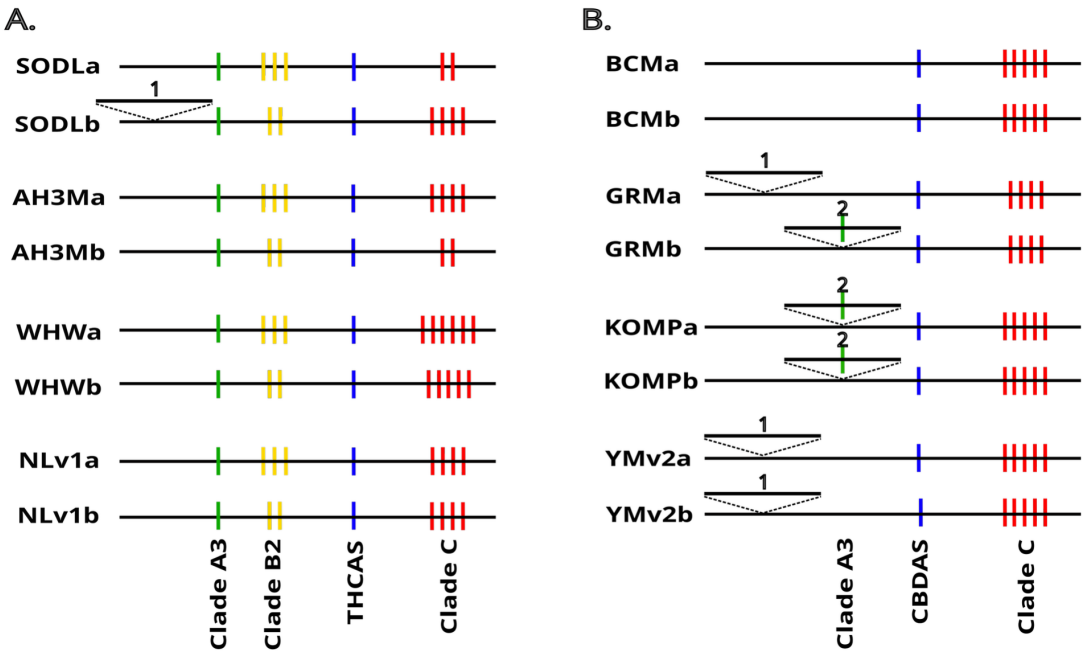

FIGURE 5 | Variability in and around the *B* locus. Summary of comparison between the two haplotype assemblies for each of the cultivars used in this study. (A) THC-type cultivars, (B) CBD-type cultivars. Two insertion sites are shown. The first of these (labeled 1) is upstream of the B locus and is nearly identical between SODLb and the three CBD-type genomes that contain it. The second of these (labeled 2) is part of the THC-type B locus inserted into a CBD-type genome.

magnitude higher than the presumably functional CBCAS and Clade C genes.

We obtained complete transcripts of the CladeA3.1 gene by mining published RNA-Seq libraries from the intermediate-type cultivar Cannbio2 (Braich et al. 2019), and THC-type cultivars Blackberry Kush and Chemdawg. Libraries used for transcript retrieval are listed in Table S2 in Supporting Information S2. A nearly complete transcript was also found in the Purple Kush transcriptome (van Bakel et al. 2011). These derived sequences are very similar, separated from each other by just a few SNPs.

2.11 | Clade A3.2: Intact Sequence in CBD-Type Cultivars

Clade A3.2 was found in the CBD-type cultivars KOMP, GRMb, and Finola. All of the other CBD-type cultivars we examined lacked the entire genome segment containing this gene (Figures 1A and 3). However, KOMP (Figure 1B) and GRMb contain a segment very similar to the THC-type *B* locus sequence that includes the Clade A3.2 gene. As noted for Finola (Van Velzen and Schranz 2021), these sequences are intact, meaning they lack the in-frame stops and the nine-base deletion found in the THC-type version. We could find no evidence of the expression of these genes. Table S1 in Supporting Information S1 contains a summary of the features of Clade A3.

2.12 | Clade A3.3: Potentially Functional but Not Expressed

Clade A3.3 was found in THC-type cultivars Ace High 3-2 and Sour Diesel, both of which were heterozygous with the other copy being Clade A3.1. The Clade A3.3 sequence carries all but one of the CBD-type SNPs (correcting both in-frame stops) and also lacks the deletion. There are an additional 29 SNPs distinguishing this gene from the other Clade A3 genes we examined. These differences mean this gene has the potential to make a functional product, but we were unable to find any evidence for transcription.

2.13 | Clade B2 Pseudogenes

CBDAS pseudogenes with multiple frameshifts and stop codons have been observed by several investigators in THC-type cultivars (Laverty et al. 2019; Weiblen et al. 2015; Cascini et al. 2019). These genes were classified as Clade B2 (Van Velzen and Schranz 2021). We found three copies of this pseudogene in several THC-type cultivars. These genes are tightly linked to THCAS. Table S1 in Supporting Information S1 contains a summary of the features of Clade B2.

We did not find Clade B2 genes in any CBD-type cultivars. This includes KOMP, GRMb, and Finola, all of which have a segment of the THC-type *B* locus that includes the Clade A3 gene. The Clade B2 cluster is located about 125 kB from the THCAS gene (Figure 4). While there is substantial variation among these genes, none has an intact open reading frame.

The middle gene in this cluster was detected as an RT-PCR product from a THC-type cultivar called Skunk #1 (Weiblen et al. 2015). We verify that this gene is expressed. In fact, Figure 7 shows that this gene is more highly expressed in the examined trichome libraries than any of the CBCAS or Clade C genes. We obtained full-length transcripts from published trichome RNA-Seq libraries from four cultivars: Lemon Skunk, Sour Diesel,

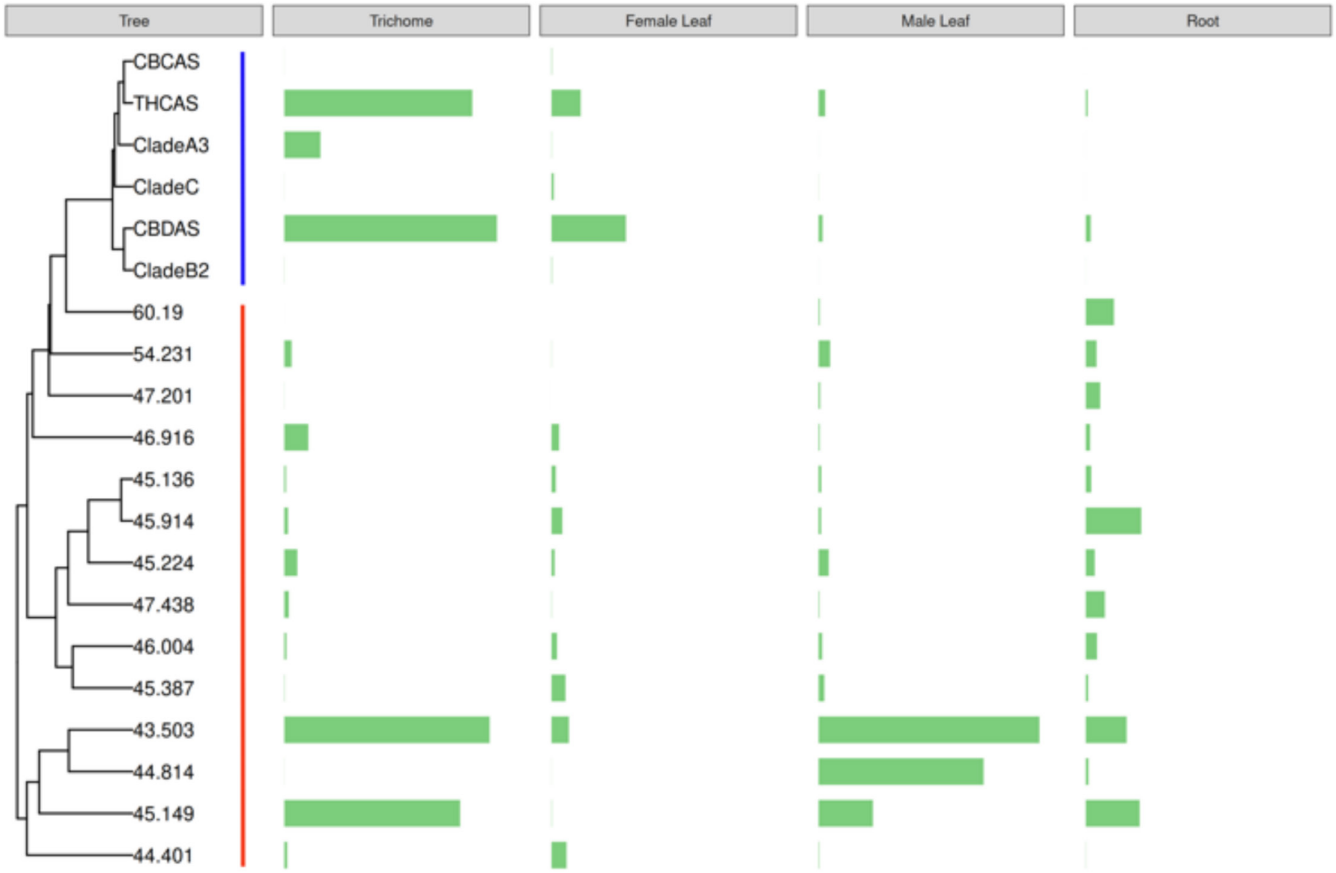

FIGURE 6 | Gene expression of the BBE family across tissues. Tissue expression of the *C. sativa* berberine bridge enzyme family. On the left side of the panel is a tree of full-length mRNA sequences constructed using the UPGMA algorithm. The cannabinoid oxidocyclase group is highlighted with a blue line, while the rest of the uncharacterized BBE proteins are highlighted with a red line. The uncharacterized proteins are denoted with their percentage identity to the canonical THCAS protein sequence. The four bar plots show gene expression levels in trichomes, female leaf, male leaf, and root tissue using the Cannbio2 rnaSEQ libraries (Braich et al. 2019). Bar height is percent total expression of the set of BBE genes for each tissue. Because the copies of the CBCAS and Clade C genes are so nearly identical, there is no reliable way to accurately determine individual gene expression values. Copies of CBCAS and Clade C genes are simply summed and collapsed in the figure.

Blackberry Kush, and Cherry Chem (Zager et al. 2019; Booth et al. 2020). These sequences are nearly identical to each other and to the sequence reported by Weiblen et al. (2015), though SODLa and AH3mb have a version differing at 25 SNPs.

2.14 | Cloning and Sequencing of Novel Cannabinoid Oxidocyclase Genes

One goal of this work was to expand the number of characterized cannabinoid oxidocyclases. Earlier work with previously sequenced PBBK (Steep Hill Labs, NCBI accession GCA_002090435.1) revealed an extensive gene family of cannabinoid oxidocyclases of unknown function. We used degenerate primers to obtain multiple full-length family members including THCAS, CBCAS, or Clade C genes (Table S2A in Supporting Information S1). PBBK was used as a control for a 6× sampling of 120 successfully recovered recombinant clones. An additional 520 recombinant clones from 20 different *C. sativa* cultivars of various chemotypes were also produced and sequenced (Table S2B in Supporting Information S1). Recombinant clones (Table S2C in Supporting Information S1) were derived from *C. sativa* intermediate- and CBD-type cultivars F-Cancer and AC/DC, respectively. These were initially assessed for cannabinoid accumulation, as confirmed by HPLC-PDA analysis using previously established chromatographic conditions (Lynch et al. 2016; Vergara et al. 2017) (Figure S4A,B).

2.15 | Cannabinoid Oxidocyclase In Vitro Enzymatic Characterization

Recombinant cannabinoid oxidocyclase enzymes were expressed in sf9 cells. These His-tagged enzymes were secreted, detected by immunochemistry, and sized at approximately 57 kDa (Figure S5), consistent with previous reports of non-glycosylated functional cannabinoid synthases (Laverty et al. 2019; Morimoto et al. 1998).

2.16 | Enzyme Activity Measurements

Functional characterization revealed significant variation in activity between the CBCAS Clade C variants (C.4, C.5, and C.6, Figure 8). CBCAS displayed the highest activity (normalized to 100%, SE = 11.5%), while Clade C variants showed significantly

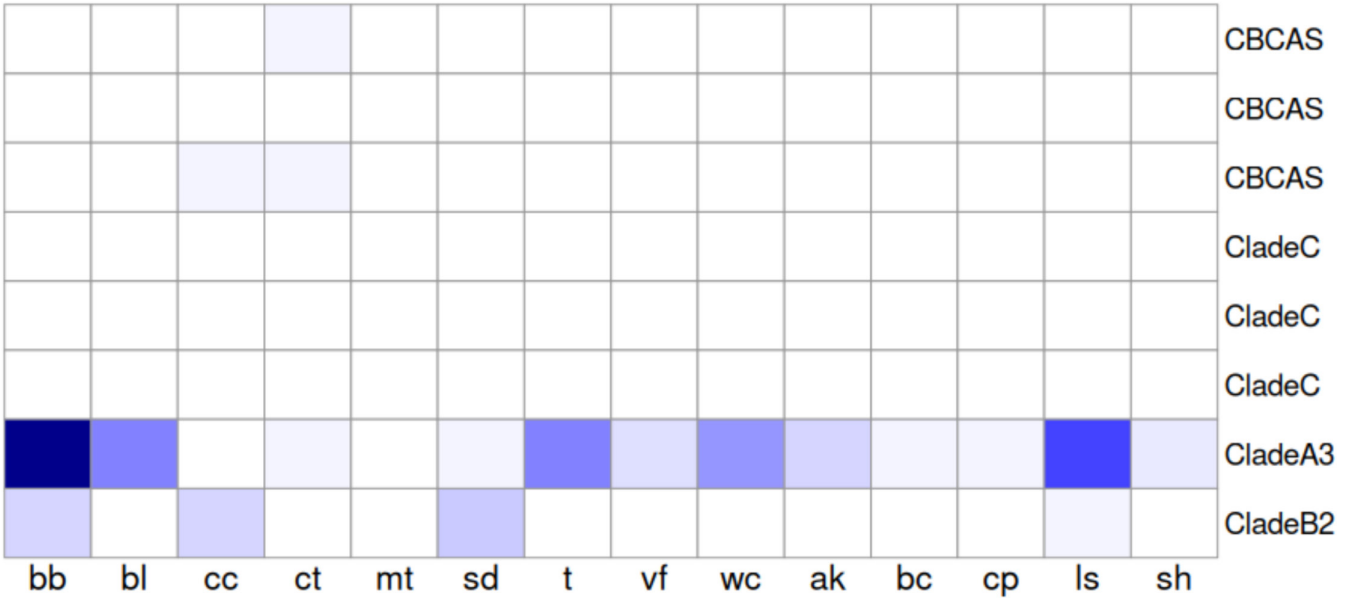

FIGURE 7 | Expression of minor cannabinoid oxidocyclase genes. Expression of the cannabinoid oxidocyclase genes excluding THCAS and CBDAS. Published expression data used in this figure are from (Zager et al. 2019) bb: Black Berry Kush, bl: Black Lime, cc: Cherry Chem, ct: Canna Tsu, mt: Mama Thai, sd: Sour Diesel, t: Terple, vf: Valley Fire, wc: White Cookies and (Booth et al. 2020) ak: Afghan Kush, bc: Blue Cheese, cp: Chocolope, ls: Lemon Skunk, sh: CBD Skunk Haze.

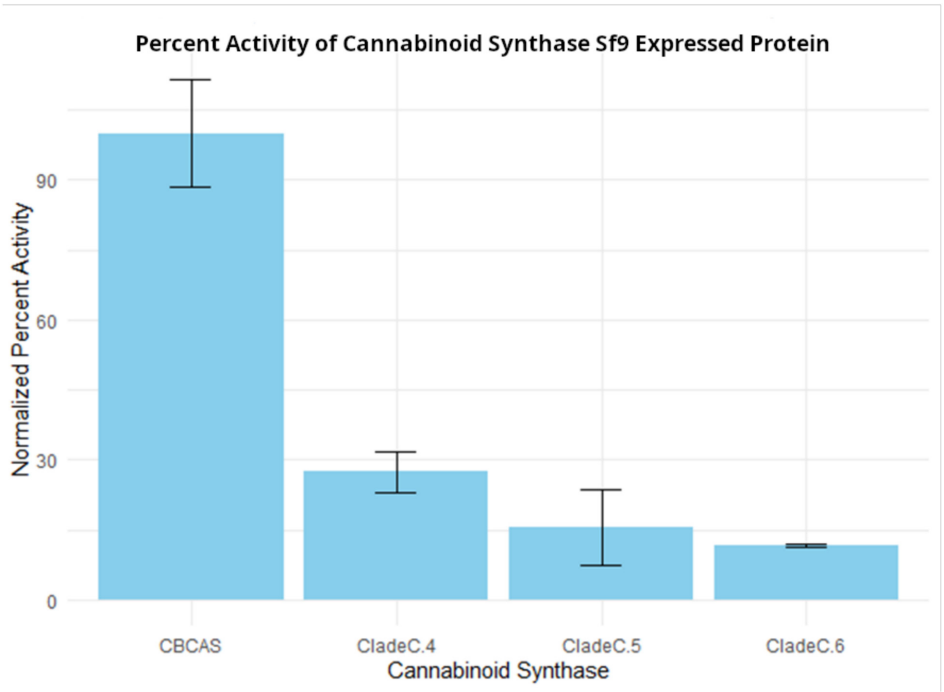

FIGURE 8 | Functional characterization of recombinant cannabinoid oxidocyclases. Purified proteins, secreted from sf9 cells infected with recombinant baculovirus, were functionally characterized. All assays were supplemented with CBGA precursor. Activity presented as percent of total CBGA converted was normalized against canonical CBCAS activity at 100% functionality. With CBCA as the dominant product of these enzymes, comparative analysis of activity for Clade C.4, Clade C.5, and Clade C.6 was carried out against CBCAS. In assays with purified protein from sf9 expression lower activity of CBCA production was observed in Clade C recombinant enzymes. Clade C.4 comparative activity was 26% of CBCAS, while Clade C.5 and Clade C.6 were 16% and 11%, respectively, of CBCAS.

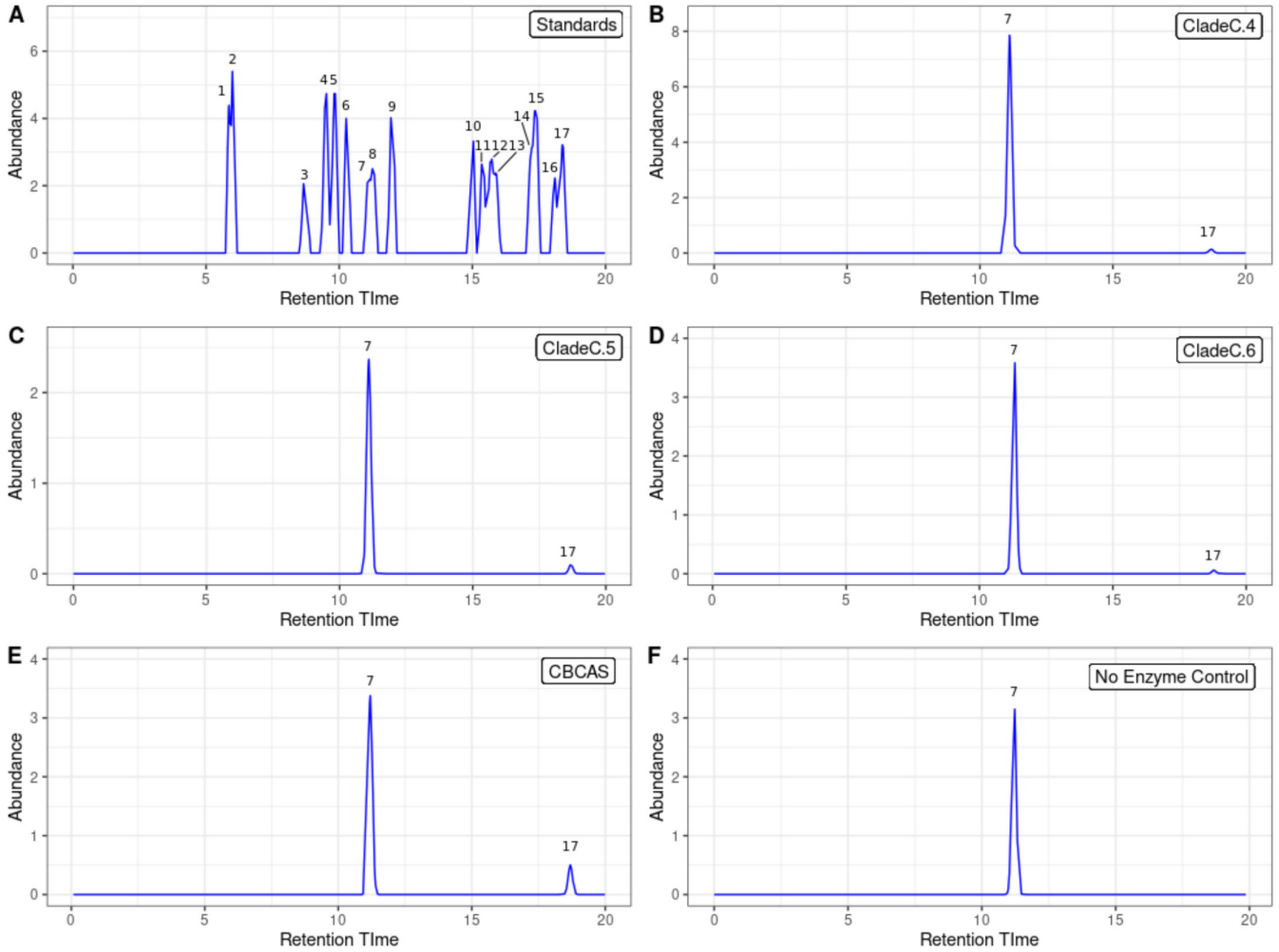

FIGURE 9 | Full chromatograms for cannabinoid oxidocyclase activity from sf9 expressed protein in vitro assays. HPLC-UV graphical chromatograms from in vitro enzymatic assays showcasing the activity of sf9 secreted recombinant cannabinoid oxidocyclases, plotting the absorbance against retention time for samples (A–F). (A) 17-Cannabinoid mix from certified reference standard: This chromatogram serves as a reference, displaying peaks corresponding to various cannabinoids with their respective retention times. Peaks are labeled with numbers (1–17) for available neutral and acidic cannabinoid standards. (B) CladeC.4: a single major peak at Peak 7, corresponding to CBGA, and a minor peak at Peak 17, corresponding to CBCA, indicating the production of a small amount of CBCA by the enzyme. (C) CladeC.5: prominent CBGA peak at Peak 7 and a minor CBCA peak at Peak 17, although with a slightly lower abundance compared to CladeC.4. (D) CladeC.6: a dominant CBGA peak at Peak 7, with a minor abundance of CBCA at Peak 17, similar to CladeC.4 and CladeC.5. (E) CBCAS: In contrast to the previous samples, while a dominant peak of CBGA is present at Peak 7, an increased CBCA production is observed for CBCA at Peak 17, indicating significant production of CBCA by CBCAS relative to tested clade C genes. (F) No enzyme control: This control reaction contained no enzyme, and only the CBGA peak at Peak 7 was observed.

reduced activity: C.4 (27.4%, SE = 4.3%), C.5 (15.6%, SE = 8.1%), and C.6 (11.6%, SE = 0.4%). This suggests that Clade C variants have lower enzymatic activity than the wild-type CBCAS, with C.6 showing the most pronounced reduction.

2.17 | Product Identification

HPLC analysis confirmed the production of CBCA from CBGA by these enzymes (Figure 9). Compound identification was conducted using a 17-cannabinoid certified reference standard mix (Figure 9A-Standards). Retention times and peak area UV absorbance spectra intensity, normalized as abundance ratios, allowed measurements at targeted wavelengths and were used for quantification (Hazekamp et al. 2005).

CBCA was identified as peak 17 in assays with CBCAS, Clade C.4, Clade C.5, and Clade C.6 (Figure 9B–E). As expected, only the substrate (CBGA) was observed in the no-enzyme control (Figure 9F). Importantly, no THCA was detected in any of the assays with Clade C enzymes.

2.18 | Enzyme Activity Measurements

Functional characterization revealed significant variation in activity between the CBCAS Clade C variants (C.4, C.5, and C.6, Figure 7). CBCAS displayed the highest activity (normalized to 100%, SE = 11.5%), while Clade C variants showed reduced activity: C.4 (27.4%, SE = 4.3%), C.5 (15.6%, SE = 8.1%), and C.6 (11.6%, SE = 0.4%). While these results suggest that

Clade C variants possess lower enzymatic activity than the wild-type CBCAS, we note the high measurement variability and nonuniform replicate set across the characterized enzymes particularly observed in Clade C.5 and CladeC.4. Limited protein availability and potential differences in activity recovery during in vitro purification contributed as technical challenges. While the downward trend in activity is consistent across all variants, these specific values should be interpreted as preliminary indicators of reduced catalytic efficiency in the Clade C lineage.

3 | Discussion

3.1 | Clade C Enzymes and the Source of THCA in CBD-Type Cultivars

A key finding of this study is that Clade C cannabinoid oxidocyclases produce CBCA as their primary product and do not generate detectable THCA. This result has important implications for understanding the persistent THCA found in CBD-type cultivars that lack functional THCAS.

An important challenge for growers is that even CBD-type plants lacking a functional THCA synthase gene still produce THCA at a fairly constant ratio of approximately 24–28:1 CBDA:THCA. This relationship is linear with a Pearson's $r = 0.98$ (Zirpel et al. 2018; Toth et al. 2020). Consequently, it is difficult to grow plants with a CBDA concentration above 8% of dry weight without THCA concentrations exceeding the 0.3% threshold, rendering the crop out of compliance (Stack et al. 2021). The origin of THCA in plants lacking functional THCAS remains an important question. Some clearly arises as a side product of CBDAS (Zirpel et al. 2018). But whether this accounts for all THCA, or whether other oxidocyclases contribute, particularly CBCA synthases, remains unclear (Onofri et al. 2015; Thomas and Kayser 2023; Thomas and Kayser 2022).

Our characterization of three Clade C enzymes confirms that none produce THCA as a side product, consistent with recent studies of CBCAS (Thomas and Kayser 2023; Thomas and Kayser 2022). Combined with the relatively low expression levels observed for Clade C genes (Figure 7), these enzymes are unlikely to contribute significantly to oil composition in most cultivars. However, the possibility remains that some cultivars may show higher Clade C expression.

The most likely source of THCA in CBD-type cultivars is CBDAS itself. In cultivars lacking a functional THCAS gene the CBD:THC ratio is typically 24–28:1 (Zirpel et al. 2018; Toth et al. 2021). This closely matches the ratio produced by purified CBDAS enzyme at its pH optimum of 4.5 (Zirpel et al. 2018). This correspondence strongly suggests that CBDAS side-product activity accounts for the extra THCA in CBD-type *C. sativa*, resolving a significant question for growers concerned with regulatory compliance.

The Clade C enzymes characterized here show varied activity levels. While CBCAS displayed the highest activity, Clade C variants (C.4, C.5, and C.6) showed progressively reduced activity at 27.4%, 15.6%, and 11.6%, respectively (Figure 8). Sequence variation among CBCAS genes isolated from different cultivars includes 28 sites with amino acid changes. While seven of these occur in the predicted signal peptide and are unlikely to affect catalytic function, the remaining mutations are distributed across the mature protein. Although none map to the predicted active site, they could potentially influence product profiles (Fulvio et al. 2021). Further investigation of these variants may reveal subtle functional differences. The observed reduction in activity across Clade C variants highlights a potential functional divergence within the CBCAS family. However, the relatively high standard error for Clade C.5 and the limited sample size for Clade C.4 suggest that further kinetic characterization is required to precisely quantify the catalytic constants (kcat and Km) of these specific recombinant proteins. Such studies will be necessary to flesh out the kinetic nuances and activity differences among these functionally related enzymes.

3.2 | BBE Family Expression and Functional Diversity

Analysis of BBE gene expression across *C. sativa* tissues reveals unexpected complexity. While THCAS and CBDAS dominate expression in trichomes as expected, several uncharacterized BBE enzymes also show substantial trichome expression at levels comparable to the main cannabinoid synthases (Figure 6). This suggests these enzymes could play important, yet-to-be-discovered roles in trichome chemistry.

The BBE family shows tissue-specific expression patterns. In female leaves, CBDAS and THCAS account for most BBE expression. In contrast, male leaves show high expression of a clade of three uncharacterized BBE genes (Figure 6). Many BBE genes show the highest expression in roots rather than aerial tissues. Several genes display dramatically different expression between male and female leaves, though functional interpretation awaits enzyme characterization.

Discovering the functions of these uncharacterized BBE enzymes will not be straightforward. The BBE family produces diverse alkaloids from varied substrates (Daniel et al. 2017), requiring systematic testing with educated substrate choices. An obvious first step would be testing with CBGA. Additional candidates might be identified by examining compounds present in tissues where specific BBE genes are highly expressed. Database searches show the uncharacterized enzymes have 50%–60% protein identity to known BBE enzymes. This is insufficient similarity for confident function assignment, but potentially useful for generating testable hypotheses.

The expression of Clade A3 and B2 pseudogenes (Figures 6 and 7) was sufficient to allow transcript assembly from RNA-Seq libraries of several THC-type cultivars. Expressed pseudogenes can play various regulatory functions (Garewal et al. 2021), though determining whether these particular pseudogenes have functional significance requires further investigation.

3.3 | The *B* Locus as a Supergene

The *B* locus exhibits defining characteristics of a supergene: discrete alternative phenotypes (THC-type vs. CBD-type), Mendelian segregation despite involving multiple genetic elements, and suppressed recombination maintaining specific chromosomal arrangements. Thompson and Jiggins (2014) define supergenes as a genetic architecture involving multiple linked functional genetic elements allowing switching between discrete, complex phenotypes maintained in a stable local polymorphism. This definition pointedly includes supergenes with even a single coding sequence, focusing on the functional architecture rather than gene number.

The puzzle of the *B* locus has been how THCAS and CBDAS segregate as a single Mendelian locus despite being clearly distinct genes. Initial models proposed they were alleles (De Meijer et al. 2003; Weiblen et al. 2015; Wenger et al. 2020). But phylogenetic analysis shows they diverged early in cannabinoid oxidocyclase family expansion (Van Velzen and Schranz 2021). The current model proposes that THCAS and CBDAS are separate loci in a region of reduced recombination, in repulsion. THC-type plants have THCAS tightly linked to CBDAS pseudogenes, while CBD-type plants have CBDAS tightly linked to THCAS pseudogenes (Grassa et al. 2021).

Our analysis of high-quality genome assemblies confirms this model with important refinements. The Clade B2 cluster adjacent to THCAS in THC-type plants can accurately be called CBDAS pseudogenes. However, most CBD-type cultivars examined lack any structure resembling a THCAS pseudogene. But exceptions are found in KOMP and one haplotype of the GRM cultivar. These loci contain a section of THC-type *B* locus, including the Clade A3 pseudogene (Figure 1). The substantial size difference between THC- and CBD-type B loci (1.5 Mb vs. 0.66 Mb) suggests a complex evolutionary history. One possibility is that this could be the result of a rare recombination event between THC-type and CBD-type genomes.

Genetic markers distinguishing THC-type from CBD-type cultivars all map within the *B* locus itself (De Meijer et al. 2003; Toth et al. 2020). Features outside the *B* locus do not correlate with chemotype.

3.4 | Mechanism of Recombination Suppression

Unlike many supergenes, where chromosomal inversions suppress recombination (Thompson and Jiggins 2014), the *B* locus shows no evidence of inversion. Instead, the mechanism appears related to high repetitive element content. The *B* locus is 92.8% covered by repeat sequences compared to 79.5% for chromosome 7 overall. The median size for unmasked regions in the *B* locus is just 105 bp, and only 11 unique stretches exceed 1500 bp—five of which contain cannabinoid oxidocyclase genes. Critically, none of these 11 segments have matches in the corresponding CBD-type cultivar.

This high repeat content likely suppresses recombination, preventing the formation of chromosomes carrying both functional CBDAS and THCAS. Such an arrangement may eventually be discovered as more cultivars are examined, but it has not been observed to date (Lynch et al. 2025). The structure effectively maintains tight linkage between THCAS and the Clade A3 and B2 pseudogenes, each showing detectable transcription in at least some cultivars. Whether this linkage has functional significance remains to be determined.

3.5 | Evolutionary Maintenance and Selective Pressures

An essential feature of supergenes is balanced selective forces maintaining both variants in the population. For the *B* locus, selection for high THCA has likely operated for millennia. The selective pressure maintaining the CBD-type variant is less immediately obvious but becomes clear when considering *Cannabis* cultivation history.

Small (2015) observed that divergent selection for high THC content and high stem fiber content represents a principal dimension of disruptive evolutionary forces in *Cannabis*. This has resulted in populations with dramatically different trait combinations: generally low THCA in fiber cultivars versus high THCA in THC-type cultivars. This divergence is readily apparent in molecular marker comparisons (Grassa et al. 2021). Fiber breeding programs typically select against THCA content (Small 2015), representing the balancing selection needed to maintain both *B* locus variants in the global *Cannabis* gene pool.

The recent surge in CBD demand adds another dimension to this selective landscape. Legal CBD-type cultivation for cannabinoid production now operates alongside traditional fiber and seed production, and alongside continuing (though often illegal) cultivation for THC content. This creates three distinct selective regimes acting on *Cannabis* populations worldwide, all of which intersect with the chemotype-determining *B* locus.

The supergene architecture of the *B* locus facilitates rapid switching between these discrete chemotype states through conventional breeding. Intermediate-type (Type II) plants serve as bridges between THC- and CBD-type populations, allowing breeders to introgress desirable traits across the chemotype boundary while maintaining the fundamental cannabinoid profile determined by *B* locus genotype. This flexibility has been crucial for developing modern varieties that maximize CBD production while remaining compliant with THC restrictions.

3.6 | Implications for Breeding and Compliance

Our findings have practical implications for *Cannabis* breeding programs. Understanding that CBDAS itself is the primary source of THCA in CBD-type cultivars means that reducing this side-product activity represents a clear breeding target. However, the consistent 24–28:1 CBD:THC ratio suggests that this is an intrinsic property of the CBDAS enzyme rather than something easily modified through traditional breeding approaches.

The supergene structure of the *B* locus also has implications for marker-assisted breeding. Markers mapping within the *B* locus can reliably predict chemotype, but the substantial structural variation we observed both within the *B* locus and in flanking regions suggests that marker development must account for this diversity. The two distinct CBD-type *B* locus structures we identified (i.e., one with and one without THC-type sequence) may represent functionally equivalent variants that could behave differently in breeding programs.

Finally, our characterization of Clade C enzymes eliminates these as potential contributors to compliance problems. CBD-type *Cannabis* breeders can focus their attention on managing the THCA side-product activity inherent in the CBDAS enzyme, rather than searching for other oxidocyclases that might be elevating THCA levels in their crops.

4 | Materials and Methods

4.1 | Cultivar Nomenclature

Throughout this paper, we will refer to specific cultivars of *C. sativa*, but it is important to note that there is no central naming authority for this species. What we are calling "cultivars" typically have whatever name the original breeder used, with no verification systems. Further, as will be apparent from our analyses, most of these cultivars are quite heterozygous. Nevertheless, this can be considered a snapshot of the current state of *C. sativa* germplasm. The cultivars used for genome comparisons in this paper are listed in Table S1 in Supporting Information S2.

4.2 | Genome Comparison and Bioinformatics Analysis

4.2.1 | Synteny Assessment

Synteny assessment was done with lastz version 1.04.22 with default parameters (Harris 2007). This was simplified by the availability of repeat-masked genomes available with the Salk genome releases (Lynch et al. 2025). The read_paf and dotplot utilities from the R pafr package (version 0.0.2) were used to visualize output.

4.2.2 | Transcriptome Construction and Gene Expression Analysis

A transcriptome was constructed using published RNA-Seq data including trichome libraries from (Zager et al. 2019) (BioProject PRJNA498707) comprising 27 libraries and the tissue expression libraries from (Braich et al. 2019) (BioProject PRJNA560453) comprising 71 libraries. This data set is heavily weighted toward trichome expression, but the Cannbio2 libraries bring substantial depth over a range of vegetative tissues.

Reads were aligned to the Jamaican Lion Dash genome assembly (GenBank accession GCA_003660325.2) using hisat2 version 2.1.0 and StringTie version 1.3.4d with –max-intronlen 10,000 (Kim et al. 2019). Sorted BAM files were created using Samtools version: 1.10 using htslib 1.10.2-3 (Li et al. 2009).

Transcriptomes were computed tissue by tissue using StringTie (Frazee et al. 2015; Pertea et al. 2016) with default parameters. The resulting output files were merged into a single transcriptome using the StringTie merge function. This resulted in 70,101 transcripts across 36,626 genes with a median length of 1552 bases.

Transcript abundances were calculated using StringTie v1.3.4d using the Jamaican Lion transcriptome as annotation. Expression values as fragments per kilobase million (FPKM) were computed with Ballgown version 2.26.0 (Frazee et al. 2015). The phylogenetic tree shown in Figure 6 was calculated using the UPGMA algorithm implemented in the phangorn R package (Schliep 2011).

4.2.3 | Extraction of Transcripts From RNA-Seq Libraries

Genomic sequences for Clade A3 and Clade B2 pseudogenes were identified in each genome assembly using BLASTn. Full-length transcripts were obtained from publicly available RNA-Seq data (Booth et al. 2020: PRJNA599437 and Zager et al. 2019: PRJNA498707). RNA-Seq libraries were mapped to the Jamaican Lion genome using hisat2 as described above.

Target gene positions were determined by BLASTn, and mapped reads were extracted from the BAM files using Samtools. This gave varying numbers of reads by library, so those with more than about 2000 reads covering the full length of the gene were assembled with CAP3 (VersionDate: 02/10/15). Full-length Clade B2 transcripts were recovered from Lemon Skunk, Sour Diesel, Blackberry Kush, and Cherry Chem. Clade A3 transcripts were recovered from Blackberry Kush and Chemdawg. Accession numbers for the libraries used are shown in Table S2 in Supporting Information S2.

4.2.4 | Molecular Cloning and Gene Characterization

4.2.4.1 | Genomic DNA Isolation. We selected several *C. sativa* cultivars from the California recreational and medicinal market for survey using PCR. Screening with degenerate primers (designed to recover sequences >85% identical to THCAS at the protein level). Genomic DNA was isolated from *C. sativa* samples with Qiagen DNA Easy Plant genomic DNA isolation kits (Qiagen) according to manufacturer instructions. Purified DNA was quantified using the Qubit HS-dsDNA System (Promega) and a Qubit Fluorometer (Thermo Fisher) per manufacturer instructions. DNA was diluted to a 5 ng/μL final working concentration for use as normalized input in PCR reactions.

4.2.5 | Degenerate Primer Set for Cloning and Sequencing

A single degenerate primer set was designed to amplify all cannabinoid oxidocyclase family members. The design was based on the mapped cannabinoid oxidocyclase family from Pineapple

Banana Bubba Kush (GenBank accession SAMN06546749). Degenerate nucleotides were used at SNP positions within the 5′ and 3′ cannabinoid oxidocyclase coding sequence ends to capture the full-length genes present in the genomic DNA of a given amplified variety (see Supplementary Table S4). The primer sets amplify gene targets of THCAS and CBCAS (Clade A), CBDAS (Clade B), and Clade C.

4.2.6 | TOPO Cloning and Sequencing

Normalized gDNA was amplified using SuperFi PCR master mix (Table S3 in Supporting Information S2) to target 1.6-kB amplicons. PCR products were cloned into TOP10 *E. coli* using the TOPO TA Cloning Kit per manufacturer instructions. Transformants underwent blue/white screening on antibiotic media. Positive colonies were cultured in LB broth, and plasmid DNA was isolated via alkaline lysis as previously described (Reid 1991). DNA samples were prepared for Sanger sequencing using M13 primers (Table S4 in Supporting Information S2) and sent to Elim Biopharmaceuticals. Reads were mapped to the cannabinoid oxidocyclase family to determine percent identity.

4.2.7 | Recombinant Protein Expression and Purification

4.2.7.1 | Sf9 Insect Cell Culture. *Spodoptera frugiperda* (sf9) insect cells were purchased from Invitrogen (Waltham, MA). Initial p0 culture was thawed and recovered per manufacturer instructions. Cells were maintained in T25 flasks with Sf-900 II SFM medium at 28C biological incubator protected from light exposure with aluminum foil coverings. Cell morphology and growth patterns were observed with an inverted microscope. Cells were passaged 1:10 every 4 days once 95%–98% uniform monolayer confluency was reached. Optimal transfections were performed between 5 and 20 passages.

4.2.7.2 | SLIC Subcloning and Generation of Recombinant Bacmids Generation. Recombinant donor pFast vectors were generated via sequence and ligation independent cloning (SLIC). Amplicons with flanking homology were purified and mixed with linearized pFast1 vector at a 1:1-M ratio. The reaction was prepared as described (Li and Elledge 2012). Recombinant bacmids were generated using the Bac-to-Bac system in DH10Bac cells, with selection performed on triple-antibiotic media (Table S5 in Supporting Information S2). Bacmid DNA was isolated using the SV genomic DNA kit.

4.2.7.3 | Sf9 Transfection of Recombinant Bacmid DNA and Recombinant Cannabinoid Oxidocyclase Expression. Sf9 cells were maintained in Sf-900 II SFM at 28°C, protected from light, and passaged 1:10 upon reaching 95%–98% confluency. Transfections utilized cells between passages 5 and 20. Purified bacmid DNA (~1 μg/μL) was transfected into Sf9 cells (8×10^5 cells/mL) using ExpiFectamine Sf Reagent following recommendations from the Bac-to-Bac expression system manual (Gibco). P0 viral stocks were harvested 5–7 days post-transfection upon observation of a lytic phenotype. P1 stocks were generated by infecting confluent Sf9 cells with P0 supernatant, facilitating the expression and secretion of 6× His-tagged proteins.

Recombinant protein was eluted stepwise with elution buffers 1–4 in 5-mL volumes. Fractions collected using buffers 3 and 4 for each recombinant enzyme were desalted using a PD-10 desalting column per manufacturer instructions (GE). Additional dialysis was performed using the Tube-O-DIALYZE system with three 20-mM sodium citrate buffer exchanges (G-biosciences, Millipore-Sigma). Purified protein was analyzed by SDS-PAGE electrophoresis and Western blot using anti-His antibodies per manufacturer guidelines (Thermo Fisher).

4.2.7.4 | Purification and Enzymatic Assays. Proteins were purified using 5-mL HisTrap Fast Flow columns. Supernatants were loaded and eluted via stepwise imidazole concentrations (Table S6 in Supporting Information S2). Fractions were desalted using PD-10 columns and dialyzed against 20-mM sodium citrate. Purified proteins were normalized and utilized in in vitro assays (Table S7 in Supporting Information S2) with CBGA substrate. Reactions were terminated with MeOH and analyzed via HPLC (Zirpel et al. 2018; Laverty et al. 2019; Lynch et al. 2016). Data were visualized in R using ggplot2, with error bars representing 95% confidence intervals.

4.2.7.5 | HPLC Analysis. HPLC analysis was performed using an instrument calibrated with certified reference cannabinoid analytical standards and quantified against a calibration curve (Cerilliant). Data were acquired and analyzed as previously described (Lynch et al. 2016).

Optimization reactions were tested in duplicate. Enzyme activity reactions were carried out under standard conditions and repeated 2–5 times: CBCAS (five replicates), Clade C.4 (two replicates), Clade C.5 (four replicates), and Clade C.6 (three replicates). Measured area mean, standard deviation, standard error, and 95% confidence intervals for enzyme activity assays were calculated and used for error bars. Data were normalized against CBCAS enzyme activity and plotted in R using ggplot2.

Author Contributions

A.K.D., T.A., and G.R. conceived and planned the work. T.A. performed all the molecular biology, including the enzyme assays. A.K.D. performed the bioinformatics analysis. A.K.D. and T.A. wrote the manuscript, and all authors took part in the editing of the manuscript.

Acknowledgments

The authors would like to thank Don Land for his support and review on enzymatic analysis and the manuscript as a whole. We would also like to extend sincere gratitude to the two anonymous reviewers who provided many useful insights.

Funding

This work was supported by the Front Range Biosciences.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The Clade C gene sequences characterized in this study were deposited in GenBank with the accession numbers OP751990, OP751991,

and OP751992. The broader set of BBE genes is available as European Nucleotide Archive accessions OZ388187-OZ388204.

Peer Review

The peer review history for this article is available in the Supporting Information for this article.

References

Booth, J. K., M. M. S. Yuen, S. Jancsik, L. L. Madilao, and A. J. E. Page. 2020. "Terpene Synthases and Terpene Variation in *Cannabis sativa*." *Plant Physiology* 184, no. 1: 130–147. https://doi.org/10.1104/PP.20.00593.

Braich, S., R. C. Baillie, L. S. Jewell, G. C. Spangenberg, and N. O. I. Cogan. 2019. "Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis." *Scientific Reports* 9, no. 1: 1–12. https://doi.org/10.1038/s41598-019-53023-6.

Cascini, F., A. Farcomeni, D. Migliorini, et al. 2019. "Highly Predictive Genetic Markers Distinguish Drug-Type From Fiber-Type *Cannabis sativa* L." *Plants* 2019, no. 11: 496. https://doi.org/10.3390/PLANTS8110496.

Daniel, B., B. Konrad, M. Toplak, et al. 2017. "The Family of Berberine Bridge Enzyme-Like Enzymes: A Treasure-Trove of Oxidative Reactions." *Archives of Biochemistry and Biophysics* 632: 88–103. https://doi.org/10.1016/J.ABB.2017.06.023.

De Meijer, E. P. M., M. Bagatta, A. Carboni, et al. 2003. "The Inheritance of Chemical Phenotype in *Cannabis sativa* L." *Genetics* 163, no. 1: 335–346. https://doi.org/10.1093/GENETICS/163.1.335.

ElSohly, M. A., Z. Mehmedic, S. Foster, C. Gon, S. Chandra, and J. C. Church. 2016. "Changes in Cannabis Potency Over the Last 2 Decades (1995–2014): Analysis of Current Data in the United States." *Biological Psychiatry* 79, no. 7: 613–619. https://doi.org/10.1016/J.BIOPSYCH.2016.01.004.

Frazee, A. C., G. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg, and J. T. Leek. 2015. "Ballgown Bridges the Gap Between Transcriptome Assembly and Expression Analysis." *Nature Biotechnology* 33, no. 3: 243–246. https://doi.org/10.1038/nbt.3172.

Fulvio, F., R. Paris, M. Montanari, et al. 2021. "Analysis of Sequence Variability and Transcriptional Profile of Cannabinoid Synthase Genes in *Cannabis sativa* L. Chemotypes With a Focus on Cannabichromenic Acid Synthase." *Plants* 10, no. 9: 1857. https://doi.org/10.3390/PLANTS10091857/S1.

Garewal, N., N. Goyal, S. Pathania, J. Kaur, and K. Singh. 2021. "Gauging the Trends of Pseudogenes in Plants." *Critical Reviews in Biotechnology* 41, no. 7: 1114–1129. https://doi.org/10.1080/07388551.2021.1901648.

Grassa, C. J., G. D. Weiblen, J. P. Wenger, et al. 2021. "A New Cannabis Genome Assembly Associates Elevated Cannabidiol (CBD) With Hemp Introgressed Into Marijuana." *New Phytologist* 230, no. 4: 1665–1679. https://doi.org/10.1111/NPH.17243.

Hanuš, L. O., S. M. Meyer, E. Muñoz, O. Taglialatela-Scafati, and G. Appendino. 2016. "Phytocannabinoids: A Unified Critical Inventory." *Natural Product Reports* 33, no. 12: 1357–1392. https://doi.org/10.1039/C6NP00074F.

Harris, R. 2007. Improved Pairwise Alignment of Genomic DNA—ProQuest. https://www.proquest.com/openview/bc77cca0fb9390b44b9ef572fb574322/1?cbl=18750&pq-origsite=gscholar.

Hazekamp, A., A. Peltenburg, R. Verpoorte, and C. Giroud. 2005. "Chromatographic and Spectroscopic Data of Cannabinoids From *Cannabis sativa* L." *Journal of Liquid Chromatography and Related Technologies* 28, no. 15: 2361–2382. https://doi.org/10.1080/10826070500187558.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping With HISAT2 and HISAT-Genotype." *Nature Biotechnology* 37, no. 8: 907–915. https://doi.org/10.1038/S41587-019-0201-4.

Laverty, K. U., J. M. Stout, M. J. Sullivan, et al. 2019. "A Physical and Genetic Map of *Cannabis sativa* Identifies Extensive Rearrangements at the THC/CBD Acid Synthase Loci." *Genome Research* 29, no. 1: 146–156. https://doi.org/10.1101/GR.242594.118/-/DC1.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352.

Li, M. Z., and S. J. Elledge. 2012. "SLIC: A Method for Sequence- and Ligation-Independent Cloning." *Methods in Molecular Biology* 852: 51–59. https://doi.org/10.1007/978-1-61779-564-0_5.

Lynch, R. C., L. K. Padgitt-Cobb, A. R. Garfinkel, et al. 2025. "Domesticated Cannabinoid Synthases Amid a Wild Mosaic Cannabis Pangenome." *Nature* 643: 1001–1010. https://doi.org/10.1038/s41586-025-09065-0.

Lynch, R. C., D. Vergara, S. Tittes, et al. 2016. "Genomic and Chemical Diversity in Cannabis." *Critical Reviews in Plant Sciences* 35, no. 5–6: 349–363. https://doi.org/10.1080/07352689.2016.1265363/SUPPL_FILE/BPTS_A_1265363_SM2607.DOCX.

Morimoto, S., K. Komatsu, F. Taura, and Y. Shoyama. 1998. "Purification and Characterization of Cannabichromenic Acid Synthase From *Cannabis sativa*." *Phytochemistry* 49, no. 6: 1525–1529. https://doi.org/10.1016/S0031-9422(98)00278-7.

Onofri, C., E. P. M. De Meijer, and G. Mandolino. 2015. "Sequence Heterogeneity of Cannabidiolic- and Tetrahydrocannabinolic Acid-Synthase in *Cannabis sativa* L. and Its Relationship With Chemical Phenotype." *Phytochemistry* 116, no. 1: 57–68. https://doi.org/10.1016/j.phytochem.2015.03.006.

Pauli, C. S., M. Conroy, B. D. Vanden Heuvel, and S. H. Park. 2020. "Cannabidiol Drugs Clinical Trial Outcomes and Adverse Effects." *Frontiers in Pharmacology* 11: 63. https://doi.org/10.3389/FPHAR.2020.00063/BIBTEX.

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg. 2016. "Transcript-Level Expression Analysis of RNA-Seq Experiments With HISAT, StringTie and Ballgown." *Nature Protocols* 11, no. 9: 1650–1667. https://doi.org/10.1038/nprot.2016.095.

Reddy, D. S., R. H. Mbilinyi, and S. Ramakrishnan. 2023. "Efficacy of the FDA-Approved Cannabidiol on the Development and Persistence of Temporal Lobe Epilepsy and Complex Focal Onset Seizures." *Experimental Neurology* 359: 114240. https://doi.org/10.1016/J.EXPNEUROL.2022.114240.

Reid, G. A. 1991. *Molecular Cloning: A Laboratory Manual, 2nd Edn: By J. Sambrook, EF Fritsch and T. Maniatis, Cold Spring Harbor Laboratory Press, 1989. $115.00 (3 Vols; 1659 Pages)*. Elsevier Current Trends.

Schliep, K. P. 2011. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics* 27, no. 4: 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Slater, G. S. C., and E. Birney. 2005. "Automated Generation of Heuristics for Biological Sequence Comparison." *BMC Bioinformatics* 6, no. 1: 31.

Small, E. 2015. "Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization." *Botanical Review* 81, no. 3: 189–294.

Stack, G. M., J. A. Toth, C. H. Carlson, et al. 2021. "Season-Long Characterization of High-Cannabinoid Hemp (*Cannabis sativa* L.) Reveals Variation in Cannabinoid Accumulation, Flowering Time, and Disease Resistance." *GCB Bioenergy* 13, no. 4: 546–561. https://doi.org/10.1111/GCBB.12793.

Thomas, F., and O. Kayser. 2022. "Natural Deep Eutectic Solvents Enhance Cannabinoid Biotransformation." *Biochemical Engineering Journal* 180: 1369–1703. https://doi.org/10.1016/j.bej.2022.108380.

Thomas, F., and O. Kayser. 2023. "Improving CBCA Synthase Activity Through Rational Protein Design." *Journal of Biotechnology* 363: 40–49. https://doi.org/10.1016/J.JBIOTEC.2023.01.004.

Thompson, M. J., and C. D. Jiggins. 2014. "Supergenes and Their Role in Evolution." *Heredity* 113: 1–8. https://doi.org/10.1038/hdy.2014.20.

Toth, J. A., L. B. Smart, C. D. Smart, et al. 2021. "Limited Effect of Environmental Stress on Cannabinoid Profiles in High-Cannabidiol Hemp (*Cannabis sativa* L.)." *GCB Bioenergy* 13, no. 10: 1666–1674.

Toth, J. A., G. M. Stack, A. R. Cala, et al. 2020. "Development and Validation of Genetic Markers for Sex and Cannabinoid Chemotype in *Cannabis sativa* L." *GCB Bioenergy* 12, no. 3: 213–222. https://doi.org/10.1111/gcbb.12667.

van Bakel, H., J. M. Stout, A. G. Cote, et al. 2011. "The Draft Genome and Transcriptome of *Cannabis sativa*." *Genome Biology* 12, no. 10: 1–18. https://doi.org/10.1186/GB-2011-12-10-R102/COMMENTS.

Van Velzen, R., and M. E. Schranz. 2021. "Origin and Evolution of the Cannabinoid Oxidocyclase Gene Family." *Genome Biology and Evolution* 13, no. 8: evab130. https://doi.org/10.1093/gbe/evab130.

Vergara, D., L. C. Bidwell, R. Gaudino, et al. 2017. "Compromised External Validity: Federally Produced Cannabis Does Not Reflect Legal Markets." *Scientific Reports* 7: 46528. https://doi.org/10.1038/srep46528.

Weiblen, G. D., J. P. Wenger, K. J. Craft, et al. 2015. "Gene Duplication and Divergence Affecting Drug Content in *Cannabis sativa*." *Wiley Online Library* 208, no. 4: 1241–1250. https://doi.org/10.1111/nph.13562.

Wenger, J. P., C. J. Dabneyiii, M. A. Elsohly, et al. 2020. "Validating a Predictive Model of Cannabinoid Inheritance With Feral, Clinical, and Industrial *Cannabis sativa*." *American Journal of Botany* 107, no. 10: 1423–1432. https://doi.org/10.1002/ajb2.1550.

Wishart, D. S., M. Hiebert-Giesbrecht, G. Inchehborouni, et al. 2024. "Chemical Composition of Commercial Cannabis." *Journal of Agricultural and Food Chemistry* 72, no. 25: 14099–14113. https://doi.org/10.1021/ACS.JAFC.3C06616.

Zager, J. J., I. Lange, N. Srividya, A. Smith, and B. Markus Lange. 2019. "Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis." *Plant Physiology* 180, no. 4: 1877–1897. https://doi.org/10.1104/PP.18.01506.

Zirpel, B., O. Kayser, and F. Stehle. 2018. "Elucidation of Structure-Function Relationship of THCA and CBDA Synthase From *Cannabis sativa* L." *Journal of Biotechnology* 284: 17–26. https://doi.org/10.1016/j.jbiotec.2018.07.031.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Comparison of a and b assemblies for the THC-type cultivars used in this study. Position of BBE genes is indicated with cannabinoid oxidocyclases in blue, and unknown BBE genes in green. Varieties shown are Northern Lights (NLv1), Sour Diesel (SODL), White Widow (WHW), and Ace High 3-2 (AH3M). **Figure S2:** Comparison of a and b assemblies from CBD-type cultivars. Position of BBE genes is indicated with cannabinoid oxidocyclases in blue, and unknown BBE genes in green. Varieties shown are Boone County (BCM), BoAx (BOAX), Golden Redwood (GRM), Kompolti (KOMP), and YunMa (YMv1). **Figure S3:** Heterozygosity across the B locus in Golden Redwood (GRM), shown by comparison to Kompoliti (KOMPa). Position of BBE genes is indicated with cannabinoid oxidocyclases in blue, and unknown BBE genes in green. **Figure S4:** Cannabinoid chemotypes. Major and minor cannabinoid chemotype from chemotype 2 and chemotype 3 California recreational *C. sativa* cultivars. **Figure S5:** Anti-His Western blot analysis of recombinant cannabinoid oxidocyclases. **Table S1:** Characteristics of Clade B2 and Clade A3 variants. **Table S2:** (A) THCAS Degenerate FL Primers and Slic Primers used for TOPO cloning, sequencing, and recombinant bacmid construct. (B) List of Cannabis cultivars screened used for TOPO cloning and sequencing. One hundred twenty clones were selected from PBBK cultivar, and 20 clones were screened from all other cultivars. (C) Randomly selected recombinant clones from Type III and Type II cultivars, AC/DC and F-Cancer, for recombinant enzyme characterization. **Table S1:** Cultivars used for synteny analysis. **Table S2:** rnaSEQ libraries used for transcript extraction. **Table S3:** Amplification conditions. **Table S4:** Cloning oligonucleotides. **Table S5:** Cloning selection. **Table S6:** Affinity purification buffers. **Table S7:** In vitro enzyme assay reaction setup. **Data S1:** Peer review.