Journal of New Results in Science

https://dergipark.org.tr/en/pub/jnrs

Research Article

Open Access

E-ISSN:1304-7981 https://doi.org/10.54187/jnrs.1522818

Genomic analysis of secondary metabolite biosynthesis gene clusters and structural characterization of terpene synthase and cytochrome P450 enzymes in *Zingiber officinale* Roscoe

Ummahan Öz[1]

Keywords:
Bioinformatic,
Gene cluster,
Protein homology,
Zingiber officinale

Abstract — This study uses bioinformatics approaches to elucidate the genetic basis of secondary metabolite biosynthesis in *Zingiber officinale* (*Z. officinale*). To this end, it identifies 44 secondary metabolite biosynthetic gene clusters and maps onto individual chromosomes, with chromosomes 1A and 8A exhibiting higher concentrations. Here, protein homology modeling provided insights into the structural characteristics of terpene synthases and Cytochrome P450 enzymes, shedding light on their potential roles in stress response and secondary metabolite production. Moreover, the identification of enzymes, such as (-)-kolavenyl diphosphate synthase TPS28 and cytochrome P450 93A3-like, opens up new possibilities for investigating the intricate pathways involved in terpene diversity and stress response mechanisms within *Z. officinale*. This study highlights the importance of understanding the molecular mechanisms underlying plant-derived bioactive compounds for pharmaceutical applications.

Subject Classification (2020):

1. Introduction

Plants synthesize two metabolites through metabolic pathways: primary and secondary metabolites [1]. Secondary metabolites play important roles in various functions, such as pharmaceutical production, plant protection, seed germination, signal transduction, and pollinator attraction [2]. Secondary metabolites can be grouped into three main categories according to their biosynthetic pathways: nitrogen-containing compounds synthesized in the tricarboxylic acid cycle pathway, phenolic compounds synthesized in the shikimate pathway, and terpenes synthesized in the mevalonic pathway [3]. Terpenes, the largest group of secondary metabolites with more than 22,000 compounds, are found in virtually all plants and are typically synthesized by isomer dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IDP) [4]. DMAPP and isopentenyl pyrophosphate (IPP) are converted to farnesyl diphosphate and geranyl diphosphate by farnesyl diphosphate synthase and geranyl diphosphate synthase enzymes, and these compounds serve as precursors for monoterpenes, sesquiterpenes, and triterpenes [5]. The enzymes terpene synthases (TPSs) are indicated to be responsible for the diversity of carbon structures in terpenoids [6].

Zingiber officinale (*Z. officinale*) Roscoe, a member of the Zingiberaceae family, is a plant native to Southeast Asia and has been utilized for over 3000 years, primarily in India, as both a flavorful spice in culinary practices and a traditional remedy in herbal medicine [7]. *Z. officinale* has anti-inflammatory, antimicrobial, antiviral,

[1]ummahanoz48@gmail.com (Corresponding Author)
[1] Department of Plant and Animal Production, Manisa Celal Bayar University, Manisa, Türkiye
Article History: Received: 26 Jul 2024 — Accepted: 26 Aug 2024 — Published: 31 Aug 2024

antioxidant, and antifungal effects. It is used in the treatment of nausea during pregnancy, fatigue or lack of energy, rheumatic diseases, and in the treatment of gastric, colorectal, ovarian, prostate, and breast cancers [8]. In addition to that, their uses have been identified in traditional medicine for conditions such as sprains, muscular aches, sore throats, cramps, vomiting, indigestion, dementia, helminthiasis, fever, infectious diseases, and hypertension [9]. *Z. officinale* contains over 400 compounds, including major biologically active compounds such as gingerols, paradols, and shogaols [10]. The components in *Z. officinale* inhibit the growth and proliferation of cancer cells, such as pancreatic, gastric, and colorectal cancer, by inducing apoptotic effects through various pathways on cancer cell lines [11]. The selection of *Z. officinale* as the material in the current article is also attributed to its diverse medicinal properties. Cytochrome P450 is a superfamily of enzymes that plays a crucial role in the metabolism of drugs [12], and *Z. officinale*, also known as ginger, is a plant used in pharmaceutical studies. For this reason, the current article provides a detailed examination of the Cytochrome P450 enzyme category. Additionally, more than half of all compounds present in the rhizome of *Z. officinale* are formed by terpenes [13]. Therefore, the enzymes belonging to the terpene synthase category, which play a role in terpene synthesis, have been detailed in the current article. Various investigations [14-25] have been conducted on the *Z. officinale*; however, as of now, no study has delved into the analysis of gene clusters related to secondary metabolite biosynthesis. As is well known, this species is a highly significant medicinal plant containing numerous important components in its composition. Identifying the gene clusters responsible for the synthesis of these components is crucial. Studying biosynthetic gene clusters is important for elucidating significant events such as synthesizing secondary metabolites, encoding drug resistance, and other critical biological processes. Identifying genes involved in the biosynthesis of secondary metabolites is essential for discovering new drug compounds and elucidating their biological activities [26].

Bioinformatics is a field that combines biology, mathematics, computer science, and statistics, playing a significant role in analyzing and interpreting large datasets [27]. Biosynthetic gene clusters typically encode the production of secondary metabolites and various software tools have been developed to detect biosynthetic gene clusters from DNA sequences [28]. In the determination of biosynthetic gene clusters for secondary metabolites, tools, such as 2metDB, antiSMASH, BAGEL, CLUSEAN, ClusterFinder, eSNaPD, EvoMining, GNP/Genome Search, GNP/PRISM, MIDDAS-M, MIPS-CG, NaPDoS, and SMURF are utilized [29]. Initially, gene annotations are obtained from the genome under study to identify conserved biosynthetic gene clusters. Following this, biosynthetic gene clusters are detected based on the core enzymes involved in the biosynthesis of secondary metabolites [30]. One of the aims of this study is to determine the gene clusters involved in the biosynthesis of secondary metabolites in *Z. officinale*.

Additionally, the study aims to utilize bioinformatics tools to identify these biosynthetic gene clusters and detect key enzymes, such as terpene synthases and Cytochrome P450. It also seeks to elucidate the protein structures of these enzymes. Identifying the individual gene clusters responsible for secondary metabolite biosynthesis on each chromosome will lay the groundwork for future research and support focused research efforts. The outcomes of this study are anticipated to significantly contribute to drug development research. This article represents a crucial step toward understanding plant-derived bioactive compounds at the molecular level and exploring their potential applications in the pharmaceutical industry.

2. Materials and Methods

To identify the gene clusters involved in the secondary metabolite biosynthesis of the *Zingiber officinale*, the species was initially scanned in the National Center of Biotechnology Information (NCBI) database. The genome assembly Zo_v1.1 was chosen due to the presence of the Reference Sequence (RefSeq). The obtained data was downloaded and uploaded to the plantiSMASH software [31] for detecting secondary metabolite biosynthetic gene clusters. PlantiSMASH software functions by identifying loci where at least two distinct enzyme classes from a minimum of three different enzyme subclasses converge, designating these as clusters.

Enzyme classes are assigned based on position-specific scoring matrices (pHMMs) tailored to each class. At the same time, sequence-based clustering uses the CD-HIT algorithm to count enzyme subclasses and assess sequence similarities above 50%.

The results obtained from the software have been downloaded to the computer, and each cluster has been separately filed. The data for each chromosome has been saved in Microsoft Excel, and the gene clusters related to secondary metabolite biosynthesis for each chromosome have been visualized in figure form. After identifying the gene clusters related to secondary metabolite biosynthesis in *Z. officinale*, enzymes in the Cytochrome P450 and Terpene synthase categories were copied into a separate Microsoft Excel file. Subsequently, on individual pages, "Enzyme ID" was assigned based on the chromosome number for each enzyme class. Subsequently, amino acid sequence data in .fasta format was downloaded from the NCBI database to determine protein homologies and uploaded to the Phyre2 (Protein Homology/Analogy Recognition Engine V 2.0) software [32]. The obtained results have been downloaded in a compressed (RAR) file format, and figures have been generated.

3. Results

According to information gathered from the NCBI platform, *Zingiber officinale* is reported to have 22 chromosomes, with a genome size of 3.1 Gb. As a result of the analyses conducted in the current article, no secondary metabolite biosynthetic gene cluster has been observed on chromosome 3B, chromosome 11A, and chromosome 11B. The secondary metabolite biosynthetic gene clusters found on other chromosomes, their respective locations, and the enzymes they contain are detailed in supplementary material (Tables 1 and 2). Based on the data acquired, the figures are presented below in sequential order of chromosomes.

Five secondary metabolite biosynthetic gene clusters have been identified on chromosome 1A. Cluster 1 is of the saccharide type, size 706.62 kb. Its location spans from 7328285 to 8034904 nt. In this cluster, enzyme categories include methyltransferase, CoA-ligase, and glycosyltransferase. Cluster 2 is of the polyketide type, spanning from 42651153 to 43781744 in location, with a size of 1130.59 kb. The enzyme categories detected within this cluster are ketosynthase, polyprenyl synthetase, and amino oxidase. Cluster 3 is of the saccharide type, while cluster 4 is of the polyketide type. Cluster 3 has a size of 179.36 kb, whereas cluster 4 is 217.54 kb. Cluster 3 is located between 146176694 and 146356055 nt. cluster 4 is found between 186274596 and 186492132 nt. In cluster 3, enzyme categories observed include CoA-ligase, glycosyltransferase, and fatty acid desaturase, while in cluster 4, aminotransferase, epimerase, and ketosynthase have been identified. Cluster 5 has been identified as the lignan type, with a size of 599.91 kb. It spans 192892652 to 193492563 and contains the enzyme categories dirigent enzymes and cytochrome P450 (Figure 1).

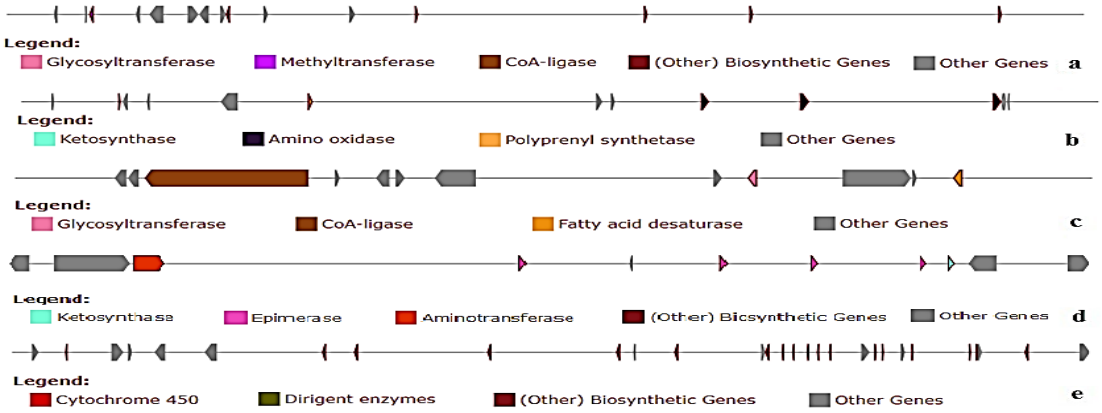

Figure 1. a) Cluster 1- saccharide, b) Cluster 2- polyketide, c) Cluster 3- saccharide, d) Cluster 4- polyketide, e) Cluster 5- lignan

Observations have revealed four clusters of secondary metabolite biosynthetic genes on Chromosome 1B, with the sequence of the first cluster being saccharide-terpene, the second cluster saccharide, the third cluster polyketide, and the fourth cluster lignan-type. Sizes of the clusters are as follows: Cluster 1 - 257.50 kb, Cluster 2 - 377.99 kb, Cluster 3 - 166.12 kb, and Cluster 4 - 483.87 kb. When evaluated in terms of locations, Cluster 1 spans from 129268146 nt. to 129525648 nt., Cluster 2 extends from 138835112 nt. to 139183102 nt., Cluster 3 covers the range from 163215785 nt. to 163381905 nt., and Cluster 4 is situated from 169767977 nt. to 170251844 nt. In addition, the enzyme categories they contain are as follows: Cluster 1 includes CoA ligase, terpene synthase, glycosyltransferase, and fatty acid desaturase; Cluster 2 consists of cytochrome P450 and glycosyltransferase; Cluster 3 involves aminotransferase, epimerase, and ketosynthase; Cluster 4 encompasses dirigent enzymes and cytochrome P450 (Figure 2).

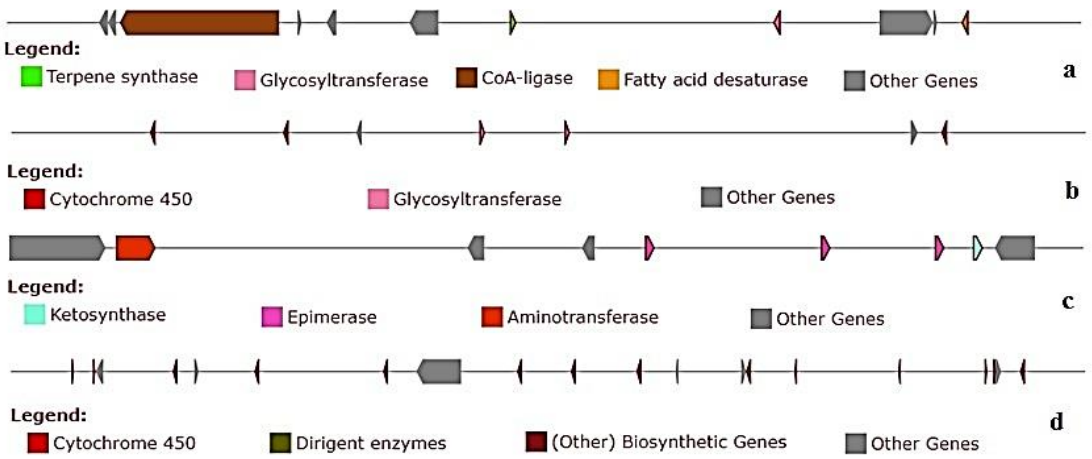

Figure 2. a) Cluster 1- saccharide-terpene, b) Cluster 2- saccharide, c) Cluster 3- polyketide, d) Cluster 4- lignan-type

According to the analysis results, two secondary metabolite biosynthetic gene clusters of polyketide and saccharide types have been identified on chromosome 2A. The size of Cluster 1 is 188.39 kb, while Cluster 2 is 327.19 kb. Cluster 1 spans from nucleotide position 77354198 to 77542590, whereas Cluster 2 is located from nucleotide position 125951799 to 126278987. The enzyme categories in Cluster 1 include ketosynthase and aminotransferase, while Cluster 2 comprises glycosyltransferase, cellulose synthase-like, and aminotransferase (Figure 3). A secondary metabolite biosynthetic gene cluster for polyketide has been identified on chromosome 2B, with a size of 188.97 kb. The location is 66468263-66657236 nt, and it contains enzymes belonging to the ketosynthase and aminotransferase categories (Figure 4).

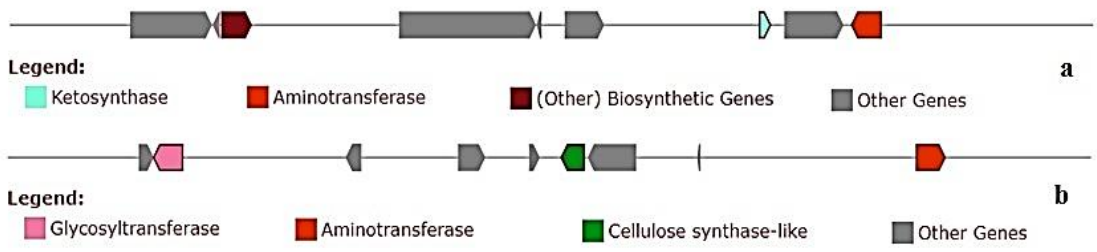

Figure 3. a) Cluster 1- polyketide, b) Cluster 2- saccharide

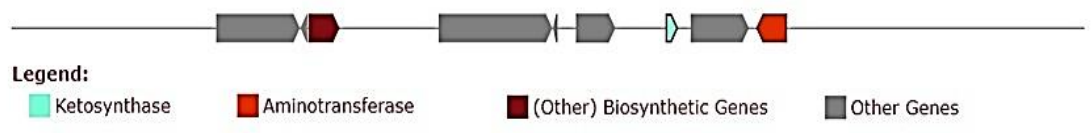

Figure 4. Polyketide gene cluster on chromosome 2B

On chromosome 3A, a secondary metabolite gene cluster of the terpene type has been identified, and its size is 186.24 kb. Furthermore, its location extends from nucleotide position 22347758 to 22533999. In addition, enzyme categories observed include dioxygenase, glycosyltransferase, and terpene synthase (Figure 5).

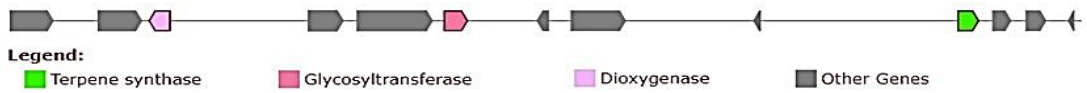

Figure 5. Terpene gene cluster on the chromosome 3A

Two secondary metabolite biosynthetic gene clusters have been identified on chromosome 4A. Cluster 1 is of the polyketide type with a size of 518.95 kb, while Cluster 2 is of the lignan-saccharide type with a size of 309.26 kb. The locations are as follows: Cluster 1 spans from nucleotide position 142297741 to 142816687, while Cluster 2 extends from nucleotide position 149885400 to 150194661. The enzyme categories are as follows: In Cluster 1, there are lipoxygenase, ketosynthase, oxidoreductase, and CoA-ligase; in Cluster 2, there are dirigent enzymes and glycosyltransferase (Figure 6). A secondary metabolite biosynthetic gene cluster of the polyketide type has been observed on chromosome 4B, with a size of 710.09 kb. This cluster is located between nucleotide positions 114712139 and 115422231 and has been found to contain enzyme categories such as ketosynthase, CoA-ligase, oxidoreductase, and lipoxygenase (Figure 7).

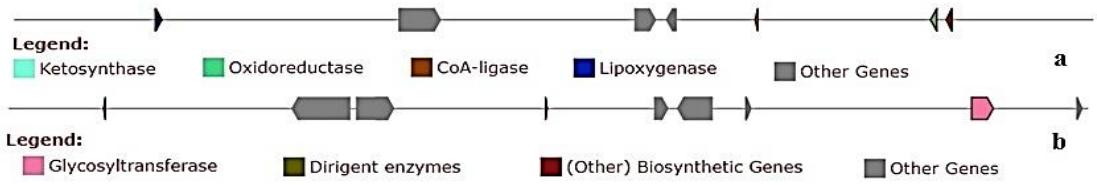

Figure 6. a) Cluster 1- polyketide, b) Cluster 2- lignan-saccharide

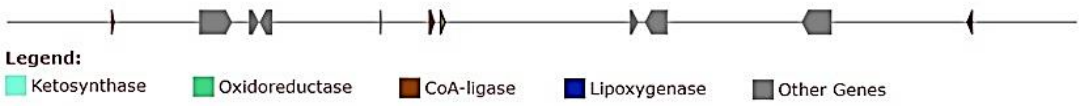

Figure 7. Polyketide gene cluster on chromosome 4B

On chromosome 5A, two distinct secondary metabolite biosynthetic gene clusters have been identified. Cluster 1 is of the polyketide type and has a size of 276.59 kb. Its location spans from nucleotide position 9574596 to 9851189 on the chromosome. Ketosynthase, methyltransferase, and fatty acid desaturase have been identified regarding enzyme categories. On the other hand, Cluster 2 is of the saccharide type and has a size of 177.40 kb. This cluster is located between nucleotide positions 124448054 and 124625455, and enzyme categories observed include glycosyltransferase and polyprenyl synthetase (Figure 8). A secondary metabolite gene cluster of the polyketide type has been analyzed on chromosome 5B, and its size is determined to be 247.96 kb. It spans from nucleotide position 10485332 to 10733290, and this cluster contains enzyme categories such as ketosynthase, methyltransferase, and fatty acid desaturase (Figure 9).

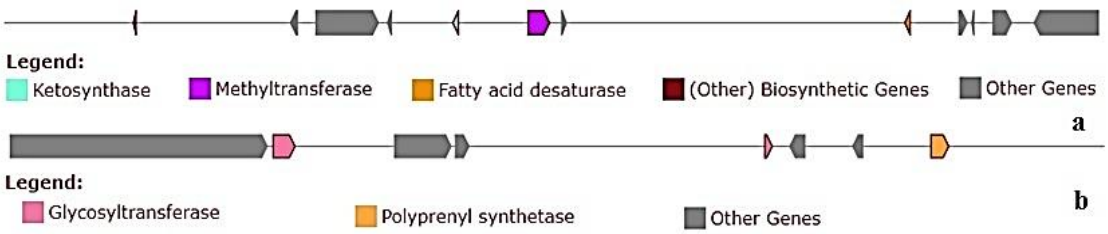

Figure 8. a) Cluster 1- polyketide, b) Cluster 2- saccharide

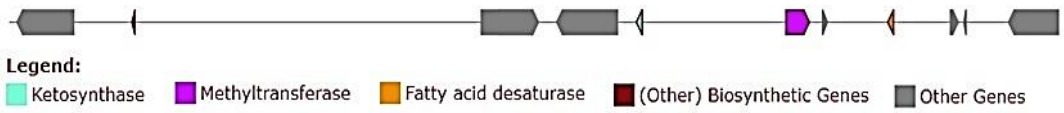

Figure 9. Polyketide gene cluster on chromosome 5B

On chromosome 6A, two types of secondary metabolite biosynthetic gene clusters have been identified, namely lignan and putative. Cluster 1 is of size 1384.62 kb, while Cluster 2 is 985.32 kb. The locations are as follows: Cluster 1 spans from nucleotide position 23617062 to 25001684, while Cluster 2 extends from nucleotide position 99578408 to 100563727. In addition, the enzyme categories are as follows: In Cluster 1, there are dirigent enzymes and Cytochrome 450, while in Cluster 2, there are lipoxygenase, oxidoreductase, and fatty acid desaturase (Figure 10). Lignan-type secondary metabolite biosynthetic gene cluster has been identified on Chromosome 6B. The size of this cluster is 727.33 kb, localized from 3845575 to 24572906 nt. Additionally, in this cluster, Scl acyltransferase, dirigent enzymes, and oxidoreductase enzyme categories have been identified (Figure 11).

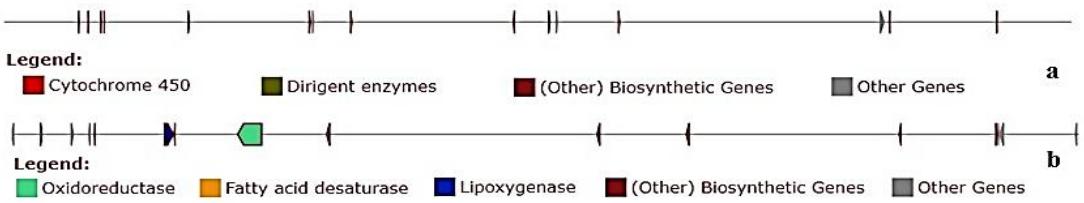

Figure 10. a) Cluster 1- lignan, b) Cluster 2- putative

Figure 11. Lignan gene cluster on chromosome 6B

Three clusters have been identified on Chromosome 7A, all of which are of saccharide type. Cluster 1 is 169.60 kb, Cluster 2 is 1514.08 kb, and Cluster 3 is 1267.58 kb in size. When evaluated in terms of their locations, Cluster 1 is observed between 84025 – 253630 nt., Cluster 2 between 46576688 – 48090770 nt., and Cluster 3 between 55068057 – 56335638 nt. Furthermore, the enzyme categories they contain are as follows: Cluster 1 includes glycosyltransferase, oxidoreductase, and BAHD acyltransferase; Cluster 2 includes Cytochrome 450, glycosyltransferase, CoA-ligase, and Scl acyltransferase; Cluster 3 includes oxidoreductase and glycosyltransferase (Figure 12). Two secondary metabolite biosynthetic gene clusters, one for saccharide and the other for polyketide type have been identified on Chromosome 7B. Cluster 1 is 355.22 kb in size, while Cluster 2 is 880.62 kb. Additionally, it has been identified that Cluster 1 extends from position 5277215 to 5632438 nucleotides, whereas Cluster 2 covers the region from position 26643208 to 27523833 nucleotides. Furthermore, the enzyme category in Cluster 1 is glycosyltransferase, while Cluster 2 includes Cytochrome 450, amino oxidase, and ketosynthase (Figure 13).

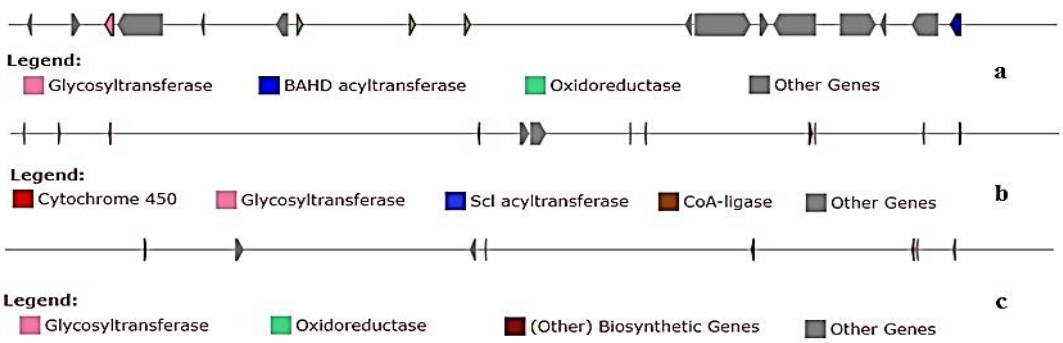

Figure 12. a) Cluster 1- saccharide, b) Cluster 2- saccharide, c) Cluster 3- saccharide

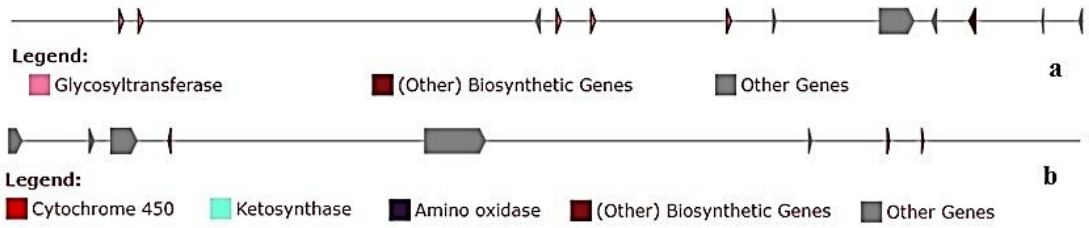

Figure 13. a) Cluster 1- saccharide, b) Cluster 2- polyketide

The analysis revealed five secondary metabolite biosynthetic gene clusters on chromosome 8A. Cluster 1, which is of the saccharide type, has a size of 240.37 kb and is located between nucleotides 10633148 and 10873514. The enzyme categories include aminotransferase, glycosyltransferase, and epimerase. Cluster 2 is of the saccharide type, with a size of 261.85 kb, and its location is between nucleotides 38297654 and 38559507. Additionally, this cluster contains enzymes in cytochrome P450, oxidoreductase, and glycosyltransferase categories. Cluster 3 is of the lignan-polyketide type, while Cluster 4 is of the saccharide type. Cluster 3 has a size of 414.08 kb, whereas Cluster 4 is 176.30 kb. Cluster 3 and Cluster 4 locations are 39117993-39532074 nt and 69732454-69908758 nt, respectively. Enzyme categories are as follows: Cluster 3 includes Ketosynthase, Dirigent enzymes, and Scl acyltransferase; Cluster 4, conversely, comprises epimerase, glycosyltransferase, and oxidoreductase. The final cluster, Cluster 5, is of the terpene-polyketide type and has a size of 819.03 kb. This cluster is located between nucleotides 130018371 and 130837398, and it contains enzymes in the categories of ketosynthase, terpene synthase, and methyltransferase (Figure 14). Four secondary biosynthetic gene clusters have been identified on Chromosome 8B, with Cluster 1 and Cluster 3 observed to be of saccharide type, Cluster 2 of polyketide type, and Cluster 4 of lignan-polyketide type. Cluster 1 is 232.74 kb, Cluster 2 is 196.19 kb, Cluster 3 is 372.31 kb and Cluster 4 is 205.09 kb in size. Cluster 1 spans from 8,488,692 to 8,720,436 nt, Cluster 2 from 27,405,491 to 27,601,678 nt, Cluster 3 from 32,315,114 to 32,687,419 nt, and Cluster 4 from 33,029,911 to 33,234,999 nt. The enzyme categories are as follows: Cluster 1 includes Aminotransferase, Glycosyltransferase, and Epimerase. Cluster 2 consists of Epimerase, Ketosynthase, and Dioxygenase. Cluster 3 encompasses Lipoxygenase, Cytochrome P450, Oxidoreductase, and Glycosyltransferase. Finally, Cluster 4 includes Ketosynthase, Dirigent enzymes, and Scl acyltransferase (Figure 15).

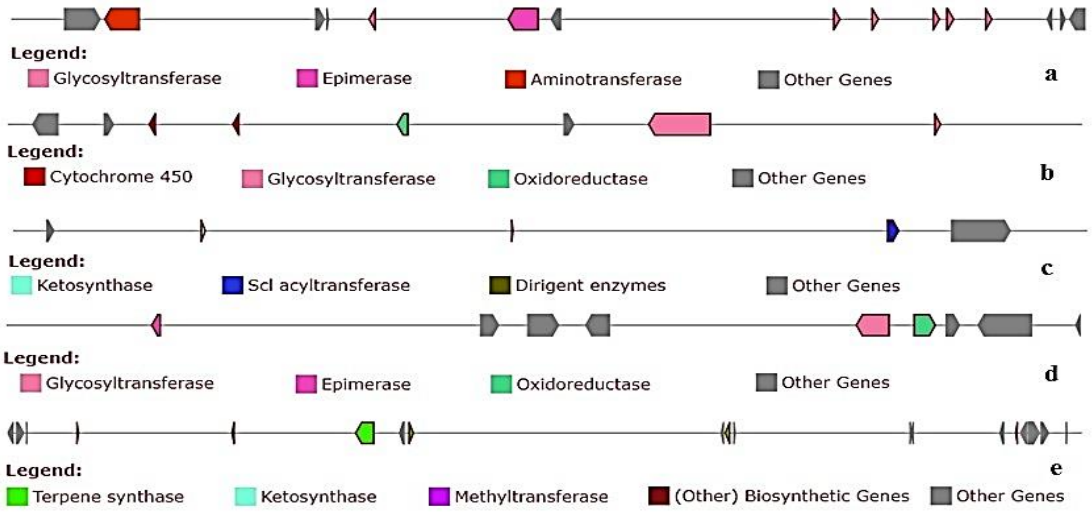

Figure 14. a) Cluster 1- saccharide, b) Cluster 2- saccharide, c) Cluster 3- lignan-polyketide, d) Cluster 4- saccharide, e) Cluster 5- terpene-polyketide

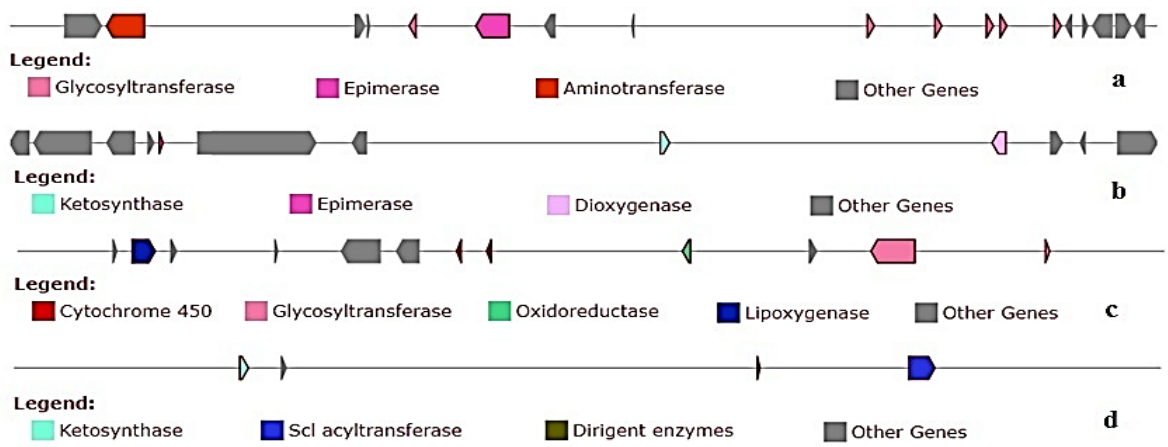

Figure 15. a) Cluster 1- saccharide, b) Cluster 2- polyketide, c) Cluster 3- saccharide, d) Cluster 4- lignan-polyketide

On Chromosome 9A, three secondary metabolite biosynthetic gene clusters have been identified. Cluster 1 is of saccharide type, Cluster 2 is of putative type, and Cluster 3 is of polyketide type. Cluster 1 is 400.98 kb, Cluster 2 is 228.58 kb, and Cluster 3 is 183.61 kb in size. When analyzed in terms of locations, it has been observed that Cluster 1 spans from nucleotide position 903,109 to 1,304,092, Cluster 2 extends from 114,121,928 to 114,350,508 nt, and Cluster 3 ranges from 133,439,172 to 133,622,782 nt. Additionally, in Cluster 1, enzyme categories include glycosyltransferase and BAHD acyltransferase; in Cluster 2, Cytochrome P450, Amino oxidase, Methyltransferase, and Scl acyltransferase are identified; while in Cluster 3, Cytochrome P450 and Methyltransferase enzyme categories have been determined (Figure 16). Chromosome 9B carries three secondary metabolite gene clusters, with the first being saccharide type and the third being polyketide type. In terms of size, Cluster 1 is 310.42 kb, Cluster 2 is 560.78 kb, and Cluster 3 is 172.30 kb. Cluster 1 is located between nucleotides 1418390 and 1728814, Cluster 2 spans from nucleotides 102414401 to 102975185, and Cluster 3 is positioned between nucleotides 117928197 and 118100500. Additionally, the enzyme categories they contain are as follows: Cluster 1 includes glycosyltransferase and BAHD acyltransferase; Cluster 2 comprises Cytochrome P450, amino oxidase, Scl acyltransferase, and glycosyltransferase; and Cluster 3 encompasses ketosynthase, Cytochrome P450, and methyltransferase (Figure 17).

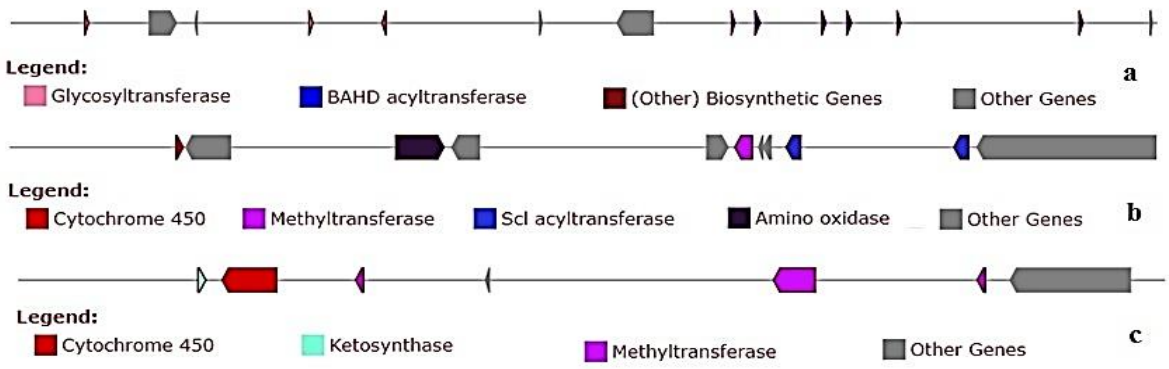

Figure 16. a) Cluster 1- saccharide, b) Cluster 2- putative, c) Cluster 3- polyketide

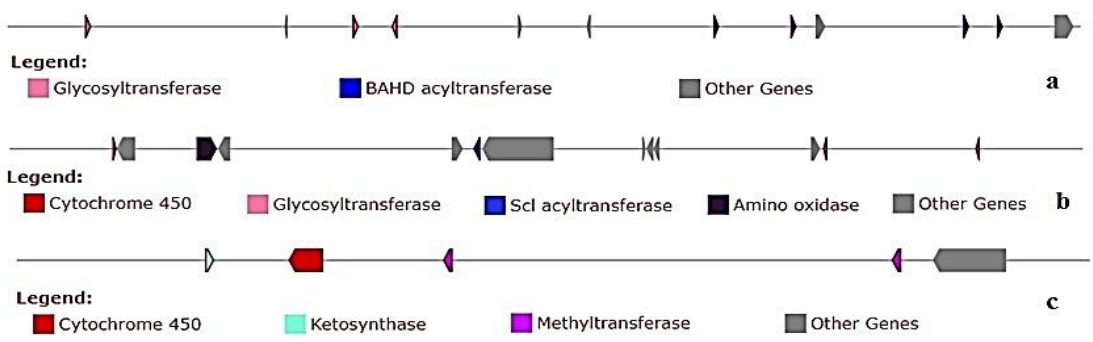

Figure 17. a) Cluster 1- saccharide, b) Cluster 2- saccharide, c) Cluster 3- polyketide

The terpene gene cluster has been detected on Chromosome 10A, with a size of 2278.99 kb. Its location spans from 89454170 to 91733163. The enzyme categories include terpene synthase and Cytochrome 450 (Figure 18). The secondary metabolite gene cluster on Chromosome 10B has been identified as terpene. It has a size of 1709.79 kb and is observed to be situated between nucleotides 86121756 and 87831544. Additionally, terpene synthase and Cytochrome 450 enzyme categories were detected within this cluster (Figure 19).

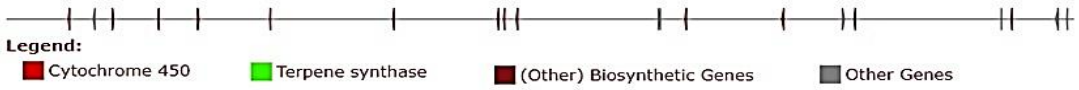

Figure 18. Terpene gene cluster on the chromosome 10A

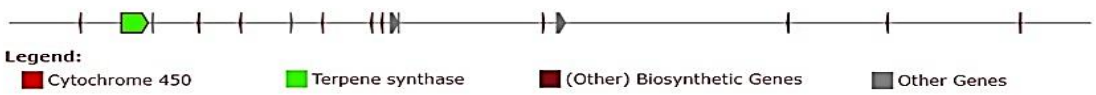

Figure 19. Terpene gene cluster on chromosome 10B

After analyzing secondary metabolite biosynthetic gene clusters, the locations of enzymes in the terpene synthase and Cytochrome P450 categories on the chromosome have been separately addressed in supplementary material (Tables 3 and 4). In the Cytochrome P450 category, the following enzymes have been observed: alpha-humulene 10-hydroxylase-like, cytochrome P450 93A3-like, flavonoid 3'-monooxygenase CYP75B137-like, alkane hydroxylase MAH1-like, cytochrome P450 734A6-like, cytochrome P450 71A9-like, cytochrome P450 77A2-like, abscisic acid 8'-hydroxylase 4-like, and cytochrome P450 94B3-like. Additionally, the presence of (-)-kolavenyl diphosphate synthase TPS28, chloroplastic-like, bifunctional isopimaradiene synthase, chloroplastic-like, (3S,6E)-nerolidol synthase 1-like, (3S,6E)-nerolidol synthase 2, chloroplastic/mitochondrial-like, and terpene synthase 13 within the category of terpene synthase enzymes has been identified.

Protein homology modeling was performed using the Phyre2 software, focusing on Cytochrome P450 and Terpene synthases. Observations reveal that the helix-loop-helix structure predominates in terpene synthases. ZoCYP1, ZoCYP2, ZoCYP3, ZoCYP4, ZoCYP5, ZoCYP6, ZoCYP8, ZoCYP10, ZoCYP11, ZoCYP12, ZoCYP13, ZoCYP14, ZoCYP15, ZoCYP17, ZoCYP19, ZoCYP33, and ZoCYP36 consist of α helices, numerous antiparallel β sheets, and long loops. ZoCYP7, ZoCYP16, ZoCYP22, ZoCYP24, ZoCYP25, ZoCYP26, and ZoCYP31 exhibit a higher presence of antiparallel β sheets. Likewise, ZoCYP23, ZoCYP29, and ZoCYP32 demonstrate the highest abundance of antiparallel β sheets, whereas ZoCYP9 and ZoCYP34 exhibit the least. Additionally, α helices, antiparallel β sheets, β turns, and long loops have been observed in ZoCYP9, ZoCYP18, ZoCYP20, ZoCYP21, ZoCYP27, ZoCYP28, and ZoCYP30. Furthermore, ZoCYP35 is characterized by numerous long loops, with a scarcity of α helix structure and antiparallel β sheets (Figure 20). After protein homology modeling, transmembrane helix prediction was performed. The analyses revealed a transmembrane helix structure in all ZoCYPs. The positions where they are located in the amino acid sequences

are given in Figure 21. When examined in terms of transmembrane helix percentage, it was determined that ZoCYP34 ranks the highest at 19%, followed by ZoCYP20 at 13%, ZoCYP25 at 13%, and ZoCYP26 at 13%. A notable point is that while the β-strand structure ranges between 6-9% in others, it is 20% in ZoCYP35.

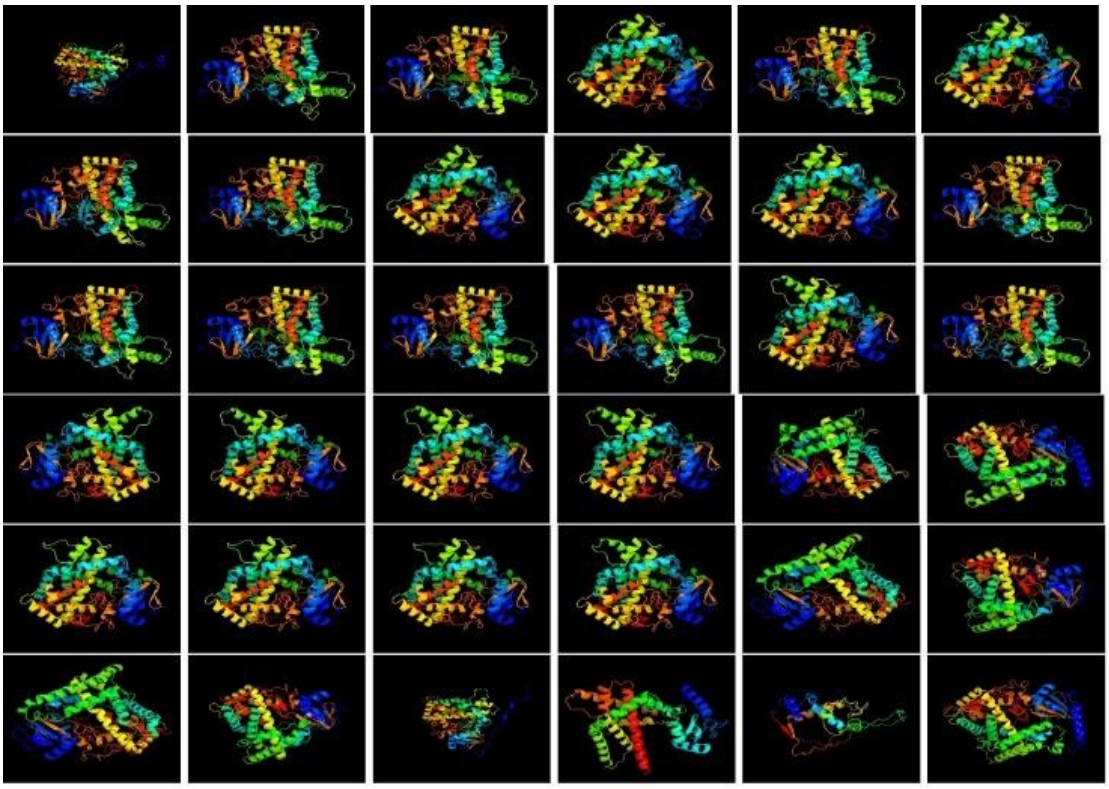

Figure 20. Predicted three-dimensional structures of ZoCYPs (sorted by protein number)

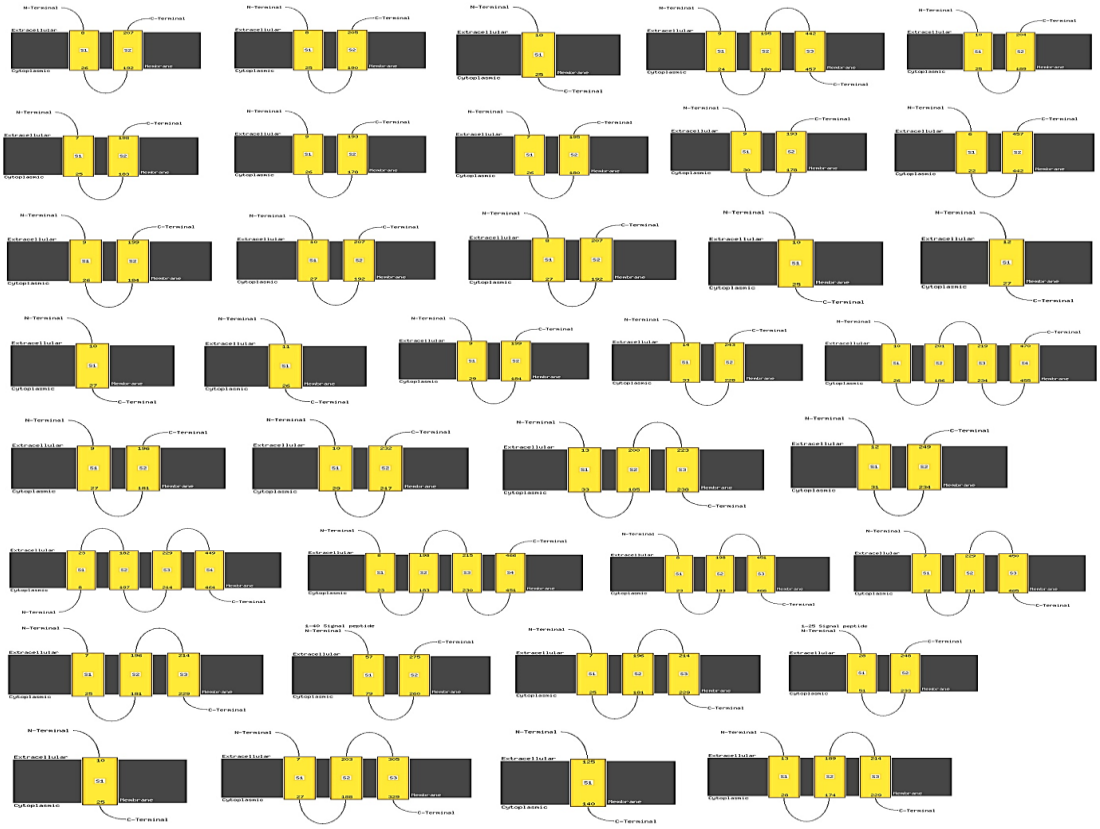

Figure 21. Transmembrane helix prediction for ZoCYPs (From ZoCYP1 to ZoCYP36 sequentially)

As a result of protein homology analysis in terpene synthases, it has been determined that the helix-loop-helix structure is dominant, and all ZoTPSs consist of α helices and long loops (Figure 22). As a result of transmembrane helix prediction analysis, a transmembrane helix structure has been identified in ZoTPS3, ZoTPS4, ZoTPS5, ZoTPS6, ZoTPS8, ZoTPS9, ZoTPS10, ZoTPS12, ZoTPS13, ZoTPS14, ZoTPS15, ZoTPS16, ZoTPS18, ZoTPS19, ZoTPS20, ZoTPS21, ZoTPS22, ZoTPS23, ZoTPS24, ZoTPS25, and ZoTPS26. The positions of these structures are detailed in Figure 23. When analyzing the amino acid sequences, ZoTPS5 exhibited the highest proportion of transmembrane helix at 13%, varying between 6% and 13% in others.

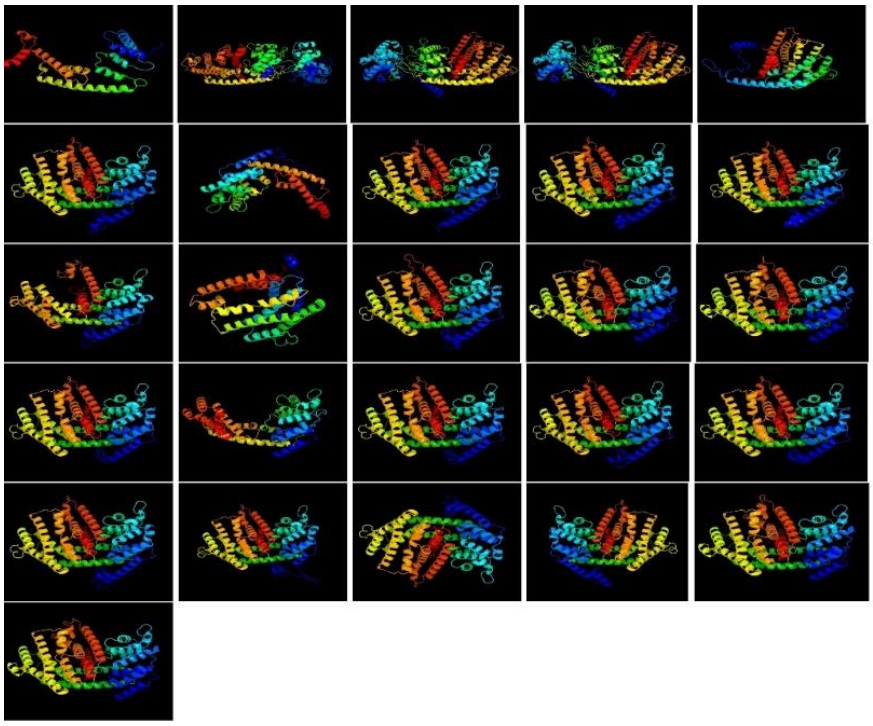

Figure 22. Predicted three-dimensional structures of ZoTPSs (sorted by protein number)

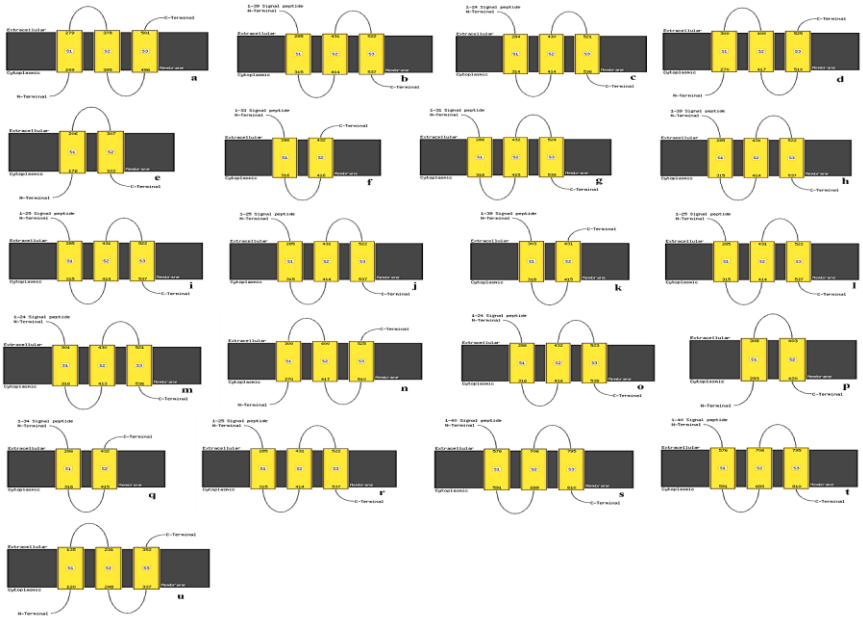

Figure 23. Transmembrane helix prediction for ZoTPSs: a)ZoTPS3, b)ZoTPS4, c)ZoTPS5, d)ZoTPS6, e)ZoTPS8, f)ZoTPS9, g)ZoTPS10, h)ZoTPS12, i)ZoTPS13, j)ZoTPS14, k)ZoTPS15, l) ZoTPS16, m)ZoTPS18, n)ZoTPS19, o) ZoTPS20, p)ZoTPS21, q)ZoTPS22, r)ZoTPS23, s)ZoTPS24, t)ZoTPS25, u)ZoTPS26

4. Discussion

Secondary metabolites in plants play important roles in innate immunity, plant growth, and development processes, developing responses to abiotic stress conditions, and signaling defense responses [33]. The current study involves the analysis of biosynthetic gene clusters responsible for these secondary metabolites, which are medically significant and involved in physiological processes within the plant. As a result of the analyses conducted, 44 secondary metabolite biosynthetic gene clusters were identified in *Z. officinale*, with the highest number of gene clusters being determined on chromosomes 1A and 8A. The observed types of secondary metabolite biosynthetic gene clusters in this plant include saccharide, polyketide, lignan, saccharide-terpene, terpene, lignan-saccharide, putative, lignan-polyketide, and terpene-polyketide. The highest number of clusters is in the saccharide type (17 clusters), while the least is in the saccharide-terpene (1), lignan-saccharide (1), and terpene-polyketide types. Research on plants' secondary metabolite biosynthetic gene cluster analyses is not extensively available. In another study [34], a total of 40 gene clusters were identified in *Citrus sinensis* (L.) Osbeck species, with 12 being saccharide, 11 putative, 6 terpene, 3 alkaloid, 3 lignan, 2 polyketide, 1 terpene-alkaloid, 1 saccharide-terpene, and 1 terpene-saccharide-polyketide type. Another study [31] found that the following numbers of biosynthetic gene clusters were detected in various plant species: *Arabidopsis thaliana* (L.) Heynh. (45), *Arabis alpina* Georgi (35), *Arachis duranensis* Krapov. and W. C. Greg. (34), *Arachis ipaensis* Krapov. & W.C.Greg. (34), *Beta vulgaris* L. (34), *Brachypodium distachyon* Roem. & Schult. (29), *Brachypodium stacei* Catalán, Joch. Müll., L. A. J. Mur and T. Langdon (43), *Brassica napus* L. (68), *Brassica oleracea* L. (34), *Brassica rapa* L. (51), *Cajanus cajan* (L.) Huth (23), *Camelina sativa* Boiss. (88), *Cicer arietinum* L. (28), *Cucumis sativus* L. (30), *Elaeis guineensis* Jacq. (18), *Fragaria vesca* L. (35), *Glycine max* (L.) Merr. (76), *Gossypium raimondii* Ulbr. (47), *Malus domestica* (Suckow) Borkh. (51), *Manihot esculenta* Crantz (36), *Medicago truncatula* Gaertn. (54), *Oryza brachyantha* A. Chev. and Roehr. (37), *Oryza sativa* Indica (54), *Oryza sativa* Japonica (46), *Panicum virgatum* L. (53), *Phaseolus vulgaris* L. (56), *Populus trichocarpa* Torr. and A. Gray ex Hook. (48), *Prunus persica* (L.) Batsch (33), *Salix purpurea* L. (33), *Sesamum indicum* L. (41), *Solanum lycopersicum* L. (45), *Solanum tuberosum* L. (51), *Sorghum bicolor* (L.) Moench (54), *Theobroma cacao* L. (48), *Vigna radiata* (L.) R. Wilczek (42), *Vitis vinifera* L. (46), *Zea mays* L. (34). When compared to other plants, *Z. officinale* has been observed to contain a greater number of secondary metabolite biosynthetic gene clusters in some plants, fewer in others, and a similar number in some. The superior aspect of the current article compared to others is the analysis of the location of each secondary biosynthetic gene cluster on individual chromosomes. In the current article, the location of each gene cluster has been determined.

Plant gene clusters possess regulatory enzymes such as Cytochrome P450s (CYP450s), UDP-glycosyl transferases, acyl transferases, short-chain alcohol dehydrogenases, transaminases, and decarboxylases to modify the backbone of signature metabolites. Reports indicate that certain CYP450s and terpene synthases exhibit non-random associations and are distributed among various gene clusters [35]. In the current article, enzyme categories such as Amino oxidase, Aminotransferase, BAHD acyltransferase, Cellulose synthase-like, CoA-ligase, Cytochrome P450, Dioxygenase, Dirigent enzymes, Epimerase, Fatty acid desaturase, Glycosyltransferase, Ketosynthase, Lipoxygenase, Methyltransferase, Oxidoreductase, Polyprenyl synthetase, Scl acyltransferase, and Terpene synthase have been identified in *Z. officinale*. The enzymes within each category and their locations on the chromosome are provided in supplementary material (Tables 1 and 2). Upon thoroughly examining this table. It is evident that apart from the terpene synthase and Cytochrome P450 categories discussed in detail in this article, *Z. officinale* also exhibits numerous important enzymes such as UDP-glycosyltransferase (uridine diphosphate-dependent glycosyltransferase, UGT), curcumin synthase 1, curcumin synthase 3-like, and others. UGTs are a crucial enzyme family involved in the biosynthesis of natural products and play significant roles in pharmaceuticals and developing responses to abiotic stress conditions [36]. In curcumin synthesis, it is mediated by the curcumin synthase gene family, which consists of three members: curcumin synthase 1, curcumin synthase 2, and curcumin synthase 3. Curcumin synthase 2 and

curcumin synthase 3 belong to the category of type III polyketide synthases and exhibit activity similar to that of curcumin synthase 1, participating in the synthesis of curcumin [37]. Curcumin is extracted from *Curcuma longa* L. and possesses numerous medical effects such as antioxidant, neuroprotective, anti-inflammatory, antiproliferative, proapoptotic, chemotherapeutic, chemopreventive, antimalarial, and antiparasitic effects [38]. Identifying curcumin synthase 1 in *Z. officinale* is an interesting and significant finding. No available data suggests the synthesis of curcumin synthase 1 from plants other than *C. longa*. Curcumin synthase 1 in *Z. officinale* has been determined to be located on chromosome 9A and chromosome 9B, with their respective localities provided in supplementary material (Tables 1 and 2). The oil of *Z.officinale* contains monoterpenoids such as cineole, b-phellandrene, (+)-camphene, geraniol, and sesquiterpenoids such as b-sesquiphellandrene, a-zingiberene, zingiberol, ar-curcumene [39]. The presence of ar-curcumene in *Z. officinale* may be associated with the data obtained in the current article. Advanced studies could be conducted to search for Curcumin Synthases on these chromosomes and compare them with those of *C. longa*. Additionally, the presence of curcumin in *Z. officinale* could be determined, and its medicinal effects could be tested on subjects to contribute to drug development efforts.

If we focus on the enzyme categories highlighted in the current article, it is indicated that Cytochrome P450 and terpene synthase gene pairs could form the basis of terpene diversity in eudicots [40]. Terpenes and terpenoids are structures that play significant roles in cellular processes such as photosynthesis, electron transport, cell wall formation, membrane fluidity, and signaling. They are synthesized through the action of terpene synthase enzymes [41]. In the current article, it has also been determined that terpene synthase enzymes are predominantly located on chromosomes 10A and 10B. When conducting studies related to terpene synthases in *Z. officinale*, it would be advantageous to focus on chromosomes 10A and 10B. The enzyme bifunctional isopimaradiene synthase, identified in *Z. officinale*, is also reported to be present in *Picea abies* (L.) H.Karst. (Norway spruce), *Picea sitchensis* (Bong.) Carrière (Sitka spruce), and *Abies balsamea* (L.) Mill. (Balsam fir) [42]. When research on isopimaradiene synthase is conducted, it is observed that studies involving this enzyme are predominantly focused on gymnosperms. The presence of this enzyme in *Z. officinale* will contribute to the literature. One of the enzymes identified in the current paper involved in terpene synthesis is (-)-kolavenyl diphosphate synthase TPS28. While this enzyme catalyzes the conversion of geranylgeranyl diphosphate to (-)- kolavenyl diphosphate, its presence is mentioned in both *Salvia divinorum* Epling & Játiva and *Scutellaria barbata* D.Don. [43-44]. In addition, the current article also identifies (3S,6E)-nerolidol synthase 1, (3S,6E)-nerolidol synthase 1-like, and (3S,6E)-nerolidol synthase 2 enzymes. Nerolidol, a sesquiterpene alcohol involved in numerous pharmacological activities, and supporting the current article, a study [45] mentions the presence of the *(3S,6E)-nerolidol synthase 1-like* gene in *Z. officinale*. Undeniably, the (3S,6E)-nerolidol synthases present in this plant contribute to the medicinal properties of *Z. officinale*. If these enzymes are to be isolated, chromosome 10A and chromosome 10B analysis should be conducted.

Cytochrome P450 enzymes participate in the biosynthesis of important secondary metabolites such as alkaloids, terpenoids, phenylpropanoids, and plant hormones while also providing tolerance to abiotic stress conditions [46]. One of these enzymes, cytochrome P450 93A3-like, has been identified in *Zingiber officinale.* Genes synthesizing the same enzyme have also been observed in *Sesamum indicum* L., an important oilseed plant, due to *Macrophomina phaseolina* interaction [47]. Additionally, it has been reported that cytochrome P450 734A6-like is expressed under salt stress in *Setaria italica* (L.) P.Beauv. and under drought stress in *Triticum aestivum* L. [48-49]. Cytochrome P450 71A9-like is up-regulated in *Triticum aestivum* under drought stress upon application of 5-aminolevulinic acid [50]. Similarly, in tomatoes, the association of cytochrome P450 71A9-like with other proteins in response development has been documented under salinity and viral stresses [51]. Additionally, CYP75B137, involved in flavonoid biosynthesis, has been reported to be up-regulated under drought stress [52]. In *Z. officinale*, a flavonoid 3'-monooxygenase CYP75B137-like enzyme has been identified, suggesting a potential similarity in function to the CYP75B137 enzyme. Another enzyme identified, alkane hydroxylase MAH1-like, increased expression when exposed to stress in a study [53],

indicating an effort to adapt to the environment. Additionally, another study [54] mentioned that alkane hydroxylase MAH1-like is involved in important functions in the biosynthesis of ketones secondary alcohols. Data also indicates this enzyme is down-regulated under heat stress in *Michelia macclurei* Dandy [55]. Abscisic acid 8'-hydroxylase 4 is an enzyme involved in abscisic acid biosynthesis [56]. Abscisic acid 8'-hydroxylase 4-like enzyme has also been identified in *Z. officinale*. This enzyme is also necessarily involved in abscisic acid biosynthesis. In the current article, another enzyme identified in the Cytochrome P450 category in *Z. officinale* is alpha-humulene 10-hydroxylase-like. In a study [57], it was mentioned that the gene named *Zoff283768* in *Z. officinale* encodes the alpha-humulene 10-hydroxylase enzyme, which catalyzes the conversion of alpha-humulene to 8-hydroxy-alpha-humulene in zerumbone biosynthesis. The reporting of the presence of the alpha-humulene 10-hydroxylase enzyme in this article supports the data obtained in the current study. When examining the enzymes in the Cytochrome P450 category in *Z. officinale*, it has been concluded that all of them are associated with stress response. This suggests the potential for these enzymes to enhance the resistance of *Z. officinale* to biotic and abiotic stress conditions.

In the current article, protein homologies of terpene synthases and enzymes in the cytochrome P450 category have also been identified. It has been observed that, unlike in ZoTPSs, antiparallel β sheets are present in the structure of ZoCYPs. Furthermore, while transmembrane helix structures have been identified in all ZoCYPs, they have been predicted in some ZoTPSs. Many crucial cellular functions are carried out through membrane proteins found in the cell membrane, and most transmembrane proteins cross the membrane with transmembrane helices [58-59]. The presence of these structures in ZoCYPs suggests significant roles in maintaining the balance between intracellular and extracellular. Additionally, it has been previously mentioned that enzymes in this enzyme category in *Z. officinale* have roles in stress response development, and it is also considered that the contribution of the transmembrane helix structure to this response development is possible.

5. Conclusion

This research focused on the gene clusters responsible for synthesizing secondary metabolites in *Zingiber officinale* and the important enzymes involved. The findings reveal the distribution of gene clusters across different chromosomes of *Z. officinale*, particularly highlighting the concentration of terpene synthases and Cytochrome P450 enzymes on specific chromosomes. This contributes to the localization and understanding of the genetic components involved in synthesizing secondary metabolites in the plant. Furthermore, the analysis of protein structures of these enzymes provides valuable insights into their functionality and interactions. For instance, protein modeling of terpene synthases elucidates their structural properties, shedding light on the potential effects on the plant's terpene synthesis. The results of this study could serve as a foundation for further research in plant biochemistry, genetics, and pharmacology. Investigating gene clusters and enzymes is an important step towards understanding the biological basis of herbal medicines and improving their utilization in the pharmaceutical industry. Deepening the understanding of the genetic structure of *Z. officinale* may also encourage applicable practices in herbal medicine development and agricultural industries to enhance productivity.

Author Contributions

The author read and approved the final version of the paper.

Conflict of Interest

The author declares no conflict of interest.

Ethical Review and Approval

No approval from the Board of Ethics is required.

Supplementary Material

https://dergipark.org.tr/en/download/journal-file/31863

References

[1] Anjali, S. Kumar, T. Korra, R. Thakur, R. Arutselvan, A. S. Kashyap, Y. Nehela, V. Chaplygin, T. Minkina, C. Keswani, *Role of plant secondary metabolites in defence and transcriptional regulation in response to biotic stress,* Plant Stress 8 (2023) Article Number 100154 19 pages.

[2] X. Zhan, Z. Chen, R. Chen, C. Shen, *Environmental and genetic factors involved in plant protection-associated secondary metabolite biosynthesis pathways*, Frontiers in Plant Science 13 (2022) Article Number 877304 14 pages.

[3] R. Jan, S. Asaf, M. Numan, Lubna, K. M. Kim, *Plant secondary metabolite biosynthesis and transcriptional regulation in response to biotic and abiotic stress conditions*, Agronomy 11(5) (2021) Article Number 968 31 pages.

[4] J. M. Al-Khayri, R. Rashmi, V. Toppo, P. B. Chole, A. Banadka, W. N. Sudheer, P. Nagella, W. F. Shehata, M. Q. Al-Mssallem, F. M. Alessa, M. I. Almaghasla, A. A. S. Rezk, *Plant secondary metabolites: The weapons for biotic stress management*, Metabolites 13 (6) (2023) Article Number 716 37 pages.

[5] S. Zhang, L. Zhang, H. Zou, L. Qiu, Y. Zheng, D. Yang, Y. Wang, *Effects of light on secondary metabolite biosynthesis in medicinal plants*, Frontiers in Plant Science 12 (2021) Article Number 781236 16 pages.

[6] E. Alicandri, A. R. Paolacci, S. Osadolor, A. Sorgonà, M. Badiani, M. Ciaffi, *On the evolution and functional diversity of terpene synthases in the pinus species: A review*, Journal of Molecular Evolution 88 (3) (2020) 253-283.

[7] Z. Yang, Z. Guo, J. Yan, J. Xie, *Nutritional components, phytochemical compositions, biological properties, and potential food applications of ginger (Zingiber officinale): A comprehensive review*, Journal of Food Composition and Analysis 128 (2024) Article Number 106057 16 pages.

[8] F. Yousfi, F. Abrigach, J. D. Petrovic, M. Sokovic, M. Ramdani, *Phytochemical screening and evaluation of the antioxidant and antibacterial potential of Zingiber officinale extracts*, South African Journal of Botany 142 (2021) 433–440.

[9] S. Xiang, Q. Jian, W. Chen, Q. Xu, J. Li, C. Wang, R. Wang, D. Zhang, J. Lin, C. Zheng, *Pharmacodynamic components and mechanisms of ginger (Zingiber officinale) in the prevention and treatment of colorectal cancer*, Journal of Ethnopharmacology 324 (2024) Article Number 117733 20 pages.

[10] I. Boarescu, R. M. Pop, P. M. Boarescu, I. C. Bocșan, D. Gheban, A. E. Bulboacă, A. D. Buzoianu, S. D. Bolboacă, *Ginger (Zingiber officinale) root capsules enhance analgesic and antioxidant efficacy of diclofenac sodium in experimental acute inflammation*, Antioxidants 12 (3) (2023) Article Number 745 19 pages.

[11] N. T. Anh Nga, S. Sathiyavimal, L. A. Al-Humaid, N. D. Al- Dahmash, J. Lee, S. Barathi, G. K. Jhanani, *Deciphering the anticancer, anti-inflammatory and antioxidant potential of Ti nanoparticles fabricated using Zingiber officinale*, Environmental Research 236 (1) (2023) Article Number 116748 10 pages.

[12] L. T. D. Würger, J. Alarcan, A. Braeuning, *Effects of marine biotoxins on drug-metabolizing cytochrome P450 enzymes and their regulation in mammalian cells*, Archives of Toxicology, 98 (2024) 1311–1322.

[13] E. Gumbarewicz, A. Jarząb, A. Stepulak, W. Kukula-Koch, *Zingiber officinale Rosc. in the treatment of metabolic syndrome disorders-A review of in vivo studies*, International Journal of Molecular Sciences 23 (24) (2022) Article Number 15545 15 pages.

[14] P. Awasthi, A. Singh, G. Sheikh, V. Mahajan, A. P. Gupta, S. Gupta, Y. S. Bedi, S. G. Gandhi, *Mining and characterization of EST-SSR markers for Zingiber officinale Roscoe with transferability to other species of Zingiberaceae*, Physiology and Molecular Biology of Plants 23 (4) (2017) 925–931.

[15] Z. Chen, N. Tang, H. Li, G. Liu, L. Tang, *Genome-wide transcriptomic analysis during rhizome development of ginger (Zingiber officinale Roscoe.) reveals hormone and transcriptional regulation involved in cellulose production*, Scientia Horticulturae 264 (2020) Article Number 109154 11 pages.

[16] S. Han, X. Han, Y. Li, F. Guo, C. Qi, Y. Liu, S. Fang, J. Yin, Y. Zhu, *Genome-wide characterization and function analysis of ginger (Zingiber officinale Roscoe) ZoGRFs in responding to adverse stresses*, Plant Physiology and Biochemistry 207 (2024) Article Number 108392 13 pages.

[17] S. Han, X. Han, C. Qi, F. Guo, J. Yin, Y. Liu, Y. Zhu, *Genome-wide identification of DUF668 gene family and expression analysis under F. solani, chilling, and waterlogging stresses in Zingiber officinale*, International Journal of Molecular Sciences 25 (2) (2024) Article Number 929 15 pages.

[18] N. A. Ismail, M. Y. Rafii, T. M. M. Mahmud, M. M. Hanafi, G. Miah, *Molecular markers: A potential resource for ginger genetic diversity studies*, Molecular Biology Reports 43 (12) (2016) 1347–1358.

[19] D. Jiang, M. Xia, H. Xing, M. Gong, Y. Jiang, H. Liu, H. L. Li, *Exploring the heat shock transcription factor (HSF) gene family in ginger: A genome-wide investigation on evolution, expression profiling, and response to developmental and abiotic stresses*, Plants 12 (16) (2023) Article Number 2999 18 pages.

[20] Y. Jiang, D. Jiang, M. Xia, M. Gong, H. Li, H. Xing, X. Zhu, H. L. Li, *Genome-wide identification and expression analysis of the TCP gene family related to developmental and abiotic stress in ginger*, Plants 12 (19) (2023) Article Number 3389 21 pages.

[21] S. Tian, Y. Wan, D. Jiang, M. Gong, J. Lin, M. Xia, C. Shi, H. Xing, H. L. Li, *Genome-Wide Identification, Characterization, and Expression Analysis of GRAS Gene Family in Ginger (Zingiber officinale Roscoe)*, Genes 14 (1) (2023) Article Number 96 19 pages.

[22] H. Xing, Y. Jiang, Y. Zou, X. Long, X. Wu, Y. Ren, Y. Li, H. L. Li, *Genome-wide investigation of the AP2/ERF gene family in ginger: evolution and expression profiling during development and abiotic stresses*, BMC Plant Biology 21 (1) (2021) 1–21.

[23] H. Xing, Y. Li, Y. Ren, Y. Zhao, X. Wu, H. L. Li, *Genome-wide investigation of microRNAs and expression profiles during rhizome development in ginger (Zingiber officinale Roscoe)*, BMC Genomics 23 (1) (2022) 1–13.

[24] X. Yao, F. Meng, L. Wu, X. Guo, Z. Sun, W. Jiang, J. Zhang, J. Wu, S. Wang, Z. Wang, X. Su, X. Dai, C. Qu, S. Xing, *Genome-wide identification of R2R3-MYB family genes and gene response to stress in ginger*, Plant Genome 17(1) (2024) Article Number e20258 12 pages.

[25] P. Zhang, D. Liu, J. Ma, C. Sun, Z. Wang, Y. Zhu, X. Zhang, Y. Liu, *Genome-wide analysis and expression pattern of the ZoPP2C gene family in Zingiber officinale Roscoe*, BMC Genomics 25 (1) (2024) 1–20.

[26] C. L. M. Gilchrist, Y. H. Chooi, *Clinker & clustermap.js: Automatic generation of gene cluster comparison figures*, Bioinformatics 37 (16) (2021) 2473–2475.

[27] M. Tu, S. Cheng, W. Lu, M. Du, *Advancement and prospects of bioinformatics analysis for studying bioactive peptides from food-derived protein: Sequence, structure, and functions*, TrAC Trends in Analytical Chemistry 105 (2018) 7–17.

[28] B.R. Terlouw, K. Blin, J. C. Navarro-Muñoz, N. E. Avalon, M. G. Chevrette, S. Egbert, S. Lee, D. Meijer, M. J. J. Recchia, Z. L. Reitz, J. A. van Santen, N. Selem-Mojica, T. Tørring, L. Zaroubi, M. Alanjary, G. Aleti, C. Aguilar, S. A. A. Al-Salihi, H.E. Augustijn, J.A. Avelar-Rivas, L.A. Avitia-Domínguez, F. Barona-Gómez, J. Bernaldo-Agüero, V.A. Bielinski, F. Biermann, T.J. Booth, V. J. Carrion Bravo, R. Castelo-Branco, F. O. Chagas, P. Cruz-Morales, C. Du, K. R. Duncan, A. Gavriilidou, D. Gayrard, K. Gutiérrez-García, K. Haslinger, E. J. N. Helfrich, J. J. J. van der Hooft, A. P. Jati, E. Kalkreuter, N. Kalyvas, K. B. Kang, S. Kautsar, W. Kim, A. M. Kunjapur, Y. X. Li, G. M. Lin, C. Loureiro, J. J. R. Louwen, N. L. L. Louwen, G. Lund, J. Parra, B. Philmus, B. Pourmohsenin, L. J. U. Pronk, A. Rego, D. A. B. Rex, S. Robinson, L. R. Rosas-Becerra, E. T. Roxborough, M. A. Schorn, D. J. Scobie, K. S. Singh, N. Sokolova, X. Tang, D. Udwary, A. Vigneshwari, K. Vind, S. P. J. M. Vromans, V. Waschulin, S. E. Williams, J. M. Winter, T. E. Witte, H. Xie, D. Yang, J. Yu, M. Zdouc, Z. Zhong, J. Collemare, R. G. Linington, T. Weber, M. H. Medema, *MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters*, Nucleic Acids Research 51 (1) (2023) 603–610.

[29] T. Weber, H. U. Kim, *The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production*, Synthetic and Systems Biotechnology 1 (2) (2016) 69–79.

[30] K. Blin, H. U. Kim, M. H. Medema, T. Weber, *Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters*, Briefings in Bioinformatics 20 (4) (2018) 1103–1113.

[31] S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, M. H. Medema, *PlantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters*, Nucleic Acids Research 45 (1) (2017) 55–63.

[32] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, *The Phyre2 web portal for protein modeling, prediction and analysis*, Nature Protocols 10 (6) (2016) 845–858.

[33] Z. Pang, J. Chen, T. Wang, C. Gao, Z. Li, L. Guo, J. Xu, Y. Cheng, *Linking Plant Secondary Metabolites and Plant Microbiomes: A Review*, Frontiers in Plant Science 12 (2021) Article Number 621276 22 pages.

[34] U. Öz, *Bioinformatics-based detection of secondary metabolite biosynthetic gene clusters and enzymes involved in terpene synthesis in Citrus sinensis (L.) Osbeck*, South African Journal Botany 168 (2024) 32–45.

[35] R. Bharadwaj, S. R. Kumar, A. Sharma, R. Sathishkumar, *Plant Metabolic Gene Clusters: Evolution, Organization, and Their Applications in Synthetic Biology*, Frontiers in Plant Science 12 (2021) 1–23.

[36] M. Wang, Q. Ji, B. Lai, Y. Liu, K. Mei, *Structure-function and engineering of plant UDP-glycosyltransferase*, Computational and Structural Biotechnology Journal 21 (2023) 5358–5371.

[37] R. Santhoshkumar, A. Yusuf, *In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of Curcuma longa*, Journal of Genetic Engineering and Biotechnology 18 (1) (2020) Article Number 24 9 pages.

[38] M. Urošević, L. Nikolić, I. Gajić, V. Nikolić, A. Dinić, V. Miljković, *Curcumin: Biological Activities and Modern Pharmaceutical Forms*, Antibiotics (Basel) 11 (2) (2022) Article Number 135 27 pages.

[39] K. Ashraf, A. Ahmad, A. Chaudhary, M. Mujeeb, S. Ahmad, M. Amir, N. Mallick, *Genetic diversity analysis of Zingiber officinale Roscoe by RAPD collected from subcontinent of India*, Saudi Journal of Biological Sciences 21 (2) (2014) 159–165.

[40] S. J. Smit, B. R. Lichman, *Plant biosynthetic gene clusters in the context of metabolic evolution*, Natural Product Reports 39 (7) (2022) 1465–1482.

[41] R. C. Rabara, C. Kudithipudi, M. P. Timko, *Identification of Terpene-Related Biosynthetic Gene Clusters in Tobacco through Computational-Based Genomic, Transcriptomic, and Metabolic Analyses*, Agronomy 13 (6) (2023) Article Number 1632 14 pages.

[42] S. Diao, Y. Zhang, Q. Luan, X. Ding, J. Sun, J. Jiang, *Identification of TPS-d subfamily genes and functional characterization of three monoterpene synthases in slash pine*, Industrial Crops and Products 188 (2022) Article Number 115609 11 pages.

[43] Y. Wang, Y. C. Liu, W. Y. Li, K. Guo, Y. Liu, S. H. Li, *Antifeedant, cytotoxic, and anti-inflammatory neo-clerodane diterpenoids in the peltate glandular trichomes and fresh leaves of Ajuga forrestii*, Phytochemistry 186 (2021) Article Number 112731 9 pages.

[44] T. Qiu, Y. Y. Li, H. Wu, H. Yang, Z. Peng, Z. Du, Q. Wu, H. Wang, Y. Shen, L. Huang, *Tandem duplication and sub-functionalization of clerodane diterpene synthase originate the blooming of clerodane diterpenoids in Scutellaria barbata*, The Plant Journal 116 (2) (2023) 375–388.

[45] Q. Wei, K. Lan, Y. Liu, R. Chen, T. Hu, S. Zhao, X. Yin, T. Xie, *Transcriptome analysis reveals regulation mechanism of methyl jasmonate-induced terpenes biosynthesis in Curcuma wenyujin*, PLoS One 17 (6) (2022) Article Number e0270309 20 pages.

[46] G. Nie, F. Wu, Z. Duan, S. Wang, B. Ao, P. Zhou, J. Zhang, *Genome-wide analysis of the cytochrome P450 superfamily suggests its roles in coumarin biosynthesis and salt stress response in Melilotus albus*, Environmental and Experimental Botany 220 (2024) Article Number 105718.

[47] N. Radadiya, N. Mangukia, V. Antala, H. Desai, H. Chaudhari, T. L. Dholaria, D. Dholaria, R. S. Tomar, B. A. Golakiya, M. K. Mahatma, *Transcriptome analysis of sesame-Macrophomina phaseolina interactions revealing the distinct genetic components for early defense responses*, Physiology and Molecular Biology of Plants 27 (8) (2021) 1675–1693.

[48] R. R. Al-Hindi, A. Bahieldin, S. Edris, M. G. Alharbi, A. A. Abulfaraj, *Expression Profiling of Transcription Factors Under Drought Stress At Seedling Stage in Bread Wheat (Triticum aestivum L.)*, Applied Ecology and Environmental Research 21 (5) (2023) 3965–3996.

[49] A. Sicilia, G. Testa, D. F. Santoro, S. L. Cosentino, A. R. Lo Piero, *RNASeq analysis of giant cane reveals the leaf transcriptome dynamics under long-term salt stress*, BMC Plant Biology 19 (1) (2019) 1–24.

[50] Y. Wang, X. Li, N. Liu, S. Wei, J. Wang, F. Qin, B. Suo, *The iTRAQ-based chloroplast proteomic analysis of Triticum aestivum L. leaves subjected to drought stress and 5-aminolevulinic acid alleviation reveals several proteins involved in the protection of photosynthesis*, BMC Plant Biology 20 (1) (2020) Article Number 96 17 pages.

[51] C. Gharsallah, S. Gharsallah Chouchane, S. Werghi, M. Mehrez, H. Fakhfakh, F. Gorsane, *Tomato contrasting genotypes responses under combined salinity and viral stresses*, Physiology and Molecular Biology of Plants 26 (7) (2020) 1411–1424.

[52] M. Salami, B. Heidari, J. Batley, J. Wang, X. L. Tan, C. Richards, H. Tan, *Integration of genome-wide association studies, metabolomics, and transcriptomics reveals phenolic acid- and flavonoid-associated genes and their regulatory elements under drought stress in rapeseed flowers*, Frontiers in Plant Science 14 (January) (2023) 1–25.

[53] U. Karki, P. Perez Sanchez, S. Chakraborty, B. Dickey, J. Vargas Ulloa, N. Zhang, J. Xu, *Intracellular trafficking and glycosylation of hydroxyproline-O-glycosylation module in tobacco BY-2 cells is*

dependent on medium composition and transcriptome analysis, Scientific Reports 13 (1) (2023) Article Number 13506 15 pages.

[54] J. Ma, Y. Geng, W. Pei, M. Wu, X. Li, G. Liu, D. Li, Q. Ma, X. S. Zang, S. Yu, J. Zhang, J. Yu, *Genetic variation of dynamic fiber elongation and developmental quantitative trait locus mapping of fiber length in upland cotton (Gossypium hirsutum L.)*, BMC Genomics 19 (2018) Article Number 882 18 pages.

[55] S. Wei, Z. Song, S. Luo, Y. Zhong, Y. Zhou, R. Lu, *Transcriptome Analysis Reveals the Heat Stress Response Genes by Fire Stimulation in Michelia macclurei Dandy*, Forests 14 (3) (2023) Article Number 610 19 pages.

[56] H. Wang, Z. Wang, M. Zhang, B. Jia, W. Heng, Z. Ye, L. Zhu, X. Xu, *Transcriptome sequencing analysis of two different genotypes of Asian pear reveals potential drought stress genes*, Tree Genetics & Genomes 14 (3) (2018) Article Number 40 14 pages.

[57] M. Huang, H. Xing, Z. Li, H. Li, L. Wu, Y. Jiang, *Identification and expression profile of the soil moisture and Ralstonia solanacearum response CYPome in ginger (Zingiber officinale)*, PeerJ 9 (2021) Article Number e11755 22 pages.

[58] J. Domanski, M. S. P. Sansom, P. J. Stansfeld, R. B. Best, *Atomistic mechanism of transmembrane helix association*, PLOS Computational Biology 16 (6) (2020) Article Number e1007919 14 pages.

[59] M. Bernhofer, B. Rost, *TMbed: transmembrane proteins predicted through language model embeddings*, BMC Bioinformatics 23 (1) (2022) Article Number 326 19 pages.