Journal of Life and Social Sciences
eISSN: 3006-2675
www.bbasrjlifess.com
J. Life Soc. Sci, 2026; Volume, 5: 61
https://doi.org/10.64013/bbasrjlifess.v2026i1.61

Original research Article *Open Access*

GENOME-WIDE CHARACTERIZATION, EVOLUTIONARY ANALYSIS, AND FUNCTIONAL PREDICTION OF THE BCCP GENE FAMILY IN *HELIANTHUS ANNUUS*

SANAM A[1*], AHMED H[2], AHMAD J[3], SHERAZI SHUH[1]

[1]*Department of Plant Breeding and Genetics, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan*
[2]*Department of Horticulture, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan*
[3]*Department of Agronomy, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan*
**Correspondence Author Email Address: abidasanam99@gmail.com*

(Received, 6th December 2025 Accepted 14th July 2026, Published 27th July 2026)

Abstract *The biotin carboxylase carrier (BCCP) is an essential component of the acetyl-CoA Carboxylase complex (ACCase), responsible for catalyzing the initial reaction of fatty acid synthesis and therefore is crucial for plant growth, development, and production of oil. Sunflower (Helianthus annuus L.) is an economically important oilseed crop; however, a genome-wide analysis of the BCCP gene family in this crop has not been reported. In this study, a genome-wide analysis of the BCCP gene family in H. annus was performed to determine its evolutionary history, duplication events, and genomic structures and predict potential functions. A total of ten HaBCCP genes were identified to be unevenly distributed across the eight sunflower chromosomes, revealing a dispersed genomic distribution largely driven by segmental duplication. Phylogenetic clustering of the BCCP genes into three distinct clades suggested that the diversification of the BCCP gene family occurred before species divergence and that the sunflower BCCP genes are orthologous to members from Arabidopsis thaliana, Brassica juncea, Gossypium hirsutum, Oryza sativa, Glycine max, and Solanum tuberosum. The Domain and Motif analysis identified highly conserved catalytic regions among numerous HaBCCP genes, while structural divergence in some genes revealed lineage-specific functional specialization. Subcellular localization analysis revealed that the majority of the BCCP proteins are targeted to chloroplasts or plastids, in line with their role in fatty acid biosynthesis, while the targeting of the other members to the cytoplasm, mitochondria, endoplasmic reticulum, and nucleus suggested wider functional diversification within the family. The cis-regulatory elements analysis revealed an abundant amount of these elements involved in hormonal signaling, light responsiveness, stress responsiveness, and development, suggesting that the HaBCCP genes are regulating diverse physiological processes. Ka/Ks analysis showed that duplicated HaBCCP gene pairs have been subjected to strong purifying selection, indicating that the function of this gene family has been conserved despite gene duplications. Additionally, intra-genomic synteny analysis confirmed segmental duplication as the primary driver of family evolution, and protein-protein interaction analysis identified a closely coordinated interaction network with HaBCCP3 and HaBCCP10 as the hub proteins. In summary, this study offers the first glimpse of the evolutionary, structural, and functional importance of the BCCP gene family in sunflower, revealing that duplication has led to family expansion while critical metabolic functions in fatty acid synthesis have been preserved. This work provides a solid genomic foundation for future functional studies of the BCCP genes and also provides potential targets for enhancing oil production and metabolic efficiency in sunflower.*

[Citation: Sanam, A., Ahmed, H., Ahmed, J., *Sherazi, S.H.U.H. (2026).* Genome-wide characterization, evolutionary analysis, and functional prediction of the BCCP gene family in *Helianthus annuus*. *J. Life Soc. Sci, 5: 61*. *https://doi.org/10.64013/bbasrjlifess.v2026i1.61]*

Keywords: *ACCase complex; phylogenetic Analysis; functional genomics; gene family evolution; oil seed crop; segmental duplication; chloroplast targeting*

Introduction

Sunflower (*Helianthus annus L.*) is a major oilseed crop with a significant contribution to the edible oil industry (Badouin et al., 2017). Sunflower oil is a nutritious and commercially important oil that can be used for human consumption and in industrial products (Badouin et al., 2017). Thus, increasing oil synthesis and identifying key molecular processes that control fat production are key goals in sunflower breeding programs. Acetyl-CoA carboxylase (ACCase) is one of the key enzyme systems involved in the biosynthesis of fatty acids and catalyzes the ATP-dependent conversion of acetyl-CoA to malonyl-CoA, which is the first committed and rate-limiting reaction in de novo fatty acid synthesis

(Sasaki and Nagano, 2004; Tong, 2005). This is a critical step in the formation of fatty acid precursors, which in turn have a significant impact on seed oil content and composition (Ohlrogge and Browse, 1995; Ohlrogge and Jaworski, 1997). Biotin carboxyl carrier protein (BCCP) is one of the subunits of plant heteromeric ACCase (Salie and Thelen, 2016). So, BCCP is a subunit that yields biotin and transports activated carboxyl groups between active sites to complete the carboxylation reaction (Cronan Jr and Waldrop, 2002; Salie and Thelen, 2016). In this way, BCCP impacts fatty acid biosynthesis, membrane lipid production and energy storage(Li-Beisson et al., 2013). Given that the synthesis of lipids supports membrane integrity, seed production and stress responsiveness, BCCP genes are regarded as being important for regulating plant growth and metabolism.

Plants commonly possess multiple BCCP genes existing as paralogs (growing through gene duplication) (Flagel and Wendel, 2009; Vision et al., 2000). Such gene copies can retain common metabolic functions while varying in gene structure, expression, regulation, and subcellular localization(Flagel and Wendel, 2009). This enables plants to regulate fatty acid production in response to changing developmental and environmental cues. Earlier research in model and crop plants like *Arabidopsis thaliana*, *Glycine max*, and *Oryza sativa* has shown that BCCP genes play roles in development, stress responses, and, of course, lipid metabolism (Salie and Thelen, 2016). Despite the significance of sunflower as an oilseed crop, the evolutionary structure and functional diversification of the BCCP gene family in H. annuus is still unknown. Whole-genome studies of gene families are an efficient way to dissect the structure, evolution and function of genes associated with key biological processes. Comparative studies of gene families using phylogenetic classification, motif and domain composition, exon-intron structure, chromosomal location, duplication events, promoter structure, and protein interaction networks can provide insights into the evolution and diversification of gene families while preserving crucial metabolic roles (Flagel and Wendel, 2009). For BCCP genes, such analyses are important to understand the evolution of core catalytic members, gene duplication and diversification, and to identify candidate genes for oil production.

Here, we conducted a genome-wide analysis of the BCCP gene family in *Helianthus annuus* to understand the structural and evolutionary characteristics. The HaBCCP genes identified were examined for phylogenetic relationships with BCCP genes from several other plant species, such as *Arabidopsis thaliana, Brassica juncea, Gossypium hirsutum, Oryza sativa, Glycine max, and Solanum tuberosum*, to evaluate evolutionary conservation and diversity. In addition, specific motifs, domains, exon-intron structures, subcellular localization, cis-regulatory elements of promoters, chromosomal distribution, gene duplication events, selection pressure, syntenic analysis, and protein-protein interaction networks were examined to gain a deeper understanding of the BCCP gene family in sunflower. The results of this study provide new insights into evolutionary conservation, diversification, and functional differentiation of HaBCCP genes, contributing to a better understanding of the lipid metabolism-related genes in sunflower. This study represents a valuable genomic resource for future molecular and functional analyses and lays the groundwork for leveraging BCCP genes in genetic improvement of sunflower oil yield and metabolic adaptation.

Materials and Methods

Identification of the BCCP genes in Helianthus annuus

BCCP protein sequences were retrieved from the UniProt database (https://www.uniprot.org/). BLASTP searches were performed in the Phytozome database (https://phytozome-next.jgi.doe.gov/).

Verification of Conserved BCCP Domains

All sequences were analysed using the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Only proteins having the conserved BCCP-related domains were retained for further analyses.

Phylogenetic Analysis

The evolutionary relationship among the BCCP proteins from Helianthus annuus, Arabidopsis thaliana, Brassica juncea, Gossypium hirsutum, Oryza sativa, Glycine max, and Solanum tuberosum was investigated through phylogenetic analysis. Protein sequences were again retrieved from the Phytozome database (https://phytozome-next.jgi.doe.gov/). The alignment of multiple sequences and the construction of a phylogenetic tree were carried out using MEGA software. The tree that was generated as thus was exported in Newick format and annotated using the Interactive Tree of Life (iTOL) platform (https://itol.embl.de/).

Prediction of Subcellular Localization

The subcellular localization of HaBCCP proteins was predicted using the WoLF PSORT server (https://wolfpsort.hgc.jp/).

Gene Structure Analysis

The intron-exon structure analysis was done using Gene Structure Display Server (GSDS v2.0) (https://gsds.gao-lab.org/). For this purpose, genomic DNA sequences and the corresponding coding sequences (CDS) were used.

Conserved Motif and Protein Domain Analysis

Conserved protein domains that were identified using the NCBI Conserved Domain Database (CDD) through the CD-Search program (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) were used to find the conserved motifs among

HaBCCP proteins using the MEME Suite (https://meme-suite.org/meme/tools/meme).

Cis Regulatory Analysis

Cis-regulatory elements were identified using the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Ka/Ks Analysis

Selection analysis acting on duplicated genes was assessed using the Ka/Ks ratio. For this purpose, synonymous (Ks) and non-synonymous (Ka) substitution rates were calculated. Divergence time (T) was estimated using the formula: $T = Ks/2\lambda$, where λ represents the neutral substitution rates for dicot plant genomes that had been reported in earlier studies.

Chromosomal localisation of HaBCCP Genes

Gene distribution across chromosomes was visualized using the MG2C v2.1 tool (http://mg2c.iask.in/mg2c_v2.1/).

Synteny Analysis

Synteny relationships among HaBCCP genes were visualized using TBtools with default parameters. Protein Protein Interaction Analysis Protein–protein interaction (PPI) analysis of HaBCCP proteins was performed using the STRING database (https://string-db.org/).

Phylogenetic Analysis

The evolutionary relationship among *Helianthus annuus, Arabidopsis thaliana, Brassica juncia, Gossypium hirsutum, Oryza sativa, Glycine max, and Solanum tuberosum* was found by carrying out phylogenetic analysis of the BCCP gene across the mentioned plant species. There were three major clades formed, which showed the divergence and conservation of this gene family during the evolution of plants (Panchy et al., 2016). In the phylogenetic tree, it is shown that the BCCP proteins from all the said plant species were distributed among these three clades rather than being clustered species-wise. It indicates that these species separated later due to evolution, while the diversification of the BCCP gene happened before it. The presence of HaBCCP members in every clade suggests that the sunflower genome preserves multiple evolutionary forms of BCCP proteins, which in turn suggests that these proteins may correspond to functionally specialized paralogs (Panchy et al., 2016).

The first clade has a relatively smaller number of BCCP proteins, including HaBCCP4. The clustering of HaBCCP4 in both dicot and monocot species indicates a highly conserved ancestral branch of the BCCP family, while the presence of OsBCCP2 within this clade indicates that even after the divergence of monocots from dicots, some BCCP members remained conserved, as shown in Figure 1. This conservation in turn suggests that the genes likely perform such biochemical functions that are necessary for the metabolism of fatty acids (Nikolau et al., 2003; Sasaki and Nagano, 2004). The second clade is important because it contains many proteins from *Helianthus annuus*, e.g., HaBCCP1, HaBCCP5, HaBCCP6, HaBCCP7, HaBCCP8, and HaBCCP9. This shows that major expansion of their BCCP in sunflower happened within this lineage, possibly through tandem or segmental duplication. This clade also includes homologs from *Arabidopsis thaliana, Brassica juncia, Gossypium hirsutum*, Glycine max, and *Solanum tuberosum*. It also suggests conserved domains and similar biochemical functions as discussed in the first clade (Nikolau et al., 2003). The third clade includes HaBCCP2 and HaBCCP3 proteins. *Gossypium hirsutum* contributed the largest number of BCCP members across all the clades, which likely reflects the polyploidy of cotton (Li et al., 2015). In contrast, the fewer members in Arabidopsis thaliana and Oryza sativa reflect relatively conserved family size and limited duplication (Panchy et al., 2016). A moderate number of BCCP genes in sunflower represents less extensive gene duplication as compared to the polyploid species, i.e., cotton and soybean (Schmutz et al., 2010). Overall, these findings prove that the BCCP gene family in *Helianthus annuus* has evolved through ancestral conservation, gene duplication, and functional divergence as shown in Figure 1. This enabled the retention of metabolic functions and also caused species-specific adaptations in lipid biosynthesis (Sasaki and Nagano, 2004).

Motif Analysis and Domain Analysis

Ten conserved motifs were found across the ten BCCP proteins in *Helianthus annuus*. There were shared conserved motifs present among most of the HaBCCP proteins. Proteins such as HaBCCP3, HaBCCP4, HaBCCP5, HaBCCP6, and HaBCCP10 showed a similar motif arrangement in the N terminal, especially motifs 1, 3, 5, and 7. It tells about the evolutionary relationships among these proteins (Bailey et al., 2009). It also suggests the presence of domains required for biotin carboxyl carrier activity. HaBCCP4, HaBCCP5, and HaBCCP6 showed highly similar motif patterns, suggesting recent duplication events, followed by limited divergence, as shown in Figure 2. HaBCCP3 and HaBCCP10 had slightly different motifs at the C-terminus, suggesting diversification in function. HaBCCP1, HaBCCP7, and HaBCCP9 can be categorised in a similar subgroup with respect to function, having motifs 2, 4, 5, 6, 8, and 9 clustered closely. The absence of motifs 1, 3, and 7 implies structural divergence. HaBCCP9 had slight differences in positioning compared to HaBCCP1 and HaBCCP9, which showed striking similarities in motif order and spacing. It all points to the supposition that HaBCCP1 and HaBCCP7 evolved from a recent duplication relative to HaBCCP9. These genes can therefore be considered a BCCP protein subfamily that has related but not identical roles in lipid metabolism in sunflower.

HaBCCP2 lacked detectable motifs, while HaBCCP8 had only 1 motif exhibiting highly reduced motif

composition. So these members can be considered highly diverged members of the BCCP gene family. It also suggests that these genes might have undergone degeneration that caused them to lose their functions as well or they might have only minimal functionality. They might be associated with neofunctionalization. Overall, the motif analysis of the *Helianthus annuus* BCCP gene family showed that there are both highly conserved and structurally different members. A main outcome of the domain analysis of *Helianthus annuus* was the identification of the domains that are associated with biotin carboxylase, especially in HaBCCP3 and HaBCCP10. HaBCCP3 and HaBCCP10 had ACC central and carboxyl transfer domains that are associated with catalytic functions of acetyl-CoA carboxylase. ACC central helps in catalytic coordination, while Carboxyl transferase transfers the activated carboxyl group during malonyl-CoA formation. HaBCCP4, HaBCCP5, and HaBCCP6 contained Biotin carb N, CPSase, and Biotin carb C domains, indicating roles in biotin-dependent carboxylation reactions.

HaBCCP1 and HaBCCP9 had an unusual combination of domains, including the pepsin retropepsin-like superfamily, RT RNase, and Integrase H2C2. These are generally associated with retrotransposons instead of the canonical BCCP proteins. So these proteins may be considered highly diverged from the canonical BCCP structure. HaBCCP2 had only the ATG superfamily domain, suggesting autophagy-related functions, and HaBCCP7 had only RVT1 and RT RNase H2 domains. So these proteins are also highly diverged. HaBCCP8 showed the most reduced structure, having only the CPSase L D2 superfamily domain and lacking the full-length BCCp protein characteristic, suggesting a truncated protein or evolutionary degeneration. So, the BCCP family has evolved through duplication, domain retention, domain loss, and domain acquisition as shown in Figure 2.

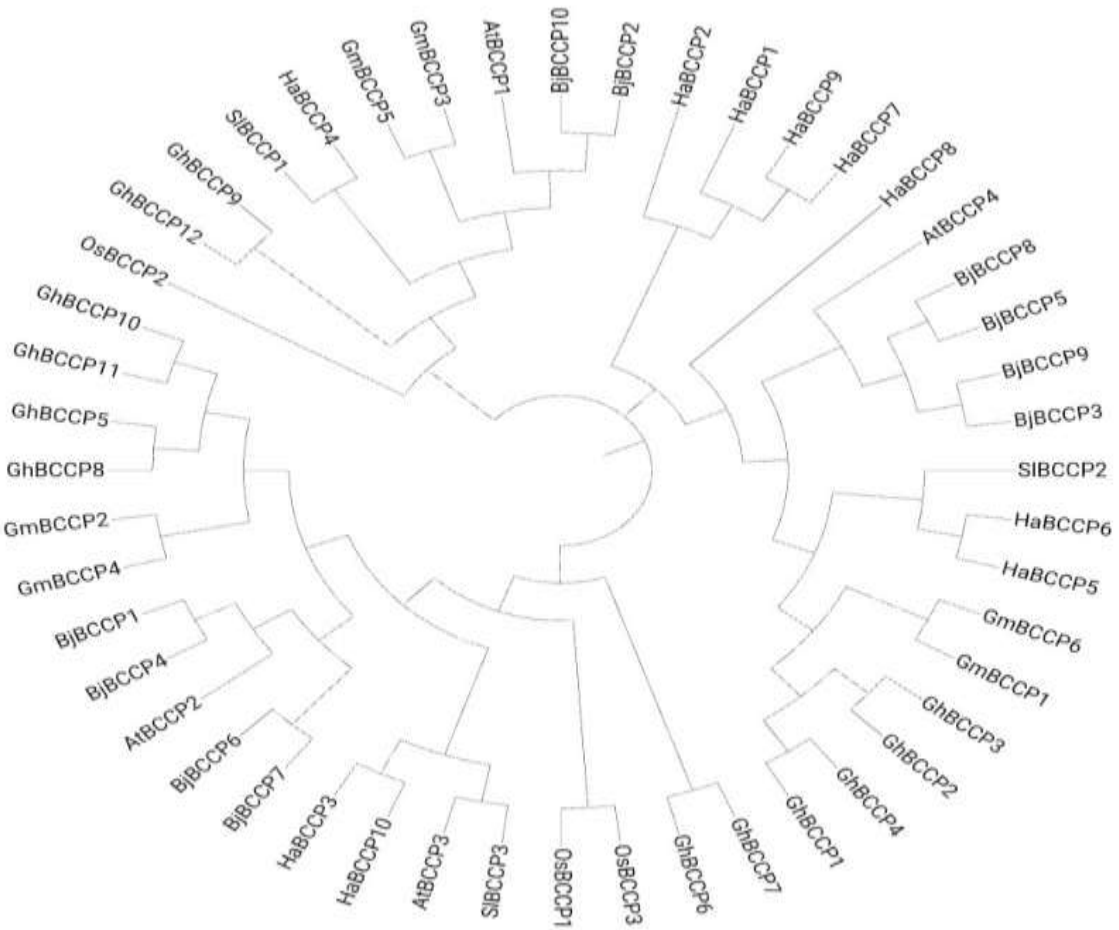

Figure 1. Polygenetic Tree

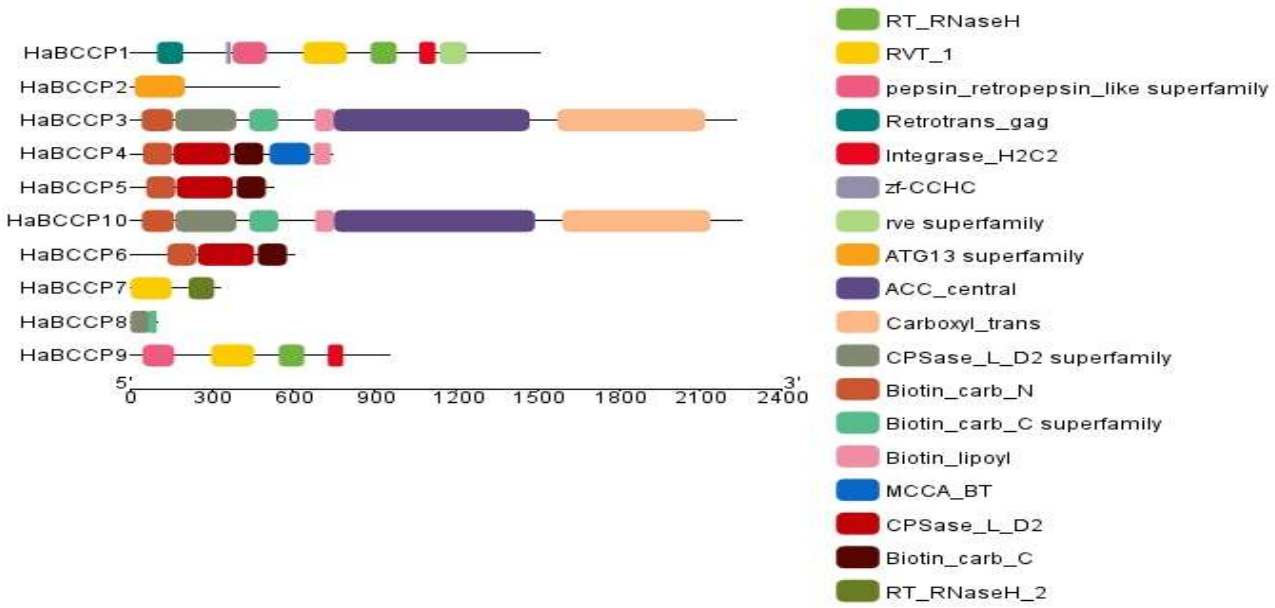

Figure 2. Domain and Motif Analysis

Intron Exon Analysis

Among the family members, HaBCCP7 had the longest genomic region with a small coding segment near the 5' region and again a long downstream non-coding region, suggesting large intron or UTR regions playing roles in transcript stability or modification, etc (Moore and Proudfoot, 2009; Rose, 2019). In contrast, HaBCCP1 was seen to be intronless, having a smooth genomic organisation that helps in efficient transcription for a functionally important BCCP protein. Many genes, including HaBCCP3, HaBCCP4, HaBCCP5, HaBCCP6, and HaBCCP10, showed highly fragmented exons because they had many small exons separated by many introns. This type of structure improves alternative splicing to produce many variants from a single gene and thus increases the versatility of the functions of proteins, as shown in Figure 3. HaBCCP10 had a genomic length of 20 kb. HaBCCP4, HaBCCP5, and HaBCCP6 exhibited multiple exons but shorter genomes, suggesting that they form a closely related subgroup within the HaBCCP gene family, as shown in Figure 3. HaBCCP2 and HaBCCP8 displayed comparatively shorter gene lengths with only a few exons and a very simple gene structure. This suggests either ancestral compaction in organisation or structural reduction following duplication. Such structural diversification underpins the adaptive capability of BCCP genes in sunflower growth, development and lipid metabolism.

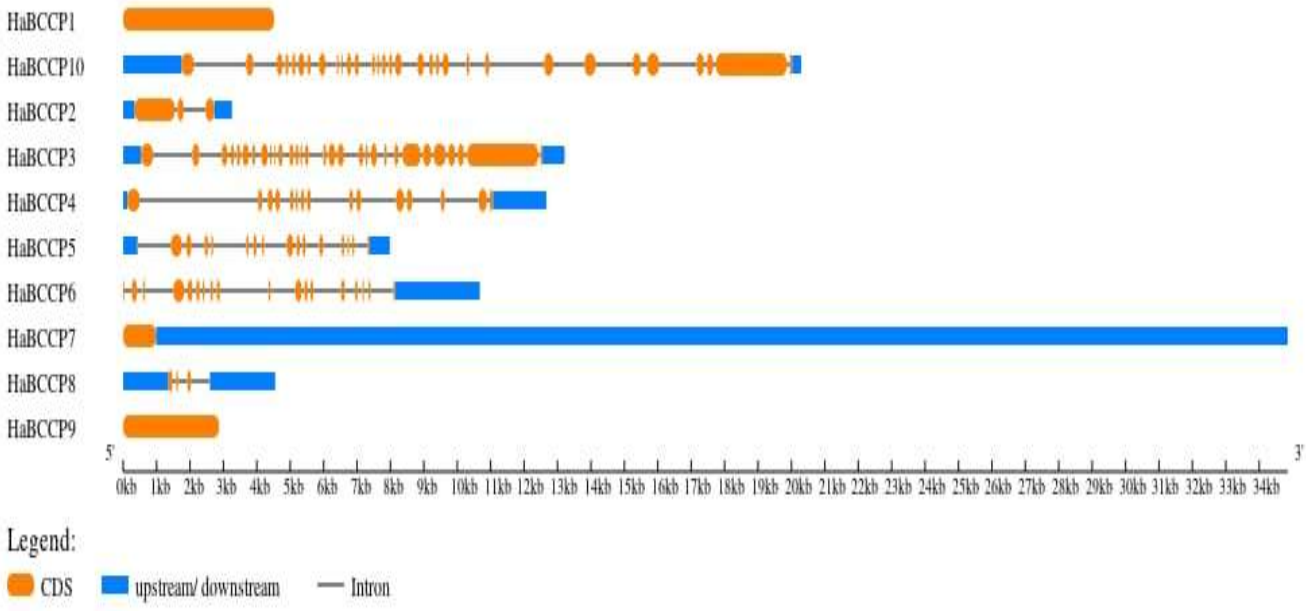

Figure 3. Intron Exon Analysis

Subcellular Localization Analysis

A major observation from this analysis was that HaBCCP1, HaBCCP2, HaBCCP4, and HaBCCP7 showed the highest localisation scores in plastids. HaBCCP4 had the highest score even among the said members, followed closely by HaBCCP1 and then HaBCCP2, as shown in Figure 4. Plastids are the primary sites of de novo fatty acid biosynthesis in plants, where acetyl-CoA catalyses the first step for this production. HaBCCP4 may be the major plastidial BCCP isoform (Ohlrogge and Browse, 1995). In addition, chloroplast localization was also observed especially for HaBCCP3 and HaBCCP10, suggesting that they might be involved in photosynthesis-related lipid metabolism. Presence of both the plastid and the chloroplast-associated fatty acid production is relevant in sunflower because lipid metabolism is important in seed oil accumulation (Li-Beisson et al., 2013). Mitochondrial localisation was also predicted by many members, e.g., HaBCCP3, HaBCCP5, HaBCCP9, and HaBCCP10, suggesting participation in carboxylation processes that are associated with mitochondria or auxiliary roles in coordination between organelles for metabolism (Montes et al., 2007). HaBCCP2, HaBCCP9, HaBCCP10, and especially HaBCCP1 showed strong nuclear localization that points to roles in signaling pathways or transcriptional regulation that might also be a part of lipid metabolism. HaBCCP6, HaBCCP7, HaBCCP8, and HaBCCP9 showed high cytoplasmic localization, with HaBCCP7 showing the highest. That entails to their participation in cytosolic metabolic pathways. In plants, cytosolic acetyl-CoA carboxylase is involved in the elongation of fatty acids. HaBCCP8 showed a relatively simple localization pattern, having limited plastid or chloroplast localization, representing a divergent member of the BCCP family. Overall, the subcellular localization analysis demonstrated a dominant plastid/chloroplast localization pattern, as shown in Figure 4. This is evidence of the conserved role of BCCP proteins in fatty acid biosynthesis. While the localization of some HaBCCP proteins to mitochondria, nucleus, ER, cytoplasm, and extracellular regions indicates diversification of functions, enhancing the adaptability of lipid metabolism in sunflower.

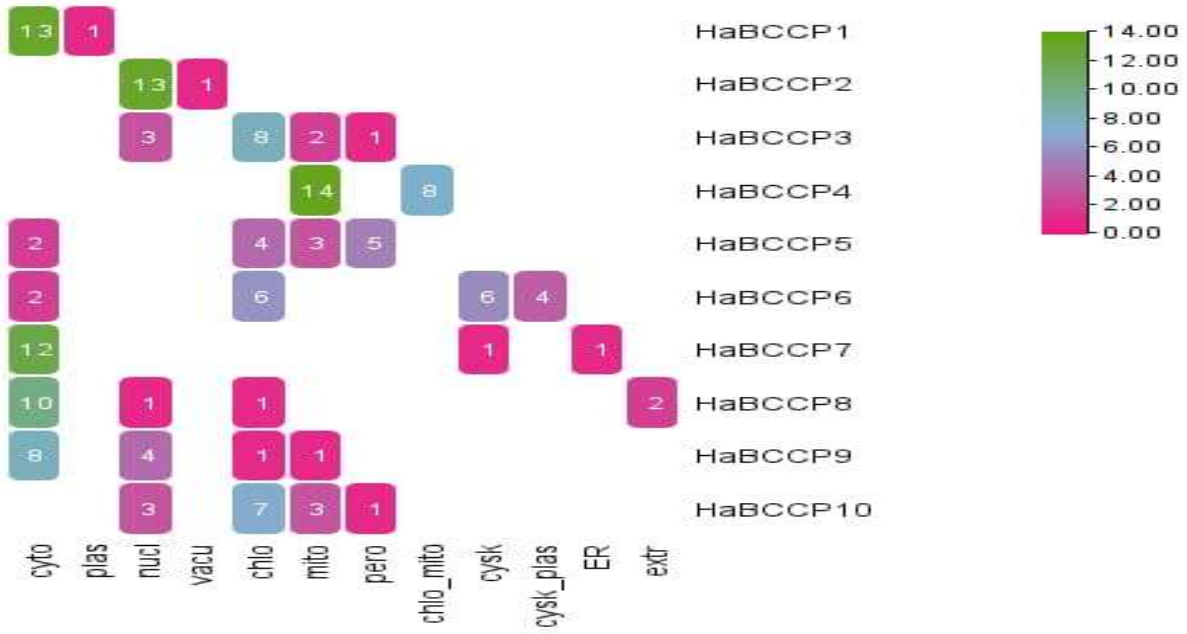

Figure 4. Subcellular localization

Cis Regulatory Analysis

The promoter regions of the ten HaBCCP genes had cis regulatory elements that are associated with light stress response, hormone signaling, developmental regulation, and transcription factor binding, suggesting multiple regulatory roles performed by these genes (Lescot et al., 2002). Almost all of the HaBCCP genes had light-responsive elements in abundance. Motifs associated with light responsiveness, e.g., G-box, Box 4, GT1-motif, TCT-motif, AE-box, ATCT-motif, and chs-CMA1a, were widely distributed throughout the promoters. G box was still present in relatively very large amounts, especially in HaBCCP8 and HaBCCP10, as shown in Figure 5. These observations are consistent with the role of BCCP proteins in the biosynthesis of fatty acids within plastids and chloroplasts. The abundance of G box in HaBCCP8 and HaBCCP10 implies that these genes are particularly important in responding to light signals and thus in the biosynthesis of lipids in the tissues that are active in performing photosynthesis (Terzaghi and Cashmore, 1995). Many promoter regions also had hormone-responsive elements, e.g., ABRE, CGTCA-motif, TGACG-motif, TCA-element, AuxRR-core, and GARE-motif. These motifs are responsive to ABA, MeJA, salicylic acid, auxin, and gibberellins. The ABRE elements were especially present in HaBCCP1, HaBCCP2,

HaBCCP3, and HaBCCP4, implying a role in ABA signalling, which is important in stress responses and seed development. It signifies the role in abiotic stress adaptation. So ABA signalling is activated especially during osmotic or drought stress (Nakashima and Yamaguchi-Shinozaki, 2013). CGTCA motif and TGACG motif play a role in methyl jasmonate responses during defence and development processes, including lipid metabolism. TCA elements respond to salicylic acid, and AuxRR motifs respond to auxin. These observations show that multiple hormones act to coordinate the regulation of fatty acid metabolism during growth, development, and defence.

Stress-responsive motifs, e.g., ARE, MBS, LTR, GC-motif, and TC-rich repeats, indicate that HaBCCP genes are active during environmental stresses, e.g., anaerobic conditions, drought, low temperature, etc. The ARE elements are associated with oxygen-limited conditions, which might be essential for sustaining lipid metabolism during hypoxia-induced metabolic stress. MBS motifs play a role in HaBCCP genes responding to water deficit stress. Drought often changes the composition of lipids present in membranes, so during this time, sustaining fatty acid biosynthesis is important. Similarly, LTF motifs play a role in cold stress responsiveness to adapt lipid metabolism during cold stress. TC-rich repeats are associated with defence responses, and GC motif elements are associated with anoxic stress regulation. The presence of these elements suggests that the BCCP family may contribute to maintaining membrane functions under stress conditions in sunflower. Many cis elements related to development were also found. These included CAT box, O2 site, circadian, RY element, and GCN4 motif as shown in Figure 5. CAT box is associated with meristems, suggesting that some HaBCCP genes are activated in growing tissues where lipid synthesis is needed for cell division. O2 site is associated with zinc metabolism, indicating a role in storage-related pathways, as in seeds. The circadian elements are linked to time-dependent processes, so might be involved in daily metabolic cycles of lipid biosynthesis. It is important for the synchronisation of lipid metabolism with photosynthesis to maximise metabolic efficiency. Overall, the cis regulatory analysis demonstrated that the BCCP gene family in *Helianthus annuus* is regulated by light signals, hormonal signals, developmental programs and stress responsiveness.

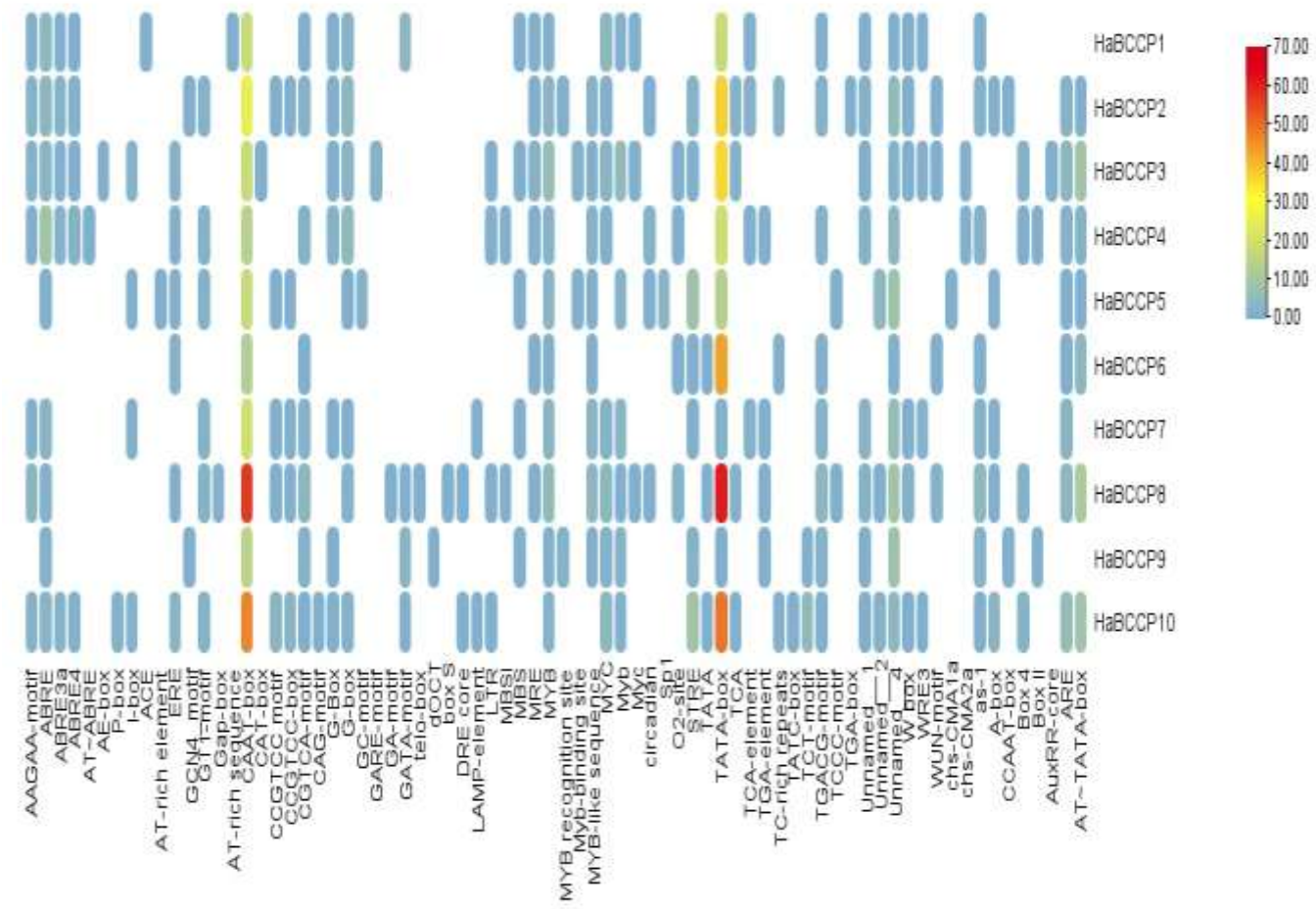

Figure 5. Cis Regulatory Analysis

Ka/Ks Analysis

All duplicated HaBCCP gene pairs exhibited Ka/Ks ratios below 1 that indicates that while duplication has increased gene copy number, natural selection has acted to conserve the functional properties of BCCP proteins (Yang and Bielawski, 2000). Gene pairs such as HaBCCP1–HaBCCP6 and HaBCCP2–HaBCCP5 displayed comparatively higher Ka/Ks values relative to other pairs, suggesting that these duplicates may have undergone greater sequence divergence after duplication, as shown in Figure 6. The estimated divergence times (MYA) further showed that the duplicated HaBCCP genes originated at different stages in sunflower evolutionary history. The relationship between Ka, Ks, and divergence time also suggests that sequence divergence has accumulated gradually under functional constraint. Overall, the Ka/Ks analysis demonstrated that the *Helianthus annuus* BCCP gene family has evolved under strong purifying selection. The duplicated gene pairs retained functionally constrained coding sequences despite varying evolutionary ages. The balance between conservation and divergence has likely enabled the BCCP gene family to maintain metabolic stability while adapting to the physiological and developmental requirements of sunflower.

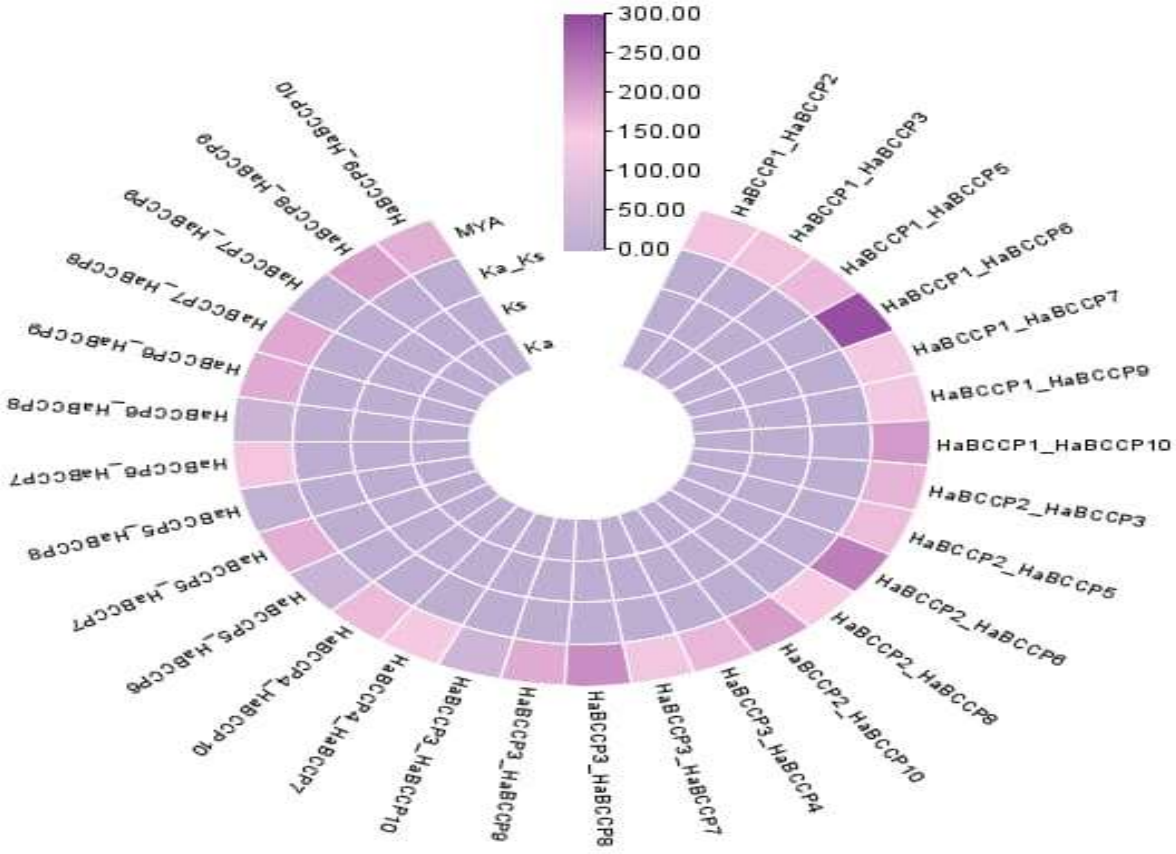

Figure 6. Ka/Ks Analysis

Synteny Analysis

The circular synteny map demonstrated interconnections among many HaBCCP genes specially HaBCCP1, HaBCCP3, HaBCCP5, HaBCCP6, and HaBCCP10. HaBCCP1 had the highest number of links to other HaBCCP loci, which suggests that it might be a central ancestral locus from which many paralogs originated, as shown in Figure 7 (Vision et al., 2000). In contrast, genes like HaBCCP8 and HaBCCP9 showed fewer syntenic links, indicating less duplicating connectivity, origin from more isolated duplication events, greater genomic rearrangement after duplication, or gene loss in surrounding regions (Panchy et al., 2016). Syntenic links present on different chromosomal segments suggest segmental duplication instead of tandem duplication. It also suggests that the HaBCCp family has undergone stepwise duplication and divergence, where some genes, which were duplicated in the ancestors, were retained and additionally gave rise to new paralogs. The combination of syntenic retention and low Ka/Ks ratios suggests that the duplicated HaBCCP genes have been preserved as functionally constrained paralogs and not only as redundant copies. In contrast, the varied number of syntenic links among different HaBCCP genes suggests that differential retention had occurred following duplication. This means that some duplicated genes have been preserved while others lost their syntenic associations because of deletion, divergence, or genomic rearrangements. Overall, the intra-genomic synteny analysis demonstrated that the *Helianthus*

annuus BCCP gene family has expanded primarily through segmental duplication events followed by selective retention of duplicated loci, as shown in Figure 7.

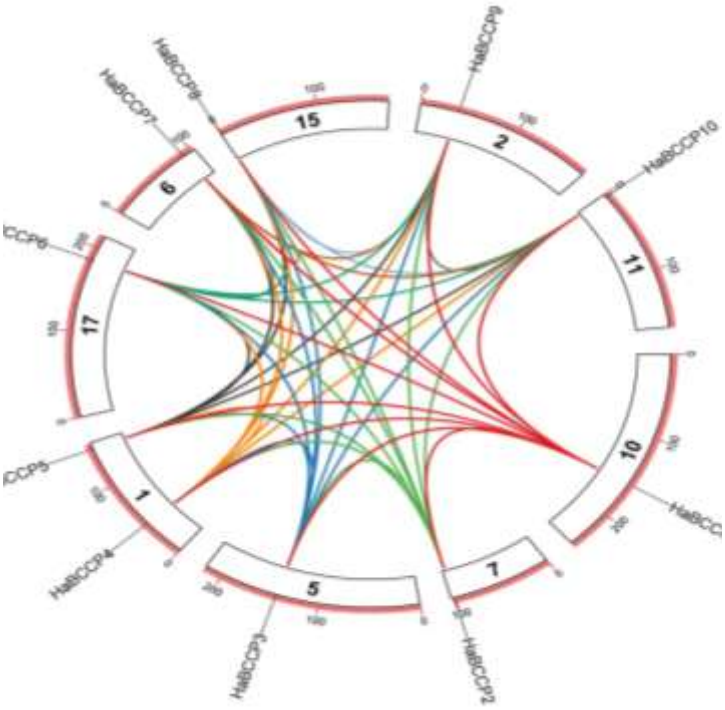

Figure 7. Synteny Analysis

Chromosomal Localization Analysis

HaBCCP genes were shown to be unevenly distributed across eight chromosomes that are chromosome 2, 4, 5, 6, 7, 15, and 17, while no BCCP genes were detected on the remaining chromosomes, as shown in Figure 8. It indicates segmental and interchromosomal duplication events rather than tandem duplication. This is also evident because tandem duplication creates closely placed paralogous genes on the same chromosome, but in this case, each chromosome had one or two BCCP loci placed at distant locations in the genome (Freeling, 2009). Chromosomes 5 and 15 each carried two HaBCCP loci, while the rest had only one. HaBCCP2 and HaBCCP3 were located on chromosome 5, whereas HaBCCP7 and HaBCCP8 were located on chromosome 15. The presence of two genes shows local duplication or duplicated segments retained after duplication in the ancestral genome. However, since these genes were not present adjacently, they show segmental duplication within the same chromosome. Chromosome 10 contained HaBCCP1, chromosome 4 contained HaBCCP4, chromosome 17 carried HaBCCP6, chromosome 6 contained HaBCCP7, chromosome 2 contained HaBCCP9, chromosome 1 had HaBCCP5, and chromosome 11 carried HaBCCP10 (Lynch and Conery, 2000). There were also variations in the placement of the genes along the chromosome arms. For example, HaBCCP8 and HaBCCP10 were placed near the upper ends of chromosomes 15 and 11, respectively, whereas HaBCCP1 and HaBCP6 were present near the lower ends. Such dispersed positioning may contribute to differential regulation of the genes in distinct regulatory landscapes. HaBCCP1 showed multiple syntenic associations with other chromosomal loci, indicating that it might be an ancestral duplication node within the sunflower BCCP family. Similarly, HaBCCP4, HaBCCP5, HaBCCP8, HaBCCP9, and HaBCCP10 also exhibited multiple interchromosomal syntenic links. Retention of duplicated genes often means that these genes are involved in important pathways of metabolic activities, as maintaining multiple copies can enhance robustness. Genes on chromosomes 10, 15, and 11 have relatively higher association, suggesting common ancestral duplication events, as shown in Figure 8. The repeated syntenic links among HaBCCP1, HaBCCP8, and HaBCCP10 indicate a relatively recent common origin. Overall, the chromosomal localization analysis demonstrated that the Helianthus anus BCCP genes are widely dispersed across the genome of sunflower, and their distribution pattern supports the segmental duplication event (Panchy et al., 2016).

Protein-protein interaction

The protein-protein interaction (PPI) analysis of the *Helianthus annuus* BCCP proteins revealed a highly interconnected network of proteins, indicating strong coordination for functions. Nearly every node seems connected to many other nodes, suggesting an integrated molecular system rather than independent functions. The coordination is most likely for processes like acetyl-CoA carboxylation, fatty acid biosynthesis, and metabolic regulation. It also leads to the assumption that HaBCCP proteins are under strong selective pressure to maintain this interaction capability, which is consistent with the Ka/Ks analysis pattern (Nikolau et al., 2003). The interaction density was the highest among the six proteins, namely HaBCCP3, HaBCCP4, HaBCCP5, HaBCCP6, HaBCCP8, and HaBCCP10, as shown in Figure 9. The isolated nodes were absent for these proteins, suggesting integrated functionality and contribution to the broader activity of the BCCP system. The central placement of HaBCCP3 and HaBCCP10 suggests that these proteins are primary functional isoforms of the BCCP family in sunflower (Szklarczyk et al., 2023). HaBCCP3 showed particularly high interactions with HaBCCP4, HaBCCP5, and HaBCCP6, suggesting that these proteins form a functional subnetwork within the BCCP network shown in Figure 9. Similarly, HaBCCP10 showed strong interactions with HaBCCP6, HaBCCP8, and HaBCCP4. The presence of conserved motifs in HaBCCP4, HaBCCP5, and HaBCCP6, along with the strong interaction of these proteins with HaBCCP3, suggests coordination in fatty acid biosynthesis. HaBCCP3 and HaBCCP10

particularly demonstrated strong interaction, as these two proteins were directly connected within the network, suggesting that they are a part of the same enzymatic complex or closely related pathways.

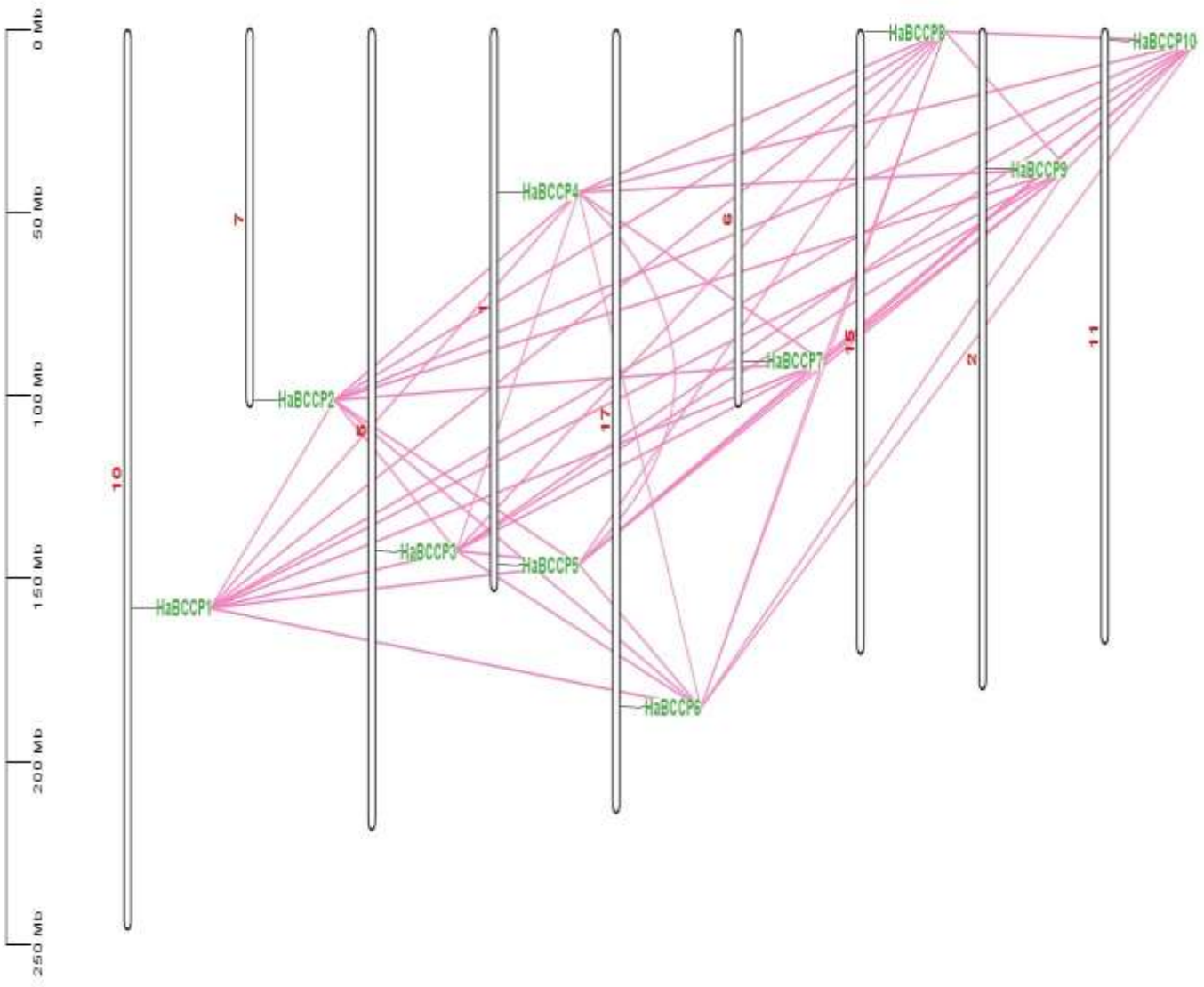

Figure 8. Chromosomal Localization Analysis

HaBCCP4, HaBCCP5, and HaBCCP6 occupied intermediate positions in the network, suggesting to be serving as supporting functional elements such as complex stabilisation or specialised roles under different physiological conditions, as shown in Figure 9. HaBCCP8 occupied a relatively peripheral position, indicating divergent functionality. The absence of HaBCCP1, HaBCCP2, HaBCCP7, and HaBCCP9 suggests either weak interaction capacities or function outside the core BCCP network. They might also have diverged significantly to escape prediction within this model of interaction. Overall, PPI analysis revealed that sunflower BCCP proteins form a network that is interconnected and coordinated as well. This network is involved in fatty acid biosynthesis and other metabolic regulatory processes. HaBCCP3 and HaBCCP10 appear to be major hubs of interaction, while other members are either supporting proteins or specialised ones. The network suggests that the BCCP family maintains a balance between functional conservation, redundancy, and specialization, ensuring efficient metabolic activity and supporting sunflower growth and oil production (Li-Beisson et al., 2013; Sasaki and Nagano, 2004)

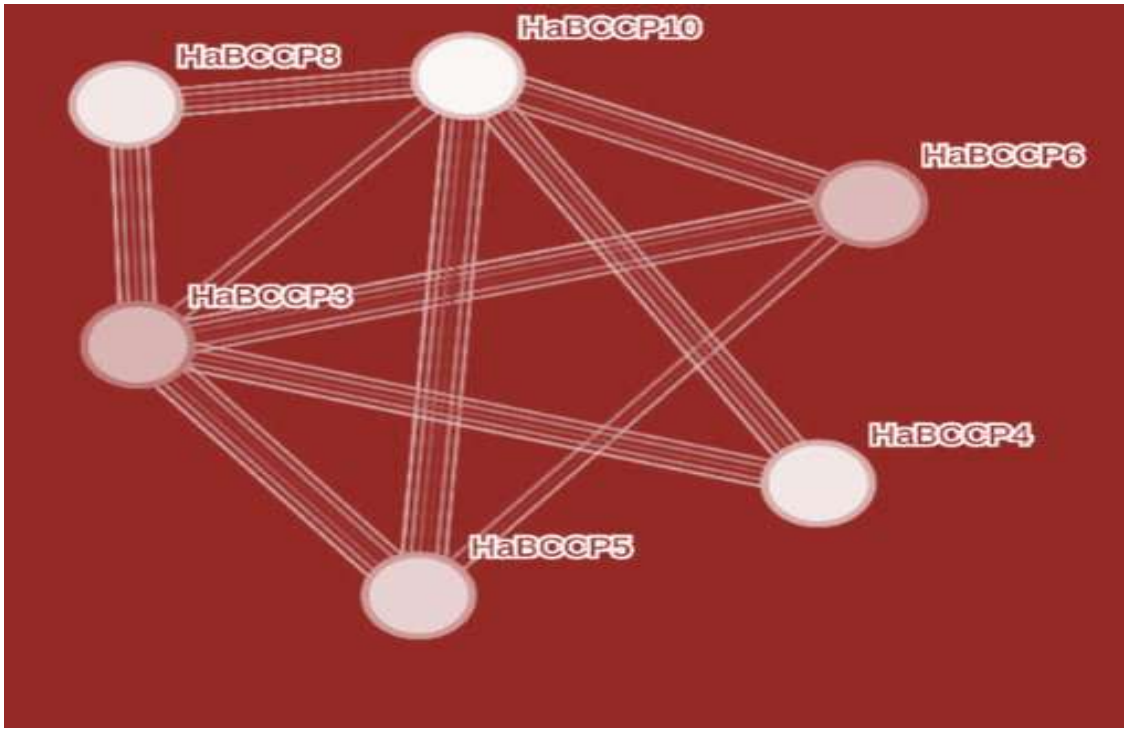

Figure 9. Protein-protein interaction

Conclusion

Here, we report the first genome-wide identification and analysis of the BCCP gene family in *Helianthus annuus* and show that the sunflower BCCP genes have substantially evolved while retaining the conserved structure and function necessary for their role in fatty acid synthesis. Ten HaBCCP genes were found to have an uneven distribution on the sunflower chromosomes with multiple duplication events, suggesting that segmental duplication is the main evolutionary force driving the diversification of the BCCP gene family in sunflower. In addition, phylogenetic analysis revealed that the HaBCCP proteins are grouped into three main clades together with their counterparts from several plant species and that the evolution of the BCCP family began before the monocot-dicot split, with the sunflower genes showing clear evolutionary connection to their orthologs in other crops. Analysis of conserved domains and motifs also showed that some HaBCCP proteins maintain the typical catalytic structure of proteins involved in acetyl-CoA carboxylation, while others have diverged, indicative of functional diversification. These features were supported by the analysis of exon-intron structures, which revealed significant diversity that reflects ongoing diversification after duplication.

Localization prediction suggests that several HaBCCP genes encode proteins with plastid and chloroplast localization, as expected for proteins involved in fatty acid biosynthesis, but also that other members of this family are targeted to other subcellular compartments, highlighting their potential expansion of function beyond the conventional plastid fatty acid synthesis pathway. The cis-element analysis of promoter sequences identified a broad spectrum of cis-regulatory motifs related to light regulation, hormone-mediated regulation, developmental regulation and adaptation to abiotic stress, suggesting that the HaBCCP genes are involved in various regulatory pathways and may act in response to developmental and environmental cues. The analyses of Ka/Ks and synteny demonstrated that the duplicated HaBCCP genes have been under strong purifying selection, suggesting that retention of core functions was important despite expansion via duplication events. However, the retention of the duplicated loci also suggests that paralogs may have evolved to play specific roles in regulatory or metabolic processes. Protein-protein interaction analysis also revealed the presence of hub proteins in HaBCCP3 and HaBCCP10, suggesting that these genes are likely to play major roles in coordinating the activities of the BCCP network in sunflower lipid metabolism. Overall, the findings demonstrate that the evolution of the BCCP gene family in sunflower has involved a delicate balance between gene duplication, domain conservation and regulatory diversification, allowing for the maintenance of core enzymatic activities while allowing for specific regulatory functions. This study provides important insights into the evolution of BCCP genes in sunflower and a solid basis for future functional research to map the specific contributions of HaBCCP genes in sunflower fatty acid synthesis, oil formation, and stress response. In the end, the HaBCCP genes identified here are attractive targets for enhancing sunflower oil yield and metabolism.

References

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., & Mayjonade, B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, **546**(7656), 148–152. https://doi.org/10.1038/nature22380

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37**(suppl_2), W202–W208. https://doi.org/10.1093/nar/gkp335

Cronan, J. E., & Waldrop, G. L. (2002). Multi-subunit acetyl-CoA carboxylases. *Progress in Lipid Research*, **41**(5), 407–435. https://doi.org/10.1016/S0163-7827(02)00007-3

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, **183**(3), 557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*, **60**(1), 433–453. https://doi.org/10.1146/annurev.arplant.043008.092122

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, **30**(1), 325–327. https://doi.org/10.1093/nar/30.1.325

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., Baud, S., Bird, D., DeBono, A., & Durrett, T. P. (2013). Acyl-lipid metabolism. *The Arabidopsis Book*, **11**, e0161. https://doi.org/10.1199/tab.0161

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., & Wu, J. (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology*, **33**(5), 524–530. https://doi.org/10.1038/nbt.3208

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, **290**(5494), 1151–1155. https://doi.org/10.1126/science.290.5494.1151

Montes, J. M., Melchinger, A. E., & Reif, J. C. (2007). Novel high-throughput phenotyping platforms in plant genetic studies. *Trends in Plant Science*, **12**(10), 433–436. https://doi.org/10.1016/j.tplants.2007.08.016

Moore, M. J., & Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, **136**(4), 688–700. https://doi.org/10.1016/j.cell.2009.02.001

Nakashima, K., & Yamaguchi-Shinozaki, K. (2013). ABA signaling in stress-response and seed development. *Plant Cell Reports*, **32**(7), 959–970. https://doi.org/10.1007/s00299-013-1418-1

Nikolau, B. J., Ohlrogge, J. B., & Wurtele, E. S. (2003). Plant biotin-containing carboxylases. *Archives of Biochemistry and Biophysics*, 414(2), 211–222. https://doi.org/10.1016/S0003-9861(03)00156-5

Ohlrogge, J., & Browse, J. (1995). Lipid biosynthesis. *The Plant Cell*, **7**(7), 957–970. https://doi.org/10.1105/tpc.7.7.957

Ohlrogge, J. B., & Jaworski, J. G. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **48**(1), 109–136. https://doi.org/10.1146/annurev.arplant.48.1.109

Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, **171**(4), 2294–2316. https://doi.org/10.1104/pp.16.00523

Rose, A. B. (2019). Introns as gene regulators: a brick on the accelerator. *Frontiers in Genetics*, **9**, 672. https://doi.org/10.3389/fgene.2018.00672

Salie, M. J., & Thelen, J. J. (2016). Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, **1861**(9), 1207–1213. https://doi.org/10.1016/j.bbalip.2016.03.003

Sasaki, Y., & Nagano, Y. (2004). Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience, Biotechnology, and Biochemistry*, **68**(6), 1175–1184. https://doi.org/10.1271/bbb.68.1175

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., & Cheng, J. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, **463**(7278), 178–183. https://doi.org/10.1038/nature08670

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., & Pyysalo, S. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, **51**(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Terzaghi, W. B., & Cashmore, A. R. (1995). Light-regulated transcription. *Annual Review of Plant Physiology and Plant Molecular Biology*, **46**(1), 445–474.

https://doi.org/10.1146/annurev.pp.46.060195.002305

Tong, L. (2005). Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cellular and Molecular Life Sciences*, **62**(16), 1784–1803. https://doi.org/10.1007/s00018-005-5096-7

Vision, T. J., Brown, D. G., & Tanksley, S. D. (2000). The origins of genomic duplications in Arabidopsis. *Science*, 290(5499), 2114–2117. https://doi.org/10.1126/science.290.5499.2114

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**(12), 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7

Statements and Declarations

Data Availability statement

All relevant data are within the manuscript file.

Author's Contribution Statement

AS, and HA collected data and wrote manuscript equally. JA, and SHUHS make final editing. SHUHS supervised the whole research work. All authors have read the final manuscript and approve its submission.

Acknowledgments

Not applicable

Funding

Not applicable

Ethical Statement

Not applicable

Conflict of interest

The investigation was undertaken without any financial conflicts of interest or any other commercial relationships that could be seen as such by any of the authors.