Original Article

Comparative computational genomic analysis between wild and domesticated species of cacao and coffee

Ronald Villamar-Torres[1], Byron Oviedo Bayas[2], Yenny Torres-Navarreete[3], Cristian Zambrano-Vega[4], Seyed Mehdi Jazayeri*[5]
[1,3]Facultad de Ciencias Agropecuarias. Universidad Técnica Estatal de Quevedo, Ecuador
[2,4]Facultad de Ciencias de la Ingeniería. Universidad Técnica Estatal de Quevedo, Ecuador
[5]Facultad de Biología, Universidad Nacional de Colombia
Email: smjazayeri@unal.edu.co

Abstract

Cacao and coffee are two tropical crops with an essential annual turnover. Alkaloid products like caffeine and theobromine show that these two species share related pathways in alkaloid metabolism. Genomic proteins of three coffee species, including Coffea arabica, C. canephora and C. eugenioides, and two cacao species, Theobroma cacao, and Herrania umbratica, were compared and analyzed based on protein clustering and GO terminology. A comparison of all five species showed orthologous proteins involved in stress, disease response, alkaloid biosynthesis, regulation and transport. Based on GO terms of orthologous proteins of 5 species, two proteins, including Tryptophan decarboxylase TDC2 and Perakine reductase, were disclosed that are involved in the biosynthesis of monoterpenoid indole alkaloids. In addition, enzymes including Caffeic acid 3-O-methyltransferase, Caffeoylshikimate esterase, Caffeoyl-CoA 3-O-methyltransferase and Fatty alcohol:caffeoyl-CoA acyltransferase were found as orthologous proteins of all five studied species. Proteins of H. umbratica and C. eugenioides were asymmetrically shared with other species as their proteins were not aligned with species of coffee for H. umbratica and species of cacao for C. eugenioides. Single-copy proteins common between 5 studied species showed functions related to growth and development, stress response, phytohormones, gene expression regulation and metabolite biosynthesis. Two proteins, including Vinorine synthase and Deacetoxyvindoline 4-hydroxylase (D4H), involved in indole alkaloid biosynthesis were found as single-copy proteins among five species. These results showed biological processes and molecular functions of proteins among cacao and coffee species. Wild species of cacao and coffee are valuable genetic resources for selection and breeding.

Keywords: Coffea arabica, Coffea canephora, Coffea eugenioides, Herrania umbratica, Theobroma cacao.

1 INTRODUCTION

Cacao and coffee are two tropical crops with an annual turnover of more than 200 billion dollars in the agriculture, food and pharmaceutical industries. Production of cacao and coffee plays a key role in the economy of many countries in South and Central America, Central Africa and South Asia (Castro Pinacy and Barrezueta Unda, 2020). Production of cacao and coffee is more than 4 and 7 million tons per year, respectively. Cacao and coffee are used as beverages, medicine, pharmaceuticals, cosmetics and alimentation ingredients.

Shared phytochemicals of cacao and coffee are phenolic acid derivatives and purine alkaloids, i.e., known mainly by caffeine and theobromine. These metabolites are used in human health and play essential roles as medicaments and treatments for diseases like neuropathology, neurodegenerative disorders, cardiovascular, physical and mental performance, mitigating stresses, antioxidant and anti-inflammatory (Baroni et al., 2021; Hamer et al., 2006; Jazayeri et al., 2021).

Address for correspondence: Seyed Mehdi Jazayeri
Facultad de Biología, Universidad Nacional de Colombia
Email: smjazayeri@unal.edu.co

Access this article online

Quick Response Code:	
	Website: www.pnrjournal.com
	DOI: 10.47750/pnr.2022.13.03.082

How to cite this article: Ronald Villamar-Torres, Byron Oviedo Bayas, Yenny Torres-Navarreete, Cristian Zambrano-Vega, Seyed Mehdi Jazayeri, Comparative computational genomic analysis between wild and domesticated species of cacao and coffee, J PHARM NEGATIVE RESULTS 2022;13:535-551.

Caffeine is a methylxanthine that has been the most widely studied and has apparent effects on neuronal network activity, promotes sustained cognitive performance and can protect neurons against dysfunction and death in animal models of stroke, Alzheimer's disease and Parkinson's disease (Camandola et al., 2019).

It is essential to select and breed cacao and coffee with more yield and tolerance to climate change issues like drought, high temperature, rainfall, and precipitation fluctuation. Crop wild relatives (CWRs) are a strong genetic diversity resource for plant breeding, selection and crop improvement and have been utilized for major gene-disease and pest resistance and abiotic stress tolerance wild and close species to crops (Jazayeri et al., 2020; von Wettberg et al., 2020). For example, Elaeis oleifera (American oil palm) is a wild close relative of E. guineensis (African oil palm) and has been used as a genetic resource for hybridization and crossing with E. guineensis. Hybrids of both species are more tolerant to drought stress (Jazayeri et al., 2015) and rot bud disease (Avila-Diazgranados et al., 2016). Some oil palm hybrids have been used for different oil fatty acid compositions and QTL studies (Montoya et al., 2014). Coffee and cacao relatives are important for hybridization, breeding and improvement, as well as for biodiversity and physiological and biochemical analyses (Aikpokpodion, 2012; Sousa Silva and Figueira, 2004).

C. arabica is an allotetraploid ($2n = 4x = 44$) hybrid species whose ancestors are C. canephora and C. eugenioides that are diploid species ($2n=2x=22$) (Scalabrin et al., 2020; Tran et al., 2018). Two Coffea species, i.e., C. arabica and C. canephora, dominate 99% of the coffee industry (Damatta et al., 2018) and are known as coffee Arabica and coffee Robusta, respectively. Genomes of three Coffea species have been a source for plant breeding for high-quality coffee (Cenci et al., 2012; Merot-L'anthoene et al., 2019; Tran et al., 2018). Despite its importance in the evolutionary history of the main economic species of coffee, few genomic and molecular studies have been reported on C. eugenioides specifically and the publications have been studies on hybrids of other Coffea species with C. eugenioides (Cenci et al., 2012; Gartner et al., 2014; Lashermes et al., 1999; Merot-L'anthoene et al., 2019; Yuyama et al., 2016).

The cacao crop belongs to T. cacao species and H. umbratica is Colombian coffee wild relative to T. cacao. There are several species of Theobroma, but they are not yet used as an alternative for T. cacao. The most closely related species of T. cacao with agronomic traits are Theobroma bicolor Humb & Bonpl. and Theobroma grandiflorum (Willd. ex Spreng.) K. Schum (González-Orozco et al., 2020). H. umbratica has been used to create intergeneric plants of cacao varieties (Doodnath, 2008). They are phylogenetically related (Whitlock and Baum, 1999). Despite its importance and potential genomic data analyses about H. umbratica is not available, only a few articles have been published about some species' genes (Salamah and Rostina, 2019; Whitlock and Baum, 1999).

Comparative genomic studies disclose similarities, differences and variations between different species. Sequence comparison of orthologous regions enables estimation of the divergence between genomes, analysis of their evolution and detection of particular features of the genomes, such as sequence rearrangements and transposable elements (Cenci et al., 2012).

Also, synergism between Zn and Cd in plants and soil suggests that Zn directly affects Cd accumulation in cacao (Arévalo-Gardini et al., 2017).

Comparative genomic studies generate valuable data for plant evolutionary and phylogenetic analyses and reveal biodiversity information used in plant selection and breeding. Plants that share genomic information are closer phylogenetically and genomic data similarities predict their characters like producing similar/same metabolites, stress tolerance level (e.g., stenothermal or eurythermal), vegetation type, ecological distribution, etc. (Mawalagedera et al., 2019; Polturak et al., 2022). Homology based on common genes from different species has provided huge information by which scientists can explain biological mechanisms and processes. Orthologs between different species determine linear descent inferred from phylogenies (Delaux et al., 2019). However, gene copy numbers in different species are also determinative in phylogeny studies and can differentiate plants and explain their traits. Gene copy number determines gene expression and transcription level, consequently resulting in the gene-related trait (Weiße et al., 2015). On the other hand, single-copy genes are important as they are critical and responsible for certain biological mechanisms, processes, and diseases that only they manage. A single-copy gene mutation can affect living organisms' life like sing-copy gene disorders in humans (Jackson et al., 2018). Single-copy gene families can be used as molecular markers for phylogenomic analysis in seed plants (Li et al., 2017a).

This study aimed to compare genomic protein datasets of 2 cacao and three coffee species, as they share metabolites like caffeine, theobromine and other alkaloids and have familiar biological characteristics like tropical distribution, tree vegetation type, etc. Therefore, five species were compared together to find common proteins among species that can be of help and importance for further breeding, selection and evolutionary studies. In addition, two cacao species, including T. cacao and its CWR H. umbratica and three coffee species, including C. arabica, C. canephora and C. eugenioides, were compared separately, as it is critical to study close species comparatively to detail genome and genes.

2 Methodology

2.1 Genomic data of cacao and coffee species

Genomic proteins of NCBI Release 100 annotation for two cacao species and three coffee species were downloaded. The RefSeq accession numbers were GCF_003713225.1 for C.

arabica with 67222 protein sequences, GCF_003713205.1 for Coffea eugenioides with 38150 proteins GCF_000208745.1 for T. cacao with 30854 proteins, and GCF_002168275.1 for H. umbratica with 27748 proteins. In addition, the 25574 protein sequences of C. canephora were downloaded under GenBank accession number GCA_900059795.1.

2.2 Comparisons between different species

Five species were compared in three different analyses, including five species, two cacao, and three coffee species. The first comparison of all five species was analyzed. Then based on the OrthoVenn2 results, two cacao species, including T. cacao and H. umbratica and three coffee species, including C. arabica, C. canephora and C. eugenioides, were analyzed separately. These different comparisons were made to understand better how proteins are clustered among 5 species and when they are compared separately, considering orthology. OrthoVenn2 was used to compare genomic protein datasets. OrthoVenn2 is a web-based program that can compare protein datasets by clustering the input proteins and generating Venn diagram, SwissProt (UniProt) data and GO enrichment for the compared datasets (Xu et al., 2019). The default parameters were employed except for E-value that was fixed at 1E-05. The SwissProt hits and protein/gene names were retrieved from UniProt (https://www.uniprot.org) database (Bateman et al., 2021) again by mapping the UniProt identifier outputted by OrthoVenn2 for each cluster to Retrieve/ID mapping tool.

2.3 Protein networks for orthologs of five species

In order to analyze single-copy proteins between 5 species, a protein network approach was performed. UniProt protein IDs of single-copy protein cluster list between 5 species were submitted to WebGestalt (WEB-based GEne SeT AnaLysis Toolkit http://www.webgestalt.org/). WebGestalt uses different analysis methods and statistical approaches to analyze a list of genes or proteins for model organisms and Arabidopsis thaliana as a plant model. The basic parameters were Organism of Interest: Arabidopsis thaliana, Method of Interest: Network Topology-based Analysis (NTA), and Functional Database: network PPI BIOGRID. Advanced parameters were as default.

2.4 GO classification for proteins of 5 species

GOSlimViewer summarized the GO terms of OrthoVenn2 results to facilitate interpretation of GO and protein functions of 5 species. GOSlimViewer is a web tool of AgBase (https://agbase.arizona.edu/cgi-bin/tools/goslimviewer_select.pl) that lists the input GO terms in a high-level summary (McCarthy et al., 2006, 2007). It receives a list of GO terms and lists them based on the main GO terms of inputs, then generates a table of GO terms and their number counted in the input list. GOSlim set parameter of GOSlimViewer was Plant.

3 Results and discussion

3.1 Comparison of protein clusters of 5 species

Five species of cacao and coffee were compared by clustering protein sequences based on OrthoVenn2 tool. In total, 189548 protein sequences of 5 species were analyzed (Table 1). For all proteins sequences, 30792 protein clusters were formed, out of which 28520 were orthologous clusters, i.e., at least containing protein of two species and 2272 single-copy gene clusters, i.e., the clusters with five proteins belonging to 5 species with one protein of each species. Single-copy proteins ranged from 3.38% for C. arabica to 8.88% for C. canephora while the other species were between these two species. In total, 25005 of all proteins in the study were not clustered and were assigned as singletons (Table 1). As the singleton number cannot show an exact estimation for comparing genes, each species' singleton percentage of proteins was calculated. Singleton percentage of T. cacao and H. umbratica were less compared to coffee species and two cacao species shared more proteins with coffee species.

On the other hand, C. canephora had the most singletons compared to coffee species suggesting that C. arabica has evolved differently from its ancestor C. canephora. If specie-specific clustered proteins are added to singletons, the results are similar except for C. arabica whose species-specific proteins reached 24.26% close to C. canephora. When T. cacao and H. umbratica were pairwise compared with each other species-specific proteins were 2978 and 5744 for H. umbratica and T. cacao, respectively. A comparison of three Coffea species showed 17553, 6445 and 6343 species-specific proteins for C. arabica, C. canephora and C. eugenioides, respectively. Therefore, these results showed that even though not close to family and species, cacao and coffee species shared more proteins than cacao or coffee species. These findings suggest that cacao and coffee are genetic resources for further studies compared to a whole dataset. However, this remains to be studied in more detail by functional genomic studies for specific common genes.

Table 1. Protein number, clusters and singleton data of each protein dataset of 5 species. C. arabica possesses the most proteins, while C. canephora has the least number of proteins. Singleton percentage was calculated by the number of singletons divided by the whole protein number of each specie. The Singleton percentage of H. umbratica was the least and the most for C. canephora. Species-specific clustered proteins SSCP, species-specific proteins SSP.

Species	*Proteins*	*Clusters*	*Singletons*	*Singleton%*	*SSCP*	*SSP*	*SSP%*
C_arabica	67222	24480	10436	15.52	5871	16307	24,26
C_eugenioides	38150	22466	4266	11.18	1775	6041	15,83

T_cacao	30854	18860	2598	8.42	1521	4119	13,35
H_umbratica	27748	17728	2061	7.43	552	2613	9,42
C_canephora	25574	18258	5644	22.07	650	6294	24,61

Protein clusters among all five species, 4,3 and 2-species comparisons and species-specific clusters are shown in Figure 1. Total common clustered proteins between all five species were 96249 (50.78% of 189548 all proteins) grouped in 12020 protein clusters distributed between 5 species as 30023 (31.19% of 96249 common proteins of 5 species) belonged to C. arabica, 17881(18.58%) to T. cacao, 17861 (18.56%) to C. eugenioides, 17762 (18.45%) to H. umbratica and 12722 (13.22%) to C. canephora. In addition, the percentage of shared proteins with other 4 species to the whole protein dataset of each species was calculated and orthologous proteins were 64.01% of H. umbratica proteins, 57,95% of T. cacao, 49,75% of C. canephora, 46,82% of C. eugenioides and 44.66% of C. arabica. These results suggest that H. umbratica is a non-crop candidate species for further studies and can be a model for cacao and coffee.

The remained protein clusters were shared between 4, 3 or 2 species, while species-specific clusters were 2074 for C. arabica, followed by C. eugenioides with 664, T. cacao with 526, C. canephora with 252 and finally, H. umbratica with 218 clusters (Figure 1). A cacao genome report has shown that T. cacao species-specific genes were 2053 when compared to Vitis vinifera, A. thaliana, Populus trichocarpa and Glycine max (Argout et al., 2011). Comparing these results with these results, more similarity between studied species was seen by Argout et al. 2011 than our results of comparing T. cacao with coffee species. This is due to phylogenetic relativeness between cacao and the studied species in Argout et al. (2011) publication; the studied species belonged to the close families to cacao, while in this current study, cacao species was compared with the coffee family that belongs to the Rubiaceae family and Gentiales order that is farther to T. cacao and H. umbratica in the phylogenetic tree.

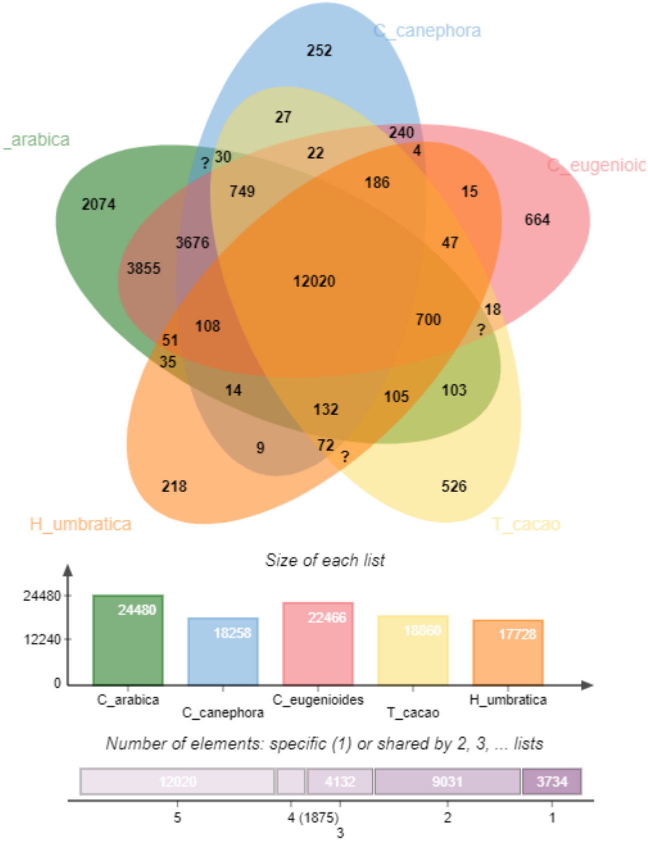


Figure 1. Venn diagram of comparison between 5 species. 12020 clusters were common among all 5 species. "?" signs are the cluster numbers that are not displayed because of space limits. They are 4102 for T. cacao and H. umbratica, 240 clusters for C. canephora and C. eugenioides and 111 for C. arabica, C. eugenioides and T. cacao.

Pairwise comparisons (Figure 2) confirmed the highest similarities between T. cacao and H. umbratica with common 22039 protein clusters and C. arabica and C. eugenioides with 21882 clusters. However, a comparison of C. canephora and C. arabica (18744 clusters) and C. canephora and C. eugenioides (18267 clusters) showed fewer proteins similarities between C. arabica and C. eugenioides. Pairwise comparisons between cacao and coffee species gave lower common clusters between the two species. However, T. cacao and C. arabica shared 18130 clusters more than common clusters of T. caca with the other two coffee species 18054 and 18045 C. canephora and C. eugenioides, respectively. On the other hand, H. umbratica with C. arabica had 18062 clusters in common, while 18036 and 18042 with C. canephora and C. eugenioides, respectively. These pairwise comparisons showed a relative contribution between cacao species and coffee species, indicating how distant cacao and coffee species are.

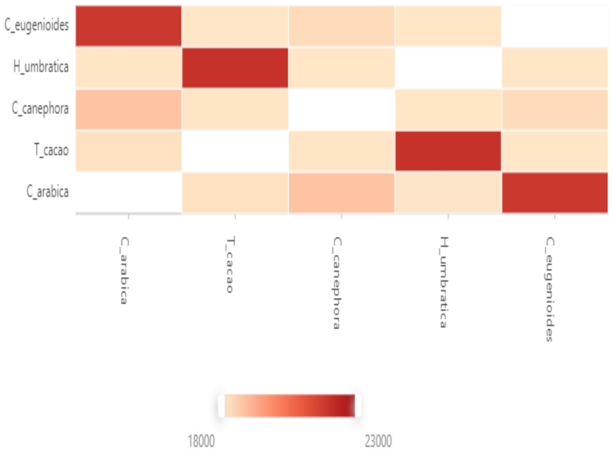


Figure 2. HeatMap of pairwise comparison between 5 species. H. umbratica and T. cacao showed the 22039 protein clusters. Between proteins of C.arabica and C. eugenioides 21882 clusters were commons. 18744 clusters were shared between C. canephora and C. arabica. 18267 clusters C. canephora and C. eugenioides.

In common protein overlap between only two-species (shared proteins between two species specifically), C. arabica shared 3855 clusters of 10906 proteins with C. eugenioides (5802 proteins of C. arabica and 5104 proteins of C. eugenioides) and 717 clusters of 2203 proteins with C. canephora (1437 proteins of C. arabica and 766 proteins of C. canephora). C. eugenioides and C. canephora shared 240 clusters of 612 proteins (256 proteins of C. canephora and 356 proteins of C. eugenioides) (Figure 3). C. arabica is a hybrid of C. canephora and C. eugenioides (Yuyama et al., 2016) and these results indicate that C. canephora and C. eugenioides have not contributed in the same way to C. arabica. Probably during the speciation of C. arabica, evolutionary events such as duplication and deletion happened that changed the gene proportion of two ancestors in C. arabica. However, C. eugenioides had more shared proteins with C. arabica suggesting it played an important role in the coffee selection and improvement programs and needs further studies to reveal its influence on coffee varieties as less genomic data is available for it than C. arabica and C. canephora. Different factors of GxE effects (environment and genetics) are responsible for these variations among three species of Coffea. GxE effects might come from deforestation, fragmentation and natural selection on Coffea populations from one side, gene drifting/drafting, and other genetic effects from another. The strength of gene drift during speciation processes affects mutations and their interactions among populations and determines parental and hybrid fitness (Dagilis et al., 2019). One study showed that C. canephora populations with low allelic diversity from Itwara and Kibale of Uganda were genetically distinct, likely implying inbreeding combined with random genetic drift. The authors suggested that because of a low proportion of constrained variance of the environmental variables, gene flow and genetic drift or bottlenecks that probably result from deforestation and fragmentation affect the genetic diversity and structure of Uganda's wild C. canephora populations (Kiwuka et al., 2021). Both forest fragmentation and forest degradation can cause increased genetic drift, reduced gene flow, and alteration of mating patterns that result in increased inbreeding and, finally, a negative impact on the genetic diversity of forest plant species (Aerts et al., 2013). Accordingly, three coffee species shared asymmetrical proteins and C. canephora role in C. arabica breeding and improvement needs further studies to understand how C. eugenioides might have been a more dominant parent of Arabica hybrids.

Pairwise comparisons that showed higher similarities between species were the comparison of C. arabica and C. eugenioides and H. umbratica and T. cacao with 10906 and 11005 proteins, respectively (Figure 3). Three species of Coffea showed 18525 common proteins clustered in 3676. A comparison of three Coffea and T. cacao with 4753 common proteins of 749 clusters and comparison including C. arabica, C. eugenioides, T. cacao and H. umbratica with 700 clusters of 4693 proteins were the highest in 4-species comparisons. On the other hand, only 9 clusters of proteins were found between H. umbratica and C. canephora. These findings suggest different distribution of orthologs of H. umbratica with Coffea species and C. canephora with other coffee species as they share fewer proteins among other species. C. eugenioides, therefore, is the species that share more proteins with two other Coffea species and H. umbratica and T. cacao suggesting it as a genetic source for further studies.

Between 3 species of Coffea 3676 clusters were formed covering 18525 proteins, while 8783 (47.41%) belonged to C. arabica, 4098 (22.12%) to C. canephora and 5644 (30.47%) to C. eugenioides. On the other hand, protein clusters shared between two cacao species were 4012, including 11005 proteins, of which 5588 (50.78%) and 5417 (49.22%) to T. cacao and H. umbratica, respectively.

These comparisons based on 4, 3 and 2 species suggest that C. canephora proteins contributed in a different pattern to coffee speciation and/or breeding programs of coffee need deeper crossing and selection studies to understand why this behavior has happened while C. arabica is s sibling of both C. canephora and C. eugenioides and the distance between cacao and coffee species are similar, but C. canephora showed different genetic contributions.

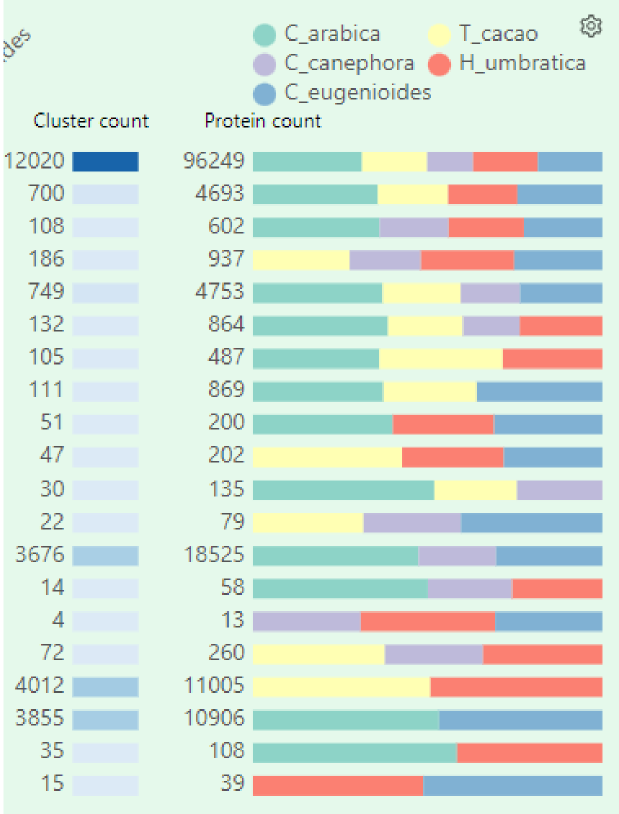


Figure 3. Cluster and protein count and their distributions for 5 species. The most counted clusters were between 5 species followed protein count for comparison of three Coffea species, two cacao species and C. arabica and C. eugenioides.

Between 5 species, there were 2272 Single-copy protein clusters. Among them, 458 clusters were unknown proteins and 491 clusters were no assigned GO terms or functions and clustered based on their sequences. Single-copy clusters are interesting as they possess only one protein of each species that can show how species are related evolutionarily. The more common single-copy clusters exist, the closer species are considered single-copy genes that have not duplicated and transferred between different species and ancestors. In one study, 2,156 single-copy genes in 34 angiosperms were identified and a high-resolution phylogenetic tree was established, indicating that single-copy genes are useful and adequate tools in plants (Li et al., 2017b). Single-copy proteins or genes show ungiven evolutionary events in which species have received only one gene copy from ancestors, deleted copies during evolution or even gene isoforms evolved to other different genes. Whatever event happened, it seems that function of genes might be determinative for gene quantity depending on plant needs to gene product or processes that it manages. In angiosperms, single-copy genes have been reported to be located in organelles generally with binding and catalytic activity, preferentially. Single-copy genes are expressed higher than non-single-copy genes. Single-copy genes indicate more conservation (Han et al., 2014).

3.2 Annotation of top common proteins between 5 species

The proteins of 12020 common clusters were analyzed for their function by UniProt database and annotation based on GO term (Table 2). Interestingly, the top clusters involved transport, resistance, defense and stress response. However, other functions like organelles and regulation were listed on the top common clusters between 5 species. The most counted cluster belonging to Q949G3 UniProt protein with 230 proteins was the PDR1 gene, a transporter protein. Pleiotropic drug resistance (PDR) genes are identified only in fungi and plants. The molecular function of PDR1 is ABC-type transporter activity and ATP binding functioning in hormone transport and regulation of secondary shoot formation. The expression of members of this family have been reported in response to various biotic and abiotic stresses. They transport a diverse array of molecules across membranes and play important roles in detoxification processes, preventing water loss, transport of phytohormones, and secondary metabolites (Nuruzzaman et al., 2014). PDR1 confers the resistance of Nicotiana attenuata against fungal pathogen Alternaria alternata (Xu et al., 2018). NpPDR1 (N. plumbaginifolia) is involved in resistance to pathogens like the fungus Botrytis cinerea, and the bacteria Pseudomonas syringae pv tabaci, P. fluorescens, and P. marginalis pv marginalis via both constitutive and jasmonic acid-dependent induced defense (Stukkens et al., 2005). PDR members are involved in shoot and root, growth regulator-mediated signaling and phytohormone transport (Migocka et al., 2012). Interestingly, in Saccharomyces cerevisiae, PDR members have been reported that may play a potential role in caffeine-hyperresistance (Sürmeli et al., 2019) and mediate caffeine efflux (Tsujimoto et al., 2015). As mechanisms and roles of PDR1 in plants are not well-known and documented, the hypothesis is that it can be a candidate gene for further study between cacao and coffee in caffeine biosynthesis, transport and accumulation.

The second most abundant protein among 5 species was UniProt Q39214 protein named RPM1. RPM1 plays a role in the restriction of pathogen growth and triggers a defense system, including the hypersensitive response (HR) (Boyes et al., 1998). In a cacao gene expression study between tolerant and susceptible genotypes of Phytophthora palmivora, expression of RPM1 was induced by P. palmivora confirming its role in disease resistance (Baruah et al., 2022). In coffee, the Non-race specific Disease Resistance 1 (NDR1) function depends on RPM1 resistance gene. NDR1 is a key regulator initiating hypersensitive signaling pathways. In C. arabica NDR1 genes have been reported as likely targets for manipulating the coffee innate immune system and achieving broad-spectrum resistance to pathogens (Cacas et al., 2011). Therefore, RPM1 is another candidate gene for improving plant resistance in cacao and coffee. Further functional genomic studies will clarify its function and mechanism in pathogen resistance/tolerance in cacao and coffee as a common resistance process.

The third top common clustered protein of all 5 species was Q9LRR4 or Putative disease resistance RPP13-like protein 1 (RPPL1) and the subsequent proteins also belonged to this Putative disease resistance RPP13-like protein family such as RPP13L3 (Q9STE7). Like the first and second proteins in the list in Table 2, GO term for this protein suggests that it might be involved in defense and stress response. A family member of RPP13-like protein in maize has been shown that is involved in abscisic acid-mediated resistance to heat stress (Yang et al., 2021). The authors found ZmRPP13-LK3 has three conserved catalytic adenylyl cyclase (AC) centers. AC synthesizes (cAMP), an essential second messenger with diverse functions. In Arabidopsis, RPP13 bestows resistance to Peronospora parasitica (causal agent of downy mildew) and encodes an NBS-LRR type R protein with a putative amino-terminal leucine zipper (Bittner-Eddy et al., 2000). Few reports on this protein family make it enjoyable to scrutinize further between cacao and coffee.

The next clustered proteins like O81832 (G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290) and Q8LGU1 (ABC transporter C family member 8 ABCC8) are also involved in salt stress response processes (Sun et al., 2013) and transport of varied molecules like metals and heavy metal tolerance (Chen et al., 2019). Interestingly, many shared proteins found among 5 species of coffee and cacao like P92555 (Uncharacterized mitochondrial protein AtMg01250) are not well-known and well-documented, making them candidate genes for resistance and adaptation to stress and disease in cacao and coffee as well as other functions like alkaloid accumulation and biosynthesis. Moreover, they help find novel mechanisms, study biodiversity and evolution, and employ them in plant selection and improvement.

In Table 2, Callose synthase 5 CALS5 was listed with a high protein count among the 5 studied species. CalS5 is involved in forming a callose-rich subintinal layer (SL) and inner cell wall and interestingly, caffeine inhibits the deposition of callose and cellulose an was id cell wall formation (Rivas-Sendra et al., 2019). Accordingly, these proteins might have multifunctional characters and be involved in varied processes where caffeine (and likely other similar-structured metabolites like theobromine) play roles like stress tolerance and consequently interact with other proteins in caffeine biosynthesis.

GO terms of common 12020 clusters were searched for “alkaloid”. Two GO terms were found as GO:0009820 alkaloid metabolic process and GO:0035835 indole alkaloid biosynthetic process assigned to UniProt proteins as Q3L181 and P93083, respectively. Q3L181 is Perakine reductase enzyme and converts raucaffrinoline to perakine (Sun et al., 2008). It has been reported that is involved in biosynthesis of monoterpenoid indole alkaloids (MIAs) in Catharanthus roseus (Murata et al., 2008). P93083 is Tryptophan decarboxylase TDC2, a cytosolic enzyme that catalyzes an early step of the terpenoid indole alkaloid biosynthetic pathway by decarboxylation of l-tryptophan to produce the protoalkaloid tryptamine. Melatonin or 5-Methoxy-N-acetyltryptamine content increased in overexpression of TDC of transgenic rice seed (Byeon et al., 2014). The biosynthesis of melatonin starts from tryptophane, an alkaloid biosynthesis starter. It belongs to protoalkaloid tryptamine compounds and is an antistress agent in plants whose quantity increases upon stress in plants. In addition, it is involved in the circadian cycle, growth and development regulation, rhizogenesis, and photosynthesis (Arnao and Hernández-Ruiz, 2019). Melatonin can increase biosynthesis of alkaloids like galanthamine and lycorine in Leucojum aestivum (Ptak et al., 2019). In rats, melatonin and caffeine showed an interaction and essential role of caffeine in reducing teratogenic effects and also reducing oxidative stress (Yilmaz et al., 2020). Therefore, Perakine reductase and Tryptophan decarboxylase TDC2 are candidate enzyme that can be considered for further studies in cacao and coffee for subsequent alkaloid biosynthesis and mechanisms in biological processes like antistress, hormonal activity.

Table 2. Top 20 Clusters of 5-species comparison. The Swiss-Prot hit. The protein and gene names obtained from UniProt database. Gene names are bold.

Protein Count	*Swiss-Prot Hit*	*Protein and gene name*	*GO Annotation*
230	Q949G3	Pleiotropic drug resistance protein 1 **PDR1**	GO:0006952; P:defense response; IEA:UniProtKB-KW
201	Q39214	Disease resistance protein **RPM1**	GO:0009626; P:plant-type hypersensitive response; IDA:TAIR
119	Q9LRR4	Putative disease resistance RPP13-like protein 1 **RPPL1**	GO:0006952; P:defense response; IEA:UniProtKB-KW
103	O81832	G-type lectin S-receptor-like serine/threonine-protein kinase **At4g27290**	GO:0048544; P:recognition of pollen; IEA:InterPro
103	P92555	Uncharacterized mitochondrial protein **AtMg01250**	GO:0005739; C:mitochondrion; IEA:UniProtKB-SubCell
101	Q9LRR4	Putative disease resistance RPP13-like protein 1 **RPPL1**	GO:0006952; P:defense response; IEA:UniProtKB-KW
87	Q9STE7	Putative disease resistance RPP13-like protein 3 **RPP13L3**	GO:0006952; P:defense response; IEA:UniProtKB-KW

83	Q8LGU1	ABC transporter C family member 8 **ABCC8**	GO:0055085; P:transmembrane transport; IBA:GO_Central
82	Q3B724	Callose synthase 5 **CALS5**	GO:0080092; P:regulation of pollen tube growth; IMP:TAIR
74	Q60EW9	FT-interacting protein 7 **FTIP7**	GO:0005886; C:plasma membrane; IEA:UniProtKB-SubCell
71	Q84JM4	Topless-related protein 3 **TPR3**	GO:0006355; P:regulation of transcription, DNA-templated; IEA:InterPro
67	Q9LV11	ATPase 11, plasma membrane-type **AHA11**	GO:0051453; P:regulation of intracellular pH; IBA:GO_Central
67	Q7XA40	Putative disease resistance protein **RGA3**	GO:0006952; P:defense response; IEA:UniProtKB-KW
66	Q940A6	Pentatricopeptide repeat-containing protein At4g19440, chloroplastic **At4g19440**	GO:0009507; C:chloroplast; IEA:UniProtKB-SubCell
64	Q9FZ87	Indole-3-acetic acid-amido synthetase **GH3.17**	GO:0009733; P:response to auxin; IMP:TAIR
62	F4I460	Myosin-8 **XI-B**	GO:0030048; P:actin filament-based movement; TAS:TAIR
56	Q9LRA7	Probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 9 **TPS9**	GO:0070413; P:trehalose metabolism in response to stress; IBA:GO_Central
55	Q9XGW1	Protein argonaute 10 **AGO10**	GO:0035019; P:somatic stem cell population maintenance; IMP:UniProtKB
53	Q7XA42	Putative disease resistance protein **RGA1**	GO:0006952; P:defense response; IEA:UniProtKB-KW
53	Q8W4P9	Protein TRANSPARENT TESTA 9 **TT9**	GO:0016192; P:vesicle-mediated transport; IMP:TAIR

The UniProt ID of 12020 common genes was submitted to UniProt Retrieve/ID mapping tool to get their functionality. This mapping generated 8890 entries from UniProtKB/SwissProt database. “Caffe” key word was searched among the list, and 8 proteins were assigned to the keyword (Table 3). Five EC codes were assigned to eight proteins and enzymes, including Caffeic acid 3-O-methyltransferase, Caffeoylshikimate esterase, Caffeoyl-CoA 3-O-methyltransferase and Fatty alcohol:caffeoyl-CoA acyltransferase was mapped to UniProt IDs. Caffeic acid 3-O-methyltransferase, Caffeoylshikimate esterase, Caffeoyl-CoA 3-O-methyltransferase enzymes are involved in phenylpropanoid and lignin biosynthesis (Saleme et al., 2017; Vanholme et al., 2013) and convert caffeic acid to CaffeolyCoA (). Fatty alcohol:caffeoyl-CoA acyltransferase is a member of clade Va of the BAHD family of acyltransferases and is involved in root wax biosynthesis in Arabidopsis (Kosma et al., 2012). BAHD family Play roles in regulating plant metabolites, cell stability, response to Phytophthora megakarya fungi and 27 non-redundant putative tcBAHD genes have been reported in T. cacao (Abdullah et al., 2021). Interestingly, these enzymes have not been reported in cacao and coffee species in the literature except the article by Abdullah et al., 2021. Therefore, they are candidate genes for further studies in cacao and coffee for their function in stress response, metabolite biosynthesis and growth and development.

Table 3. Proteins mapped to UniProtKB/SwissProt database and assigned to Caffe keyword. The proteins are enzymes involved in biosynthesis of diverse metabolites like lignin, waxes, alkaloids.

UniProt ID	Protein names
Q9XGP7	Tricin synthase 1 (EC 2.1.1.175) (Caffeoyl-CoA 3-O-methyltransferase ROMT15)
Q9C5D7	Probable caffeoyl-CoA O-methyltransferase At4g26220 (EC 2.1.1.104) (Trans-caffeoyl-CoA 3-O-methyltransferase) (CCoAMT) (CCoAOMT)
Q9C942	Caffeoylshikimate esterase (EC 3.1.1.-) (Lysophospholipase 2) (LysoPL2)
Q43609	Caffeic acid 3-O-methyltransferase (CAOMT) (COMT) (EC 2.1.1.68) (S-adenosysl-L-methionine:caffeic acid 3-O-methyltransferase)
Q9FFQ7	Fatty alcohol:caffeoyl-CoA acyltransferase (EC 2.3.1.-) (Fatty alcohol:caffeoyl-CoA caffeoyl transferase)
Q43095	Caffeoyl-CoA O-methyltransferase (EC 2.1.1.104) (Trans-caffeoyl-CoA 3-O-methyltransferase) (CCoAMT) (CCoAOMT)
P46484	Caffeic acid 3-O-methyltransferase (CAOMT) (COMT) (EC 2.1.1.68) (S-adenosysl-L-methionine:caffeic acid 3-O-methyltransferase)
Q9C9W3	Putative caffeoyl-CoA O-methyltransferase At1g67980 (EC 2.1.1.104) (Trans-caffeoyl-CoA 3-O-methyltransferase) (CCoAMT) (CCoAOMT)

3.3 Annotation for single-copy common proteins

Single-copy protein clusters were assigned to many various functions. However, as no statistical analysis could differentiate them, only some of them were scrutinized for more details. For example, stress-related GO terms were 55 among 2272 protein clusters mainly with salt, oxidative stress response activity as GO:0006979 response to oxidative stress, GO:0009651 response to salt stress, GO:0097501 stress response to the metal ion, GO:0006970 response to osmotic stress, GO:0061408 positive regulation of transcription from RNA polymerase II promoter in response to heat stress, GO:0031098 stress-activated protein kinase signaling cascade and GO:0034976 response to endoplasmic reticulum stress.

Single-copy proteins revealed several functions Q70PR7 and O04847 were found under alkaloid metabolic process. Q70PR7 activity is Vinorine synthase based on UniProt. Vinorine synthase is an enzyme of Acetyltransferase that catalyzes the formation of vinorine (Bayer et al., 2004). Vinorine is a precursor of the antiarrhythmic monoterpenoid indole alkaloid ajmaline. Deacetoxyvindoline 4-hydroxylase (D4H) enzyme generates vindoline (Bose Mazumdar and Chattopadhyay, 2015). These two enzymes are involved in the indole alkaloid biosynthesis pathway (KEGG map 00901). The final product of Vinorine is Ajmaline and vindoline reaches vincristine. However, these alkaloids are not directly involved in metabolism pathways of caffeine and theobromine but might be indirectly involved in alkaloid conversion. Few research publications are available on their relevance to purine alkaloid metabolism, making them interesting proteins for further studies in cacao and coffee.

The proteins related to GO term regulation of transcription, DNA-templated (GO:0006355) were the most counted among single-copy proteins of all 5 species with 41 hits followed by a translation (GO:0006412) with 31 proteins, RNA modification (GO:0009451) with 26, protein ubiquitination (GO:0016567) with 25, protein phosphorylation (GO:0006468) with 24, response to oxidative stress (GO:0006979) with 22, protein transport (GO:0015031) with 21, an integral component of membrane (GO:0016021) with 20, response to salt stress (GO:0009651) with 20, and metal ion binding (GO:0046872) with 20. Mostly the proteins assigned to these GO terms were transcription factors (like MYB98, Scarecrow-like protein 15, Zinc finger protein, Auxin-induced protein, TCP11, Homeobox-leucine zipper protein GLABRA, ILR3, BHLH, MAGPIE, PHD finger protein Alfin), ribosomal proteins, etc. They involve growth and development, stress response, phytohormones, gene expression regulation, and metabolite biosynthesis.

These functionalities suggest that single-copy genes/proteins are important to reveal specific functions of genes and proteins and might be determinative for pathways that are involved in and in the organisms that have them. As only one copy of single-copy genes exists in each species, biological reaction or process might be specific to the species and/or close species. In addition, it might handle certain biological functions that other genes/proteins cannot. For example, interestingly, none of these two alkaloid enzymes, i.e., Vinorine synthase and D4H, have not been found in A. thaliana. D4H has been reported only in some alkaloid plants including Catharanthus roseus and Cannabis sativa and few species Vigna angularis, Vitis vinifera, Gossypium arboreum, Mucuna pruriens and Vinorine synthase has been found only in A. lyrata and few plant species (data retrieved from NCBI). This suggests the specific role and existence of single-copy genes/proteins in plants and further studies on cacao and coffee to reveal how they function.

3.4 GO enrichment of common proteins between all 5 species

GO enrichment for 5 studied species outputted by OrthoVenn2 revealed 36 GO terms (Table 4). Go enrichment showed a relation between common clustered proteins of 5 species and responses to stresses and biosynthesis of alkaloids and secondary metabolites as well as other biological processes like signaling, protein modification, and regulation. As expected, Alkaloid metabolic process, indole glucosinolate metabolic process, UDP-glucosyltransferase activity GO terms that are involved in alkaloid metabolism directly or indirectly were among enriched GO terms among 5 species. In addition, UDP-glucosyltransferase (UGT) glycosylate many secondary metabolites like alkaloids (Speeckaert et al., 2020). The GO terms counted were RNA modification, defense response, rRNA processing, plasma membrane, protein serine/threonine kinase activity, protein autophosphorylation, zinc finger binding and signal transduction, indicating regulatory and signaling function. These suggest that cacao and coffee species are a good model for each other to be compared for stress responses, signaling and regulatory activities. However, functional genomics and further comparative studies are required to target and list the genes/proteins between cacao and coffee species to understand how similar they respond to stresses, manage regulatory processes and signaling and in what biological mechanisms their common genes are involved.

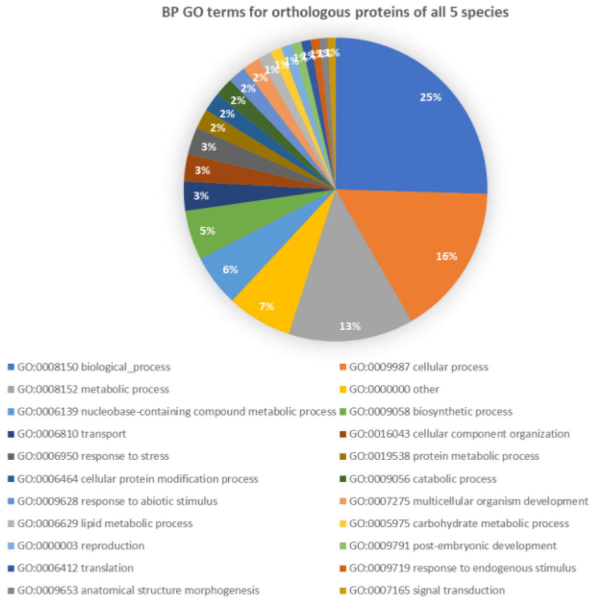
Table 4. GO enrichment for common protein clusters of all 5 species. They showed various activities like enzymatic activity, response to stresses, signaling and regulation.

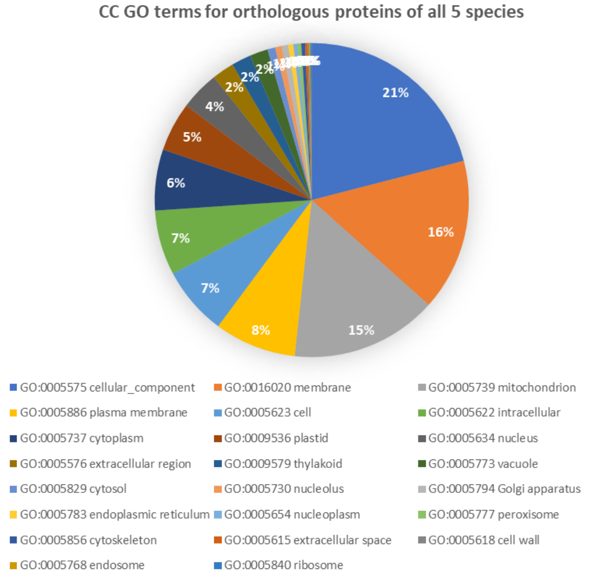
GO ID	*Count*	*Name*	*GO category*	*p-value*
GO:0016705	23	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	1.8668986312064635e-23
GO:0009820	6	alkaloid metabolic process	BP	5.25505878754441e-18

GO:0006952	87	defense response	BP	9.57508966212676e-17
GO:0009626	10	plant-type hypersensitive response	BP	1.7666248443478174e-16
GO:0006364	81	rRNA processing	BP	4.895344355105526e-12
GO:0016758	5	transferase activity, transferring hexosyl groups	MF	1.2768206373961726e-11
GO:0004674	58	protein serine/threonine kinase activity	MF	1.175103804599375e-10
GO:0016491	43	oxidoreductase activity	MF	2.9689846165511735e-09
GO:0046777	49	protein autophosphorylation	BP	4.122247150585193e-09
GO:0009791	2	post-embryonic development	BP	2.5287573158387284e-08
GO:0050832	9	defense response to fungus	BP	3.749207848181802e-08
GO:0035251	5	UDP-glucosyltransferase activity	MF	1.1766324859875054e-07
GO:0019748	2	secondary metabolic process	BP	1.9768578195130687e-07
GO:0009451	159	RNA modification	BP	3.43693935704881e-07
GO:0007018	25	microtubule-based movement	BP	4.295515902035284e-07
GO:0045735	10	nutrient reservoir activity	MF	8.859932494596022e-07
GO:0042343	2	indole glucosinolate metabolic process	BP	9.103780183400925e-07
GO:0016747	4	transferase activity, transferring acyl groups other than amino-acyl groups	MF	3.014296461342774e-06
GO:0048046	6	apoplast	CC	1.4589348257422275e-05
GO:0007166	2	cell surface receptor signaling pathway	BP	1.8374204242546806e-05
GO:0005886	77	plasma membrane	CC	2.592606768634989e-05
GO:0000272	8	polysaccharide catabolic process	BP	3.290283090942899e-05
GO:0042254	16	ribosome biogenesis	BP	6.949379962074648e-05
GO:1900994	2	(-)-secologanin biosynthetic process	BP	7.995745893382696e-05
GO:0009607	21	response to biotic stimulus	BP	0.00014482738927661313
GO:0009826	46	unidimensional cell growth	BP	0.00017690416475511168
GO:0009734	12	auxin-activated signaling pathway	BP	0.00027606450123583264
GO:0019252	19	starch biosynthetic process	BP	0.00028453079430714373
GO:0008270	46	zinc ion binding	MF	0.00034460315362313534
GO:0007165	41	signal transduction	BP	0.0003449461817400814
GO:0009407	6	toxin catabolic process	BP	0.0003808870044474163
GO:0008194	14	UDP-glycosyltransferase activity	MF	0.0003883476405828039
GO:0010182	24	sugar mediated signaling pathway	BP	0.000401550353235148
GO:0006310	2	DNA recombination	BP	0.0005429696370070311
GO:0005634	31	nucleus	CC	0.0005665636542377944
GO:0045454	28	cell redox homeostasis	BP	0.0006628074143379066

3.5 GO analysis for all shared proteins and single-copy proteins of 5 species

GO GOSlimViewer classified terms of shared proteins between all 5 species. The results for BP, MF and CC, are shown in Figure 4. Each category's first top GO terms were general terms, including biological_process, molecular_function and cellular_component. Forty-five GO terms were classified for BP, 22 for MP and 23 for CC. Metabolic and biosynthetic process, stress response, transport, and cell-related GO terms were the most assigned GO terms of BP category. For MF category, binding terms related to nucleic acids (DNA, RNA) and proteins and metabolite-related GO terms were mainly GO terms. GO terms of CC category at the top of the list were membrane, mitochondrion and plasma membrane. These results indicate proteins of cacao and coffee are similar in metabolite biosynthesis and transport activities, suggesting their similarity in biological processes and functions. However, further analyses on specific genes shared between different species are needed to reveal the exact similarity function between cacao and coffee species.





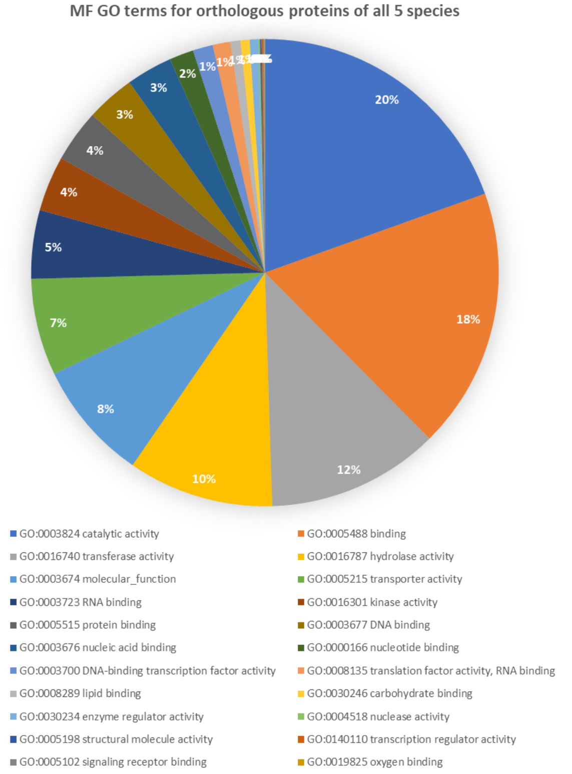


Figure 4. GO analysis for common proteins between all 5 species. The top GO terms for each category were used. Another category for BP is the sum of all GO terms with less than 200 assigned numbers.

GO terms for single-copy proteins were analyzed by GOSlimViewer. The first three most abundant GO terms for the BP category were general terms followed by biosynthetic process, nucleobase-containing compound metabolic process, response to stress, transport, cellular component organization, and catabolic process (Figure 5). For the MF category, binding-related GO terms were the most abundant and enzymatic activity terms, including transferase activity and hydrolase activity, were also among the first most abundant GO terms. Considering BP, CC and MF categories, main terms belonged to organelles, catalytic, binding activities. This is following the report that single-copy genes in angiosperms mainly bind and are catabolic (Han et al., 2014). However, our results showed stress response, biosynthesis and transport activities among most abundant GO terms of single-copy genes of 5 studied species. Copy number of genes has been shown that is critical for stress response in A. halleri and A. thaliana. A. halleri has more copies of genes involved in heavy metal tolerance and accumulation (Jazayeri et al., 2019; Suryawanshi et al., 2016). Therefore, these GO terms for single-copy proteins showed that they might be responsible for key processes by which cacao and coffee respond to stresses in cacao and coffee.

Interestingly, signaling-related terms were not only two among all GO terms between cacao and coffee, suggesting that signaling genes possess more than one copy and isoform. This is because of the ubiquitous character of signaling genes and their involvement in varied processes simultaneously, like photosynthesis and stress response networks where a rapid reaction is required to protect the plant (Kollist et al., 2019). However, comparative studies will remain to disclose how cacao and coffee single-copy genes and proteins are involved in signaling and stress responses.

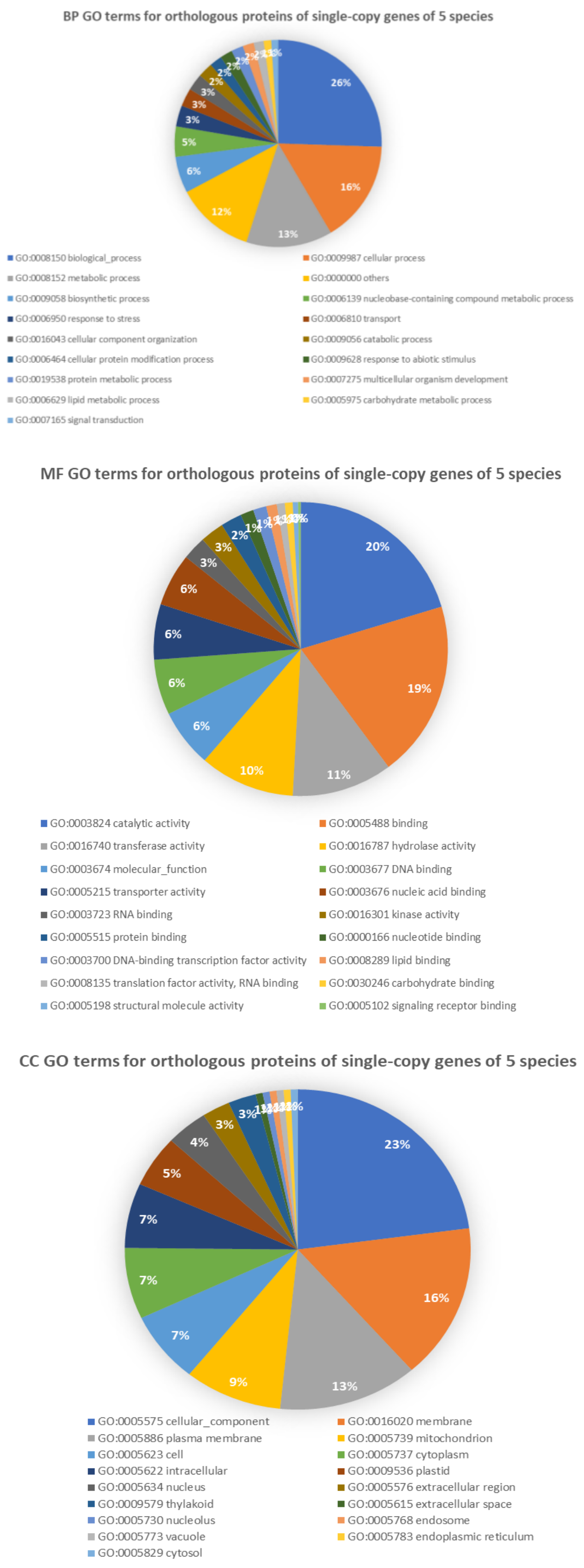


Figure 5. GO analysis for single-copy proteins of 5 species for 3 categories of BP, MF and CC. The first most abundant GO terms were general terms. The terms more than 50 were used and GO terms less than 50 were grouped as others. For CC category all terms were used to create the figure.

3.6 Protein network of single-copy protein orthologs of 5 species

Single-copy proteins of all 5 species were distributed in 2272 clusters that, after filtering of N/A clusters, formed a list of 1751 proteins. This list was submitted to WebGestalt tool to create a gene network. Only plant data of WebGestalt to create a gene network was A. thaliana. Using 1751 UniProt proteins as input, 1280 proteins were identified as network seeds with Arabidopsis gene identifier. Out of 1280 Arabidopsis seeds, 638 were mapped in the selected network and finally, 360 proteins were used for creating the expanded sub-network (Figure 6), top-ranking genes of the main sub-network (Table 5) and GO enrichment (Table 6). Seven sub-networks were assigned from which the main sub-network (PNC: principal network component) contained 10 top-ranking gene neighbors and were chosen for further analyses. These top-ranking neighbors are the genes based on which the sub-network was formed (Table 5). These genes belong to TFs, degradation and regulation functions and transport and are involved in varied plant processes like growth and development, stress tolerance and signaling. Interactions between these top ten protein seeds of the PNC were 75 edges for UBQ3 followed by GRF members with 62, NTL9 with 22, as the highest ones.

The first gene with the greatest edges of the PNC was UBQ3, encoding ubiquitin attached to proteins destined for degradation and is a universal gene involved in many processes. This high number of edges for UBQ3 might be due to its ubiquitous and omnipresent character. However, it is expressed in higher levels in vegetative tissue and modulated by UV-B and light/dark treatments (Brosché et al., 2002; McNeilly et al., 2018). Another ubiquitin-related gene of the PNC was SMALL UBIQUITIN-LIKE MODIFIER 3 SUMO3 and its family members SUMO1 and SUMO2. They belong to the SUMO family that shows remarkable sequence conservation and have been used to diversify plant species, including tomato, potato, strawberry, grape vine, papaya, cacao and sweet orange (Hammoudi et al., 2016). SUMOylation is a biological reaction that refers to the covalent attachment of SUMO to target proteins that alters their function, location, and protein-protein interactions (Hu et al., 2022). Copy number of SUMO genes and its increased levels activate immunity and alter flowering time in Arabidopsis, suggesting this protein family is a key gene in protein regulation of immunity. Furthermore, SUMO3 is involved in SA-induced defense gene expression and sumoylates NPR1 which play important roles to biotic and abiotic stresses and has been reported in cacao and coffee (Backer et al., 2019). Therefore, SUMO family members are candidates for studying cacao and coffee diversification and stress response.

Growth-regulating factors (GRFs) are highly conserved transcription factors. GRF target genes showed broad regulatory functions in plant growth and development, phytohormone biosynthesis and signaling, and the cell cycle GRFs regulate many developmental processes and plant responses to biotic and abiotic stimuli and growth-defense antagonism and modulate the levels of defense- and development-related hormones in opposite directions (Piya et al., 2020). GRFs and miR396 modulate in various plant mechanisms and processes like growth and development, drought, salinity, cold, and UV stress. In T. cacao species 10 members of GRF have been reported (Omidbakhshfard et al., 2015). RAE1 belongs to nucleoporins and is involved in unction in mRNA export and spindle assembly (Tamura et al., 2010). In the generated sub-network, UBQ3 and SUMO3 had interactions with all GRFs suggesting an involvement in same processes. However, as GRFs are TF, they might regulate UBQ3 and its target genes or UBQ3 might be involved in GRF activation/suppression. This remains for further functional genomic studies as no evidence is available for this hypothesis in the literature.

AT4G30840 is a member of Transducin/WD40 repeat-like protein family that controls seed germination, growth and biomass accumulation (Gachomo et al., 2014), cell wall formation (Guerriero et al., 2015). Transducin/WD40 repeat-like superfamily protein and a member of E3 ubiquitin-protein ligase have been reported in cacao genotypes induced in Phytophthora palmivora infection that triggers tolerance to this pathogen (Baruah et al., 2022). As seen in cacao, this gene and UBQ3 might manage disease resistance in coffee.

Of note, as they are single-copy genes found in 5 compared species, they might play key roles in the processes they are involved in, as mentioned above. However, the network was formed based on Arabidopsis data and further studies on them in both cacao and coffee species will be useful to reveal the functions of these genes in mechanisms and processes like disease and stress tolerance and to find how they interact with other single-copy genes and genes with higher copy number in cacao and coffee.

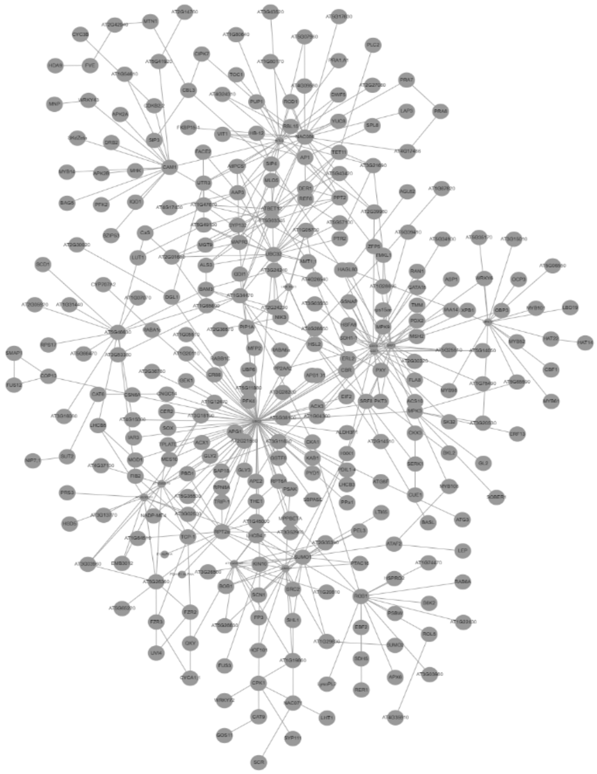


Figure 6. The main sub-network or Principal network component PNC of single-copy proteins of 5 studied species created by WebGestalt. It contained 10 top-ranking proteins as the main edges based on which the PNC was constructed. This network was created based on single-copy proteins showing their importance of comparison between different species.

Table 5. Top ranking neighbors of single-copy proteins of all 5 species. WebGestalt listed these 10 genes based on orthologs of Arabidopsis for UniProt IDs for single-copy proteins.

Gene Symbol	***Annotation***	***Interaction edge***	***Random Walk Probability***
UBQ3	POLYUBIQUITIN 3	75	1.50E-02
GRF3	GROWTH-REGULATING FACTOR 3	23	4.00E-03
GRF1	GROWTH-REGULATING FACTOR 1	21	3.58E-03
GRF8	GROWTH-REGULATING FACTOR 8	18	2.23E-03
NTL9	NAC TRANSCRIPTION FACTOR-LIKE 9	22	1.99E-03
AT1G68185	Ubiquitin-like superfamily protein	14	1.98E-03
AT4G30840	NUP43 Transducin/WD40 repeat-like superfamily protein	11	1.98E-03
MPK10	MAP KINASE 10	15	1.93E-03
SUMO3	SMALL UBIQUITIN-LIKE MODIFIER 3	17	1.84E-03
AT1G80670	RAE1, RNA EXPORT FACTOR 1	12	1.75E-03

The GO enrichment for single-copy proteins of 5 species showed almost all GO terms for metabolite transport and only one GO term indicated regulation of DNA transcription. This might be due to different steps done by WebGestalt to reach a final list based on Arabidopsis data, while many UniProt proteins outputted by OrthoVenn2 were assigned to other plants. However, considering statistical analyses to get the final GO enrichment results as proof for data analysis, this result is interesting as single-copy orthologs were involved in one biological process category, transport (Table 6). This suggests that single-copy proteins are specific among cacao and coffee species but their exact function remains to be disclosed further.

Table 6. GO enrichment for single-copy proteins of 5 species generated by WebGestalt. Almost all are related to metabolite transport. WebGestalt listed them based on raw and adjusted p-value.

GO ID	*GO Name*	*Gene number in GO category*	*Gene number of input in GO category*	*Raw P-value*	*Adjusted P-value*
GO:0015713	phosphoglycerate transmembrane transport	6	4	0	0.0513
GO:0015717	triose phosphate transport	6	4	0	0.0513
GO:0035436	triose phosphate transmembrane transport	6	4	0	0.0513
GO:0042873	aldonate transmembrane transport	6	4	0	0.0513
GO:0015718	monocarboxylic acid transport	11	5	0	0.0513
GO:1901264	carbohydrate derivative transport	26	7	0.0001	0.0647
GO:2000142	regulation of DNA-templated transcription, initiation	17	5	0.0005	0.2526
GO:0015849	organic acid transport	69	10	0.0006	0.2526
GO:0046942	carboxylic acid transport	69	10	0.0006	0.2526
GO:0015711	organic anion transport	95	12	0.0006	0.2526

4 Conclusions

Protein comparisons between species of cacao and coffee showed highly shared proteins. The orthologs of 5 species are involved in transport, regulation, and metabolite biosynthesis. Single-copy proteins of 5 studied species revealed that these proteins are of help for further functional studies due to their specific function related to plant species. Two cacao species shared almost more than half of their genes with coffee species. This is of special interest as these species grow in tropical zones under similar ecological conditions and produce the same metabolites as caffeine and theobromine. As an interesting finding, transport-related proteins were common among 5 studied species, especially single-copy proteins. Therefore, further comparative functional genomic studies can reveal more details about their similarities, common biological mechanisms and pathways involved in transport, biosynthesis of metabolites and their function in stress response. Comparison between 3 Coffea species showed that C. eugenioides share more proteins with C. arabica and C. canephora as crop coffees. This suggest that C. eugenioides is still a natural wild source for coffee breeding programs. Proteins involved in resistance and stress responses were highly overrepresented among the studied species suggesting that they may be candidates for further studies for breeding and selection of cacao and coffee using the data from wild species.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest for this article's research, authorship, and publication.

Funding and acknowledgment

The authors thank the Universidad Técnica Estatal de Quevedo (UTEQ), Ecuador, for financing the research, through the FOCICYT-2019 Seventh Call for Competitive Funds. In addition, the authors also thank Professor M.Sc. Raquel Guerrero Chuez, for her contribution to improving this manuscript.

REFERENCES

1. Abdullah, Faraji S., Heidari P., Poczai P. (2021) The BAHD Gene Family in Cacao (Theobroma cacao, Malvaceae): Genome-Wide Identification and Expression Analysis. Frontiers in Ecology and Evolution. Frontiers Media SA 9: 696.
2. Aerts R., Berecha G., Gijbels P., Hundera K., Van Glabeke S., Vandepitte K., Muys B., Roldán-Ruiz I., Honnay O. (2013) Genetic variation and risks of introgression in the wild Coffea arabica gene pool in south-western Ethiopian montane rainforests. Evolutionary Applications. Wiley-Blackwell 6: 243.
3. Aikpokpodion P.O. (2012) Phenology of flowering in cacao (Theobroma cacao) and its related species in Nigeria. African Journal of Agricultural Research. Academic Journals 7: 3395–3402.
4. Arévalo-Gardini E., Arévalo-Hernández C.O., Baligar V.C., He ZL (2017) Heavy metal accumulation in leaves and beans of cacao (Theobroma cacao L.) in major cacao growing regions in Peru. Science of the Total Environment. Elsevier B.V. 605–606: 792–800.

5. Argout X. et al. (2011) The genome of Theobroma cacao. Nature Genetics. Nature Publishing Group 43: 101–108.
6. Arnao M.B., Hernández-Ruiz J. (2019) Melatonin and reactive oxygen and nitrogen species: a model for the plant redox network. Melatonin Research. ST Bio-life 2: 152–168.
7. Avila-Diazgranados R.A., Daza E.S., Navia E., Romero H.M. (2016) Response of various oil palm materials (Elaeis guineensis and Elaeis oleifera × Elaeis guineensis interspecific hybrids) to bud rot disease in the southwestern oil palm-growing area of Colombia. Agronomía Colombiana. Universidad Nacional de Colombia, Facultad de Agronomía 34: 74–81.
8. Backer R., Naidoo S., van den Berg N. (2019) The NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1) and Related Family: Mechanistic Insights in Plant Disease Resistance. Frontiers in Plant Science. Frontiers Media SA 10.
9. Baroni L., Sarni A.R., Zuliani C. (2021) Plant Foods Rich in Antioxidants and Human Cognition: A Systematic Review. Antioxidants 2021, Vol. 10, Page 714. Multidisciplinary Digital Publishing Institute 10: 714.
10. Baruah I.K., Ali S.S., Shao J., Lary D., Bailey B.A. (2022) Changes in Gene Expression in Leaves of Cacao Genotypes Resistant and Susceptible to Phytophthora palmivora Infection. Frontiers in Plant Science. Frontiers Media SA 12: 3281.
11. Bateman A. et al. (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. Oxford Academic 49: D480–D489.
12. Bayer A., Ma X., Stöckigt J. (2004) Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase. Bioorganic & medicinal chemistry. Bioorg Med Chem 12: 2787–2795.
13. Bittner-Eddy P.D., Crute I.R., Holub E.B., Beynon J.L. (2000) RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. The Plant journal : for cell and molecular biology. Plant J 21: 177–188.
14. Bose Mazumdar A., Chattopadhyay S. (2015) Sequencing, De novo Assembly, Functional Annotation and Analysis of Phyllanthus amarus Leaf Transcriptome Using the Illumina Platform. Frontiers in Plant Science. Frontiers Media SA 6: 1199.
15. Boyes D.C., Nam J., Dangl J.L. (1998) The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. Proceedings of the National Academy of Sciences of the United States of America. 95: 15849–15854.
16. Brosché M., Schuler M.A., Kalbina I., Connor L., Strid Å. (2002) Gene regulation by low level UV-B radiation: identification by DNA array analysis. Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology. Photochem Photobiol Sci 1: 656–664.
17. Byeon Y., Park S., Lee H.Y., Kim Y.S., Back K. (2014) Elevated production of melatonin in transgenic rice seeds expressing rice tryptophan decarboxylase. Journal of Pineal Research. John Wiley & Sons, Ltd 56: 275–282.
18. Cacas J.L., Petitot A.S., Bernier L., Estevan J., Conejero G., Mongrand S., Fernandez D. (2011) Identification and characterization of the Non-race specific Disease Resistance 1 (NDR1) orthologous protein in coffee. BMC Plant Biology. BioMed Central 11: 1–17.
19. Camandola S., Plick N., Mattson M.P. (2019) Impact of Coffee and Cacao Purine Metabolites on Neuroplasticity and Neurodegenerative Disease. Neurochemical research. NIH Public Access 44: 214.
20. Castro Pinacy C.V., Barrezueta Unda S. (2020) Aspectos sociales y económicos: caso productores de café en la provincia El Oro | Castro Pincay | Revista Metropolitana de Ciencias Aplicadas. Revista Metropolitana de Ciencias Aplicadas. 3.
21. Cenci A., Combes M.C., Lashermes P. (2012) Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. Plant molecular biology. Plant Mol Biol 78: 135–145.
22. Chen C., Cao Q., Jiang Q., Li J., Yu R., Shi G. (2019) Comparative transcriptome analysis reveals gene network regulating cadmium uptake and translocation in peanut roots under iron deficiency. BMC Plant Biology. BioMed Central 19.
23. Dagilis A.J., Kirkpatrick M., Bolnick DI (2019) The evolution of hybrid fitness during speciation. PLOS Genetics. Public Library of Science 15: e1008125.
24. Damatta F.M., Avila R.T., Cardoso A.A., Martins S.C.V., Ramalho J.C. (2018) Physiological and Agronomic Performance of the Coffee Crop in the Context of Climate Change and Global Warming: A Review. Journal of Agricultural and Food Chemistry. American Chemical Society 66: 5264–5274.
25. Delaux P.-M. et al. (2019) Reconstructing trait evolution in plant evo-devo studies. Current Biology. 29.
26. Doodnath R. (2008) Study of intergeneric hybridization between Theobroma Cacao L. and Herrania species. The University of The West Indies.
27. Gachomo E.W., Jimenez-Lopez J.C., Baptiste L.J., Kotchoni S.O. (2014) GIGANTUS1 (GTS1), a member of Transducin/WD40 protein superfamily, controls seed germination, growth and biomass accumulation through ribosome-biogenesis protein interactions in Arabidopsis thaliana. BMC Plant Biology. 14.
28. Gartner G.A.L., McCouch S.R., Botero MDPM (2014) Segregation analysis of molecular markers in a population derived from Coffea liberica Hiern x C. eugenioides L. Acta Agronómica. Universidad Nacional de Colombia 63: 153–163.
29. González-Orozco C.E., Galán A.A.S., Ramos P.E., Yockteng R. (2020) Exploring the diversity and distribution of crop wild relatives of cacao (Theobroma cacao L.) in Colombia. Genetic Resources and Crop Evolution. Springer Science and Business Media B.V. 67: 2071–2085.
30. Guerriero G., Hausman J.F., Ezcurra I. (2015) Wd40-repeat proteins in plant cell wall formation: Current evidence and research prospects. Frontiers in Plant Science. Frontiers Media SA 6: 1112.
31. Hamer M., Williams E.D., Vuononvirta R., Gibson E.L., Steptoe A. (2006) Association between coffee consumption and markers of inflammation and cardiovascular function during mental stress. Journal of Hypertension. Lippincott Williams and Wilkins 24: 2191–2197.
32. Hammoudi V., Vlachakis G., Schranz M.E., van den Burg H.A. (2016) Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. New Phytologist. 211: 172–185.
33. Han F., Peng Y., Xu L., Xiao P. (2014) Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics. BioMed Central Ltd. 15: 1–9.
34. Hu X., Xiao X., Zhang C.-L., Wang G.-L., Zhang Y.-L., Li Y.-Y., You C.-X. (2022) Organization and regulation of the apple SUMOylation system under salt and ABA. Plant Physiology and Biochemistry. Elsevier Masson.
35. Jackson M., Marks L., May G.H.W., Wilson J.B. (2018) The genetic basis of disease. Essays in Biochemistry. Portland Press Ltd 62: 643.
36. Jazayeri S.M., Oviedo-Bayas B., Guerrero-Chuez R., Torres-Navarrete Y., Villamar-Torres R.O. (2021) Environmental Factors Enhance Production of Plant Secondary Metabolites Toward More Tolerance and Human Health: Cocoa and Coffee Two Model Species. Innovations in Biotechnology for a Sustainable Future. Springer International Publishing 155–183.
37. Jazayeri S.M., Rivera Y.D., Camperos-Reyes J.E., Romero H.M. (2015) Physiological effects of water deficit on two oil palm (Elaeis guineensis Jacq.) genotypes. Agronomía Colombiana. 33: 164–173.
38. Jazayeri S.M., Villamar-Torres R.O., Zambrano-Vega C., Cruzatty L.C.G., Oviedo-Bayas B., Santos M. do A., Maddela N.R., Ghafoor SMHAS, Viot C. (2020) Transcription factors and molecular markers revealed asymmetric contributions between allotetraploid Upland cotton and its two diploid ancestors. Bragantia.
39. Jazayeri S.M.M., García Cruzatty L.C., Villamar-Torres R., Cruzatty L.C.G., Villamar-Torres R. (2019) Genomic comparison among three arabidopsis species revealed heavy metal responsive genes. Journal of Animal and Plant Sciences. 29.
40. Kiwuka C. et al. (2021) Genetic diversity of native and cultivated Ugandan Robusta coffee (Coffea canephora Pierre ex A. Froehner):

Climate influences, breeding potential and diversity conservation. PLOS ONE. Public Library of Science 16: e0245965.

41. Kollist H., Zandalinas S.I., Sengupta S., Nuhkat M., Kangasjärvi J., Mittler R. (2019) Rapid Responses to Abiotic Stress: Priming the Landscape for the Signal Transduction Network. Trends in Plant Science. Elsevier Current Trends 24: 25–37.
42. Kosma D.K., Molina I., Ohlrogge J.B., Pollard M. (2012) Identification of an Arabidopsis Fatty Alcohol:Caffeoyl-Coenzyme A Acyltransferase Required for the Synthesis of Alkyl Hydroxycinnamates in Root Waxes. Plant Physiology. Oxford University Press 160: 237.
43. Lashermes P., Combes M.C., Robert J., Trouslot P., D'Hont A., Anthony F., Charrier A. (1999) Molecular characterisation and origin of the Coffea arabica L. genome. Molecular and General Genetics MGG 1999 261:2. Springer 261: 259–266.
44. Li Z. et al. (2017a) Single-Copy Genes as Molecular Markers for Phylogenomic Studies in Seed Plants. Genome Biology and Evolution. Oxford Academic 9: 1130–1147.
45. Li Z. et al. (2017b) Single-copy genes as molecular markers for phylogenomic studies in seed plants. Genome Biology and Evolution. 9: 1130–1147.
46. Mawalagedera S.M.U.P., Callahan D.L., Gaskett A.C., Rønsted N., Symonds M.R.E. (2019) Combining Evolutionary Inference and Metabolomics to Identify Plants With Medicinal Potential. Frontiers in Ecology and Evolution. Frontiers Media S.A. 7: 267.
47. McCarthy F.M. et al. (2006) AgBase: a functional genomics resource for agriculture. BMC genomics. BioMed Central 7: 229.
48. McCarthy F.M., Bridges S.M., Wang N., Magee G.B., Williams W.P., Luthe D.S., Burgess S.C. (2007) AgBase: a unified resource for functional analysis in agriculture. Nucleic Acids Research. Oxford University Press 35: D599.
49. McNeilly D., Schofield A., Stone S.L. (2018) Degradation of the stress-responsive enzyme formate dehydrogenase by the RING-type E3 ligase Keep on Going and the ubiquitin 26S proteasome system. Plant molecular biology. Plant Mol Biol 96: 265–278.
50. Merot-L'anthoene V. et al. (2019) Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of Coffea arabica L. .
51. Migocka M., Papierniak A., Warzybok A., Kłobus G. (2012) CsPDR8 and CsPDR12, two of the 16 pleiotropic drug resistance genes in cucumber, are transcriptionally regulated by phytohormones and auxin herbicide in roots. Plant Growth Regulation. Kluwer Academic Publishers 67: 171–184.
52. Montoya C. et al. (2014) Genetic Architecture of Palm Oil Fatty Acid Composition in Cultivated Oil Palm (Elaeis guineensis Jacq.) Compared to Its Wild Relative E. oleifera (H.B.K) Cortés. PLOS ONE. Public Library of Science 9: e95412.
53. Murata J., Roepke J., Gordon H., De Luca V. (2008) The Leaf Epidermome of Catharanthus roseus Reveals Its Biochemical Specialization. The Plant Cell. Oxford Academic 20: 524–542.
54. Nuruzzaman M., Zhang R., Cao H.Z., Luo Z.Y. (2014) Plant Pleiotropic Drug Resistance Transporters: Transport Mechanism, Gene Expression, and Function. Journal of Integrative Plant Biology. John Wiley & Sons, Ltd 56: 729–740.
55. Omidbakhshfard M.A., Proost S., Fujikura U., Mueller-Roeber B. (2015) Growth-regulating factors (GRFs): A small transcription factor family with important functions in plant biology. Molecular Plant. 8: 998–1010.
56. Piya S., Liu J., Burch-Smith T., Baum T.J., Hewezi T. (2020) A role for Arabidopsis growth-regulating factors 1 and 3 in growth-stress antagonism. Journal of experimental botany. J Exp Bot 71: 1402–1417.
57. Polturak G., Liu Z., Osbourn A. (2022) New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. Current Opinion in Green and Sustainable Chemistry. Elsevier 33: 100568.
58. Ptak A., Simlat M., Morańska E., Skrzypek E., Warchoł M., Tarakemeh A., Laurain-Mattar D. (2019) Exogenous melatonin stimulated Amaryllidaceae alkaloid biosynthesis in in vitro cultures of Leucojum aestivum L. Industrial Crops and Products. Elsevier 138: 111458.
59. Rivas-Sendra A., Corral-Martínez P., Porcel R., Camacho-Fernández C., Calabuig-Serna A., Seguí-Simarro J.M. (2019) Embryogenic competence of microspores is associated with their ability to form a callosic, osmoprotective subintinal layer. Journal of Experimental Botany. Oxford Academic 70: 1267–1281.
60. Salamah A., Rostina I. (2019) Analysis of AGAMOUS Gene Expression in Hibiscus rosasinensis L. Single Pink, Crested Peach, and Double Orange Flowers. Journal of Physics: Conference Series. IOP Publishing 1245: 12001.
61. Saleme M. de LS et al. (2017) Silencing CAFFEOYL SHIKIMATE ESTERASE Affects Lignification and Improves Saccharification in Poplar. Plant Physiology. Oxford University Press 175: 1040.
62. Scalabrin S. et al. (2020) A single polyploidization event at the origin of the tetraploid genome of Coffea arabica is responsible for the extremely low genetic variation in wild and cultivated germplasm. Scientific Reports 2020 10:1. Nature Publishing Group 10: 1–13.
63. Sousa Silva C.R., Figueira A. (2004) Phylogenetic analysis of Theobroma (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. Plant Systematics and Evolution 2004 250:1. Springer 250: 93–104.
64. Speeckaert N. et al. (2020) Characterization of the UDP-glycosyltransferase UGT72 Family in Poplar and Identification of Genes Involved in the Glycosylation of Monolignols. International Journal of Molecular Sciences. Multidisciplinary Digital Publishing Institute (MDPI) 21: 1–24.
65. Stukkens Y., Bultreys A., Grec S., Trombik T., Vanham D., Boutry M. (2005) NpPDR1, a Pleiotropic Drug Resistance-Type ATP-Binding Cassette Transporter from Nicotiana plumbaginifolia, Plays a Major Role in Plant Pathogen Defense. Plant Physiology. Oxford University Press 139: 341.
66. Sun L., Ruppert M., Sheludko Y., Warzecha H., Zhao Y., Stöckigt J. (2008) Purification, cloning, functional expression and characterization of perakine reductase: the first example from the AKR enzyme family, extending the alkaloidal network of the plant Rauvolfia. Plant Molecular Biology 2008 67:5. Springer 67: 455–467.
67. Sun X.L., Yu Q.Y., Tang L.L., Ji W., Bai X., Cai H., Liu X.F., Ding X.D., Zhu Y.M. (2013) GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress. Journal of plant physiology. J Plant Physiol 170: 505–515.
68. Sürmeli Y., Holyavkin C., Topaloğlu A., Arslan M., Kısakesen H.İ., Çakar Z.P. (2019) Evolutionary engineering and molecular characterization of a caffeine-resistant Saccharomyces cerevisiae strain. World journal of microbiology & biotechnology. World J Microbiol Biotechnol 35.
69. Suryawanshi V., Talke I.N., Weber M., Eils R., Brors B., Clemens S., Krämer U. (2016) Between-species differences in gene copy number are enriched among functions critical for adaptive evolution in Arabidopsis halleri. BMC Genomics. BioMed Central 17.
70. Tamura K., Fukao Y., Iwamoto M., Haraguchi T., Hara-Nishimura I. (2010) Identification and Characterization of Nuclear Pore Complex Components in Arabidopsis thaliana. The Plant Cell. Oxford University Press 22: 4084.
71. Tran H.T.M., Ramaraj T., Furtado A., Lee L.S., Henry R.J. (2018) Use of a draft genome of coffee (Coffea arabica) to identify SNPs associated with caffeine content. Plant Biotechnology Journal. Wiley-Blackwell 16: 1756.
72. Tsujimoto Y., Shimizu Y., Otake K., Nakamura T., Okada R., Miyazaki T., Watanabe K. (2015) Multidrug resistance transporters Snq2p and Pdr5p mediate caffeine efflux in Saccharomyces cerevisiae. OUP. Taylor & Francis 79: 1103–1110.
73. Vanholme R. et al. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. Science (New York, N.Y.). Science 341: 1103–1106.
74. von Wettberg E., Davis T.M., Smýkal P. (2020) Editorial: Wild Plants as Source of New Crops. Frontiers in Plant Science. Frontiers Media S.A. 11: 1426.
75. Weiße A.Y., Oyarzún D.A., Danos V., Swain P.S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. Proceedings of the National Academy of Sciences of the United States

of America. National Academy of Sciences 112: E1038–E1047.
76. Whitlock B.A., Baum D.A. (1999) Phylogenetic relationships of Theobroma and Herrania (Sterculiaceae) based on sequences of the nuclear gene Vicilin. Systematic Botany. American Society of Plant Taxonomists Inc. 24: 128–138.
77. Xu L. et al. (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic acids research. Nucleic Acids Res 47: W52–W58.
78. Xu Z., Song N., Ma L., Fang D., Wu J. (2018) NaPDR1 and NaPDR1-like are essential for the resistance of Nicotiana attenuata against fungal pathogen Alternaria alternata. Plant Diversity. Kunming Institute of Botany, Chinese Academy of Sciences 40: 68.
79. Yang H. et al. (2021) A new adenylyl cyclase, putative disease-resistance RPP13-like protein 3, participates in abscisic acid-mediated resistance to heat stress in maize. Journal of experimental botany. J Exp Bot 72: 283–301.
80. Yilmaz S., Göçmen A.Y., Üner A.K., Akyüz E., Tokpinar A. (2020) The protective role of melatonin against the effect of caffeine on embryonic kidney. Turk Hij Den Biyol Derg. 77: 51–58.
81. Yuyama P.M., Reis Júnior O., Ivamoto S.T., Domingues D.S., Carazzolle M.F., Pereira G.A.G., Charmetant P., Leroy T., Pereira L.F.P. (2016) Transcriptome analysis in Coffea eugenioides, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. Molecular Genetics and Genomics. Springer Verlag 291: 323–336.