Unraveling Cross-Cellular Communication in *Cannabis sativa* Sex Determination: A Network Ontology Transcript Annotation (Nota) Analysis.

Leonardo R. Orozco[1,2*], Audrey E. Weaver[1,2,+], Christopher C. Pauli[1,3+], Christopher J. Grassa[4,+], Daniela Vergara[1,5,+], Anthony Baptista[2,6,7,+], Kristin White[1,+], Benjamin F. Emery[8], Natalie R.M. Castro[2,8], Rafael F. Guerrero[9], Brian C. Keegan[8], and Nolan C. Kane[1,2,+]

[1]University of Colorado Boulder, Ecology and Evolutionary Biology, Boulder, CO 80304, USA
[2]SciAnno Mosaics LLC, Colorado Boulder, USA
[3]Tryptomics LLC, Colorado Longmont, USA
[4]Harvard University, Department of Organismal and Evolutionary Biology, Cambridge, MA, 02138-02142, USA
[5]Cornell Cooperative Extension, Harvest NY, Geneva NY, USA
[6]The Alan Turing Institute, The British Library, London, NW1 2DB, United Kingdom
[7]Cancer Bioinformatics, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom
[8]University of Colorado Boulder, Information Science, Boulder, CO, 80304, USA
[9]North Carolina State University, Department of Biological Sciences, 112 Derieux Pl, Raleigh, NC, 27695, USA
[*]Corresponding Author leonardo.orozco@colorado.edu
[+]these authors contributed equally to this work

ABSTRACT

Cannabis sativa L. (marijuana, hemp) is a dioecious angiosperm species currently evolving sex chromosomes. Genetic mechanisms, primarily controlled by an XY chromosome system, appear to dictate sex expression in *C. sativa*. However, sexual expression is also governed by the interplay of hormone regulatory gene networks, influenced by both genetic and environmental factors. Within the species, some populations exhibit dioecy, monoecy, or a gradient of both. Dioecious individuals produce exclusively male or female flowers, while monoecious plants bear both male and female flowers. However, through interruption of phytohormone signal transduction via abiotic stressors, male and female Cannabis are able to produce flowers of the opposite sex. This mechanism appears to be mediated via long-range signaling cascades enabling alternative cell wall embryo-genesis programming. When the pollen produced by these masculinized XX (genetically female) plants is applied for breeding, the seeds produced all lack a Y chromosome, and are thus considered "feminized" seeds. Previous transcriptomic analysis have identified genes associated with masculinization through the application of phytohormone signal disruption using silver thiosulfate treatment. We used Jack and Nota (Network Ontology Transcript Annotation), tools developed by SciAnno Mosaics and CU Boulder, to analyze transcriptomic data from *C. sativa* treated with colloidal silver to induce masculinization. Jack streamlined differential gene expression analysis and functional annotation of the resulting transcripts. Using Nota's multilayer network analysis (random walk with restart), we identified candidate sex-determining genes. Nota and Jack operate on the principle of building a reference-free biology, enabling discovery and annotation of gene-trait associations without relying on pre-existing genomic references. Our findings highlight Nota's power to dissect the genetic architecture of complex traits, particularly in systems like *C. sativa* where sex determination is influenced by multiple factors.

Introduction

Cannabis sativa L. (marijuana and hemp) is currently recognized as the largest emerging crop worldwide and is in the process of maturing into a versatile and modern cash crop. This flowering plant[1], has been important in human history as one of the first plants domesticated by multiple ancient cultures and for multiple uses, including medicine, food, cloth, rope, and intoxication[2,3]. *Cannabis sativa* is particularly known for its production of cannabinoids, which are unique secondary metabolites that possess medicinal and psychoactive properties[4,5]. In addition, cannabis plants have the capacity to produce strong, high-quality fibers, as well as grain, known for their high nutritional value and wide range of uses[6]. Together, these products make this species one of the most valuable and rapidly growing crops worldwide[7].

Cannabis sativa possesses a diploid genome (2n=20) characterized by nine autosomes and a pair of sex chromosomes, with

Males, the heterogametic sex, harboring X and Y chromosomes and Females carrying two X chromosomes[8]. Although much of the scientific literature portrays *C. sativa* as a dioecious plant, where sex determination depends on the presence or absence of the X or Y chromosome[8], there are numerous populations, in wild or ruderal populations and among certain hemp varieties, where monoecious plants are prevalent. These monoecious plants also seem to have two X chromosomes[9,10], although their X chromosomes may be different than those of dioecious populations. Dioecious varieties remain prevalent for medicinal and recreational purposes, where males and pollination are avoided. The interfertility between monoecious and dioecious varieties demonstrates their close evolutionary relationship.

Recreational *C. sativa* is cultivated for cannabinoid and terpene compounds produced by glandular trichomes, secretory organs concentrated in female Cannabis inflorescence tissue[11]. These compounds have important medicinal and psychoactive properties, mainly due to effects on the endocannabinoid system[12–15]. When heated, Δ -9-tetrahydrocannabinolic acid (THCA) and Cannabidiolic acid (CBDA), convert to neutral and psychoactive forms Δ -9-THC and CBD, respectively[16]. These two cannabinoids, as well as a suite of terpenoids, form the basis of cannabis derived medicinal and recreational products. In female cannabis plants, pollination shifts metabolism from cannabinoid and terpenoid biosynthesis to seed production, reducing the carbon flux available for these valuable compounds. Consequently, males and monoecious individuals are commercially undesirable for recreational growers.[17]. Unlike recreational cannabis, hemp strains, primarily monoecious, are cultivated for fiber or grain, where stalk and seed production is prioritized over cannabinoid and terpene biosynthesis[18], therefore pollination is not a remarkable concern for hemp cultivation[19].

However, the complex interplay between genetic and environmental factors in sex determination is particularly evident in monoecious cannabis plants. These individuals, which bear both male and female flowers, often represent a more plastic sexual phenotype, potentially allowing them to adapt to a wider range of environmental conditions. Beyond the genetic influence of the X and Y chromosomes, environmental stimuli also modulate sex expression[19–21]. Factors such as temperature variations, photoperiod, and the administration of hormones or chemicals interacting with plant hormonal pathways can all influence sex expression in *C. sativa*[22]. Therefore, chemical agents can reprogram sex determination in undifferentiated floral meristem cells, demonstrating substantial plasticity in sex expression.[23,24]. Genetically Female plants, when treated to yield male flowers, produce pollen used for male-free pollination yielding "feminized seeds", seedlings lacking a Y chromosome[25]. Feminized seeds and masculinization treatment agents are commodities of high demand for recreational grow operations, highlighting cannabinoid production.

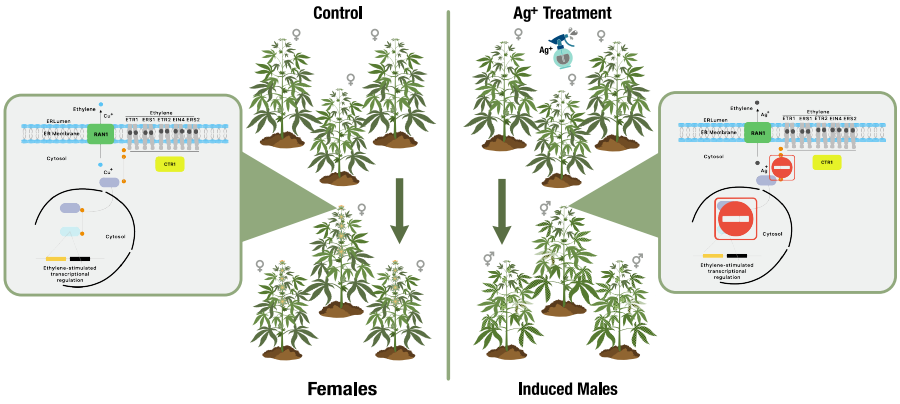

Figure 1. Colloidal Silver Interruption of Ethylene Signal Transduction: Silver ions (Ag^+) inhibit ethylene (C_2H_4) signal transduction in cannabis by out-competing copper ions (Cu^{2+}) at the ethylene receptor sites. This prevents receptor activation and reduces ethylene-related gene expression. The control panel shows uninterrupted ethylene signaling in female cannabis plants, while the treatment panel illustrates how Ag^+-induced hormonal stress disrupts this process, enabling the dioecious female plant to leverage its sexual plasticity for seed production under stress.

Plant hormones (phytohormones) are valuable tools for breeders, enabling them to promote general development and enhance crop immune defenses[26]. Ethylene is essential for various developmental processes, including fruit ripening, as well as shoot generation, seed germination, root nodulation, and flower senescence[27,28]. Ethephon, a synthetic compound widely used in agriculture, up-regulates the ethylene signaling pathway, impacting flowering and plant development[29]. In fruiting plants, ethephon application can accelerate ripening and, notably, has been used to feminize monoecious C. sativa varieties[30]. While ethephon provides an exogenous source of ethylene, heavy metal ions, such as silver, can interfere with the plant's own ethylene production. Silver disrupts metal ion homeostasis, leading to imbalances that affect the transcriptional programming of ethylene biosynthesis[31].

Silver nitrate, in particular, has been effective in inducing male flowers in diverse plant systems, including mulberry[28]. In *C. sativa*, both silver nitrate and silver thiosulfate, when used to reverse sex expression in female plants, yield male flowers capable of producing viable pollen grains for pollination[32]. These applied genetic modifications inhibiting the ethylene signaling pathway in cannabis, suggest that the ethylene signaling pathway may serve as a modulator of sexual determination in monoecious plants. However, the ethylene signaling pathway interacts with other phytohormone signaling pathways through cross-talk (shared functionality) between the gene networks driving them[33].

We followed and extended the methods of Zhu et al. 2023, who predicted drought stress-related genes in rice[34] by employing a random-walk with restart analysis on a multilayer biological network[35] comprised of a gene co-expression network and a species representative cannabis protein-protein interaction network.

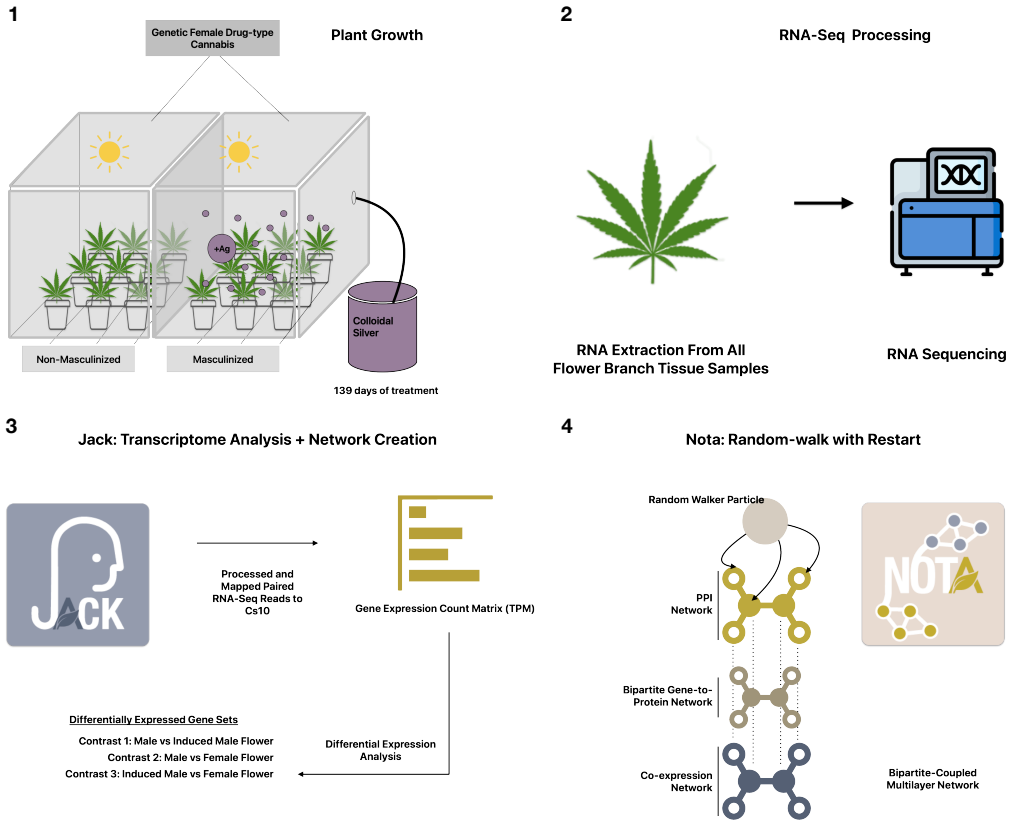

Figure 2. Experimental Work Flow (1) Plant Growth and Treatment: Genetically Female cannabis plants were grown under standard conditions. A subset of plants was treated with colloidal silver to induce masculinization. (2) RNA Sequencing: RNA was extracted and sequenced from Male, Female, and Masculinized Female flowers after 139 days of treatment.(3) Bioinformatic and Statistical Analysis: Raw RNA-seq reads were processed and mapped to the C. sativa reference genome (Cs10) using Jack. Differential expression analysis was performed using DESeq2 to identify genes differentially expressed between the three groups (Male vs. Induced Male, Male vs. Female, and Induced Male vs. Female). (4) Nota: Network Ontology Transcript Annotation, a tool employing a personalized random walk with restart algorithm, was applied to the network to identify key proteins and pathways associated with sexual plasticity.

In our approach, the tool Nota implements a personalized page rank algorithm adapted from the multixrank python package[36] which completes a random walk with restart using seed nodes determined from our differential expression analysis

to infer additional phytohormone stress-related genes in *Cannabis sativa*.

Random walk is a powerful approach to exploring the topology of networks. By simulating the movement of a walker randomly traversing nodes through edges in a network, random walks are able to capture several structural properties of networks[37], including connectivity[38], community structure[39], and node centrality[40]. Inspired by the PageRank algorithm[41], which was initially developed for ranking web pages in search results by simulating the behavior of an internet user following hyperlinks or restarting on arbitrary pages, Random Walk with Restart (RWR) was first introduced by Pan et al. 2004[42]. In the RWR approach used by Nota, the walker particle, at each step, can navigate from one node to one of its neighbors or restart its walk from a node randomly sampled from a set of seed nodes. By enabling restart from one or several seed nodes, RWR simulates a diffusion process in which the objective is to determine the steady state of an initial probability distribution[43]. This steady state represents a measure of proximity between the seed(s) and all the network nodes, rather than a centrality measure as in PageRank. Nota's RWR quantifies the extent to which the influence or information from the seed nodes has spread throughout the network, identifying nodes that are closely connected to the seed(s). Recently, this method has been extended to more complex networks, such as multilayer networks[48], which include multiplex and bipartite networks. Multilayer networks enable the integration of complementary and heterogeneous information into one network. A multiplex network is composed of several layers that consist of the same nodes but differently informed edges, displaying the complex relationships between nodes. For instance, a protein multiplex network can display protein-protein interactions, molecular complexes, and pathways[36]. A bipartite network is composed of interactions between two different kinds of nodes, such as proteins and genes, if they are annotated differently in the study. In our study, we utilized the network construction capabilities of Nota to curate our multilayer biological network. This multilayer is composed of a gene co-expression network and a species representative protein-protein interaction network, connected by a bipartite network that associates each protein with its corresponding gene (1:1).

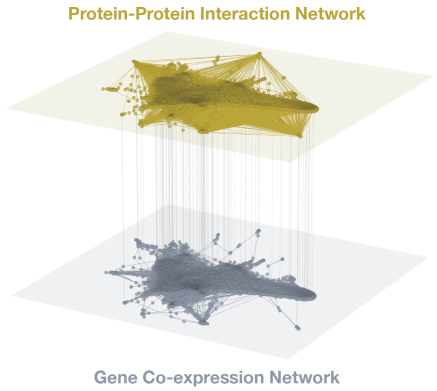

Figure 3. Multilayer Networks Visualization We applied open source methods to visualize the multilayer networks by integrating different graph layers (protein-protein interaction data, co-expression data, and a gene-to-protein gene annotation file generated maps)[44–47].

To investigate the genetics of sex expression in *Cannabis*, we used multilayer networks to analyze gene expression differences among genetically XX Female, genetically XY Male, and hormonal-treatment induced genetically XX Male flowers grown in the same controlled environment. Our analysis is guided by the hypothesis that the high metabolic cost associated with Cannabis masculinization must be accommodated through transcriptional mediation and phytohormone signaling, which regulate general physiology in response to environmental stimuli. Additionally, we hypothesize that several cellular cross-talk systemsincluding protein synthesis, plant development, stress response, and flowering regulatory pathwaysmediate the sexual plasticity of *Cannabis* under environmental signaling pressures.

Results

Differential Expression Analysis

Consistent with similar transcriptomic analysis[20,21,49], genes identified as differentially expressed in our treatment contrasts of Induced males (IM) against true genetic Females were genes associated with; seed, pollen and flower development, ethylene pathway related genes, and phytohormone signaling related genes. We constructed our DESeq2 object, to compare our Male, Female and IM samples. We parsed differentially expressed genes from pandas data frames of the tabulated result object for each treatment contrast, Male vs IM (CN1), Female vs Male (CN2), and IM vs Female (CN3) CNby setting statistical significance parameters as reported in the Methods, for all three experimental contrasts (Table 3). We performed independent analyses of our RNA-seq data and the data from Adal et al. (2021), in parallel using identical differential expression parameters in DESeq2, which revealed a consistent set of up- and down-regulated genes across comparable pairwise contrasts (Table 2). This consistency was observed despite the use of different masculinization agents (colloidal silver in our study vs. silver thiosulfate in Adal et al., 2021), suggesting a shared set of transcriptional responses underlying sexual plasticity in Cannabis sativa. Here, we report additional findings from our differential gene expression analysis: genes associated with developmental metal homeostasis, protein transport/modification, lipid metabolism, and general abiotic stress-related genes, which may be additional components of the genetic architecture underlying sexual plasticity in cannabis.

CN(CN)	CN1: M vs IM	CN2: M vs F	CN3: IM vs F
CN1: Male vs Induced Males	3,324	2,728	34
CN2: Male vs Female	2,728	3,504	111
CN3: Induced Male vs Female	34	111	133

Table 1. Total number of similarly differentially expressed genes (DEGs; Absolute log-fold change ≥ 1 and a p-adj less than or equal to 0.05) identified in pairwise contrasts of our 3 cannabis flower tissue sample pools (Male/Female/Induced Female). Overlapping DEGs calculated in a pairwise comparison of each CN with 6932 total DEGs identified across all three contrasts.

	CN1		CN2		CN3	
Dataset	Up	Down	Up	Down	Up	Down
Orozco	790	2,534	1,022	2,482	123	10
Orozco and Adal	69	56	1,002	6	1	0
Adal	1,753	494	19,973	41	744	1,689

Table 2. Counts of unidirectional DEGs (log-fold change 1, p-adj 0.05) identified post independent processing of both the Adal et al. 2021 and our RNA-seq samples into counts using Jack and identical differential expression analysis parameters via DESeq2.

Colloidal Silver Disruption of Interconnected Phytohormone Gene Network Signaling

Across all contrasts, genes directly and indirectly related to phytohormone signaling pathways: ethylene(ETH), abscisic (ABA), jasmonic (JA), auxin (AX), and gibberellin (GB) were differentially expressed. Cytochrome P450 704B1 (LOC115700892), involved in gibberellin biosynthesis, was consistently up-regulated in both true males (CN2) and IMs (CN3), supporting the role of gibberellins in promoting male reproductive development[50]. This aligns with studies in other plant species in which gibberellins have been shown to influence floral sex differentiation and promote the development of male reproductive structures[50]. MYB101 transcription factor (LOC115699028), which was up-regulated in true males but down-regulated in IMs, emerged as a key male transcription factor. MYB101 regulates pollen tube-synergid interactions during fertilization[51] and has been linked in *Arabidopsis thaliana* to ABA-mediated stress responses, such as stomatal closure and transcriptional regulation[52,53]. Its differential expression further underscores the disruption of phytohormone signaling in IMs caused by colloidal silver treatment.

Ethylene-responsive transcription factor ERF106 showed a complex pattern, being up-regulated in IMs (CN1, CN3) but down-regulated in true males (CN2) compared to females. This suggests that colloidal silver treatment may disrupt ethylene signaling, potentially contributing to the induction of male traits[54]. The differential expression of genes functionally linked to metal ion transport and homeostasis, such as myrcene synthase (LOC115716063) and metal-nicotianamine transporter YSL7 (LOC115701423) in the CNbetween true Male and true Female samples (CN2), further highlights the potential role of metal ions and metal ion homeostasis in secondary metabolism and sex differentiation in response to disruption of the interconnected phytohormone gene network crosstalk. Metal ions play essential roles in various biological processes, and their homeostasis is crucial for proper plant development and stress response[55].

The role of ethylene in sex determination is well-documented in various plant species, and its disruption can lead to alterations in sex expression[56]. The disruption of ethylene signaling by silver ions (Ag^+), as shown earlier (Figure 1), not only inhibits ethylene receptor activity but also appears to disrupt the regulation of key proteins associated with oxidative stress and lignin metabolism. For example, expression of the blue copper protein (LOC115699453), was up-regulated in true males but down-regulated in IMs, likely due to incomplete enzymatic activity of the ethylene receptors, being blocked by the applied silver ions[31]. IMs exhibited down-regulated expression of genes linked to auxin signaling and desiccation responses, such as desiccation-related protein PCC13-62-like (LOC115725418)[57], reflecting stress-adaptive transcriptional programs that respond to oxidative stress and hormonal imbalance. Furthermore, the down-regulation of ethylene receptors (LOC115721785) and ethylene-responsive transcription factors (LOC115723033) in IMs (CN3) reinforces the disruption of ethylene signaling pathway by colloidal silver. IAA-amino acid hydrolase ILR1-like 4 (LOC115707590), involved in auxin metabolism, was down-regulated in IMs and up-regulated in Females (CN3). Auxin has been reported as a key regulator of plant development, playing a role in sex differentiation[58,59] thus indicating a potential role for auxin in the prioritization of male embyrogenic tissue reprogramming in treated samples.

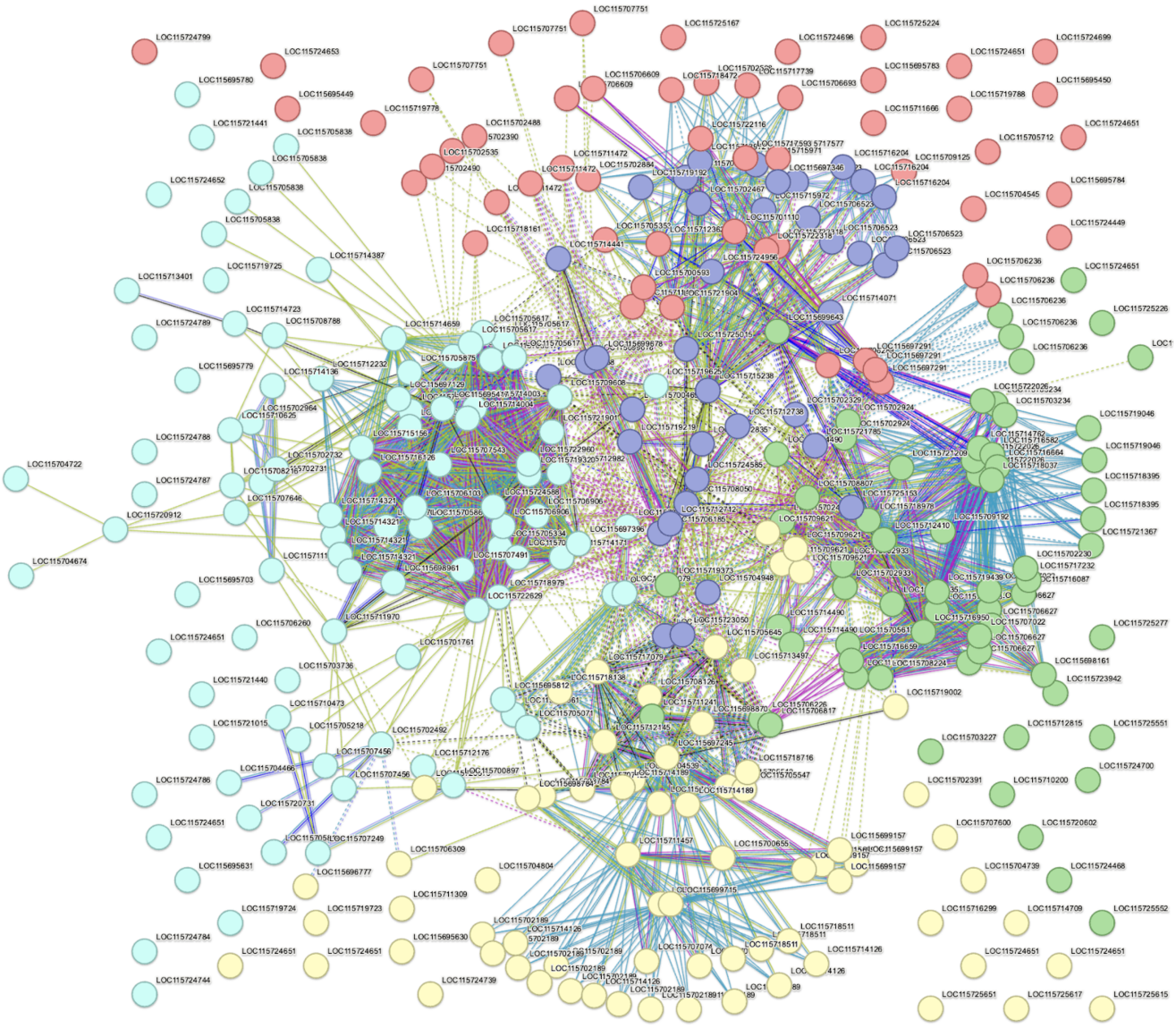

Figure 4. Protein-Protein Interaction Network between Five Major Phytohormone signaling pathways in *Cannabis sativa* (reference genome cs10). Green nodes: Abscisic acid signaling pathway, Yellow nodes: Jasmonic acid signaling pathway, Teal nodes: Auxin signaling pathway, Purple nodes: Ethylene/Brassinosteroid signaling pathway, Red nodes: Cytokinin signaling pathway. Edges are color-coded based on the type of interaction: Blue edges represent known interactions from curated databases, Magenta edges indicate experimentally determined interactions, Green edges signify predicted interactions from gene neighborhood, Red edges denote gene fusions, and Blue edges signify gene co-occurrence. Black edges represent co-expression, Yellow edges indicate text mining, and Light Blue edges denote protein homology. Dashed lines often represent inter-cluster connections, indicating relationships or interactions between different clusters (K-means 5 clustering applied).

Stress Response

The induction of stress response genes was a common theme across all contrasts, suggesting that both natural sex differentiation and colloidal silver-induced sex reversal involve stress-related pathways. Aldehyde oxidase GLOX-like (LOC115709228), involved in detoxifying reactive aldehydes and its Arabidopsis t. homolog is known to be regulated by MYB80 transcription factor in the development of tapetum and pollen[60], was up-regulated in both true males (CN2) and IMs (CN1, CN3), indicating a potential role for oxidative stress response in general plant male development. Oxidative stress is a common consequence of environmental perturbations and developmental transitions, and its management is crucial for maintaining cellular homeostasis[61]. Up-regulation of genes like Callose synthase 5 (LOC115714911) in IMs (CN3) and F-box/kelch-repeat protein (LOC115718742, At5g26960) in both IMs (CN1) and females (CN2) suggesting that various stress response pathways may be activated

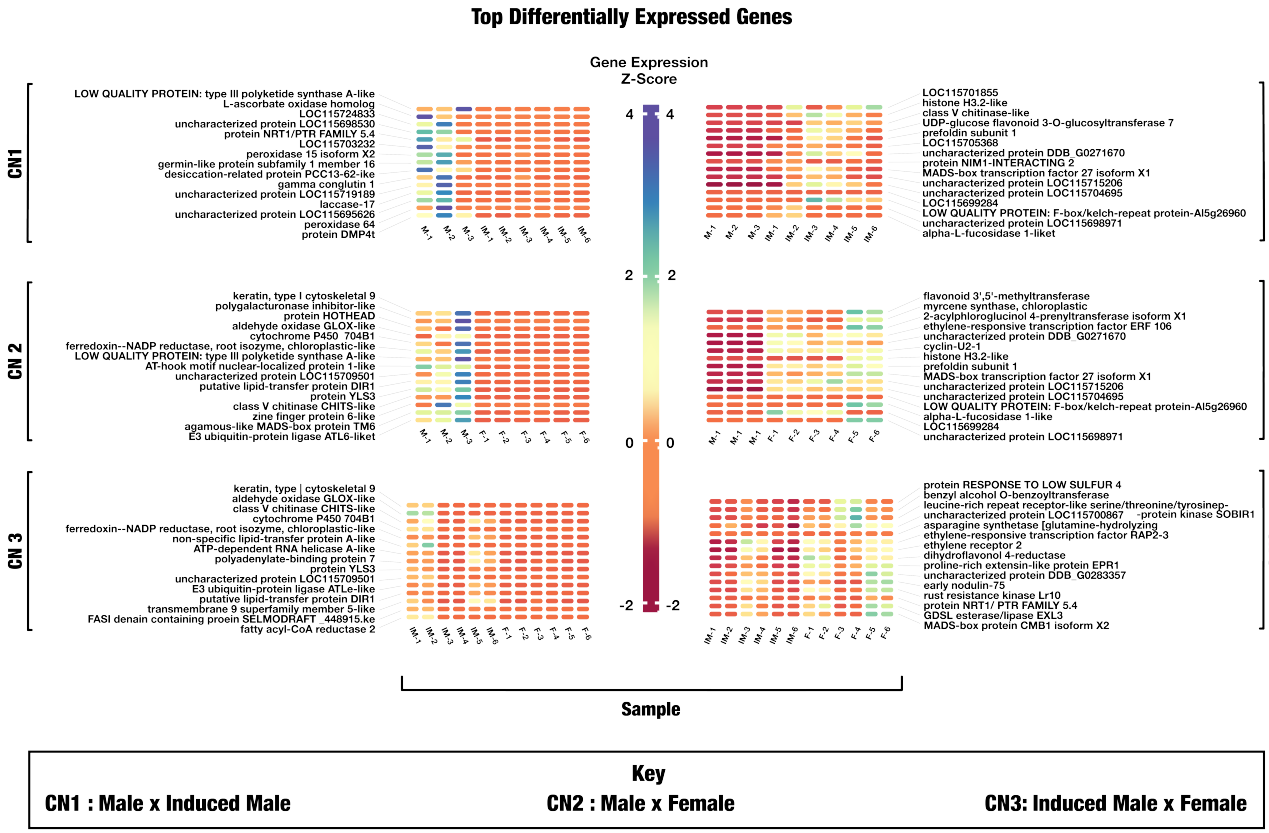

Figure 5. Heatmaps of differentially expressed genes across three contrasts Z-score normalized gene expression values are shown for the top differentially expressed genes (DEGs) identified in pairwise comparisons: CN1: Male vs. Induced Male, CN2: Male vs. Female, and CN3: Induced Male vs. Female. Protein annotations were derived using the Cs10G2P file and ontology inference. Substantial up-regulation of genes related to pollen development and phytohormone signaling in Induced Male samples suggests cross-talk between stress-response hormones. Expression levels in Female and Induced Male samples (CN3) indicate conservation of female expression patterns in Induced Male plants.

concurrently during sex differentiation and in response to colloidal silver treatment.

IMs, on the other hand, up-regulated genes such as F-box/kelch-repeat protein (LOC115718742, At5g26960), which modulates proteasomal activity under stress conditions and was down-regulated in True Males. This gene is also involved in protein ubiquitination, a critical process for proteostasis under stress conditions[62]. The up-regulation of stress-related genes in IMs likely reflects a stress-adaptive mechanism to regulate protein turnover during physiological adjustments induced by colloidal silver treatment or environmental signal cues. The up-regulation of stress-related genes in IMs likely reflects a stress-adaptive mechanism to regulate protein turnover during physiological adjustments induced by colloidal silver treatment or environmental signal cues.

Protein Ubiquitination, Folding, and Degradation

Genes involved in protein stability (ubiquitination) folding and degradation were differentially expressed, indicating a potential role for proteostasis in sex differentiation and response to colloidal silver. E3 ubiquitin-protein ligase ATL6-like (LOC115718105) was up-regulated in true males (CN2) and IMs (CN3), suggesting a role for protein degradation in male development. True males exhibited lower expression of ubiquitination-related genes, including F-box protein SKIP14 (LOC115699558), which is also associated with negative regulation of ABA signaling[63]. The ubiquitin-proteasome system is a key regulator of protein turnover and plays a crucial role in various cellular processes; hormone signaling, the regulation of chromatin structure and transcription, tailoring morphogenesis, responses to environmental challenges, self recognition and battling pathogens[64]. Chaperones assist in protein folding and assembly, ensuring proper protein function and cellular homeostasis[65]. Prefoldin subunit 1 (LOC115705604) of the pre-foldin chaperone complex, is involved in protein folding and was up-regulated in IMs (CN1) and down-regulated in True males (CN2). Chaperones assist in protein folding and assembly, ensuring proper protein function and cellular homeostasis[65] therefore indicating a potential role for chaperone complex related

proteins in the response to colloidal silver and sex differentiation in cannabis. These findings suggest that proteostasis is involved in the molecular mechanism driving sexual development, an interesting pairing with the high transcription factor modularity that has been sexual development.

Transcriptional Regulation

Transcription factors play a central role in orchestrating the complex gene expression changes associated with sex differentiation and the response to colloidal silver. Agamous-like MADS-box protein TM6 (LOC115714657) was up-regulated in true males (CN2), while MADS-box transcription factor 27 isoform X1 (LOC115704295) was up-regulated in IMs (CN1, CN3) and females (CN2) and down-regulated in true males (CN1,CN2). These findings complement the role that MADS-box transcription factors play in regulating distinct aspects of female floral development. MADS-box genes are well-known for their roles in floral organ identity and flowering time regulation, and their differential expression can considerably impact plant development[66]. Transcription factor bHLH91 (LOC115713370) was up-regulated in IMs (CN3), as were transcription factor ABORTED MICROSPORES (LOC115706375), and PHD finger protein MALE MEIOCYTE DEATH 1 (LOC115721367), a probable transcription factor required for chromosome organization and progression during male meiosis[67]. Both of these transcription factors were down-regulated in females (CN3) and up-regulated in IMs, highlighting their potential roles in male reproductive development and the response to colloidal silver in cannabis. Fittingly the high transcriptional re-configuration observed in our samples was purportedly used to shift our samples cell wall and lipid metabolism.

Cell Wall Modification and Lipid Metabolism

Changes in cell wall-related genes were observed across the contrasts, indicating that cell wall modifications may be important for both natural and induced sex differentiation. Polygalacturonase inhibitor-like (LOC115700758) was up-regulated in both true males (CN2) and IMs (CN3), suggesting a role for pectin modification in male reproductive development. Cell wall modifications are essential for plant growth and development, and alterations in cell wall composition can influence various physiological processes including plant morphogenesis[68,69]. The protein HOTHEAD (LOC115717125), involved in cuticle development, was up-regulated in true males (CN2), while alpha-L-fucosidase 1-like (LOC115709178), a lysosomal enzyme that breaks down fucose-containing glycoproteins and glycolipids[70,71], was down-regulated in true males (CN2) and up-regulated in females. These additional findings suggest that different cell wall modifications may be associated with male and female cannabis morphogenesis. Similarly, Laccase-17 (LOC115719496) and Peroxidase 64 (LOC115723402), are both implicated in lignin metabolism, and showed up-regulated expression in true males while being down-regulated in IMs. Laccase-17 and Peroxidase 64 are involved in lignin degradation, detoxification, and oxidative stress responses[72,73], suggesting that IMs may be actively modifying lignin development as part of the physiological shift post-treatment. This reduction in lignin-related gene expression in IMs highlights potential metabolic cost associated with transitioning from female to male reproductive structures. To infer additional functional components of masculinization with regard to these differentially expressed genes, we next performed a Random-Walk-with-Restart (RWR) analysis.

Random-Walk-with-Restart Inference Results

Using Nota, we conducted a personalized Random Walk with Restart (RWR) analysis on our bipartite-coupled monoplex networks, employing three sets of seeds derived from differentially expressed genes (DEGs) identified across three experimental contrasts. Our network inference identified additional candidate genes involved in the genetic architecture underlying masculinization in female cannabis treated with colloidal silver. The RWR algorithm calculates the geometric mean for each seed set, ensuring that the resulting scores and inferences sum to 1. Complete inference result tables are available in the supplementary material (add name). Here, we report the topologically related gene inferences for each contrast, excluding seed nodes and genes with a geometric mean score below our confidence thresholds (CN1: 157, CN2: 184, CN3: 236). Of these inferences, some overlapped when check agaisnt our DEG lists (CN1: 42, CN2: 77, CN3: 23). The presence of overlap of inferred genes and DEG's validated the RWR algorithm's ability to infer candidate genes which are seen experimentally to be involved in masculinization of female cannabis with our treatment. The majority of the other inferred genes were presently expressed in each CN but did not meet our criteria for differentially expressed. We generated four ontology enriched networks, one comprising the non-unique genes shared across all three inference sets and separate ontology reports for each CN inferences by selecting them as a local cluster network within our representative cannabis protein-protein interaction network available on StringDB[74].

Gene Network Ontology of Shared Non-Unique Inferences

Across all three inference sets, we identified 577 genes with confident geometric mean scores, of which 433 were non-unique and shared across multiple sets. Our shared network cluster consists of 433 nodes and 470 edges, with an average node degree of 2.17 and a local clustering coefficient of 0.287. The overall protein-protein interaction (PPI) enrichment was highly significant (p-value < 1.0e-16). To identify the most meaningful biological functions within this set of inferred genes, we refined the

network by retaining only edges between proteins with high-confidence interaction scores (0.9). We then applied k-means clustering to group functionally related proteins, ensuring that all disconnected sub-graphs were accounted for. This approach allowed us to highlight the most biologically relevant ontology profiles within the network (Figure 6).

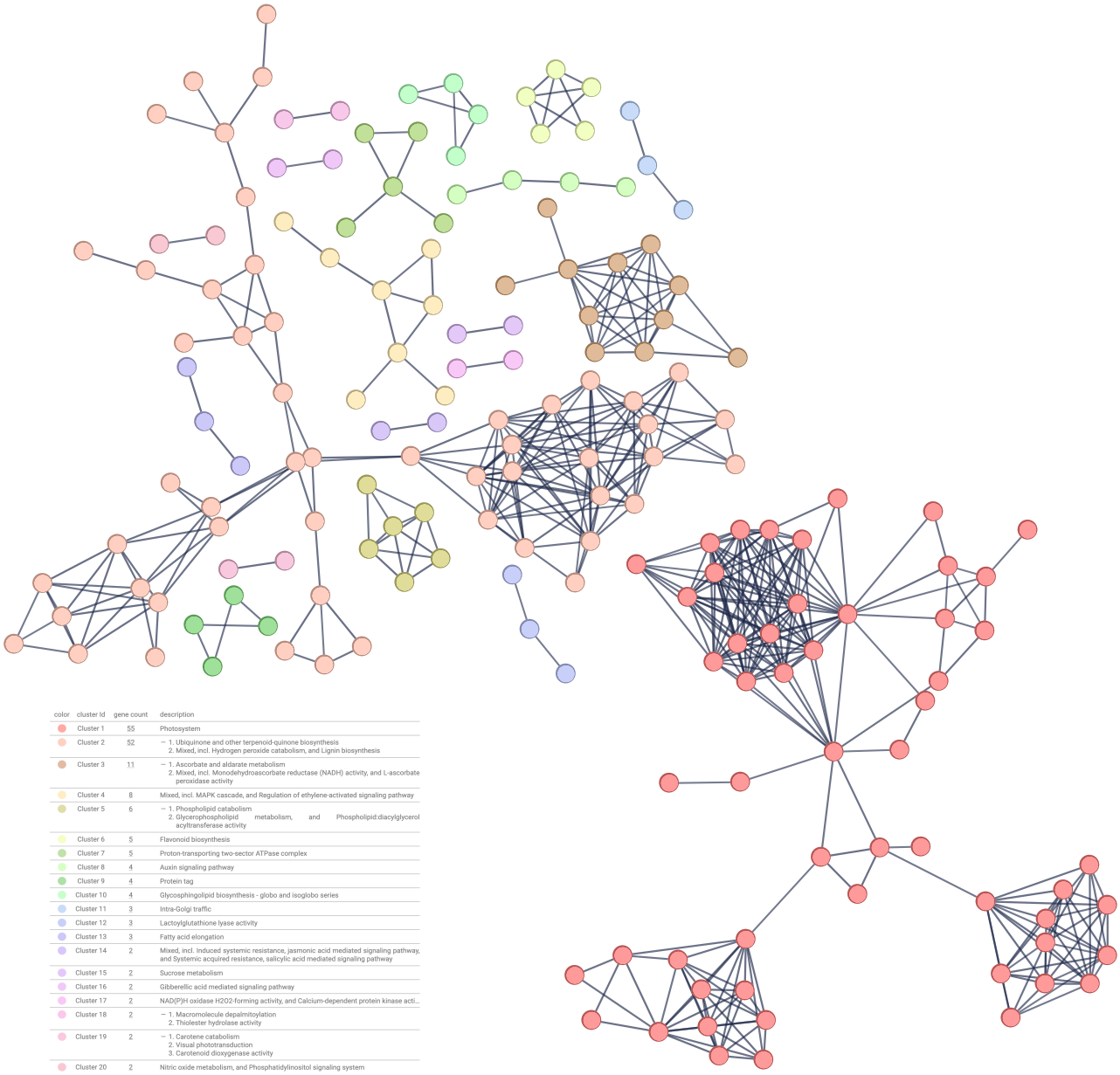

Figure 6. K-Mean 20 clustering of Shared Non-unique Inferences This network was made by highlighting all 577 inferred genes to to our species representative PPI. We selected proteins with high-confidence (>/=0.9) interactions shared between them, post this filtering we clustered the 433 non-unique genes. We hid all the non-connected nodes and removed node labels for visual clarity. Edge color opacity is correlated with interaction scores (darker correlates to higher scoring interactions).

This unique set of inferences exhibited gene ontology (GO) enrichment for biological processes related to photosynthesis, spliceosomal snRNP assembly, cellular component assembly involved in morphogenesis, and secondary metabolism. Additionally, we identified enrichment for molecular functions associated with ligase and oxidoreductase activity. Consistently, many enriched cellular component terms were linked to photosystem I and photosystem II, the mediator complex, the SMN-Sm protein, and spliceosome complexes. Several genes were associated with both primary metabolism (e.g., pyruvate, tyrosine, cysteine, and methionine metabolism) and secondary metabolic pathways, including ubiquinone and other terpenoid-quinone

biosynthesis, ascorbate and aldarate metabolism, pyruvate metabolism, and tyrosine metabolism (KEGG, cite). Furthermore, we observed enrichment in reactome pathways such as fatty acid metabolism and signaling by receptor tyrosine kinases.

Gene Network Ontology of CN One Inferences

The network consisted of; 157 genes, 224 edges, an average node degree of 2.85, an average local clustering coefficient of 0.324, and an overall PPI enrichment p-value of < 1.0e-16. These genes were inferred to be related to the genes of interest found in differential expression results from the comparison between true male and IM samples. Following the filter of proteins sharing high confidence interaction scores we ran k-means clustering (accounting for the minimum disconnected graphs in the network) and identified 10 clusters consisting of mixed functional grouping. We found that several of these genes clustered to primary metabolism associated genes (**phenylpropanoid metabolism, the remodeling of DAG and TAG as well as ubiquinone and other-terpenoid biosynthesis related genes**). Additionally we observed clusters of secondary metabolism related genes (**flavonoid, sucrose, ascorbate/aldarate metabolism, including redox-related enzymes**). Additionally several of these inferred genes clustered to terms related to photosynthesis and light-associated processes (**Photosystem-related, visual phototransduction, carotenoid dioxygenase activity**). Consistent with our DEG results for CNone, several of these inferences clustered to hormonal and stress-response pathways (**Induced systemic resistance, jasmonic acid and salicylic acid signaling pathways, Regulation of jasmonic acid signaling and ubiquitin-like conjugation**). Clusters 7 and 10 are related to energy metabolism and transcriptional regulation (**glycolysis, Transcription factors: MYC2, TIFY 10c**). These functional clusters are consistent with those identified in our our CN1 DEG results.

Gene Network Ontology of CN Two Inferences

This high-confidence interaction network consisted of; 184 genes, 198 edges, an average node degree of 2.15, an average local clustering coefficient of 0.334, and an overall PPI enrichment p-value of < 1.0e-16. These inferred genes were found to be topologically related to the differentially expressed genes identified in the CN of true male and true female samples. Following the same procedure for CN one inferences we identified 14 functional clusters (Table S3). Largely these inferences clustered to transcriptional and regulatory complex terms. As well as primary and secondary metabolism terms (**phenylpropanoid metabolism, including 4-coumarate-CoA ligase activity, acyl chain remodeling, and ubiquinone/terpenoid-quinone biosynthesis, vitamin E, flavonoid, and glycosphingolipid biosynthesis**). Consistent with the findings of the DEG results for CN2, CN2 inferences clustered into groups related to lipid and membrane dynamic terms (**Phospholipid catabolism, glycerophospholipid metabolism, phospholipid:diacylglycerol acyltransferase activity, Acyl-CoA biosynthesis, S-acyltransferase activity and Fatty acid elongation**). Three cluster groups (8,12,13) categorically align with energy metabolism and cellular regulation GO terms. We found that genes in cluster 7 were associated to protein modification and degradation. Both cluster 11 and 14 were associated with cellular trafficking and stress response. These categorization highlights are similar to those identified in the CNDEG results section. These inferences provide additional key differences between male and female cannabis at the metabolic, regulatory, and structural levels.

Gene Network Ontology of CN Three Inferences

This inferred gene network consisted of; 236 genes, 85 edges, an average node degree of 0.75, an average local clustering coefficient of 0.184, and an overall PPI enrichment p-value of 1.28e-11. These additional 236 inferred genes of interest are related to the differentially expressed genes identified in the CN of true female and IM samples. Here we identified 13 functional cluster (Table S4). In categorizing our clusters we found that genes belonging to clusters 1, 2 and 3 can be grouped into terms associated with DNA repair and transcriptional regulation (**DNA damage checkpoint signaling and non-homologous end-joining, mediator complex, Spliceosomal snRNP assembly and Sm-like protein family complex**). Consistent with the categories of the CN3 DEG results, several clusters (2,4,10,11,12) had genes which grouped to terms related to oxidative stress, and hormone-mediated-developmental signal transduction/integration (**Lactoylglutathione lyase activity, NAD(P)H oxidase HO-forming activity and calcium-dependent protein kinase activity, MAPK cascade, regulation of ethylene-activated signaling pathway, Gibberellic acid-mediated signaling pathway**). We had several clusters (5,6,8,9) which grouped to lipid metabolism and energy production (**Fatty acid degradation,glyceraldehyde-3-phosphate dehydrogenase (NAD+) (non-phosphorylating) activity, and formaldehyde metabolism, Fatty acid synthase activity, acyl-CoA biosynthesis, and the tricarboxylic acid (TCA) cycle enzyme complex, Glycosphingolipid biosynthesis, 4-coumarate-CoA ligase activity and acyl chain remodeling of DAG and TAG**). Clusters 7 and 13 both had genes which grouped to terms associated with protein modification and metabolic regulation (**Protein tag-related processes, Macromolecule depalmitoylation and thiolester hydrolase activity**). The categorization of these clusters continued to highlight the functional categories identified in the CN3 DEG results that highlighted key molecular processes differentiating IM cannabis tissue from true female cannabis tissue. These inferences provide additional insight into the regulatory, metabolic, and stress-response pathways involved in the masculinization process.

Discussion

Across all differential expression contrasts, the transcriptional landscape revealed notable differences in genes associated with phytohormones, metal ion homeostasis, lipid metabolism, cell wall remodeling, transcriptional regulation, stress responses, and both primary and secondary metabolism. CN1 (Male x IM) highlighted the disrupted reproductive function and stress adaptation in IMs, while CN2 (Male x Female) underscored the distinct hormonal and structural pathways differentiating male and female reproductive strategies. CN3 (IM x Female) revealed the transcriptional cost and incomplete physiological transition from female to IM phenotypes. The Random Walk with Restart (RWR) inference results across all contrasts identified functional categories consistent with our differential expression analysis. These findings support our hypothesis that several systems respond to environmental and hormonal signals to induce sexual plasticity of cannabis inflorescence tissue.

A key observation was the reprogramming of embyrogenic tissue development through long-range intracellular cascade signaling, primarily mediated by lipid metabolism. The introduction of high-affinity silver ions through colloidal silver treatment disrupted metal ion homeostasis and phytohormone network cross-signaling, triggering widespread transcriptional reprogramming and increased energy production. Notably, treated flowers exhibited gene expression changes related to lipid-protein transport, lipid metabolism, and cell wall development/modification. This physiological shift, coupled with hormonal imbalance, initiated a cascade of cross-cellular signaling, ultimately leading to extensive protein ubiquitination and proteolysis. Our findings further identified genes associated with cell wall morphogenesis and embyrogenic tissue homeostasis, suggesting that colloidal silver treatment perturbs plant hormone signal transduction through the disruption of metal ion equilibrium. This disruption likely activates multiple environmental response suites that drive cellular cross-talk. Collectively, these results provide insight into the genetic and physiological processes underlying sexual plasticity in *Cannabis sativa* and its modulation by environmental cues.

The overall mechanistic story emerging from our data suggests our treatment involves stress-induced signaling, which reprograms cellular processes, leading to the expression of genes involved in lipid metabolism, protein transport, and cell wall remodeling. These changes appear to trigger a cascade of intercellular communication events, with intriguing parallels to independently derived metazoan developmental pathways. In mammalian development, HH ligand signaling is known to mediate sex differentiation in embryonic development, our results suggest that *Cannabis sativa* employs a similar cross-cellular cascade molecular mechanism to generate non-genetically encoded reproductive structures.

Metal Ion Homeostasis

From our differential expression analysis we observed interesting expression profiles of metal ion related genes such as the blue copper protein (LOC115699453), probable metal-nicotianamine transporter YSL7 (LOC115701423) and myrcene synthase (LOC115716063) were up-regulated in our males but not in our treated female tissue samples. These metal homeostasis related gene expression profiles suggest that our treatment disrupted metal ion homeostasis in our treated samples. Within our CN one RWR inference set we observed 40 genes associated with metal ion binding ontology. Cytochrome P450 CYP73A100 (XP_030493605) was an inferred gene related to the findings of our CN1 DEG's, the cytochrome P450 enzyme family are known to play crucial roles in the biosynthesis of secondary metabolism, defense and response to environmental stressors[75]. This observation is interesting as it seems that metal ion related genes from both the differential expression analysis and the RWR analysis for the comparison of true males and true females against our IM samples have persisted. Our findings highlight the potential role of metal ions and metal ion homeostasis in secondary metabolism and sex differentiation in response to disruption of the interconnected phytohormone gene network crosstalk. The application of silver ions to plant tissue is known to disrupt the ethylene signaling pathway[32] which is likely due to transcriptomic compensation to satiate incomplete enzymatic activity of the ethylene receptors dependent on copper ion binding, being blocked by the applied silver ions.

Phytohormone Network Signaling Crosstalk

Here in our analysis we observed the disruption of the sexual development of treated female cannabis, by an initial change in the metal ion homeostasis which directly affects the ethylene biosynthesis signaling pathway. However it is known that the signaling pathways of phytohormone networks are highly interconnected suggesting that gene members are shared between these networks[33](Figure 4). We did identify several phytohormone related genes in our differential expression analysis from varying phytohormone signaling networks besides the ethylene signaling pathway. Taken together, the findings of each of our DEG contrasts demonstrate that up-regulation of stress-responsive, phytohormone-regulated, and developmental genes in IMs reflects an adaptive transcriptional landscape shaped by colloidal silver treatment. The interplay between the ethylene, GA, and JA signaling networks appears to mediate the genetic reprogramming that underlies sexual plasticity. In our inference set for CN1 we identified several proteins related to the Giberellic, Jasmonic and Brassinosteroid mediated signaling pathways (XP_030496910 DELLA protein GAIP-B, XP_030501877 transcription factor MYC2, XP_030488074 somatic embryo-genesis receptor kinase 2). These results elucidate the possible molecular mechanisms that cannabis employs during transcriptional regulation and phytohormone signaling to adaptively mitigate cellular stress and regulate reproductive organ development

under abiotic stress conditions. One possibility is that abiotic stress acts to modulate the genetic architecture of sexual plasticity, leading genetically XX Females to produce male flowers, perhaps as an evolutionary adaptation to ensure seed production, a major life history event in plants. However, we motion that further functional experimentation is required to validate the mediation of sex determination shared between major phytohormone gene networks.

Lipid Metabolism and Embryogenesis Development

A key finding from our study is the considerable up-regulation of cell wall and lipid metabolism/transport-related genes in male samples compared to Female and IM samples. In our differential expression results for CN2 we found two lipid transport proteins (LOC115718067 putative lipid-transfer protein DIR1, LOC115707560 non-specific lipid-transfer protein A-like) were up-regulated in male tissue samples while being down-regulated in female tissue samples. Additionally in our CN1 RWR inferences we identified a gene involved in phytohormone signaling but also clustered to di-terpenoid biosynthesis and cellular response to lipid metabolism (XP_030496910 DELLA protein GAIP-B) and protein XP_030487533 ABC transporter G family member 9 which is linked to ABC transporters in lipid homeostasis (MAP-1369062). Also protein (XP_030497001 REF/SRPP-like protein At1g67360) that is directly involved in lipid droplet formation, and clustered to proteins associated with the regulation of fertilization, seed dormancy process and pollen germination. These results support the idea that lipid-based signaling may be sex-specific in reproductive tissue differentiation.

Several of our CN1 inferences were ontologically associated to both primary cell wall biogenesis and embryo-genesis related terms. COBRA-like protein 6 (XP_030478181) was another inferred CN1 related protein clustered to proteins involved in Plant-type primary cell wall biogenesis. Universal stress protein PHOS34 (XP_030483164) and kiwellin-like (XP_030480936) clustered to anther development and spermatogenesis, respectively. Additionally, somatic embryo-genesis receptor kinase 2 protein (XP_030488074) clustered to receptor serine/threonine kinase binding and flower morphogenesis. We also observed proteins associated with with cellular component assembly involved in morphogenesis and pollen wall exine assembly in our CN2 set. Our CN3 inference set also had several proteins associated to spliceosomal snRNP assembly, Glycosphingolipid biosynthesis, Ubiquitin mediated proteolysis, MAPK signaling pathway in plants, further emphasizing the interplay between lipid metabolism, signal transduction and reproductive development.

Intriguingly, this cellular activity draws parallels with mechanisms observed in mammalian systems, where Hedgehog (Hh) ligands—Sonic Hedgehog (SHH), Indian Hedgehog (IHH), and Desert Hedgehog (DHH)—are pivotal for developmental signaling[76]. In mammals, secreted morphogens can bind to lipid rafts on the surface of secreting cells. These rafts, with their characteristic glycosphingolipids and serine/tyrosine protein receptor composition, facilitate signaling that influences development in nearby cells. Alternatively, they can be released via proteolysis or packaged into vesicles or lipoprotein particles to act on distant cells. Regardless of their distinct biological roles in mammals, all Hh ligands undergo proteolytic processing and lipid modification during their transit to the cell surface[77]. The similarities between mammalian lipid-based developmental signaling and the lipid-mediated processes identified in Cannabis sativa suggest that lipid rafts could play a crucial role in modulating sexual plasticity in embryo-genic tissue. Our findings support the hypothesis that lipid rafts act as mediators of intercellular communication, enabling long-range signaling cascades that facilitate the reprogramming of inflorescence tissue in response to environmental stimuli. These results highlight the molecular mechanisms which may be mediating sex-determining phytohormone signaling pathways modulating floral identity genes, however in the following section we relate our findings to the cannabis masculinization related literature more broadly.

Genetic Mechanisms Driving Sex Expression in Cannabis sativa

Interestingly, we did not observe differential expression in genes identified within the male-specific, non-recombining region of the Cannabis sativa Y chromosome[78] or other male-associated sequences[49,79]. This may be due to these genes being expressed at very low levels, not expressed in these tissues (particularly pseudogenes), or exhibiting close sequence similarity with paralogs on other chromosomes, making differential expression challenging to detect.

While sex determination in C. sativa is primarily controlled at the genetic level, hormonal signaling plays a critical role in modulating sex expression[20,21]. Adal et al. (2021) provides a broad overview of how phytohormones such as abscisic acid, auxin, cytokinin, ethylene, and gibberellin influence sex expression, which is consistent with our findings. However, our results suggest that these pathways are intricately linked to lipid and cell membrane metabolism, cell wall biogenesis, and embryogenic development, extending beyond phytohormone influence alone.

Building upon this, Monthony et al. (2024) specifically examined ethylene-related genes (CsERGs) in sex determination and sexual plasticity, leveraging transcriptomic data that includes IM flowers. Their study introduces the Karyotype Concordant (KC), Floral Organ Concordant (FOC), and Unique Ethylene-Related Gene (uERG) expression patterns, which classify CsERG expression across different reproductive phenotypes. Their results show that FOC expression remains consistent between individuals sharing floral organ sex but differing in sex chromosome karyotype, while uERGs display distinct expression across female, male, and IM flowers, suggesting a more complex regulatory role.

Our findings reinforce this framework but highlight a broader interplay of additional phytohormone-related genes in sex determination. Specifically, our results suggest that ethylene-inhibiting silver ion treatment acts as a catalyst for masculinization by disrupting metal ion homeostasis. This disruption appears to initiate a cascade of transcriptional and metabolic shifts, enabling embryogenic differentiation and the reprogramming of reproductive identity in cannabis flowers.

Further supporting this hypothesis, García-de Heer et al. (2024) suggest that hormone pathway manipulation can induce sex reversion, with monoecy in Cannabis sativa likely being linked to spatial variability in sex-influencing hormones, which act upstream to repress or promote floral identity genes[19]. This aligns with our findings, reinforcing the idea that sex expression is tightly linked to flowering and influenced by both genetic and hormonal factors. Although C. sativa has a chromosomal system (XX for females and XY for males), environmental conditions and hormone fluctuations can override genetic determinants, leading to the formation of male flowers on genetically female plants.

Marker-assisted breeding, SNP mapping, and QTL analysis are increasingly used to improve seed production and fiber quality in commercial hemp cultivation (Adal et al. 2021). Male-associated MADC2 PCR markers have been applied to distinguish male phenotypes from dioecious and monoecious females, providing a genetic framework for breeding programs. However, a key limitation of these approaches is their reliance on known molecular markers, which may not capture the full regulatory complexity of sex determination in non-model species.

Advancing Sex Expression Analysis with Nota

A key innovation in this study was the application of Nota for gene network annotation in C. sativa. Unlike previous studies that align C. sativa transcriptomic data to closely related model organisms (such as Arabidopsis thaliana in Monthony et al. 2024), we used a multilayer network inference approach. Our method integrated high-order interactions and functional annotations derived from species-representative proteomes via the STRING database, providing a more tailored and robust framework for gene annotation in non-model organisms. We suggest that readers search for all functional terms of interest with our Cannabis Species Representative STRING PPI.

Our findings suggest that Nota offers a flexible and species-adaptive alternative for functional annotation, particularly for systems with limited reference genomes or established molecular markers. By incorporating high-order interactions, Nota improves upon traditional annotation methods by capturing complex regulatory networks rather than relying solely on primary sequence homology. We encourage the transcriptomics community to further explore functional annotation through network-based approaches, which can enhance the resolution of biological inference in non-model species like C. sativa.

For further exploration of phytohormone network clustering, visit the STRINGDB synthetic proteomes linked in each respective contrast section within the results.

Jack and Nota

A key innovation in this study was the application of Nota, a platform designed to integrate network inference algorithmsincluding link prediction, node classification, and community detectionto predict and annotate candidate genes likely related to complex traits of interest. These network analyses incorporate various data types, including gene expression and protein-protein interaction (PPI) networks, to make predictions about individual gene function and the relatedness of complex trait architectures. This innovative platform allowed us to connect transcriptomic evidence with inferred relationships in the phytohormone signaling network, as demonstrated by the highly interconnected protein-protein interaction network of the reference cannabis genome (cs10) (Figure 4). This is achieved through Nota's use of multilayer network-based methods, enabling the identification of gene modules and community structures that relate to complex traits more efficiently.

Nota's ability to identify higher-order interactions between transcriptomic and proteomic datasets was pivotal in our study. By integrating these datasets within a multilayer framework, we were able to uncover the complex cross-talk mechanisms between different biological layers, providing a comprehensive view of the regulatory networks driving sexual plasticity in Cannabis sativa. This approach uncovered novel relationships, such as the clustering of stress-responsive and developmental genes, which may otherwise have been overlooked. Nota can incorporate various mono-plex (gene co-expression and or protein-protein interaction (PPI)) networks into a bipartite coupled multiplex gene expression and or protein-protein interaction (PPI) networks, to infer additional functional annotation of complex trait architectures.

For example, we visualized the gene co-expression layer using a filtered network representing a 100,000-edge sample of the 1.36 million-edge network (Figure 7), laid out with the ForceAtlas2 spring layout algorithm to reveal key community structures. Similarly, the protein layer was visualized using the sampled cannabis species representative PPI network, with a multi-scale backbone disparity filter applied to reduce the networks density and uncover its structure (Figure 4).

Nota enables users to efficiently analyze gene expression data through its integration of advanced bioinformatics pipelines and Jack a user-friendly Google Gemini powered bioinformatic analysis assistant. Traditional gene discovery workflows often require months of data analysis, model building, and validation. By leveraging network inference and bioinformatic tools, Nota considerably reduces the time researchers spend searching for genes associated with complex traits, without sacrificing the rigor of traditional genetic analysis. Nota empowers users to prioritize experimentally validated genes of interest, facilitating

a deeper and more precise annotation of gene network architectures for complex traits requiring only novice coding literacy. This accessibility is transformative for geneticists, molecular biologists, and plant and animal breeders who employ functional annotation tools which rely solely on pair-wise interactions for analyzing and annotating gene expression data and identifying candidate genes.

The goal of Jack is to combine the distributed data infrastructures within the Bioinformatics community to both increase researcher efficiency and data accessibility to the public. The infrastructure for data aggregation is present through multiple government institutions, however, is not presented in a way which is easy to access for multiple stakeholders. Further development is guided by principles to increase accessibility to public resources through data stewardship[80]. In addition, the functionalities of Jack support the long-term data preservation between multiple entities, to support researchers in better data practices overall and continue to support open-science data processes. Increasing accessibility in such a way is beneficial for multiple facets of the socio-economic sphere as found by[80,81]. Jack decreases the technical barriers to this open data access, thus decreasing the overall financial cost of researchers who do not reside in the global north.

Future Importance

Understanding the genetic and hormonal basis of sex determinism provides the foundation for future research to selectively control hermaphroditic tendencies in Cannabis through marker-assisted breeding systems or chemical applications to control gene expression. These findings could also lead to the development of genetic markers to identify plants with greater hermaphroditic tendencies early, mitigating the economic risks to cultivators.

Our findings also indicate that sexual plasticity is not related exclusively to sex linked chromosomes, but instead to a multifaceted collection of cellular processes instigated by abiotic stressors, conducted by phytohormone signal transduction, and executed by modulated protein development and interactions. Aspects of the inferences presented in this study are supportive of perspectives which are novel or yet to be validated. These findings provide a basis for further testing to verify the inferred protein interactions from our RWR results. While standards for network creation and interaction score thresholds are determined throughout our methods to ensure the most accurate inferences, we are conducting this study using network analysis techniques which can result in missing or incorrect predictions. Thus, exploration of the functionality of the inferred genes and experimentally verifying sequences and interactions is critical research to be continued. Our representative proteome is a demonstration of new methods for improved annotation and gene ontology within *Cannabis* without needing to develop new model organisms.

Overall, genetic analysis in *Cannabis sativa* is a multifaceted field that employs various techniques to dissect the genetic architecture of key traits. Despite challenges such as heterogeneous environments, varying genome assembly quality, and legal constraints, integrating genomic, transcriptomic, and inferred functional proteomic data, combined with Network inference, promises to improve our understanding of Cannabis biology and facilitate crop improvement.

Methods

Plant Growth and Treatment

The three crosses selected contain the same paternal parent Sour Strawberry, which had been crossed with Super Lemon Haze (SxSS), Flaming Cookies (FCxSS), and Golden Goat (GGxSS) (Table 1). These seeds were grown to produce four clones, one of which was pre-maturely flowered to ensure the sex of the plant was female, while the other three clones were grown to maturity under isolated and standardized growing conditions. The controlled conditions included water, and nutrient schedules, as well as temperature, humidity, and placement under the lights in the grow tent (Figure3). The plants were systematically rotated around the tent each week to receive equal light and physical data were recorded each week (Table 3). The plants were fed 5.5 grams of VegBloom (Hydroponic Research, San Diego, California) per gallon of water, with an average of 10 gallons of water used to water the 12 plants. To induce male flowers on these female plants, treatment branches on each plant were labeled and then sprayed daily with 50 PPM colloidal silver (Silver Mountain Minerals, West Jordan, Utah) beginning the first day of flower and continuing until mature male flowers were produced on every plant. Subjectively, there appeared to be a high similarity in the early growth of the plants; however, as the plants grew, more noticeable differences became prevalent.

Library Preparation and Sequencing

These samples were sent to the University of Colorados BioFrontier Sequencing Core for preparation of these libraries. Twelve libraries were prepared between the three selected individuals with two technical replicates of each control and treatment branch. BioFrontiers performed quality assessment using the BioAnalyzer/TapeStation and Qubit, mRNA isolation using Poly-A beads, and reverse-transcriptase (RT)-PCR to index the libraries for multiplex sequencing. These steps were performed using the NEXTflex Rapid Directional mRNA-Seq Kit Bundles (Bioo Scientific, Austin, Texas) according to standard protocol included in the kit. To sequence these libraries, Illumnias NextSeq Sequencer (2X150, 300-cycle, paired-end reads) produced approximately 130 million reads between the 12 samples with 10.8 million reads on average per library.

Differential Expression Analysis

RNA was extracted from treatment and control flowers and leaves at three time points: pre-flower, 7-days-in-flower, and 30-days-in-flower. The plant material was cut from the plant and directly stored in a ribonuclease (RNase) inhibitory solution to prevent degradation of RNA before extraction using the RNeasy mini kit (Qiagen, Redwood City, CA) following the manufacturers protocol using the RLC buffer. This protocol extracted total RNA libraries from the treatment and control branches of each plant, including ribosomal RNA (rRNA), transfer RNA (tRNA), messenger RNA (mRNA), and other small RNAs. These RNA samples were stored at -80řC in the extraction buffer until library preparation began to prevent degradation.

Raw paired-end RNA-seq libraries were generated for 15 samples total: 6 female flowers, 6 IM flowers, and 3 male flower samples. Jack, Nota's network inference assistive pipeline, generated count files and aided in the construction of gene co-expression, protein-protein interaction and bipartite networks. Here Jack was used for the processing of our paired RNA-seq sample files. Paired FASTQ files were interleaved into individual FASTQ files per sample and aligned to the cs10 reference genome index generated using the STAR: ultrafast universal RNA-seq aligner software[82]. The average percentage of uniquely mapped reads of our samples was calculated to be 73 percent with a mapping range of 60.8 - 78.1 percent. Post generating bam files for each sample via STAR, the bam files were processed into a gene expression matrix using featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features[83].

Due to the complexity and user-specificity required for differential gene expression analysis, and a plethora of tools and resources available[84–86], we maintained this statistical component of transcriptomic analysis independent from Nota. Our gene count sample profiles were analyzed using the DESeq2[84] R package. Raw count data and associated sample information were imported from CSV files. The DESeq2 dataset was constructed by integrating count data with sample metadata, specifying the experimental design with the sex 'ConditionType' as a variable of interest. Prior to differential expression analysis, genes exhibiting low expression (fewer than one count across all samples) were filtered out to remove non-expressed genes from our statistical inference. Differential expression analysis was conducted using the DESeq2 function, which fits a negative binomial model to our gene counts matrix. The model was adjusted to account for differences in condition types by redesigning the experimental setup within the analysis framework.

Statistical significance of differentially expressed genes was determined using an adjusted p-value (alpha) less than 0.05 with an absolute log2 fold change thresholds greater than 1 to focus on all expressed genes without bias towards up or down-regulated genes. Results were extracted and tabulated including log2 fold changes, p-values, and adjusted p-values. Normalized gene expression data were computed using DESeq2 'normalization' function for subsequent analyses and visualization purposes.

Expression data between the Orozco and Adal samples were filtered again using the same thresholds for adjusted p value and log2 fold change, then merged based on Locus IDs present in both sample sets for each contrast, respectively. Directionality of regulation was then determined based on the sign of the Log 2 Fold Change value for each gene and appended to the data frame. Using pandas, each CN was compared between the two datasets and genes that were differentially expressed in the same direction were conserved to assess similarity between experiments (Table 2).

Heat maps were generated to visualize the z-score transformed expression data of each sample to illustrate patterns of expression across different conditions (Figure 3). Prior to this, these data containing Gene IDs, protein names, and regulation direction, for each CN were then read into R and formatted for calculation of Z score. contain the protein name as the row title and the expression value for each row was appended to the data frame and normalized using the R package dplyr[87]. Consistency of order, Z score range, and CN specific samples were ensured within the script using dedicated functions. Heat maps were created using the package ComplexHeatmap[88] with visual specifics, formatting, and color pallets produced by circlize[89]. The protein names displayed in heat maps were collected from the primary basic annotation list by parsing the protein name associated with each Locus ID.

Network Construction

The multilayer network we curated consists of two interlayer networks (Layer A and Layer B) of equal node cardinality but different edge structures. Layer A represents a filtered gene co-expression network constructed from RNA-seq reads as specified below. Subsequently, Layer B is the curated species representative protein-protein interaction network generated using STRINGDB as described below. Layers A and B are interconnected through a bipartite layer, where the bipartite nodes correspond to the union of nodes in Layer A and Layer B. In our network, the bipartite layer is representative of the previously identified correlation of encoding genes to translated proteins. The edges in the bipartite layer link nodes in Layer A to their associated counterparts in Layer B, facilitating inter-layer connectivity and facilitating the basis for further prediction.

Gene Co-Expression Network

Due to our small sample size, we incorporated 15 samples from the Adal et al. 2021 cannabis masculinization transcriptomic analysis[20]. We combined and processed all RNA-seq reads together using Jack and generated a multi-experiment gene count matrix. Post-hoc of our differential expression analysis with our combined datasets, we constructed a co-expression network using 30 RNA-seq samples. Prior to the creation of the co-expression network, we removed all zero sum rows and normalized

the gene count matrix. The co-expression network was constructed using the Jack function which calculates the Pearson correlation coefficient between all pairs of genes in a gene count matrix. The Pearson correlation function measures the linear relationship between two variables (in this case, gene expression levels across samples) and is given by:

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ik}-\bar{x_i})(x_{jk}-\bar{x_j})}{\sqrt{\sum_{k=1}^{n}(x_{ik}-\bar{x_i})^2 \sum_{k=1}^{n}(x_{jk}-\bar{x_j})^2}}$$

where r_{ij} is the Pearson correlation coefficient between gene i and gene j, x_{ik} is the expression level of gene i in sample k, $\bar{x_i}$ is the mean expression level of gene i across all samples, and n is the number of samples. This function returns a value between -1 and 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

Given the size of the correlation matrix (generated from 30 samples and thousands of genes), Jack utilized multi-threading with the `ThreadPoolExecutor` module in Python. This enabled the processing of different gene-pair correlation calculations in parallel, considerably reducing the overall computation time by distributing the workload across multiple threads. Each thread processed a subset of gene pairs and identified significant correlations independently, which were then aggregated into the final co-expression network.

Using Jack, the network was constructed as an undirected graph using the NetworkX library[44], where nodes represent individual genes and edges represent significant co-expression relationships between gene pairs. Isolated nodes, which did not show significant co-expression with any other gene, were removed to improve the clarity and focus of the network.

Once the correlation matrix was calculated, Jack applied a stringent filtering criterion by retaining only gene pairs with an absolute correlation value ($|r_{ij}|$) greater than or equal to 0.5, indicating a moderate to strong co-expression relationship. Gene pairs below this threshold were discarded to focus on biologically meaningful interactions. To fit the parameters which allows Nota to employ network inference across non-homogeneous mono and multiplex networks we built a function within Jack that allows users to convert correlation matrices into distance matrices with a positive weight range.

Additionally, we used Nota's graph filtration function, which filters a weighted similarity matrix to generate an adjacency matrix, allowing for the extraction of significant connections in a graph. It applies a topological filtering method based on a paper by De Vico Fallani et al. 2017[90,91]. Summary network statistics, (nodes and edges pre and post filtering), were computed and stored as a tabulated document. The final network was saved in Graph Modeling Language (GML) format for subsequent analysis and visualization (Figure 8 and Figure 5). We employed additional modularity-based community detection algorithms to visualize these large networks that assign nodes to communities based on link structure[40,92].

Species Representative Protein-Protein Interaction Network

Using Nota and Jack, we curated and filtered a protein-protein interaction (PPI) network for *Cannabis sativa* generated using STRINGDB, a widely-used protein annotation tool and interaction network generator[74]. To ensure the accuracy of the network, we required a comprehensive but non-redundant proteome. We used Jack to construct this synthetic representative proteome from publicly available data and tools, using the Y chromosome from the BCMb Cannabis line (Salk Institute) and the Pink Pepper genome (RefSeq Accession GCF029168945.1) as reference genomes[93]. All available *Cannabis sativa* protein sequences were downloaded from NCBI using Jack's Gemini API functionality, excluding those submitted solely on the basis of in silico predictions from genome sequences.

To further refine the data, Jack reduced sequence redundancy by employing a two-stage alignment protocol using `miniprot`[93]. In the first stage, protein sequences were aligned to the reference genome, and overlapping alignments were parsed to retain a single representative protein sequence for each locus based on the highest chain score. In the second stage, Jack assigned representative multi-copy protein sequences to a single genomic locus using the highest chain score. These steps yielded a 1-to-1 representative proteome for *Cannabis sativa*, which was used to generate the initial PPI network Cannabis Species Representative STRING PPI. Here again we applied the correlation to distance matrix conversion function in Jack to maintain all the interaction scores.

To curate and filter the PPI network, we utilized the attribute of Nota that processes the PPI data generated by STRINGDB which incorporates preliminary filtering. Nota then converted the PPI network using tools found in NetworkX, where nodes represent proteins and edges represent protein interactions, into a GML graph object. After filtering, isolated nodes (proteins without significant interactions), reversed self-loop interactions (B-to-A) were removed by Nota to decrease the density of the network. The final network was exported in GML format for visualization, as well as a TSV edge list for further analysis. Summary statistics, including the number of nodes and edges in the filtered network, were saved to a TSV file (Table 3.).

Bipartite Coupled Multilayer Network

By utilizing Nota to symmetrically connect the PPI network and the Co-expression network into a multilayer network, each node represents a gene or the derived proteins (using a Locus ID to protein ID key for cannabis). The gene and protein layers of

the multilayer network are connected by a bipartite layer; each protein is connected to its encoding gene. The bottom layer displays the gene co-expression network, including only the largest connected component, extracted with `graph-tool`[45] (Figure 3). Nodes are positioned using the ForceAtlas2 layout in Gephi, and plotted as layers using Matplotlib and NetworkX with code adapted from Klein[44,46,47]. Spring force layout is determined using the topology of the gene co-expression network, with the same positions repeated for the protein-protein interaction network.

CN(CN)	Number of Nodes	Number of Edges
Co-Expression	22,658	30,455,297
Co-Expression Alpha 0.3	22,630	3,431,798
Filtered Co-Expression	12,873	1,360,160
PPI	23,704	3,285,936
Filtered PPI	12,873	1,275,017
Bipartite Network	25,746	12,873

Table 3. Total number of nodes and edges in the co-expression and PPI networks before and after filtration. Co-expression is entirely unfiltered, Co-expression Alpha 0.3 uses filtration factor alpha of 0.3 for preliminary filtration. Filtered Co-expression and Filtered PPI truncates nodes and corresponding edges which are not present in both the co-expression and PPI networks to create the Bipartite Network. Nodes and edges in bipartite network included for posterity. Compute significance scores (alpha) for weighted edges in a NetworkX graph as defined in Serrano et al. 2009, The resulting graph contains only edges that survived from the filtering with the alpha_t threshold. Nota performs a cut of the graph previously filtered through a proprietary function. PPI network was only filtered to match the cs10G2P 1-to-1 map.

Network Ontology Transcript Annotation: Random-Walk-With-Restart

Random-walk-with-restart is a powerful approach for exploring the topology of networks. By simulating the movement of a walker randomly traversing nodes through edges in a network, random walks are capable of capturing several structural properties of networks. In Nota's RWR, the random walker, at each step, can navigate from one node to one of its neighbors or restart its walk from a node randomly sampled from a set of seed nodes. Nota, by enabling restart from one or several seed nodes, simulates a diffusion process in which the objective is to determine the steady state of an initial probability distribution. This steady state represents a measure of proximity between the seed(s) and all the network nodes. Here we used the top 60 differentially expressed genes in each experimental CN and filtered these 180 genes to keep those which existed in the symmetric bipartite-coupled multilayer network as contrast-specific seed nodes (CN1: 30, CN2: 36, CN3:37). Using RWR, Nota was able to predict genes which were related to the seed nodes through high order interactions. The inferred genes were then scores and mapped for further functional analysis.

Jack

Annotated gene expression data were computed using the tabulated results from differential expression analysis for each contrast, filtered to retain significant data with adjusted p-values 0.05 and log2 fold change values > 1. Genes with the top 30 and bottom 30 log2 fold change values were designated as up-regulated and down-regulated, respectively. After this filtration, each Locus ID was cross-referenced with known cannabis gene and protein names for primary basic annotation, using annotation provided in the cs10 reference genome and using Jack's Gemini API functionality, we were able to generate a reliable list of available scientific journal references along with additional primary basic annotation for our cannabis transcripts and their respective encoded proteins. When a protein name was unavailable, we applied RWR, which predicted functional annotation by assigning a most-similar protein to an uncharacterized protein(s), and through sequential clustering to the Cannabis Species Representative STRING PPI now publicly available on STRINGDB. We were able to assign function confidently and, when available, ontology enrichment associated to its respective cluster.

References

1. Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303, DOI: 10.3732/ajb.0900346 (2010).
2. Li, H.-L. An archaeological and historical account of cannabis in China. *Econ. Bot.* **28**, 437–448, DOI: 10.1007/BF02862859 (1973).
3. Russo, E. B. History of cannabis and its preparations in saga, science, and sobriquet. *Chem. & Biodivers.* **4**, 1614–1648, DOI: 10.1002/cbdv.200790144 (2007).

4. Elsohly, M. A. & Slade, D. Chemical constituents of marijuana: the complex mixture of natural cannabinoids. *Life Sci.* **78**, 539–548, DOI: 10.1016/j.lfs.2005.09.011 (2005).
5. Russo, E. B. & McPartland, J. M. Cannabis is more than simply delta(9)-tetrahydrocannabinol. *Psychopharmacology* **165**, 431–432; author reply 433–434, DOI: 10.1007/s00213-002-1348-z (2003).
6. Ahmed, A. T. M. F., Islam, M. Z., Mahmud, M. S., Sarker, M. E. & Islam, M. R. Hemp as a potential raw material toward a sustainable world: A review. *Heliyon* **8**, e08753, DOI: 10.1016/j.heliyon.2022.e08753 (2022).
7. Burton, R. A., Andres, M., Cole, M., Cowley, J. M. & Augustin, M. A. Industrial hemp seed: from the field to value-added food ingredients. *J. Cannabis Res.* **4**, 45, DOI: 10.1186/s42238-022-00156-7 (2022).
8. van Bakel, H. *et al.* The draft genome and transcriptome of Cannabis sativa. *Genome Biol.* **12**, R102, DOI: 10.1186/gb-2011-12-10-r102 (2011).
9. Peil, A., Flachowsky, H., Schumann, E. & Weber, W. E. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (Cannabis sativa L.). *TAG. Theor. applied genetics. Theor. und angewandte Genet.* **107**, 102–109, DOI: 10.1007/s00122-003-1212-5 (2003).
10. Hillig, K. W. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genet. Resour. Crop. Evol.* **52**, 161–180, DOI: 10.1007/s10722-003-4452-y (2005).
11. Tanney, C. A. S., Backer, R., Geitmann, A. & Smith, D. L. Cannabis Glandular Trichomes: A Cellular Metabolite Factory. *Front. Plant Sci.* **12** (2021).
12. Gertsch, J. *et al.* Beta-caryophyllene is a dietary cannabinoid. *Proc. Natl. Acad. Sci. United States Am.* **105**, 9099–9104, DOI: 10.1073/pnas.0803601105 (2008).
13. Russo, E. B. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br. J. Pharmacol.* **163**, 1344–1364, DOI: 10.1111/j.1476-5381.2011.01238.x (2011).
14. Swift, W., Wong, A., Li, K. M., Arnold, J. C. & McGregor, I. S. Analysis of cannabis seizures in NSW, Australia: cannabis potency and cannabinoid profile. *PloS One* **8**, e70052, DOI: 10.1371/journal.pone.0070052 (2013).
15. Volkow, N. D., Baler, R. D., Compton, W. M. & Weiss, S. R. B. Adverse health effects of marijuana use. *The New Engl. J. Medicine* **370**, 2219–2227, DOI: 10.1056/NEJMra1402309 (2014).
16. Hart, C. L., van Gorp, W., Haney, M., Foltin, R. W. & Fischman, M. W. Effects of acute smoked marijuana on complex cognitive performance. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **25**, 757–765, DOI: 10.1016/S0893-133X(01)00273-1 (2001).
17. Ruzgas, R., Tilvikien, V., Barauskait, K., Viril, A. & ydelis, R. Ethephons effects on sexual expression and cannabinoid production in monoecious and dioecious varieties of hemp (*Cannabis sativa* L.): A field trial. *Ind. Crop. Prod.* **210**, 118064, DOI: 10.1016/j.indcrop.2024.118064 (2024).
18. Schafroth, M. A. *et al.* 9-cis-Tetrahydrocannabinol: Natural Occurrence, Chirality, and Pharmacology. *J. Nat. Prod.* **84**, 2502–2510, DOI: 10.1021/acs.jnatprod.1c00513 (2021).
19. Garcia-de Heer, L., Mieog, J., Burn, A. & Kretzschmar, T. Why not XY? Male monoecious sexual phenotypes challenge the female monoecious paradigm in Cannabis sativa L. *Front. Plant Sci.* **15**, 1412079, DOI: 10.3389/fpls.2024.1412079 (2024).
20. Adal, A. M., Doshi, K., Holbrook, L. & Mahmoud, S. S. Comparative RNA-Seq analysis reveals genes associated with masculinization in female Cannabis sativa. *Planta* **253**, 17, DOI: 10.1007/s00425-020-03522-y (2021).
21. Monthony, A. S., de Ronne, M. & Torkamaneh, D. Exploring ethylene-related genes in Cannabis sativa: implications for sexual plasticity. *Plant Reproduction* DOI: 10.1007/s00497-023-00492-5 (2024).
22. Ram, H. Y. M. & Jaiswal, V. S. Induction of female flowers on male plants ofCannabis Sativa L. by 2-chloroethanephosphonic acid. *Experientia* **26**, 214–216, DOI: 10.1007/BF01895593 (1970).
23. Moliterni, V. M. C., Cattivelli, L., Ranalli, P. & Mandolino, G. The sexual differentiation of Cannabis sativa L.: A morphological and molecular study. *Euphytica* **140**, 95–106, DOI: 10.1007/s10681-004-4758-7 (2004).
24. Truta, E., S., O., Surdu, S., Zamfirache, M. & OPRICA, L. Some aspects of sex determinism in hemp. *Analele Stiintifice ale Univ. Al. I. Cuza din Iasi. Genet. si Biol. Mol.* **VIII**, 31–39 (2007).
25. Vergara, D. *et al.* New York State Cannabis sativa L. Production Manual. (2023).
26. Mukherjee, A. *et al.* The bioactive potential of phytohormones: A review. *Biotechnol. Reports* **35**, e00748, DOI: 10.1016/j.btre.2022.e00748 (2022).

27. Grant, S. *et al.* Genetics of sex determination in flowering plants. *Dev. Genet.* **15**, 214–230, DOI: 10.1002/dvg.1020150304 (1994). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/dvg.1020150304.

28. Kumar, V., Giridhar, P. & Gokare, R. AgNO3 - A potential regulator of ethylene activity and plant growth modulator. *Electron. J. Biotechnol. (ISSN: 0717-3458) Vol 12 Num 2* **12**, DOI: 10.2225/vol12-issue2-fulltext-1 (2009).

29. Ma, Q. & Dong, C.-H. Regulatory functions and molecular mechanisms of ethylene receptors and receptor-associated proteins in higher plants. *Plant Growth Regul.* **93**, 39–52, DOI: 10.1007/s10725-020-00674-5 (2021).

30. Moon, Y.-H. *et al.* Effect of Timing of Ethephon Treatment on the Formation of Female Flowers and Seeds from Male Plant of Hemp (Cannabis sativa L.). *Korean J. Plant Resour.* **33**, 682–688, DOI: 10.7732/KJPR.2020.33.6.682 (2020).

31. Veen, H. Silver thiosulphate: An experimental tool in plant science. *Sci. Hortic.* **20**, 211–224, DOI: 10.1016/0304-4238(83)90001-8 (1983).

32. Mohan Ram, H. Y. & Sett, R. Induction of fertile male flowers in genetically female Cannabis sativa plants by silver nitrate and silver thiosulphate anionic complex. *Theor. Appl. Genet.* **62**, 369–375, DOI: 10.1007/BF00275107 (1982).

33. Yin, L. *et al.* Transcription Factor Dynamics in Cross-Regulation of Plant Hormone Signaling Pathways. *bioRxiv* 2023.03.07.531630, DOI: 10.1101/2023.03.07.531630 (2023).

34. Zhu, L. *et al.* Drought Stress-Related Gene Identification in Rice by Random Walk with Restart on Multiplex Biological Networks. *Agriculture* **13**, 53, DOI: 10.3390/agriculture13010053 (2023). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

35. Valdeolivas, A. *et al.* Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* **35**, 497–505, DOI: 10.1093/bioinformatics/bty637 (2019).

36. Baptista, A., Brière, G. & Baudot, A. Random walk with restart on multilayer networks: from node prioritisation to supervised link prediction and beyond. *BMC Bioinforma.* **25**, 70, DOI: 10.1186/s12859-024-05683-z (2024).

37. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diffusion on networks. *Phys. Reports* **716-717**, 1–58, DOI: 10.1016/j.physrep.2017.07.007 (2017). ArXiv:1612.03281 [cond-mat, physics:physics].

38. Costa, L. d. F. & Travieso, G. Exploring complex networks through random walks. *Phys. Rev. E* **75**, 016102, DOI: 10.1103/PhysRevE.75.016102 (2007). Publisher: American Physical Society.

39. K, M., T, C. & Ak, S. RRW: repeated random walks on genome-scale protein networks for local cluster discovery. *BMC bioinformatics* **10**, DOI: 10.1186/1471-2105-10-283 (2009). Publisher: BMC Bioinformatics.

40. Newman, M. E. J. A measure of betweenness centrality based on random walks. *Soc. Networks* **27**, 39–54, DOI: 10.1016/j.socnet.2004.11.009 (2005). Place: Netherlands Publisher: Elsevier Science.

41. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Syst.* **30**, 107–117, DOI: 10.1016/S0169-7552(98)00110-X (1998).

42. Pan, J.-Y., Yang, H.-J., Faloutsos, C. & Duygulu, P. Automatic multimedia cross-modal correlation discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, 653–658, DOI: 10.1145/1014052.1014135 (Association for Computing Machinery, New York, NY, USA, 2004).

43. Gómez, S. *et al.* Diffusion Dynamics on Multiplex Networks. *Phys. Rev. Lett.* **110**, 028701, DOI: 10.1103/PhysRevLett.110.028701 (2013). Publisher: American Physical Society.

44. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. *scipy* DOI: 10.25080/TCWV9851 (2008).

45. Peixoto, T. P. The graph-tool python library (2014).

46. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. & Eng.* **9**, 90–95, DOI: 10.1109/MCSE.2007.55 (2007). Conference Name: Computing in Science & Engineering.

47. Klein, B. jkbren/matplotlib-multilayer-network (2025). Original-date: 2020-07-30T12:20:22Z.

48. Baptista, A., Gonzalez, A. & Baudot, A. Universal multilayer network exploration by random walk with restart. *Commun. Phys.* **5**, 1–9, DOI: 10.1038/s42005-022-00937-9 (2022). Publisher: Nature Publishing Group.

49. Prentout, D. *et al.* An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of Cannabis sativa. *Genome Res.* **30**, 164–172, DOI: 10.1101/gr.251207.119 (2020).

50. Plackett, A. R. *et al.* Analysis of the Developmental Roles of the Arabidopsis Gibberellin 20-Oxidases Demonstrates That GA20ox1, -2, and -3 Are the Dominant Paralogs[C][W]. *The Plant Cell* **24**, 941–960, DOI: 10.1105/tpc.111.095109 (2012).

51. Liang, Y. *et al.* MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in Arabidopsis thaliana fertilization. *PLoS genetics* **9**, e1003933, DOI: 10.1371/journal.pgen.1003933 (2013).

52. Hugouvieux, V., Kwak, J. M. & Schroeder, J. I. An mRNA Cap Binding Protein, ABH1, Modulates Early Abscisic Acid Signal Transduction in Arabidopsis. *Cell* **106**, 477–487, DOI: 10.1016/S0092-8674(01)00460-3 (2001). Publisher: Elsevier.

53. Hugouvieux, V. *et al.* Localization, Ion Channel Regulation, and Genetic Interactions during Abscisic Acid Signaling of the Nuclear mRNA Cap-Binding Protein, ABH1. *Plant Physiol.* **130**, 1276–1287, DOI: 10.1104/pp.009480 (2002).

54. Müller, M. & Munné-Bosch, S. Ethylene Response Factors: A Key Regulatory Hub in Hormone and Stress Signaling. *Plant Physiol.* **169**, 32–41, DOI: 10.1104/pp.15.00677 (2015).

55. Clemens, S. Molecular mechanisms of plant metal tolerance and homeostasis. *Planta* **212**, 475–486, DOI: 10.1007/s004250000458 (2001).

56. Chailakhyan, M. K. & Timiriazev, K. A. GENETIC AND HORMONAL REGULATION OF GROWTH, FLOWERING, AND SEX EXPRESSION IN PLANTS. *Am. J. Bot.* **66**, 717–736, DOI: 10.1002/j.1537-2197.1979.tb06276.x (1979). Publisher: Wiley.

57. Piatkowski, D., Schneider, K., Salamini, F. & Bartels, D. Characterization of Five Abscisic Acid-Responsive cDNA Clones Isolated from the Desiccation-Tolerant Plant Craterostigma plantagineum and Their Relationship to Other Water-Stress Genes. *Plant Physiol.* **94**, 1682–1688, DOI: 10.1104/pp.94.4.1682 (1990).

58. Brumos, J. *et al.* Local Auxin Biosynthesis Is a Key Regulator of Plant Development. *Dev. Cell* **47**, 306–318.e5, DOI: 10.1016/j.devcel.2018.09.022 (2018).

59. Sundberg, E. & Østergaard, L. Distinct and Dynamic Auxin Activities During Reproductive Development. *Cold Spring Harb. Perspectives Biol.* **1**, a001628, DOI: 10.1101/cshperspect.a001628 (2009).

60. Phan, H. A., Iacuone, S., Li, S. F. & Parish, R. W. The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in Arabidopsis thaliana. *The Plant Cell* **23**, 2209–2224, DOI: 10.1105/tpc.110.082651 (2011).

61. Mittler, R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* **7**, 405–410, DOI: 10.1016/s1360-1385(02)02312-9 (2002).

62. Kuroda, H., Yanagawa, Y., Takahashi, N., Horii, Y. & Matsui, M. A Comprehensive Analysis of Interaction and Localization of Arabidopsis SKP1-LIKE (ASK) and F-Box (FBX) Proteins. *PLOS ONE* **7**, e50009, DOI: 10.1371/journal.pone.0050009 (2012). Publisher: Public Library of Science.

63. Li, Y. *et al.* The Arabidopsis F-box E3 ligase RIFP1 plays a negative role in abscisic acid signalling by facilitating ABA receptor RCAR3 degradation. *Plant, Cell & Environ.* **39**, 571–582, DOI: 10.1111/pce.12639 (2016).

64. Vierstra, R. D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10**, 385–397, DOI: 10.1038/nrm2688 (2009).

65. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332, DOI: 10.1038/nature10317 (2011).

66. Theissen, G. *et al.* A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149 (2000).

67. Yang, X., Makaroff, C. A. & Ma, H. The Arabidopsis MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis. *The Plant Cell* **15**, 1281–1295 (2003).

68. Qiu, D., Xu, S., Wang, Y., Zhou, M. & Hong, L. Primary Cell Wall Modifying Proteins Regulate Wall Mechanics to Steer Plant Morphogenesis. *Front. Plant Sci.* **12**, 751372, DOI: 10.3389/fpls.2021.751372 (2021).

69. Majda, M. & Robert, S. The Role of Auxin in Cell Wall Expansion. *Int. J. Mol. Sci.* **19**, 951, DOI: 10.3390/ijms19040951 (2018).

70. Carter, C. *et al.* The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins. *The Plant Cell* **16**, 3285–3303, DOI: 10.1105/tpc.104.027078 (2004).

71. Zeleny, R. *et al.* Molecular cloning and characterization of a plant alpha1,3/4-fucosidase based on sequence tags from almond fucosidase I. *Phytochemistry* **67**, 641–648, DOI: 10.1016/j.phytochem.2006.01.021 (2006).

72. Berthet, S. *et al.* Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of Arabidopsis thaliana stems. *The Plant Cell* **23**, 1124–1137, DOI: 10.1105/tpc.110.082792 (2011).

73. Hoffmann, N., Benske, A., Betz, H., Schuetz, M. & Samuels, A. L. Laccases and Peroxidases Co-Localize in Lignified Secondary Cell Walls throughout Stem Development. *Plant Physiol.* **184**, 806–822, DOI: 10.1104/pp.20.00473 (2020).

74. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646, DOI: 10.1093/nar/gkac1000 (2023).

75. Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. & Jugulam, M. Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* **9**, 454, DOI: 10.3390/antiox9050454 (2020).

76. Gallet, A. Hedgehog morphogen: from secretion to reception. *Trends Cell Biol.* **21**, 238–246, DOI: 10.1016/j.tcb.2010.12.005 (2011).

77. Callejo, A., Torroja, C., Quijada, L. & Guerrero, I. Hedgehog lipid modifications are required for Hedgehog stabilization in the extracellular matrix. *Dev. (Cambridge, England)* **133**, 471–483, DOI: 10.1242/dev.02217 (2006).

78. Carey, S. B. *et al.* The evolution of heteromorphic sex chromosomes in plants, DOI: 10.1101/2024.12.09.627636 (2024). Pages: 2024.12.09.627636 Section: New Results.

79. Faux, A.-M. & Bertin, P. Modelling approach for the quantitative variation of sex expression in monoecious hemp (Cannabis sativa L.). *Plant Breed.* **133**, 782–787, DOI: 10.1111/pbr.12208 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pbr.12208.

80. Arzberger, P. *et al.* Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Sci. J.* **3**, DOI: 10.2481/dsj.3.135 (2006).

81. Sá, C. & Grieco, J. Open Data for Science, Policy, and the Public Good. *Rev. Policy Res.* **33**, 526–543, DOI: 10.1111/ropr.12188 (2016). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ropr.12188.

82. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. (Oxford, England)* **29**, 15–21, DOI: 10.1093/bioinformatics/bts635 (2013).

83. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. (Oxford, England)* **30**, 923–930, DOI: 10.1093/bioinformatics/btt656 (2014).

84. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550, DOI: 10.1186/s13059-014-0550-8 (2014).

85. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419, DOI: 10.1038/nmeth.4197 (2017).

86. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323, DOI: 10.1186/1471-2105-12-323 (2011).

87. Wickham, H. *et al.* dplyr: A Grammar of Data Manipulation (2023).

88. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinforma. (Oxford, England)* **32**, 2847–2849, DOI: 10.1093/bioinformatics/btw313 (2016).

89. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812, DOI: 10.1093/bioinformatics/btu393 (2014).

90. De Vico Fallani, F., Latora, V. & Chavez, M. A Topological Criterion for Filtering Information in Complex Brain Networks. *PLOS Comput. Biol.* **13**, e1005305, DOI: 10.1371/journal.pcbi.1005305 (2017).

91. Serrano, M. A., Boguna, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.* **106**, 6483–6488, DOI: 10.1073/pnas.0808904106 (2009). ArXiv:0904.2389 [physics].

92. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**, 8577–8582, DOI: 10.1073/pnas.0601602103 (2006). Publisher: Proceedings of the National Academy of Sciences.

93. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014, DOI: 10.1093/bioinformatics/btad014 (2023).

Acknowledgments

Support for this work was provided by the Institute of Cannabis Research at Colorado State University Pueblo ICR-FY22-Kane (R.F.G., D.V., N.C.K.), and NIH R35 GM147107 (R.F.G.). We also acknowledge the illustrative figure design provided by Maisie G. Troilett and Ruben Orozco-Ortiz, as well as the Google Cloud computing resources generously donated in part by SciAnno Mosaics LLC. Anthony Baptista acknowledges support from the CRUK City of London Centre Award [$CTRQQR - 202100004$].

Author Contributions Statement

D.V., R.G., C.P and N.K. conceived the experiments, C.P. and K.W. collected the data in the experiments, L.O., A. B, and A.W. constructed Nota and analyzed the results. L.O., A.W., N.R.M, and C.J.G constructed Jack. B.C.K, B.F.E generated Network visualizing figures. All authors reviewed the manuscript.

Data Access Statement

All the RNA-seq data generated in this analysis were deposited to the NCBI SRA under the Submission(SUB15016853) title "Genetic Networks Underlying Sexual Plasticity in Cannabis sativa." These data are associated to the bioproject PRJNA121253. The Adal et al. 2021 dataset is available via the bioproject PRJNA66938. For beta access to Nota and Jack please visit the SciAnno Mosaics website. A direct link to our species representative Cannabis sativa STRINGDB protein-protein interaction is network. Interaction networks in "Gene Network Ontology of Shared Non-Unique Inferences" and "Gene Network Ontology of CN One Inferences" can be found in the network hyperlinks at the beginning of their respective sections.

Plant Material Processing Compliance Statement

The authors processed the plant DNA, RNA, and other data in accordance with local and federal law, under the Research and Development permit 08 - 104227 to the Regents of the University of Colorado.

Financial and Non-Financial Conflict of Interest Statement

The authors declare no competing financial interests. Leonardo R. Orozco, Audrey E. Weaver, Anthony D. Baptista, and Nolan C. Kane, are inventors of Nota and members of SciAnno Mosaics LLC Boulder, Colorado, USA, which holds exclusive licensing rights to Nota through the University of Colorado Boulder. We acknowledge that this ownership has not influenced the research or findings reported in this publication. Provisional Patent Application No.: 63/743,561 Title: Bioinformatics System to Assist in Genetic Annotation of Model and Non-model Organisms.

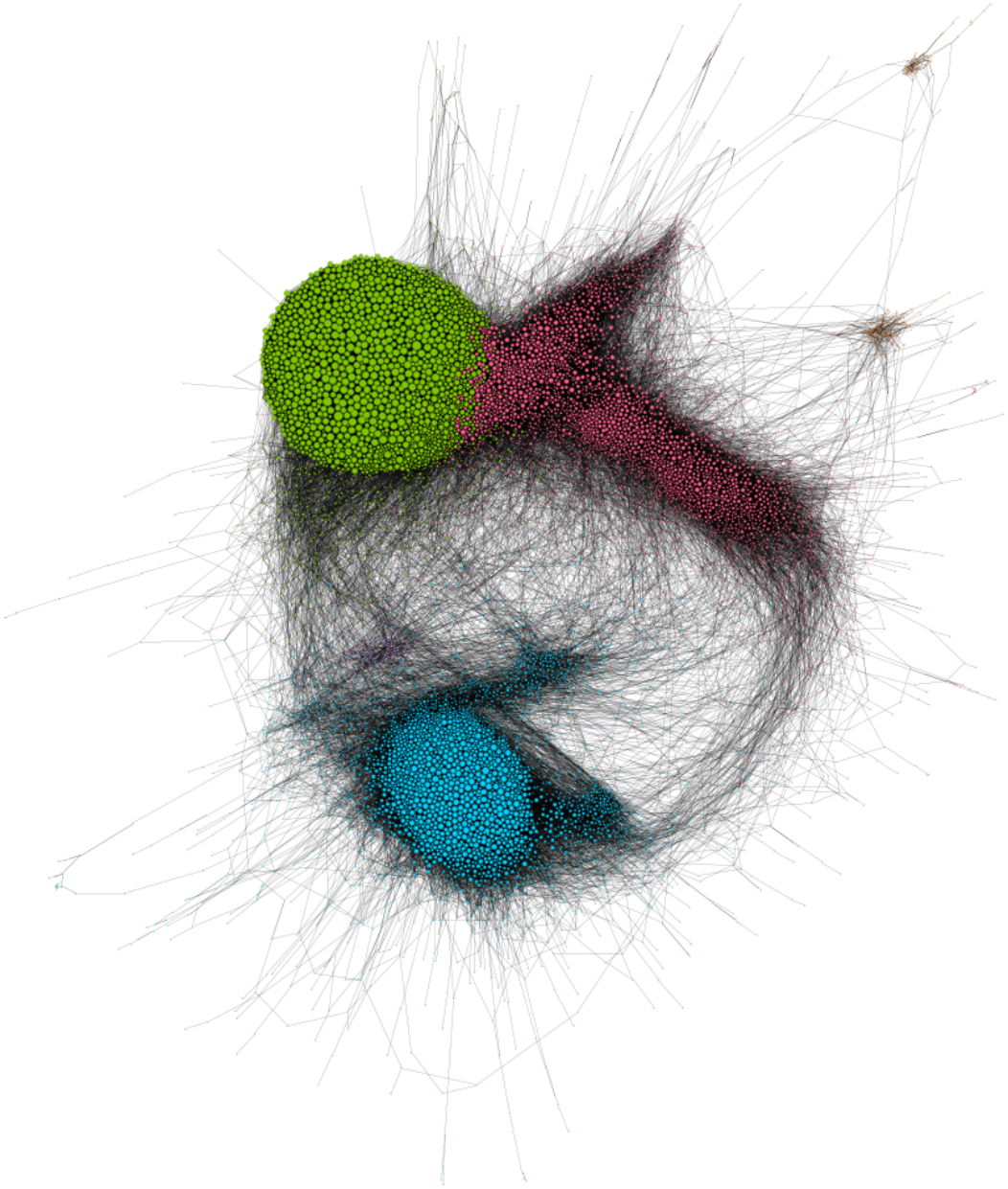

Figure 7. Filtered Co-expression Gene Network A 100,000 edge sample of the 1.36 million edge gene co-expression network (11,533 nodes) laid out using the ForceAtlas2[?] spring layout algorithm. Nodes are sized by degree and colored by modularity class.

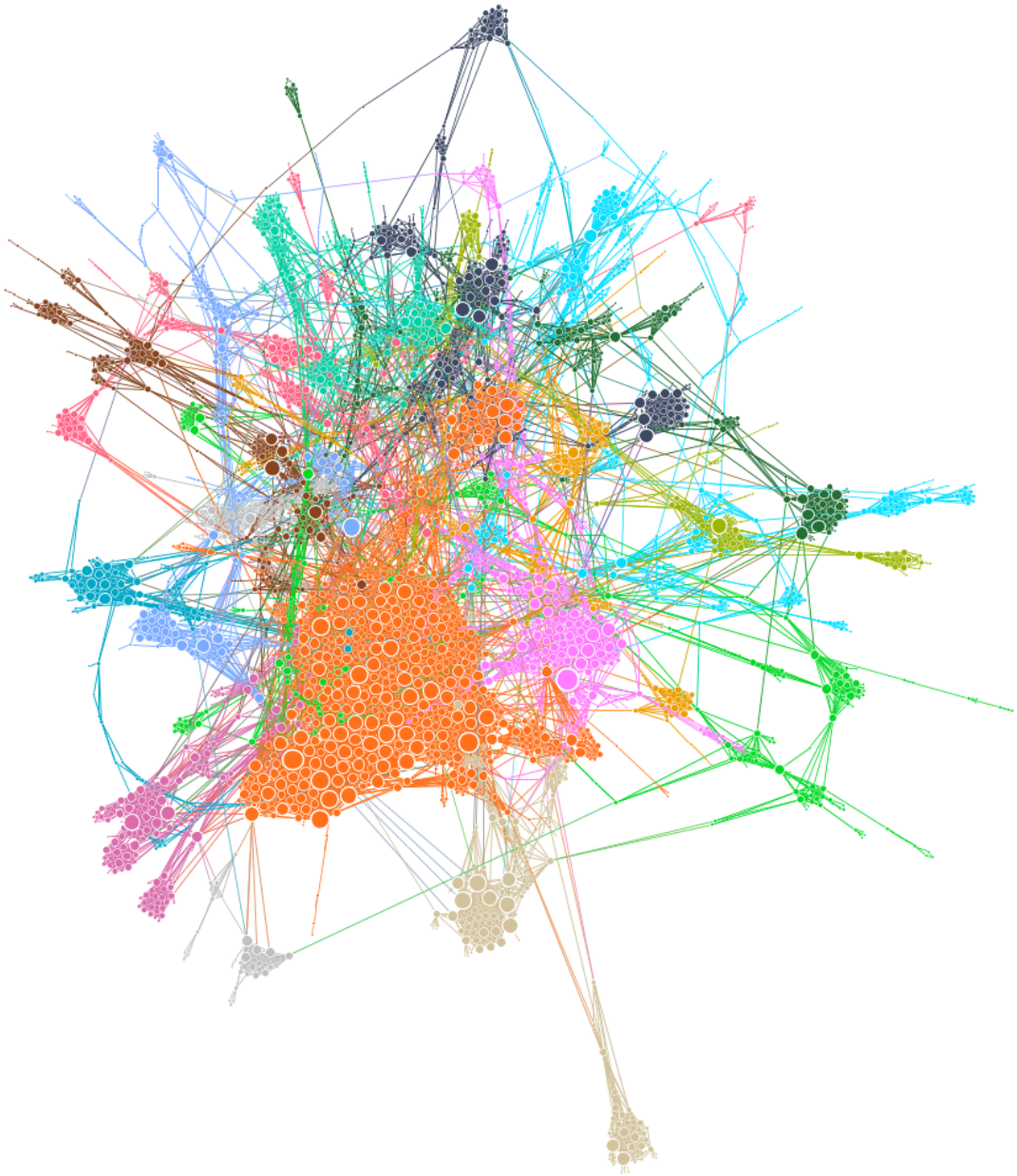

Figure 8. Sampled Cannabis Species Representative Protein-Protein Interaction Network Because the weighted network (12,736 and 1.28 million edges) was over dense for visualization, the multi-scale backbone with a disparity filter set at 0.05 significance reduced the network to find underlying structure. Graph laid out in ForceAtlas2 and nodes are sized by degree and colored by modularity class.