Exploring the Medicinally Important Scondary Metabolites Landscape Through the Lens of Transcriptome Data in Fenugreek (Trigonella Foenum Graecum L.)

Mahantesha B.N. Naika
University of Horticultural Sciences-Bagalkot
Nitish Sathyanarayanan
National Centre for Biological Sciences (TIFR)
Radha Sivarajan Sajeevan
National Centre for Biological Sciences (TIFR)
Teerna Bhattacharyya
National Centre for Biological Sciences (TIFR)
Pritha Ghosh
National Centre for Biological Sciences (TIFR)
Iyer S. Meenakshi
National Centre for Biological Sciences (TIFR)
Mahita Jarjapu
National Centre for Biological Sciences (TIFR)
Adwait G. Joshi
National Centre for Biological Sciences (TIFR)
K. Harini
National Centre for Biological Sciences (TIFR)
K. Mohamed Shafi
National Centre for Biological Sciences (TIFR)
Neha V. Kalmankar
National Centre for Biological Sciences (TIFR)
Snehal D. Karpe
National Centre for Biological Sciences (TIFR)
Bhavika Mam
National Centre for Biological Sciences (TIFR)
Shaik Naseer Pasha
National Centre for Biological Sciences (TIFR)
Ramanathan Sowdhamini (mini@ncbs.res.in)
National Centre for Biological Sciences (TIFR)

Article

Keywords: Trigonellafoenum-graecum, transcriptome, secondary metabolite, differential expression, metabolite quantification

Posted Date: March 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1441723/v1

Abstract

Fenugreek (*Trigonella foenum-graecum* L.) is a self-pollinated leguminous crop belonging to the Fabaceae family. It is a multipurpose crop used as herb, spice, vegetable and forage. It is a traditional medicinal plant in India attributed with several nutritional and medicinal properties including antidiabetic and anticancer. We have performed a combined transcriptome assembly from RNA sequencing data derived from leaf, stem and root tissues. Around 2,09,831 transcripts were deciphered from the assembly of 92% completeness and an N50 of 1382 bases. Whilst secondary metabolites of medicinal value, such as trigonelline, diosgenin, 4-hydroxyisoleucine and quercetin, are distributed in several tissues, we report transcripts that bear sequence signatures of enzymes involved in the biosynthesis of such metabolites and are highly expressed in leaves, stem and roots. One of the antidiabetic alkaloid, trigonelline and its biosynthesising enzyme, is highly abundant in leaves. These findings are of value to nutritional and the pharmaceutical industry.

Highlights

- Fenugreek (AFG-1) transcriptome from leaf, stem and root tissue enables the exploring landscape of enzymes involved in synthesis of medicinally important secondary metabolites.
- Relative expression of these transcripts across tissues was quantified and validated.
- Selected secondary metabolites have several important medicinal values including antidiabetic, anticancer, anti-inflammatory, antioxidant properties.
- Trigonelline, 4-hydroxyisoleucine and diosgenin which have antidiabetic and antiobesity properties were quantified in leaf, stem and root tissues.
- Trigonelline and its biosynthesising enzyme are most abundant in the leaf tissue.

1. Introduction

Fenugreek (*Trigonella feonum-graecum* L.) is a dicotyledon leguminous self-pollinated crop with chromosome number 2*n* = 16 belonging to the Fabaceae (Leguminosae) family [1]. It is native to India and eastern Mediterranean countries. The genus *Trigonella* includes 140 species distributed globally, among which 107 species occur in Asia [2]. In India mainly two types are commonly cultivated i.e., common fenugreek (*Trigonella foenum-graecum* L.) for leaves, stem and seeds, and Kasuri fenugreek (*Trigonella corniculata* L.) for leaves. Fenugreek is one of the oldest known medicinal plants of India and is traditionally used in ayurvedic medicines. It is considered to be a multipurpose crop for its food and non-food uses such as herb, spice, fodder, and its nutraceutical, pharmaceutical and therapeutic properties [3]. In 2017, Venkata and co-workers [4] reviewed fenugreek for its broad pharmaceutical usage on preclinical and clinical research on human diseases and antipathogenic properties.

Although there has been considerable advancement in the field of drug development, medicinal plants are being increasingly utilised for treating various diseases owing to their therapeutic properties and safety.

Fenugreek has a high content of dietary fibre, vitamin and mineral content as well as bioactive compounds such as modified amino acids, phenolic acids, alkaloids and saponins (Fig. 1). It has multiple health benefits, such as antidiabetic, anticancer, galactagogue, helps in digestion, hepatoprotective effects, regulatory functions, antioxidant properties, and works against anorexia, antilithogenic, hepatoprotective effect, antipathogenic properties and several other medicinal properties [5–7]. It is also Generally Recognized As Safe (GRAS) while using as a flavour by the U.S. Food and Drug Administration (FDA) [8].

A recent review reports the presence of more than 100 phytochemicals in the fenugreek seeds [9]. Among the several important secondary metabolites produced by fenugreek, diosgenin, trigonelline and 4-hydroxyisoleucine are three compounds which have been implicated in the anti-diabetic properties of the plant (Fig. 1). Diosgenin, a steroidal saponin, is an important starting material for the preparation of several steroidal drugs in the pharmaceutical industry and has shown high potential in the treatment of various types of disorders such as cancer, hypercholesterolemia, inflammation, and several types of infections [10]. The pharmacological activities of trigonelline, an alkaloid, include hypoglycaemic, hypolipidemic, neuroprotective, antimigraine, sedative, memory-improving, antibacterial, antiviral, and anti-tumour activities [11]. 4-hydroxyisoleucine (4-HIL), a modified amino acid is another bioactive compound abundant in fenugreek seeds. The antidiabetic property of 4-HIL is attributed to its ability to induce insulin secretion in a glucose-dependent manner [12,13]. There are several other important classes of metabolites such as volatiles (eugenol, linalool), flavonoids (quercetin, rutin, vitexin and isovitexin) which render fenugreek multiple medicinal properties [7,14].

There have been a limited number of studies for identification of potential genes involved in biosynthesis of bioactive compounds attributing the nutritional and pharmaceutical properties of the fenugreek as well as differential expression of these genes across the tissues of the plant. The transcriptome studies in fenugreek will pave the way for identification of candidate genes involved in medicinal as well as environmental stress tolerance properties of the plant and these can be taken up further in molecular breeding programs [15,16]. Transcriptomics provide the details of genes associated with functions at the cellular and tissue level; analysis of transcriptomic data from different tissues of the organism is particularly useful in understanding differential gene expression [17]. Towards this goal, we have performed transcriptome sequencing from leaf, stem and root tissues of fenugreek and carried out a differential expression analysis observed across these tissues. We use this data to explore the secondary metabolite landscape of the plant by deciphering genes involved in synthesis of several nutritionally and medicinally important metabolites. We have also quantified the expression of selected genes and performed quantification of the selected metabolites.

2. Materials And Methods

2.1 Sequencing and assembly of the transcriptome

Seeds of *T. foenum-graecum* (Ajmer Fenugreek-1 (AFG-1) variety) were procured from College of Horticulture, GKVK Campus, Bangalore with due permission. All the studies on *T. foenum-graecum* were carried out in accordance with relevant institutional, national and international guidelines and legislation. For transcriptome sequencing, RNA from three different tissues- leaf, stem and root, was isolated from one month old plants using the Plant Spectrum Plant RNA extraction kit (Sigma-Aldrich). The samples, with a technical replicate for each tissue, were sequenced using Illumina HiSeq 1000 platform with 100 base pairs (bp) read length. Trimmomatic was used to process the reads using default settings [18]. A total of 226.5 million reads were obtained after quality processing from six libraries. These reads were assembled *de novo* using Trinity with default parameters [19]. BUSCO was used to assess the completeness of the transcriptome assembly [20].

2.2 Transcriptome functional annotation

TransDecoder (v3.0.0) was used to identify the candidate coding regions within transcript sequences obtained from Trinity [19]. A length cut-off of 100 bases was used to predict Open Reading Frames (ORFs). Function assignment for amino acid sequences was obtained using BLASTP [21] against UniProt viridiplantae database with an E-value cut-off of 10^{-5} [22] and domain identification was performed using hmmscan of the HMMER suite (v.3.1b2) [23] against Pfam (v.31.0) using an E-value cut-off of 10^{-5} [24]. For transcripts, BLASTX [21] was performed against the above sequence database with the default parameters. Additional annotation was rendered using SignalP (v.4.0) [25], TMHMM (v.2.0) [26] and RNAmmer (v.1.2) [27] to extract information on signal peptides, transmembrane regions and RNA sequences. Gene ontology (GO) terms [28], orthologous relationships (eggNOG) [29] and pathway information (KEGG) [30] were also mined.

2.3 Detection of orthologous relationships

Orthology analysis of *T. foenum-graecum* was carried out with the well-studied model plants *Arabidopsis thaliana* and *Oryza sativa* and plants which belong to the same tribe as fenugreek (Trifolieae) viz. *Trifolium pratense* and *Medicago truncatula* using OrthoMCL [31]. An E-value cut-off of 10^{-5} was used for the all-against-all BLAST runs. Additionally, 34 different plant species proteomes from Phytozome (v.12) [32] were used for large scale orthology analysis using ProteinOrtho (v5.11) [33] (Supplementary Table 1) with an E-value cut-off of 10^{-5}.

The concatenated nucleotide sequence alignments for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL), maturase K (matK) and internal transcribed spacer 1 and 2 (ITS1 and ITS2) were used for constructing species phylogenetic tree for the species belonging to subclass Rosids. In case of partial sequences, the alignment was trimmed. The individual sequences were aligned using MAFFT[34] with maximum allowance for aligning gapped regions and the separate alignments were concatenated. Phylogenetic reconstruction was performed using the Neighbour-Joining method implemented in MAFFT with 100 bootstraps. The orthologue distribution obtained from ProteinOrtho analysis along with their phylogenetic tree for comparison amongst these closely related species was plotted. Through orthogroup

analysis, the in-paralogous groups and singletons of *T. foenum-graecum* common to both methods were identified. The sequences were compared with each other with respect to their expression profiles in different tissues, Pfam domains [22] and GO terms [28].

2.4 Differential Expression Analysis

The reads from the three tissues were mapped onto the *de novo* transcriptome assembly using Bowtie2 (v2.3.0). The expression counts were obtained using eXpress (v1.5.1) and RSEM [35–37]. Transcripts per Million (TPM) values estimated for all the transcripts and compared across the replicates as well as across the three tissues. Among these replicates, the greater of the two values was reported.

2.5 Identification and analysis of genes involved in synthesis of secondary metabolites

The genes involved in synthesis of secondary metabolites belonging to alkaloids, saponins, volatiles and flavonoids, present in *T. foenum-graecum* were identified using strategy adapted from Joshi et al (2020) [38]. The pathway information and related functional annotation was extracted from multiple sources such as PlantCyc, MetaCyc, KEGG and literature [11,30,39–42]. The sequences for the enzymes involved in the synthesis of secondary metabolites serving as start points for sequence search were identified from UniProt [22]. This was followed with creation and curation of multiple sequence alignment profiles including functionally important residues (FIR), typically used for selecting trusted homologues [21,43]. Such profiles were implemented to search the transcriptome data for potential hits and were ascertained by presence of the FIRs. The phylogenetic trees were constructed using the neighbor-joining method from MEGA (v10) [44] for 1000 bootstraps, to assess co-clustering of the candidate genes with the curated start points. The phylogenetic trees obtained were visualized and edited in Figtree (v1.4.2) [45] and iToL [46].

2.6 Expression analysis of genes involved in production of important secondary metabolites

The genes identified to be involved in production of secondary metabolites trigonelline and diosgenin were validated using quantitative reverse transcription PCR (qRT-PCR). The details of methods used for RNA isolation, qRT-PCR quantification are described in the Supplementary Text (section 1.1).

2.7 Quantification of diosgenin, trigonelline and 4-hydroxyisoleucine in different tissue samples

The quantification of diosgenin, trigonelline and 4-hydroxyisoleucine was carried out from leaf, stem and root tissue samples using HLPC-PDA method. The details of this method are described in the methods part of Supplementary Text (section 1.2).

3. Results

We carried out transcriptome sequencing of three tissues (leaf, stem and root) of fenugreek plant and constructed a combined assembly along with functional annotation. We investigated the orthology relationship with other plant species and also assessed the differential gene expression across tissues. Further, we explored the secondary metabolite landscape of fenugreek considering metabolites with reported medicinal properties from several classes of compounds (alkaloids, saponins, volatiles, flavonoids, etc.) and mined for the enzymes involved in their synthesis.

3.1 Transcriptome assembly and functional annotation

Combined transcriptome assembly was obtained from three different tissues (leaf, stem and root) with N50 of 1382 bases. The 169.5 Mb assembly contained 209,831 transcripts. This assembly was 92.2% complete (95.49% using partial core genes) as assessed through the BUSCO [20].

The candidate coding regions within the fenugreek transcriptome were predicted using Transdecoder program. A set of 64,670 ORFs were predicted from the assembly which varied from 200 to 2000 base pairs. We obtained annotations, using BLASTX against UniProt Viridiplantae database, for 26% of transcripts (Table 1). Whereas, among the predicted ORFs, we obtained annotation for 90% of ORFs, through BLASTP using E-value threshold of 10^{-5} and 50% query coverage cut-off. Most of these ORFs (76%) were associated with *Medicago truncatula*, a closely related species, as the best hit.

Table 1
Transcriptome assembly and annotation statistics

Details of the transcriptome assembly						
Assembly	**Total number of transcripts**	**Number of transcripts > 1 Kb**	**N50 (bp)**	**Size (Mb)**	**GC%**	**BUSCO completeness**
Combined assembly	209831	48685	1382	169.56	36.58	92.2% (95.49% including partial genes)
Functional annotation of transcripts and predicted amino acid sequences						
Annotation of transcripts (209831):						
BLASTX	54577 (26%)					
GO	60798 (30%)					
rRNAs (RNAmmer)	8 (0.04%)					
Annotation of predicted amino acid sequences (Total ORFs: 64670):						
Sequence hits (BLASTP)	58504 (90%)					
Protein domains (Pfam-hmmscan)	42524 (66%)					
Signal peptides (SignalP v.4.0)	3743 (6%)					
Transmembrane helices (TMHMM v.2.0)	12417 (19%)					
Orthologous groups (eggNOG/COG)	37243 (58%)					
Association with KEGG Pathway	38930 (60%)					
GO terms	38790 (59%)					

Among the predicted ORFs, 66% (42524) could be associated with Pfam domains, representing 4400 unique Pfam domains. The leucine-rich repeat (LRR), pentatricopeptide repeat (PPR), tetratricopeptide repeats (TPR) and several transcription factor binding (WD40, F-box, AAA) domain families were the most abundant annotated domains observed. Similar to other angiosperms, several other Pfam domains such as, protein serine/threonine kinases, protein tyrosine kinases, calcium binding EF-hand and ankyrin repeats were found in abundance (Supplementary Fig. 1A). We could associate GO terms to more than 50% of total ORFs. The top enriched GO terms were classified into molecular function (MF), biological

process (BP) and cellular component (CC) (Supplementary Fig. 1B). The enriched terms under MF included transporter activity, transcription regulator, antioxidant activity and nutrient reservoir. Whereas, some of the most enriched BP terms were metabolic process, response to stimuli and immune system process. It was observed that transmembrane helices were present in 19% of ORFs, signal peptides were predicted in 6% of ORFs, whereas eight rRNAs were identified. Of the total number of amino acid sequences obtained from the transcriptome assembly, 59% were mapped to functional annotation databases, namely eggNOG (orthogroups), GO and KEGG pathways (Table 1; complete annotation data available in Supplementary Data 1).

3.2 Detection of orthologous relationships

The orthologs were detected using OrthoMCL and ProteinOrtho for fenugreek and closely related plants (Figs. 2A and 2B and Supplementary Fig. 2). Among the five plants used for OrthoMCL analysis, 8822 orthogroups were common to all whereas, for ProteinOrtho, 114 orthogroups were common across 39 plants considered for the analysis. It was observed that the orthogroup distribution was highly similar across all the plants from subfamily Faboideae (*G. max, P. vulgaris, T. pratens, M. truncatula* and *T. foenum-graecum*). We constructed the phylogenetic tree using selected plants from subclass Rosids and *O. sativa*. (Fig. 2B).

We observed 20,230 fenugreek specific paragroups and singletons, among which only the longest ORFs (15,712, 24% of all predicted gene products) were assessed for their functional annotation (hereafter referred to as singletons). A total of 6300 of the singletons in fenugreek proteins show presence of 1771 unique Pfam families. PPR, LRR, protein kinases, protein tyrosine kinases, zinc knuckle and reverse transcriptase domain were among the most abundant Pfam domains in the singleton proteins (Supplementary Fig. 3A). Almost half of the fenugreek singletons (7235) were found to have Gene Ontology (GO) annotation. Nucleotide/nucleic acid binding, metal ion binding were abundant in molecular functions within GO terms (Supplementary Fig. 3B). Translation, regulation of transcription and defence response were the most abundant biological process (Supplementary Fig. 3C). The most common cellular component term within these GO terms was ‘integral component of membrane’ (Supplementary Fig. 3D).

3.3 Differential expression analysis

The transcript abundance across all three tissues (leaf, stem and root) was assessed. The top 20 significantly expressing transcripts were identified from each tissue and are represented as a heatmap (Fig. 3). While observing variation between highly expressed transcripts in each tissue, we could identify six transcripts (protease inhibitor, metallothionein, *dehydrin B*, *thioredoxin H*, plant invertase inhibitor and translationally controlled tumor-like protein) highly expressed across three tissues. Considering the leaf sample, the highly expressed transcripts were mostly related to photosynthesis, for example, ribulose bisphosphate carboxylase, chlorophyll a-b binding protein, carbonic anhydrase etc. In the stem, apart from photosynthesis related transcripts, few cytochrome genes and pathogenesis related protein families were highly expressed. In the root, the top three highly expressed annotated transcripts included

metallothionein, *dehydrin B* and *Nmr-A* like family of genes. These genes are well documented to be involved in drought tolerance in crops [47–50]. The tissue-wide expression values for enzymes involved in synthesis of several secondary metabolites considered in this work are mentioned in the subsequent section.

3.4 Identification and analysis of genes involved in synthesis of secondary metabolites

We explored the secondary metabolites landscape in fenugreek with an aim to identify the transcripts involved in their biosynthesis. The candidate metabolites and the enzymes involved in their biosynthesis were chosen following an extensive literature survey. These spanned several classes of secondary metabolites: alkaloids, saponins, volatiles and flavonoids.

3.4.1 Alkaloids

Alkaloids are heterocyclic nitrogen containing compounds displaying a wide array of pharmacological properties. These compounds are not only associated with plant growth and defence but are a storage source of nitrogen [51]. We considered trigonelline and glycine-betaine from the alkaloid class found in fenugreek. We pursued identification of enzymes involved in their biosynthesis by adopting the strategy described in Joshi et al (2020) [38].

Trigonelline is a derivative of nicotinic acid, first isolated from fenugreek (*Trigonella foenum-graecum*). Trigonelline has several medicinal properties such as hypoglycaemic, neuroprotective, antibacterial, and anti-tumour activities [52]. It is synthesised through purine nucleotide cycling, wherein, nicotinate is methylated to methyl nicotinate (trigonelline) by nicotinate N-methyltransferase (NNMT) (EC: 2.1.1.7). We selected and curated protein sequences for this enzyme found in other plants, as a start point to carry out sequence search in fenugreek transcriptome. There are five functionally important residues (FIRs), a substrate-binding motif Asn 21, Tyr 120, His 124 a catalytic motif Thr 264 and a SAM binding motif in the NNMT enzyme [42]. Although 12 hits were obtained, a single hit (Tfoe_c19814_g2_i1_m7872) was identified based on co-clustering with *Medicago truncatula* sequence (NCBI RefSeq ID: XP_013464471.1) (Supplementary Fig. 4E) with a sequence identity of 92.6 %and a 100 %query coverage. Further this hit was validated through mapping of the FIRs (Supplementary Fig. 4A-D). Based on the TPM values, this transcript was expressed highest in the leaf tissue sample (Fig. 4, Supplementary Table 3). This trend was also similar to the expression level estimation through qRT-PCR (Supplementary Fig. 5A). Trigonelline, as a compound, was quantified across different tissues of fenugreek as described further in section 3.5.

Glycine betaine is one of the major osmoprotectants in plants and also contributes to maintaining membrane and cell stability. It plays a protective role against inflammation, carcinogenesis and neurodegenerative diseases [53–55]. It was identified in the extract of seeds of fenugreek [56]. It is synthesized in plants by the oxidation of choline through a two-step process (Supplementary Data 2).

The first step is catalyzed by choline monooxygenase (CHMO) (EC: 1.14.15.7) which is an unusual iron-sulfur containing enzyme. We identified one hit (Tfoe_c38951_g1_i2_m30577) from fenugreek which contained both the motifs (Rieske 2Fe-2S and non-heme binding region). The second step catalyzes the formation of glycine-betaine from betaine-aldehyde by betaine aldehyde dehydrogenase (BADH) (EC: 1.2.1.8) [57]. We identified one hit for BADH (Tfoe_c82300_g1_i1_m72031) and mapped the FIRs (proton acceptor: Glu260, Nucleophile: Cys294 and cation-pi interacting Trp285, Trp456). The hits for both the enzymes in fenugreek co-clustered with the *M.truncatula* proteins (Supplementary Data 2). Both these transcripts were relatively abundant in the leaf tissue sample (Supplementary Table 3).

3.4.2 Saponins

Saponins are amphipathic glycosides grouped phenomenologically by the soap-like foaming they produce when shaken in aqueous solutions. These metabolites are found in the plant families- Sapindaceae, Aceraceae, Hippocastanaceae and Fabaceae [58]. We focussed on the biosynthetic pathway of diosgenin, a steroidal sapogenin attributed to treatment of metabolic disorders such as diabetes and hypercholesterolemia and also implicated in treatment of cancer, inflammation and several infections [10]. Tissue-wide diosgenin content varies in fenugreek plant and is reported to be present in highest quantity in seeds. Moreover, its quantity differs across several cultivars and is considered one of the criterion in crop improvement programs [59].

From the proposed biosynthetic pathway of diosgenin, as per Ciura et al. 2017 [60], (Supplementary Data 2), putative sequences corresponding to the last two enzymes (sterol-3β-glucosyltransferase and β-glucosidase) were identified from the set of protein sequences obtained from the transcriptome assembly. Sterol-3β-glucosyltransferase (EC: 2.4.1.173) contains two motifs- putative steroid binding domain (PSBD) motif and plant secondary product glucosyltransferase (PSPG) motif [61]. The hits obtained from the searches against the proteins sequences were filtered for the presence of these two motifs and of the four hits that contained this motif, from the FIR analysis, two transcripts (Tfoe_c26467_g1_i1_m12030 and Tfoe_c33087_g1_i3_m19456) expressing significantly in all three tissues of the plant were considered as true hits (referred henceforth as sterol-3β-glucosyltransferase-1 and − 2). While sterol-3β-glucosyltransferase-1 (Tfoe_c26467_g1_i1_m12030) clustered with a sequence from a closely related plant, *M. truncatula*, the sterol-3β-glucosyltransferase-2 (Tfoe_c33087_g1_i3_m19456) clustered with the enzyme sequence from *Glycine soja*. Sterol-3β-glucosyltransferase-1 expressed well in all tissues of the plant (leaf, stem and root) while sterol-3β-glucosyltransferase-2 showed relatively higher expression in leaves and this was further validated using qRT-PCR (Supplementary Fig. 5B, 5C).

The last enzyme in the biosynthetic pathway of diosgenin, β-glucosidase (EC: 3.2.1.21), is known to contain two catalytic glutamate residues which are part of the (I/V)TENG and TFNEP motifs and these have been identified from crystal structures of β-glucosidase from rice in apo and bound forms (2RGL and 2RGM) [62]. We identified the β-glucosidase in fenugreek (Tfoe_c120410_g1_i1_m78951) and confirmed the presence of these motifs. Although, we found a total of 22 such hits which had > 70% query

coverage and > 30% sequence identity, with the query β-glucosidase sequences from Viridiplantae, the selected candidate protein Tfoe_c120410_g1_i1_m78951 clustered with the enzyme from *M. truncatula* and was thus considered a true hit. This transcript showed high expression in root followed by stem and the same relative expression pattern was observed from qRT-PCR validation (Supplementary Fig. 5D, Supplementary Table 3). The metabolic pathway, sequence motif, phylogenetic tree and the expression data for all diosgenin synthesising enzymes is documented in Supplementary Data 2.

3.4.3 Volatiles

The volatiles are a class of compounds which typically include aromatic organic compounds. These compounds are used in perfumes, flavours, topical formulations, etc. The medicinal properties of essential volatiles have been known since ancient times. We have explored the enzymes involved in the biosynthesis of eugenol and linalool which are volatile oils found in fenugreek [63,64]. The therapeutic potential of eugenol and isoeugenol, which are phenolic compounds, is attributed to antioxidant potency and anti-inflammatory benefits [65,66]. Linalool, a terpene, has been reported to show anticancer, analgesic, anxiolytic and other neuroprotective properties [67]. Eugenol and isoeugenol are phenylpropene volatiles that have a phenyl ring with a propenyl side chain synthesized from coniferyl alcohol precursor in multiple enzymatic reactions (Supplementary Data 2). We studied three enzymes in the eugenol synthesis pathway: Coniferyl alcohol acetyltransferase, eugenol synthase and isoeugenol synthase [68]. Coniferyl alcohol acetyltransferase (EC: 2.3.1.224) associated with two proteins (Tfoe_c37673_g1_i1_m27189, Tfoe_c102741_g1_i1_m76275), however, Tfoe_c37673_g1_i1_m27189 was selected to be the true candidate gene (marked in Supplementary Data 2) since it co-clustered with the *M. truncatula* protein and was also supported by the presence of the FIR. Similarly, the candidate proteins from fenugreek for eugenol synthase (EC: 1.1.1.318) and the isoeugenol synthase (EC: 1.1.1.319) co-clustered with clades belonging to known enzymes (highlighted in Supplementary Data 2). The TPM values across tissues for each of the enzymes involved in synthesis of eugenol are represented in Supplementary Table 3. The transcripts corresponding to both the enzymes were found to express highest in root compared to the other tissues.

In the linalool synthesis pathway we studied three enzymes: Farnesyl-pyrophosphate synthase, (3S)-linalool synthase and (3R)-linalool synthase [69,70]. (3S)-linalool and (3R)-linalool are two enantiomeric forms of naturally occurring monoterpenoids. The candidate protein from fenugreek for farnesyl-pyrophosphate synthase (EC: 2.5.1.1) (Tfoe_c5299_g1_i1_m1972) co-clustered with another legume *Lupinus albus* (Fabaceae family). The (3S)-linalool synthase (EC: 4.2.3.25) (Tfoe_c11093_g1_i1_m4195, Tfoe_c11093_g1_i2_m4196) and (3R)-linalool synthase (EC: 4.2.3.26) (Tfoe_c44084_g4_i2_m55446) co-clustered with other sequence homologues (as highlighted in the Supplementary Data 2). The TPM values for each of the enzymes involved in synthesis of linalool are represented in Supplementary Table 3. While the first enzyme, i.e., farnesyl-pyrophosphate synthase is expressed highest in root compared to the other tissues, (3S)-linalool synthase and (3R)-linalool synthase is expressed highly in leaves.

3.4.4 Flavonoids

Flavonoids, an important class of secondary metabolites with a polyphenolic structure, widely found in fruits, vegetables and certain beverages. This class of compounds has been linked to a variety of nutraceutical, pharmaceutical, and medicinal properties, including antioxidative, anti-inflammatory, antimutagenic, and anticarcinogenic properties [71]. We studied biosynthetic pathways of four flavonoids (quercetin, vitexin, isovitexin and rutin) known to be present in fenugreek [63]. These metabolites are synthesised from coumaroyl-CoA through different enzymatic reactions. The flavonoids share a common flavonoid backbone structure with differences in their hydroxylation patterns.

Quercetin biosynthesis pathway involves six enzymes (4-Coumarate-CoA ligase (EC: 6.2.1.12), chalcone synthase (EC: 2.3.1.74), chalcone flavone isomerase (EC: 5.5.1.6), flavanone 3-hydroxylase (EC: 1.14.11.9) / flavonol synthase (EC: 1.14.20.6), tricin synthase (EC: 2.1.1.175) and flavonoid 3'-monooxygenase (EC: 1.14.14.82)) that convert coumaroyl-CoA to quercetin through several intermediates [17]. For each of the six enzymes involved in the pathway, multiple transcript hits were identified in fenugreek as mentioned in Supplementary Table 3. The sequence hits for these enzymes were validated as true hits, based on co-clustering with well annotated sequence of same sub-family observed in the phylogenetic tree (as described in Methods section 2.5). It was observed that the selected enzymes are highly expressed in roots of the plant followed by abundance of expression in the stem (Supplementary Table 3, Supplementary Data 2).

We further studied biosynthetic pathways for vitexin and isovitexin, which are two enantiomeric metabolites. The enzymes involved their pathway are identical [72], and they produce these metabolites in a mixture. A three-step reaction is known to be involved in the conversion of naringenin to vitexin/isovitexin. The first two reactions are enzymatic while the third step (hydration reaction) [73], although, has been hypothesized to be an enzymatic one, there is no definitive enzyme that has been isolated for it. For the first two enzymes (naringenin 2-hydroxylase (EC: 1.14.14.162) and C-glucosyltransferase (EC: 2.4.1.360)) known in the reaction, four and 13 hits were obtained from the fenugreek transcriptome Supplementary Table 3. These hits were validated using phylogenetic analysis of the enzymes from the same sub-family (Supplementary Data 2). The highest expression of the naringenin 2-hydroxylase in the pathway was observed in the stem, while for C-glucosyltransferase, although there were 13 hits identified fenugreek, most of them were highly expressed in leaf.

The third flavonoid, rutin, is synthesised in plants from taxol. The pathway involves two enzymatic reactions. While vitexin/isovitexin are obtained by C-glycosylation, rutin is obtained by O-glycosylation of the substrates. The enzymes involved in the pathway are flavonol-3-o-glucosyltransferase (EC: 2.4.1.91) and flavonol-3-o-glucoside rhamnosyltransferase (EC: 2.4.1.15) [74]. There were 11 hits identified for flavonol-3-o-glucosyltransferase, and most of them had higher expression value in leaf and root tissue. In case of flavonol-3-o-glucoside rhamnosyltransferase a single hit (Tfoe_c121323_g1_i1_m79650) was identified with a high expression in leaf (Supplementary Table 3, Supplementary Data 2).

3.5 Quantification of metabolites in different tissue samples

Fenugreek is a traditionally well-known plant for its medicinal properties and the presence of complex mixtures of key phytochemicals renders this plant such benefits [4,63]. In the literature, details of these chemical constituents in fenugreek tissue extracts have mainly been demonstrated by bio-analytical techniques. We quantified, presence of important metabolites- diosgenin, trigonelline and 4-hydroxyisoleucine (4-HIL) in different tissues (leaf, stem, root and seed) of fenugreek, by HPLC and mass spectrometry. Although high-quality RNA could not be purified from the seed tissue, we used it nonetheless for metabolite quantification. Diosgenin and trigonelline were quantified using HPLC-PDA and 4-HIL was quantified using LC-MS analysis from extracts of different tissues as described in Supplementary Methods section 1.2. Supplementary Figs. 6A and B show the standard curve of diosgenin and trigonelline respectively. The concentration of each compound was quantitatively determined by comparison of the peak area of the standard with that of the samples. Supplementary Figs. 7A-D and 7E-H refers to the HPLC peaks which showed the presence of trigonelline and diosgenin, and Supplementary Fig. 7L-N refers to LC profile from LC-MS data of 4-HIL. As seen in Supplementary Table 4, seed gave the highest concentration of diosgenin i.e. 1.56 µg/ml of the tissue followed by the root (1.036 µg/ml). Leaf tissue showed the highest concentration of trigonelline i.e. 4.715 µg/ml of the tissue followed by the seed (2.698 µg/ml). 4-HIL was detected at a very high concentration of 84.77 µg/ml in the dry seed and 77.48 µg/ml from the leaf. The concentration of 4-HIL was observed to significantly decrease with water-soaked seeds post 12 hours and 24 hours. The concentrations of these medicinally important metabolites vary across several cultivars of fenugreek [75,76].

4. Discussion

We have surveyed here the landscape of secondary metabolites in fenugreek, which are medicinally important, through the lens of transcriptome data. Apart from being a food crop of semi-arid nature, fenugreek is not only a food source but has been used in traditional medicines in several parts of the world. The medicinal properties of this drought tolerant leguminous plant can largely be attributed to an array of secondary metabolites produced in it [63]. In this study we have identified candidate genes involved in the biosynthesis of such metabolites belonging to four major classes, alkaloids, saponins, volatiles and flavonoids. The process adopted for identification of such genes in the transcriptome has been previously implemented in *Ocimum tenuiflorum*, *Moringa oleifera* [17,38,77]. It involved creation and curation of a knowledgebase for candidate gene products from other (closely) related plants and applying it through a rigorous sequence search and validation process to ascertain the transcripts in the fenugreek transcriptome assembly. The raw data from leaf, stem and root tissue samples was processed and assembled into a set of transcripts and their level of expression was compared across these tissues in measure of TPM values. In particular, several drought responsive genes such as metallothionein, dehydrin B and Nmr-A like family of genes as described in the results section 3.3 were found to be among highly expressed transcripts, especially in the root tissue sample.

During the process of identification of candidate genes involved in biosynthesis of secondary metabolites, it is vital to select appropriate start points for the rigorous sequence search and subsequent

validation. We rely on multiple sources such as PlantCyc, MetaCyc, KEGG pathway and metabolites cited in literature [11,30,39–42] to create such a knowledge base as start points. It further helps in identification of FIRs which resolves selection of true candidate genes along with the co-clustering of proteins from fenugreek transcripts with the selected start points as described for the trigonelline synthesising enzyme NNMT (section 3.4.1, Fig. 4, Supplementary Fig. 4E).

In the case of the diosgenin pathway, which is well studied in fenugreek [78], there are two enzymes which were explored in the current study. Sterol-3β-glucosyltransferase (the penultimate enzyme) is associated with four hits in fenugreek. However, based on identification of PSBD and PSPG motifs and co-clustering with sequences from closely related leguminous plants, ensured identification of two true candidate genes (Tfoe_c26467_g1_i1_m12030, Tfoe_c33087_g1_i3_m19456) (section 3.4.2). This was further supported through the TPM values observed across the tissue to ascertain the final candidate genes. Additionally, the qRT-PCR quantification also corroborated with the trend observed through the TPM values across the tissues. The last enzyme in diosgenin biosynthesis, β-glucosidase, despite presence of (I/V)TENG and TFNEP motifs in all of the associated 22 hits, only a single hit was selected as the candidate gene based on co-clustering with *M. truncatula* protein sequence. There are several transcriptome studies available in NCBI for fenugreek (BioProject IDs: PRJNA544308, PRJNA508420, PRJNA383660). We confirmed the presence of both the diosgenin pathway enzymes in their corresponding SRA datasets. We observed more than 98% sequence identity and close to 90% query coverage for the candidate genes from our study. This suggested that the candidate genes identified in our study agree well with the publicly available data for fenugreek.

Several fenugreek cultivars are widely cultivated across India. Traditionally, the AFG-1 variety, described in this study, has been one of the cultivars commonly grown across India, with bold seeds and good yield [79]. Many of the fenugreek cultivars grown across India face the powdery mildew disease, including the AFG-1 variety [80]. Ongoing work in UHS, Bagalkot, India (unpublished), on Indian fenugreek cultivars for identification of biotic stress resistant genotype, has observed that among selected cultivars, Kasuri methi showed high immune response towards the biotic stress arising from the powdery mildew disease. This could be further utilized to understand disease resistance mechanisms in fenugreek cultivars. The current transcriptome data will assist in establishing the genetic determinants for the nutritional and medicinal properties among the cultivars. It will also help us to estimate the important secondary metabolites and develop varieties with high nutraceutical value. This will further pave the way to develop and breed biotic and abiotic tolerant cultivars of fenugreek.

5. Conclusions

Fenugreek is an important multi-purpose crop of high nutritional and medicinal value commercially grown in many parts of the world. The assembly and analysis of the transcriptome of the Ajmer Fenugreek-1 (AFG-1) variety of *T. foenum-graecum* and subsequent functional annotation helped us identify enzymes involved in the biosynthesis of secondary metabolites such as alkaloids, saponins, flavonoids and volatile compounds. Some of these secondary metabolites are of immense value in the

pharmaceutical and nutraceutical industries. Diosgenin and trigonelline are known to be specific to fenugreek and are antidiabetic, antioxidant, anti-inflammatory, antiobesity along with other several medicinal values. The candidate genes encoding enzymes for synthesis of such metabolites were identified in the transcriptome data followed by their differential expression analysis across different tissues. These genes were also validated through qRT-PCR for tissue-level expression. Additionally, these metabolites were quantified across the tissues.

From the transcriptome assembly, we identified highly expressed transcripts across leaf, stem and root tissue. Among them, *metallothionein*, *dehydrin B* and *Nmr-A* and *Nmr-B* like family of genes, which are associated with drought tolerance, underlines semi-arid nature of fenugreek. Identification of genes involved in the biosynthesis of diosgenin and trigonelline in fenugreek will aid in bioengineering these enzymes for large scale production of these compounds. Further, this transcriptome can be utilised for cross comparison of several cultivars and through molecular breeding programs lead to crop improvement.

Abbreviations

Tfoe: *Trigonellafoenum-graecum*; AFG-1: Ajmer Fenugreek – 1; ORF: Open Reading Frame; TPM: Transcripts Per Million; FIR: Functionally Important Residues; qRT-PCR: quantitative Reverse Transcriptase Polymerase Chain Reaction; 4-HIL: 4-Hydroxyleucine; GO: Gene Ontology; NNMT: Nictonate N-Methyl Transferase; BADH: Betaine Aldheyde Dehydrogenase; CHMO: Choline Monooxygenase; PSBP: Putative Sterol Binding Domain; PSPG: Plant Secondary Product Glycosyltransferase, HPLC-PDA: High Performance Liquid Chromatography-Photo Diode Array; LC-MS: Liquid Chromatography-Mass Spectrometry.

Declarations

Acknowledgements: We acknowledge the initial discussions regarding the transcriptome analysis pipeline with A. Gandhimathi, DhingraSurbhi, Mathew Oommen K., Malhotra Sony, ManoharanMalini, Mutt Eshita, Shingate Prashant N., SukhwalAnshul, Sunitha Margaret S., UpadhyayAtul K. andUpadhyayulaRaghavender S. We also thank Dr. PadmaRamakrishnan from Centre for cellular and molecular platforms, Bangalore for use of mass spectrometry and HPLC facilities. We thank College of Horticulture, GKVK campus, Bangalore for plant seed materials. We also thank Department of Crop Physiology, UAS, Bangalore for use of experimental facilities.

Funding: This work was supported by the Department of Biotechnology, Government of India grant (BT/PR10550/BID/7/479/2013), the JC Bose fellowship, Science and Engineering Research Board, Department of Science and Technology, Government of India (SB/S2/JC-071/2015) and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021) to RS. We also thank NCBS (TIFR) for infrastructural and financial support.

Author's contribution:RS: Conceived and designed experiments, provided resources and tools and critically reviewed the manuscript. MBNN, SN, RSS, MI, KH, KN, SNP:Data generation. MBNN, SN, RSS, BT, GP, IM, JM, JAG, KH, KMS, KN, KSD, MB, SNP:Data analysis and presentation. MBNN, SN, RSS, BT, GP, IM, JM, JAG, KH, KMS, KN, KSD, MB, SNP and RS:Writing of manuscript.

Data availability:NCBI BioProject: PRJNA734905, BioSample: (leaf replicates: SAMN19548760, SAMN19548761; stem replicates: SAMN19548765, SAMN19548764; root replicates: SAMN19548762, SAMN19548763), SRA accession: (leaf replicates: SRR14721911, SRR14721912; stem replicates: SRR14721915, SRR14721916; root replicates: SRR14721913, SRR14721914). The Supplementary Data is available at the following link:http://caps.ncbs.res.in/download/tfoe_data/. The details are indexed in Supplementary Text file (section 2).

Competing interests: The author(s) declare that they have no competing interests.

References

1. Martin, E., Akan, H., Ekici, M. &Ayta, Z. Karyotype analyses of ten sections of Trigonella (Fabaceae). *Comparative Cytogenetics***5**, 105–121 (2011).
2. Ranjbar, M. & Zahra, H. Chromosome numbers and biogeography of the genus Trigonella (Fabaceae). *Caryologia* (2016) doi:10.1080/00087114.2016.1169090.
3. Ahmad, A., Alghamdi, S. S., Mahmood, K. & Afzal, M. Fenugreek a multipurpose crop: Potentialities and improvements. *Saudi Journal of Biological Sciences***23**, 300–310 (2016).
4. NagulapalliVenkata, K. C., Swaroop, A., Bagchi, D. &Bishayee, A. A small plant with big benefits: Fenugreek (Trigonellafoenum-graecum Linn.) for disease prevention and health promotion. *Molecular nutrition & food research***61**, (2017).
5. Srinivasan, K. Fenugreek (Trigonellafoenum-graecum): A review of health beneficial physiological effects. *Food Reviews International* (2006) doi:10.1080/87559120600586315.
6. Yadav, U. C. S. &Baquer, N. Z. Pharmacological effects of Trigonellafoenum-graecum L. in health and disease. *Pharmaceutical biology***52**, 243–254 (2014).
7. Syed, Q. A. *et al.* Nutritional and therapeutic properties of fenugreek (Trigonellafoenum-graecum): a review. *International Journal of Food Properties* (2020) doi:10.1080/10942912.2020.1825482.
8. Vaughn, C. J. Drugs and Lactation Database: LactMed. *Journal of Electronic Resources in Medical Libraries* (2012) doi:10.1080/15424065.2012.735134.
9. Yao, D. *et al.* Advances on application of fenugreek seeds as functional foods: Pharmacology, clinical application, products, patents and market. *Critical reviews in food science and nutrition***60**, 2342–2352 (2020).
10. Jesus, M., Martins, A. P. J., Gallardo, E. & Silvestre, S. Diosgenin: Recent Highlights on Pharmacology and Analytical Methodology. *Journal of Analytical Methods in Chemistry***2016**, 4156293 (2016).

11. Zhou, J., Chan, L. & Zhou, S. Trigonelline: a plant alkaloid with therapeutic potential for diabetes and central nervous system disease. *Current medicinal chemistry***19**, 3523–3531 (2012).
12. Broca, C. *et al.* 4-Hydroxyisoleucine: experimental evidence of its insulinotropic and antidiabetic properties. *American Journal of Physiology-Endocrinology and Metabolism***277**, E617–E623 (1999).
13. Avalos-Soriano, A., de la Cruz-Cordero, R., Rosado, J. L. & Garcia-Gasca, T. 4-Hydroxyisoleucine from Fenugreek (Trigonellafoenum-graecum): Effects on Insulin Resistance Associated with Obesity. *Molecules (Basel, Switzerland)***21**, (2016).
14. Tewari, D. *et al.* Fenugreek (Trigonellafoenum-graecum L.) Seeds Dietary Supplementation Regulates Liver Antioxidant Defense Systems in Aging Mice. *Nutrients* vol. 12 (2020).
15. Choudhary, S. *et al.* Transcriptome profiling of coriander: a dual purpose crop unravels stem gall resistance genes. *Journal of Genetics* (2019) doi:10.1007/s12041-019-1064-7.
16. Choudhary, S., Naika, M. B. N. &Meena, R. D. Development and characterization of genic SSR-FDM for stem gall disease resistance in coriander (Coriandrumsativum L.) and its cross species transferability. *Molecular Biology Reports* (2021) doi:10.1007/s11033-021-06396-9.
17. Pasha, S. N. *et al.* The transcriptome enables the identification of candidate genes behind medicinal value of Drumstick tree (Moringaoleifera). *Genomics* (2020) doi:10.1016/j.ygeno.2019.04.014.
18. Bolger, A. M., Lohse, M. &Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)***30**, 2114–2120 (2014).
19. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols***8**, 1494–1512 (2013).
20. Nishimura, O., Hara, Y. &Kuraku, S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics***33**, 3635–3637 (2017).
21. Altschul, S. F. BLAST Algorithm. in *Encyclopedia of Life Sciences* (2005). doi:10.1038/npg.els.0005253.
22. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research***49**, D480–D489 (2021).
23. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research***39**, (2011).
24. Sonnhammer, E. L. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function and Genetics***28**, 405–420 (1997).
25. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods***8**, 785–786 (2011).
26. Krogh, A., Larsson, B., von Heijne, G. &Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* (2001) doi:10.1006/jmbi.2000.4315.
27. Lagesen, K. *et al.*RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research***35**, 3100–3108 (2007).

28. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* (2000) doi:10.1038/75556.
29. Huerta-Cepas, J. *et al.*eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research***47**, D309–D314 (2019).
30. Kanehisa, M. &Goto, S. KEGG: kyotoencyclopedia of genes and genomes. *Nucleic acids research***28**, 27–30 (2000).
31. Li, L., Stoeckert, C. J. &Roos, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research***13**, 2178–2189 (2003).
32. Goodstein, D. M. *et al.*Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research***40**, (2012).
33. Lechner, M. *et al.*Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics***12**, 124 (2011).
34. Katoh, K. &Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution***30**, 772–780 (2013).
35. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics***12**, 323 (2011).
36. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. &Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology* (2011) doi:10.1186/gb-2011-12-3-r22.
37. Langmead, B. &Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* (2012) doi:10.1038/nmeth.1923.
38. Joshi, A. G. *et al.* A knowledge-driven protocol for prediction of proteins of interest with an emphasis on biosynthetic pathways. *MethodsX* (2020) doi:10.1016/j.mex.2020.101053.
39. Schläpfer, P. *et al.* Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiology***173**, 2041–2059 (2017).
40. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research***42**, (2014).
41. Mizuno, K. *et al.* Conversion of nicotinic acid to trigonelline is catalyzed by N-methyltransferase belonged to motif B' methyltransferase family in Coffeaarabica. *Biochemical and biophysical research communications***452**, 1060–1066 (2014).
42. Li, W. *et al.* A Novel N-Methyltransferase in Arabidopsis Appears to Feed a Conserved Pathway for Nicotinate Detoxification among Land Plants and Is Associated with Lignin Biosynthesis .*Plant Physiology***174**, 1492–1504 (2017).
43. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology***7**, 539 (2011).
44. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution***35**, 1547–1549 (2018).

45. Rambaut, A. FigTree v1.4.2, A Graphical Viewer of Phylogenetic Trees. *Available from http://tree.bio.ed.ac.uk/software/figtree/* (2014).
46. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)***23**, 127–128 (2007).
47. Joshi, R. & Karan, R. Physiological, biochemical and molecular mechanisms of drought tolerance in plants. in *Molecular Approaches in Plant Abiotic Stress* (2013). doi:10.1201/b15538-16.
48. Yang, Z., Wu, Y., Li, Y., Ling, H. Q. & Chu, C. OsMT1a, a type 1 metallothionein, plays the pivotal role in zinc homeostasis and drought tolerance in rice. *Plant Molecular Biology* (2009) doi:10.1007/s11103-009-9466-1.
49. Lopez, C. G., Banowetz, G. M., Peterson, C. J. &Kronstad, W. E. Dehydrin expression and drought tolerance in seven wheat cultivars. *Crop Science* (2003) doi:10.2135/cropsci2003.0577.
50. Hassan, N. M., El-Bastawisy, Z. M., El-Sayed, A. K., Ebeed, H. T. &NematAlla, M. M. Roles of dehydrin genes in wheat tolerance to drought stress. *Journal of Advanced Research* (2015) doi:10.1016/j.jare.2013.11.004.
51. Waller, G. R. &Nowacki, E. K. The Role of Alkaloids in Plants BT - Alkaloid Biology and Metabolism in Plants. in (eds. Waller, G. R. &Nowacki, E. K.) 143–181 (Springer US, 1978). doi:10.1007/978-1-4684-0772-3_5.
52. Zhou, J., Chan, L. & Zhou, S. Trigonelline: A Plant Alkaloid with Therapeutic Potential for Diabetes and Central Nervous System Disease. *Current Medicinal Chemistry* (2012) doi:10.2174/092986712801323171.
53. Zhao, G. *et al.* Betaine in inflammation: Mechanistic aspects and applications. *Frontiers in Immunology* (2018) doi:10.3389/fimmu.2018.01070.
54. Lee, J. E. *et al.* Choline and betaine intake and the risk of colorectal cancer in men. *Cancer Epidemiology Biomarkers and Prevention* (2010) doi:10.1158/1055-9965.EPI-09-1295.
55. Sun, J., Wen, S., Zhou, J. & Ding, S. Association between malnutrition and hyperhomocysteine in Alzheimer's disease patients and diet intervention of betaine. *Journal of Clinical Laboratory Analysis* (2017) doi:10.1002/jcla.22090.
56. Hassanein, R. A. Effect of heat shock on some biochemical and molecular criteria of fenugreek (Trigonellafoenum-graceum L.). *Journal of Medicinal Plants Research* (2012) doi:10.5897/jmpr11.1624.
57. Rathinasabapathi, B. *et al.* Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: Prosthetic group characterization and cDNA cloning. *Proceedings of the National Academy of Sciences of the United States of America* (1997) doi:10.1073/pnas.94.7.3454.
58. Wink, M. Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany* (2013) doi:10.1016/j.sajb.2013.06.006.
59. Giridhar, K., Surya Kumari, S., Rajani, A., Sarada, C. &Naram Naidu, L. Identification of potential genotypes of fenugreek in rainfedvertisols for yield and diosgenin content. *Indian Journal of*

Agricultural Research (2016) doi:10.18805/ijare.v0iOF.8603.

60. Ciura, J., Szeliga, M., Grzesik, M. &Tyrka, M. Next-generation sequencing of representational difference analysis products for identification of genes involved in diosgenin biosynthesis in fenugreek (Trigonellafoenum-graecum). *Planta* (2017) doi:10.1007/s00425-017-2657-0.
61. Ramirez-Estrada, K. *et al.* Tomato UDP-glucose sterol glycosyltransferases: A family of developmental and stress regulated genes that encode cytosolic and membrane-associated forms of the enzyme. *Frontiers in Plant Science* (2017) doi:10.3389/fpls.2017.00984.
62. Chuenchor, W. *et al.* Structural Insights into Rice BGlu1 β-Glucosidase Oligosaccharide Hydrolysis and Transglycosylation. *Journal of Molecular Biology* (2008) doi:10.1016/j.jmb.2008.01.076.
63. Wani, S. A. & Kumar, P. Fenugreek: A review on its nutraceutical properties and utilization in various food products. *Journal of the Saudi Society of Agricultural Sciences***17**, 97–106 (2018).
64. Al-Daghri, N. M. *et al.* Fenugreek extract as an inducer of cellular death via autophagy in human T lymphoma Jurkat cells. *BMC Complementary and Alternative Medicine* (2012) doi:10.1186/1472-6882-12-202.
65. Yogalakshmi, B., Viswanathan, P. &Anuradha, C. V. Investigation of antioxidant, anti-inflammatory and DNA-protective properties of eugenol in thioacetamide-induced liver injury in rats. *Toxicology***268**, 204–212 (2010).
66. Rauscher, F. M., Sanders, R. A. & Watkins, J. B. Effects of isoeugenol on oxidative stress pathways in normal and streptozotocin-induced diabetic rats. *Journal of Biochemical and Molecular Toxicology* (2001) doi:10.1002/jbt.13.
67. Pereira, I., Severino, P., Santos, A. C., Silva, A. M. &Souto, E. B. Linalool bioactive properties and potential applicability in drug delivery systems. *Colloids and Surfaces B: Biointerfaces* (2018) doi:10.1016/j.colsurfb.2018.08.001.
68. Koeduka, T. *et al.* Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proceedings of the National Academy of Sciences of the United States of America* (2006) doi:10.1073/pnas.0603732103.
69. Raguso, R. A. &Pichersky, E. A day in the life of a linalool molecule: Chemical communication in a plant-pollinator system. Part 1: Linalool biosynthesis in flowering plants. *Plant Species Biology* (1999) doi:10.1046/j.1442-1984.1999.00014.x.
70. Carrau, F. M. *et al.* De novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts. *FEMS Microbiology Letters* (2005) doi:10.1016/j.femsle.2004.11.050.
71. Panche, A. N., Diwan, A. D. & Chandra, S. R. Flavonoids: an overview. *Journal of nutritional science***5**, e47–e47 (2016).
72. Du, Y., Chu, H., Chu, I. K. & Lo, C. CYP93G2 Is a Flavanone 2-Hydroxylase Required for C-Glycosylflavone Biosynthesis in Rice .*Plant Physiology***154**, 324–333 (2010).
73. Kerscher, F. & Franz, G. Biosynthesis of Vitexin and Isovitexin: Enzymatic Synthesis of the C-GlucosylflavonesVitexin and Isovitexin with an Enzyme Preparation from Fagopyrumesculentum M. Seedlings. *ZeitschriftfürNaturforschung C***42**, 519–524 (1987).

74. Lucci, N. &Mazzafera, P. Rutin synthase in fava d'anta: purification and influence of stressors. *Canadian Journal of Plant Science***89**, 895–902 (2009).
75. Paramesha, M., Priyanka, N., Crassina, K. & Shetty, N. P. Evaluation of diosgenin content from eleven different Indian varieties of fenugreek and fenugreek leaf powder fortified bread. *Journal of Food Science and Technology* (2021) doi:10.1007/s13197-021-04967-z.
76. Hassanzadeh, E., Chaichi, M. R., Mazaheri, D., Rezazadeh, S. &Badi, H. A. N. Physical and chemical variabilities among domestic Iranian Fenugreek (Trigonellafoenum-graceum) seeds. *Asian Journal of Plant Sciences* (2011) doi:10.3923/ajps.2011.323.330.
77. Upadhyay, A. K. *et al.* Genome sequencing of herb Tulsi (Ocimumtenuiflorum) unravels key genes behind its strong medicinal properties. *BMC Plant Biology* (2015) doi:10.1186/s12870-015-0562-x.
78. Ciura, J., Szeliga, M., Grzesik, M. &Tyrka, M. Next-generation sequencing of representational difference analysis products for identification of genes involved in diosgenin biosynthesis in fenugreek (Trigonellafoenum-graecum). *Planta***245**, 977–991 (2017).
79. ICAR-NRCSS. List of varieties/technologies developed at ICAR-NRCSS.
80. Rathi, A. S., Gupta, P. P. &Jhorar, B. S. Yield losses due to powdery mildew disease in fenugreek (Trigonellafoenum-graecum L.). *Journal of Spices and Aromatic Crops***11**, 143–145 (2002).

Figures

Figure 1

A snapshot of selected secondary metabolites produced by *Trigonellafoenum-graecum* (shown in image is the AFG-1 variety). The abundance of these compounds make fenugreek a repertoire of nutraceutical and pharmaceutical properties.

Figure 2

Orthology analyses A) OrthoMCL: Venn diagram representing the common and unique orthogroups for the five proteomes studied. B) Proteinortho results showing the distribution of orthogroups between few selected plants from the subclass Rosids (including *Arabidopsis thaliana* and few plants from Faboideae and Rosaceae) and *Oryzasativa*. Each coloured section represents the number of species with which the orthogroups are shared

Figure 3

Heatmap for top 20 highly expressed transcripts from leaf, stem and root tissues

Figure 4

Trigonellinebiosynthesis pathway and expression of enzyme (represented as heatmap in green), Nicotinate N-methyltransferase (log2 scale) in each tissues.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tif
- SupplementaryFigure2.tif
- SupplementaryFigure3.tif
- SupplementaryFigure4.tif
- SupplementaryFigure5.tif
- SupplementaryFigure6.tif
- SupplementaryFigure7.tif
- tfoeScientificReportssupplementarytext.docx