Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Characterizing the circadian connectome of Ocimum tenuiflorum using an integrated network theoretic framework

**Vikram Singh**
Central University of Himachal Pradesh School of Life Sciences
**Vikram Singh** ( vikramsingh@cuhimachal.ac.in )
Central University of Himachal Pradesh https://orcid.org/0000-0002-7644-4084

**Research Article**

**Keywords:** Random walk with restart (RWR), Graphlet degree vector (GDV), Circadian rhythms, Protein interaction network (PIN), Ocimum tenuiflorum, Tulsi.

**Posted Date:** July 6th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1691807/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Characterizing the circadian connectome of *Ocimum tenuiflorum* using an integrated network theoretic framework

Vikram Singh† and Vikram Singh*

Centre for Computational Biology and Bioinformatics, Central University of Himahcal Pradesh, Dharamshala, 176206, Himachal Pradesh, India.

*Corresponding author(s). E-mail(s): vikramsingh@cuhimachal.ac.in;
Contributing authors: vikram.singh7571@gmail.com;

**Abstract**

Across the three domains of life, circadian clock is known to regulate vital physiological processes, like, growth, development, defence *etc.* by anticipating environmental cues. In this work, we report an integrated network theoretic methodology comprising of random walk with restart (RWR) and graphlet degree vectors (GDV) to characterize genome wide core circadian clock and clock associated raw candidate proteins in a plant for which protein interaction information is available. As a case study, we have implemented this framework in *Ocimum tenuiflorum* (Tulsi); one of the most valuable medicinal plants that has been utilized since ancient times in the management of a large number of diseases. For that, **24** core clock (CC) proteins were mined in **56** template plant genomes to build their hidden Markov models (HMMs). These HMMs were then used to identify **24** core clock proteins in *O. tenuiflorum*. The local topology of the interologous Tulsi protein interaction network (TulsiPIN) was explored to predict the CC associated raw candidate proteins. Statistical and biological significance of the raw candidates was determined using permutation and enrichment tests. A total of **66** putative CC associated proteins were identified and their functional annotation was performed.

**Keywords:** Random walk with restart (RWR), Graphlet degree vector (GDV), Circadian rhythms, Protein interaction network (PIN), *Ocimum tenuiflorum*, Tulsi.

**Key message:** An integrated method utilizing specific features of proteins' sequences and their interactome is proposed to characterize genome wide core circadian clock (CC) and circadian clock associated (CCa) candidate proteins in plants

# 1 Introduction

The earth's rotation around its own axis and revolution around the sun generates repetitive changes in the environmental conditions, like, dark-light cycle, rhythmicity of light intensity, temperature, seasons etc. Guided by these environmental rhythms, in living organisms exist biological rhythms called circadian rhythms which are endogenously generated, self sustained, and temperature compensated, and have a period approximately equal to 24 hours in free running conditions (Alabadí et al., 2001; Harmer et al., 2000; McClung, 2006; R.B. Singh, Singh, & Ramaswamy, 2010). These rhythms directly influence the physiological processes of all the living beings including plants that are helpful in diurnal modulation of biological events (Greenham & McClung, 2015). The involvement of circadian clock in regulating vital plant processes, like, nutrient biogenesis, root growth, photosynthesis, hormonal signalling, sugar metabolism, flowering time, and plant immunity has been documented in a large number of studies (Covington, Maloof, Straume, Kay, & Harmer, 2008; Dodd et al., 2005; Lu, McClung, & Zhang, 2017; Wang et al., 2011; Yakir, Hilman, Harir, & Green, 2007). These biological rhythms are entrained to their external environment by an internal time keeper called circadian clock that keeps track of environmental fluctuations and allow temporal separation of internal biological events accordingly (Hotta et al., 2007). Circadian clock is crucial for optimizing plant's adaptive responses, in different physiological processes including growth, development and defence, to varying environmental conditions and improve overall fitness of the plant (Dodd et al., 2005; Yerushalmi & Green, 2009).

Recent advancements in high-throughput next generation sequencing platforms have revolutionized the practice of modern biology by generating a huge volume of molecular sequence data (Sayers et al., 2021). Since the expression of a large proportion of an organism's genes is circadian controlled, these genome scale technologies allow us to characterize core circadian clock (CC) and clock associated genes at systems scale (Hughes et al., 2017). Various statistical approaches including LS (Glynn, Chen, & Mushegian, 2006), ARSER (Yang & Su, 2010), JTK_CYCLE (Hughes, Hogenesch, & Kornacker, 2010), RAIN (Thaben & Westermark, 2014), MetaCycle (Wu, Anafi, Hughes, Kornacker, & Hogenesch, 2016), BIO_CYCLE (Agostinelli, Ceglia, Shahbaba, Sassone-Corsi, & Baldi, 2016) for characterizing oscillatory genes along with their estimated rhythmic parameters *i.e.* phase, amplitude and waveform have been proposed (Mei et al., 2021). All these approaches utilize temporal transcriptomic data to characterize circadian regulated genes and estimate their rhythmic parameters, however, in this study we report a novel integrated network framework that completely rely on network topology to characterize rhythmic genes in plants. It has been shown that circadian oscillators are ubiquitous among all the three domains of life with a very little elemental (gene) conservation across them, however, their architectures and mechanisms are remarkably convergent across domains (Edgar et al., 2012; Eelderink-Chen et al., 2021; McClung, 2013). Furthermore, topologically similar proteins in protein interaction networks (PINs) have been shown to perform similar functions (Davis, Yaveroğlu, Malod-Dognin, Stojmirovic, & Pržulj, 2015).

Several approaches have shown that architectural properties of proteins in an interactome complement their structural and sequence properties, and could be successfully utilized for inferring the functional annotations of interacting proteins (Milenković & Pržulj, 2008; Nabieva, Jim, Agarwal, Chazelle, & Singh, 2005; Sharan & Ideker, 2006). This motivated us to exploit the topology of protein interaction network to identify circadian regulated genes.

Recently, genome (Rastogi et al., 2015; Upadhyay et al., 2015), tanscriptome (Rastogi et al., 2014) and proteome (V. Singh, Singh, & Singh, 2019) wide studies have reported a draft catalogue of genes, transcripts and proteins present in *Ocimum tenuiflorum* (Tulsi) along with important mechanistic insights about the secondary metabolite biosynthetic pathways. However, an exhaustive systems scale study presenting core circadian clock and clock associated genes along with complex interplay among them at molecular level is still lacking in *O. tenuiflorum*. Tulsi is a member of family Lamiaceae, an aromatic herb that is widely recognized for array of health benefits it offers (Cohen, 2014). We find references of the holy basil in Ayurveda, an ancient medicinal system, for its usage in treating variety of diseases and disorders either independently or as a formulation in combination with several other herbs (Pattanayak, Behera, Das, & Panda, 2010). The plant is reared for its oils and every part of this plant produces a number of secondary metabolites which possess significant pharmaceutical properties (A.K. Pandey, Singh, & Tripathi, 2014).

In this study, we propose a systems scale approach to mine and characterize core circadian clock proteins giving rise to these rhythms in *O. tenuiflorum*. We leveraged random walk with restart (RWR) and graphlet degree vector (GDV) methods to explore the local topology of protein interaction network of Tulsi (TulsiPIN) and identified raw candidate proteins associated with the core oscillator. Furthermore, statistical and biological significance of the raw candidates was determined and their functional annotations were determined.

# 2 Materials and methods

## 2.1 Data collection

From the literature, 24 core circadian clock (CC) proteins of *Arabidopsis thaliana* (Lu et al., 2017; Nohales & Kay, 2016) were identified and their sequences were retrieved from the TAIR database (Lamesch et al., 2012). Proteome data of template 56 template plants, for which genome information is available, was downloaded from UniProt database https://www.uniprot.org/. All the CC proteins were then queried against each template proteome using Blast-P to identify orthologous proteins. Top hits with at least 40% sequence identity, 50% query coverage, and an *E-value* $\leq 10^{-5}$ were selected for every sequence, and were aligned using Clustal Omega (Sievers et al., 2011). HMMER http://hmmer.janelia.org/ was used to construct Hidden Markov models for each CC protein. These HMMs were then used to scan *O. tenuiflorum* proteome obtained from TulsiDB http://caps.ncbs.res.in/Ote/ to identify putative CC proteins in Tulsi. Each identified CC protein was subjected to InerProScan v5.48-83 https://www.ebi.ac.uk/interpro/ to identify their respective domains (G. Singh, Singh, & Singh, 2022). Protein-protein interaction network of *O. tenuiflorum* called TulsiPIN developed in our previous work (V. Singh et al., 2019) was considered for network scale studies. TulsiPIN is an exhaustively constructed network consisted of 13,604 nodes and 327,374 edges that possess both scale free and small world properties.

## 2.2 Mining core circadian clock associated (CCa) proteins using local topological measures

Two local network topology based algorithms, namely, random walk with restart (RWR) and graphlet degree vector (GDV) were leveraged to identify core circadian clock associated proteins.

### 2.2.1 Random walk with restart (RWR) based mining

RWR is a ranking algorithm that has been successfully applied to estimate the likelihood of genes/proteins being associated with a given biological process (X) based on their proximity to already known proteins belonging to X in the input network (Köhler, Bauer, Horn, & Robinson, 2008). The algorithm initiates walks from one or more seed nodes (CC proteins) to some random node(s) in the network and returns the probability of each protein being a CCa protein. In this study, RWR algorithm was applied on TulsiPIN using 9 CC proteins (which could be mapped to TulsiPIN) as seed nodes. Initially a vector $P_0$, containing $13,604$ elements each corresponding to a node in TulsiPIN, was constructed in which each CC protein was initialized with a number $\frac{1}{9}$ and other proteins with zero. The vector is then updated using the equation $P_{i+1} = (1-r)A^T P_i + rP_0$ for $i^{th}$ iteration until $P_i$ is stabilized. Here, $r$ represents restart probability that is set as (0.8), and $A$ is TulsiPIN's adjacency matrix every column of that sums up to one. The iterations stop when Manhattan distance between $P_i + 1$ and $P_i$ is less than $10^{-6}$ *i.e.* $P_{i+1} - P_{iL_1} < 10^{-6}$. This yielded a probability vector containing the probability values corresponding to each protein of TulsiPIN quantifying their likelihood of being CC associated. Accordingly, a threshold value of $10^{-5}$ was adopted and all the proteins having probabilities higher than the selected threshold, called raw CCa proteins, were considered for further study (Chen et al., 2015).

### 2.2.2 Graphlets degree vector (GDV) based mining

Small, non-isomorphic, induced sub-graphs of a large network are defined as graphlets (Pržulj, Corneil, & Jurisica, 2004). Based on the node position within a graphlet, each sub-graph can be differentiated into symmetrically equivalent sets of nodes (symmetry groups) called automorphism orbits (Pržulj, 2007). A total of 15 automorphism orbits are possible for up to 4 node graphlets. Each node of an undirected network can be represented with a vector of length 15, each element of which represents the degree of corresponding orbit, called graphlet degree vector (GDV). Finally a vector of 11 non-redundant orbits (Yaveroğlu et al., 2014) containing orbit degrees, representing wiring around that node, was computed for every node in the network. Since current PPI networks have been shown to be largely incomplete and lower degree nodes have higher likelihood of being present in those incomplete parts, so we remove proteins with degree lower than three (Milenković & Pržulj, 2008). Among the 24 CC proteins only 10 could be mapped to TulsiPIN. Since one of them was having degree lower than three so removed from the further study. For each of 9 CC proteins, a pairwise similarity ($S_{GDV}$) was computed with every TulsiPIN protein having degree more than three that is defined as

$$S_{GDV} = 1 - \frac{\sum |u_i - v_i|}{\sum |u_i + v_i|}$$

where $u$ and $v$ represent GDV of CC proteins and TulsiPIN proteins respectively, while $i$ represents the $i^{th}$ orbit. All the proteins having similarity above 94% with the protein of interest were considered as CCa proteins. Previously a signature similarity threshold within the range 90% and 95% has been shown to be optimal (Milenković & Pržulj, 2008). So to select the upper signature similarity threshold, we started with a threshold value of 96% and decreased the threshold value until we obtained at least one TulsiPIN protein similar with each CC protein other than themselves.

## 2.3 Statistical and functional validation of CCa proteins

As these algorithms exploit the topology of the network to estimate the likelihood, this may output some false positive proteins as CCa proteins *e.g.* hubs will be ranked higher than those having degree lower than the average degree of network (Chen et al., 2015). To reduce these false positive proteins and assess the reliability of each prediction, two methods, namely, permutation test and enrichment test were applied.

### 2.3.1 Permutation or randomisation test

To test the statistical significance of CCa proteins identified by RWR algorithm, we sampled 1000 random sets of 9 proteins each from 13, 604 TulsiPIN proteins. RWR algorithm was then performed on TulsiPIN using each of these samples as seed. A *p-value* or permutation FDR for every CCa proteins predicted was computed based on the following equation $p(g) = \frac{\theta_{RWR}}{1000}$, where $\theta$ is the number of random sets for which probability of $g$ is higher than $p(g)$ obtained using CC proteins as seed. On the other hand to test the statistical significance of CCa proteins obtained from GDV algorithm, we first constructed an ensemble of 1000 random networks with same number of nodes and similar degree sequence (Miller & Hagberg, 2011) and then computed graphlet degree vectors for each protein of every network. Then pairwise $S_{GDV}$ similarity scores were computed for each of the CC protein with other proteins in TulsiPIN. Similar to RWR, a permutation FDR for every CCa protein predicted by GDV algorithm was computed using $p(g) = \frac{\theta_{GDV}}{1000}$, where $\theta$ is the number of random networks for which the similarity score of protein pair between $g$, a CCa gene, and some CC gene is higher than the similarity score obtained for same pair in real network. Higher *p-value* of a gene indicates the gene is less likely to be associated with CC. Thus CCa proteins with *p-value* less than 0.05 were selected, from both methods, and the selected proteins were called candidate CCa proteins for convenience.

### 2.3.2 Enrichment test

Enrichment test is based on the premise that interacting proteins are more likely to share common functions (G. Singh, Singh, & Singh, 2021; V. Singh et al., 2019). So, in this section, we used PPI information form TulsiPIN and Gene Ontology (GO) and KEGG Orthology (KO) annotation information of interacting proteins to infer the validity of candidate CC associated proteins. It has been shown earlier that similar proteins may perform similar molecular functions, localize in same cellular milieu, and may be involved in similar biological processes (G. Singh et al., 2021; V. Singh et al., 2019). Thus candidate CCa proteins having similar functional annotations as that of

CC proteins may be the novel CCa proteins. To quantify the relationship between a protein and an annotation term (GO or KEGG pathways), ontology enrichment

score that is the negative log of probability value obtained from hypergeometric test was computed (Carmona-Saez, Chagoyen, Tirado, Carazo, & Pascual-Montano, 2007). Given $N$ total TulsiPIN proteins among which $M$ are annotated by one annotation term ($A$), if we randomly draw a set of proteins ($H(g)$) containing any protein $g$ and its immediate neighbors then the probability that at least $m$ of them are annotated with $A$ is given by

$$S(g, A) = -log_{10}\left(\sum_{k=m}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\right)$$

where $n$ is the number of proteins present in set $H(g)$ and $m$ is the number of $H(g)$ proteins annotated by $A$. A vector $ES(g)$ can be created by computing $S$ for all the annotation terms of $g$. These vectors can be used to quantify the similarity between two genes $g$ and $g'$ as follows

$$\Gamma(g, g') = \frac{ES(g).ES(g')}{ES(g).ES(g')}$$

Large value of $\Gamma$ indicate high functional similarity between the two genes $g$ and $g'$. We compute maximum annotation similarity (MAS), that represents the largest $\Gamma$ value among all possible pairs between a candidate protein $p$ and set of CC genes. A threshold value of 0.75 was used to select candidate CCa proteins.

## 2.4 Functional annotation of CCa proteins

Each protein identified to be associated with core circadian clock of *O. tenuiflorum* was then queried against NCBI non-redundant (nr), UniProt (Apweiler et al., 2004) and TAIR (Berardini et al., 2015) databases using Blast-P (Altschul, Gish, Miller, Myers, & Lipman, 1990) program to assign putative functions. Default parameters except *E-value*, that is set to be $10^{-5}$, were used to assign homology. AgriGo (Du, Zhou, Ling, Zhang, & Su, 2010) and WEGO (Ye et al., 2006) were used to predict Gene Ontology (GO) terms for each CCa protein. KEGG ontology (KO) terms were predicted for each CCa protein using KEGG (Kanehisa & Goto, 2000) database. All the 13 programs aggregated in InterProScan v5.48-83 were utilised to predict protein family, functional domains and sites in each CCa protein (Jones et al., 2014).

# 3 Results and Discussion

*O. tenuiflorum* plants have been reported to produce secondary metabolites (essential oils) which are mixtures of diverse aromatic compounds. These chemical compounds possess enormous medicinal properties that are being exploited continuously since ancient times (Cohen, 2014). Under varying environmental conditions, plants are continuously exposed to biotic as well as abiotic stress that hamper plant growth and development, and have significant impact on the oil production (P. Pandey, Irulappan, Bagavathiannan, & Senthil-Kumar, 2017; Rastogi et al., 2019). Circadian clock is the main regulatory mechanism that controls the growth and development of any plant (Dodd et al., 2005). A precisely tuned circadian clock plays substantial role in increasing the fitness and survival of a plant by temporally coordinating various vital processes, like, metabolism, physiology *etc.* in accordance with the external conditions (Dodd et al., 2005; Lu et al., 2017). Thus, we attempt to identify the protein set comprising the core circadian clock and its associated proteins in *O. tenuiflorum* through an exhaustive network science based computational framework presented in Figure 1.

## 3.1 Mining and functional characterization of core circadian clock (CC) proteins

From literature, we enlisted 24 core circadian clock proteins and attempt to identify them in *Ocimum tenuiflorum*. Proteomes of 56 template plants were scanned using Blast-P for putative ortholgs of selected clock proteins that resulted in 2,543 top hits at an *E-value* $< 10^{-5}$. Further, top hits with sequence identity $< 40\%$ and query coverage $< 50\%$ were filtered out resulting in 963 hits. Separate HMM profiles for each of the CC protein were built by extracting orthologous sequences of that protein from the respective template plants which satisfy the above criteria. These HMMs were used to scan *O. tenuiflorum* proteome, and on the based of top hits, all the 24 candidate CC proteins were identified (Table S1). Transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LONG ELONGATED HYPOCOTYL (LHY) are partially redundant (Mizoguchi et al., 2002), also the transcription regulator LIGHT-REGULATED WD1 (LWD1) is partially redundant with LWD2, LUX ARRHYTHMO (LUX) with NOX, REVEILLE 6 (RVE6) with RVE8 (McClung, 2011; Nohales & Kay, 2016), so the best hits returned the same proteins for them. These genes are important because they express during different times of the day or form complexes to regulate the expression of other genes and their functions are synergistic (Lu, Knowles, Andronis, Ong, & Tobin, 2009; McClung, 2011), so we selected the second best hit for LWD2, RVE4, PRR5, RVE6, NOX, PSEUDO-RESPONSE REGULATOR 3 (PRR3) and LHY proteins. Functions of all the 24 CC proteins were predicted by querying them against NCBI non-redundant, UniProt and TAIR databases using Blast-P. The results confirmed the identities of predicted clock proteins and show that most of the proteins possess their respective functions (Table S2). Furthermore, protein family and domain databases integrated by InterProScan also confirmed our results (Table S2).

Functional annotation of CC proteins is followed by their gene ontology (GO) annotations (Table S2). A total of 86 GO terms were successfully identified for these CC proteins; with 67 terms classified into 13 categories of biological processes, 2 terms were classified into 2 categories of cellular components and 17 GO terms were associated with 2 categories of molecular functions. 'Regulation of biological process', 'Cellular process', 'Response to stimulus' are highly enriched among biological processes followed by 'signaling', 'rhythmic process'. Among molecular functions, 'binding' and 'transcription regulator activity' were found to be the most enriched categories (Fig. 3). Successful assignment of "rhythmic process" associated gene ontology (GO) terms to the identified CC proteins further confirms the role of these proteins as core circadian clock of *O. tenuiflorum*. Furthermore, for finding the CC proteins regulated pathways, KEGG database was used that predicted KO ids of 16 proteins enriching 8 pathways, of these 'Circadian rhythm in plants' is found to be most enriched (Table S2). Other pathways include 'Ubiquitin system' followed by 'Plant hormone signal transduction' and 'Transcription factors'.

## 3.2 Prediction and functional characterization of core circadian clock (CCa) associated proteins

It is known that different centrality measures address different topological aspects of a network and therefore capture the snapshot of a network, in terms of selected architectural metric (Lü et al., 2016). For example, connectivity-based measures, assigns higher weights to high degree nodes, on the other hand, middle nodes connecting

communities will be ranked higher by measures employing betweenness centrality. Many graph theoretic methods assign functions to unannotated proteins based on their closeness or shared neighborhood in the PINs (Davis et al., 2015; Nabieva et al., 2005; Vazquez, Flammini, Maritan, & Vespignani, 2003). However, it has been shown that functional similarities among proteins in a PIN are independent of their network location *i.e.* sharing same local neighborhood rather it depends on similarities of the wiring patterns around them (Milenković & Pržulj, 2008). So, in this work, two local network topology based algorithms, namely, random walk with restart (RWR) and graphlet degree vector (GDV) have been leveraged to identify CCa proteins. GDV based approach does not consider only the current node rather it takes into consideration the wiring pattern around that node *i.e.* it incorporates the information about local neighbourhood (Pržulj, 2007). In order to find the circadian clock associated proteins all the identified proteins were mapped to genome wide interologous protein interaction network of Tulsi (TulsiPIN) (V. Singh et al., 2019). A total of 10 proteins were successfully mapped to TulsiPIN proteins which were found to interact with 248 other proteins and form 317 interactions. Since one protein has degree lower than 3 so removed from the study. Remaining nine CC proteins were used as seed nodes to the RWR algorithm applied on TulsiPIN. This produced a probability value corresponding to each protein representing its likelihood of being a novel circadian clock associated protein. All the proteins having probability value larger than $10^{-5}$ were considered for further study resulting in 687 nodes called $\text{raw}_{RWR}$ proteins (Table S3). Because all RWR proteins can not be circadian clock related, some of them have been selected due to network architecture alone. To filter them out we employed permutation test and computed a *p-value* called the permutation FDR for every RWR protein. All the nodes having *p-values* less than 0.05 were selected resulting in 251 candidate CCa proteins called $\text{candidate}_{RWR}$ proteins and are presented in Table S3. Similarly, we computed pairwise GDV similarity between CC proteins and other TulsiPIN proteins. This resulted in 146 proteins called $\text{raw}_{GDV}$ proteins, having $S_{GDV} > 94$, predicted to be CC associated by GDV algorithm. Pairwise similarities were obtained from ensemble of 1000 random networks preserving the degree sequence of TulsiPIN. A total of 146 proteins were found to be statistically significant having *p-value* $< 0.05$ and were called as $\text{candidate}_{GDV}$ proteins (Table S3). Furthermore, to add another layer of confidence and to select only those proteins, which are having similar functions to CC proteins, we leveraged Gene Ontology and pathways information to compute the functional similarity ($\Gamma$) between each CC protein and every candidate protein. Among 251 $\text{candidate}_{RWR}$ proteins a total of 43 were found to have $\Gamma > 0.75$ in both GO and pathways information based methods. On the other hand 25 out of 146 $\text{candidate}_{GDV}$ proteins were found to have $\Gamma$ score 0.75 or above in both the methods (Table S3). We combined the proteins obtained from the two methods (RWR and GDV) and found a total of 66 unique candidate proteins among which 2 were common to both the methods. We constructed circadian connectome of Tulsi by mapping the CC proteins and CCa proteins to TulsiPIN and found that these 90 proteins regulating $2,412$ other proteins and participate in a total of $8,017$ interactions (Fig. 4).

Similar to CC proteins, all the 66 CCa proteins were queried against NCBI non-redundant, UniProt and TAIR databases using using Blast-P. The functional annotation of CCa proteins revealed that most of the proteins are involved in various aspects of growth, development and defence of *O. tenuiflorum* (Table S4). We broadly classified them into five categories, namely, Gene expression (23), metabolism (14), cell signaling (14), transport (9), post translational modification (5) and protein

folding (1). It has been shown that approximately one-third of genes in *Arabidopsis thaliana* and ≈ 89% of its transcripts are regulated by varying environmental conditions, like, temperature, light *etc* (Michael & McClung, 2003; Michael et al., 2008). These studies revealed a tight circadian regulation of *A. thaliana* genes at transcriptional level (Michael & McClung, 2003) which is consistent with our results that 20 of CC associated proteins are involved in gene expression and regulation (Fig 5).

## 3.3 Circadian regulatory pathway of *O. tenuiflorum*

The basic model of circadian regulation is comprised of a core oscillator that is a complex web of multiple coupled transcriptional and translational feedback loops (TTFL). The core oscillator perceive external cues through input pathways and temporally organise various physiological processes of the plant (Srivastava et al., 2019).

### 3.3.1 The input module of core clock

Proteins Ote100059320041, Ote100047200201 and Ote10021351017 have been predicted to be light perceiving photoreceptors namely Phytochrome A, Phytochrome B, Adagio protein 1 (ADO1) or ZEITLUPE (ZTL) protein. It is known that phytochromes primarily perceive red and far red light while cryptochromes and ZEITLUPE (ZTL) family members are blue light photoreceptors. Among the five phytochromes (PHYA-E) in *A. thaliana*, when subjected to continuous red light, PHYA and PHYB have been shown to be primary contributors to photoperception and entrainment of circadian rhythms (Somers, Devlin, & Kay, 1998). Further these signals are transduced to the core oscillator through an intermediate protein CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) (Sheerin et al., 2015). Proteins Ote237772210011 and Ote100260480091 have been characterized as an E3 ubiquitin ligase COP1. It triggers light gated proteasomal degradation of core clock components ELF3 and GI hence acts as a direct link between the clock and light perceiving photoreceptors (Sheerin et al., 2015; Yu et al., 2008). The expression of ELONGATED HYPOCOTYL5 (HY5), FAR RED ELONGATED HYPOCOTYL3 (FHY3) and FAR RED IMPAIRED RESPONSE1 (FAR1) is under tight regulation of PHYA signaling. During the day they binds directly to ELF4 promoter and induce its expression (Li et al., 2011). The ADO1 and other two members this family possess an F-Box domain, a PAS-like LOV domain and six kelch repeats. F-Box domain links target protein for proteasomal degradation by an E3 Ubiquition ligase complex called skp-cullin-F-Box that is comprised of by interactions between F-Box protein with SKP1, CULLIN, and RBX1 (Han, Mason, Risseeuw, Crosby, & Somers, 2004). It has been shown that ADO1 receive input signal by interacting with carboxy terminus of blue light photoreceptors phyB and cry1 (Jarillo et al., 2001) and mediates blue light dependent degradation of TOC1 and PRR5 through GI mediated stabilization (Kim et al., 2007). Ote100194510162 has been characterized as a serine/threonine phosphatase 5 protein. It has been shown to dephosphorylate a specific serine residue on pfr phytochromes. This dephosphorylation enhances phytochrome stability as well as increase its ability to bind downstream signal transducers and thus enhances phytochrome mediated photoresponses (Ryu et al., 2005). Proteins Ote237887910022 and Ote100030530111 have been characterized as serine/threonine phosphatase 7 that has been show to implicate in cryptochromes mediated signaling (Møller, Kim, Kunkel, & Chua, 2003). Arabidopsis's serine/threonine phosphatase

7 (AtPP7) has been shown to act as a nuclear localization signaling molecule that dephosphorylates nuclear signaling intermediates (probably unknown) when exposed to blue light. It acts in the downstream of cryptochrome and in the response of blue light ensuring stable signaling (Møller et al., 2003).

### 3.3.2 The core circadian clock

Proteins CCA1 (Ote100217520081) and LHY (Ote100050210101) represent MYB-related transcription factors that heterodimerize at dawn and repress the expression of morning phased PRR family genes including PRR9 (Ote100275800021), PRR7 (Ote100120330051) and PRR5 (Ote100011940071). In addition they repress expression of evening phased genes including TOC1 (Ote100167620091) and the evening complex (EC) members LUX (Ote100252500041), ELF3 (Ote100215050121), ELF4 (Ote100004890041) (Gendron et al., 2012; Pruneda-Paz, Breton, Para, & Kay, 2009). The PRR family members are two-component response regulators similar to ARR family His-Asp phosphorelay systems except that the aspartate site that accept phosphate group is missing. Circadian control at transcriptional level of all the members of PRR family has been documented and they have shown to express sequentially in order of PRR9, PRR7, PRR5, PRR3, and PRR1 within a period of 24*h* (Matsushika, Makino, Kojima, & Mizuno, 2000). At dawn CCA1 and LHY negatively regulate PRR1, CCA1 HIKING EXPEDITION (CHE) (Ote237864970011) and their own expression. But during the day CHE levels rise that repress CCA1 and finally at evening TOC1 reset the clock cycle by repressing the expression of CHE (Pruneda-Paz et al., 2009). Similarly, PRR7 has also been shown to repress CCA1 and LHY expressions and regulate expression of other clock components (Nakamichi et al., 2010). It acts as master regulator of various physiological processes of plant including response to abiotic stress (Liu, Carlsson, Takeuchi, Newton, & Farre, 2013), photoperiodic flowering and temperature compensation (Nakamichi et al., 2010). TOC1 (PRR1) also repress the expression of PRR5, LUX, EARLY FLOWERING 4 (ELF4) and GIGANTEA (GI) (Ote100199680211) (Nakamichi et al., 2010). EC component express from evening onwards throughout night and are required for rhythmic expression of CCA1 and LHY. LUX downregulate the expression of PRR9 and itself by binding the LUX binding sites (LBS) present in their promoters. EC is believed to indirectly upregulate the expression of CCA1 and LHY by repressing the expression of PRR9, PRR7, TOC1 (Pokhilko et al., 2012). REVEILLE family members including RVE8 (Ote100018670061), RVE4 (Ote100114610051) along with coactivators NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED 1 (LNK1) (Ote100153880201) and LNK2 (Ote100241100111) induce the expression of PRR5, TOC1, LUX and ELF4 (Farinas & Mas, 2011; Xie et al., 2014). Fig 2 presents the core clock identified in Tulsi.

### 3.3.3 The output pathways regulated by core clock

Proteins Ote100001960091 and Ote100189530101 have been identified as Ethylene receptor 1 and ethylex'ne response sensor 1, respectively. Both of which encode histidine kinases similar to members of two-component regulator family and have been shown to be integral components in ethylene signaling (Clark, Larsen, Wang, & Chang, 1998). The gaseous hormone has a large number of implications, like, fruit-ripening, senescence, stress response, seed germination and many more and its

signaling has been shown to be under circadian control (Thain et al., 2004). Protein Ote100058660011 has been predicted as Gibberellin (GA) receptor GID1B that acts as a soluble gibberellin receptor. GA is a diterpenoid hormone that regulates various stages of plant growth and development, like, seed germination, development of fruit, pollen tube and flowering time. GID1 binds with biologically active GAs like GA1, GA3 and GA4 targets a GRAS family protein DELLA that negatively regulate GA signaling for proteasomal degradation (Ueguchi-Tanaka et al., 2005). Proteins Ote100237240051 and Ote100260840121 are predicted to be UDP-N-acetylglucosamine–peptide N-acetylglucosaminyltransferase SPINDLY (SPY) that negatively regulate Gibberellin signaling to inhibit hypocotyl elongation. It has been shown that SPY and GI function together and GI acts as a light gated negative regulator of SPY function to control various physiological processes, like, elongation of hypocotyl, flowering time and cotyledon movements (Tseng, Salomeé, McClung, & Olszewski, 2004). Ote100003860081, predicted as Two-component response regulator ARR3, behaves like a bacterial two component system that negatively regulate cytokine signaling and affects the phosphorylation state of ARRs by transducing the signal through a multi-step His-Asp phosphotransfer process (To et al., 2004). Ote100257670011 has been characterized as Histidine kinase 3 protein that acts as a positive regulator of cytokine signaling (Nishimura et al., 2004). Reactive oxygen species (ROS) production, signaling and scavenging has been shown to be linked with circadian clock output and plant response to oxidative stress (Lai et al., 2012). Ote100224900091 has been characterized as ascorbate oxidase that oxidizes ascorbate to MDA and reduce $O_2$ to $H_2O$ (Gapper & Dolan, 2006). Protein Ote100214260021 has been characterized as Regulator of nonsense transcripts UPF2 that is a participates in nonsense mediated decay (NMD) pathway and regulates the degradation of mRNA containing premature termination codons (Shi, Baldwin, & Wu, 2012). It has been shown to be involved in photoperiod dependent transcriptional regulation of jasmonic acid and salicylic acid induced defence genes LOX2 and VSP2 in response to biotic and abiotic stresses (Shi et al., 2012).

# 4 Summary and Conclusions

Considering the importance of circadian rhythms in regulating physiological processes of the plant, current study reports a novel network based methodology that exploits the topological information of a network to identify circadian connectome including the core clock (CC) proteins and clock associated proteins Although we have implemented the proposed framework in *O. tenuiflorum*, however, we can easily adopt it to any other organism having an endogenous clock. It will increase our understanding of various components of circadian circuitry responsible for the circadian rhythms as well as the underlying mechanisms sustaining these rhythms which is largely unknown in Tulsi. In this study we propose a methodology to identify 24 known core circadian clock genes in Tulsi that leverage in house built HMMs from 56 template plants. Further, we propose a hybrid methodology that combines RWR and GDV local topological measures to identify novel proteins associated with core clock. Based on RWR, GDV and two statistical significance tests a total of 70 putative core clock associated genes proteins were obtained. Functional annotation of the candidate proteins was performed and we found that a large number of proteins are involved in regulation of gene expression. We obtained crucial proteins transducing red, far-red and blue light signals to the core oscillator. Furthermore, we obtained

proteins involved in different physiological process, like, gibberllin signaling, ethylene signlaing, cytokine signaling, abiotic and biotic stress response etc. We hope that the proposed framework for reporting genome-wide identification and characterization of core circadian clock and associated proteins in *O. tenuiflorum* shall pave the way for prioritizing the CC or any other biological process associated proteins in plants.

## Acknowledgments

VS[†] thanks Council of Scientific and Industrial Research (CSIR), India for providing Junior Research Fellowship (JRF). **Funding:** Authors recieved no specific funding for this research work.

## Authors Contributions

VS[*] conceptualized and designed the research framework. VS[†] performed the computational experiments. VS[†] and VS[*] analyzed the data and interpreted results. VS[†] and VS[*] wrote and finalized the manuscript.

## Competing Interests:

The authors declare that they have no conflict of interests.

## Data and materials availability:

All data is available in the manuscript or the supplementary materials.

## References

Agostinelli, F., Ceglia, N., Shahbaba, B., Sassone-Corsi, P., Baldi, P. (2016). What time is it? deep learning approaches for circadian rhythms. *Bioinformatics*, *32*(12), i8–i17.

Alabadí, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Más, P., Kay, S.A. (2001). Reciprocal regulation between toc1 and lhy/cca1 within the arabidopsis circadian clock. *Science*, *293*(5531), 880–883.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., ... others (2004). Uniprot: the universal protein knowledgebase. *Nucleic Acids Research*, *32*(suppl_1), D115–D119.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E. (2015). The arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M., Pascual-Montano, A. (2007). Genecodis: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology*, *8*(1), 1–8.

Chen, L., Chu, C., Lu, J., Kong, X., Huang, T., Cai, Y.-D. (2015). A computational method for the identification of new candidate carcinogenic and non-carcinogenic chemicals. *Molecular BioSystems*, *11*(9), 2541–2550.

Clark, K.L., Larsen, P.B., Wang, X., Chang, C. (1998). Association of the arabidopsis ctr1 raf-like kinase with the etr1 and ers ethylene receptors. *Proceedings of the National Academy of Sciences*, *95*(9), 5401–5406.

Cohen, M. (2014). Tulsi - Ocimum sanctum: A herb for all reasons. *Journal of Ayurveda and Integrative Medicine*.

10.4103/0975-9476.146554

Covington, M.F., Maloof, J.N., Straume, M., Kay, S.A., Harmer, S.L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology*, *9*(8), 1–18.

Davis, D., Yaveroğlu, Ö.N., Malod-Dognin, N., Stojmirovic, A., Pržulj, N. (2015). Topology-function conservation in protein–protein interaction networks. *Bioinformatics*, *31*(10), 1632–1639.

Dodd, A.N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F., ... Webb, A.A. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science*, *309*(5734), 630–633.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., Su, Z. (2010). agrigo: a go analysis toolkit for the agricultural community. *Nucleic Acids Research*, *38*(suppl_2), W64–W70.

Edgar, R.S., Green, E.W., Zhao, Y., Van Ooijen, G., Olmedo, M., Qin, X., ... others (2012). Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, *485*(7399), 459–464.

Eelderink-Chen, Z., Bosman, J., Sartor, F., Dodd, A.N., Kovács, Á.T., Merrow, M. (2021). A circadian clock in a nonphotosynthetic prokaryote. *Science Advances*, *7*(2), eabe2086.

Farinas, B., & Mas, P. (2011). Functional implication of the myb transcription factor rve8/lcl5 in the circadian control of histone acetylation. *The Plant Journal*, *66*(2), 318–329.

Gapper, C., & Dolan, L. (2006). Control of plant development by reactive oxygen species. *Plant Physiology*, *141*(2), 341–345.

Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., Kay, S.A. (2012). Arabidopsis circadian clock protein, toc1, is a dna-binding transcription factor. *Proceedings of the National Academy of Sciences*, *109*(8), 3167–3172.

Glynn, E.F., Chen, J., Mushegian, A.R. (2006). Detecting periodic patterns in unevenly spaced gene expression time series using lomb–scargle periodograms. *Bioinformatics*, *22*(3), 310–316.

Greenham, K., & McClung, C.R. (2015). Integrating circadian dynamics with physiological processes in plants. *Nature Reviews Genetics*, *16*(10), 598–610.

Han, L., Mason, M., Risseeuw, E.P., Crosby, W.L., Somers, D.E. (2004). Formation of an scfztl complex is required for proper regulation of circadian timing. *The Plant Journal*, *40*(2), 291–301.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.-S., Han, B., Zhu, T., ... Kay, S.A. (2000). Orchestrated transcription of key pathways in arabidopsis by the circadian clock. *Science*, *290*(5499), 2110–2113.

Hotta, C.T., Gardner, M.J., Hubbard, K.E., Baek, S.J., Dalchau, N., Suhita, D., ... Webb, A.A. (2007). Modulation of environmental responses of

plants by circadian clocks. *Plant, Cell & Environment*, *30*(3), 333–349.

Hughes, M.E., Abruzzi, K.C., Allada, R., Anafi, R., Arpat, A.B., Asher, G., ... others (2017). Guidelines for genome-scale analysis of biological rhythms. *Journal of Biological Rhythms*, *32*(5), 380–393.

Hughes, M.E., Hogenesch, J.B., Kornacker, K. (2010). Jtk_cycle: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, *25*(5), 372–380.

Jarillo, J.A., Capel, J., Tang, R.-H., Yang, H.-Q., Alonso, J.M., Ecker, J.R., Cashmore, A.R. (2001). An arabidopsis circadian clock component interacts with both cry1 and phyb. *Nature*, *410*(6827), 487–490.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., ... others (2014). Interproscan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.

Kanehisa, M., & Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.

Kim, W.-Y., Fujiwara, S., Suh, S.-S., Kim, J., Kim, Y., Han, L., ... Somers, D.E. (2007). Zeitlupe is a circadian photoreceptor stabilized by gigantea in blue light. *Nature*, *449*(7160), 356–360.

Köhler, S., Bauer, S., Horn, D., Robinson, P.N. (2008). Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, *82*(4), 949–958.

Lai, A.G., Doherty, C.J., Mueller-Roeber, B., Kay, S.A., Schippers, J.H., Dijkwel, P.P. (2012). Circadian clock-associated 1 regulates ros homeostasis and oxidative stress responses. *Proceedings of the National Academy of Sciences*, *109*(42), 17129–17134.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., ... others (2012). The arabidopsis information resource (tair): improved gene annotation and new tools. *Nucleic Acids Research*, *40*(D1), D1202–D1210.

Li, G., Siddiqui, H., Teng, Y., Lin, R., Wan, X.-y., Li, J., ... others (2011). Coordinated transcriptional regulation underlying the circadian clock in arabidopsis. *Nature Cell Biology*, *13*(5), 616–622.

Liu, T., Carlsson, J., Takeuchi, T., Newton, L., Farre, E.M. (2013). Direct regulation of abiotic responses by the a rabidopsis circadian clock component prr 7. *The Plant Journal*, *76*(1), 101–114.

Lu, H., McClung, C.R., Zhang, C. (2017). Tick tock: circadian regulation of plant innate immunity. *Annual Review of Phytopathology*, *55*, 287–311.

Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports*, *650*, 1–63.

Lu, S.X., Knowles, S.M., Andronis, C., Ong, M.S., Tobin, E.M. (2009). Circadian clock associated1 and late elongated hypocotyl function synergistically in the circadian clock of arabidopsis. *Plant Physiology*, *150*(2), 834–843.

Matsushika, A., Makino, S., Kojima, M., Mizuno, T. (2000). Circadian waves of expression of the aprr1/toc1 family of pseudo-response regulators in arabidopsis thaliana: insight into the plant circadian clock. *Plant and Cell Physiology*, *41*(9), 1002–1012.

McClung, C.R. (2006). Plant circadian rhythms. *The Plant Cell*, *18*(4), 792–803.

McClung, C.R. (2011). The genetics of plant clocks. *Advances in Genetics*, *74*, 105–139.

McClung, C.R. (2013). Beyond arabidopsis: the circadian clock in non-model plant species. , *24*(5), 430–436.

Mei, W., Jiang, Z., Chen, Y., Chen, L., Sancar, A., Jiang, Y. (2021). Genome-wide circadian rhythm detection methods: systematic evaluations and

practical guidelines. *Briefings in Bioinformatics*, *22*(3), bbaa135.

Michael, T.P., & McClung, C.R. (2003). Enhancer trapping reveals widespread circadian clock transcriptional control in arabidopsis. *Plant Physiology*, *132*(2), 629–639.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., ... others (2008). Network discovery pipeline elucidates conserved time-of-day–specific cis-regulatory modules. *PLoS Genetics*, *4*(2), e14.

Milenković, T., & Pržulj, N. (2008). Uncovering biological network function via graphlet degree signatures. *Cancer Informatics*, *6*, CIN–S680.

Miller, J.C., & Hagberg, A. (2011). Efficient generation of networks with given expected degrees. *International workshop on algorithms and models for the web-graph* (pp. 115–126).

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.-R., ... Coupland, G. (2002). Lhy and cca1 are partially redundant genes required to maintain circadian rhythms in arabidopsis. *Developmental Cell*, *2*(5), 629–641.

Møller, S.G., Kim, Y.-S., Kunkel, T., Chua, N.-H. (2003). Pp7 is a positive regulator of blue light signaling in arabidopsis. *The Plant Cell*, *15*(5), 1111–1119.

Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., Singh, M. (2005). Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics*, *21*(suppl_1), i302–i310.

Nakamichi, N., Kiba, T., Henriques, R., Mizuno, T., Chua, N.-H., Sakakibara, H. (2010). Pseudo-response regulators 9, 7, and 5 are transcriptional repressors in the arabidopsis circadian clock. *The Plant Cell*, *22*(3), 594–605.

Nishimura, C., Ohashi, Y., Sato, S., Kato, T., Tabata, S., Ueguchi, C. (2004). Histidine kinase homologs that act as cytokinin receptors possess overlapping functions in the regulation of shoot and root growth in arabidopsis. *The Plant Cell*, *16*(6), 1365–1377.

Nohales, M.A., & Kay, S.A. (2016). Molecular mechanisms at the core of the plant circadian oscillator. *Nature Structural & Molecular Biology*, *23*(12), 1061–1069.

Pandey, A.K., Singh, P., Tripathi, N.N. (2014, sep). Chemistry and bioactivities of essential oils of some Ocimum species: an overview. *Asian Pacific Journal of Tropical Biomedicine*, *4*(9), 682–694. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S2221169115300897

10.12980/APJTB.4.2014C77

Pandey, P., Irulappan, V., Bagavathiannan, M.V., Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science*, *8*, 537.

Pattanayak, P., Behera, P., Das, D., Panda, S. (2010). Ocimum sanctum Linn. A reservoir plant for therapeutic applications: An overview. *Pharmacognosy Reviews*.

10.4103/0973-7847.65323

Pokhilko, A., Fernández, A.P., Edwards, K.D., Southern, M.M., Halliday, K.J., Millar, A.J. (2012). The clock gene circuit in arabidopsis includes a repressilator with additional feedback loops. *Molecular Systems Biology*, *8*(1), 574.

Pruneda-Paz, J.L., Breton, G., Para, A., Kay, S.A. (2009). A functional genomics approach reveals che as a component of the arabidopsis circadian clock. *Science*, *323*(5920), 1481–1485.

Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*, *23*(2), e177–e183.

Pržulj, N., Corneil, D.G., Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics*, *20*(18), 3508–3515.

Rastogi, S., Kalra, A., Gupta, V., Khan, F., Lal, R.K., Tripathi, A.K., ... Shasany, A.K. (2015). Unravelling the genome of Holy basil: An

"incomparable" "elixir of life" of traditional Indian medicine. *BMC Genomics*.

10.1186/s12864-015-1640-z

Rastogi, S., Meena, S., Bhattacharya, A., Ghosh, S., Shukla, R.K., Sangwan, N.S., ... Shasany, A.K. (2014). De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. *BMC Genomics*.

10.1186/1471-2164-15-588

Rastogi, S., Shah, S., Kumar, R., Vashisth, D., Akhtar, M.Q., Kumar, A., ... Shasany, A.K. (2019). Ocimum metabolomics in response to abiotic stresses: Cold, flood, drought and salinity. *PloS One*, *14*(2), e0210903.

Ryu, J.S., Kim, J.-I., Kunkel, T., Kim, B.C., Cho, D.S., Hong, S.H., ... others (2005). Phytochrome-specific type 5 phosphatase controls light signal flux by enhancing phytochrome stability and affinity for a signal transducer. *Cell*, *120*(3), 395–406.

Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., ... others (2021). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *49*(D1), D10.

Sharan, R., & Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, *24*(4), 427–433.

Sheerin, D.J., Menon, C., zur Oven-Krockhaus, S., Enderle, B., Zhu, L., Johnen, P., ... Hiltbrunner, A. (2015). Light-activated phytochrome a and b interact with members of the spa family to promote photomorphogenesis in arabidopsis by reorganizing the cop1/spa complex. *The Plant Cell*, *27*(1), 189–201.

Shi, C., Baldwin, I.T., Wu, J. (2012). Arabidopsis plants having defects in nonsense-mediated mrna decay factors upf1, upf2, and upf3 show photoperiod-dependent phenotypes in development and stress responses. *Journal of Integrative Plant Biology*, *54*(2), 99–114.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., ... others (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*,

*7*(1), 539.

Singh, G., Singh, V., Singh, V. (2021). Genome-wide interologous interactome map (teagpin) of camellia sinensis. *Genomics*, *113*(1), 553–564.

Singh, G., Singh, V., Singh, V. (2022). Systems scale characterization of circadian rhythm pathway in camellia sinensis. *Computational and Structural Biotechnology Journal*.

Singh, R.B., Singh, V., Ramaswamy, R. (2010). Stochastic synchronization of circadian rhythms. *Journal of Systems Science and Complexity*, *23*(5), 978–988.

Singh, V., Singh, G., Singh, V. (2019). Tulsipin: an interologous protein interactome of ocimum tenuiflorum. *Journal of Proteome Research*, *19*(2), 884–899.

Somers, D.E., Devlin, P.F., Kay, S.A. (1998). Phytochromes and cryptochromes in the entrainment of the arabidopsis circadian clock. *Science*, *282*(5393), 1488–1490.

Srivastava, D., Shamim, M., Kumar, M., Mishra, A., Maurya, R., Sharma, D., ... Singh, K. (2019). Role of circadian rhythm in plant system: An update from development to stress response. *Environmental and Experimental Botany*, *162*, 256–271.

Thaben, P.F., & Westermark, P.O. (2014). Detecting rhythms in time series with rain. *Journal of Biological Rhythms*, *29*(6), 391–400.

Thain, S.C., Vandenbussche, F., Laarhoven, L.J., Dowson-Day, M.J., Wang, Z.-Y., Tobin, E.M., ... Van Der Straeten, D. (2004). Circadian rhythms of ethylene emission in arabidopsis. *Plant Physiology*, *136*(3), 3751–3761.

To, J.P., Haberer, G., Ferreira, F.J., Deruere, J., Mason, M.G., Schaller, G.E., ... Kieber, J.J. (2004). Type-a arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. *The Plant Cell*, *16*(3), 658–671.

Tseng, T.-S., Salomeé, P.A., McClung, C.R., Olszewski, N.E. (2004). Spindly and gigantea interact and act in arabidopsis thaliana pathways involved in light responses, flowering, and rhythms in cotyledon movements. *The Plant Cell*, *16*(6), 1550–1563.

Ueguchi-Tanaka, M., Ashikari, M., Nakajima, M., Itoh, H., Katoh, E., Kobayashi, M., ... others (2005). Gibberellin insensitive dwarf1 encodes a soluble receptor for gibberellin. *Nature*, *437*(7059), 693–698.

Upadhyay, A.K., Chacko, A.R., Gandhimathi, A., Ghosh, P., Harini, K., Joseph, A.P., ... others (2015). Genome sequencing of herb tulsi (ocimum tenuiflorum) unravels key genes behind its strong medicinal properties. *BMC plant biology*, *15*(1), 1–20.

Vazquez, A., Flammini, A., Maritan, A., Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. *Nature Biotechnology*, *21*(6), 697–700.

Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., ... Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. *Nature*, *470*(7332), 110–114.

Wu, G., Anafi, R.C., Hughes, M.E., Kornacker, K., Hogenesch, J.B. (2016). Metacycle: an integrated r package to evaluate periodicity in large scale data. *Bioinformatics*, *32*(21), 3351–3353.

Xie, Q., Wang, P., Liu, X., Yuan, L., Wang, L., Zhang, C., ... others (2014). Lnk1 and lnk2 are transcriptional coactivators in the arabidopsis circadian oscillator. *The Plant Cell*, *26*(7), 2843–2857.

Yakir, E., Hilman, D., Harir, Y., Green, R.M. (2007). Regulation of output from the plant circadian clock. *The FEBS Journal*, *274*(2), 335–345.

Yang, R., & Su, Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, *26*(12), i168–i174.

Yaveroğlu, Ö.N., Malod-Dognin, N., Davis, D., Levnajic, Z., Janjic, V., Karapandza, R., ... Pržulj, N. (2014). Revealing the hidden language of complex networks. *Scientific Reports*, *4*(1), 1–9.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., ... others (2006). Wego: a web tool for plotting go annotations. *Nucleic Acids Research*, *34*(suppl_2), W293–W297.

Yerushalmi, S., & Green, R.M. (2009). Evidence for the adaptive significance of circadian rhythms. *Ecology Letters*, *12*(9), 970–981.

Yu, J.-W., Rubio, V., Lee, N.-Y., Bai, S., Lee, S.-Y., Kim, S.-S., ... others (2008). Cop1 and elf3 control circadian function and photoperiodic flowering by regulating gi stability. *Molecular Cell*, *32*(5), 617–630.

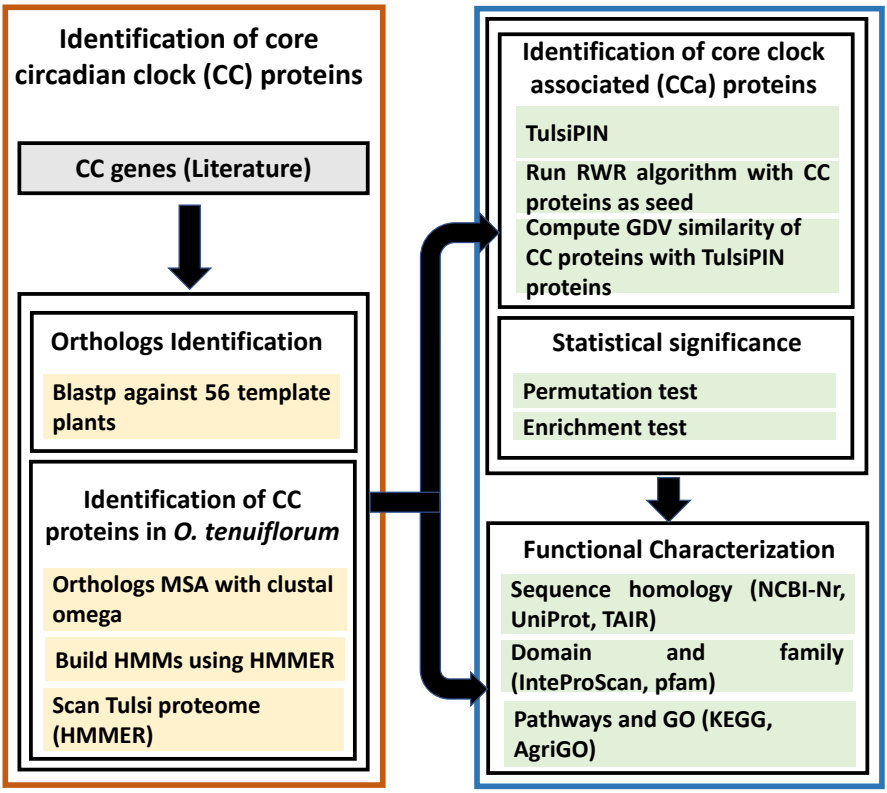



**Fig. 1** Overall methodology for genome wide identification, characterization of core circadian clock (CC) and core clock associated (CCa) proteins in *O. tenuiflorum*

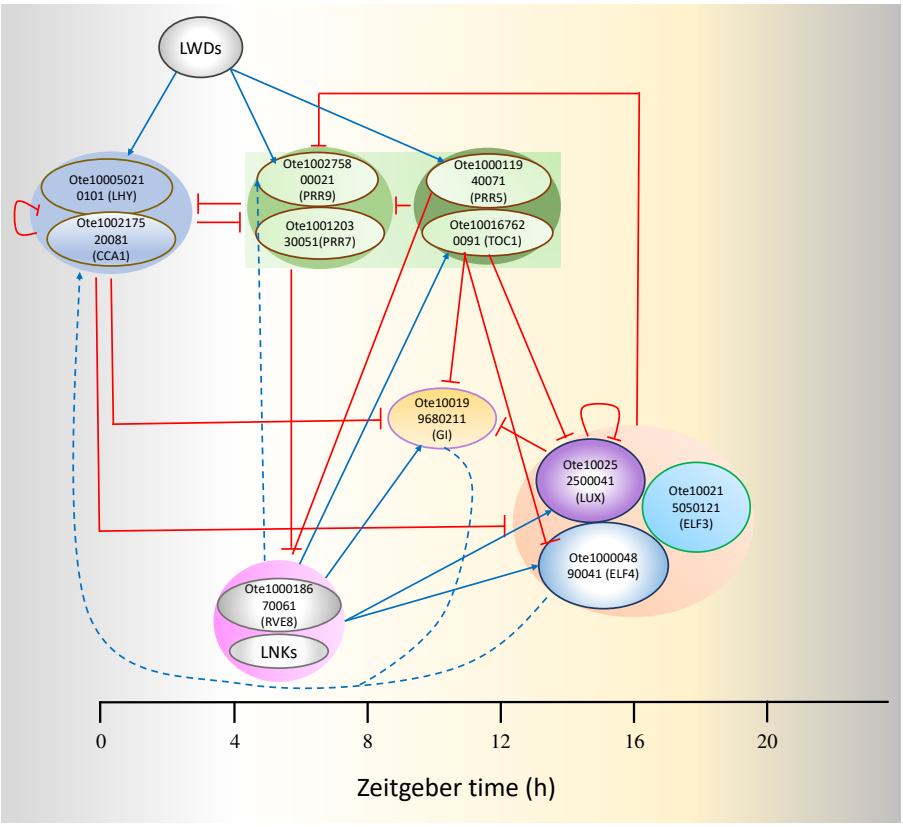



**Fig. 2** The core clock module of *O. tenuiflorum*

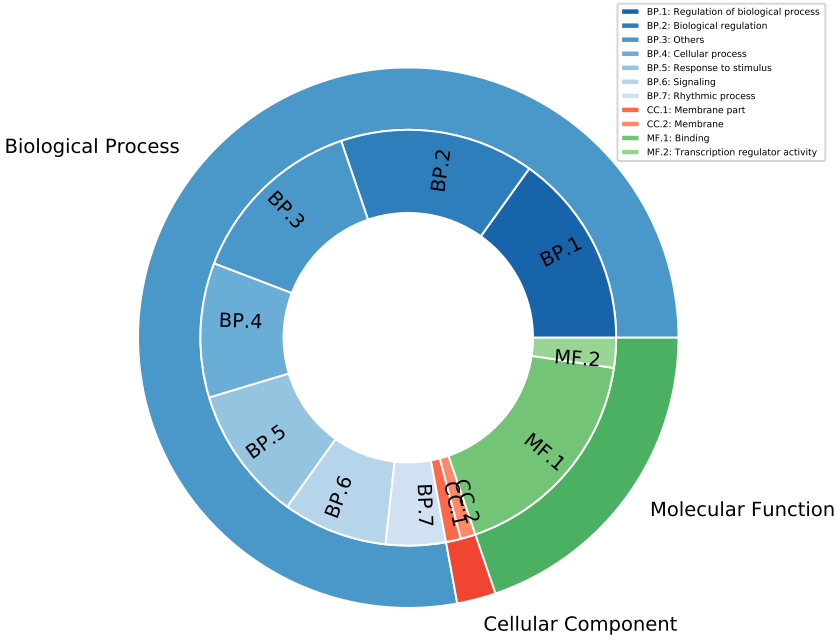



**Fig. 3** Gene ontology annotations of core circadian clock (CC) proteins of *O. tenuiflorum*

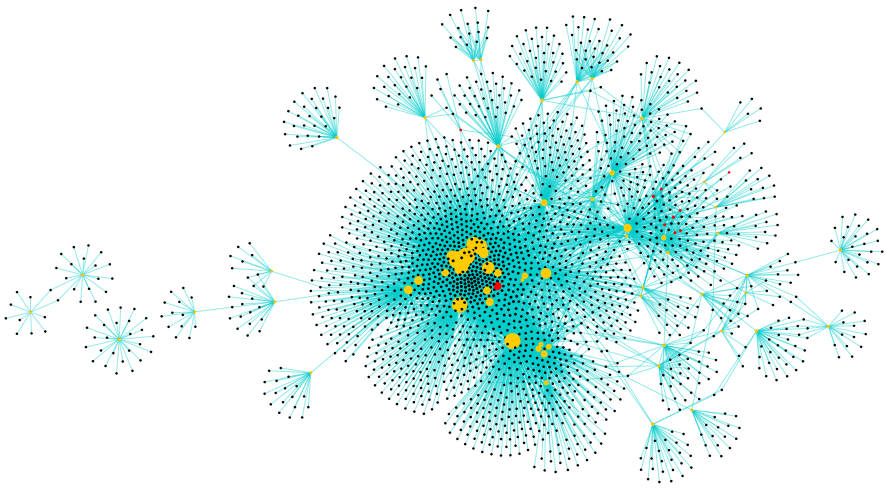


**Fig. 4** Sub network of proteins obtained by combining core circadian clock (CC) and core circadian associated (CCa) proteins (including first interactors). Yellow colored nodes represent CCa proteins, red colored nodes are CC proteins

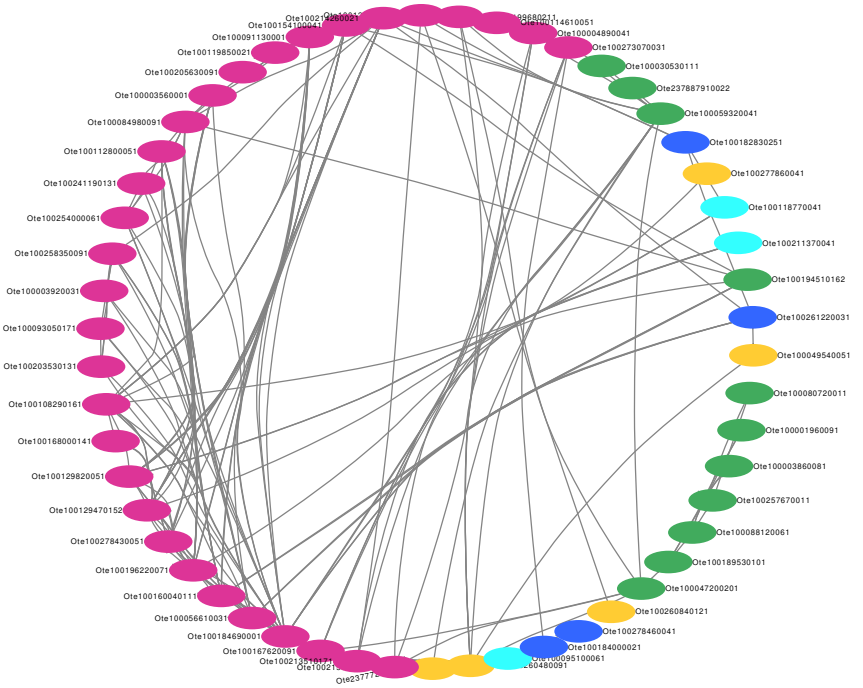



**Fig. 5** Sub network of circadian clock circuitry including both core circadian clock (CC) and core circadian associated (CCa) proteins only (excluding first interactors). Nodes are colored according to the broad functional class of each protein where hot pink represents proteins involved in gene expression, green (signaling), blue (metabolism), yellow (post translational modification) and cyan represents transport proteins.

# Supplementary Files

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

- S1.xlsx
- S2.xlsx
- S3.xlsx
- S4.xlsx