BioTechnologia vol. 103 (2) • pp. 113–129 • 2022
Journal of Biotechnology, Computational Biology and Bionanotechnology RESEARCH PAPERS
http://doi.org/10.5114/bta.2022.116205

Physicochemical properties and homology studies of the floral meristem identity gene LFY in nonflowering and flowering plants

ROSHNI PULUKKUNADU THEKKEVEEDU [1], SMITHA HEGDE [2*]

[1] Department of Post Graduate Studies and Research in Biotechnology, St. Aloysius College, Mangalore, India
[2] Nitte University Centre for Science Education and Research, Mangalore, India

Abstract

In flowering plants, the LEAFY (LFY) gene controls floral meristem activity. In early land plants such as mosses and ferns, it, however, has a minimum role in cell division and development of diploid sporophyte. Homology modeling, an accurate and efficient protein structure prediction method, was used to construct a 3D model of the LEAFY protein in nonflowering and flowering plants. The present study examines the following species: Charophyte green algae, *Physcomitrella*, *Ceratopteris*, *Picea*, and *Arabidopsis*, as they are the popularly used model organisms for developmental studies. LEAFY protein sequences from the model organisms were aligned by multiple sequence alignment. 3D models of the LEAFY protein from all the model organisms was constructed using the PHYRE2 program with 100% confidence, and the constructed models were evaluated using the MolProbity tool. On the basis of the conserved regions, Charophyte green algae shared 38–46% sequence similarity with *Physcomitrella* sp., 37–46% similarity with *Ceratopteris* sp., 33–41% similarity with *Picea* sp., and 32–38% similarity with *Arabidopsis* sp. The Motif Finder server identified the protein family domain FLO_LFY and LFY_SAM, whose function is floral meristem development. Secondary structure prediction analysis indicated that the LEAFY protein belongs to the alpha (α) protein class, which is stable against mutation and thus limits structural changes in the LEAFY protein. The study findings reveal two distinct clusters of the LFY gene from the common ancestor green algae. One cluster is present in nonflowering plants that include mosses, pteridophytes, and gymnosperms, and the other cluster is present in flowering plants that include orchids, monocots, dicots, and angiosperms.

Key words: *Arabidopsis*, *Picea, Ceratopteris*, *Physcomitrella*, homology modeling, LFY

Introduction

The homologous gene LFY (LEAFY) regulates cell proliferation and flower development in plants. It is widely distributed in algae, mosses, ferns, gymnosperms, and flowering plants (Villimová, 2012). LFY homologs reported in aquatic Charophyte green algae are closely related to those found in land plants. LFY encodes a plant-specific transcription factor that functions as an activator or a repressor, depending on the cofactor it interacts with (Siriwardana and Lamb, 2012). In *Physcomitrella patens*, LFY regulates cell division in gametophytes and sporophyte (Tanahashi et al., 2005), whereas LFY homologs in the fern *Ceratopteris richardii* function in shoot development (Plackett et al., 2018). LFY is a floral meristem identity gene that controls multiple aspects of inflorescence development in the flowering plant *Arabidopsis thaliana* (Weigel and Nilsson, 1995), and it is active during reproductive structure development in gymnosperms (Dornelas and Rodriguez, 2005; Moyroud et al., 2017). An increase in the expression of LFY results in early flowering, and a mutation in LFY causes a transition of flowers into leaves and shoots (Weigel et al., 1992). Charophytes (algae) (Domozych et al., 2016), *P. patens* (moss) (Cove et al., 2009), *C. richardii* (fern) (Hickok et al., 1995; Renzaglia and Warne, 1995), *Picea abies* (Spruce) (Nystedt et al., 2013), and *Arabi-*

* Corresponding author: Nitte University Centre for Science Education and Research, Paneer Campus, Deralakatte, Mangalore 575018, India; e-mail: smitha.hegde@nitte.edu.in

dopsis thaliana (flowering plant) (Ezhova, 1999) are model organisms for genetic, developmental, and evolutionary studies. In the present study, LFY homologs of few model plants were analyzed to determine molecular differences, i.e., their transition changes during evolution from simple to complex structures in plants. The diverging lineage of the LFY gene, which is modified and altered during the evolution of floral meristems, will help us to understand the origin of flower development or the lack of it in plants. The present study attempted to corroborate the molecular changes of LFY and the evolution of flowering in plants from Charophyte green algae to angiosperms.

Prediction of structure is imperative to study the biochemical and cellular functions of proteins. X-ray crystallography, NMR spectroscopy, and electron microscopy are the techniques currently used for protein structure prediction; however, these methods are time-consuming and require expensive wet lab tools (Venkatesan et al., 2013). Computational tools have been used for the past 30 years in protein structure prediction and continue to help researchers in experimental investigations on a large scale (Nagano, 1973; Gupta et al., 2014; Kc, 2017; Kuhlman and Bradley, 2019). Computational techniques have improved the success rate in protein prediction methods in the last decade. The prediction methods are categorized into comparative modeling (homology modeling) (Šali and Blundell, 1993; Lam et al., 2017), threading (Panchenko et al., 2000; Skolnick and Kihara, 2001; Xu et al., 2007), and *ab initio* modeling (free modeling) (Ortiz et al., 1998; Simons et al., 2001; Lee et al., 2017). The most accurate method for protein structure prediction is homology modeling, which is used to construct 3D models of unknown target sequences based on known structures (templates) with sequence similarity >30% (Cavasotto and Phatak, 2009) collected from databases by using software or web servers (Eswar et al., 2003; Pieper et al., 2006). The process of homology modeling involves template search of related structures for query sequence, multiple sequence alignment of targets and template structures, construction of a 3D model, and finally, evaluation of the model (Hasani and Barakat, 2017; Studer et al., 2019). The steps are repeated to obtain an optimum model (Martí-Renom et al., 2000). The 3D structures of proteins are more conserved than their amino acid sequences, and minor changes in sequences usually result in a slight variation in their 3D structure (Lesk and Chothia, 1986).

In the present study, the homologous sequences of the LFY gene were compared to analyze its phylogenesis and physicochemical properties of the gene product; moreover, protein structure prediction and construction of 3D models of protein and their evaluation were performed. The LFY homologous sequences from algae to flowering plant model systems were elaborated. A comparative analysis of the homologous genes was conducted virtually to study their structure, function, phylogeny, and proteins.

Materials and methods

Data mining from database and phylogeny construction

Complete and partial LEAFY protein sequences of the LFY gene of all available plant families (supplementary Table 1) were collected from GenBank NCBI (National Centre for Biotechnology Information) (http://www.ncbi.nlm.nih.gov) in FASTA format after clicking protein search, followed by building datasets of sequences using Notepad from the Protein database of related organisms in NCBI [LEAFY in plants – Protein – NCBI (nih.gov)]. A phylogenetic tree was constructed using MEGA 7 (Molecular Evolutionary Genetic Analysis version 7) software with the neighbor-joining method with bootstrapping values for 1000 replicates. The evolutionary distance was computed using the p-distance method. All the LEAFY protein sequences were copied onto MEGA 7 software and aligned using ClustalW. Complete and partial sequences of the LEAFY protein of charophyte green algae – *Klebsormidium subtile* and *Coleochaete scutata* (2 sequences), *P. patens* (2 sequences), *Ceratopteris* sp. (3 sequences), *Picea* sp. (2 sequences), and *A. thaliana* (3 sequences) were used (supplementary Table 1 and Table 2). The aligned sequences were then exported to conduct phylogenetic analysis. The phylogeny analysis was conducted using Neighbor Joining Tree method. It was further tested with the Bootstrap method. A total of 1000 bootstrap replications were used to compute and construct a phylogenetic tree.

Motif search

The Motif Finder server (https://www.genome.jp/tools/motif/) was used to analyze the family or the protein domain of the protein sequences. FASTA sequences of the protein were entered as a query sequence in the

Motif Finder server. The results showed LEAFY protein sequences of all model organisms with conserved domains from Pfam databases. These sequences were confirmed in the CDD database (NCBI) webserver (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) by using the protein accession number of the respective organisms.

Multiple sequence alignment

FASTA sequences of LEAFY proteins were aligned using the Clustal Omega Multiple Sequence Alignment Program [CLUSTAL O (1.2.4)] (https://www.ebi.ac.uk/Tools/msa/clustalo/) using complete and partial sequences of LEAFY proteins. FASTA sequences of LEAFY proteins were also used for screening sequence alignment of conserved, conservative mutated, semi-conservative mutated, and nonconservative mutated sequences among the species analyzed.

Physicochemical analysis

The amino acid sequences of LEAFY proteins were determined for analyzing physicochemical properties in the ProtParam tool – ExPASy (http://web.expasy.org/protparam) that computes various parameters such as molecular weight (MW), isoelectric point (pI), instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY). Each amino acid sequence was imported into the ProtParam tool ExPASy and computed for analyzing various parameters of LEAFY protein sequences.

Structure prediction and analysis

The secondary structure of the LEAFY protein was predicted using the PSI-blast-based secondary structure tool PREDiction (PSIPRED 4.0), and the 3D model was generated using PHYRE 2 (ProteinHomology/analogy RecognitionEngineV2.0) (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) program. Multiple sequences of each model plant were batch-processed while uploading onto the PHYRE2 server. The LEAFY protein sequence was submitted in FASTA format and processed for the following information: 1) summary and sequence analysis details, 2) secondary structure and disorder prediction, 3) domain analysis, and 4) detailed template alignment information with figures. Secondary structure was predicted as α-helix, β-strand, coils, and disorder in the structure analysis report.

Protein modeling and validation

Protein homology modeling for LEAFY proteins was performed with the SWISS-MODEL https://swissmodel.expasy.org/ web server. Protein modeling involved the following: 1) identification of template structure, 2) alignment of the target sequence and template structure, 3) model building, and 4) model quality evaluation.

The best models of LEAFY protein structures for all the plant models were evaluated for quality by using the structure validation tool MolProbity (http://molprobity.biochem.duke.edu/). The PDB code of the protein structure was further validated using their web server. The analysis revealed the distribution of residues in the torsion angles (φ and ψ) of Ramachandran plot with the φ and ψ values between +180 and −180 on the *x*-axis and *y*-axis, respectively. The percentage of residues in the favored, allowed region and outliers from the Ramachandran plot analysis were also detected.

Results and discussion

Evolution of the LFY gene in the plant family

Forty-one LEAFY protein sequences of different plant families were downloaded from GenBank (Table 1), and phylogenetic analysis was conducted. The phylogenetic tree (Fig. 1) illustrates the relationship among 41 amino acid sequences of LEAFY proteins analyzed in MEGA7 with 1000 bootstrap replicates using the neighbor-joining method. The bootstrap percentage specifies the reliability of each node of the phylogenetic tree, and an estimate of $<70\%$ reliability on tree topology is not considered to be acceptable (Hall, 2013). The p-distance method was used to compute evolutionary distances with the differences in the number of amino acids per site.

LEAFY protein sequences of Charophyte green algae were placed at the base of the phylogenetic tree and grouped into two distinct clusters. One cluster of LEAFY protein sequences was present in flowering plants starting from orchids to *Arabidopsis*, such as *Vanilla planifolia, Phalaenopsis* hybrid cultivar, *Tricyrtis formosana, Oryza sativa* Japonica Group, *Zea mays, Allium cepa, Amborella trichopoda, Nymphaea odorata, Chrysanthemum indicum, Litchi chinensis, Mangifera indica, Populus balsamifera, Magnolia virginiana, Annona squamosa, Brassica rapa*, and *A. thaliana*. The second cluster contained all nonflowering plants such as mosses, pteridophytes, gymnosperms, and cycads. It is reported that

Table 1. List of LEAFY homolog of different plant families from GenBank NCBI

No.	Organisms	Protein accession number	Number of amino acid
1	*Ceratozamia mexicana*	AIG12601.1	375
2	*Lepidozamia peroffskyana*	AIG12598.1	380
3	*Dioon spinulosum*	AIG12603.1	380
4	*Microcycas calocoma*	AIG12609.1	376
5	*Encephalartos arenarius*	AIG12597.1	380
6	*Bowenia spectabilis*	AIG12608.1	375
7	*Macrozamia lucida*	AIG12599.1	377
8	*Stangeria eriopus*	AIG12610.1	363
9	*Cycas revoluta*	AIG12606.1	377
10	*Ginkgo biloba*	ADD64700.1	402
11	*Picea abies*	AAV49504.1	386
12	*Picea sitchensis*	AKA55658.1	380
13	*Angiopteris lygodiifolia*	BAB93543.1	344
14	*Sceptridium robustum*	BAB88864.1	350
15	*Psilotum nudum*	BAB88863.1	372
16	*Ceratopteris thalictroides*	ABF74516.1	237
17	*Ceratopteris pteridoides*	ABF74512.1	237
18	*Ceratopteris richardii*	ABF74513.1	237
19	*Physcomitrella patens*	BAD91044.1	349
20	*Physcomitrella patens*	BAD91043.1	348
21	*Vanilla planifolia*	AOA52645.1	491
22	*Phalaenopsis* hybrid cultivar	ACS94257.1	437
23	*Tricyrtis formosana*	BAN62610.1	411
24	*Oryza sativa* Japonica Group	AHX83809.1	389
25	*Zea mays*	AAO43173.1	393
26	*Allium cepa*	AFR67540.1	372
27	*Allium cepa*	AVT42847.1	370
28	*Amborella trichopoda*	AGV98899.1	391
29	*Nymphaea odorata*	AAF77609.1	387
30	*Chrysanthemum indicum*	ARR73986.1	412
31	*Litchi chinensis*	AGR45584.1	388
32	*Mangifera indica*	ADX97320.1	383
33	*Populus balsamifera*	AEK06015.1	377
34	*Magnolia virginiana*	ACV88634.1	389
35	*Annona squamosa*	AKV57239.1	411
36	*Brassica rapa*	ANJ12320.1	417
37	*Arabidopsis thaliana*	AAM27932.1	424
38	*Arabidopsis thaliana*	AAM27931.1	424
39	*Arabidopsis thaliana*	AAM27941.1	424
40	*Coleochaete scutata*	AHJ90705.1	328
41	*Klebsormidium subtile*	AHJ90707.1	495

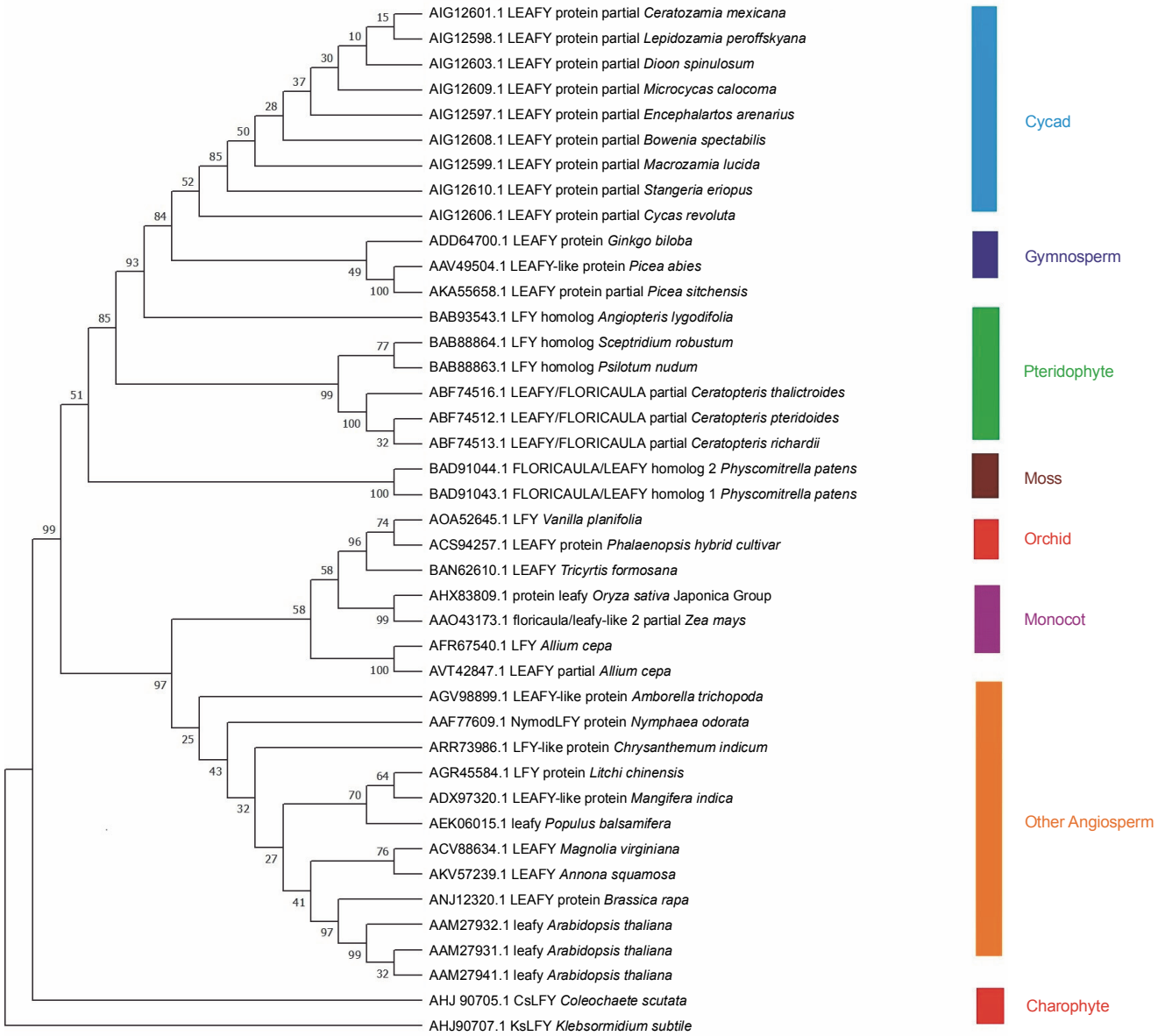

Fig. 1. Phylogenetic tree constructed in MEGA 7 with 41 LEAFY protein sequences; bootstrap values are listed next to the branch

the plant-specific transcription factor LFY evolved in Streptophyte algae (Wilhelmsson et al., 2017). The LEAFY protein sequence of pteridophytes is closer to that of cycads and gymnosperms than to that of orchids, monocots, and other angiosperms, which indicates structural and functional similarity to LEAFY from gymnosperms. The tree shows an early alteration in the LFY homolog, which segregated and evolved into two clusters of flowering and nonflowering plants. It is also reported that ancient gene duplication and sub-functionalization processes influenced the evolution of the LEAFY gene (Gao et al., 2019).

Description of the candidate LEAFY protein

The LEAFY protein sequences from 2 species of Charophyte green algae (*K. subtile* – 495 aa and *C. scutata* – 328 aa), 2 moss species (*P. patens* – 349 aa, *P. patens* – 348 aa), 2 gymnosperm species (*P. abies* – 386 aa and *P. sitchensis* – 380 aa), 3 pteridophytes species (*C. thalictroides* – 237 aa, *C. pteridoides* – 237 aa, *C. richardii* – 237 aa), and 3 angiosperm species (*A. thaliana* – 424 aa) were collected from the NCBI protein database (Table 2) and analyzed.

The LFY transcription factor gene evolved from algae (charophyte) (Sayou et al., 2014; Brunkard et al., 2015;

Table 2. Description and list of LEAFY homolog of *Charophyte* green algae, *Physcomitrella* sp., *Ceratopteris* sp., *Picea* sp., and *Arabidopsis* sp. from GenBank NCBI

Organism	Protein accession number	Protein	Number of amino acid
Klebsormidium subtile	AHJ90707.1	KsLFY	495
Coleochaete scutata	AHJ90705.1	CsLFY	328
Physcomitrella patens	BAD91044.1	FLORICAULA/LEAFY homolog2	349
Physcomitrella patens	BAD91043.1	FLORICAULA/LEAFY homolog2	348
Ceratopteris thalictroides	ABF74516.1	LEAFY/FLORICAULA	237
Ceratopteris richardii	ABF74513.1	LEAFY/FLORICAULA	237
Ceratopteris pteridoides	ABF74512.1	LEAFY/FLORICAULA	237
Picea abies	AAV49504.1	PaLFY	386
Picea sitchensis	AKA55658.1	LEAFY	380
Arabidopsis thaliana	AAM27931.1	Leafy	424
Arabidopsis thaliana	AAM27932.1	Leafy	424
Arabidopsis thaliana	AAM27941.1	Leafy	424

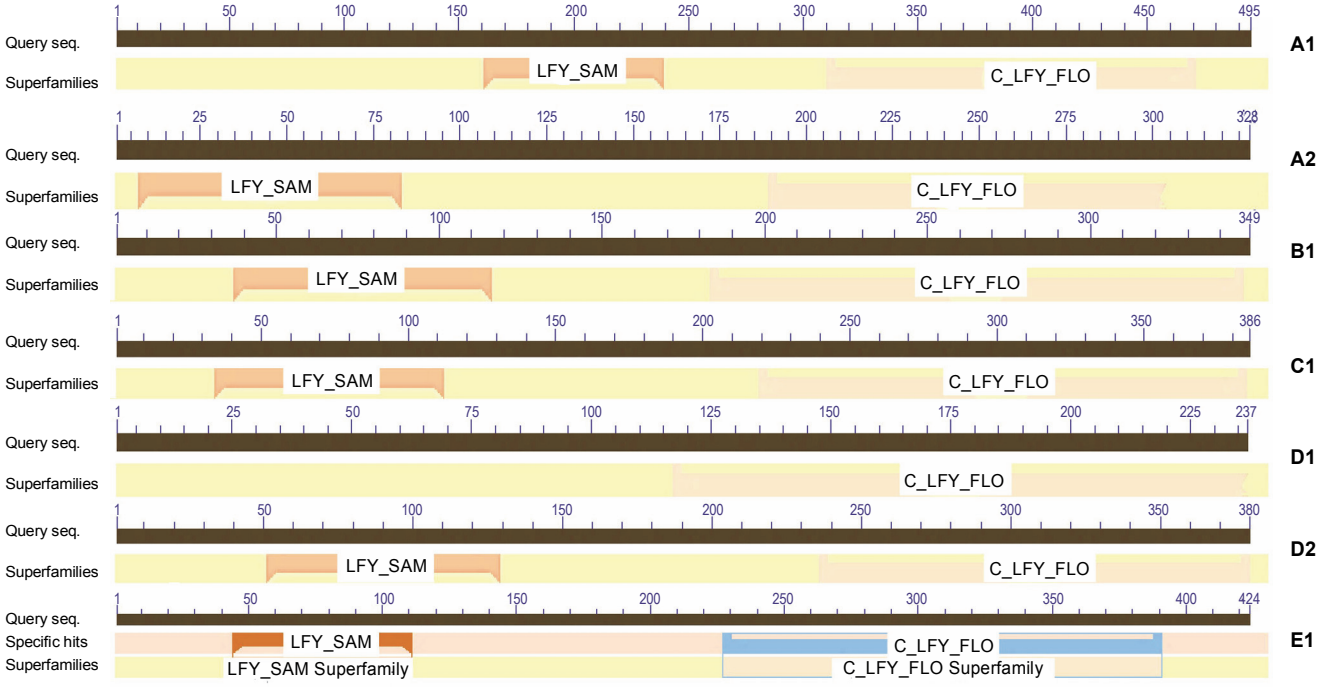

Fig. 2. Conserved domain analysis from Conserved Domain Database – NCBI of: A1 – *Klebsormidium subtile*, A2 – *Coleochaete scutata*, B1 – *Physcomitrella patens*, C1 – *Ceratopteris* sp., D1 – *Picea abies*, D2 – *Picea sitchensis*, E1 – *Arabidopsis thaliana*

Gao et al., 2019). LEAFY/ LFY homologs in different model organisms, such as FLO/LFY genes (PpLFY1, PpLFY2), regulate the first zygotic cell division in *P. patens* (Tanahashi et al., 2005). LFY maintains apical stem cell activity in gametophyte and sporophyte during shoot development in *C. richardii* (Plackett et al., 2018). In gymnosperms, LFY and the paralog of LFY – NEEDLY (NLY) regulate male and female reproductive structures (Silva et al., 2016), and their expression levels were characterized in *Picea* sp. (Vázquez-Lobo et al., 2007). Moreover, LFY, the plant-specific transcription factors, are conserved as floral meristem identity genes, which

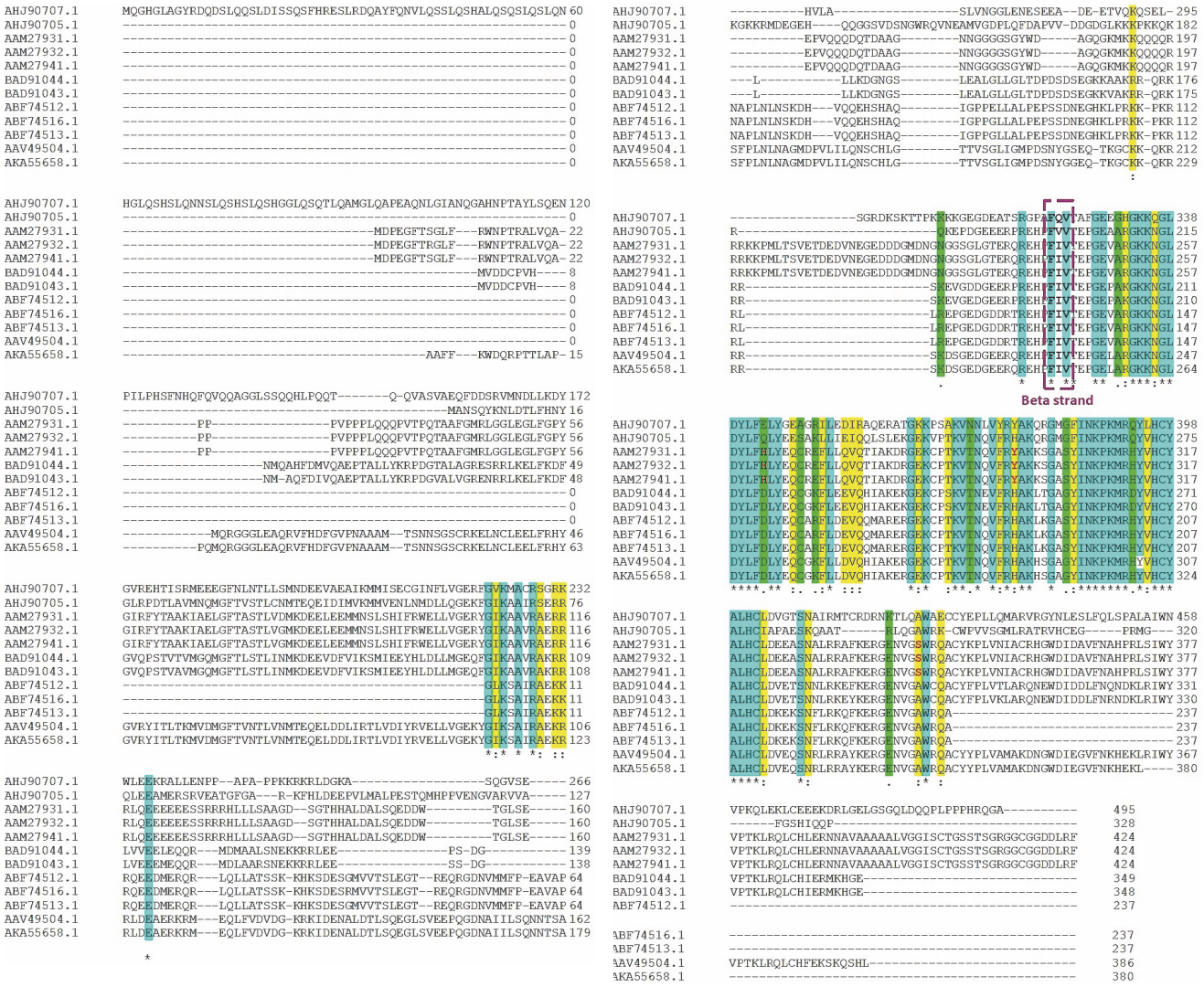

Fig. 3. Multiple Sequence Alignment of FLO_LFY protein sequences of length 237, 348, 349 and 424 residues obtained using Clustal Omega; (*) – denotes conserved sequence which is highlighted in blue, (:) – denotes conservative mutation which is highlighted in yellow [variation in sequence marked in red], (.) – denotes semi-conservative mutation which is highlighted in green [variation in sequence marked in red] and ( ) – denotes non-conservative mutation

control inflorescence and floral organ development in *Arabidopsis* (Wang et al., 2004).

Domain analysis

Motif search of protein sequences was conducted (Table 3), and it was found that the LEAFY protein sequences shared two domains, namely N-terminal Sterile Alpha Motif (SAM_ LFY) and C-terminal DNA binding domain (C_LFY_FLO) (GenomeNet bioinformatics tool). The result was confirmed using the CDD database (NCBI) web server (Fig. 2). The SAM domain mediates LFY oligomerization that helps to access low-affinity binding sites or closed chromatin regions (Sayou et al., 2016), and the biochemical properties of SAM domains are conserved throughout the evolution of all plant species. The crystal structure of the LFY- DNA binding domain resembles that of helix-turn-helix proteins and dimerizes on DNA, which triggers major developmental switches in plants (Hamès et al., 2008). The domains bind to short stretches of DNA called transcription factor binding sites (TFBS) that regulate gene expression. Domain analysis reported two domains of LEAFY proteins in all model organisms, except in fern species. In ferns, it was noted that only the C-terminal C_LFY_FLO, DNA binding domain is in partial confidence with other LEAFY protein sequences screened. The Pfam report in

Table 3. Domain and functional analysis of LFY/LEAFY homologous using motif find server

Protein ID	Pfam	Position (independent *E*-value)	Description
AHJ90707.1	1.C_ LFY_FLO	311–471 (7.7e-49)	**PF17538**, DNA Binding Domain (C-terminal) Leafy/Floricaula
	2.SAM_LFY	162–238 (3.6e-19)	**PF01698**, Floricaula /Leafy protein SAM domain
AHJ90705.1	1.C_ LFY_FLO	189–287 (3.6e-46)	**PF17538**, DNA Binding Domain (C-terminal) Leafy/Floricaula
	2.SAM_LFY	5–82 (3.5e-27)	**PF01698**, Floricaula /Leafy protein SAM domain
BAD91044.1	1.C_ LFY_FLO	183–347 (5.8e-86)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	38–115 (1e-27)	**PF01698**, Floricaula/Leafy protein SAM domain
BAD91043.1	1.C_ LFY_FLO	182–346 (6.9e-86)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	37–114 (2.4e-28)	**PF01698**, Floricaula/Leafy protein SAM domain
ABF74516.1	1.C_ LFY_FLO	119–237 (8.6e-68)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
ABF74512.1	1.C_ LFY_FLO	119–237 (8.6e-68)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
ABF74513.1	1.C_ LFY_FLO	119–237 (8.6e-68)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
AAV49504.1	1.C_ LFY_FLO	218–383 (3.6e-95)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	35–112 (1.4e-37)	**PF01698**, Floricaula/Leafy protein SAM domain
AKA55658.1	1.C_ LFY_FLO	235–380 (4.8e-83)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	52–129 (1.4e-37)	**PF01698**, Floricaula/Leafy protein SAM domain
AAM27932.1	1.C_ LFY_FLO	229–393 (1.3e-107)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	45–123 (4.6e-43)	**PF01698**, Floricaula/ Leafy protein SAM domain
AAM27931.1	1.C_ LFY_FLO	229–393 (1.3e-107)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	45–123 (4.6e-43)	**PF01698**, Floricaula/Leafy protein SAM domain
AAM27941.1	1.C_ LFY_FLO	229–393 (1.3e-107)	**PF17538**, DNA Binding Domain (C – terminal) Leafy/Floricaula
	2.SAM_ LFY	45–123 (4.6e-43)	**PF01698**, Floricaula/Leafy protein SAM domain

Table 4. Details of LEAFY protein sequences

Organism	Protein ID	Number of amino acid	Sequence
Klebsormidium subtile	AHJ90707.1	495 aa	complete sequence
Coleochaete scutata	AHJ90705.1	328 aa	complete sequence
Physcomitrella patens	BAD91044.1	349 aa	complete sequence
Physcomitrella patens	BAD91043.1	348 aa	complete sequence
Ceratopteris thalictroides	ABF74516.1	237 aa	partial sequence
Ceratopteris richardii	ABF74513.1	237 aa	partial sequence
Ceratopteris pteridoides	ABF74512.1	237 aa	partial sequence
Picea abies	AAV49504.1	386 aa	complete sequence
Picea sitchensis	AKA55658.1	380 aa	partial sequence
Arabidopsis thaliana	AAM27931.1	424 aa	complete sequence
Arabidopsis thaliana	AAM27932.1	424 aa	complete sequence
Arabidopsis thaliana	AAM27941.1	424 aa	complete sequence

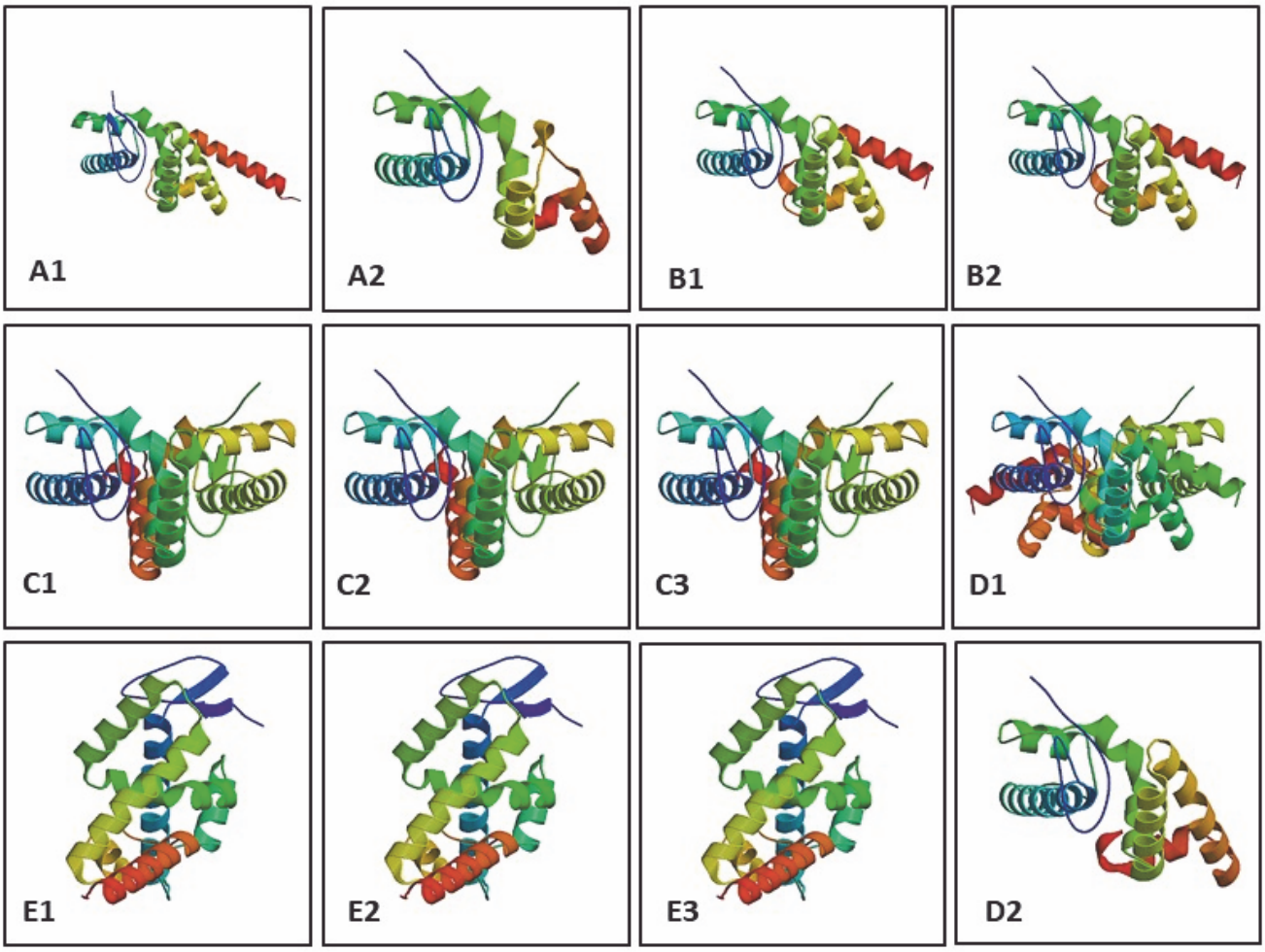

Fig. 4. Constructed 3D models of LEAFY protein: A1 – *Klebsormidium subtile*, A2 – *Coleochaete scutata*, B1–B2 – *Physcomitrella patens*, C1 – *Ceratopteris thalictroides*, C2 – *Ceratopteris richardii*, C3 – *Ceratopteris pteridoides*, D1 – *Picea abies*, D2 – *Picea sitchensis*, E1–E3 – *Arabidopsis thaliana* using SWISS-MODEL (represented in rainbow colour from N → C)

CDD revealed that these development LEAFY proteins are homologs of floricaula (FLO) and LEAFY (LFY), which function in floral meristem identity (Table 4). A mutation in these protein sequences affected the flower and leaf formation (Weigel et al., 1992; Hofer et al., 1997; Grandi et al., 2012; Monniaux et al., 2017).

Analysis of sequence similarity

Multiple sequence alignment methods align and compare DNA, RNA, or protein sequences for evolutionarily relatedness. The aligned sequences provide valuable information regarding the structural, functional, and evolutionary history, often leading to a common ancestor (Edgar and Batzoglou, 2006; Chatzou et al., 2016). LFY primary sequences were aligned using the Clustal Omega [CLUSTAL O (1.2.4)] program to find the conserved region(s). The sequences were aligned by inserting a gap or space into the sequence to extend to the same length after alignment (Wang and Jiang, 1994; Tran and Wallinga, 2017). Charophyte green algae shared 38–46% sequence similarity with *Physcomitrella* sp., 37–46% similarity with *Ceratopteris* sp., 33–41% similarity with *Picea* sp., and 32–38% similarity with *Arabidopsis* sp. Fifty conserved (similar amino acid sequences), 22 conservative mutated (mutation results in the replacement of amino acid with a similar biochemical property), and 9 semi-conservative mutated (mutation results in the replacement of amino acid with a similar shape but dissimilar biochemical property) amino acids were identified from the sequence alignment of LEAFY proteins. In *Arabidopsis* sp., in conservative mutated amino acid sequence, aspartic acid (D) was replaced with

Table 5. Physicochemical characterization of LEAFY proteins in the ProtParam tool

Organism	Protein ID	Number of amino acid	Molecular weight	Isoelectric point (pI)	Instability index (II)	Aliphatic index (AI)	Grand average of hydropathicity (GRAVY)
Klebsormidium subtile	AHJ90707.1	495 aa	55211.84	6.51	62.51	72.18	−0.754
Coleochaete scutata	AHJ90705.1	328 aa	37142.57	8.67	47.39	67.47	−0.694
Physcomitrella patens	BAD91044.1	349 aa	40090.81	6.40	44.90	74.87	−0.702
Physcomitrella patens	BAD91043.1	348 aa	40134.89	6.78	47.65	76.47	−0.727
Ceratopteris thalictroides	ABF74516.1	237 aa	27286.98	9.44	40.39	62.57	−1.080
Ceratopteris richardii	ABF74513.1	237 aa	27259.90	9.27	40.31	62.57	−1.076
Ceratopteris pteridoides	ABF74512.1	237 aa	27187.84	9.36	39.58	62.57	−1.063
Picea abies	AAV49504.1	386 aa	44116.15	8.55	48.15	76.01	−0.674
Picea sitchensis	AKA55658.1	380 aa	43161.95	8.05	45.83	74.16	−0.665
Arabidopsis thaliana	AAM27931.1	424 aa	47168.86	6.48	55.47	66.08	−0.685
Arabidopsis thaliana	AAM27932.1	424 aa	47157.79	6.22	55.52	66.08	−0.683
Arabidopsis thaliana	AAM27941.1	424 aa	47099.75	6.34	55.68	66.08	−0.676

Table 6. Secondary structure prediction of LEAFY proteins using the PHYRE 2 programme

Protein ID	Alpha helix [%]	Beta strand [%]	Disordered [%]
AHJ90707.1	62	2	35
AHJ90705.1	63	1	43
BAD91044.1	66	3	31
BAD91043.1	66	4	32
ABF74516.1	72	0	58
ABF74513.1	71	1	50
ABF74512.1	72	1	50
AAV49504.1	61	2	41
AKA55658.1	60	1	47
AAM27931.1	58	2	44
AAM27932.1	58	3	44
AAM27941.1	58	3	44

histidine (H) at position 296 and alanine (A) was replaced with serine (S) at position 343, while in semi-conservative mutated amino acid sequence, histidine was replaced with tyrosine (T) at position 262, which were different from the respective amino acid sequences in the other non-flowering plant species tested (Fig. 3).

LFY orthologs are found in all land plants, and the LFY gene performs various functions in multiple species as it evolves after gene duplication events (Silva et al., 2016). LFY homologs are involved in regulating cell division, and expansion and arrangement in free-sporing land plants such as ferns or fern allies and bryophytes. They also regulate both floral identity and cell division in gymnosperms and angiosperms (Moyroud et al., 2010).

Physicochemical properties of LEAFY proteins

The physicochemical properties of proteins influence their affinity, interaction, and adaptability to a biological

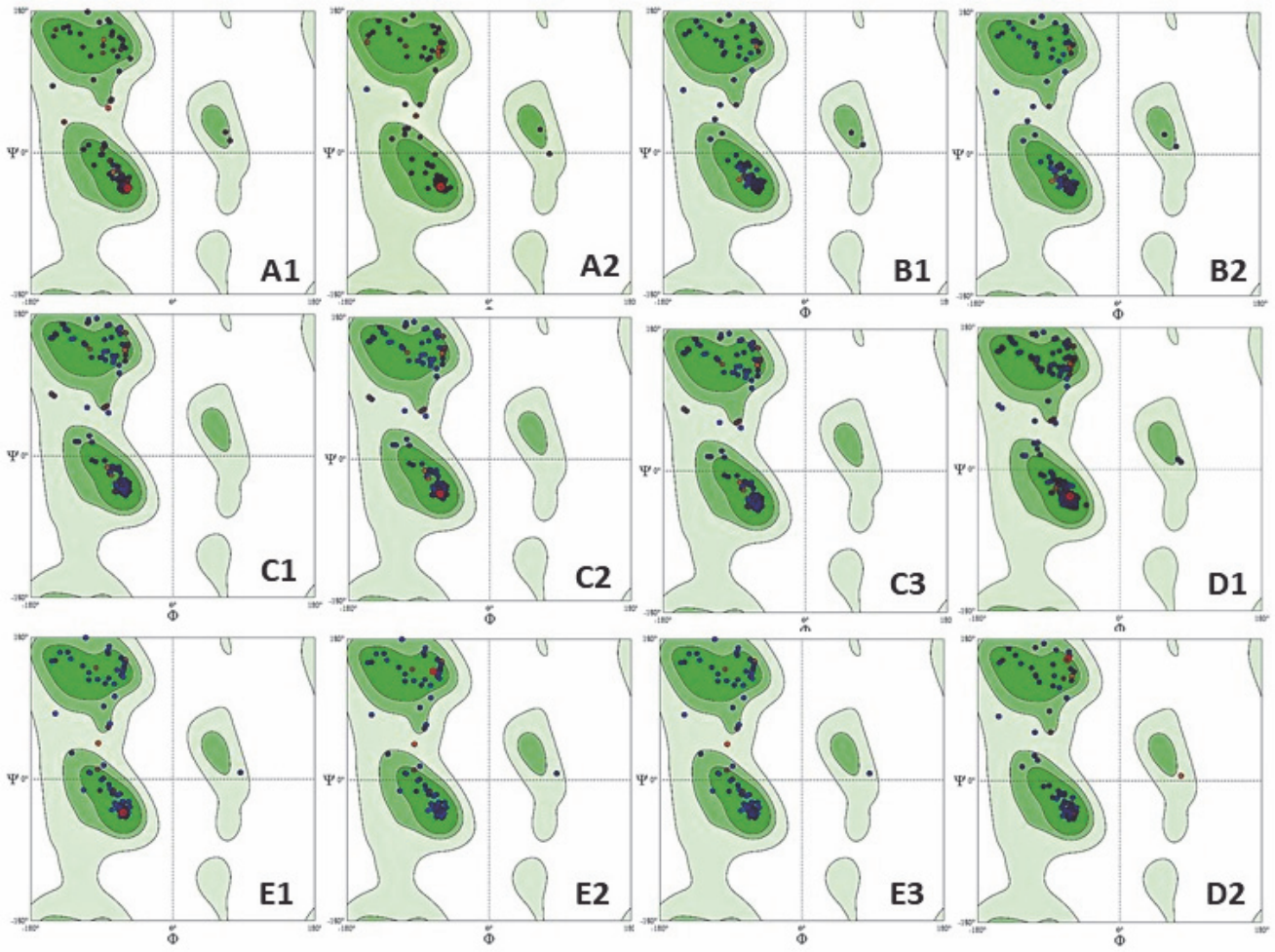

Fig. 5. Ramachandran plot analysis of protein model of: A1 – *Klebsormidium subtile*, A2 – *Coleochaete scutate*, B1–B2 – *Physcomitrella patens,* C1 – *Ceratopteris thalictroides,* C2 – *Ceratopteris richardii,* C3 – *Ceratopteris pteridoides*, D1 – *Picea abies,* D2 – *Picea sitchensis,* E1–E3 – *Arabidopsis thaliana*

system (Panda and Chandra, 2012; Dhar et al., 2020). Physicochemical characterization of the amino acid sequence includes MW, pI, II, AI, and GRAVY (Kaur et al., 2020), which were estimated using Expasy's ProtParam (Table 5). This comparative analysis helped us to identify the occurrence of diversity of LFY protein sequences across Charophyte green algae, *Physcomitrella* sp.*, Ceratopteris* sp., *Picea* sp., and *Arabidopsis* sp. pI is the pH value at which there are no net charges on the protein, and the protein remains stable without migration in the electric field and remains firm and stable at this pI. pI is crucial in protein separation and characterization (Pergande and Cologna, 2017). The pI of the LEAFY protein was lower and acidic for *Arabidopsis* sp.*, Physcomitrella* sp., and *K. subtile,* whereas it was higher and alkaline for *Coleochaete scutata*, *Ceratopteris* sp., and *Picea* sp. II reveals the stability of the protein in both *in vivo* and *in vitro* conditions. The II value above 40 indicates that the protein is unstable, while the value below 40 indicates that it is stable (Guruprasad et al., 1990; Gamage et al., 2019). Our findings reveal that the II value of LEAFY proteins ranged from 39.58 to 62.51, which indicates that the structure of proteins is unstable. The AI index is essential to determine the thermal stability potential of the amino acid sequence, and thermal stability increases with a higher AI value (Panda and Chandra, 2012). The AI of LEAFY proteins ranged from 62.57 to 74.87, which indicates that the protein is thermally stable at a wide range of temperature (20–45 °C) (Enany, 2014; Ikai, 1980). Similar to stability studies, it is essential to evaluate the hydrophobic and hydrophilic nature of proteins by using the GRAVY score. The negative GRAVY value indicates that the protein is hydrophilic, and a positive value indicates

they are hydrophobic; the value usually ranges from −2 to +2 (Kyte and Doolittle, 1982; Kaur and Pati, 2018). The GRAVY score of the LEAFY proteins for all model organisms was negative, which indicated a high number of interactions with water.

Secondary structure prediction and analysis

Secondary structure formation is the initial step in protein folding to attain its functional shape (Pirovano and Heringa, 2010). The most accurate and reliable prediction of protein sequence structure is a challenging aspect of computational biology. The PHYRE2 program uses homology modeling techniques that help in structure prediction, function prediction, domain analysis, and mutation analysis (Kelley et al., 2015). As an essential step in the prediction of tertiary structures, PHYRE2 was first used for determining secondary structures such as an alpha helix, beta-strands, and irregular coil regions in the polypeptide chain of amino acids, which determine protein activity, interactions, and functions at the molecular level (Kelley et al., 2015). Homologous sequences for the LEAFY protein of each model organism (query sequence) were detected from multiple sequence alignment using PSI-Blast from the PHYRE2 server. The secondary structure prediction and disorder region prediction was made by Psi-Pred and Diso-Pred programs. In the predicted secondary structures of LFY proteins, it was found that the percentage of alpha-helix (α) structure ranged from 58 to 72% and that of beta-strands (β) ranged from 0 to 4% (Table 6). Proteins with alpha-helix (α) ≥ 40% and beta-strands (β) ≤ 5% were categorized as alpha protein class (Chou, 1995). Thus, these secondary structures belong to alpha protein classes. In alpha helices and beta strands, the potential to tolerate mutation differs significantly. Helices are more robust to mutation than strands or coils due to the noncovalent interactions of residues in the secondary structure units without a structural change (Abrusán and Marsh, 2016). The contact density among residues determines the acceptance of mutation without destabilizing the protein fold (England and Shakhnovich, 2003; Shakhnovich et al., 2005; Nemtseva et al., 2019). Thus, mutations result in lesser structural change.

The protein region without a secondary structure is a disordered region that affects the 3D structure of a protein. The disordered region binds with the partner molecule (nucleic acid, another protein, etc.) and thus exists as a structured protein (Dyson and Wright, 2005; Ishida and Kinoshita, 2008; Uversky, 2019). It often plays a functional role and is commonly involved in transcription, translation, and cell signaling (Van Der Lee et al., 2014; Hsu et al., 2020). Mutations in the disordered regions result in inappropriate protein folding (Uversky et al., 2005; Dyson, 2011); thus, the prediction of disordered regions is pivotal for the structure and function analysis of a protein sequence. The Diso-Pred server predicted the presence of 31% to 58% disordered regions in the tested LEAFY homologs, with the highest value found for *C. thalictroides* and the lowest value for *P. patens*, indicating that the LEAFY homolog is dynamic in the fern *C. thalictroides* and stable in the moss *P. patens* (Table 6).

Homology modeling

The determination of three-dimensional structure is essential as it provides insights into biochemical functions and protein interactions (Ittisoponpisan et al., 2019). The model was constructed by identifying sequence similarity (homologous sequence) with a target sequence and alignment with a suitable template from PDB (Fiser, 2010). The protein model was constructed using SWISS-MODEL based on the sequence and alignment with the most appropriate structural template for the LEAFY protein of model organisms from the PDB database, with GMQE and QMEAN Z-score values (Biasini et al., 2014). GMQE (Global Model Quality Estimation) estimates the model's accuracy and is expressed as a number between 0 and 1. The QMEAN Z-score reports the reliability of the model quality estimation, and a QMEAN Z-score of around 0 indicates a good quality model (Benkert et al., 2011; Biasini et al., 2014; Waterhouse et al., 2018). Protein structure with a sequence homology of > 40% shares similarity with other protein structures, whereas sequence homology < 25% results in significant structural differences. Thus, a reliable protein structure cannot be predicted based on homology modeling when sequence homology is < 25% (Venclovas, 2011). Here, 32 template matches of the LEAFY protein (target sequence) for *K. subtile* and 43 template matches for *Arabidopsis* sp. were reported, in which the best and highest sequence similarity was reported for the template 2vy2.1A, which is the LEAFY protein structure of *A. thaliana* complex with DNA from Ag – I promoter (Hamès et al., 2008). Twenty-eight templates

Table 7. Template identification results for each LEAFY protein sequence in the SWISS-MODEL tool

Protein ID	GMQE	QMEAN	Template	Sequence identity [%]	Description
AHJ90707.1	0.21	−1.22	2vy2.1A	41.07	Protein Leafy
AHJ90705.1	0.24	−1.35	4bhk.1.A	54.48	Floricaula/Leafy homolog1
BAD91044.1	0.26	−0.63	4bhk.1.A	97.62	Floricaula/Leafy homolog1
BAD91043.1	0.26	−0.54	4bhk.1.A	100	Floricaula/Leafy homolog1
ABF74516.1	0.30	−0.24	4bhk.1.A	80.17	Floricaula/Leafy homolog1
ABF74513.1	0.29	−0.28	4bhk.1.A	80.17	Floricaula/Leafy homolog1
ABF74512.1	0.30	−0.24	4bhk.1.A	80.17	Floricaula/Leafy homolog1
AAV49504.1	0.26	−0.42	4bhk.1.A	79.17	Floricaula/Leafy homolog1
AKA55658.1	0.26	−0.42	4bhk.1.A	79.59	Floricaula/Leafy homolog1
AAM27931.1	0.24	−0.28	2vy2.1.A	100	Protein Leafy
AAM27932.1	0.24	−0.28	2vy2.1.A	100	Protein Leafy
AAM27941.1	0.24	−0.28	2vy2.1.A	100	Protein Leafy

Table 8. MolProbity results of 3D models of LEAFY protein generated after structure assessment

Protein ID	Ramachandran favoured [%]	Ramachandran outliers [%]
AHJ90707.1	97.52	0.00
AHJ90705.1	96.75	0.81
BAD91044.1	98.08	0.00
BAD91043.1	98.08	0.00
ABF74516.1	98.18	0.00
ABF74513.1	98.18	0.00
ABF74512.1	98.18	0.00
AAV49504.1	98.72	0.00
AKA55658.1	97.79	0.00
AAM27931.1	98.14	0.00
AAM27932.1	98.14	0.00
AAM27941.1	98.14	0.00

match of the LEAFY protein for *C. scutata*, 50 templates match for *Physcomitrella* sp., 8–9 templates match for *Ceratopteris* sp., 30 templates match for *P. abies*, and 50 templates match for *P. sitchensis* were reported, in which the best and highest sequence similarity was reported for the template 4bhk.1. A FLORICAULA/ LEAFY HOMOLOG 1 codes for the transcription factor LEAFY in mosses, which interacts with DNA (Sayou et al., 2014) (Table 7). The 3D models for the LEAFY protein in each model organism were constructed based on the template, represented by rainbow color from N-terminal to C-terminal (Fig. 4).

Structure evaluation

The structural validation of the predicted protein models is crucial as the predicted structures may con-

tain substantial errors. Because the structure is related to function, the generated model should be error-free. The structure evaluation was conclusively performed by Ramachandran plot analysis (Carugo and Djinović-Carugo, 2013). The MolProbity tool was accessed with the PDB files of protein structures for Ramachandran analysis, which helps to determine the protein geometry (Chen et al., 2010). Ramachandran plot generates the graphical representation of the allowed and forbidden regions of torsion angles, phi (φ) and psi (ψ), by plotting phi (φ) on the *x*-axis and psi (ψ) on the *y*-axis. Torsion angles of amino acid residues in the protein structure form secondary structures corresponding to the allowed and disallowed regions (Saravanan and Selvaraj, 2017). The dark-colored region in the Ramachandran plot is considered as the most favorable, the light-colored region as favorable, and the white region is disallowed and regarded as forbidden in the four quadrants of the Ramachandran plot structure. The four-quadrant plot helps in analysis of possible combination of torsion angles of the proposed protein. An optimal quality structure contains all the combinations of torsion angles in the allowed region, whereas if all sets of torsion angles occupy a forbidden region, it reflects a poor-quality homology model, resulting in steric hinderance (Røgen, 2021). The conformation of phi-psi torsion angles of the predicted LEAFY protein structure was satisfactory, as > 96% of all residues were present in the allowed region (Table 8), indicating a good quality model. There were no outliers in the Ramachandran plot for all the plant species, except for *C. scutata* (Fig. 5); thus, it can be considered as a good quality model suitable for further application (Muhammed and Aki Yalcin, 2019).

Conclusions

The present study revealed that LFY genes are conserved in Charophyte green algae, moss, fern, gymnosperms, and angiosperms. Domain analysis showed that the LEAFY proteins in all plant species shared two conserved domains, namely C_LFY_FLO and SAM_LFY. The physicochemical characterization reported that the LEAFY protein has an unstable structure, indicating its dynamic nature. The protein is thermally stable and hydrophilic in nature. In LEAFY protein sequences, most conserved, conservative mutated, and semi-conservative mutated sequences were predicted as helical structures. Beta strands were conserved in all plant species with only sequence differences in charophyte green algae, which is a unique variation in LFY evolution. The 3D models generated from the LEAFY protein sequences were of good quality and will help to corroborate structural and functional analysis. The results of phylogenetic analysis indicated a very early mutation that led to the formation of two distinct clusters, one leading to angiosperms and the other to gymnosperms. The LFY gene of the gymnosperms showed homology with that of mosses and pteridophytes as compared to that of orchids, monocots, and other flowering plants.

Acknowledgments

Our sincere thanks to Rev Fr Dr Praveen Martis SJ, Principal; Dr Shreelalitha Suvarna, HOD, Department of Post Graduate Studies and Research in Biotechnology; and Rev Fr Dr Leo D'Souza SJ, Director, Dr. Küppers Laboratory of Applied Biology, St Aloysius College (Autonomous) for providing the research facilities.

References

Abrusán G., Marsh J.A. (2016) *Alpha helices are more robust to mutations than beta-strands*. PLoS Comput. Biol. 12(12): e1005242.

Benkert P., Biasini M., Schwede T. (2011) *Toward the estimation of the absolute quality of individual protein structure models*. Bioinformatics 27(3): 343–350.

Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Schwede T. (2014) *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information*. Nucl. Acids Res. 42(W1): W252–W258.

Brunkard J.O., Runkel A.M., Zambryski P.C. (2015) *Comment on "A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity"*. Science 347(6222): 621–621.

Carugo O., Djinović-Carugo K. (2013) *A proteomic Ramachandran plot (PRplot)*. Amino Acids 44(2): 781–790.

Cavasotto C.N., Phatak S.S. (2009) *Homology modelling in drug discovery: current trends and applications*. Drug Discov. Today 14(13–14): 676–683.

Chatzou M., Magis C., Chang J.M., Kemena C., Bussotti G., Erb I., Notredame C. (2016) *Multiple sequence alignment modeling: methods and applications*. Brief. Bioinformat. 17(6): 1009–1023.

Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., Richardson D.C. (2010) *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta Crystallogr. D: Biol. Crystallogr. 66 (Pt 1): 12–21.

Chou K.C. (1995) *A novel approach to predicting structural protein classes in a (20–1) – D amino acid composition space*. Proteins. 21(4): 319–344.

Cove D.J., Perroud P.F., Charron A.J., McDaniel S.F., Khandelwal A., Quatrano R.S. (2009) *The moss Physcomitrella patens: a novel model system for plant development and genomic studies*. [in:] *Emerging Model Organisms: a laboratory manual*, vol. 1, NY: CSHL Press: 592.

Dhar S., Sood V., Lohiya G., Devenderan H., Katti D.S. (2020) *Role of physicochemical properties of protein in modulating the nanoparticle-bio interface*. J. Biomed. Nanotechnol. 16(8): 1276–1295. http://doi.org/10.1166/jbn. 2020. 2958

Domozych D., Popper Z.A., Sorensen I. (2016) *Charophytes: evolutionary giants and emerging model organisms*. Front. Plant Sci. 7: 1470.

Dornelas M.C., Rodriguez A.P.M. (2005) *A FLORICAULA/ LEAFY gene homolog is preferentially expressed in developing female cones of the tropical pine Pinus caribaea var. caribaea*. Genet. Mol. Biol. 28(2): 299–307.

Dyson H.J. (2011) *Expanding the proteome: disordered and alternatively folded proteins*. Q. Rev. Biophys. 44(4): 467–518.

Dyson H.J., Wright P.E. (2005) *Intrinsically unstructured proteins and their functions*. Nat. Rev. Mol. Cell Biol. 6(3): 197–208.

Edgar R.C., Batzoglou S. (2006) *Multiple sequence alignment*. Curr. Opin. Struct. Biol. 16(3): 368–373.

Enany S. (2014) *Structural and functional analysis of hypothetical and conserved proteins of Clostridium tetani*. J. Infect. Public Health 7(4): 296–307.

England J.L., Shakhnovich E.I. (2003) *Structural determinant of protein designability*. Phys. Rev. Lett. 90(21): 218101.

Eswar N., John B., Mirkovic N., Fiser A., Ilyin V.A., Pieper U., Stuart A.C., Marti-Renom M.A., Madhusudhan M.S., Yerkovich B., Sali A. (2003) *Tools for comparative protein structure modeling and analysis*. Nucl. Acids Res. 31(13): 3375–3380.

Ezhova T.A. (1999) *Arabidopsis thaliana (L.) Heynh. kak model'nyĭ ob"ekt dlia izycheniiageneticheskogo kontrolia morfogeneza* [*Arabidopsis thaliana (L.) Heynh. as a model object for studying genetic control of morphogenesis*]. Genetika 35(11): 1522–1537.

Fiser A. (2010) *Template-based protein structure modeling*. [in:] *Methods in molecular biology (methods and protocols)*. Totowa: Humana Press.

Gamage D.G., Gunaratne A., Periyannan G.R., Russell T.G. (2019) *Applicability of instability index for in vitro protein stability prediction*. Protein Pept. Lett. 26(5): 339–347.

Gao B., Chen M., Li X., Zhang J. (2019) *Ancient duplications and grass-specific transposition influenced the evolution of LEAFY transcription factor genes*. Commun. Biol. 2(1): 1–10.

Grandi V., Gregis V., Kater M.M. (2012) *Uncovering genetic and molecular interactions among floral meristem identity genes in Arabidopsis thaliana*. Plant J. 69(5): 881–893.

Gupta C.L., Akhtar S., Bajpai P. (2014) *In silico protein modeling: possibilities and Gupta limitations*. EXCLI J. 13: 513–515.

Guruprasad K., Reddy B.B., Pandit M.W. (1990) *Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence*. Protein Eng. Des. Sel. 4(2): 155–161.

Hall B.G. (2013) *Building phylogenetic trees from molecular data with MEGA*. Mol. Biol. Evol. 30(5): 1229–1235.

Hamès C., Ptchelkine D., Grimm C., Thevenon E., Moyroud E., Gérard F., Martiel J.L., Benlloch R., Parcy F., Müller C.W. (2008) *Structural basis for LEAFY floral switch function and similarity with helix turn helix proteins*. EMBO J. 27(19): 2628–2637.

Hasani H.J., Barakat K. (2017) *Homology modeling: an overview of fundamentals and tools*. Int. Rev. Model. Simul. 10(2): 1–14.

Hickok L.G., Warne T.R., Fribourg R.S. (1995) *The biology of the fern Ceratopteris and its use as a model system*. Int. J. Plant Sci. 156(3): 332–345.

Hofer J., Turner L., Hellens R., Ambrose M., Matthews P., Michael A., Ellis N. (1997) *UNIFOLIATA regulates leaf and flower morphogenesis in pea*. Curr. Biol. 7(8): 581–587.

Hsu C.C., Buehler M.J., Tarakanova A. (2020) *The order-disorder continuum: linking predictions of protein structure and disorder through molecular simulation*. Sci. Rep. 10(1): 1–14.

Ikai A. (1980) *Thermostability and aliphatic index of globular proteins*. J. Biochem. 88(6): 1895–1898.

Ishida T., Kinoshita K. (2008) *Prediction of disordered regions in proteins based on the meta approach*. Bioinformatics 24(11): 1344–1348.

Ittisoponpisan S., Islam S.A., Khanna T., Alhuzimi E., David A., Sternberg M.J. (2019) *Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated?* J. Mol. Biol. 431(11): 2197–2212.

Kaur A., Pati P.K., Pati A.M., Nagpal A.K. (2020) *Physicochemical characterization and topological analysis of pathogenesis-related proteins from Arabidopsis thaliana and Oryza sativa using in-silico approaches*. PLoS One 15(9): e0239836.

Kaur G., Pati P.K. (2018) *In silico physicochemical characterization and topology analysis of respiratory burst oxidase homolog (Rboh) proteins from Arabidopsis and rice*. Bioinformation 14(3): 93.

Kc D.B. (2017) *Recent advances in sequence-based protein structure prediction*. Brief Bioinform. 18(6): 1021–1032.

Kelley L.A., Mezulis S., Yates C.M., Wass M.N., Sternberg M.J. (2015) *The Phyre2 web portal for protein modeling, prediction and analysis*. Nat. Protoc. 10(6): 845–858.

Kuhlman B., Bradley P. (2019) *Advances in protein structure prediction and design*. Nat. Rev. Mol. Cell Biol. 20(11): 681–697.

Kumar S., Stecher G., Tamura K. (2016) *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets*. Mol. Biol. Evol. 33: 1870–1874.

Kyte J., Doolittle R.F. (1982) *A simple method for displaying the hydropathic character of a protein*. J. Mol. Biol. 157(1): 105–132.

Lam S.D., Das S., Sillitoe I., Orengo C. (2017) *An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences*. Acta Crystallogr. D: Struct. Biol. 73(8): 628–640.

Lee J., Freddolino P.L., Zhang Y. (2017) *Ab initio protein structure prediction*. [in:] *From protein structure to function with bioinformatics*. Ed. Rigden D.J., Dordrecht: Springer: 3–35.

Lesk A.M., Chothia C.H. (1986) *The response of protein structures to amino-acid sequence changes*. Phil. Trans. R. Soc. A. 317(1540): 345–356.

Martí-Renom M.A., Stuart A.C., Fiser A., Sánchez R., Melo F., Šali A. (2000) *Comparative protein structure modeling of genes and genomes*. Annu. Rev. Biophys. Biomol. Struct. 29(1): 291–325.

Monniaux M., McKim S.M., Cartolano M., Thévenon E., Parcy F., Tsiantis M., Hay A. (2017) *Conservation vs divergence in LEAFY and APETALA1 functions between Arabidopsis thaliana and Cardamine hirsuta*. New Phytol. 216(2): 549–561.

Moyroud E., Kusters E., Monniaux M., Koes R., Parcy F. (2010) *LEAFY blossoms*. Trends Plant Sci. 15(6): 346–352.

Moyroud E., Monniaux M., Thévenon E., Dumas R., Scutt C.P., Frohlich M.W., Parcy F. (2017) *A link between LEAFY and B gene homologues in Welwitschia mirabilis sheds light on ancestral mechanisms prefiguring floral development*. New Phytol. 216(2): 469-481.

Muhammed M.T., Aki Yalcin E. (2019) *Homology modeling in drug discovery: overview, current applications, and future perspectives*. Chem. Biol. Drug Des. 93(1): 12–20.

Nagano K. (1973) *Logical analysis of the mechanism of protein folding. I. Predictions of helices, loops and beta-structures from primary structure*. J. Mol. Biol. 75: 401–420.

Nemtseva E.V., Gerasimova M.A., Melnik T.N., Melnik B.S. (2019) *Experimental approach to study the effect of mutations on the protein folding pathway*. PloS One 14(1): e0210361.

Nystedt B., Street N.R., Wetterbom A., Zuccolo A., Lin Y.C., Scofield D.G., Vezzi F., Delhomme N., Giacomello S., Alexeyenko A., Vicedomini R. (2013) *The Norway spruce genome sequence and conifer genome evolution*. Nature 497(7451): 579–584.

Ortiz A.R., Kolinski A., Skolnick J. (1998) *Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments*. J. Mol. Biol. 277(2): 419–448.

Panchenko A.R., Marchler-Bauer A., Bryant S.H. (2000) *Combination of threading potentials and sequence profiles improves fold recognition*. J. Mol. Biol. 296(5): 1319–1331.

Panda S., Chandra G. (2012) *Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates*. Bioinformation 8(18): 891.

Pergande M.R., Cologna S.M. (2017) *Isoelectric point separations of peptides and proteins*. Proteomes. 5(1): 4.

Pieper U., Eswar N., Davis F.P., Braberg H., Madhusudhan M.S., Rossi A., Marti-Renom M., Karchin R., Webb B.M., Eramian D. Shen M.Y., (2006) *MODBASE: a database of annotated comparative protein structure models and associated resources*. Nucl. Acids Res. 34(suppl_1): D291–D295.

Pirovano W., Heringa J. (2010) *Protein secondary structure prediction*. [in:] *Data mining techniques for the life sciences*. Ed. Carugo O., Eisenhaber F., Humana Press: Springer Protocols: 327–348.

Plackett A.R., Conway S.J., Hazelton K.D.H., Rabbinowitsch E.H., Langdale J.A., Di Stilio V.S. (2018) *LEAFY maintains apical stem cell activity during shoot development in the fern Ceratopteris richardii*. eLife 7: e39625.

Renzaglia K.S., Warne T.R. (1995) *Ceratopteris: an ideal model system for teaching plant biology*. Int. J. Plant Sci. 156(3): 385–392.

Røgen P. (2021) *Quantifying steric hindrance and topological obstruction to protein structure superposition*. Algorithms Mol. Biol. 16(1): 1–19.

Šali A., Blundell T.L. (1993) *Comparative protein modelling by satisfaction of spatial restraints*. J. Mol. Biol. 234(3): 779–815.

Saravanan K.M., Selvaraj S. (2017) *Dihedral angle preferences of amino acid residues forming various non-local interactions in proteins*. J. Biol. Phys. 43(2): 265–278.

Sayou C., Monniaux M., Nanao M.H., Moyroud E., Brockington S.F., Thévenon E., Wong G.K.S. (2014) *A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity*. Science 343(6171): 645–648.

Sayou C., Nanao M.H., Jamin M., Posé D., Thévenon E., Grégoire L., Schmid M. (2016) *A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor*. Nat. Commun. 7: 11222.

Shakhnovich B.E., Deeds E., Delisi C., Shakhnovich E. (2005) *Protein structure and evolutionary history determine sequence space topology*. Genome Res. 15(3): 385–392.

Silva C.S., Puranik S., Round A., Brennich M., Jourdain A., Parcy F., Zubieta C. (2016) *Evolution of the plant reproduction master regulators LFY and the MADS transcription factors: the role of protein structure in the evolutionary development of the flower*. Front. Plant Sci. 6: 1193.

Simons K.T., Strauss C., Baker D. (2001) *Prospects for ab initio protein structural genomics*. J. Mol. Biol. 306(5): 1191–1199.

Siriwardana N.S., Lamb R.S. (2012) *The poetry of reproduction: the role of LEAFY in Arabidopsis thaliana flower formation*. Int. J. Dev. Biol. 56(4): 207–221.

Skolnick J., Kihara D. (2001) *Defrosting the frozen approximation: PROSPECTOR – a new approach to threading*. Proteins: Struct. Funct. Bioinf. 42(3): 319–331.

Studer G., Tauriello G., Bienert S., Waterhouse A.M., Bertoni M., Bordoli L., Schwede T., Lepore R. (2019) *Modeling of protein tertiary and quaternary structures based on evo-*

lutionary information. [in:] *Computational methods in protein evolution*. Ed. Sikosek T., Humana Press, New York: Springer Protocols: 301–316

Tanahashi T., Sumikawa N., Kato M., Hasebe M. (2005) *Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens.* Development 132(7): 1727–1736.

Tran Q., Wallinga M. (2017) *A novel method for multiple sequence alignment using morphing techniques*. J. Health Inform. Manag. 1(2): 2.

Uversky V.N. (2019) *Intrinsically disordered proteins and their "mysterious" (meta) physics.* Front. Phys. 7: 10.

Uversky V.N., Oldfield C.J., Dunker A.K. (2005) *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signalling*. J. Mol. Recognit. 18(5): 343–384.

Van Der Lee R., Buljan M., Lang B., Weatheritt R.J., Daughdrill G.W., Dunker A.K., Fuxreiter M., Gough J., Gsponer J., Jones D.T., et al. (2014) *Classification of intrinsically disordered regions and proteins*. Chem. Rev. 114(13): 6589–6631.

Vázquez-Lobo A., Carlsbecker A., Vergara-Silva F., Alvarez-Buylla E.R., Piñero D., Engström P. (2007) *Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera Picea, Podocarpus, and Taxus: implications for current evo-devo hypotheses for gymnosperms*. Evol. Dev. 9(5): 446–459.

Venclovas Č. (2011) *Methods for sequence–structure alignment*. [in:] *Homology modeling*. Ed. Orry A., Abagyan R. Humana Press: Springer Protocols: 55–82.

Venkatesan A., Gopal J., Candavelou M., Gollapalli, S., Karthikeyan K. (2013) *A computational approach for protein structure prediction*. Healthc. Inform. Res. 19(2): 137–147.

Villimová V. (2012) *Evolution of gene network controlling plant reproductive development*. Biosci. Master Rev. Univ. Lyon.

Wang L.L., Liang H.M., Pang J.L., Zhu M.Y. (2004) [*Regulation network and biological roles of LEAFY in Arabidopsis thaliana in floral development*] Yi Chuan; Hereditas 26(1): 137–142.

Wang L., Jiang T. (1994) *On the complexity of multiple sequence alignment*. J. Comput. Biol. 1(4): 337–348.

Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Schwede T. (2018) *SWISS-MODEL: homology modelling of protein structures and complexes*. Nucl. Acids Res. 46(W1): W296–W303.

Weigel D., Alvarez J., Smyth D.R., Yanofsky M.F., Meyerowitz E.M. (1992) *LEAFY controls floral meristem identity in Arabidopsis*. Cell. 69(5): 843–859.

Weigel D., Nilsson O. (1995) *A developmental switch sufficient for flower initiation in diverse plants*. Nature 377(6549): 495–500.

Wilhelmsson P.K., Mühlich C., Ullrich K.K., Rensing S.A. (2017) *Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae*. Genome Biol. Evol. 9(12): 3384–3397.

Xu Y., Liu Z., Cai L., Xu D. (2007) *Protein structure prediction by protein threading*. [in:] *Computational methods for protein structure prediction and modeling*. Ed. Xu Y., Xu D., Liang J. New York: Springer: 1–42.