

# ANALISIS *IN SILICO* PROTEIN BIOSINTESIS LIGNIN PADA TANAMAN SORGUM (*Sorghum bicolor* L. Moench) [*In Silico Analysis of Lignin Biosynthesis Protein in Sorghum (Sorghum bicolor* L. *Moench)*]

Vincentia Esti Windiastr<sup>1,3✉\*</sup>, Sintho Wahyuning Ardie<sup>2</sup>, dan Satya Nugroho<sup>3</sup>

<sup>1</sup>Program Studi Pemuliaan dan Bioteknologi Tanaman, IPB

<sup>2</sup>Departemen Agronomi dan Hortikultura, IPB

<sup>3</sup>Pusat Riset Rekayasa Genetika, BRIN

\*Email: vinc001@brin.go.id

## ABSTRACT

As one of the natural products, lignin can be utilized as raw materials for industries, biofuel, and bioenergy. With the increasing demand for lignin, it is important to find a renewable and environmentally friendly source to produce lignin. Sorghum has the potential to be used as a biomass source to produce lignin because it annually produces 19–47 ton/ha lignocellulose. However, to utilize them optimally, it is important to increase the lignin content in the crops. Genetic engineering for lignin content modification is key to producing lignin optimally. Unfortunately, the information about lignin biosynthesis genes in sorghum is still limited. This experiment has aim to explore more information on the lignin biosynthesis proteins in sorghum (*Sorghum bicolor* L. Moench) and their respective homologs from other monocot plants, via studying their phylogenetic trees, genetic distances, domain compositions, and conserved regions. 10 proteins from 10 monolignol biosynthesis protein families in sorghum, based on Sorghum Functional Database (SFDB), had been analyzed using some bioinformatics tools, such as Blast-p, Mega-X, Conserved Domain Database, InterPro and MEME. The results of the analysis illustrate that the clustering of lignin biosynthetic proteins follows the divergence of monocot plants, where sorghum proteins clustered together with *Mischanthus*, sugarcane, and maize. The best bootstrap was achieved by the COMT protein phylogeny tree. The alignment of COMT family protein from sorghum (XP\_002445083.1) showed that this protein has similar functional characters with other COMT family proteins from other representative monocot plants and also from Arabidopsis plant. This suggests that XP\_002445083.1 is a putative COMT protein from sorghum.

**Keywords:** biomass, lignin biosynthesis protein, genetic distance, phylogeny, monocots

## ABSTRAK

Lignin adalah bahan alam yang dapat dimanfaatkan sebagai bahan mentah dalam industri biofuel dan bioenergi. Peningkatan produksi lignin mendorong tersedianya sumber lignin terbaharukan dan ramah lingkungan. Sorgum berpotensi digunakan sebagai sumber biomassa untuk produksi lignin. Produksi lignoselulosa dari sorgum per tahunnya mencapai 19–47 ton/ha. Meskipun demikian, untuk memanfaatkan sorgum sebagai sumber lignin secara optimal diperlukan peningkatan kandungan total lignin pada tanaman sorgum. Rekayasa genetika dalam modifikasi kandungan lignin dapat menjadi jawaban. Namun, informasi mengenai gen biosintesis lignin masih terbatas. Untuk memperoleh informasi protein biosintesis lignin pada sorgum (*Sorghum bicolor* L.Moench) dan protein homolognya dari tanaman monokotil yang lain, analisis bioinformatika mengenai protein biosintesis lignin telah dilakukan. Analisis dilakukan menggunakan program Blast-p, Mega-X, Conserved Domain Database, InterPro dan MEME. Hasil dari analisis yang dilakukan menggambarkan bahwa protein biosintesis lignin membentuk kluster searah dengan arah divergensi taksonomi tanaman monokotil. Protein lignin dari sorgum mengelompok bersama *Mischanthus*, tebu dan jagung. Pohon kekerabatan protein COMT menghasilkan bootstrap tertinggi untuk protein lignin sorgum di antara pohon filogeni protein-protein yang lain. Hasil penyejajaran keluarga protein COMT sorgum (XP\_002445083.1) dengan protein COMT dari *Mischanthus*, tebu, jagung dan Arabidopsis menunjukkan bahwa protein XP\_002445083.1 mempunyai karakter fungsional yang saling mirip. Sehingga dapat disimpulkan bahwa XP\_002445083.1 diduga adalah protein COMT dari sorgum.

**Kata Kunci:** biomassa, jarak genetik, monokotil, pohon kekerabatan, protein biosintesis lignin

## PENDAHULUAN

Sumber-sumber energi yang terbarukan dan ramah lingkungan mulai banyak dicari beberapa tahun terakhir. Pemanfaatan biomassa sebagai sumber energi adalah salah satu alternatif untuk mendapatkan energi yang terbarukan dan ramah lingkungan. (Baskar *et al.*, 2012). Residu pertanian, terutama dari tanaman monokotil berupa jerami dan brangkasan memiliki potensi yang besar sebagai sumber biomassa. *World Bioenergy Association* (WBA) melaporkan bahwa pada tahun 2019 limbah pertanian di dunia tersedia sebesar 9,4 milliar ton per tahun, dengan tanaman serealia, terutama jagung, padi dan gandum sebagai kontributor utama (WBA, 2019). Biomassa yang berupa lignoselulosa potensial dimanfaatkan sebagai bahan mentah dalam industri karena kemelimpahannya, sifatnya

yang terbarukan serta dapat dimanfaatkan dalam berbagai macam bentuk (Baskar *et al.*, 2012).

Lignoselulosa terdiri atas selulosa, hemiselulosa dan lignin (Baskar *et al.*, 2012). Pemanfaatan lignin dalam berbagai bidang industri, mulai dari bioenergi, biofuel dan biopolimer, membutuhkan lignin dengan kriteria kandungan dan komposisi lignin tertentu (Irmer, 2017; Grossman dan Vermerris, 2019). Jumlah total lignin, struktur dan komposisi lignin sangat bervariasi antar-spesies, jaringan dan tipe sel tanaman (Vanholme *et al.*, 2019). Berdasarkan kelompok tanaman, lignin dibagi menjadi tiga kelas secara luas yaitu: lignin kayu keras (Angiospermae), lignin kayu lunak (Gymnospermae) dan lignin rerumputan (Graminaceous) (Bajpai, 2017). Komposisi lignin

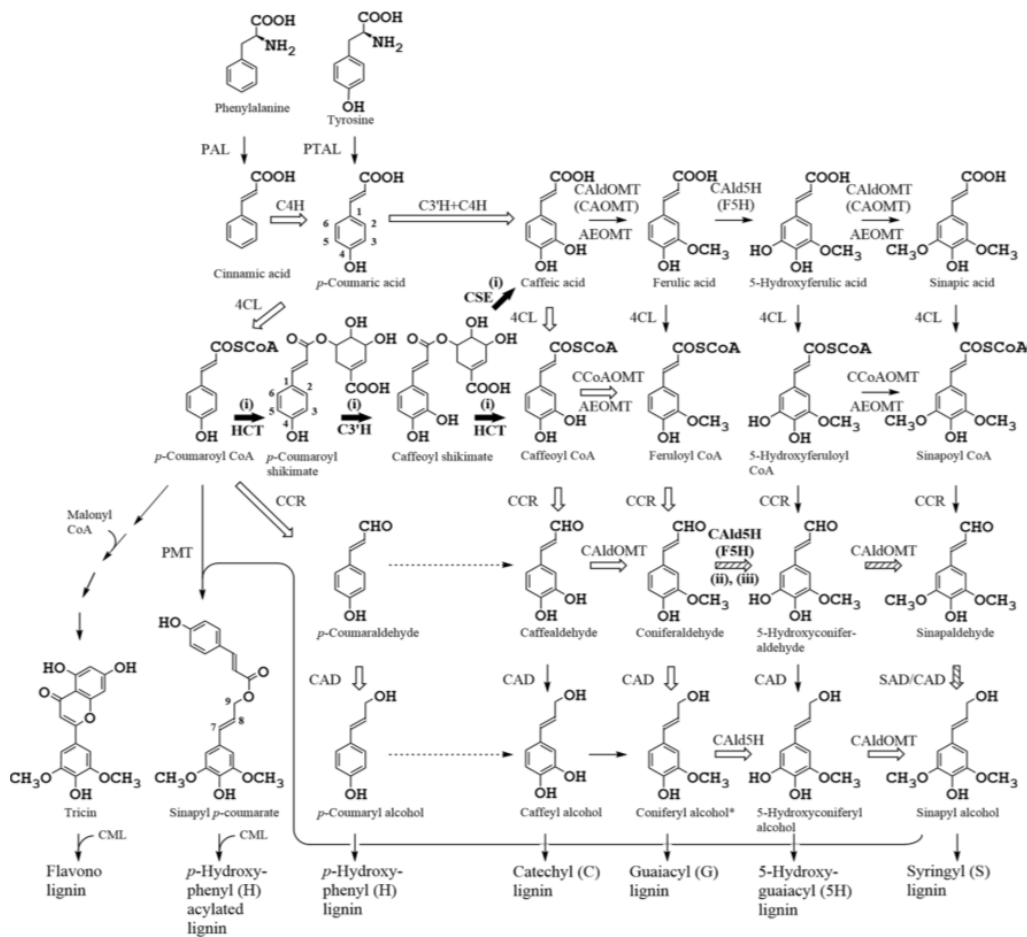
\*Kontributor Utama

\*Diterima: 27 Juni 2023 - Diperbaiki: 5 Juli 2023 - Disetujui: 24 Agustus 2023

kayu keras tanaman Angiospermae dari kelas dikotil, sebagian besar terdiri atas monolignol unit Guaiacyl (G) dan Syringyl (S) dengan sedikit jejak unit p- Hydroxyphenyl (H), sedangkan pada tanaman rerumputan (*Graminaceous*) mempunyai komposisi yang mirip dengan dikotil dengan unit G dan S namun total unit H lebih banyak dari pada dikotil. Lignin pada monokotil biasanya diasilasi dengan p-coumarat (Bonnawitz dan Chapple, 2010; Umezawa, 2018).

Penambangan lignin dari biomassa pertanian, yang sebagian besar adalah tanaman monokotil, menghadapi tantangan karena total kandungan lignin tanaman monokotil lebih rendah dari pada tanaman dikotil yang telah biasa digunakan sebagai

sumber lignin untuk industri (Vanholme, 2019). Hal ini dapat diatasi dengan cara memanipulasi gen-gen kunci yang terkait dengan biosintesis lignin untuk mendapatkan tanaman dengan total kandungan lignin yang lebih tinggi. Namun sayangnya, informasi mengenai gen-gen tersebut masih terbatas terutama pada tanaman monokotil. Informasi mengenai gen-gen lignin yang berasal dari tanaman monokotil dengan biomassa besar seperti sorgum, tebu, jagung dan *Mischanthus* akan bermanfaat dalam pengembangan tanaman biomassa.



**Gambar 1.** Jalur utama biosintesis lignin menjadi monolignol beserta gen-gen yang berperan dalam biosintesis lignin. Anak panah tebal mewakili jalur utama untuk biosintesis monolignol. (*The main pathway of monolignol biosynthesis along with the genes that play roles in lignin biosynthesis. Thick arrows represent major pathways for monolignol biosynthesis.*)

Keterangan (abbreviations): C4H, Cinnamate 4-hydroxylase; C3'H, P-coumarate 3-hydroxylase; CALDOMT, 5-hydroxyconiferaldehyde O-methyltransferase; 4CL, 4-hydroxycinnamate CoA Ligase; HCT, Hydroxycinnamoyl CoA : Shikimate/Quinate Hydroxycinnamoyl Transferase; CCOAOMT, Caffeoyl CoA O-methyltransferase; CCR, Cinnamoyl Coa Reductase; CALD5H, Coniferaldehyde 5-hydroxylase; CAD, Cinnamyl Alcohol Dehydrogenase; SAD, Sinapyl Alcohol Dehydrogenase; F5H, Ferulate 5-hydroxylase; CAOMT, Caffeic Acid O-methyltransferase; AeOMT, hydroxycinnamic acids/Hydroxycinnamoyl CoA esters O-methyltransferase; C3'H, p-coumaroyl Shikimate/Quinate 3-hydroxylase; PAL, Phenylalanine Ammonia Lyase; PTAL, Bifunctional Ammonia Lyase; CSE, Caffeoyl Shikimate Esterase; PMT, p-Coumaroyl CoA:monolignol transferase. CML, Canonical Monolignols (p-Coumaryl Alcohol, Coniferyl Alcohol, and Sinapyl Alcohol) (Umezawa 2018).

Sorgum (*Sorghum bicolor* L. Moench) merupakan salah satu tanaman sereal yang potensial dimanfaatkan sebagai sumber lignoselulosa. Sorgum dapat menghasilkan lignoselulosa sebesar 19–47 ton/ha per tahun, sehingga mulai dimanfaatkan sebagai sumber energi alternatif dan bahan baku beberapa industri (Hattori dan Morita, 2010). Potensi biomassa sorgum tidak hanya dari sisi jumlah produksi biomassa per tahun, namun juga karena sorgum merupakan model genetik dan referensi genom khususnya untuk tanaman rerumputan C4. Sorgum telah memiliki *platform* terintegrasi untuk pemuliaan-genetika-genomik. Selain itu sorgum banyak dimanfaatkan di seluruh dunia baik untuk biji-bijian, pakan ternak, gula maupun tanaman bioenergi. Hal ini karena plasma nutfah sorgum yang beragam dan kemampuan adaptasinya yang luas. Sebagai model genetik, sorgum berpotensi untuk digunakan sebagai referensi dalam penemuan gen dan analisis genom fundamental, seperti studi gen terkait biosintesis lignin (Olson *et al.*, 2012; Mullet *et al.*, 2014).

Studi mengenai jalur biosintesis lignin dan gen-gen kunci yang berperan di dalamnya penting untuk dilakukan karena merupakan informasi yang penting untuk aplikasi selanjutnya, seperti manipulasi kandungan dan komposisi lignin dalam suatu tanaman dengan memanfaatkan bioteknologi. Fenilalanin adalah precursor lignin. Dalam tahap biosintesis lignin (Gambar 1), fenilalanin akan diubah menjadi cinnamic acid oleh enzim *Phenylalanine Ammonia Lyase* (PAL) yang terkespresi dari gen dengan nama yang sama. Selain enzim *PAL*, enzim-enzim kunci yang berperan dalam biosintesis lignin adalah *C4H* (*Cinnamate 4-Hydroxylase*), *C3H* (*p-Coumarate 3-hydroxylase*), *4CL* (*4-Hydroxycinnamate CoA Ligase*), *HCT* (*Hydroxycinnamoyl CoA Shikimate/Quinate Transferase*), *CCoAOMT* (*Caffeoyl CoA O-methyltransferase*), *CCR* (*Cinnamoyl CoA Reductase*), *CAD* (*Cinnamyl Alcohol Dehydrogenase*), *F5H* (*Ferulate 5-hydroxylase*), *COMT* (*Caffeic Acid O-methyltransferase*) (Umezawa 2018).

Penelitian ini bertujuan untuk mendapatkan informasi tentang keanekaragaman dan kekerabatan genetik protein-protein hasil ekspresi gen biosintesis lignin antar-tanaman monokotil dan mengidentifikasi protein yang terlibat dalam biosintesis lignin pada sorgum sehingga dapat digunakan untuk memodifikasi kandungan lignin.

## BAHAN DAN METODE

### Bahan Penelitian

Penelitian ini merupakan studi bioinformatika yang dilakukan dengan menggunakan sekuen-sekuen gen dan protein yang terlibat dalam proses

biosintesis lignin pada tanaman sorgum (*PAL*, *C4H*, *4CL*, *CCR*, *HCT*, *C3H*, *CCoAOMT*, *F5H*, *COMT* dan *CAD*). Sekuen-sekuen yang digunakan dalam studi ini diambil dari Sorghum Functional Genomics Database (SFGD, <http://structuralbiology.cau.edu.cn/sorghum/index.html>; Tian *et al.*, 2016).

### Identifikasi Protein Biosintesis Lignin pada Tanaman Sorgum dan Homolognya pada Tanaman Monokotil

Untuk mengetahui nama aksesori protein yang dipilih, dilakukan Blastn (*Basic Local Alignment Search Tool - nucleotide*). Hal ini dikarenakan, data yang diperoleh dari SFGD berupa *Sorghum ID* (Tabel 1). Sekuen-sekuen protein tersebut dijadikan sebagai *query* untuk mengidentifikasi protein-protein homolognya pada tanaman-tanaman monokotil lainnya menggunakan BLASTp di laman GenBank. Hasilnya diunduh dan kemudian dimasukkan dalam BLASTp (*Basic Local Alignment Search Tool*) di laman GenBank. Dari hasil yang ditampilkan dipilih beberapa protein identik yang berasal dari spesies tanaman yang lain dan dan selanjutnya diunduh dalam format fasta (.fasta).

### Analisis Kekerabatan

Unduhan hasil dari BLASTp disejajarkan secara clustal dengan MEGA-X (*Molecular Evolutionary Genetics Analysis*) (Kumar *et al.*, 2018) dan disimpan dalam format mega (.meg). Kemudian diagram kekerabatan dikonstruksi dengan dasar metode Maximum Likelihood berdasarkan matriks metode JTT dalam MEGA-X dengan bootstrap 500. Untuk mengetahui estimasi jarak genetik antar-protein antar-spesies tanaman monokotil, dilakukan analisis jarak kekerabatan dengan metode pairwise distance dengan bootstrap 1000 kali dan model Poisson.

### Analisis Fungsi Protein XP\_002445083.1 sebagai Kandidat Protein Caffeic Acid O-methyltransferase tanaman sorgum (SbCOMT)

Protein XP\_002445083.1 yang terpilih kemudian diuji, dengan cara membandingkan domain penyusun protein dan juga motif terkonservasinya dengan keluarga protein COMT dari tanaman monokotil lainnya yang telah diketahui peranannya dalam biosintesis lignin. Protein COMT dari Arabidopsis (NP001319057.1) digunakan sebagai perwakilan dari jenis tanaman dikotil. Identifikasi domain-domain penyusun protein masing-masing protein menggunakan *Conserved Domain Database* di NCBI. Untuk mengetahui area sekuen asam amino yang terkonservasi dilakukan pengecekan dengan secara online pada laman InterPro (<https://www.ebi.ac.uk/>

interpro/) dengan memasukkan urutan asam amino dari protein yang diinginkan. Sedangkan untuk mengetahui motif terkonservasi pada masing-masing protein digunakan program MEME (<http://meme-suite.org/meme/>, Bailey dan Elkan, 1994).

#### HASIL Identifikasi Protein Biosintesis Lignin pada Tanaman Sorgum dan Homolognya pada Tanaman Monokotil

Penelusuran gen-gen dan protein-protein penyandi biosintesis lignin pada laman SFGD telah dilakukan dan didapatkan *sorghum ID* gen biosintesis lignin dan sekuen protein-protein yang diinginkan (Tabel 1). Berdasarkan *sorghum ID* tersebut dilakukan perunutn melalui BLAST dan didapatkan nomor-nomor aksesi protein biosintesis lignin tanaman sorgum dari database GenBank (Tabel 1). Sekuen asam amino dari protein biosintesis lignin pada sorgum digunakan sebagai sekuen dasar (*query*) untuk pencarian asam amino

keluarga protein yang sama dari tanaman monokotil lainnya melalui metode BLASTP. Hasil BLASTP untuk setiap protein adalah berupa 100 protein dengan sekuen asam amino yang mempunyai identitas paling mirip. Dari hasil tersebut dipilih protein lignin pada tanaman monokotil dengan skor tertinggi untuk dikoleksi sekuen asam aminonya. Selain sekuen asam amino, didapatkan pula jumlah *hits*, sekuen protein yang mempunyai kemiripan dan berhasil disejajarkan dengan sekuen referensi (Tabel 2), dari spesies-spesies tanaman monokotil lainnya, seperti *Mischantus*, padi, jagung, millet, dll.

Jumlah hit tertinggi didapatkan dari protein COMT tanaman jagung, yaitu sebanyak 78 dan protein CCoAOMT sebanyak 63. Pada sorgum CCR dan COMT adalah dua protein yang mempunyai jumlah *hit* tertinggi, masing-masing 17 dan 16 hit. Sedangkan jumlah hit pada tanaman *Mischantus giganteus* dan padi Japonica (*Oryza sativa Japonica*) termasuk yang terendah. Bahkan pada padi Japonica terdapat 4 protein yang tidak diketahui jumlah hit-nya.

**Tabel 1.** Nama *Sorghum ID* berdasarkan database SFGD dan aksesi GenBank dari setiap protein penyandi lignin pada tanaman sorgum (*Sorghum IDs based on SFGD and GenBank accession codes of each lignin encoding proteins from sorghum*)

Protein	Sorghum ID	Nama Aksesi GenBank (Genbank accession codes)
PAL	Sobic.006G148800	XP_002448193.1
C4H	Sobic.002G126600	XP_002458683.1
4CL	Sobic.004G062500	XP_002451647.1
CCR	Sobic. <u>003G203500</u>	XP_021311799.1
HCT	Sobic.004G212300	XP_002452435.1
C3H	Sobic.009G181800	XP_002440001.1
CCoAOMT	Sobic.010G052200	XP_002436550.1
F5H	Sobic.001G196300	XP_002464383.1
COMT	Sobic.007G047300	XP_002445083.1
CAD	Sobic.002G195400	XP_002462348.1

**Tabel 2.** Jumlah hit masing-masing protein biosintesis lignin dari masing-masing spesies monokotil pada database GenBank. (*Hit counts of each lignin biosynthesis protein from each monocots species in GenBank database*)

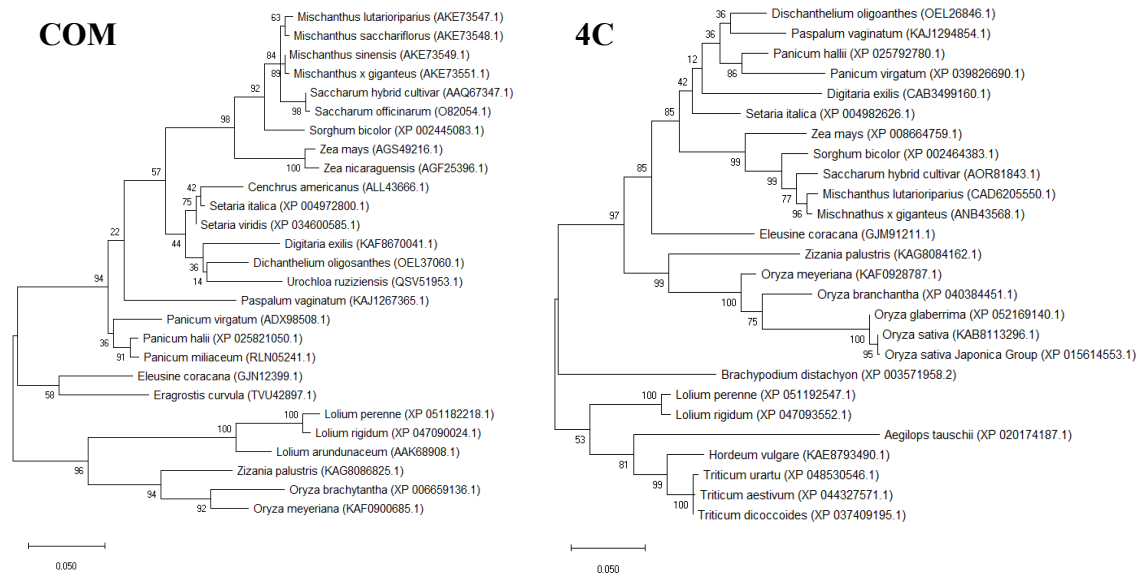
Spesies	PAL	C4H	4CL	CCR	HCT	C3H	CcoA-OMT	F5H	CO-MT	CAD
<i>Sorghum bicolor</i>	3	3	6	17	8	8	8	3	16	8
<i>Mischantus x giganteus</i>	2	1	1	1	1	2	1	1	1	td
<i>Mischantus lutarioriparius</i>	4	2	4	5	4	4	2	2	3	1
<i>Paspalum vaginatum</i>	2	1	2	3	2	3	1	1	1	2
<i>Panicum miliaceum</i>	3	1	2	td	3	2	2	td	2	4
<i>Panicum hallii</i>	4	2	4	6	4	4	2	2	2	4
<i>Panicum virgatum</i>	8	3	10	8	2	12	7	5	9	6
<i>Digitaria exilis</i>	5	2	5	1	5	8	4	5	3	10
<i>Setaria viridis</i>	12	2	4	2	4	2	2	2	2	6
<i>Setaria italica</i>	10	2	4	2	4	2	Td	2	2	6
<i>Oryza sativa</i>	3	td	2	2	4	1	1	1	td	2
<i>Oryza sativa Japonica group</i>	12	td	14	11	13	6	5	5	td	10
<i>Oryza sativa Indica group</i>	1	td	td	2	2	2	1	td	td	2
<i>Zea mays</i>	2	2	12	12	19	16	63	9	78	9

(PAL: Phenylalanine Ammonia Lyase, C4H: Cinnamate 4-hydroxylase, 4CL: 4-hydroxycinnamate CoA Ligase, CCR: Cinnamoyl Coa Reductase, HCT: Hydroxycinnamoyl CoA : Shikimate/Quinate Hydroxycinnamoyl Transferase, C3H: p-coumaroyl Shikimate/Quinate 3-hydroxylase, CcoAMOMT: Caffeoyl CoA O-methyltransferase, F5H: Ferulate 5-hydroxylase, COMT: Caffeic Acid O-methyltransferase, CAD: Cinnamyl Alcohol Dehydrogenase, td: tidak diketahui)

**Analisis Kekerbatan**

Analisis kekerabatan dilakukan dengan pembuatan filogram (*phylogram, phylogeny dendrogram*). Pada semua filogram protein biosintesis lignin, terlihat pengelompokan yang selaras dengan kelompok taksonomi monokotil. Protein sorgum saat berkelompok dengan protein monokotil lainnya menunjukkan nilai bootstrap yang cukup kuat (lebih dari 80) (Kress, 2002). Bootstrap tertinggi (97) adalah protein COMT sorgum yang berkelompok bersama dengan *Mischanthus*, tebu dan jagung (kelompok kedua). Pada filogram protein 4CL, sorgum juga masih mengelompok bersama dengan *Mischanthus*, tebu dan jagung (kelompok pertama).

Analisis jarak genetik dilakukan untuk setiap protein lignin pada beberapa tanaman monokotil. Hasil analisis tersebut disarikan pada tabel 3. Tanaman *Mischanthus* dan jagung dijadikan sebagai perwakilan tanaman yang berkerabat dekat dengan sorgum, sedangkan tanaman *D. exilis* sebagai perwakilan tanaman yang berkerabat jauh dari sorgum. Hasil analisis jarak genetik dari masing-masing protein antar tanaman monokotil tersebut menunjukkan Filogram Protein COMT dan 4CL (*COMT and 4CL Protein Phylogenetic Trees*).



**Gambar 2.** Filogram kekerabatan genetik dari protein COMT (kiri) dan 4CL (kanan) pada beberapa tanaman monokotil. *Keterangan: Filogram COMT terdiri dari 27 spesies, filogram 4CL terdiri dari 26 spesies. Nilai pada cabang menunjukkan hasil analisis bootstrap 1000 kali (COMT (left) and 4CL (right) protein phylogenetic trees of monocots plants. Exp: COMT phylogram consists of 27 species, and 4CL phylogram consists of 26 species)*

**Tabel 3.** Jarak genetik protein biosintesis lignin antara sorgum dan *Mischanthus x giganteus*, *Mischanthus lutarioriparius*, jagung dan *Digitaria exilis* (*Genetic distances of lignin biosynthesis proteins between sorghum and Mischanthus x giganteus, Mischanthus lutarioriparius, maize or Digitaria exilis*)

Protein	Jarak genetik ( <i>genetic distances</i> )				rerata ( <i>means</i> )
	Sb - Mg	Sb - Ml	Sb - Zm	Sb - De	
PAL	0,00714	0,01001	0,02448	0,03717	0,12
C4H	0,03424	0,02811	0,05077	0,05285	0,09
4CL	0,03677	0,03864	0,07480	0,08750	0,13
C3H	0,02385	0,01385	0,05006	0,03187	0,12
HCT	0,23938	0,01810	0,03189	0,08947	0,13
CCR	0,40392	0,39315	0,46563	0,49936	0,42
CCoAOMT	0,00389	0,01156	0,02772	0,05513	0,12
F5H	0,03693	0,04090	0,15673	0,13244	0,15
COMT	0,04255	0,03967	0,09309	0,11814	0,14
CAD	td	0,04256	0,08423	td	0,21

Keterangan : Sb: sorgum, Mg: *Mischanthus giganteus*, Ml: *Mischanthus lutarioriparius*, Zm: jagung, De: *Digitaria exilis*.

tanaman sorgum cenderung berdekatan secara genetik dengan tanaman *Mischanthus* dan jagung dibandingkan dengan tanaman yang lain karena angka jarak genetiknya lebih rendah dari pada angka rerata jarak genetik. Jarak genetik terdekat antara protein yang dianalisis adalah pada protein PAL sorgum dengan protein PAL *M.giganteus*. Sedangkan jarak ketiga tanaman itu paling jauh ada pada protein CCR untuk sorgum – *D. exilis* dan sorgum – jagung.

**Analisis Fungsi XP\_002445083.1 sebagai Kandidat Caffeic Acid O-methyltransferase tanaman sorgum (SbCOMT)**

Hasil analisis protein terduga COMT pada sorgum (XP\_002445083.1) pada *Conserved Domain Database* dari NCBI memperlihatkan bahwa protein ini mempunyai komposisi domain yang mirip dengan NP001319057.1 (*AtCOMT*). Protein XP\_002445083.1 mempunyai domain dimerisasi dan Methyltransferase 2 yang tergabung dalam super famili protein S-adenosyl-L-methionine (SAM atau AdoMet). Sedangkan NP001319057.1 terdiri dari domain dimerisasi dan

AdoMet\_Mtases yang juga tergabung dalam superfamily SAM. Selain itu, kedua protein

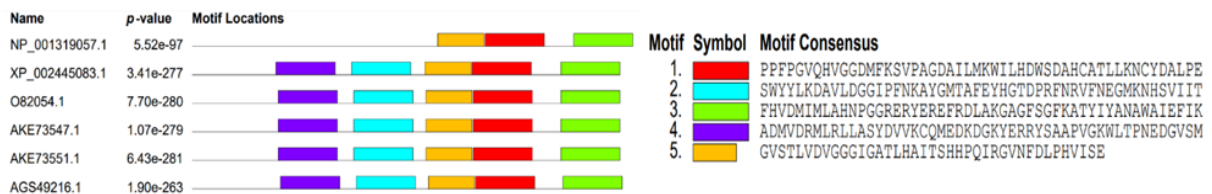
tersebut juga diperkirakan mempunyai fungsi katalitik yang sama.

NP_001319057.1	TDERFSKLFNQTG--FTIAVVKKALEVY	GFK-GVNVLDVGGGGV	TLGVVTSKYPNIK	235
AGS49216.1	TDPRFNRFNFMKNSVITKLLDFY	GFE-GVSTLVDVGGGGV	TLHAITSRHPHIS	227
XP_002445083.1	TDPRFNRFNFMKNSVITKLLDFY	GFE-GVSTLVDVGGGGV	TLHAITSRHPHIS	224
O82054.1	TDPRFNRFNFMKNSVITKLLDFY	GFE-GVSTLVDVGGGGV	TLHAITSRHPHIS	225
AKE73547.1	TDPRFNRFNFMKNSVITKLLDFY	GFE-GVSTLVDVGGGGV	TLHAITSRHPHIS	225
AKE73551.1	TDPRFNRFNFMKNSVITKLLDFY	GFE-GVSTLVDVGGGGV	TLHAITSRHPHIS	225

**Gambar 3.** Area terkonservasi famili protein COMT (Gfdesv.StLVDVGGGGIG, kotak merah) dari Arabidopsis (NP001319057.1), jagung (AGS49216.1), sorgum (XP\_002445083.1), tebu (O82054.1), *M. lutarioriparius* (AKE73547.1), *M.x giganteus* (AKE73551.1). (Conserved area of COMT family protein (Gfdesv.StLVDVGGGGIG, red box) from Arabidopsis (NP001319057.1), jagung (AGS49216.1), sorgum (XP\_002445083.1), tebu (O82054.1), *M. lutarioriparius* (AKE73547.1), *M.x giganteus* (AKE73551.1)).

Sebagai bahan pertimbangan lainnya, dilakukan pula penyejajaran area terkonservasi antara protein terduga COMT dari sorgum dan COMT-like lainnya yang telah diketahui, seperti COMT dari Arabidopsis (NP001319057.1), jagung (AGS49216.1), tebu (O82054.1), *Mischanthus lutarioriparius* (AKE73547.1) dan *M.x giganteus* (AKE73551.1). Area terkonservasi pada famili protein COMT dari 6 tanaman tersebut adalah Gfdesv.StLVDVGGGGIG. Huruf besar menunjukkan bahwa pada semua protein akan menunjukkan asam amino yang sama, Sedangkan huruf kecil memperlihatkan adanya variasi asam amino pada penyejajarannya (Gambar 3). Semakin banyak perbedaan asam amino pada area terkonservasi memperlihatkan jarak genetik yang semakin jauh.

Lima besar motif asam amino terkonservasi pada famili protein COMT telah didapatkan (Gambar 4). Pada famili protein COMT ini, sekuens asam amino motif 1 (PPFPGVQHVGDDMFKSVPAGDAILMKWILHDWSDAHCATLLKNCYDALPE), motif 3 (FHVDMIMLAHNPGGRERYERFRDLAKGAGFSGFKATYIYANAWAIEFIK) dan motif 5 (GVSTLVDVGGGGIGATLHAITSRHPHIS) terkonservasi pada semua tanaman (Arabidopsis, sorgum, tebu dan *Mischanthus*) dengan urutan motif yang sama. Sedangkan sekuens asam amino dengan motif 2 (SWYYLKDAVLDDGGIPFNKAYGMTAFEYHGTDPRFNRFNFMKNSVIT) dan motif 4 (ADMVDRMLRLLASYDVVKCQMEDKDGKYERRYSAAPVGKWLTPNEDGVSM) hanya terdapat pada tanaman monokotil, namun tidak ditemukan pada Arabidopsis.



**Gambar 4.** Motif asam amino terkonservasi penyusun protein COMT dan lokasi masing-masing motif pada sekuens protein COMT Arabidopsis (NP001319057.1), sorgum (XP\_002445083.1), tebu (O82054.1), jagung (AGS49216.1), *Mischanthus lutarioriparius* (AKE73547.1), *M. x giganteus* (AKE73551.1). (Conserved motives of COMT proteins and their location order derives from COMT Arabidopsis (NP001319057.1), sorgum (XP\_002445083.1), tebu (O82054.1), jagung (AGS49216.1), *Mischanthus lutarioriparius* (AKE73547.1), *M. x giganteus* (AKE73551.1)).

**PEMBAHASAN**

Pada penelitian ini digunakan sekuens protein sorgum sebagai query (dasar pencarian) karena penelitian ini berfokus pada tanaman sorgum sebagai tanaman pertanian berbiomassa tinggi yang berpotensi sebagai bahan produksi lignin. Informasi genomik mengenai keluarga gen dan protein di database sorgum (SFGD) dapat dijadikan sebagai alternatif untuk mendapatkan informasi mengenai sekuens-sukuens gen dan protein selain di NCBI.

Hasil Blast-P masing-masing protein lignin sorgum memperlihatkan bahwa 100 ranking teratas protein yang mempunyai identitas yang mirip adalah berasal dari kelompok monokotil, seperti *Mischanthus*, padi, jagung, millet, dll. Hal ini dikarenakan protein query yang digunakan berasal dari sorgum, yang juga adalah tanaman monokotil. Dari hasil tersebut diketahui pula hit dari suatu protein pada tanaman lainnya. Hit adalah jumlah protein dengan identitas yang mirip dengan protein

*query*. Jumlah hits yang ada tidak mencerminkan jumlah keseluruhan variasi protein, karena hasil BLAST-p hanya menunjukkan 100 besar ranking protein dengan identitas yang mirip.

Pada penelitian ini, diketahui bahwa protein yang mempunyai jumlah hit tertinggi adalah CCR dan COMT, masing-masing 17 dan 16 hit. Hal ini selaras dengan penjelasan Xu *et al.* (2009) yakni bahwa pada umumnya gen CCR dan COMT mempunyai anggota keluarga gen terbanyak pada tanaman, baik monokotil maupun dikotil. Jumlah hit tertinggi didapatkan dari protein COMT tanaman jagung, yaitu sebanyak 78. Jumlah tersebut lebih banyak jika dibandingkan dengan hasil penelitian Chen *et al.* (2014) yang menemukan paling tidak ada 55 alel gen COMT yang eksotis pada tanaman jagung. Jumlah total hits terbanyak dari suatu tanaman monokotil pada keseluruhan protein lignin ada pada tanaman jagung dengan total 269 hits. Meskipun demikian jagung bukanlah tanaman yang paling dekat dengan sorgum, melainkan *Mischantus*. Koski dan Golding (2001) menjelaskan bahwa jumlah hits terbanyak bukan berarti berdekatan secara filogenetik. Masih perlu dilakukan telaah lain, seperti tingkat kemiripan maupun jumlah variasi gen dalam suatu tanaman. Pada beberapa tanaman, jumlah hits dengan protein tertentu tidak diketahui (td). Hal ini dikarenakan identitas keluarga protein dari tanaman tersebut mempunyai kemiripan yang rendah dengan protein *query*nya sehingga tidak masuk ke dalam list protein yang mempunyai kemiripan dengan protein *query*.

Pada filogram yang berhasil disusun protein lignin dari tanaman sorgum hampir selalu membentuk kelompok bersama dengan *Mischanthus*, tebu dan terkadang juga bersama dengan jagung. Tanaman millet (*Setaria*, *Digitaria*) dan rumput-rumputan (*Panicum*, *Paspalum*) membentuk kelompok di luar kelompok sorgum, *Mischanthus*, dan jagung. Hasil analisis jarak genetik dari masing-masing protein antar tanaman monokotil menunjukkan hal yang selaras dengan analisis kekerabatan. Tanaman sorgum cenderung berdekatan secara genetik dengan tanaman *Mischanthus* dan jagung dibandingkan dengan tanaman *Digitaria exilis*. Tanaman *D. exilis* adalah tanaman monokotil yang tidak bergabung pada sub-suku *Saccharineae*, sehingga jarak genetik proteinnya dengan sorgum lebih jauh dari pada *Mischanthus*. Sedangkan sorgum, tebu, *Mischanthus* dan jagung termasuk dalam suku *Andropogoneae*, sedangkan millet-miletan tidak termasuk dalam suku *Andropogoneae*. Namun mereka tergabung dalam super suku *Andropogonodae*. Hal tersebut sesuai dengan Xu *et al.* (2009) yang menyatakan bahwa divergensi keluarga gen biosintesis lignin

berkorelasi dengan keragaman tanaman yang membawa gen. Selain itu karena perluasan keluarga gen sebagian besar terjadi setelah divergensi takson monokotil dan dikotil. Sehingga protein yang merupakan hasil ekspresi dari gen, juga mengikuti sifat divergensi tersebut.

Pada analisis fungsional, protein COMT dari sorgum dipilih sebagai protein yang dianalisis. Protein sorgum dengan nilai bootstrap tertinggi saat berkelompok dengan kelompok protein dari tanaman yang lain ada pada protein COMT, yakni sebesar 97. Pada filogram protein COMT, sorgum membentuk kelompok dengan tanaman monokotil lainnya dengan bootstrap 99. Hasil penelusuran fungsi protein COMT-like sorgum (XP\_002445083.1) pada *Conserved Domain Database* NCBI menunjukkan adanya domain dimerisasi dan Methyltransferase 2 yang tergabung dalam super famili protein S-adenosyl-L-methionine (SAM atau AdoMet\_MTases). AdoMet-MTases adalah enzim yang menggunakan S-adenosyl-L-methionine (SAM atau AdoMet) sebagai substrat untuk methyltransfer, menciptakan produk S-adenosyl-L-homocysteine (AdoHcy) (Struck *et al.*, 2012). Saat dibandingkan dengan fungsi dari protein COMT *Arabidopsis* (NP001319057.1) yang telah diketahui secara pasti, ternyata kedua protein ini mempunyai domain fungsional yang sama. Suatu protein dianggap mempunyai fungsi biologis yang sama apabila protein tersebut mempunyai komposisi domain yang sama dengan suatu famili protein (Wang *et al.*, 2021). Oleh karena kemiripan ini, maka kedua protein ini mempunyai klasifikasi yang sama yaitu *class I SAM-dependent methyltransferase*.

Area terkonservasi pada protein COMT-like sorgum (XP\_002445083.1) mempunyai kesesuaian saat disejajarkan dengan protein COMT dari *Arabidopsis* (NP001319057.1), jagung (AGS49216.1), tebu (O82054.1), *M. lutarioriparius* (AKE73547.1), *M.x giganteus* (AKE73551.1). Hal ini dapat diartikan, protein COMT-like sorgum berasal dari famili gen yang sama dengan protein COMT *Arabidopsis* (NP001319057.1), jagung (AGS49216.1), sorgum (XP\_002445083.1), tebu (O82054.1), *M. lutarioriparius* (AKE73547.1), *M.x giganteus* (AKE73551.1). Dengan kesesuaian tersebut, protein XP\_002445083.1 dapat dikelompokkan sebagai keluarga protein COMT, khususnya pada tanaman sorgum.

Validasi analisis fungsional lainnya adalah melalui analisis motif. Penelusuran motif terkonservasi pada keluarga protein COMT, didapatkan lima motif terkonservasi. Pada semua protein COMT yang dianalisis terdapat tiga motif terkonservasi yang sama, secara berurutan yaitu motif 5, 1 dan 3. Sedangkan motif 2 dan 4 tidak ditemui pada protein COMT *Arabidopsis*



(NP001319057.1), namun dapat ditemui pada protein COMT pada tanaman tebu (O82054.1), jagung (AGS49216.1), *Mischanthus lutarioriparius* (AKE73547.1) dan *M. x. giganteus* (AKE73551.1). Dengan demikian, ada kemungkinan bahwa motif asam amino 2 dan 4 adalah motif yang spesifik untuk famili protein COMT yang berasal dari tanaman monokotil. Kesesuaian motif terkonservasi pada protein terduga COMT sorgum (XP\_002445083.1) dengan protein COMT tanaman yang lain menandakan, bahwa protein XP\_002445083.1 adalah keluarga protein COMT.

## KESIMPULAN

Protein biosintesis lignin pada tanaman monokotil mempunyai pola kekerabatan genetik yang selaras dengan divergensi takson dalam monokotil. Kelompok protein biosintesis lignin (PAL, C4H, 4CL, C3H, CCoAOMT, F5H, COMT dan CAD) dari tanaman sorgum mempunyai kekerabatan yang erat dengan kelompok protein biosintesis dari genus *Mischanthus* dan tanaman jagung bila dibandingkan dengan monokotil lainnya dikarenakan kedua tanaman tersebut masih berkerabat dekat dengan sorgum.

Protein XP\_002445083.1 dari sorgum dapat dikelompokkan sebagai keluarga protein COMT karena mempunyai fungsi biologis, area terkonservasi dan motif terkonservasi yang sama dengan keluarga protein COMT dari tanaman lainnya, baik itu *Arabidopsis* maupun *Mischanthus*, jagung dan tebu. Protein COMT dari sorgum mempunyai potensi untuk digunakan pada penelitian gen kunci yang berperan dalam biosintesis lignin berdasarkan kekerabatan dan kesetaraan fungsi dengan famili protein COMT dari tanaman lainnya.

## DAFTAR PUSTAKA

- Bajpai, P., 2017. *Carbon fibre from Lignin*. Springer Nature. Singapore. pp. 77
- Bailey, T.L and C. Elkan., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. California, USA. pp. 28–36.
- Baskar, C., Baskar, S and R.S., Dhillon., 2012. Biomass conversion: The interface of biotechnology chemistry and material science. Springer International Publisher. Berlin. pp. 199
- Bonnawitz, N.D and C., Chapple., 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet*. 44, pp. 337–363.
- Chen, Y., Blanco, M., Ji, Q., Frei, U.K and T., Lübberstedt., 2014. Extensive genetic diversity and low linkage disequilibrium within the COMT locus in maize exotic populations. *Plant Sci J*. 221-222, pp. 69–80.
- Grossman, A and W., Vermerris., 2019. Lignin-based polymers and nanomaterials. *Curr Opin Biotechnol*. 56, pp 112–120.
- Hattori, T and S., Morita., 2010. Energy crops for sustainable bioethanol production; which, where and how? *Plant Prod Sci*. 13(3), pp. 221–234.
- Irmer, J., 2017. Lignin-a natural resource with huge potential. BIOPRO Baden-Württemberg GmbH. Bioecomy BW. (diakses 29 Agustus 2022). <https://www.bioekonomie-bw.de/en/articles/dossiers/lignin-a-natural-resource-with-huge-potential>.
- Koski, L.B and Golding, G.B., 2001. The closest BLAST hit is often not the nearest neighbor. *J Mol Evol*. 52, pp 540–542.
- Kumar, S., Stecher, G., Li, M., Knyaz, C and K., Tamura., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evo*. 35(6), p.1547.
- Mullet, J., Morishige, D., McCormick, R., Truong, S., Hilley, J., McKinley, B., Anderson, R., Olson, S.N and W., Rooney., 2014. Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *J Exp Bot*. 65(13), pp. 3479–3489.
- Olson, S.N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B.A., Zhang, Y., Hall, S., Packer, D and J., Mullet., 2012. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuel Bioprod Biorefin*. 6(6), pp. 640–655.
- Struck, A.W., Thompson, M.L., Wong, L.S and Micklefield, J., 2012. S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chem Bio Chem*. 13(18), pp 2642–2655.
- Tian, T., You, Q., Zhang, L., Yi, X., Yan, H., Xu, W and Z., Su., 2016. SorghumFDB: sorghum functional genomics database with multidimensional network analysis. *Database*. 2016, baw099.
- Umezawa, T., 2018. Lignin modification in planta for valorization. *Phytochem rev*. 17(6), pp. 1305–1327.
- Vanholme, R., De Meester, B., Ralph, J and W., Boerjan., 2019. Lignin biosynthesis and its integration into metabolism. *Curr Opin Biotechnol*. 56, pp. 230–239.
- Wang, Y., Zhang, H., Zhong, H and Xue, Z., 2021. Protein domain identification methods and online resources. *Comput Struct Biotechnol*

- J. 19*, pp. 1145–1153.
- [WBA]. World Bioenergy Association., 2019. Global bioenergy statistic 2019. [diakses 10 Agustus 2021]. <https://www.greengrowthknowledge.org/research/global-bioenergy-statistics-2019>.
- Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K.L and W, Shi., 2009. Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC bioinformatics. 10*(11), pp 1–15.