IN SILICO ANALYSES OF GENE FAMILIES INVOLVED IN XYLOGLUCAN SYNTHESIS IN BARLEY (HORDEUM VULGARE L.)

by
Vedat Cenk Güler

Submitted to Graduate School of Natural and Applied Sciences
in Partial Fulfillment of the Requirements
for the Degree of Master of Science in
Biotechnology

Yeditepe University
2022

IN SILICO ANALYSES OF GENE FAMILIES INVOLVED IN XYLOGLUCAN SYNTHESIS IN BARLEY (HORDEUM VULGARE L.)

APPROVED BY:

Assist. Prof. Dr. Andrew John Harvey ……
(Thesis Supervisor)
(Yeditepe University)

Assist. Prof. Dr. Bahar Soğutmaz Özdemir ……
(Yeditepe University)

Assist. Prof. Dr. Merve Seven ……
(Bahçeşehir University)

DATE OF APPROVAL: …/…/2022

I hereby declare that this thesis is my own work and that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I have fully cited and referenced all material and results as required by these rules and conduct, and this thesis study does not contain any plagiarism. If any material used in the thesis requires copyright, the necessary permissions have been obtained. No material from this thesis has been used for the award of another degree.

I accept all kinds of legal liability that may arise in case contrary to these situations.

Name, Last name Vedat Cenk Güler

Signature

ACKNOWLEDGEMENTS

First of all, I would like to express my special gratitude to my supervisor, Professor Andrew John Harvey for his invaluable assistance, support, guidance and patience. I am so happy to involve in such an important project and work under his supervisory who always believed in me.

Secondly, I would like to say that I deeply thank to a person who were with me during this process. I would like to thank my senior in laboratory, PhD candidate Mehtap Aydın for her great assistance and support during my protectional steps. Moreover, I want to thank Professor Bahar Soğutmaz Özdemir for her assistance and moral support to me. At last, I would like to thank all my instructors and other fellow in this Master program for their advice and friendship.

Finally, I am sincerely grateful to my beloved mother and father, Nuray Güler and Vedat Güler, and my lovely friend Ece Nur Erol for their moral support during my thesis progress.

ABSTRACT

IN SILICO ANALYSES OF GENE FAMILIES INVOLVED IN XYLOGLUCAN SYNTHESIS IN BARLEY (HORDEUM VULGARE L.)

Barley (*Hordeum vulgare L.*) is a cereal that has always been one of the key plants to cultivate since ancient times. Today, it is the fifth most produced crop in the world. Because of its economic importance, barley has been investigated in many studies and scientific works. Xyloglucans (XyGs) is an important component of barley cell walls, and is involved in the process of cell growth. The backbone of XyGs is synthesized by cellulose synthase-like C (CSLC) enzymes, whilst xylosyl substitution, added by xyloglucan xylosyl transferases (XXT) is the most common side chain. Many different sidechains patterns have been observed on xyloglucan, which are added through the catalytic activities of enzymes such as xyloglucan galactosyl transferase (XGT), xyloglucan o-acetyl transferase (O-AcT), and fucosyl transferase (FUT) transferase. To date, they are not any publications that have systematically investigated all the genes involved in XyG synthesis, in any plant. This study aimed to analyze, at the *in silico* level, the gene families that are involved in XyG synthesis in barley. Since there are many different polysaccharides present in barley, the second goal was to determine the co-expression patterns of all these genes, in order to find the most likely candidates involved in XyG synthesis. For this study, five different gene groups, *CSLC*, *XXT*, *XGT*, *O-AcT*, and FUT were investigated. Protein sequence alignments and phylogenetic trees were used to determine the level of relatedness within each group. In addition, their expression levels were investigated using public RNA-seq data, with calculations of their FPKM values. Values were analyzed using Pearson Correlation to find co-expression levels between each of the genes. Since CSLC enzymes are crucial to xyloglucan structure, the analysis focused onto co-expression with these genes. *HvCSLC3* in particular showed the highest number of correlating genes, including *XXT*s, *XGT*s, *O-AcT*s, and somewhat surprisingly, several *FUT*s. Tissues from 4-6 weeks barley plants generally gave higher FPKM values for the studied genes, so the need for xyloglucan might be increase at that time interval in barley. This study should lead to further investigations into the biosynthesis of XyG in barley, not only at mRNA/protein levels, but also in polysaccharide/carbohydrate levels and fine structural analyses.

ÖZET

ARPADA (HORDEUM VULGARE L.) KSİLOGLUKAN SENTEZİNDE ROL ALAN GEN AİLELERİ ÜZERINE İN SİLİKO ÇALIŞMALAR

Arpa, antik zamanlardan beri her dönemde ekilip biçilen en önemli bitkilerden bir tahıldır. Günümüzde en çok üretilen ekinlerde beşinci sırada yer almaktadır. Ekonomik öneminden ötürü, arpa birçok çalışmada ve bilmsel araştırmada tetkik edilmiş ve araştırma konusu olmuştur. Ksiloglukan (XyG) arpa hücre duvarlarının önemli bir bileşeni olup, hücre büyümesi sürecinde de yer almaktadır. XyG omurgası selüloz sentaz-benzer C enzimleri tarafından sentezlenirken, ksiloglukan ksilosil transferaz (XXT) tarafından gerçekleştrilen ksilosil eklenmesi en çok görülen yan zincir yapısıdır. Ksiloglukan galaktosil transferaz (XGT), ksiloglukan o-asetil transferaz (O-AcT), ve fukosil transferaz (FUT) gibi enzimlerin katalitik aktiviteleriyle birçok farklı yan zincir örüntüleri ksiloglukan yapısında gözlenebilir. Bugüne kadar, hiçbir bitkide ksiloglukan sentezinde görev alan genleri sistematik olarak inceleyen bir çalışma literatürde bulunmamaktadır. Bu çalışma, arpada farklı XyG sentezine dahil olan farklı gen ailelerinin in siliko analizini amaçlamıştır. Arpada birçok farklı polisakkarit mevcut olmasından dolayı, ikincil hedef olarak XyG sentezinde görev alan en olası aday genleri belirlemek amacıyla bu genlerin ortak ifade paternlerinin belirlenmesi olmuştur. Bu çalışmada *CSLC*, *XXT*, *XGT*, *O-Act* ve *FUT* gen grupları üzerinde çalışılmıştır. Protein dizi karşılaştırmaları ve filogenetik ağaçlar her grup arasındaki yakınlık seviyesini belirlemede kullanılmıştır. Bunun yanı sıra, halka açık bir RNA sekans verisinden FPKM değerleri hesaplanarak ekspresyon seviyeleri araştırılmıştır. Değerler Pearson Korelasyonuna dayanarak incelenip, her bir gen için ortak ekspresyon seviyeleri bulunmuştur. Ksiloglukan yapısının CSLC enzimlerinin ksiloglukan yapısına ötürü, analiz bu genlerin ortak ekspresyonuna odaklanılmıştır. *HvCSLC3* geni ve birçok farklı *XXT*, *XGT*, *O-AcT*, ve şaşırtıcı bir biçimde, birkaç *FUT* ile görece en yüksek korelasyon değerlerini göstermiştir. Çalışılan genler arasında 4-6 haftalık arpa bitkilerinden alınan dokulardan en yüksek FPKM değerleri elde edilmiştir ve bu zaman aralığında ksiloglukana olan ihtiyaç arpada artmış olabilir. Bu çalışma, ileride arpadaki XyG biyosentezi üzerine yapılacak araştırmalara ve çalışmalara öncülük edecektir ve sadece mRNA/protein seviyesinde değil, ayrıca polisakkarit/karbonhidrat seviyesinde ve yapısal analizlerle de incelenmesinin önünü açacaktır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT v

ÖZET vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xii

LIST OF SYMBOLS/ABBREVIATIONS xiv

1. INTRODUCTION 1

1.1. HORDEUM VULGARE 1

1.2. PLANT CELL WALL 2

1.3. PLANT CELL WALL STRUCTURE 3

1.4. CELLULOSE 5

1.5. MIXED-LINKED B-GLUCANS 8

1.6. XYLAN 10

1.7. MANNAN 10

1.8. PECTIN 11

1.9. XYLOGLUCAN 12

1.10. THE OBJECTIVE OF THESIS 16

2. MATERIALS AND METHODS 17

2.1. IDENTIFICATION OF XYLOGLUCAN SYNTHESIS RELATED GENES IN HORDEUM VULGARE 17

2.2. PHYLOGENETIC ANALYSIS AND SEQUENCE ALIGNMENT 17

2.3. GENOMIC LOCATION DETERMINATION OF TARGETED GENES 18

2.4. EXPRESSION PATTERN ANALYSIS IN RNA-SEQ DATA 18

3. RESULTS 20

3.1. BIOINFORMATIC ANALYSES OF XYLOGLUCAN SYNTHESIS RELATED GENE FAMILIES IN HORDEUM VULGARE 20

3.1.1. Identification of Barley Gene Groups and their Locations on Genome 20

3.1.2. Multiple Sequence Alignments 27

3.1.3. Phylogenetic Analyses 38

3.1.4. Gene Expression and Pearson Correlation Analyses 44

4. DISCUSSION 58

5. CONCLUSION 69

APPENDIX A 76

LIST OF FIGURES

Figure 1.1. Production/Yield quantities of barley in Turkey, between 2010 and 2019……2

Figure 1.2. Structure of primary cell wall……...…….5

Figure 1.3. Structural model of a cellulose microfibril……....6

Figure 1.4. Rosette formation formed by CESA proteins. Each rosette unit contains 6 rosette subunit that is formed by 6 CESA proteins……..…7

Figure 1.5. Phylogenetic tree of cellulose synthase (CESA) and cellulose synthase-like (CSL) genes and families in plants……..8

Figure 1.6. Molecular comprehension of mixed linked (1;3,1;4)-β-D-glucan…….9

Figure 1.7. Molecular comprehensions of different mannans…….…..11

Figure 1.8. Chemical structures and nomenclatures of XyG sidechains. Elemental β-(1,4)-glucosyl backbones have been represented as black color, numbers are representing for linkage point for hydroxyl group of the saccharide. Arrows are representing the direction of glycosidic bonds……..……..14

Figure 1.9. Schematic structures of xyloglucan (XyG) with different oligosaccharide units. Blue circles are represented glucosyl units; linked together with β-(1,4) linkage. Xylosyl substitutes (orange stars) are joined with α-(1-6) linkage. Galactosyl (yellow circles) and fucosyl (red triangle) are introduced in xyloglucan structure via β and α-(1-2) linkages…...15

Figure 3.1. Chromosomal map of barley genome. CSLC, XXT, XGT, and O-AcT genes that determined for XyG biosynthesis were listed on figure. Length of chromosomes were adjusted and projected for real base pair length of those chromosomes……26

Figure 3.2. Multiple sequence alignment results of Cellulose synthase-like C proteins of *Hordeum vulgare, Arabidopsis thaliana, Oryza sativa,* and *Populus trichocarpa*……37

Figure 3.3. Phylogenetic tree of aligned Cellulose synthase-like C proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa* with bootstrap values……39

Figure 3.4. Phylogenetic tree of aligned xylosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon* with bootstrap values……40

Figure 3.5. Phylogenetic tree of aligned galactosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, with bootstrap values……41

Figure 3.6. Phylogenetic tree of aligned acetyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon* with bootstrap values……43

Figure 3.7. FPKM scores of genes that significantly correlated with HvCSLC1, tissue by tissue……53

Figure 3.8. Expression level (as percentages) of genes that significantly correlated with HvCSLC1, tissue by tissue……53

Figure 3.9. FPKM scores of genes that significantly correlated with HvCSLC2, tissue by tissue……54

Figure 3.10. Expression level (as percentages) of genes that significantly correlated with HvCSLC2, tissue by tissue……54

Figure 3.11. FPKM scores of genes that significantly correlated with HvCSLC3, tissue by tissue……55

Figure 3.12. Expression level (as percentages) of genes that significantly correlated with HvCSLC3, tissue by tissue……55

Figure 3.13. FPKM scores of genes that significantly correlated with HvCSLC4, tissue by tissue……56

Figure 3.14. Expression level (as percentages) of genes that significantly correlated with HvCSLC4, tissue by tissue……56

Figure 3.15. FPKM scores of genes that significantly correlated with HvCSLC5, tissue by tissue…….57

Figure 3.16. Expression level (as percentages) of genes that significantly correlated with HvCSLC5, tissue by tissue……57

Figure 4.1. A demonstrated visual of "XXGGG" composition, could be the most abundant xyloglucan oligosaccharide composition one among all tissue and time intervals in barley, with name of genes that possibly involve and co-express for its biosynthesis, at highest level and possibility……...68

Figure A.1. Multiple sequence alignment results of xyloglucan xylosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon*……81

Figure A.2. Multiple sequence alignment results of xyloglucan galactosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*……...88

Figure A.3. Multiple sequence alignment results of xyloglucan O-acetyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon*……95

LIST OF TABLES

Table 2.1. Information list of RNA-seq data of predicted genes from an assembly of whole-genome-shotgun sequence from barley. Each tissue has 3 replicates from 2 different batches……..18

Table 3.1. List of identified Cellulose synthase-like C (CSLC) genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes…….20

Table 3.2. List of xylosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes……21

Table 3.3. List of galactosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes…….21

Table 3.4. List of acetyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes……22

Table 3.5. List of fucosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes……24

Table 3.6. List of reference barley genes for expression values with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes…….42

Table 3.7. List of FPKM values of CSLC barley genes for expression values from NGS barley RNA-seq data…….43

Table 3.8. List of FPKM values of xylosyl transferase barley genes for expression values from NGS barley RNA-seq data……...43

Table 3.9. List of FPKM values of galactosyl transferase barley genes for expression values from NGS barley RNA-seq data……...44

Table 3.10. List of FPKM values of acetyl transferase barley genes for expression values from NGS barley RNA-seq data……...45

Table 3.11. List of FPKM values of fucosyl transferase barley genes for expression values from NGS barley RNA-seq data……...46

Table 3.12. List of FPKM values of reference barley genes for expression values from NGS barley RNA-seq data…….47

Table 3.13. List of correlated FPKM values of genes versus targeted barley genes for expression values from NGS barley RNA-seq data……...48

Table 3.14. List of correlation values that above 0.7 from barley CSLC expression values..50

Table 4.1. HvCSLC1 co-expression pattern in LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow……..……...62

Table 4.2. HvCSLC2 co-expression pattern in CAR5, LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow……...62

Table 4.3. HvCSLC3 co-expression pattern in EMB, ROO, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow…….63

Table 4.4. HvCSLC4 co-expression pattern in LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow……...64

Table 4.5. HvCSLC5 co-expression pattern in LEM, PAL, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow……...65

LIST OF SYMBOLS/ABBREVIATIONS

Ara*f*	Arabinofuranose
At	*Arabidopsis thaliana*
CESA	Cellulose synthase protein family
CSLC	Cellulose synthase-like family C
DNA	Deoxyribonucleic acid
FPKM	Fragments per kilobase of transcript per million mapped reads
FUT	Fucosyl transferase
Gal	Galactose
Glc	Glucose
Glcp	Glucopyranose
GT	Galactosyl transferase
Hv	Barley (*Hordeum vulgare*)
IRX	Xylosyl transferase
MLG	Mixed-linked glucan
MUR	Xyloglucan galactosyltransferase, from *Arabidopsis thaliana*
MXE	Endotransglucosylase
LDL	Low-density lipoprotein
Man	Mannose
MSA	Multiple sequence alignment
O-AcT	Xyloglucan O-acetyltransferase
RNA	Ribonucleic acid
RWA	Reduced wall acetylation protein family
TBL	Trichome birefringence-like protein family
UDP	Uridine diphosphate
UV	Ultraviolet Light
XGT	Xyloglucan galactosyltransferase
XXGT	Xyloglucan xylosyl-galactosyltransferase
XXT	Xyloglucan xylosyltransferase
XyG	Xyloglucan

1. INTRODUCTION

1.1. HORDEUM VULGARE

Barley (*Hordeum vulgare*) is the one of the key crops to place in the of the Old-World agriculture period. It is the fifth place in the most produced crop in the world, after maize, wheat, rice and soybeans. A study has suggested that about ten thousand years ago, barley was cultivated in the Fertile Crescent side of the Middle East [1]. In addition, besides from the Middle East, Tibet was also a major cultivation site for the plant. Barley was used before for primarily as a human food source. However, it has been developed and evolved as a feed grain for animals, and also is important as a malting and brewing product. Nowadays, only two percent of produced barley is used for human food, and almost sixty-five percent of them is used as a feed product [2]. Because of its high drought and heat resistance, it is the most adapted and cultivated cereal around the world than other crops.

Many researches have shown that barley has many potential health benefits for humans. For instance, barley contains a high level of β-glucans, so it can have effects such as a decrease in blood cholesterol level [3]. Moreover, it contains many tocols such as tocopherols and tocotrienols that have an effect on reducing LDL levels in the blood [4]. For these reasons, it is suggested that consuming barley could also reduce possible cardiovascular disorders. The highest barley production in the world belongs to the Russian Federation (about 20 million metric tons in 2019), followed by France, Germany, Canada, and Ukraine. According to the same source, Turkey is listed as ninth among barley producing countries. In Turkey, 7.6 million metric tons of barley were produced in 2019 [5], graphic version is located below as Figure 1.1.

Until now, there are thirty-one barley species/sub-species have been identified. A study has suggested that all of these species came from the same ancestral species which is called *Hordeum spontaneum* [6].

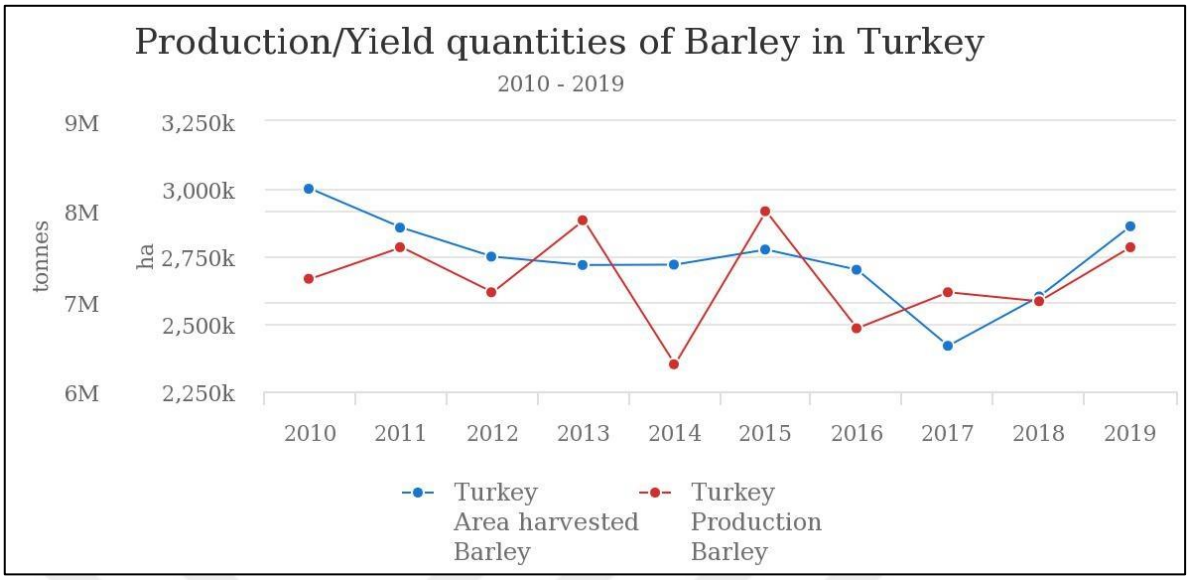


Figure 1.1. Production/Yield quantities of barley in Turkey, between 2010 and 2019 [5]

1.2. PLANT CELL WALL

Different ecosystems around the world contain a various number of plant species that are distinct from each other according to their structure, habitats that they belonged, and many adaptations. One of a very first common features of land plants is the cell wall. It has a complex and dynamic structure that takes a key role in processes such as growth, cell differentiation, resistance to the external threats, and cellular communication. The cell wall has also a selectively permeable structure that can lead to cell-cell interactions. In plants, there have more than 30 cell types that can be differentiated from each other based on their size and position in the body of the plant. However, the cell wall ensures and preserves the plant's unique structure and body type because of its strength [7]. For these reasons, cell wall structure can be categorized according to their position, cell type and growth phase periods [8].

The cell wall thickness is between 0.1-1 μm, depending on cell type that it surrounds. Because of the turgor pressure, plant cells resist pressure that coming from inside and this can be led to the upright structure of the plant. That's why plant cells wall can be enlarged and stretched about 100-fold in volume [9]. When the plant cell is growing, there is tensile wall stress. Cell walls can be extended because of that. On the other hand, the cell wall does not interfere in cellular growth. It just adapts itself to the needs of the plant cell.

Extracellular pathogen attack can be interrupted by the cell walls of the plant cells. They can also prevent any harmful interaction from pathogens Cell wall elements can activate the defense mechanism of cell with signal transduction processes and eliminate pathogen threats. The research showed that it is glycans that are responsible for signal transduction [10].

The structure of cell wall contains about 90 percent carbohydrate and rest of them are proteins such as glycoproteins. There are many polymers that involved in the cell wall structure such as cellulose, pectin and structural proteins. Because of this enriched plant cell region, they are the source of many biogeochemical cycles such as the carbon flow process. They can enrich soil structure, acting just like fertilizer. The plant cell wall contains many types of polysaccharides that they can be used as renewable biofuels [11]. Moreover, humans can consume these polysaccharides for their high-fiber diets. In the end, extracted polysaccharides from plant cell walls can be modified for making and producing paper, fibers, textile, and also timber products.

1.3. PLANT CELL WALL STRUCTURE

Even though a diversity can be observed on cell wall structures, but generally they consist of a cellulosic and non-cellulosic matrix of complex molecules. In addition, the cell wall contains structural polymers such as lignin and glycoproteins [12]. Broadly speaking, the cell walls consist of three main parts which are the primary cell wall, middle lamella, and secondary cell wall. The middle lamella is formed at the cytokinesis period that during cell division, and it separates the two cell walls. The primary cell wall has been formed especially during in cell growth and development periods. It is the utmost outer layer of the cell during cell expansion, it has to show resistance to and pressures and external forces. The secondary cell wall is a thicker structure than primary one, partly because of consisting much more lignin content. Due to the lignin component, it stabilizes cell's mechanical condition, structural shape and also helps pathogen resistance [13]. Moreover, secondary cell wall has a role in absorbing UV radiation, and preventing other competing plants' growth located nearby. After the cell growth has completely ceased and the cell has reached its maximum volume, the secondary wall will be formed. At the end, secondary cell wall contains much more phenolics such as lignin and ferulic acid, comparing with primary cell wall.

Both dicots and monocots' primary cell walls are comprised of cellulose, non-cellulosic saccharides, phenolics and structural proteins. However, the quantity of these components can vary in between them. For instance, the percentage of structural proteins in primary cell wall is about 10 percent in dicots, but only one percent in monocots [14]. Another example is that monocots' primary cell wall consists of around five percent xyloglucan, but dicots have much more xyloglucan in their primary cell wall, about 20-25 percent. On the other hand, monocots can contain β-glucans in their primary cell wall, unlike dicots. At the end, monocots comprise more cellulose in their cell wall (in between 20-30 percent), matched against the dicots (15-25 percent) [15].

After cellulose, there are two other polysaccharide groups that are located in the cell wall structure which are pectins and non-cellulosic β-linked polysaccharides. These complex polymers that fill the space between the cellulosic microfibrils are known as the matrix polysaccharides. The main difference in between pectins and non-cellulosic β-linked polysaccharides is that non-cellulosic β-linked polysaccharides show more rigid structure, but also flexible feature. That's why, they are acting and taking a role as backbone material in the cell wall. Depends on the sugar type (and also structure), naming and categorizing of matrix polysaccharides is arranged. Those sugar-based polysaccharides called as *glycans*. For instance, if the backbone is from xylose sugar, then it is called *xylan*, and if the backbone is comprised of glucose then it is known as a *glucan*. Substitutions onto the polysaccharide chain can also affect naming. When a glucan backbone is substituted with with xylose sugar, it will be called as *xyloglucan*. Detailed cell wall structure is visualized in Figure 1.2.

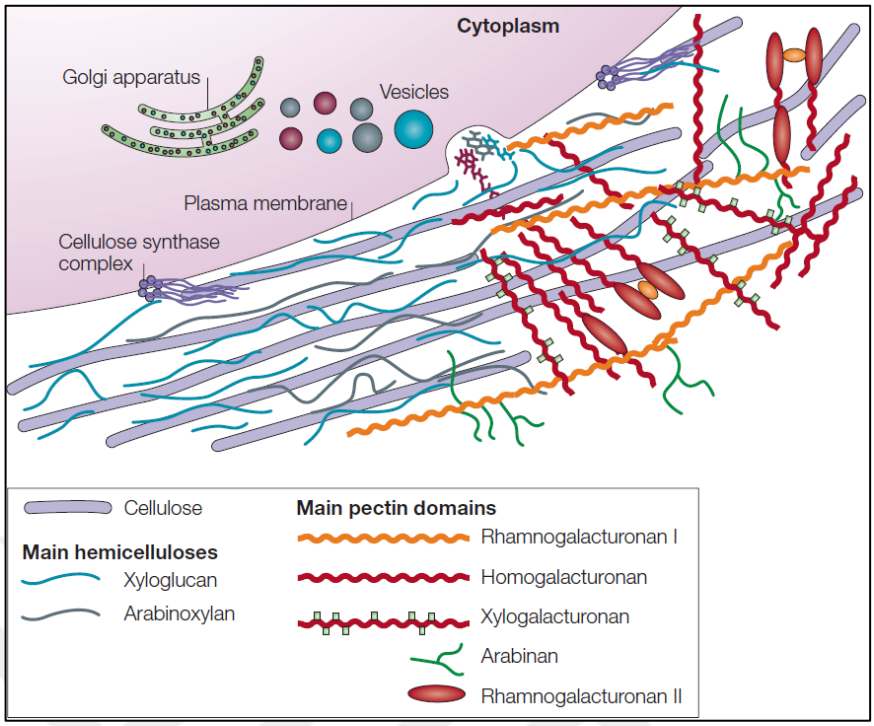


Figure 1.2. Structure of primary cell wall [7]

1.4. CELLULOSE

In plants, it is cellulose that can be observed as the most abundant polysaccharide [16]. Not only plants, it is also found some other living organisms such as bacteria, algae, and tunicates [17]. It is the fundamental molecule for plant cell walls. Its duty in the cell wall can be explained as acting like bridge for connecting other molecules and structures. Cellulose is made up with β-1,4-linked glucosyl chains that connects each other parallelly and linearly. This parallel structure of glucan residues leads to the enduring structure of cell wall. Extensive hydrogen bonding forms up between the glucose chains.

Cellulose is synthesized at the plasma membrane's cytoplasmic side in plant cells. This membrane- bound complex synthesizes and extracts formations known as the cellulose synthase complex. These protein-based cellulose synthase molecules create products that are made up with between 16 to 32 glucan chains [18], just like in Figure 1.3. In this way, microfibril formation is formed. The diameter of this microfibril can be changed due to its location on the cell wall. Generally, thinner microfibrils (5-10 nm) are located in primary cell wall, but thicker ones (30-50 nm) are consisted in secondary cell wall structure. Because

of van der Waals's forces on glucan formations, microfibrils will be aggregated and formed crystalline cellulose. In this way, cellulose cannot be soluble in water and microfibrils can provide much rigidity to the cell wall.

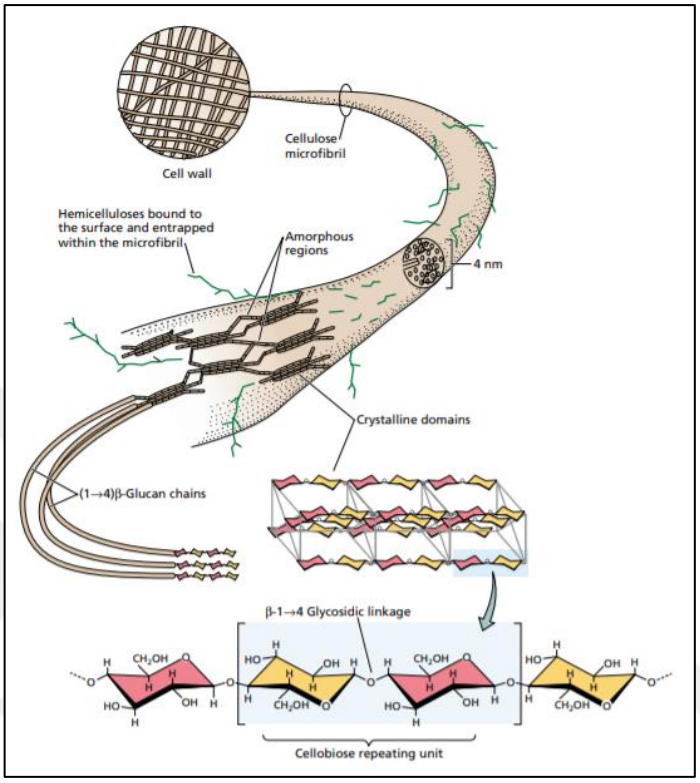


Figure 1.3. Structural model of a cellulose microfibril [19]

Cellulose synthase proteins (CESA) are responsible from formation of glucan chains at the plasma membrane. As a substrate material, UDP-glucose is used for forming β-1,4-glucan chains. Thanks to genetic studies, in different plant species, many numbers of *CESA* genes were discovered and identified. For instance, *Arabidopsis thaliana*, one of the model plants for genetic studies, has 10 different *CESA* genes. Not all of them are working on producing the cellulose synthesis complex. However, it is concluded that at least 3 different *CESA* genes must be work together to produce these complexes [20]. For example, generally, *CESA1*, *CESA3*, and *CESA6* are responsible for formation of cellulose structure in primary cell wall, whereas another 3 genes, *CESA4*, *CESA7*, and *CESA8* are involved synthesis process in secondary cell wall in *Arabidopsis thaliana* [21].

Synthesized CESA proteins are gathered in different protein complex is named rosette in plant cell’s plasma membrane. For rosette structure formation, homodimers are located to

form a hexagon with six particles. Each homodimer part consists three different CESA protein. At next, hexagon formation will be arranged in rosette formation, in other words, symmetry in six-fold. Folding and arranging are held in plant cell's Golgi apparatus and endoplasmic reticulum. When arranging and folding are completed, complex will be transferred to plasma membrane for microfibril synthesis [22]. Figure 1.4. is a visualized form of Rosette formation structure.

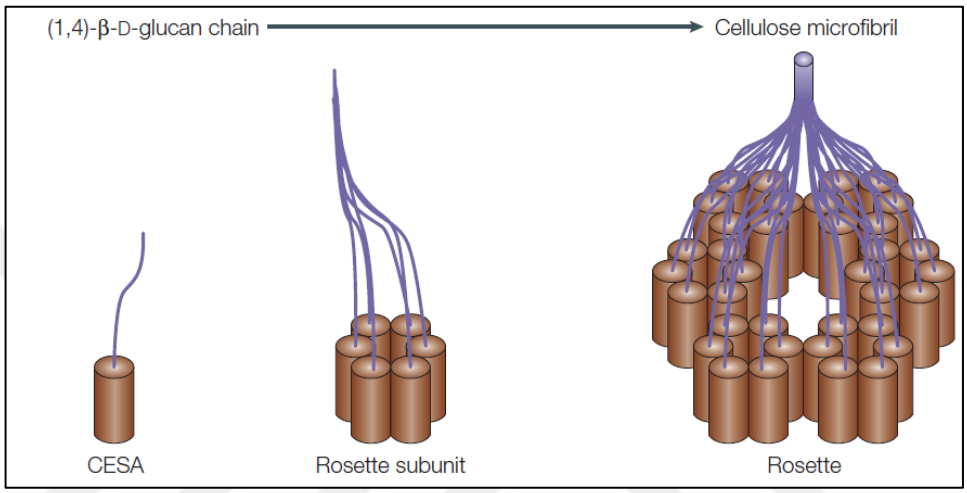


Figure 1.4. Rosette formation formed by CESA proteins. Each rosette unit contains 6 rosette subunit that is formed by 6 CESA proteins [7]

There is a protein superfamily named as Cellulose Synthase-like (CSL) that covers the CESA protein family. All of them are counted as glycosyltransferases. Cellulose Synthase-like superfamily has 9 different subgroup that basically they are called CSLA, CSLB, CSLC, CSLD, CSLE, CSLF, CSLG, CSLH, and CSLJ [23]. CSL proteins can show basic characteristic of transferases like β-glycosyltransferases. However, N-terminal regions are not present like CESA proteins. Due to high similarity in between CESA and CSL proteins, CSL proteins have involved the process of non-cellulosic β-linked polysaccharides synthase with β-D-glycans. It is thought that they are related with the production of mannan, xylan, and xyloglucan in the Golgi apparatus.

CSL superfamily is divided in according to their responsibilities. For instance, CSLA, CSLC, CSLF, and CSLH protein families are responsible from glucan backbones synthesis of non-cellulosic polysaccharides [24]. For the mannan backbone synthesis, CSLA is involved [25]. CSLB and CSLG family proteins are only observed in eudicot plants.

Moreover, CSLH, CSLF, and CSLJ are specific protein families for the monocot *Poaceae* family. In addition, it was demonstrated that at least one member of the CSLC group is specialized for xyloglucan backbone synthesis [26]. There are some unclear thoughts about CSL superfamily, such some of the proteins' functions are still unknown and still there are many investigations about how do they involve in these 1,4-β-D-glucan synthesis [27]. Detailed phylogenetic tree is explained below in Figure 1.5.

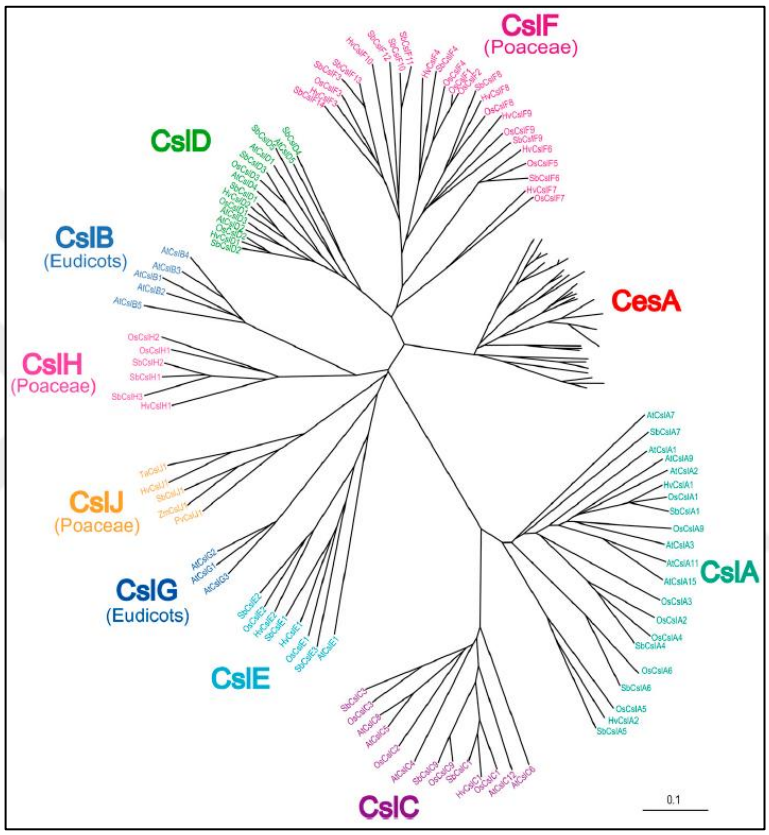


Figure 1.5. Phylogenetic tree of cellulose synthase (*CESA*) and cellulose synthase-like (*CSL*) genes and families in plants [28]

1.5. MIXED-LINKED B-GLUCANS

In cereals and grasses' cell walls, there have been observed a glucan that has been called (1;3,1;4)-β-D-glucan, in other words, mixed-linked glucan (MLG) [29]. It is an unbranched, linear group that they are not arranged or formed in ordered and repeated sequences. When the two or more (1,4)-β-D-glucosyl residues are linked, a different one, (1,3)-β-D-glucosyl has been attached [30]. Two different parts have been identified in MLG, which are tetra saccharide part that contains two β-D-Glcp-(1,4), one β-D-Glcp-(1,3) and Glcp; and

trisaccharide that contains one β-D-Glcp-(1,4), one β-D-Glcp-(1,3) and Glcp. Depends on species, the percentage of (1,3)-β-D-glucosyl can be changed, in between 5 and 20 percent. This structure is figured in Figure 1.6.

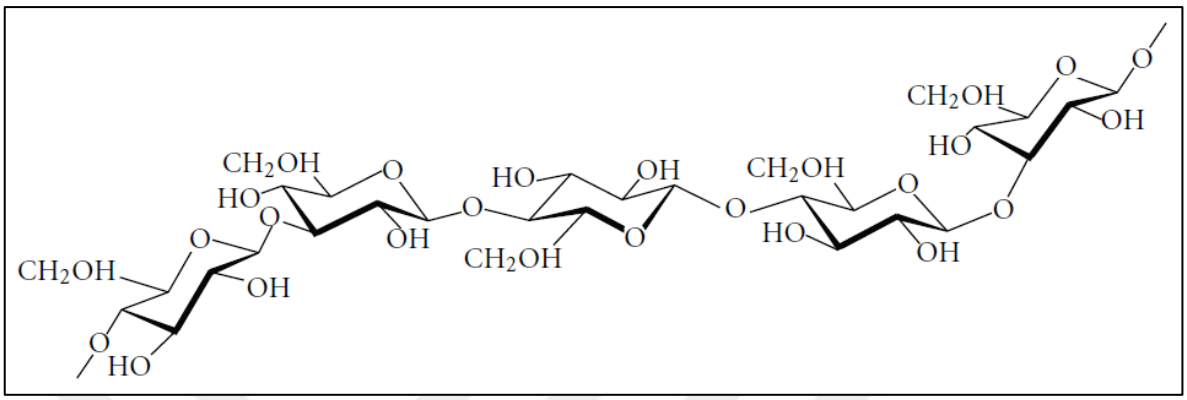


Figure 1.6. Molecular comprehension of mixed linked (1;3,1;4)- β-D-glucan [31]

In *Poaceae*, MLG is taking an important role in cell wall growth and also building in primary cell wall structure. MLG is found in these plants' endosperm tissue, and also expanding tissues. The study suggested that after the germination, those MLG residues have been hydrolyzed, so they have been used as carbohydrate storage [32]. Because of MLG's asymmetric structure, those polysaccharides can be more soluble even polymerization at a high level. This situation is also leaded to more flexible structure for primary cell walls in vascular plants. There is an unclear point for the location that MLGs have been synthesized. The study discussed that they are synthesized in Golgi apparatus [33]. On the other hand, the research showed that MLGs have been synthesized in maze coleoptiles' Golgi apparatus [34].

There is an enzyme group that is responsible from MLG transferase activity. They have been named as mixed-linkage glucan:xyloglucan endotransglucosylase (MXE). These enzymes have been identified in *Equisetum* genus [35]. In addition, MLG can be found especially in these species aged tissues, so MXEs could have a role in cell wall thickening in *Equisetum arvense* [36].

MLGs are not specific to the *Poaceae*, they could be observed and detected across many monocots. Because these plants have enzymes that can be capable of synthesizing 1,3-β and 1,4-β glucan linkages. That's why, MLGs have been synthesized from gene families called

CSLF, *CSLH*, AND *CSLJ* [37]. Moreover, these enzymes enroll MLG synthesis in these plants' Golgi apparatus [38, 39].

1.6. XYLAN

Xylans are classified as non-cellulosic β-linked polysaccharides that generally consists of β-1,4 xylosyl units that is linked with linearly. Its xylose backbone would contain substitutes like arabinose and glucuronic acid. They have been found in both parts of the cell wall, but mainly in secondary one. In addition, they are located much more in cereal grains and grass species' cell wall structures, about 30 percent of polysaccharide content in cell wall [40]. Because of different substitute has been linked to xylose residue, naming will be changed. If the α-1,2- or 1,3-L-arabinofuranosyl (Ara*f*) will be substituted, a new xylan formation is called arabinoxylan. This arabinose residue is going to lead to make this polysaccharide form more soluble. In this way, aggregation in intercellular level will be prevented [41]. Those arabinoxylans can be found in many types of cereal species endosperm's cell walls.

Several enzymes have been contributed the synthesis of xylan. Especially in *Arabidopsis thaliana,* two different protein families are involved for xylan synthesis which are GT43 and GT47. GT43 family members, which are IRX9, IRX9L, IRX14 and IRX14L, are leaded to xylan chain elongation. The other glycosyltransferases from GT47 family, which are IRX10 and IRX10L, is responsible from backbone elongation [42]. On the other hand, the study shows that the main glycosyltransferases for xylan synthesis are IRX9, 10, and 14, but due to overexpression of other three enzymes, their functions can be changed [43].

Because it is consisted with polysaccharides like pentoses, xylan is getting attention as source for biofuel production. These pentoses would be needed different enzymes for degradation, so fermentation process will be difficult. That's why, the study is focused on manipulation of synthesis of xylan to understand cell wall structure much easier [44] and facilitate other polymers that are located in cell wall [45].

1.7. MANNAN

Mannans are member of non-cellulosic β-linked polysaccharides that is consisted with β-1,4-mannose solely or combination of glucose and mannose sugars with β-1,4-linked. Some other substitutes like α-1,6-galactosyl could be linked with mannose backbone, and it is named as galactomannan. In addition, these galactosyl residues might be interacted with heteromannan backbone, which is called gluctomannan, and when it is linked with that, the name will be changed as galactoglucomannan. These interactions can lead to increase in solubility, in other words, galactosyl residues could affect solubility positively [46]. Moreover, from its O-2 and O-3 sides, mannose backbone would be acetylated [47]. Molecular comprehensions are stated in Figure 1.7. below.

Mannans can be found in many terrestrial plants, and also algae. They have been used in different purposes such as storage and structural in seeds [48]. Especially in coniferous plants' seeds and softwood parts, galactoglucomannans are very abundant. In tomato seed endosperm, mannan can provide rigidity, and also be affected on embryonic root process [49]. Although it is found in different plants' seeds, their functions are not clarified yet. They are located in cell wall of *Arabidopsis thaliana*. However, even the deletion of synthase genes of mannan, there has no phenotypic change been observed [50].

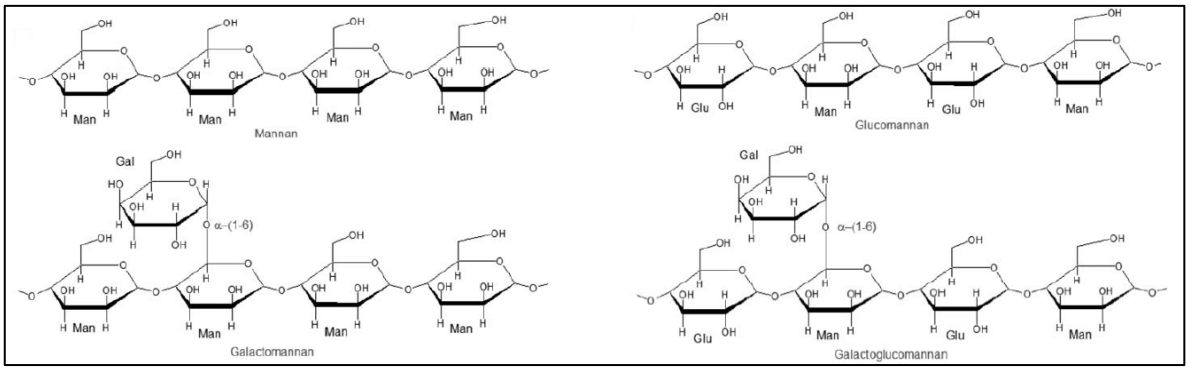


Figure 1.7. Molecular comprehensions of different mannans [51]

1.8. PECTIN

Pectins are structurally complex heteropolysaccharides that are consisted with an oxidized form of galactose which is called galacturonic acid. Mainly they are located in middle

lamella and primary parts of cell wall. Grasses contain around two percent of pectins in their primary cell walls. On the other hand, woody tissues are consisted much more pectins, up to 10 percent of primary cell wall content [52]. Not only these tissues, but pectins can also be found in many plant tissues such as growing and elongation parts of plants. In addition, dicots' primary cell wall is consisted much more pectin, up to 35 percent of whole content, comparing with monocots [53].

It is the important point that understanding of pectin synthesis and its structure is fundamental for understanding and also modifying of plant cell wall. Because of it has industrial importance, it has been investigated in many researches. Many important functions have been found such as growth and development, defense, structure and cell wall porosity [54]. In the industry level, it has been using for a stabilizer agent in food and cosmetics. In addition, there are some other studies related with pharmaceutical levels, like reducing LDL level of blood, and even cancer progress [55].

1.9. XYLOGLUCAN

In primary cell walls, xyloglucan (XyG) is the major non-cellulosic β-linked polysaccharide. Especially in flowering plants' primary cell walls, xyloglucan comprise up to 20-25 percent. On the other hand, this percentage is deeply decreased in the primary cell walls of monocots that are categorized in non-graminaceous, about 2%. [56]. Its structure is highly branched that is consisted with 1,4-β-linked glucans and 1,6-α-xylosyl residues [57]. XyGs can be linked to cellulose with hydrogen bonds that are involved in cell growth process. Moreover, XyGs can be formed into in secondary cell wall to create thick and non-lignified structure [58]. That are the reasons that they are important and essential for cell wall.

Not only glucans or xylosyl residues, but there are also some other saccharides that are involved xyloglucan structure. For instance, galactosyl residues can be attached to xylosyl groups from β-(1,2) linkage, and then fucosyl residues can attach to those galactosyl residues. In addition, O-acetyl groups could be also attached to galactosyl groups, but at O-2 position [59]. Those linkage positions and branching points could affect xyloglucan's connection and adsorption to cellulose in both ways. For example, fucosylated xyloglucan residues could adsorb much effectively, comparing with galactosylated xyloglucan connection to cellulose. The study showed that if the XyG residue has lower molecular

weight than others, then it can be connected to cellulose much easier than heavier ones [60]. Moreover, due to different substitute, bond energy changes can affect the adsorption levels and binding capacities.

XyG residues are embedded in microfibrils of cellulose with hydrogen bonds and also van der Waals force. The minimum number of glucose units that needs to be linked to the cellulose is around 12-16 for xyloglucan [61]. In the Golgi apparatus, xyloglucan is synthesized and with secretory vesicles, it is transported to the plant cell's plasma membrane. In the outer surface of the plasma membrane, transported xyloglucan residues are interacted with microfibrils to connect with cellulose that is located in primary cell wall [62]. However, there are some other approaches for understanding xyloglucan-microfibril interaction. It is believed that it is the xyloglucan that can connect other cellulose microfibrils together with binding spontaneously. Another thought is that xyloglucans might be stuck in microfibrils, so they should connect with cellulose directly or some other polysaccharides that are not located on cellulose. This situation can be caused to tether formation for microfibrils. In the end, another idea is suggested that normally microfibrils are surrounded with xyloglucans, so other XyGs could bind other non-cellulosic polysaccharides for cellulose connection [63].

Because of many saccharides and other substitutes, there are various XyG sidechain patterns have been formed. They have named with a single letter nomenclature, firstly classified by Fry *et al*. (1993) [64]. To form a xyloglucan residue, generally it is started with β-D-Glucosyl that can be grown up to four residue-long. According to recent study, there are 19 different XyG sidechains have been observed and classified, just like showed in Figure 1.8. [65]. All of them have indifferent letter naming for nomenclature. For instance, if there is only β-D-Glucosyl, that sidechain is called G. If α-D-xylosyl residue is linked with glucosyl backbone from α-1-6 linkage point, that sidechain is named as X. Other mainly observed substitutes are D- and L-galactose, D-xylose, L-fucose, and L-arabinofuranose.

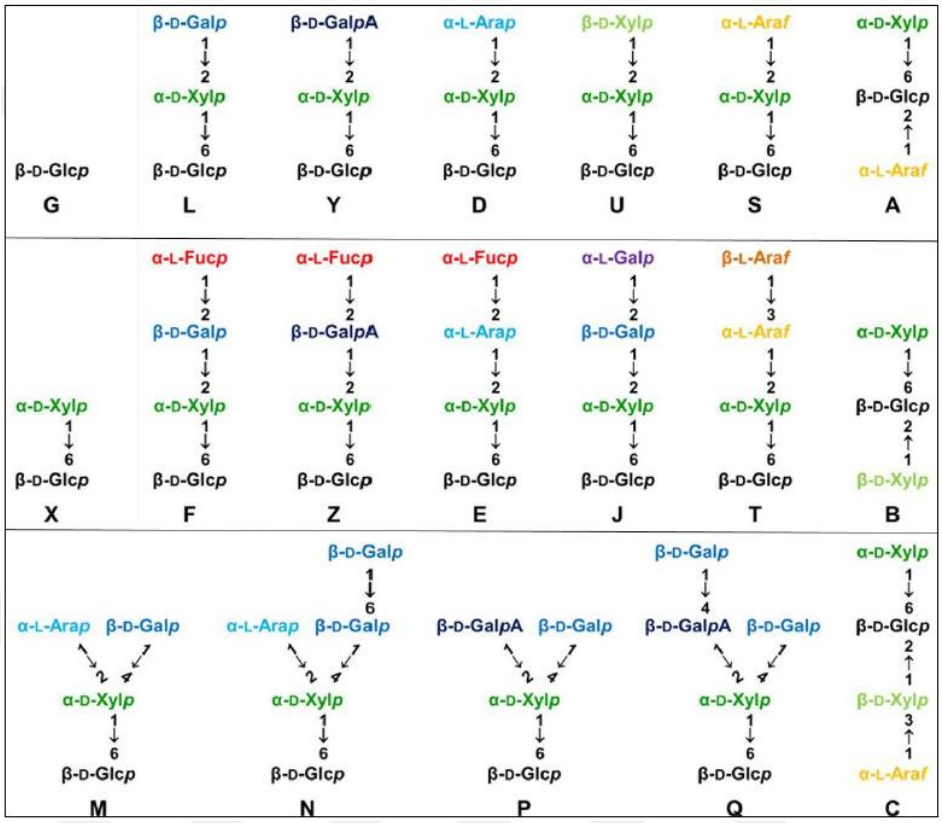


Figure 1.8. Chemical structures and nomenclatures of XyG sidechains. Elemental β-(1,4)-glucosyl backbones have been represented as black color, numbers are representing for linkage point for hydroxyl group of the saccharide. Arrows are representing the direction of glycosidic bonds [65]

From species to species, the structure of xyloglucan backbones can be varied. For instance, many kinds of vascular plants, XXXG is repeated in their XyG [66], whereas grasses can produce much fewer xylosyl contained xyloglucan backbones such as XXGG and XXGGG [67]. In the developmental stage of vascular plants, their XyG residues might include XXLG, XXFG, and XLFG [68]. On the other hand, *Poaceae* family plants such as barley (*Hordeum vulgare*) and rice (*Oryza sativa*) have XXGG or XXGGG type xyloglucan backbones, their unbranched glucose groups consist acetyl substitutes, rather than xylosyl substitute [69]. These formations are schematized below in Figure 1.9.

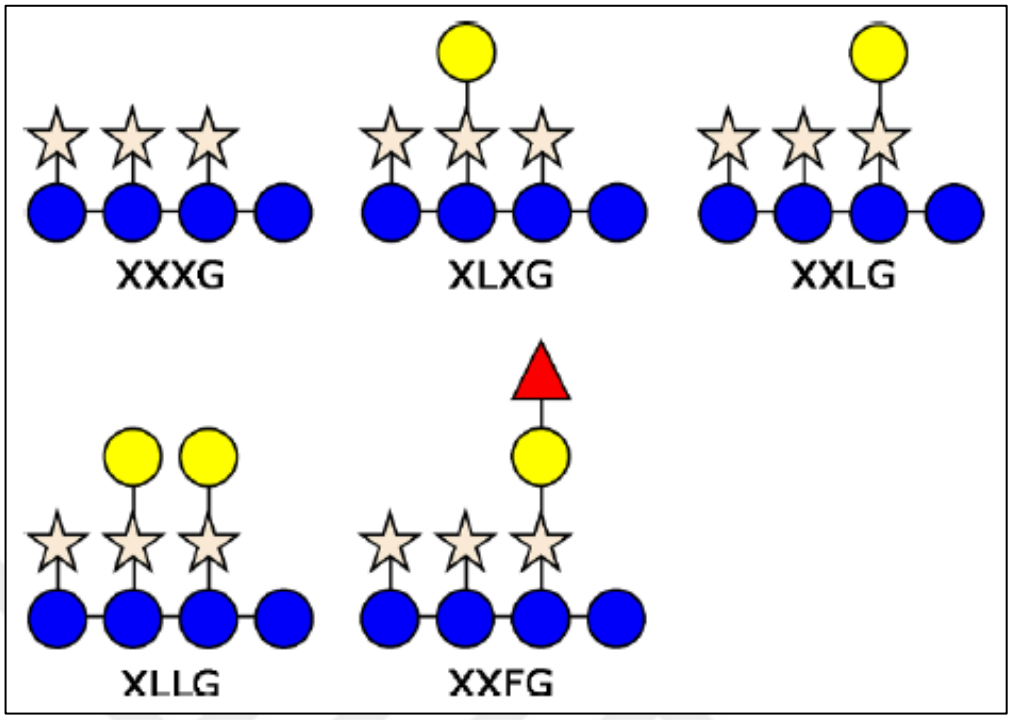


Figure 1.9. Schematic structures of xyloglucan (XyG) with different oligosaccharide units. Blue circles are represented glucosyl units; linked together with β-(1,4) linkage. Xylosyl substitutes (orange stars) are joined with α-(1-6) linkage. Galactosyl (yellow circles) and fucosyl (red triangle) are introduced in xyloglucan structure via β and α-(1-2) linkages [70]

Glucan synthase synthesizes xyloglucan's glucan backbone, and other side chains' substitutes have been produced with their own synthase enzymes, in other words, glycosyl transferases. For example, for xyloglucan synthesis in *Arabidopsis,* at least five different transferases have been involved. For xylosyl transfer, three different enzymes called XXT1, XXT2, and XXT5 are involved for transferring them to glucose residues. MUR3 are introduced for attaching galactose to xylosyl substitute, named as galactosyltransferase. There is a fucosyltransferase classified as FUT1, transferred fucose group to galactosyl group [71]. Not only glucan synthases themselves, cellulose synthase-like C (CSLC) synthase clade could introduce in β-1,4-glucan biosynthesis for XyG [72]. For instance, AtCSLC4 is responsible from xyloglucan backbone synthesis in *Arabidopsis thaliana* [73].

1.10. THE OBJECTIVE OF THESIS

The objective of this master's thesis is to investigate the genes that are involved in xyloglucan synthesis processes in barley. For this purpose, comparisons to xyloglucan synthesis genes from *Arabidopsis thaliana* and *Oryza sativa* species will be performed. Xyloglucan (XyG) is a β-linked glucosyl polysaccharide that is found in the cell walls of most plants, and it is very important for the structural organization of cell wall and also growth and development control. For these reasons, XyG biosynthesis should be investigated and enzymes that are involved in the process should be studied for a better understanding of the biosynthesis process. There has been some research on one of the main enzymes that is involved in XyG synthesis in barley, specifically Cellulose synthase-like C (CSLC) genes that are thought to be responsible for the formation of the (1,4)- β-linked glucosyl backbone. Recent studies have shown that there are several other types of enzymes that are involved in XyG synthesis which include xyloglucan xylosyltransferase, galactosyl transferase, and acetyl transferase. These genes have not yet been studied in detail in other species, have not been studied at all in barley. Moreover, making multiple sequence alignments and phylogenetic trees of them, comparing with *Arabidopsis* and rice species can help reveal important amino acid residues of the enzymes through conservation of sequences. Furthermore, another aim is that to determine these genes' location on chromosomes, their sizes and structures (intron/exon). The expression patterns of all identified genes will be examined from publicly available RNA-seq data. Co-expression analyses will be investigated both for XyG synthesis genes as well as the genes for XyG modifying enzymes such as xyloglucan endotransglycosylase/hydrolases. If necessary, some of these genes will be selected for RT-qPCR analyses, so plant growth and mRNA isolation would be applied in order to obtain cDNA for analysis. Results of this study will help to determine more details about XyG synthesis in barley and the number of genes involved, mainly by comparison with model plant species such as *Arabidopsis thaliana* and *Oryza sativa*. Knowledge of the genes involved, their tissue specificity and expression levels in the biosynthesis process of XyG will aid in the understanding of the potential roles of this import polysaccharide.

2. MATERIALS AND METHODS

2.1. IDENTIFICATION OF XYLOGLUCAN SYNTHESIS RELATED GENES IN *HORDEUM VULGARE*

As a covering up all possible related genes that are responsible from Xyloglucan (XyG) synthesis in barley, four different gene group have been targeted which are Cellulose synthase-like C (CSLC), xylosyl transferases, galactosyl transferases, and acetyl transferases. For detection of these genes, first of all, word-searching method has been used from NCBI database (https://www.ncbi.nlm.nih.gov/) for genes and also their proteins. For further targeting, detected genes' sequences have been used in BLAST search on ENSEMBL Plants (https://plants.ensembl.org/). Even further, some other species' homolog genes have been investigated for any possible prediction from *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon*. In addition, some other genome annotation public databases have been used such as Phytozome and IPK Galaxy Blast Suite (https://phytozome-next.jgi.doe.gov/; https://galaxy-web.ipk-gatersleben.de/) for identification. Whole targeted genes that belong four different subgroups have been detected and identified. Finally, all candidate genes' genomic sequences and protein sequences have been recorded and listed in FASTA format.

2.2. PHYLOGENETIC ANALYSIS AND SEQUENCE ALIGNMENT

The resulting barley Cellulose synthase-like C (CSLC), xyloglucan xylosyl transferase (XXT), xyloglucan galactosyl transferase (XGT), and xyloglucan O-acetyl transferase (O-AcT) protein sequences, together with the rice and *Arabidopsis* thaliana same groups, have been used to generate multiple sequence alignments (MSA) via Clustal X software (version 2.1) [74]. Comparison in between each of the species have been made and recorded. Furthermore, those sequences have been included into subsequent phylogenetic analyses. Phylogenetic trees have been visualized with Dendroscope software (version 3.7.5) [75]. Visuals have been recorded.

2.3. GENOMIC LOCATION DETERMINATION OF TARGETED GENES

All targeted genes' location on chromosomes have been determined (their locus and also strand). After their start-stop points have been designated with BLAST searches, their introns/exons have been identified for expression analysis. Number of base pairs for every gene have been detected, and every intron/exon position on chromosome determined on ENSEMBL Plants (https://plants.ensembl.org/).

2.4. EXPRESSION PATTERN ANALYSIS IN RNA-SEQ DATA

RNA-seq data form barley using next generation sequencing (NGS) have been performed from on barley cultivar Morex, was obtained from the James Hutton Institute, Dundee, Scotland. The files were downloaded from NCBI from Bioproject PRJEB14349. There were 16 different tissues from which the mRNA was sequenced from two batches. Each batch has three different replicates that gives a total 96 sample that have been investigated.

Table 2.1. Information list of RNA-seq data of predicted genes from an assembly of whole-genome-shotgun sequence from barley. Each tissue has 3 replicates from 2 different batches

Tissue code	Tissue description	Paired-end reads
EMB	Embryo (germinating)	142,000,000
ROO	Root (10 cm seedlings)	192,000,000
LEA	Shoot (10 cm seedlings)	135,000,000
INF1	Inflorescence (0.5cm)	154,000,000
INF2	Inflorescence (1-1.5cm)	176,000,000
NOD	Tillers (3rd internode)	557,000,000
CAR5	Grain (5 DPA)	149,000,000
CAR15	Grain (15 DPA)	166,000,000
ETI	Etiolated (10 day seedlings)	134,000,000
LEM	Lemma (6 weeks pa)	162,000,000
LOD	Lodicule (6 weeks pa)	142,000,000
PAL	Palea (6 weeks pa)	138,000,000
EPI	Epidermis (4 weeks)	133,000,000
RAC	Rachis (5 weeks pa)	144,000,000
ROO2	Root (4 week seedling)	144,000,000

SEN	Senescing leaf (2 months)	117,000,000

Expression values are represented as Fragments Per Kilobase of transcript per Million mapped reads (FPKM), and have been calculated and received with using Cufflinks software (version 2.2.1) [76]. All FPKM values that belong each of the gene have been correlated each other on Microsoft Excel 2016 (version 2102) [77]. Applied correlation method is Pearson that statistical analysis was based on that method. Correlation values have been investigated for co-expression analyses.

3. RESULTS

3.1. BIOINFORMATIC ANALYSES OF XYLOGLUCAN SYNTHESIS RELATED GENE FAMILIES IN HORDEUM VULGARE

3.1.1. Identification of Barley Gene Groups and their Locations on Genome

From careful screening of available barley genomic and transcript sequences, it was revealed that there are at least 5 Cellulose synthase-like C (CSLC), 10 xylosyl transferase (XXT), 17 galactosyl transferase (XGT), 33 acetyl transferase (O-AcT) genes, and 11 fucosyl transferases (FUT) that may be related with the synthesis of xyloglucan in barley. Totally, 76 different genes were identified as being potentially involved in the synthesis of xyloglucan in barley. All of the genes chromosomal locations were detemined and listed. Moreover, UniProt entry names were recorded and listed. All of those are tabled in Table 3.1., 3.2., 3.3., 3.4., and 3.5.

Table 3.1. List of identified Cellulose synthase-like C (CSLC) genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	Transcript ID	UniProt	Location	Strand
HvCSLC1	HORVU1Hr1G077230.2	F2DH86	Chr1H: 516783758-516787746	Forward
HvCSLC2	HORVU5Hr1G098490	-	Chr5H: 607885227-607888529	Reverse
HvCSLC3	HORVU3Hr1G000340.19	A0A287JV44	Chr3H: 680445-683426	Reverse
HvCSLC4	HORVU3Hr1G074570.5	A0A287LP59	Chr3H: 558854879-558859329	Forward
HvCSLC5	HORVU1Hr1G067470.1	A0A287G208	Chr1H: 478192712-478195887	Reverse

Table 3.2. List of xylosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	Transcript ID	UniProt	Location	Strand
HvXXT.1	HORVU6Hr1G040950.2.1	-	Chr6H: 222383217-222386248	Reverse
HvXXT.2	HORVU4Hr1G054910.1	F2EHF5	Chr4H: 460072407-460075545	Forward
HvXXT.3	HORVU5Hr1G042000.1	A0A287QYN1	Chr5H: 319144604-319146715	Forward
HvXXT.4	HORVU4Hr1G054120.2	F2ELV2	Chr4H: 450349845-450351444	Forward
HvXXT.5	HORVU7Hr1G102070.1	M0XCY6	Chr7H: 611130689-611132570	Reverse
HvXXT.6	HORVU4Hr1G054020.3	A0A287P3K4	Chr4H: 449910612-449911931	Reverse
HvXXT.7	HORVU6Hr1G071920.1	A0A287UMH0	Chr6H: 499927483-499929208	Forward
HvXXT.8	HORVU4Hr1G089670.1	A0A287Q3U7	Chr4H: 643380058-643381950	Forward
HvXylT_2.1	HORVU4Hr1G080470.1	A0A287PTZ8	Chr4H: 615749297-615750856	Forward
HvXylT_2.2	HORVU7Hr1G072980.4	M0X092	Chr7H: 411034504-411038129	Reverse

Table 3.3. List of galactosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	Transcript ID	UniProt	Location	Strand
HvXGT.1	HORVU1Hr1G053890.1	A0A287FMP6	Chr1H: 397672141-397673986	Reverse
HvXGT.2	HORVU4Hr1G002040.1	A0A287MX50	Chr4H: 3587964-3589369	Forward
HvXGT.3	HORVU2Hr1G094620.6	A0A287J079	Chr2H: 665452593-665454809	Reverse

HvXGT.4	HORVU5Hr1G020990.2	M0UPI8	Chr5H: 98708098-98709735	Reverse
HvXGT.5	HORVU4Hr1G079030.2	F2DD42	Chr4H: 610304842-610307561	Forward
HvXGT.6	HORVU0Hr1G005330.3	A0A287DW96	ChrUn: 29773904-29775163	Reverse
HvXGT.7	HORVU4Hr1G079010.1	A0A287PSH4	Chr4H: 609995681-609997761	Forward
HvXGT.8	HORVU3Hr1G002760.1	A0A287JXL7	Chr3H: 7035844-7037699	Forward
HvXGT.9	HORVU3Hr1G002770.1	A0A287JXN2	Chr3H: 7046764-7048586	Forward
HvXGT.10	HORVU4Hr1G078990	-	Chr4H: 609974781-609976927	Forward
HvXGT.11	HORVU3Hr1G002750.2	A0A287JXV8	Chr3H: 7021088-7022293	Reverse
HvXGT.12	HORVU4Hr1G079020.2	F2E230	Chr4H: 610203034-610205099	Reverse
HvXGT.13	HORVU3Hr1G048310	-	Chr3H: 332584888-332587185	Forward
HvXGT-like.1	HORVU1Hr1G038500.2	M0XL78	Chr1H: 267243137-267245981	Forward
HvXGT-like.2	HORVU3Hr1G016840.1	A0A287K957	Chr3H: 41605847-41607417	Forward
HvXGT-like.3	HORVU1Hr1G038510.1	M0VCI3	Chr1H: 267633848-267635918	Reverse
Hv_GALT_UCW98	HORVU7Hr1G024690.2	A0A287VR30	Chr7H: 40113928-40120154	Reverse

Table 3.4. List of acetyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	**Transcript ID**	**UniProt**	**Location**	**Strand**
Hv-AcT.1	HORVU1Hr1G072050.2	A0A287G658	Chr1H: 495875162-495876936	Forward
Hv-AcT.2	HORVU2Hr1G005330.5	A0A287GYQ2	Chr2H: 11646308-11647704	Forward

Hv-AcT.3	HORVU2Hr1G005330.1	F2EK02	Chr2H: 11638191-11640163	Forward
Hv-AcT.4	HORVU2Hr1G006540.1	-	Chr2H: 13693507-13694973	Reverse
Hv-AcT.5	HORVU2Hr1G012990	-	Chr2H: 28304849-28306765	Forward
Hv-AcT.6	HORVU7Hr1G076490.1	M0VIN7	Chr7H: 447269089-447270827	Forward
Hv-AcT.7	HORVU2Hr1G005310.2	A0A287GYP3	Chr2H: 11625325-11626704	Forward
Hv-AcT.8	HORVU3Hr1G085100.2	A0A287M173	Chr3H: 610947834-610949907	Reverse
Hv-AcT.9	HORVU1Hr1G073330.1	M0VNQ6	Chr1H: 503056222-503058634	Reverse
Hv-AcT.10	HORVU3Hr1G023780.1	A0A287KFL0	Chr3H: 88690770-88692993	Forward
Hv-AcT.11	HORVU7Hr1G000970.2	A0A287VC14	Chr7H: 2071822-2073387	Forward
Hv-AcT.12	-	-	Chr2H: 11795210-11796844	Forward
Hv-O-AcT.1	HORVU3Hr1G093240.1	A0A287MBH0	Chr3H: 641437481-641441165	Reverse
Hv-O-AcT.2	HORVU3Hr1G058870.2	M0XCC7	Chr3H: 445089296-445090583	Reverse
Hv-O-AcT.3	HORVU3Hr1G058830.3	F2CSH9	Chr3H: 444736521-444738710	Forward
Hv-O-AcT.4	HORVU7Hr1G042930.1	M0UPV6	Chr7H: 126843568-126844800	Reverse
Hv-O-AcT.5	HORVU6Hr1G072710.2	A0A287UN13	Chr6H: 504830160-504831801	Forward
Hv-O-AcT.6	HORVU7Hr1G043010.1	M0YE64	Chr7H: 127188009-127189797	Forward
Hv-O-AcT.7	HORVU1Hr1G029950.1	F2CXD9	Chr1H: 168736997-168744269	Reverse
Hv-O-AcT.8	HORVU6Hr1G072660.1	M0YYI6	Chr6H: 504771329-504773540	Forward
Hv-O-AcT.9	HORVU7Hr1G042960.1	M0V430	Chr7H: 127099919-127101877	Forward

Hv-O-AcT.10	HORVU7Hr1G039420.3	A0A287W5J8	Chr7H: 103316974-103319245	Reverse
Hv-O-AcT.11	HORVU7Hr1G039180.2	F2E2E6	Chr7H: 102290457-102292396	Forward
Hv-O-AcT.12	HORVU7Hr1G089640.1	M0VC71	Chr7H: 545402365-545406093	Forward
Hv-O-AcT.13	HORVU7Hr1G079700.2	A0A287X1G7	Chr7H: 470463536-470465383	Reverse
Hv-O-AcT.14	HORVU7Hr1G039080.1	F2E8N4	Chr7H: 101642714-101645559	Forward
Hv-O-AcT.15	HORVU7Hr1G035050.1	M0YLV2	Chr7H: 75333619-75337088	Forward
Hv-O-AcT.16	HORVU5Hr1G113770.2	M0VZ57	Chr5H: 642696361-642697923	Reverse
Hv-O-AcT.17	HORVU5Hr1G058980.1	A0A287RDN1	Chr5H: 460758928-460760923	Reverse
Hv-O-AcT.18	HORVU5Hr1G048030.1	F2DDT7	Chr5H: 372899242-372906729	Reverse
Hv-O-AcT.19	HORVU5Hr1G041030.1	F2DVF6	Chr5H: 308430853-308436832	Forward
Hv-O-AcT.20	HORVU4Hr1G090420	-	Chr4H: 645139609-645141372	Reverse
Hv-O-AcT.21	HORVU3Hr1G087740.1	A0A287M4J2	Chr3H: 622654280-622660060	Forward

Table 3.5. List of fucosyl transferase genes with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	**Transcript ID**	**UniProt**	**Location**	**Strand**
HvFUT.1	HORVU6Hr1G078500.1	A0A287UTE3	Chr6H: 534738671-534741608	Forward
HvFUT.2	HORVU7Hr1G026180.1	A0A287VRY8	Chr7H: 44495358-44497411	Reverse
HvFUT.3	HORVU7Hr1G035550.1	A0A287W0V3	Chr7H: 79246546-79248297	Forward
HvFUT.4	HORVU7Hr1G035560.1	A0A287W100	Chr7H: 79293982-79295704	Forward
HvFUT.5	HORVU7Hr1G035570.1	A0A287W0V8	Chr7H: 79304174-79306786	Reverse

HvFUT.6	HORVU7Hr1G035590.1	A0A287W126	Chr7H: 79324382-79326849	Forward
HvFUT.7	HORVU7Hr1G035620.3	M0UTP4	Chr7H: 79512802-79514885	Reverse
HvFUT.8	HORVU5Hr1G068430.1	A0A287RMK1	Chr5H: 517657373-517659219	Forward
HvFUT.9	HORVU6Hr1G078540.1.1	-	Chr6H: 535111409-535129068	Forward
HvFUT.10	HORVU6Hr1G078630.1	A0A287UTI7	Chr6H: 535343863-535346193	Reverse
HvFUT.11	HORVU6Hr1G078650.1	A0A287UU06	Chr6H: 535400289-535403895	Forward

After all targeted genes that related with XyG biosynthesis in barley had been found their genomic locations exactly, a visual of chromosomal map were prepared (in Figure 3.1.). In this visual, genes that are from CSLC, XXT, XGT, and O-AcT gene families' exact locations had been marked on barley's chromosomes. It was also analyzed for checking that are these genes neighboring to each other, and is there any correlation in between their positions to each other for their expression patterns. Especially for CSLC, there is no strong/high correlation in between the genes that located very close to these CSLCs, so in this case there is no any output for strong co-expression of neighboring genes. Moreover, there was no observed clustering in between these gene families' members to each other, according to their exact positions on barley chromosomes.

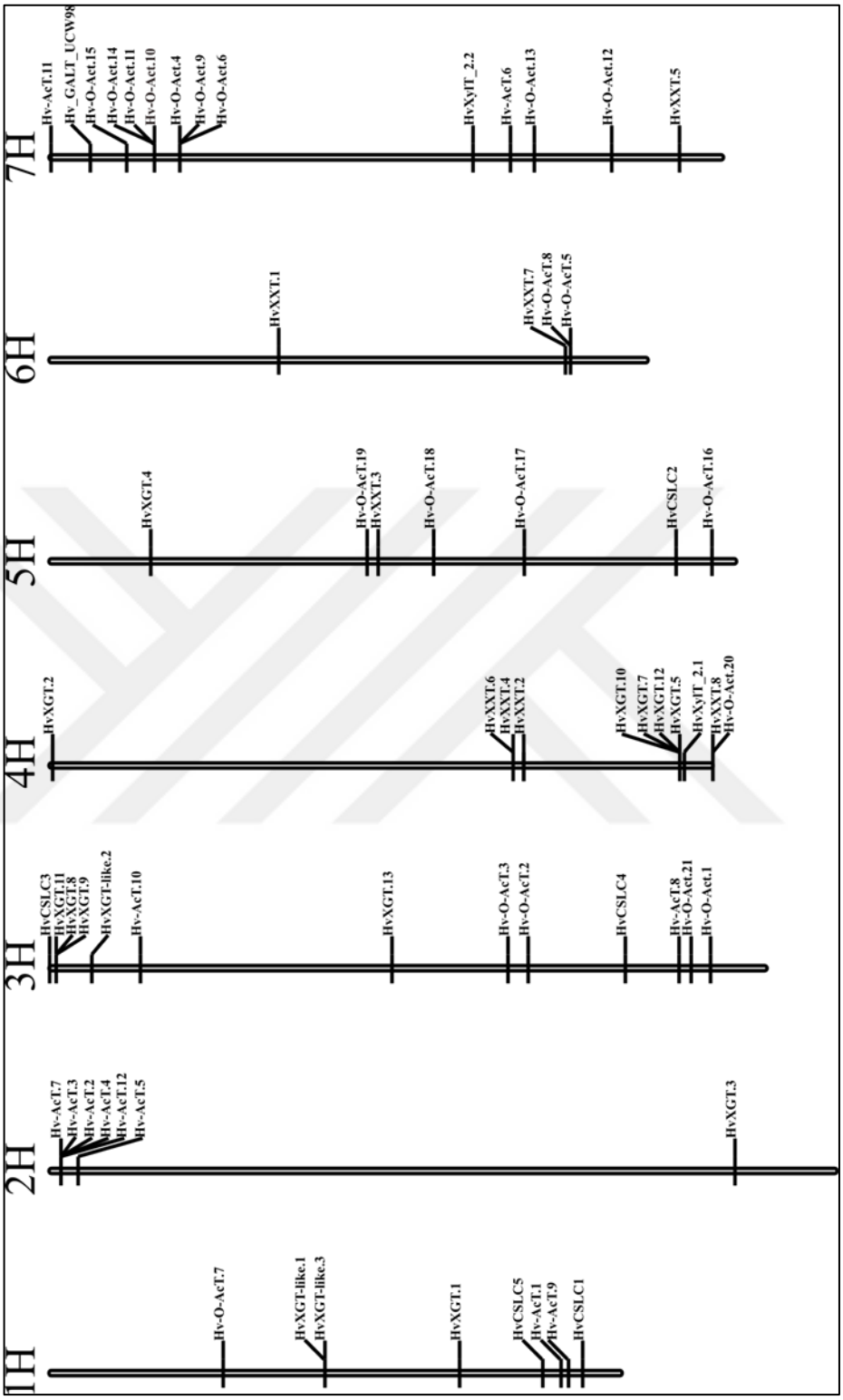
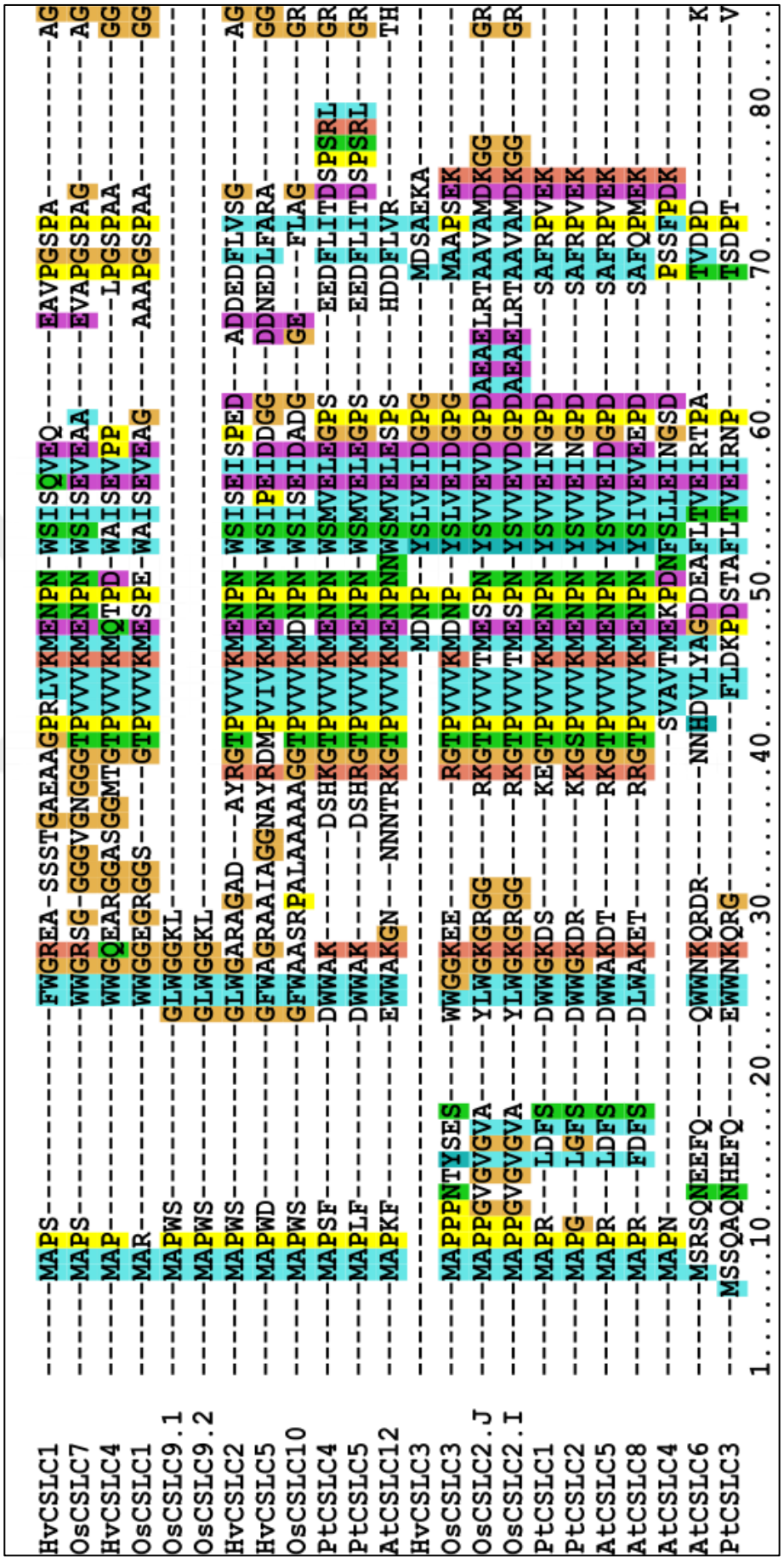
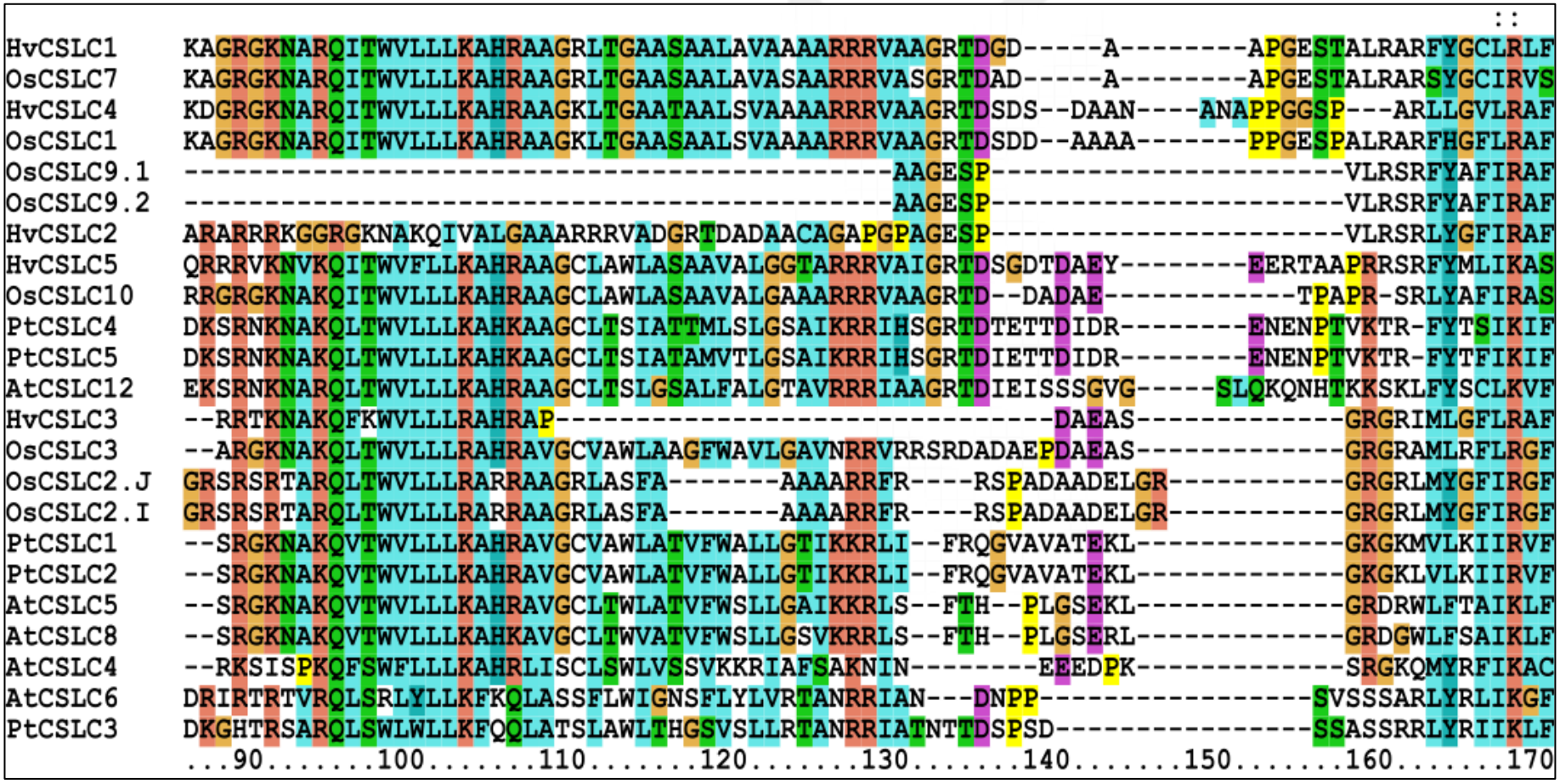


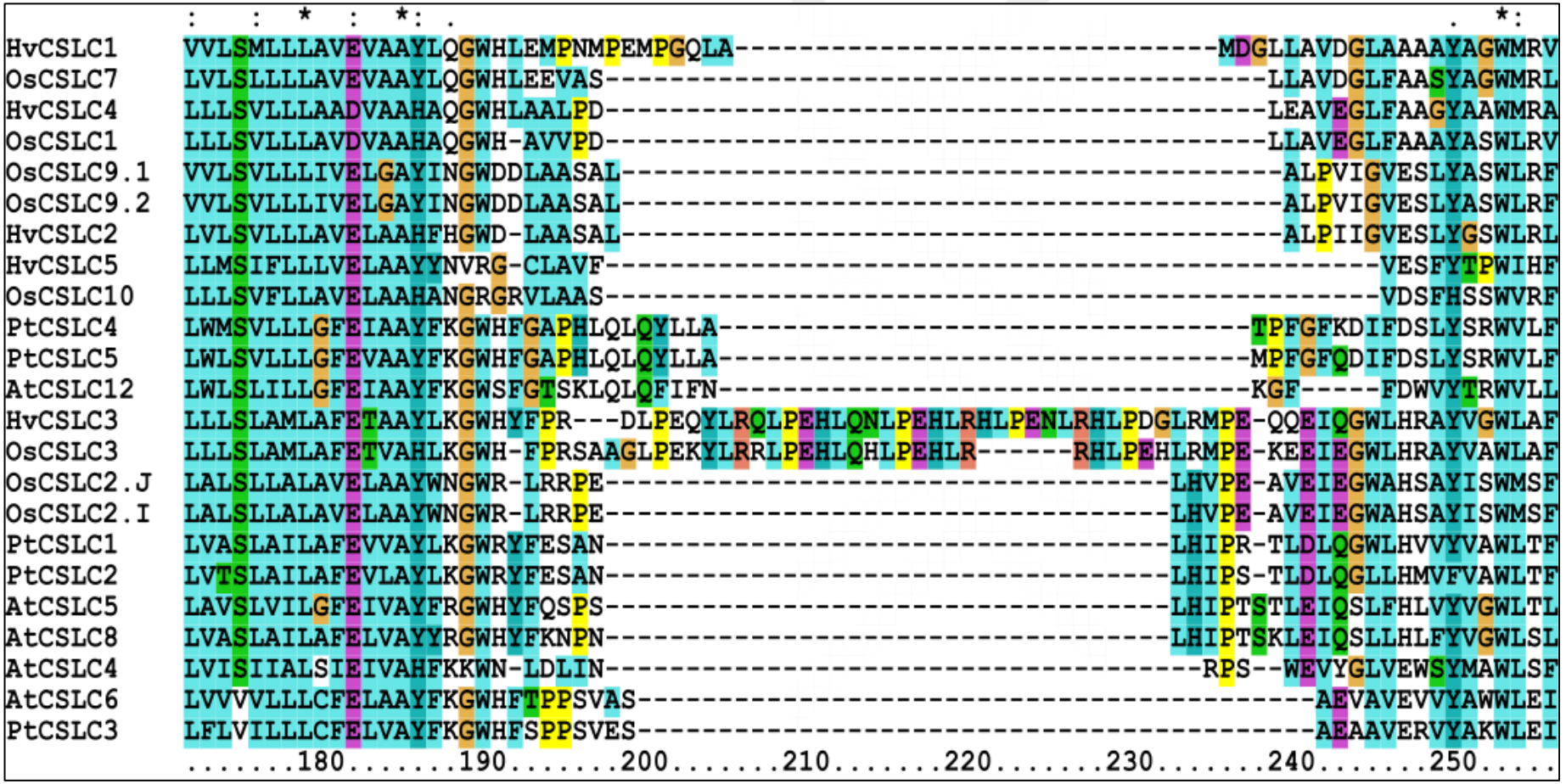
Figure 3.1. Chromosomal map of barley genome. CSLC, XXT, XGT, and O-AcT genes that determined for XyG biosynthesis were listed on figure. Length of chromosomes were adjusted and projected for real base pair length of those chromosomes

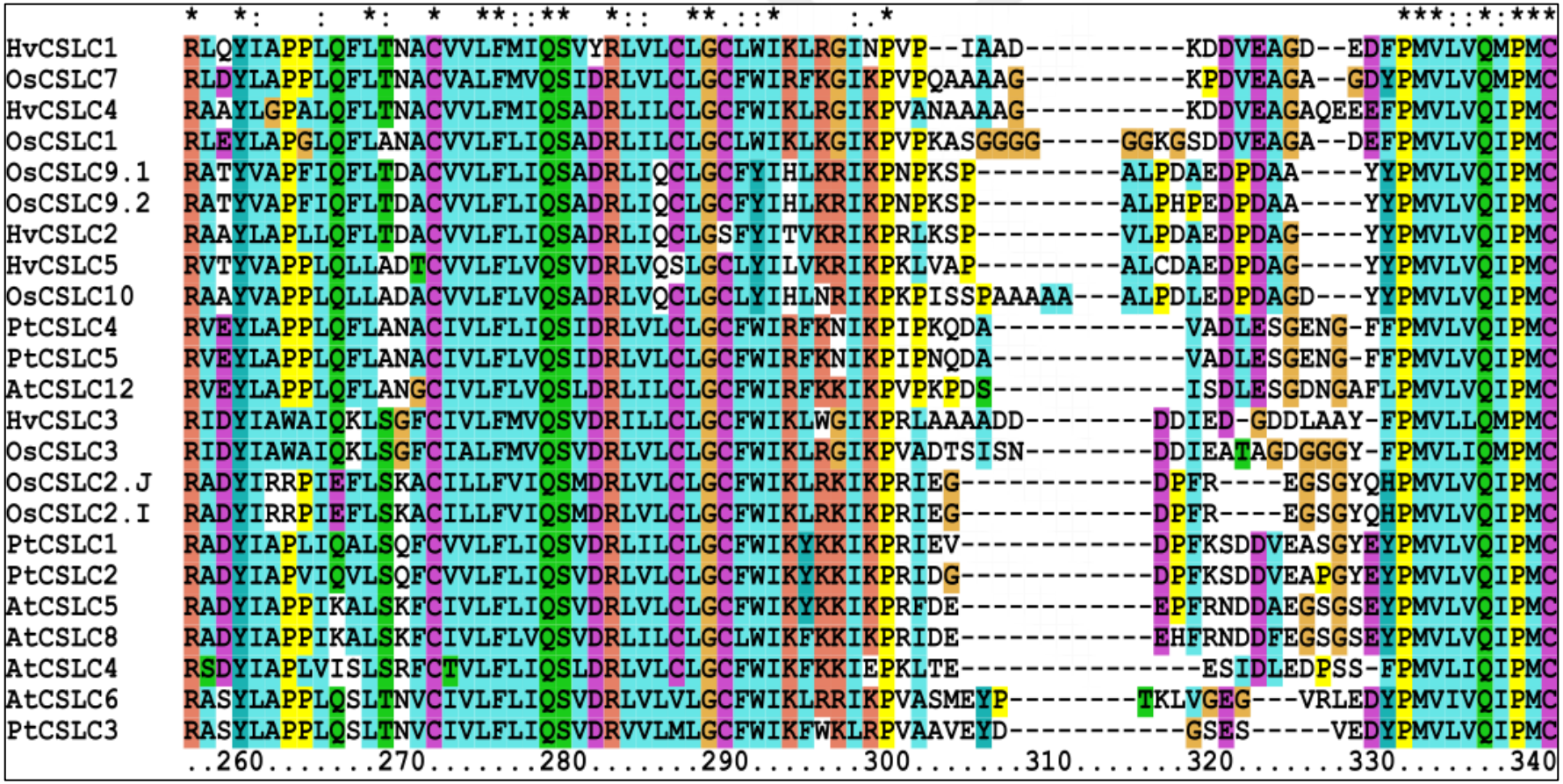
3.1.2. Multiple Sequence Alignments

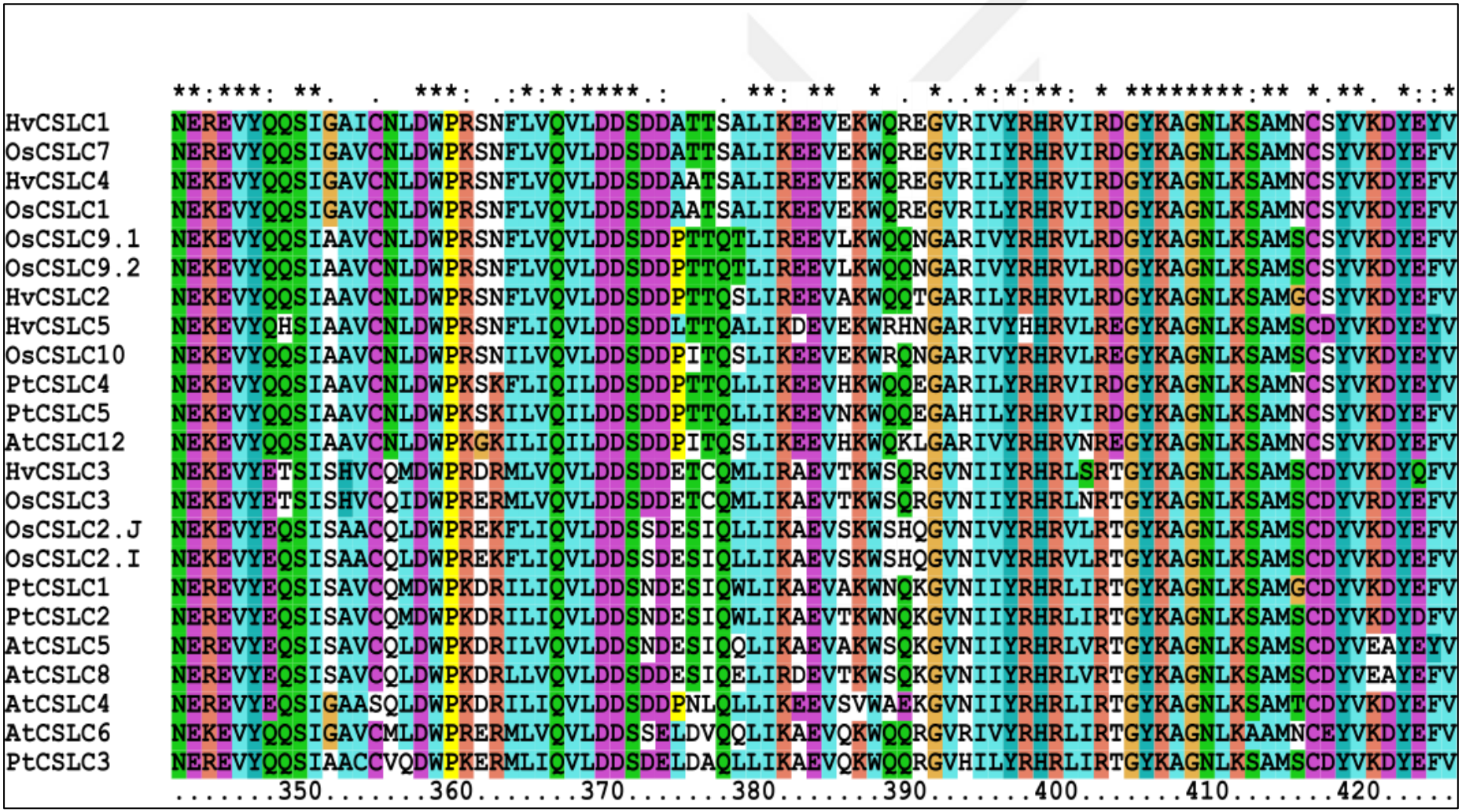
There are 65 different genes' protein sequences have been received from NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/) and also ENSEMBL Plants for barley. In addition, for alignment with other species' protein sequences such as *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon* were received from same online tools. Alignments were done and visualized on Clustal X software. Moreover, for adjusting alignments, a few protein sequences were predicted with online tools called FGENESH and FGENESH+ [78]. All protein sequences have been recorded in FASTA format. In total, 23 cellulose synthase-like c, 26 xylosyl transferase, 54 galactosyl transferases, and 76 acetyl transferases have been detected and identified for multiple sequence alignment from 5 different species. Due to high level similarity and same motifs, CSLC one is visualized below in Figure 3.2., and others are including in Appendix part.

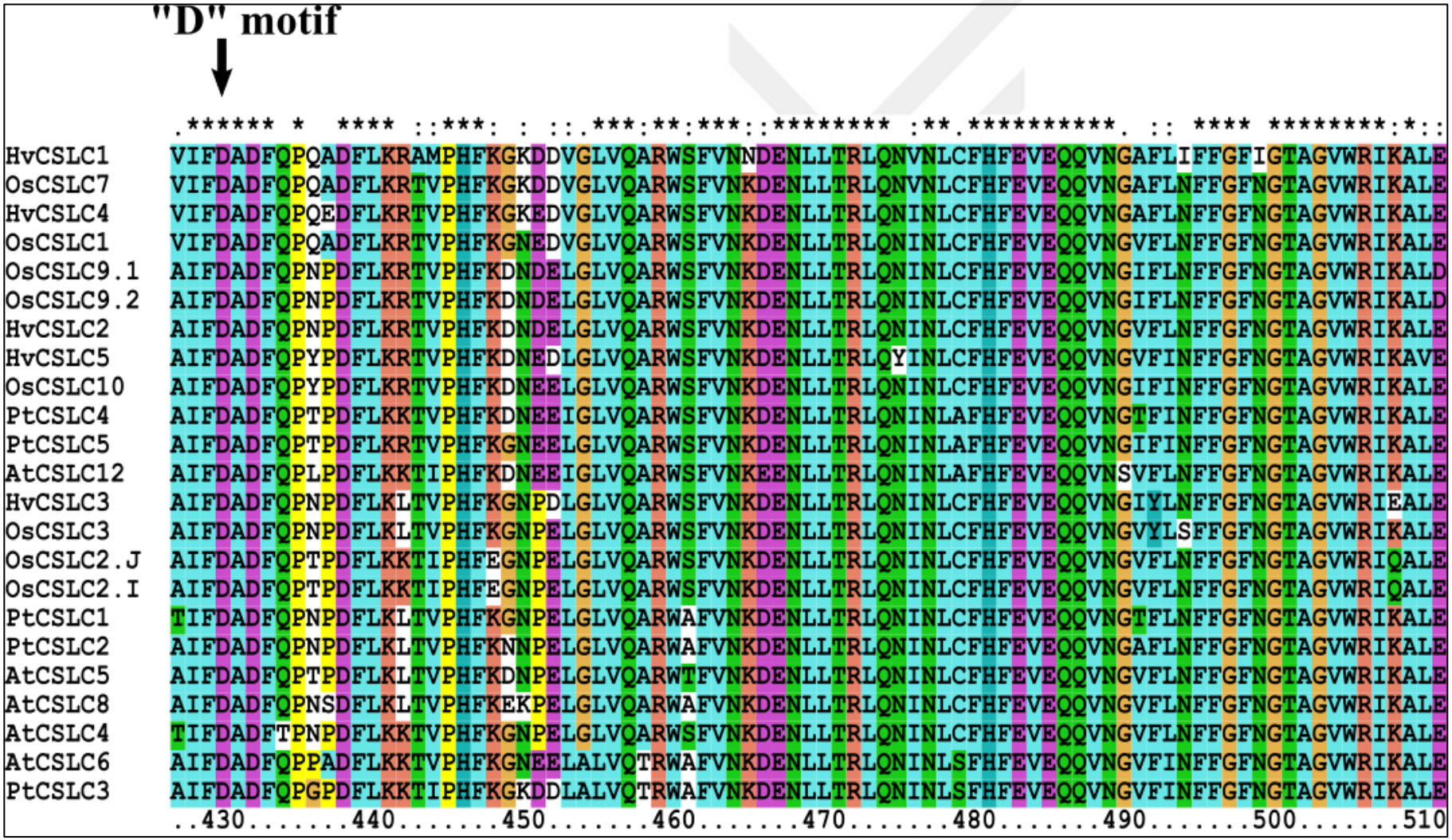

HvCSLC1
OsCSLC7
HvCSLC4
OsCSLC1
OsCSLC9.1
OsCSLC9.2
HvCSLC2
HvCSLC5
OsCSLC10
PtCSLC4
PtCSLC5
AtCSLC12
HvCSLC3
OsCSLC3
OsCSLC2.J
OsCSLC2.I
PtCSLC1
PtCSLC2
AtCSLC5
AtCSLC8
AtCSLC4
AtCSLC6
PtCSLC3

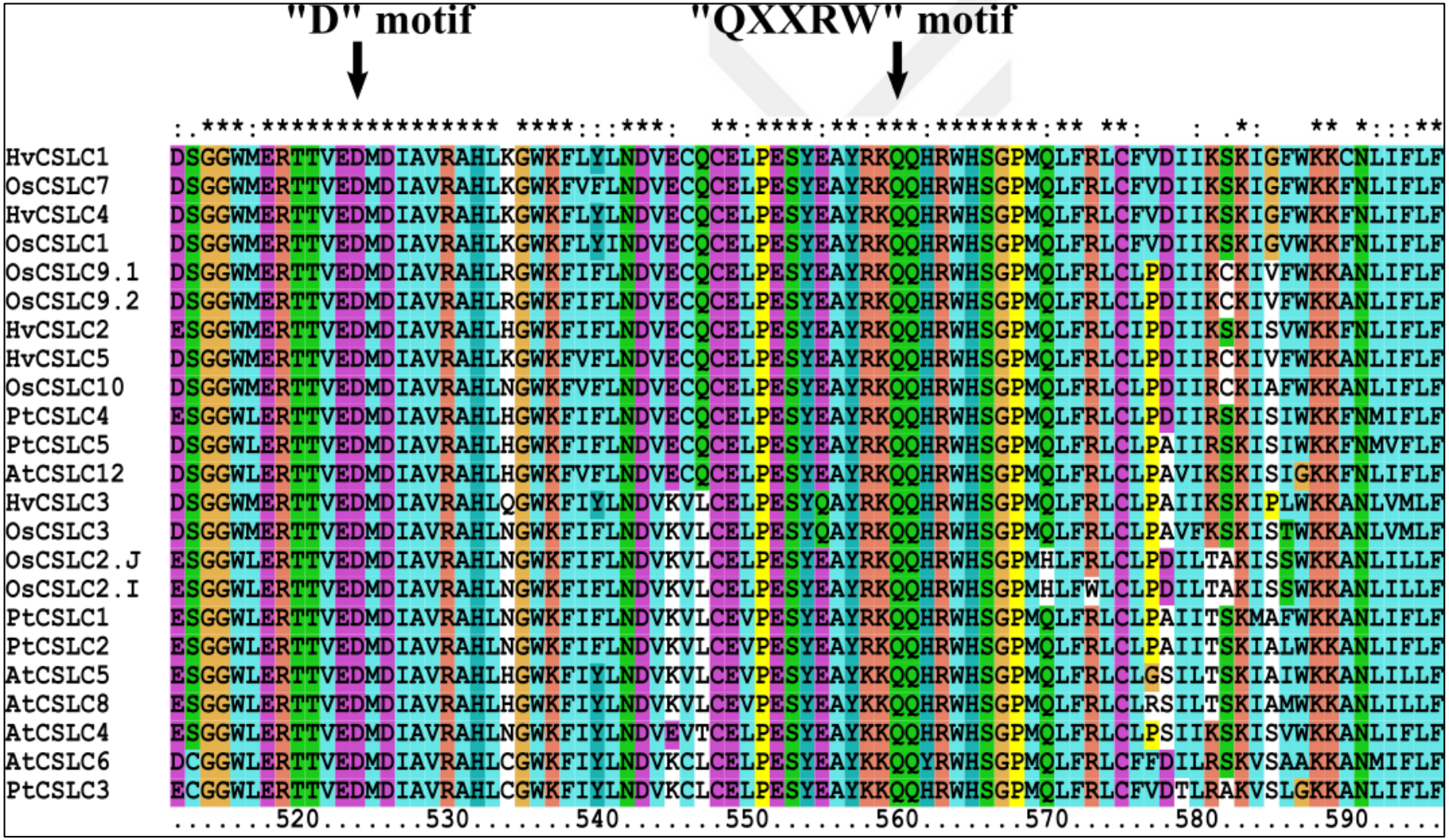

HvCSLC1 KAGRGKNARQITWVLLLKAHRAAGRLTGAASAALAVAAAARRRVAAGRTDGD-----A--------APGESTALRARFYGCLRLF
OsCSLC7 KAGRGKNARQITWVLLLKAHRAAGRLTGAASAALAVASAARRRVASGRTDAD-----A--------APGESTALRARSYGCIRVS
HvCSLC4 KDGRGKNARQITWVLLLKAHRAAGKLTGAATAALSVAAAARRRVAAGRTDSDS--DAAN----ANAPPGGSP---ARLLGVLRAF
OsCSLC1 KAGRGKNARQITWVLLLKAHRAAGKLTGAASAALSVAAAARRRVAAGRTDSDD--AAAA-------PPGESPALRARFHGFLRAF
OsCSLC9.1 ----------AAGESP----------VLRSRFYAFIRAF
OsCSLC9.2 ----------AAGESP----------VLRSRFYAFIRAF
HvCSLC2 ARARRRKGGRGKNAKQIVALGAAARRRVADGRTDADAACAGAPGPAGESP----------VLRSRLYGFIRAF
HvCSLC5 QRRRVKNVKQITWVFLLKAHRAAGCLAWLASAAVALGGTARRRVAIGRTDSGDTDAEY--------EERTAAPRRSRFYMLIKAS
OsCSLC10 RRGRGKNAKQITWVLLLKAHRAAGCLAWLASAAVALGAAARRRVAAGRTD--DADAE----------TPAPR-SRLYAFIRAS
PtCSLC4 DKSRNKNAKQLTWVLLLKAHKAAGCLTSIATTMLSLGSAIKRRIHSGRTDTETTDIDR--------ENENPTVKTR-FYTSIKIF
PtCSLC5 DKSRNKNAKQLTWVLLLKAHKAAGCLTSIATAMVTLGSAIKRRIHSGRTDIETTDIDR--------ENENPTVKTR-FYTFIKIF
AtCSLC12 EKSRNKNARQLTWVLLLKAHRAAGCLTSLGSALFALGTAVRRRIAAGRTDIEISSSGVG-----SLQKQNHTKKSKLFYSCLKVF
HvCSLC3 --RRTKNAKQFKWVLLLRAHRAP----------DAEAS----------GRGRIMLGFLRAF
OsCSLC3 --ARGKNAKQLTWVLLLRAHRAVGCVAWLAAGFWAVLGAVNRRVRRSRDADAEPDAEAS----------GRGRAMLRFLRGF
OsCSLC2.J GRSRSRTARQLTWVLLLRARRAAGRLASFA-------AAAARRFR----RSPADAADELGR----------GRGRLMYGFIRGF
OsCSLC2.I GRSRSRTARQLTWVLLLRARRAAGRLASFA-------AAAARRFR----RSPADAADELGR----------GRGRLMYGFIRGF
PtCSLC1 --SRGKNAKQVTWVLLLKAHRAVGCVAWLATVFWALLGTIKKRLI--FRQGVAVATEKL----------GKGKMVLKIIRVF
PtCSLC2 --SRGKNAKQVTWVLLLKAHRAVGCVAWLATVFWALLGTIKKRLI--FRQGVAVATEKL----------GKGKLVLKIIRVF
AtCSLC5 --SRGKNAKQVTWVLLLKAHRAVGCLTWLATVFWSLLGAIKKRLS--FTH--PLGSEKL----------GRDRWLFTAIKLF
AtCSLC8 --SRGKNAKQVTWVLLLKAHKAVGCLTWVATVFWSLLGSVKRRLS--FTH--PLGSERL----------GRDGWLFSAIKLF
AtCSLC4 --RKSISPKQFSWFLLLKAHRLISCLSWLVSSVKKRIAFSAKNIN--------EEEDPK----------SRGKQMYRFIKAC
AtCSLC6 DRIRTRTVRQLSRLYLLKFKQLASSFLWIGNSFLYLVRTANRRIAN---DNPP----------SVSSSARLYRLIKGF
PtCSLC3 DKGHTRSARQLSWLWLLKFQQLATSLAWLTHGSVSLLRTANRRIATNTTDSPSD----------SSASSRRLYRIIKLF
...90.....100.....110.....120.....130.....140.....150.....160.....170

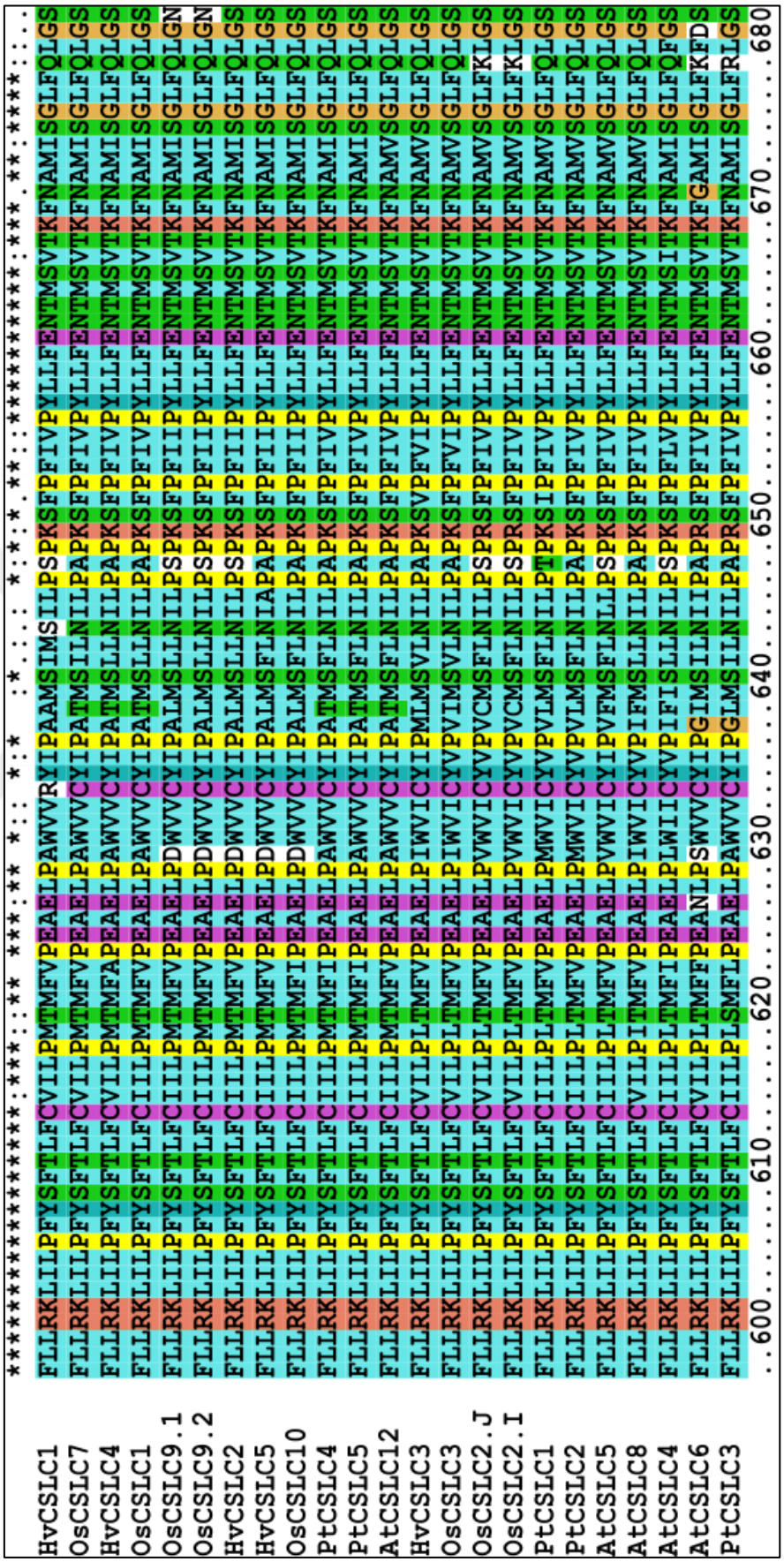

HvCSLC1 VVLSMLLLAVEVAAYLQGWHLEMPNMPEMPGQLA----------MDGLLAVDGLAAAAYAGWMRV
OsCSLC7 LVLSLLLLAVEVAAYLQGWHLEEVAS----------LLAVDGLFAASYAGWMRL
HvCSLC4 LLLSVLLLAADVAAHAQGWHLAALPD----------LEAVEGLFAAGYAAWMRA
OsCSLC1 LLLSVLLLAVDVAAHAQGWH-AVVPD----------LLAVEGLFAAAYASWLRV
OsCSLC9.1 VVLSVLLLIVELGAYINGWDDLAASAL----------ALPVIGVESLYASWLRF
OsCSLC9.2 VVLSVLLLIVELGAYINGWDDLAASAL----------ALPVIGVESLYASWLRF
HvCSLC2 LVLSVLLLAVELAAHFHGWD-LAASAL----------ALPIIGVESLYGSWLRL
HvCSLC5 LLMSIFLLLVELAAYYNVRG-CLAVF----------VESFYTPWIHF
OsCSLC10 LLLSVFLLAVELAAHANGRGRVLAAS----------VDSFHSSWVRF
PtCSLC4 LWMSVLLLGFEIAAYFKGWHFGAPHLQLQYLLA----------TPFGFKDIFDSLYSRWVLF
PtCSLC5 LWLSVLLLGFEVAAYFKGWHFGAPHLQLQYLLA----------MPFGFQDIFDSLYSRWVLF
AtCSLC12 LWLSLILLGFEIAAYFKGWSFGTSKLQLQFIFN----------KGF-----FDWVYTRWVLL
HvCSLC3 LLLSLAMLAFETAAYLKGWHYFPR---DLPEQYLRQLPEHLQNLPEHLRHLPENLRHLPDGLRMPE-QQEIQGWLHRAYVGWLAF
OsCSLC3 LLLSLAMLAFETVAHLKGWH-FPRSAAGLPEKYLRRLPEHLQHLPEHLR------RHLPEHLRMPE-KEEIEGWLHRAYVAWLAF
OsCSLC2.J LALSLLALAVELAAYWNGWR-LRRPE----------LHVPE-AVEIEGWAHSAYISWMSF
OsCSLC2.I LALSLLALAVELAAYWNGWR-LRRPE----------LHVPE-AVEIEGWAHSAYISWMSF
PtCSLC1 LVASLAILAFEVVAYLKGWRYFESAN----------LHIPR-TLDLQGWLHVVYVAWLTF
PtCSLC2 LVTSLAILAFEVLAYLKGWRYFESAN----------LHIPS-TLDLQGLLHMVFVAWLTF
AtCSLC5 LAVSLVILGFEIVAYFRGWHYFQSPS----------LHIPTSTLEIQSLFHLVYVGWLTL
AtCSLC8 LVASLAILAFELVAYYRGWHYFKNPN----------LHIPTSKLEIQSLLHLFYVGWLSL
AtCSLC4 LVISIIALSIEIVAHFKKWN-LDLIN----------RPS--WEVYGLVEWSYMAWLSF
AtCSLC6 LVVVVLLLCFELAAYFKGWHFTPPSVAS----------AEVAVEVVYAWWLEI
PtCSLC3 LFLVILLLCFELVAYFKGWHFSPPSVES----------AEAAVERVYAKWLEI
.....180.....190.....200.....210.....220.....230.....240.....250.....

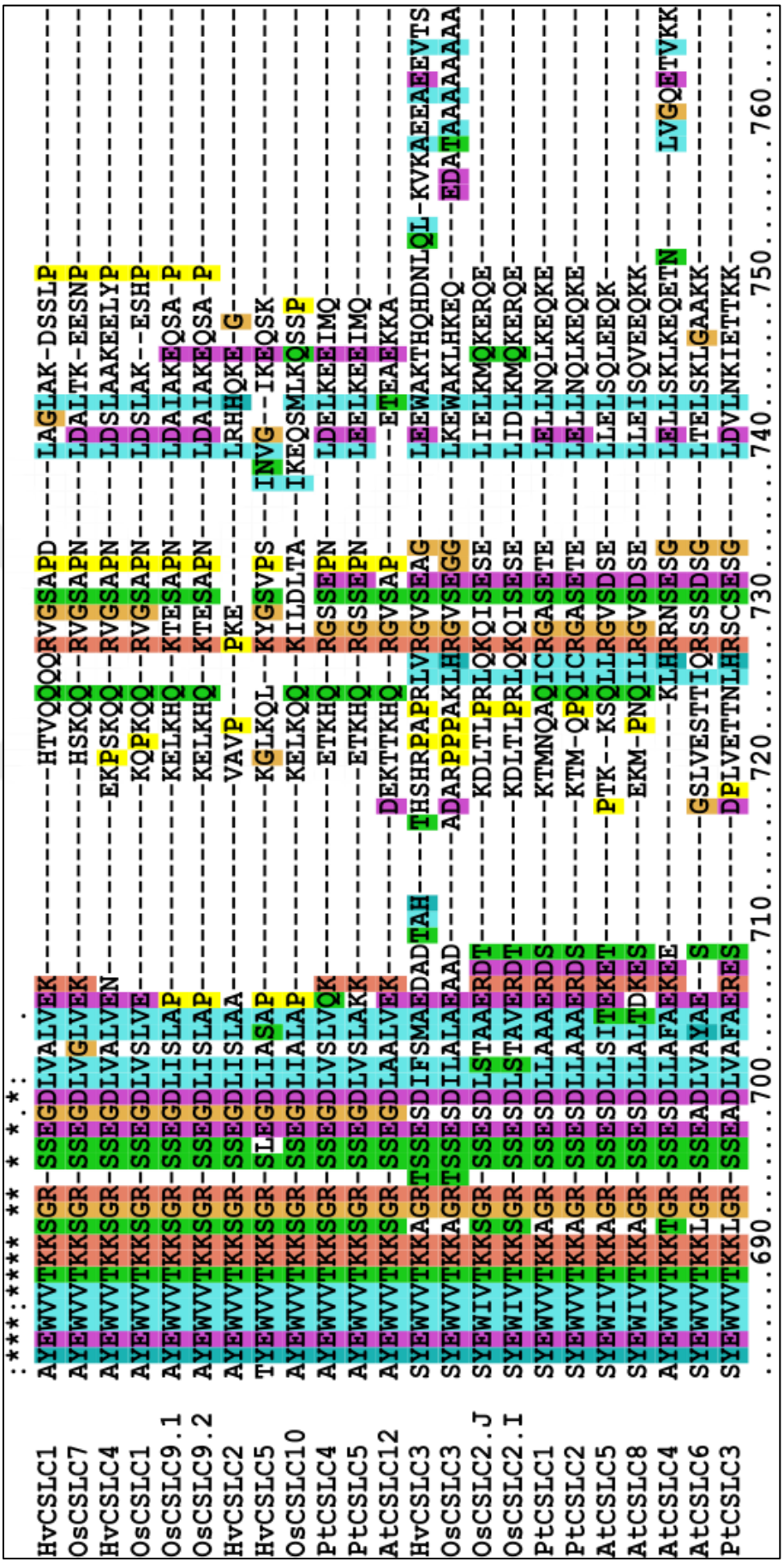

HvCSLC1
OsCSLC7
HvCSLC4
OsCSLC1
OsCSLC9.1
OsCSLC9.2
HvCSLC2
HvCSLC5
OsCSLC10
PtCSLC4
PtCSLC5
AtCSLC12
HvCSLC3
OsCSLC3
OsCSLC2.J
OsCSLC2.I
PtCSLC1
PtCSLC2
AtCSLC5
AtCSLC8
AtCSLC4
AtCSLC6
PtCSLC3
..260.....270.....280.....290.....300.....310.....320.....330.....340


HvCSLC1 NEREVYQQSIGAICNLDWPRSNFLVQVLDDSDDATTSALIKEEVEKWQREGVRIVYRHRVIRDGYKAGNLKSAMNCSYVKDYEYV
OsCSLC7 NEREVYQQSIGAVCNLDWPKSNFLVQVLDDSDDATTSALIKEEVEKWQREGVRIIYRHRVIRDGYKAGNLKSAMNCSYVKDYEFV
HvCSLC4 NEKEVYQQSIGAVCNLDWPRSNFLVQVLDDSDDAATSALIREEVEKWQREGVRILYRHRVIRDGYKAGNLKSAMNCSYVKDYEFV
OsCSLC1 NEKEVYQQSIGAVCNLDWPRSNFLVQVLDDSDDAATSALIKEEVEKWQREGVRILYRHRVIRDGYKAGNLKSAMNCSYVKDYEFV
OsCSLC9.1 NEKEVYQQSIAAVCNLDWPRSNFLVQVLDDSDDPTTQTLIREEVLKWQQNGARIVYRHRVLRDGYKAGNLKSAMSCSYVKDYEFV
OsCSLC9.2 NEKEVYQQSIAAVCNLDWPRSNFLVQVLDDSDDPTTQTLIREEVLKWQQNGARIVYRHRVLRDGYKAGNLKSAMSCSYVKDYEFV
HvCSLC2 NEKEVYQQSIAAVCNLDWPRSNFLVQVLDDSDDPTTQSLIREEVAKWQQTGARILYRHRVLRDGYKAGNLKSAMGCSYVKDYEFV
HvCSLC5 NEKEVYQHSIAAVCNLDWPRSNFLIQVLDDSDDLTTQALIKDEVEKWRHNGARIVYHHRVLREGYKAGNLKSAMSCDYVKDYEYV
OsCSLC10 NEKEVYQQSIAAVCNLDWPRSNILVQVLDDSDDPITQSLIKEEVEKWRQNGARIVYRHRVLREGYKAGNLKSAMSCSYVKDYEYV
PtCSLC4 NEKEVYQQSIAAVCNLDWPKSKFLIQILDDSDDPTTQLLIKEEVHKWQQEGARILYRHRVIRDGYKAGNLKSAMNCSYVKDYEYV
PtCSLC5 NEKEVYQQSIAAVCNLDWPKSKILVQILDDSDDPTTQLLIKEEVNKWQQEGAHILYRHRVIRDGYKAGNLKSAMNCSYVKDYEFV
AtCSLC12 NEKEVYQQSIAAVCNLDWPKGKILIQILDDSDDPITQSLIKEEVHKWQKLGARIVYRHRVNREGYKAGNLKSAMNCSYVKDYEFV
HvCSLC3 NEKEVYETSISHVCQMDWPRDRMLVQVLDDSDDETCQMLIRAEVTKWSQRGVNIIYRHRLSRTGYKAGNLKSAMSCDYVKDYQFV
OsCSLC3 NEKEVYETSISHVCQIDWPRERMLVQVLDDSDDETCQMLIKAEVTKWSQRGVNIIYRHRLNRTGYKAGNLKSAMSCDYVRDYEFV
OsCSLC2.J NEKEVYEQSISAACQLDWPREKFLIQVLDDSSDESIQLLIKAEVSKWSHQGVNIVYRHRVLRTGYKAGNLKSAMSCDYVKDYEFV
OsCSLC2.I NEKEVYEQSISAACQLDWPREKFLIQVLDDSSDESIQLLIKAEVSKWSHQGVNIVYRHRVLRTGYKAGNLKSAMSCDYVKDYEFV
PtCSLC1 NEREVYEQSISAVCQMDWPKDRILIQVLDDSNDESIQWLIKAEVAKWNQKGVNIIYRHRLIRTGYKAGNLKSAMGCDYVKDYEFV
PtCSLC2 NEREVYEQSISAVCQMDWPKDRILIQVLDDSNDESIQWLIKAEVTKWNQKGVNIIYRHRLIRTGYKAGNLKSAMSCDYVKDYDFV
AtCSLC5 NEREVYEQSISAVCQLDWPKDRILVQVLDDSNDESIQQLIKAEVAKWSQKGVNIIYRHRLVRTGYKAGNLKSAMSCDYVEAYEYV
AtCSLC8 NEREVYEQSISAVCQLDWPKDRLLVQVLDDSDDESIQELIRDEVTKWSQKGVNIIYRHRLVRTGYKAGNLKSAMSCDYVEAYEFV
AtCSLC4 NEREVYEQSIGAASQLDWPKDRILIQVLDDSDDPNLQLLIKEEVSVWAEKGVNIIYRHRLIRTGYKAGNLKSAMTCDYVKDYEFV
AtCSLC6 NEKEVYQQSIGAVCMLDWPRERMLVQVLDDSSELDVQQLIKAEVQKWQQRGVRIVYRHRLIRTGYKAGNLKAAMNCEYVKDYEFV
PtCSLC3 NEREVYQQSIAACCVQDWPKERMLIQVLDDSDELDAQLLIKAEVQKWQQRGVHILYRHRLIRTGYKAGNLKSAMSCDYVKDYEFV


"D" motif
.****** * **** ::***: : ::.***:**:***::******** :**.**********. :: **** ********:*::
HvCSLC1 VIFDADFQPQADFLKRAMPHFKGKDDVGLVQARWSFVNNDENLLTRLQNVNLCFHFEVEQQVNGAFLIFFGFIGTAGVWRIKALE
OsCSLC7 VIFDADFQPQADFLKRTVPHFKGKDDVGLVQARWSFVNKDENLLTRLQNVNLCFHFEVEQQVNGAFLNFFGFNGTAGVWRIKALE
HvCSLC4 VIFDADFQPQEDFLKRTVPHFKGKEDVGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGAFLNFFGFNGTAGVWRIKALE
OsCSLC1 VIFDADFQPQADFLKRTVPHFKGNEDVGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIKALE
OsCSLC9.1 AIFDADFQPNPDFLKRTVPHFKDNDELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGIFLNFFGFNGTAGVWRIKALD
OsCSLC9.2 AIFDADFQPNPDFLKRTVPHFKDNDELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGIFLNFFGFNGTAGVWRIKALD
HvCSLC2 AIFDADFQPNPDFLKRTVPHFKDNDELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIKALE
HvCSLC5 AIFDADFQPYPDFLKRTVPHFKDNEDLGLVQARWSFVNKDENLLTRLQYINLCFHFEVEQQVNGVFINFFGFNGTAGVWRIKAVE
OsCSLC10 AIFDADFQPYPDFLKRTVPHFKDNEELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGIFINFFGFNGTAGVWRIKALE
PtCSLC4 AIFDADFQPTPDFLKKTVPHFKDNEEIGLVQARWSFVNKDENLLTRLQNINLAFHFEVEQQVNGTFINFFGFNGTAGVWRIKALE
PtCSLC5 AIFDADFQPTPDFLKRTVPHFKGNEELGLVQARWSFVNKDENLLTRLQNINLAFHFEVEQQVNGIFINFFGFNGTAGVWRIKALE
AtCSLC12 AIFDADFQPLPDFLKKTIPHFKDNEEIGLVQARWSFVNKEENLLTRLQNINLAFHFEVEQQVNSVFLNFFGFNGTAGVWRIKALE
HvCSLC3 AIFDADFQPNPDFLKLTVPHFKGNPDLGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGIYLNFFGFNGTAGVWRIEALE
OsCSLC3 AIFDADFQPNPDFLKLTVPHFKGNPELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVYLSFFGFNGTAGVWRIKALE
OsCSLC2.J AIFDADFQPTPDFLKKTIPHFEGNPELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIQALE
OsCSLC2.I AIFDADFQPTPDFLKKTIPHFEGNPELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIQALE
PtCSLC1 TIFDADFQPNPDFLKLTVPHFKGNPELGLVQARWAFVNKDENLLTRLQNINLCFHFEVEQQVNGTFLNFFGFNGTAGVWRIKALE
PtCSLC2 AIFDADFQPNPDFLKLTVPHFKNNPELGLVQARWAFVNKDENLLTRLQNINLCFHFEVEQQVNGAFLNFFGFNGTAGVWRIKALE
AtCSLC5 AIFDADFQPTPDFLKLTVPHFKDNPELGLVQARWTFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIKALE
AtCSLC8 AIFDADFQPNSDFLKLTVPHFKEKPELGLVQARWAFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIKALE
AtCSLC4 TIFDADFTPNPDFLKKTVPHFKGNPELGLVQARWSFVNKDENLLTRLQNINLCFHFEVEQQVNGVFLNFFGFNGTAGVWRIKALE
AtCSLC6 AIFDADFQPPADFLKKTVPHFKGNEELALVQTRWAFVNKDENLLTRLQNINLSFHFEVEQQVNGVFINFFGFNGTAGVWRIKALE
PtCSLC3 AIFDADFQPGPDFLKKTIPHFKGKDDLALVQTRWAFVNKDENLLTRLQNINLSFHFEVEQQVNGVFINFFGFNGTAGVWRIKALE
..430.....440.....450.....460.....470.....480.....490.....500.....510


"D" motif
"QXXRW" motif
HvCSLC1
OsCSLC7
HvCSLC4
OsCSLC1
OsCSLC9.1
OsCSLC9.2
HvCSLC2
HvCSLC5
OsCSLC10
PtCSLC4
PtCSLC5
AtCSLC12
HvCSLC3
OsCSLC3
OsCSLC2.J
OsCSLC2.I
PtCSLC1
PtCSLC2
AtCSLC5
AtCSLC8
AtCSLC4
AtCSLC6
PtCSLC3
520
530
540
550
560
570
580
590


HvCSLC1
OsCSLC7
HvCSLC4
OsCSLC1
OsCSLC9.1
OsCSLC9.2
HvCSLC2
HvCSLC5
OsCSLC10
PtCSLC4
PtCSLC5
AtCSLC12
HvCSLC3
OsCSLC3
OsCSLC2.J
OsCSLC2.I
PtCSLC1
PtCSLC2
AtCSLC5
AtCSLC8
AtCSLC4
AtCSLC6
PtCSLC3


HvCSLC1
OsCSLC7
HvCSLC4
OsCSLC1
OsCSLC9.1
OsCSLC9.2
HvCSLC2
HvCSLC5
OsCSLC10
PtCSLC4
PtCSLC5
AtCSLC12
HvCSLC3
OsCSLC3
OsCSLC2.J
OsCSLC2.I
PtCSLC1
PtCSLC2
AtCSLC5
AtCSLC8
AtCSLC4
AtCSLC6
PtCSLC3

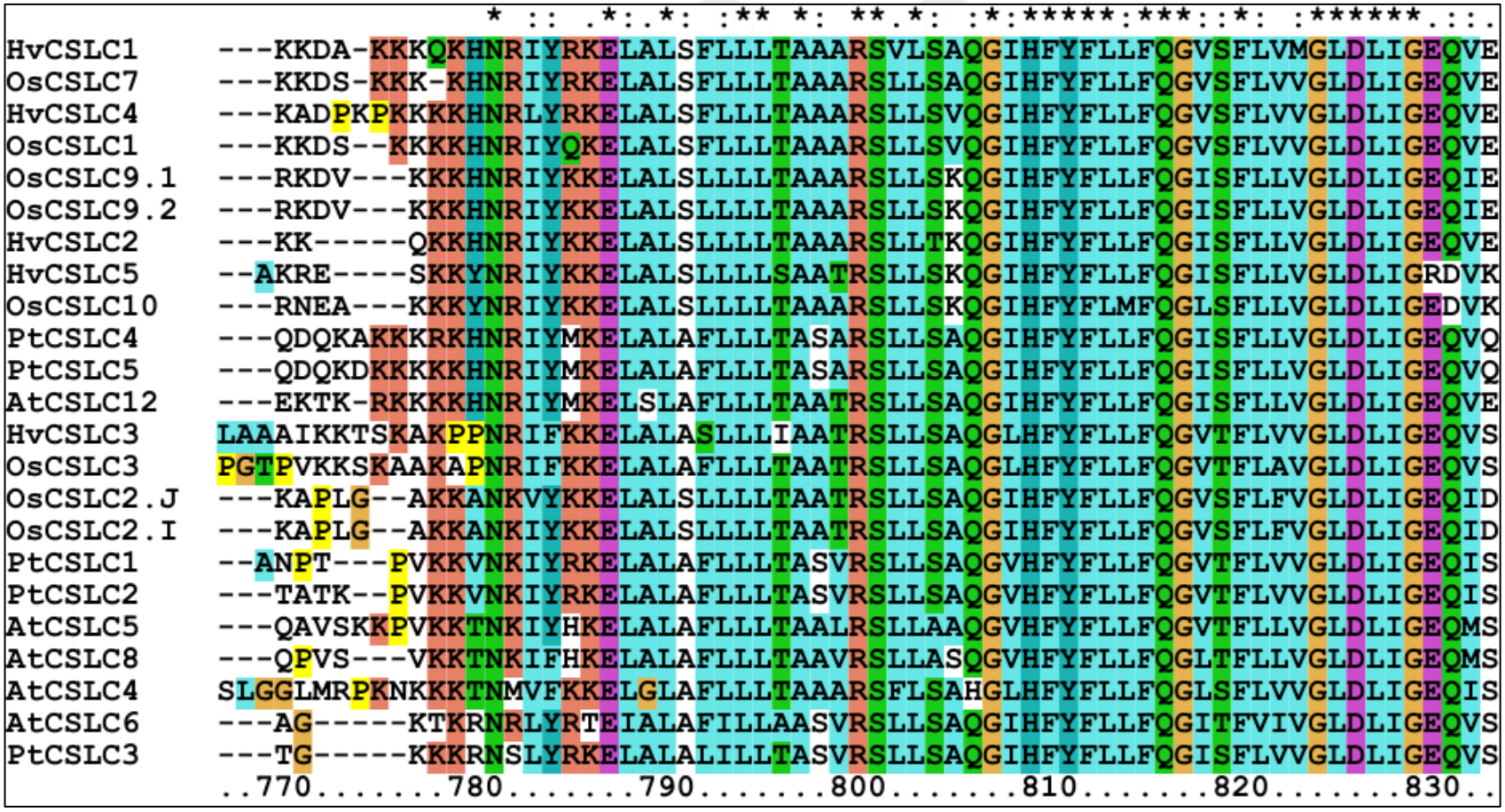

HvCSLC1 ---KKDA-KKKQKHNRIYRKELALSFLLLTAAARSVLSAQGIHFYFLLFQGVSFLVMGLDLIGEQVE
OsCSLC7 ---KKDS-KKK-KHNRIYRKELALSFLLLTAAARSLLSAQGIHFYFLLFQGVSFLVVGLDLIGEQVE
HvCSLC4 ---KADPKPKKKKHNRLYRKELALSFLLLTAAARSLLSVQGIHFYFLLFQGVSFLVVGLDLIGEQVE
OsCSLC1 ---KKDS--KKKKHNRIYQKELALSFLLLTAAARSLLSVQGIHFYFLLFQGVSFLVVGLDLIGEQVE
OsCSLC9.1 ---RKDV---KKKHNRIYKKELALSLLLLTAAARSLLSKQGIHFYFLLFQGISFLLVGLDLIGEQIE
OsCSLC9.2 ---RKDV---KKKHNRIYKKELALSLLLLTAAARSLLSKQGIHFYFLLFQGISFLLVGLDLIGEQIE
HvCSLC2 ---KK-----QKKHNRIYKKELALSLLLLTAAARSLLTKQGIHFYFLLFQGISFLLVGLDLIGEQVE
HvCSLC5 --AKRE----SKKYNRIYKKELALSLLLLSAATRSLLSKQGIHFYFLLFQGISFLLVGLDLIGRDVK
OsCSLC10 ---RNEA---KKKYNRIYKKELALSLLLLTAAARSLLSKQGIHFYFLMFQGLSFLLVGLDLIGEDVK
PtCSLC4 ---QDQKAKKKRKHNRIYMKELALAFLLLTASARSLLSAQGIHFYFLLFQGISFLLVGLDLIGEQVQ
PtCSLC5 ---QDQKDKKKKKHNRIYMKELALAFLLLTASARSLLSAQGIHFYFLLFQGISFLLVGLDLIGEQVQ
AtCSLC12 ---EKTK-RKKKKHNRIYMKELSLAFLLLTAATRSLLSAQGIHFYFLLFQGISFLLVGLDLIGEQVE
HvCSLC3 LAAAIKKTSKAKPPNRIFKKELALASLLLIAATRSLLSAQGLHFYFLLFQGVTFLVVGLDLIGEQVS
OsCSLC3 PGTPVKKSKAAKAPNRIFKKELALAFLLLTAATRSLLSAQGLHFYFLLFQGVTFLAVGLDLIGEQVS
OsCSLC2.J ---KAPLG--AKKANKVYKKELALSLLLLTAATRSLLSAQGIHFYFLLFQGVSFLFVGLDLIGEQID
OsCSLC2.I ---KAPLG--AKKANKIYKKELALSLLLLTAATRSLLSAQGIHFYFLLFQGVSFLFVGLDLIGEQID
PtCSLC1 --ANPT---PVKKVNKIYRKELALAFLLLTASVRSLLSAQGVHFYFLLFQGVTFLVVGLDLIGEQIS
PtCSLC2 ---TATK--PVKKVNKIYRKELALAFLLLTASVRSLLSAQGVHFYFLLFQGVTFLVVGLDLIGEQIS
AtCSLC5 ---QAVSKKPVKKTNKIYHKELALAFLLLTAALRSLLAAQGVHFYFLLFQGVTFLLVGLDLIGEQMS
AtCSLC8 ---QPVS---VKKTNKIFHKELALAFLLLTAAVRSLLASQGVHFYFLLFQGLTFLLVGLDLIGEQMS
AtCSLC4 SLGGLMRPKNKKKTNMVFKKELGLAFLLLTAAARSFLSAHGLHFYFLLFQGLSFLVVGLDLIGEQIS
AtCSLC6 ---AG-----KTKRNRLYRTEIALAFILLAASVRSLLSAQGIHFYFLLFQGITFVIVGLDLIGEQVS
PtCSLC3 ---TG-----KKKRNSLYRKELALALILLTASVRSLLSAQGIHFYFLLFQGISFLVVGLDLIGEQVS
..770.....780.....790.....800.....810.....820.....830..

Figure 3.2. Multiple sequence alignment results of Cellulose synthase-like C proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*

Alignment results show that between these species, there are many similar regions or conserved sequences were detected. Especially in CSLC protein family, even the gap and residue regions were highly matched and fitted. This situation can clearly be observed in Figure 3.2. Moreover, these similarities can be detected in the different plant classes (monocotyledons/dicotyledon). At last, highly conserved motifs that researched in recent study [79] were detected and identified among all sequences, marked on figures.

3.1.3. Phylogenetic Analyses

Aligned protein sequences were further investigated to make connections in phylogenetic ways to see and understand genetic relativities. Four different phylogenetic trees were made and bootstrapped using Clusxtal X software, and visualized using with Dendroscope software (version 3.7.5) [75]. Phylogenetic trees were edited for clarity in cladogram form. Illustrated trees show genetic relatedness and significant difference in between monocotyledon and dicotyledon plant species. Moreover, bootstrap values were recorded and listed on trees, out of 1000.

In Figure 3.3., there is a difference in between monocot and dicot plant species. However, all of the bootstrap values are above 500. At the end of root line, monocots were coupled with another monocot, and dicot species' proteins are always making branching with dicot ones. In other words, barley CSLCs are closely related the each other as well as to another monocot species, the rice sequences. Arabidopsis CSLCs show much more relatedness to the black cottonwood sequences, which is another eudicot.

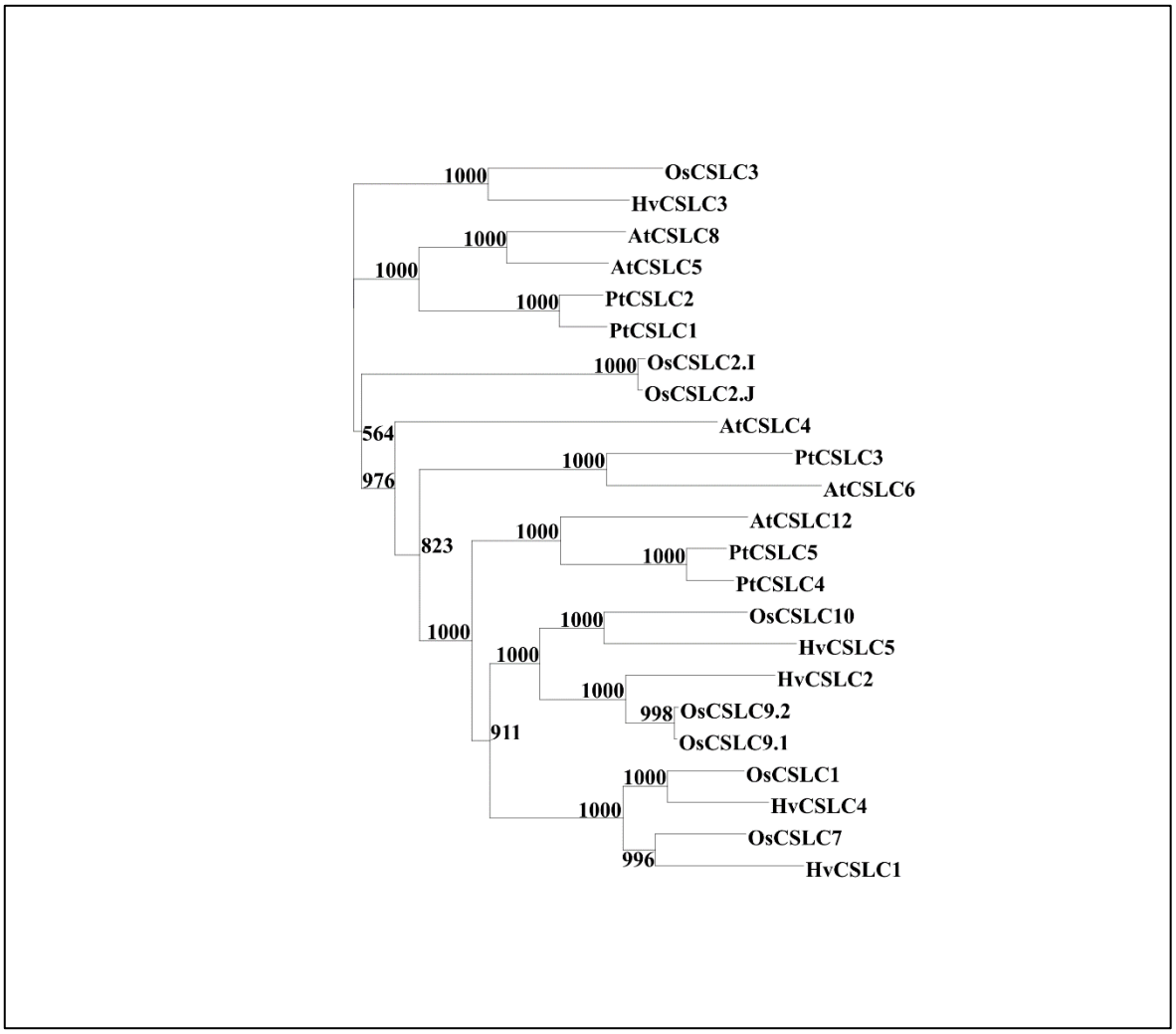


Figure 3.3. Phylogenetic tree of aligned Cellulose synthase-like C proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa* with bootstrap values

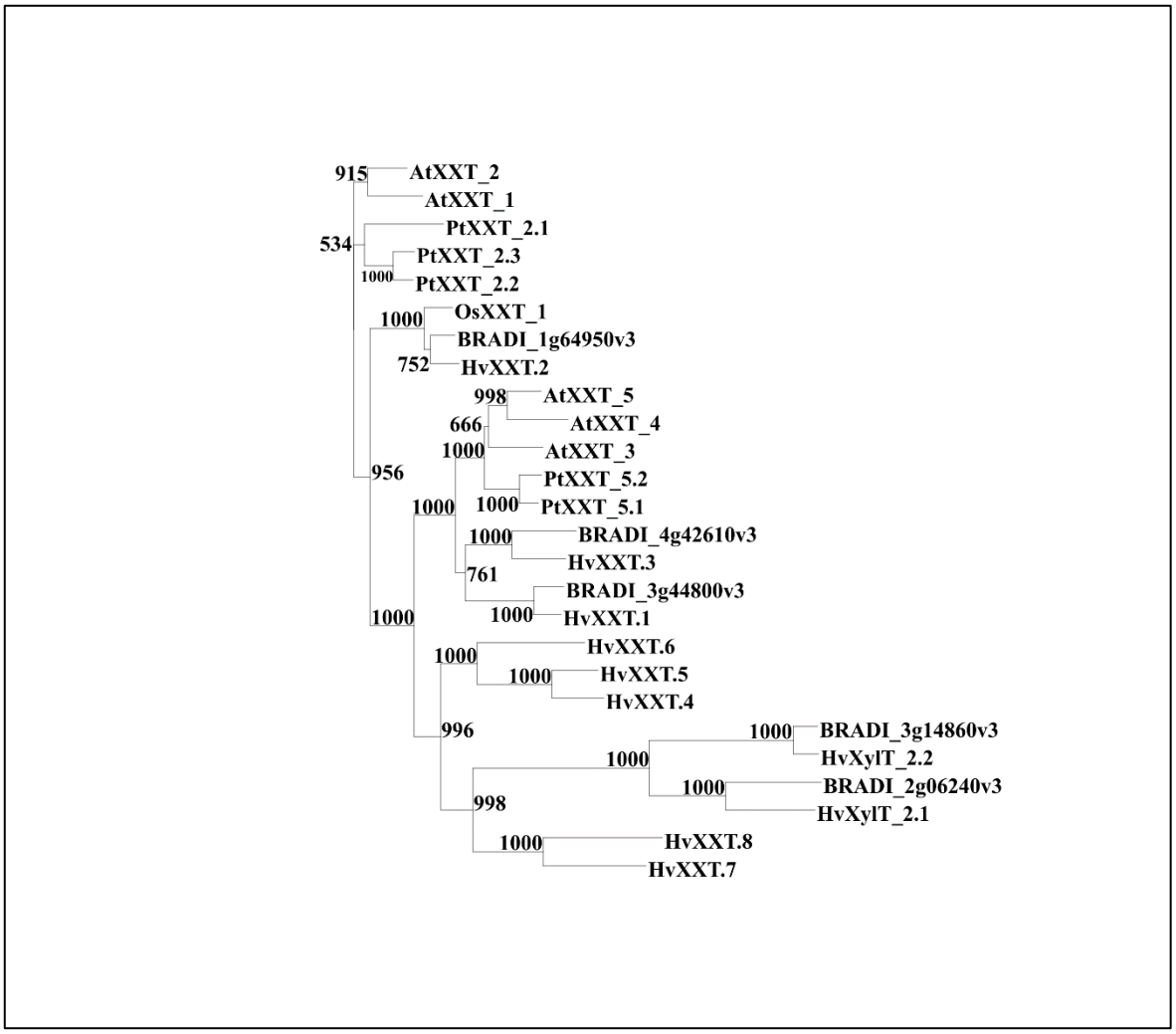


Figure 3.4. Phylogenetic tree of aligned xylosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa,* and *Brachypodium distachyon* with bootstrap values

Just like the previous phylogenetic tree, this Figure 3.4. tree shows that at the end of the branches, monocot sequences are always similar with other monocots. In this situation, *Brachypodium distachyon* xylosyl transferases were always paired with and therefore very similar to barley xylosyl transferases.

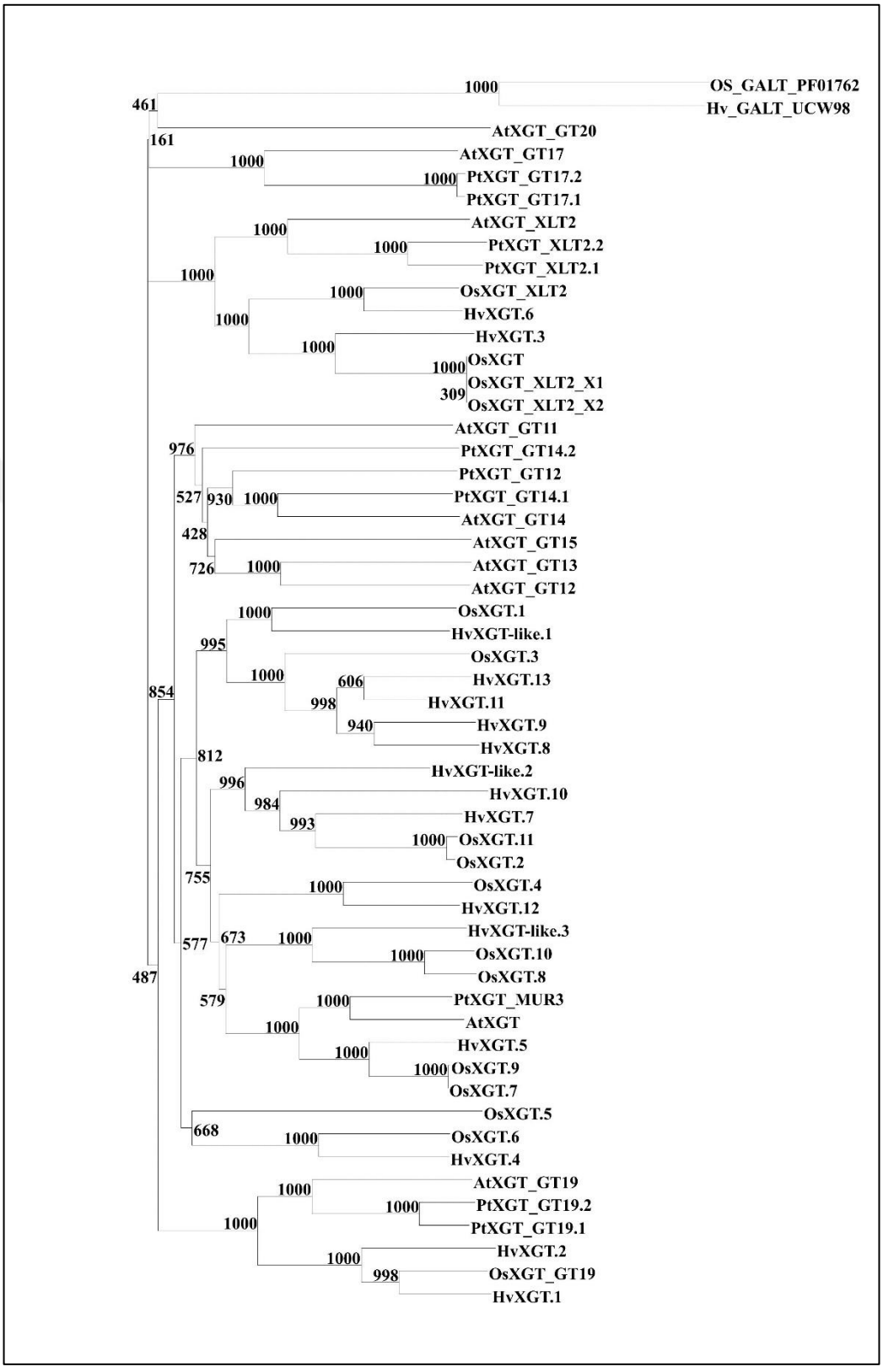


Figure 3.5. Phylogenetic tree of aligned galactosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, with bootstrap values

In phylogenetic tree for galactosyl transferases (Figure 3.5.), monocot plant genes and eudicot plant genes were diverted on this figure. On the other hand, at the bottom side of the tree, there are some proteins that showed very close relations, looking for their bootstrap results. These are HvXGT.1 and 2 (xyloglucan galactosyl transferases from barley, with variations), PtXGT_GT19.1 and 2 (xyloglucan galactosyl transferases from black cottonwood), OsXGT_GT19 (xyloglucan galactosyl transferases from rice), and AtXGT_GT19 (xyloglucan galactosyl transferases from *Arabidopsis thaliana*). Even the class difference, they located next to the each other on the tree, so GT19 family might has evolutionary conserved regions among species. In this way, they might take a role in xyloglucan synthesis all of these species.

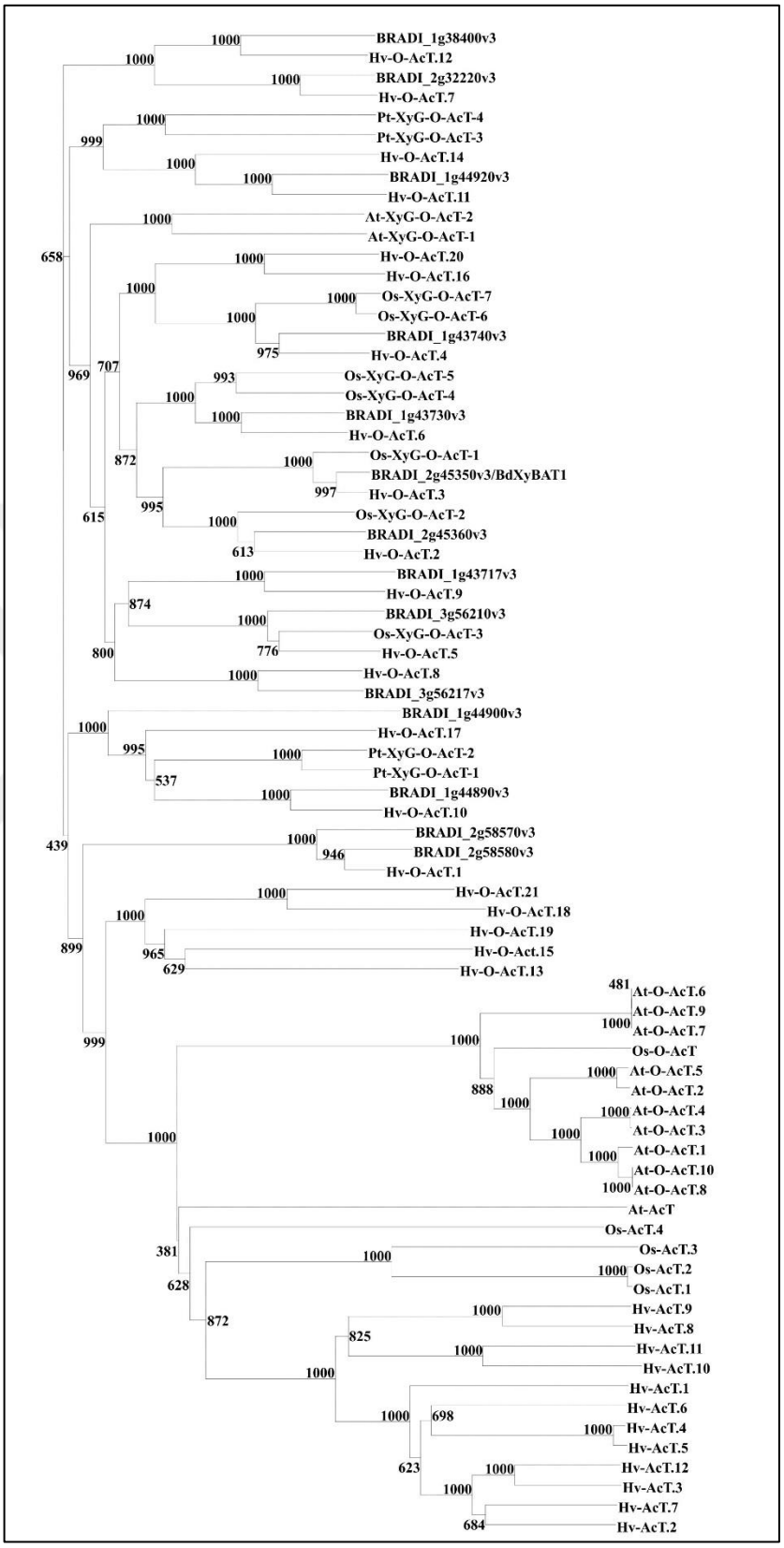


Figure 3.6. Phylogenetic tree of aligned acetyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon* with bootstrap values

The last phylogenetic tree for acetyl transferases (Figure 3.6.), a similar situation can be observed between the proteins. In this time, especially barley's and *Brachypodium* xyloglucan o-acetyltransferases, they are highly matched on the Figure 3.9. The spectacular result is coming from the pre-worked protein, named "BdXyBAT1" on the tree. The letter to Molecular Plant journal [80] explains that this specific 6-O-Acetyltransferase modulates xyloglucan backbone formation and even also xylosylation, though. This protein shows highly similarity with Hv-O-AcT.3 (xyloglucan o-acetyltransferase, with a variation).

3.1.4. Gene Expression and Pearson Correlation Analyses

After the alignments and phylogenetic trees were made and investigated, with the information of intron/exon regions of targeted barley genes (total 76 genes), expression levels were checked and investigated with RNA-seq data from on barley cultivar Morex, was obtained from the James Hutton Institute, Dundee, Scotland. Expression scores were noted and received as FPKM value scores, with Cufflinks software (version 2.2.1) [76]. For every tissue, there were 3 technical replicates from 2 different biological batches, so every tissue has 6 total samples. After calculating FPKM values, they were averaged tissue by tissue, and gene by gene, then results were tabulated. In addition, there were 3 barley genes were added for examination of expression levels as reference, which are Actin, Tubulin alpha-2 chain, and ADP-Ribose. They were listed and tabled below in Table 3.7., 3.8., 3.9., 3.10., 3.11, and 3.12. In total, 79 barley genes were investigated and recorded their FPKM values recorded to examine their expression patterns. In the table (Table 3.6.), the order of the genes is according to their chromosomal locations.

Table 3.6. Reference barley genes for expression values with their transcript ID from ENSEMBL Plants, UniProt entry names, and locations on barley chromosomes

Gene	Locus	UniProt	Location	Strand
HvACTIN	HORVU1Hr1G074350.1	M0X0A5	Chr1H: 507686919-507690316	Reverse
HvTUB	HORVU1Hr1G081280.1	Q96460	Chr1H: 527410688-527413835	Forward
HvADP-Rib	HORVU3Hr1G079700.1	A0A287LUS2	Chr3H: 584539968-584543097	Forward

Table 3.7. FPKM values of CSLC barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	CAR5	CAR15	EMB	EPI	ETI	INF1	INF2	LEA
HvCSLC1	11,5827833	0,5225415	19,3921	2,89806	6,438622	7,966965	5,18225333	6,6209483
HvCSLC2	19,8682333	1,21668217	10,9661833	27,2238	20,46168	41,49385	23,32985	7,1792417
HvCSLC3	0	0,282804	5,92175	0	0	0,0071421	0,00439488	0
HvCSLC4	25,3050333	0,79699667	28,7846833	6,00588	11,88626	13,27635	7,85886333	10,07095
HvCSLC5	0	0,00386182	0,02024193	0,03137	0,065519	0	0,01633583	0,0113456

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
HvCSLC1	12,1800833	56,5917	10,414495	8,99744	14,74898	5,975955	33,3640667	1,79633
HvCSLC2	52,3723667	62,5415833	5,58814333	46,8035	57,90227	7,467555	38,0483833	5,1465667
HvCSLC3	1,27665133	0,03331387	0,00514657	0,35074	0,004725	3,568135	20,0435667	0
HvCSLC4	17,7936667	55,82945	19,934	35,4719	58,4573	6,440785	38,0312167	1,0481142
HvCSLC5	0,09715493	0,00338197	0,00360147	1,23381	0	0,0037294	0,0084587	0,0049763

Table 3.8. FPKM values of xylosyl transferase barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	CAR5	CAR15	EMB	EPI	ETI	INF1	INF2	LEA
HvXXT.1	17,2615333	3,22772833	14,70093333	3,63917	6,04677	9,37778	7,97457667	5,562743333
HvXXT.2	18,0372	1,76264667	22,74398333	8,143332	6,623665	18,4933	8,40439	6,683068333
HvXXT.3	0,04123472	0,335016	10,22028833	0,035913	0	0,0062	0	0
HvXXT.4	0,07114138	0,04238092	0,806579833	0,153077	0,15518	0,01592	0,03055293	0,36501405
HvXXT.5	0,06531747	0	0,245328417	0,942535	1,961361	0,01989	0,02687168	0,819023667
HvXXT.6	0	0,06061122	2,001138333	0	0	0,04739	0,01719963	0,013419517
HvXXT.7	11,0659967	3,32462833	0,506164333	0,012365	0,055537	2,12142	1,27758267	0,904415333
HvXXT.8	18,6317667	5,613235	1,891241667	0,704006	0,74467	1,8507	0,8207605	0,384365917
HvXylT_2.1	0	0	0,034310133	0	0	0	0	0
HvXylT_2.2	7,42913667	1,54498	6,488401667	5,240648	7,711817	3,85499	4,94150333	2,942961667

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
HvXXT.1	6,50908167	13,8243	13,78721667	8,436195	11,7986	5,66544	11,3582167	3,70738
HvXXT.2	32,1783	51,0779167	19,9520333	22,21695	62,05865	8,66113	87,96508333	2,214616667
HvXXT.3	0,05979945	0,03338573	0,024259517	0,079032	0,00867	3,73577	13,3630833	0
HvXXT.4	1,348325	0,257962	1,980829333	0,645876	1,147925	3,99575	11,4182667	0,147828433
HvXXT.5	12,4991167	0,4315255	80,30293333	21,48605	10,22498	0,57667	0,15964383	0,012000783
HvXXT.6	0,04383193	0,01012845	0	0,007981	0,021413	0,67665	2,09739167	0,008874483
HvXXT.7	0,530365	0,56700683	0,06724755	2,455498	1,701392	0,42072	0,28637025	0
HvXXT.8	3,31850333	20,6437667	2,148313667	2,96952	0,629176	1,18094	3,37057	0,0308846
HvXylT_2.1	0	0	0	0	0	0,0227	0,02547933	0
HvXylT_2.2	7,935105	8,35836167	10,29208	10,24295	8,088662	2,59312	7,80228	1,946275

Table 3.9. FPKM values of galactosyl transferase barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	CAR5	CAR15	EMB	EPI	ETI	INF1	INF2	LEA
HvXGT.1	3,82264167	0,45983083	2,94529667	0,707359	1,0955928	5,04183833	3,52215	0,83838733
HvXGT.2	0	0	0	0	0	0	0	0
HvXGT.3	0,099943	0,0152136	0,65223417	0	0,0261387	0,04137567	0,08725	0
HvXGT.4	0,420307	0,01939333	0,01205057	0,017123	0,0387678	0	0	0,01042995
HvXGT.5	41,7829333	2,50853	7,00635833	2,45812	3,1479167	7,19985667	8,05749	2,10424833
HvXGT.6	0	0	0	0	0	0	0	0
HvXGT.7	0,13532618	0,05497657	1,21434133	0	0	0	0	0
HvXGT.8	0	0,04743855	0,32861533	0	0	0	0	0
HvXGT.9	0,27490067	0,04515678	0,369651	0	0	0	0	0
HvXGT.10	0,14447807	0,07004783	1,5341715	0	0,0686953	0,01450258	0,31322	0,0165586
HvXGT.11	0,23526642	0	0,0199565	0	0	0	0	0
HvXGT.12	0,00750168	0,09953333	0,69665717	0,004423	0	0	0	0
HvXGT.13	0,1578945	0,03582658	0,0388327	0,03639	0,0706143	0,12164467	0,15587	0,0966393
HvXGT-like.1	3,08590333	0,0213654	1,16477217	0,446113	0,4162355	1,033454	0,40817	0,76873083
HvXGT-like.2	0	0	0,72684983	0	0	0	0	0
HvXGT-like.3	0,11358825	0,05556687	1,40085833	0	0,098704	0,1697114	0,11226	0,09584562
Hv_GALT_UCW98	2,48355	0,64100567	2,74875667	1,935142	1,3721867	1,79418	1,40098	0,686275

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
HvXGT.1	1,646545	5,10276833	6,12264	1,811172	4,4879267	1,58097267	2,93185	0,663497
HvXGT.2	0,01652802	0	0	0,159144	0	0	0,03009	0
HvXGT.3	0,0811035	3,1757	0,1323593	0,026971	0,8552958	0,38284467	1,01886	0
HvXGT.4	0,18775783	0,03354905	0,18300248	0,244397	0,0480518	0,01124212	0,03095	0,02128355
HvXGT.5	10,8984083	56,9759167	0,89580133	6,026855	16,86605	3,12113833	8,32613	0,30494167
HvXGT.6	0	0	0	0	0	0	0	0
HvXGT.7	0	0	0	0	0,0048954	0,947234	4,31595	0
HvXGT.8	0	0	0	0	0	0,1145767	0,67571	0
HvXGT.9	0	0	0	0	0	0,396196	1,32885	0
HvXGT.10	0	0,0097115	0	0	0	1,125944	4,49037	0
HvXGT.11	0,00657687	0	0	0	0,0082871	0	0,22169	0
HvXGT.12	0,0036181	0,02180412	0	0	0,0043542	0,87613383	2,02779	0
HvXGT.13	0,019482	0,00811947	0,07586248	0,047704	0,0562706	0	0,02795	0,0392785
HvXGT-like.1	0,20311655	9,45138667	0,31461217	0,129433	0,3711873	0,6874665	1,88527	2,40340833
HvXGT-like.2	0	0	0	0	0	0,19837178	0,4198	0
HvXGT-like.3	0,00401335	0,11636363	0,09254432	0	0,0509115	0,99702783	2,71002	0,585396
Hv_GALT_UCW98	1,71360167	3,576075	2,70338833	3,18352	3,2989567	0,6601935	4,4066	1,66767333

Table 3.10. FPKM values of acetyl transferase barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	CAR5	CAR15	EMB	EPI	ETI	INF1	INF2	LEA
Hv-AcT.1	2,22631333	0,09922952	1,548923	0,4036415	0,94917267	1,818472	2,35877	0,53274067
Hv-AcT.2	0	0,01215902	0,02753503	0	0	0	0	0,01044823
Hv-AcT.3	0,01616083	0,51225633	73,1082	0,00485523	0,01015642	0,025554	0,0059117	0,00844745
Hv-AcT.4	0	0	8,510065	0	0,02238303	0	0,01993473	0
Hv-AcT.5	0	0	4,61615	0	0	0	0	0
Hv-AcT.6	0,1791005	0,12713003	0,96387717	0,11012602	0,12520613	0,054793	0,04865413	0,23429967
Hv-AcT.7	2,280235	0,08954065	0,1898975	0,05367315	0,11887022	0,010004	0,02943663	0,16444817
Hv-AcT.8	2,45923667	1,033642	12,8205667	12,3592017	2,23229	0,307117	0,46741233	3,0839
Hv-AcT.9	5,57525667	1,15502	7,56574167	59,6511333	43,6635833	7,727807	10,3395267	6,72517333
Hv-AcT.10	0,11088783	0,01305198	0,56785517	1,09881883	0,15073653	0,173657	0,11090167	0,709542
Hv-AcT.11	0,11683668	0,0252411	1,37582333	0,22820972	1,72297667	0,310973	0,02612342	0,87583067
Hv-AcT.12	0	0	0,09373148	0	0	0	0	0
Hv-O-AcT.1	13,9788833	2,22486333	9,70005	6,82251167	6,078265	7,903035	8,08373667	3,65799667
Hv-O-AcT.2	0	0,03433752	0,295774	0	0	0	0	0
Hv-O-AcT.3	94,6692833	3,48055167	90,2036833	10,9833133	20,4071833	37,8381	29,6591167	33,2835833
Hv-O-AcT.4	0,02755183	0,0774695	0,028634	0	0,02671887	0	0	0
Hv-O-AcT.5	0	0	3,04244167	0	0	0	0	0,063581
Hv-O-AcT.6	1,97548167	0,42696412	5,37473333	0,00591962	0,05897562	13,55262	11,1811083	0,631813
Hv-O-AcT.7	20,2646667	4,29919833	10,74468	8,665295	14,3482067	27,9585	34,8281333	7,91909333
Hv-O-AcT.8	0,11699108	0,36211133	1,50619833	0,31787283	0,19477167	0,176217	0,04612818	1,1901415
Hv-O-AcT.9	0	0	1,40119467	0	0,07571165	0,073512	0,03352017	0,00702013
Hv-O-AcT.10	3,49764833	0,51200267	5,038	0,4301555	1,06129533	9,54194	9,3788	0,992736
Hv-O-AcT.11	0,08092362	0,04076592	5,94738333	0,0049947	1,95331783	0,176456	0,1112223	0,983081
Hv-O-AcT.12	16,4394167	2,45676	11,7777983	2,35605167	3,35909833	7,272617	2,19195167	2,32539667
Hv-O-AcT.13	0,73586083	0,03240112	1,01658217	0,03388817	0,60559	0,210863	0,11236163	0,07193567
Hv-O-AcT.14	0,59139617	0,68073567	0	0	0	0	0	0
Hv-O-AcT.15	10,271485	0,79572333	4,16837833	2,992	3,78631667	7,241722	6,272925	1,57775867
Hv-O-AcT.16	0,01148082	0	0,2198686	0	0,04373295	0	0	0,00975753
Hv-O-AcT.17	0,42036485	0,06116483	0,7899875	0,01013842	0,02213365	0,354552	0,11062688	0,06659293
Hv-O-AcT.18	22,0291167	7,02021	24,1816333	9,43211	10,6004867	13,944	10,4736233	9,07193833
Hv-O-AcT.19	6,020345	0,598896	4,56092167	8,76048	4,94442	5,788357	6,48939667	4,10459
Hv-O-AcT.20	0	0	0	0	0	0	0	0,02890333
Hv-O-AcT.21	1,69680667	1,936915	14,4353333	3,43033167	3,29017333	0,040826	0,02046493	7,79410167

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
Hv-AcT.1	0,06247535	0,37972483	0,24272127	0,15325017	1,93375667	0,792256	0,76620283	0,96302483
Hv-AcT.2	0	0	0,00320957	0	0	0	0	0
Hv-AcT.3	0	0	0,0157713	0	0,00516835	7,317882	0,14519317	0
Hv-AcT.4	0	0	0,00817432	0	0	0,359805	1,05798833	0
Hv-AcT.5	0	0	0	0	0	0,093669	0,13251963	0
Hv-AcT.6	0,05389718	0,33258317	0,17023217	0,0171921	0,0506655	1,041751	0,950893	0,83424433
Hv-AcT.7	0,52553517	0,4392555	0,60516267	0,4614595	0,20904683	0,185573	0,02088	0,00900673

Hv-AcT.8	5,98840333	1,63980333	0,3975325	4,00034	15,7872167	5,96437	15,92175	0,62454867
Hv-AcT.9	44,7370833	17,51275	76,82125	22,1691	33,1199833	2,18695	3,55161667	2,36625167
Hv-AcT.10	0,72360033	0,34109	0,06426003	0,53671867	1,05824367	0,447779	13,4502167	0,79613
Hv-AcT.11	0,34939133	0,01611948	0	0,23658617	0,03468892	1,638237	5,98682333	1,6159945
Hv-AcT.12	0	0	0	0	0	0	0	0
Hv-O-AcT.1	7,68119333	19,0920167	9,23115333	12,2693	17,0423667	4,147837	12,2259667	5,14374667
Hv-O-AcT.2	0	0	0	0	0	0,56208	1,80618833	0
Hv-O-AcT.3	19,56195	38,5779833	73,273	15,3959167	79,8164667	23,63587	28,9832833	9,94062833
Hv-O-AcT.4	0,03813032	0	0	0,02790643	0	0	0,01921797	0,00961732
Hv-O-AcT.5	0	0	0,01516012	0	0,06745148	0,558689	1,32724	0
Hv-O-AcT.6	0,1157275	0,35177008	0,38783193	0,40555683	10,997795	2,701652	3,2231	0
Hv-O-AcT.7	9,84364667	14,0105467	26,1972667	8,66142	17,80805	7,299847	7,31116667	12,98815
Hv-O-AcT.8	0,50016483	0,0063704	2,52115333	0,19428967	0,50879467	0,298558	0,78051717	0,20566617
Hv-O-AcT.9	0	0,02667978	0,01151457	0	0	0,511978	1,97946	0
Hv-O-AcT.10	1,02187133	2,35503167	3,29694	2,06453667	4,52741833	1,765087	1,94825167	0,24707467
Hv-O-AcT.11	37,8214333	4,10695	73,9908	51,6787	28,57145	4,987413	4,59204667	0,01227478
Hv-O-AcT.12	3,24334833	33,6539167	5,33276667	4,065025	21,4527167	5,51362	24,14755	2,16127167
Hv-O-AcT.13	0,19444728	0,61916532	0,68440165	0,45730617	4,82375333	0,649341	1,92641167	0
Hv-O-AcT.14	0	0	0	0	0	0	0	0
Hv-O-AcT.15	6,97597667	17,8707	8,384975	6,50397833	12,7294217	1,135329	5,72092667	0,16056667
Hv-O-AcT.16	0	0	51,7329	0	0,00736093	0,941366	0,25680817	0
Hv-O-AcT.17	0,1335374	1,15594283	0,29503892	0,48522783	1,22282333	0,855053	3,05794667	0,00671628
Hv-O-AcT.18	71,3048667	35,2282	67,2335	110,289883	39,6652667	12,83655	51,6729333	7,32732333
Hv-O-AcT.19	2,98996167	4,21403333	4,62596333	3,04633167	6,64986333	2,841737	4,11900333	3,71507167
Hv-O-AcT.20	0	0	0	0	0	0	0	0,01316133
Hv-O-AcT.21	7,86908667	2,31259167	0,1403315	8,83483833	0,3719385	15,35702	17,5118667	4,75386333

Table 3.11. FPKM values of fucosyl transferase barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	**CAR5**	**CAR15**	**EMB**	**EPI**	**ETI**	**INF1**	**INF2**	**LEA**
HvFUT.1	0,611117167	0,3095985	0,58272567	47,1752167	10,3450217	0,15412245	0,06929453	8,914285
HvFUT.2	0,235562667	0	0,7883715	0	0,00447463	0	0	0
HvFUT.3	0	0	0,03757363	0	0,2246677	0	0	0,13647695
HvFUT.4	0	0	0,02374867	0	0	0	0,0063362	0
HvFUT.5	0,057683517	0,00822823	0,40894667	0	0	0	0	0
HvFUT.6	0,049168733	0	0	0	0	0	0	0
HvFUT.7	0,043483167	0,04801417	1,330882	0,01572463	0,04375742	0,21265595	0,00830817	0,17997425
HvFUT.8	0,0392094	0,02465777	0,14744123	0,04449525	0,00646042	0,14298137	0	0
HvFUT.9	0,713094833	0,28163517	2,391415	17,5472167	9,73887167	0,2689235	0,25343067	17,4011733
HvFUT.10	0,007099283	0,08164602	0,18082078	0	0	0,08284835	0,06882803	0
HvFUT.11	0,7015615	0,23864932	14,5778833	5,75070333	7,16611167	3,952295	0,80056167	4,95831333

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
HvFUT.1	8,10704	8,64051167	2,76036833	4,79002167	3,94124	0,38771372	0,2543215	20,4056
HvFUT.2	0	0,0085812	0	0	0	0,22435183	0,5851575	0
HvFUT.3	0,004746433	0,0123761	0	0,00582912	0	0,0943282	0,0748936	0
HvFUT.4	0,01272945	0,04705343	0	0,02027072	0,03226535	0,01363977	0,15934587	0
HvFUT.5	0,277696	0	0	0,51080883	0	0,643602	0,72185583	0
HvFUT.6	0,60449155	0,01324883	0	1,10767	0	0	0,00850777	0
HvFUT.7	0,003801033	0,02377268	0,00586387	0,00513017	0,00475493	0,33920167	1,10656383	0,00570625
HvFUT.8	0,727911667	2,39068567	0,00714195	0,763921	3,112025	0,0322209	1,41271217	0
HvFUT.9	6,21599	1,28229133	1,09567033	2,60927667	0,79049983	2,26658167	1,08323417	30,2602333
HvFUT.10	0,0546777	0,01337382	0	0,02192553	0,02541348	0,25506283	1,20537467	0
HvFUT.11	0,363548167	7,60668167	5,24002833	0,333052	2,07480833	8,952145	21,3576167	5,17561833

Table 3.12. FPKM values of reference barley genes for expression values from NGS barley RNA-seq data

Gene/Tissue	CAR5	CAR15	EMB	EPI	ETI	INF1	INF2	LEA
HvACTIN	47,08025	73,0077667	76,93515	28,8642	9,63949667	10,63076	7,04272667	15,0900333
HvADP-Rib	83,1514167	25,4791667	122,096167	34,5597167	31,0369167	71,13475	56,5808167	49,9299333
HvTUB	1162,17167	285,496833	996,7345	160,045833	628,404167	1270,27667	1174,19167	211,404833

Gene/Tissue	LEM	LOD	NOD	PAL	RAC	ROO	ROO2	SEN
HvACTIN	47,5131667	118,778433	15,46701	47,9020333	51,8085333	129,772783	245,811833	11,6668133
HvADP-Rib	47,2284	138,687	118,33355	94,9585333	111,9332	59,2634167	82,4950667	34,7503833
HvTUB	801,422667	1428,28833	1697,31667	1423,45167	1726,46833	431,446667	1295,61167	158,1715

After getting the expression values as FPKM, correlation was applied onto whole data set for examining expression patterns. Especially CSLC genes were targeted, due to its main object on xyloglucan synthesis in barley so co-expression analysis could be done and tabled in Table 3.12.

Table 3.13. Correlated FPKM values of CSLC genes versus targeted barley genes for expression values from NGS barley RNA-seq data

	HvCSLC1	HvCSLC2	HvCSLC3	HvCSLC4	HvCSLC5
HvCSLC1	1				
HvCSLC2	0,579147995	1			
HvCSLC3	0,392187193	0,062961596	1		
HvCSLC4	0,765216201	0,738937005	0,23933941	1	
HvCSLC5	-0,08272768	0,278092229	-0,09003785	0,19649603	1
HvXXT.1	0,576212853	0,281915228	0,19807331	0,68427371	-0,05676124
HvXXT.2	0,729358384	0,665638749	0,67153585	0,81426618	-0,02277258
HvXXT.3	0,363983745	-0,06890401	0,92264164	0,22362508	-0,11740981
HvXXT.4	0,348736718	0,088110277	0,93916224	0,23796415	-0,07925656
HvXXT.5	-0,06256758	-0,10487496	-0,14557265	0,0969181	0,17293652
HvXXT.6	0,344553754	-0,09241757	0,87241651	0,21245305	-0,12363384
HvXXT.7	-0,09654255	-0,03555282	-0,17853573	0,0749828	0,0618161
HvXXT.8	0,63387084	0,303431351	-0,07791388	0,43858496	-0,06482549
HvXylT_2.1	0,249102426	-0,19800355	0,73220326	0,11946627	-0,1306311
HvXylT_2.2	0,455393516	0,550968174	0,12329212	0,68895098	0,40801392
Hv-O-AcT.1	0,745767406	0,739296491	0,12831606	0,9339647	0,15825895
Hv-O-AcT.2	0,355908127	0,028692314	0,98002084	0,19177999	-0,10865144
Hv-O-AcT.3	0,241224341	0,034639803	0,01453362	0,49506536	-0,23033376
Hv-O-AcT.4	-0,17695137	-0,1867874	0,08082639	-0,1833873	0,17281705
Hv-O-AcT.5	0,253903436	-0,1739043	0,5950158	0,18191749	-0,10885606
Hv-O-AcT.6	-0,04598515	0,271006674	0,01573894	0,18047476	-0,18914984
Hv-O-AcT.7	-0,06899052	0,085136974	-0,30856048	0,01220037	-0,19786672
Hv-O-AcT.8	-0,01115875	-0,37733013	0,17435878	0,03041421	-0,15203392
Hv-O-AcT.9	0,376841188	-0,0471048	0,93593216	0,22853537	-0,12540516
Hv-O-AcT.10	0,015052564	0,193562618	-0,07341961	0,12007921	-0,10835351
Hv-O-AcT.11	-0,01691468	0,189273608	-0,12276327	0,2900088	0,46404756
Hv-O-AcT.12	0,899987062	0,578805532	0,39337313	0,85511387	-0,17167649
Hv-O-AcT.13	0,266821758	0,428963424	0,25783133	0,71599684	-0,08697405
Hv-O-AcT.14	-0,19664201	-0,31920694	-0,13884832	-0,1913817	-0,11726351
Hv-O-AcT.15	0,726345583	0,745573079	-0,09571782	0,82856225	0,01471849
Hv-O-AcT.16	-0,04502072	-0,27976821	-0,09598278	-0,0189573	-0,08127731
Hv-O-AcT.17	0,656619365	0,379639355	0,86784096	0,63011601	-0,05168547
Hv-O-AcT.18	0,249169542	0,490854091	0,16166192	0,5045551	0,71947334
Hv-O-AcT.19	-0,01876403	0,236218332	-0,13503567	0,14310399	-0,22024066
Hv-O-AcT.20	-0,19728841	-0,35623145	-0,14004224	-0,2774634	-0,10052571
Hv-O-AcT.21	0,189583504	-0,105557	0,74032201	0,05384751	0,15718659
Hv-AcT.1	-0,08888306	0,027175177	-0,04042607	0,10778619	-0,29820445
Hv-AcT.2	-0,02486686	-0,43889691	0,10701873	-0,0748124	-0,12439642
Hv-AcT.3	0,111156342	-0,23231165	0,21693565	0,09105906	-0,07330948
Hv-AcT.4	0,167481161	-0,19819255	0,3285314	0,13578846	-0,07763127
Hv-AcT.5	0,133642772	-0,2060733	0,23639931	0,11725943	-0,06837798
Hv-AcT.6	0,225833154	-0,34353294	0,6293672	-0,0320384	-0,24170006
Hv-AcT.7	0,085558773	0,022705918	-0,17802107	0,19987499	0,05296288

Hv-AcT.8	0,635085494	0,297273233	0,59750327	0,45390533	-0,05687355
Hv-AcT.9	-0,08637084	-0,14228086	-0,27425921	0,06405981	0,04504282
Hv-AcT.10	0,383110851	0,176499518	-0,12515701	0,26982981	-0,0664855
Hv-AcT.11	0,258525597	-0,06414898	0,92200597	0,75821043	-0,11915308
Hv-AcT.12	0,124735597	-0,20466519	0,20690556	0,11412912	-0,10559999
HvXGT.1	0,491652013	0,381237158	0,00548196	0,57643669	-0,15836705
HvXGT.10	0,372131861	0,002660497	0,99184102	0,2106975	-0,12234404
HvXGT.11	0,269770322	0,042333232	0,6088309	0,2483019	-0,1222118
HvXGT.12	0,34529133	-0,03835145	0,96901134	0,1742766	-0,12612991
HvXGT.13	-0,32271042	-0,15287758	-0,29677596	-0,1830734	-0,09481789
HvXGT.2	0,000600626	0,322361753	0,09022998	0,25474025	0,97989498
HvXGT.3	0,947007671	0,536285937	0,20924675	0,72164058	-0,14659817
HvXGT.4	-0,02930725	0,115781746	-0,16036388	0,19825185	0,37694441
HvXGT.5	0,735508767	0,521669293	-0,09248665	0,62793572	-0,10161247
HvXGT.6	-	-	-	-	-
HvXGT.7	0,383575683	0,027417218	0,99630277	0,22921576	-0,11141767
HvXGT.8	0,382534264	-0,01688221	0,9808214	0,23120965	-0,11624085
HvXGT.9	0,359959813	-0,01365906	0,97049266	0,213473	-0,13129308
HvXGT-like.1	0,827968379	0,370895454	0,01648024	0,49005742	-0,17201488
HvXGT-like.2	0,27834599	-0,16657972	0,68773293	0,1772135	-0,11443344
HvXGT-like.3	0,342646917	-0,09718155	0,95204162	0,15631976	-0,1649235
Hv_GALT_UCW98	0,707722582	0,601618934	0,50570771	0,83973276	0,23144643
HvFUT.1	-0,17931996	-0,00043481	-0,22805486	-0,2549896	-0,03507303
HvFUT.2	0,311857597	-0,15562019	0,74257859	0,21634262	-0,13847982
HvFUT.3	-0,04410004	-0,24736615	0,18223128	-0,1637055	-0,10022962
HvFUT.4	0,632674885	0,372586461	0,9157917	0,52851996	-0,00687611
HvFUT.5	0,197963546	0,076328556	0,73283477	0,1781996	0,35912691
HvFUT.6	-0,06004442	0,404125133	-0,09345507	0,17921165	0,90078844
HvFUT.7	0,312567897	-0,12943629	0,79266585	0,18046072	-0,14357091
HvFUT.8	0,672053383	0,785079143	0,18068748	0,89271732	0,04724975
HvFUT.9	-0,33728813	-0,32350932	-0,19990344	-0,4535309	-0,08600552
HvFUT.10	0,363176673	0,085880737	0,98264879	0,20513342	-0,10224887
HvFUT.11	0,476953207	-0,07008607	0,84782392	0,21273438	-0,25302859
HvACTIN	0,61157143	0,215013584	0,85691695	0,41714598	-0,06139368
HvADP-Rib	0,693708002	0,420242613	0,14791349	0,82310791	0,13142898
HvTUB	0,494589825	0,571324646	0,13900036	0,76173167	0,21864403

According to Pearson Correlation, when the positive association is observed among quantitative variables, strength of relationship must be considered. If some of values placed 0.7 or above, it is decided that the relationship in between variables is generally accepted as strong and significant [81]. For this reason, on Table 3.13, significant results from CSLC expression values are listed below and for further investigation, these values are used for expression pattern analyses.

Table 3.14. Correlation values that above 0.7 from barley CSLC expression values

	HvCSLC1			HvCSLC3
HvCSLC1	1		**HvCSLC3**	1
HvXGT.3	0,947007671		**HvXGT.7**	0,996302766
Hv-O-AcT.12	0,899987062		**HvXGT.10**	0,991841015
HvXGT-like.1	0,827968379		**HvFUT.10**	0,982648794
HvCSLC4	0,765216201		**HvXGT.8**	0,9808214
HvOAcT.1	0,745767406		**Hv-O-AcT.2**	0,980020839
HvXGT.5	0,735508767		**HvXGT.9**	0,970492663
HvXXT.2	0,729358384		**HvXGT.12**	0,969011343
Hv-O-AcT.15	0,726345583		**HvXGT-like.3**	0,952041619
Hv_GALT_UCW98	0,707722582		**HvXXT.4**	0,939162241
			HvOAcT.9	0,935932164
	HvCSLC4		**HvXXT.3**	0,922641637
HvCSLC4	1		**HvAcT.11**	0,922005974
Hv-O-AcT.1	0,933964701		**HvFUT.4**	0,915791702
HvFUT.8	0,89271732		**HvXXT.6**	0,872416511
Hv-O-AcT.12	0,855113867		**Hv-O-AcT.17**	0,867840965
Hv_GALT_UCW98	0,839732763		**HvACTIN**	0,856916948
Hv-O-AcT.15	0,828562251		**HvFUT.11**	0,84782392
HvADP-Rib	0,823107907		**HvFUT.7**	0,792665851
HvXXT.2	0,814266176		**HvFUT.2**	0,742578593
HvCSLC1	0,765216201		**Hv-O-AcT.21**	0,740322014
HvTUB	0,761731669		**HvFUT.5**	0,732834769
Hv-AcT.11	0,758210429		**HvXylT_2.1**	0,732203262
HvCSLC2	0,738937005			
HvXGT.3	0,721640579			**HvCSLC2**
Hv-O-AcT.13	0,715996841		**HvCSLC2**	1
			HvFUT.8	0,785079143
	HvCSLC5		**Hv-O-AcT.15**	0,745573079
HvCSLC5	1		**Hv-O-AcT.1**	0,739296491
HvXGT.2	0,97989498		**HvCSLC4**	0,738937005
HvFUT.6	0,90078844			
Hv-O-AcT.18	0,719473338			

According to Pearson Correlation results, for barley CSLC genes, there are 9 different genes for CSLC1, 4 genes for CSLC2, 22 genes for CSLC3, 13 genes for CSLC4, and 3 genes for CSLC5 are listed as significant. Correlation results show that those genes are co-expressed with CSLC in significant level, so they might have strong relationship in each other. For this reason, tissue by tissue FPKM values were investigated for co-expression level analyses and those values were adjusted to 100% expression level for better understand the results.

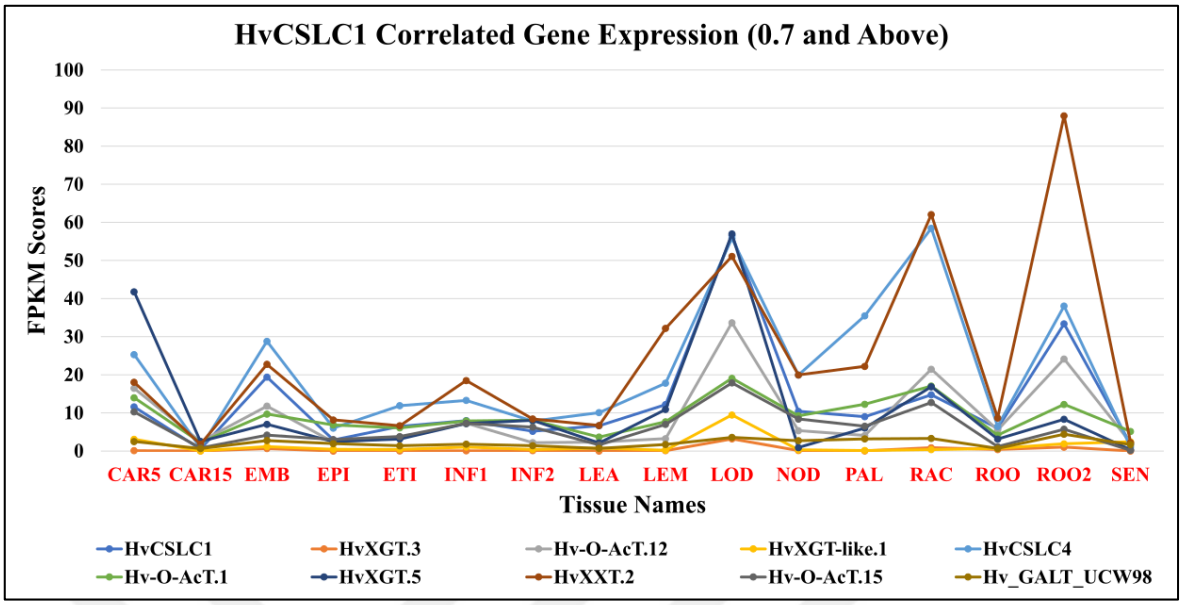


Figure 3.7. FPKM scores of genes that significantly correlated with HvCSLC1, tissue by tissue

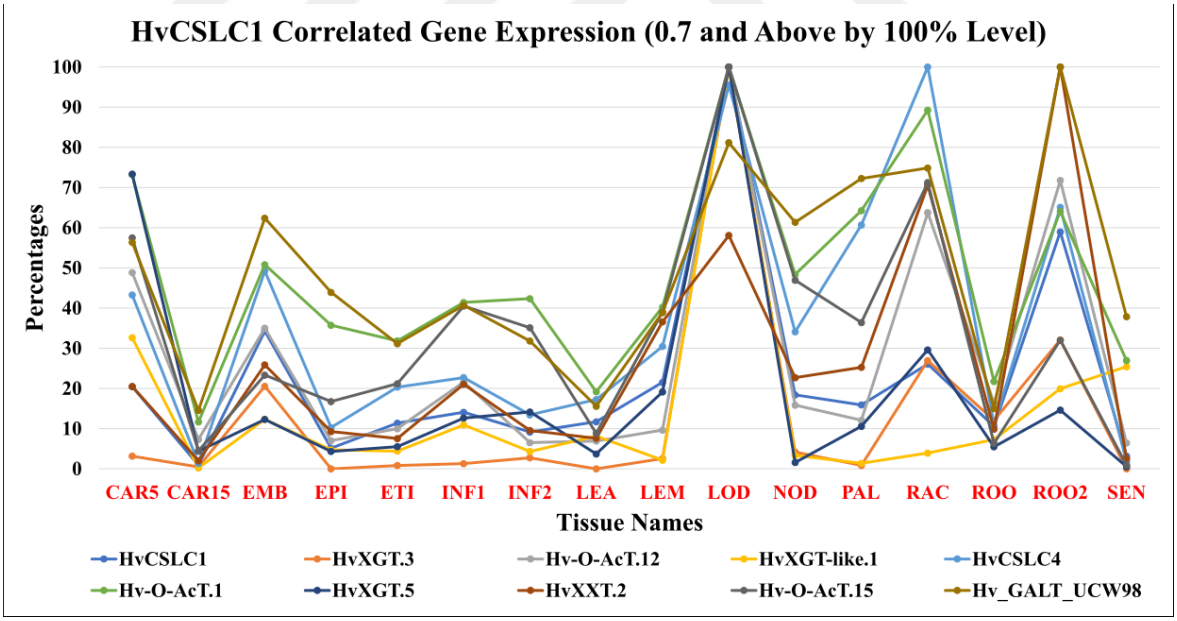


Figure 3.8. Expression level (as percentages) of genes that significantly correlated with HvCSLC1, tissue by tissue

For HvCSLC1, in Grain-1, Lodicule, Rachis, and Root-2 tissues, high level of co-expression could be observed in Figure 3.7. and 3.8. Besides, in Embryo and Grain tissue, co-expression could be come out above average, according to graphs that located above.

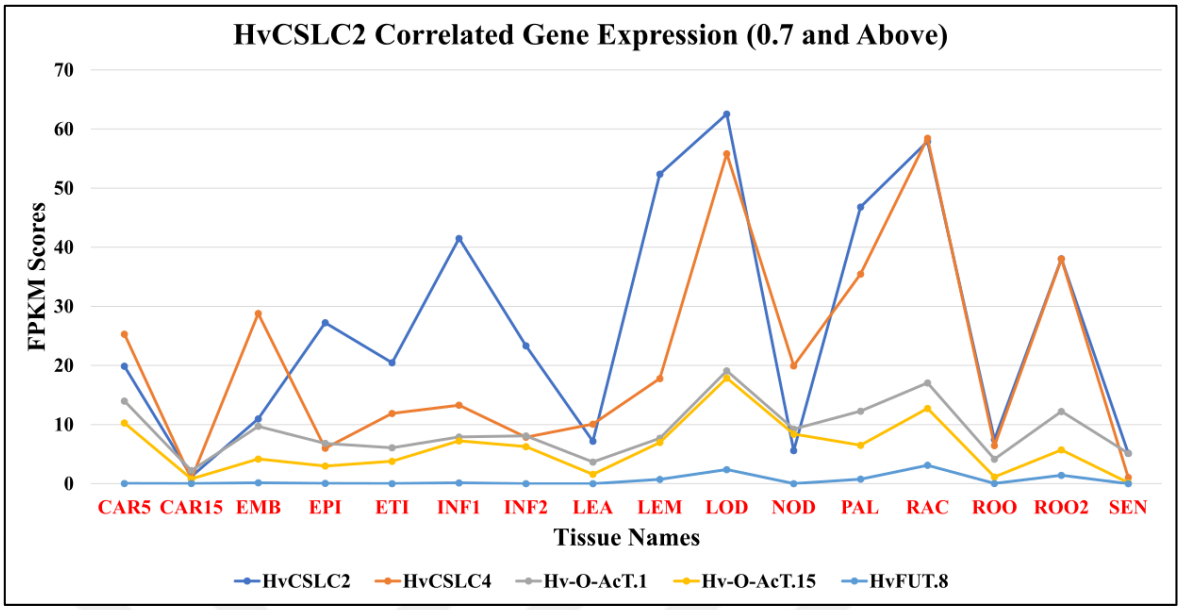


Figure 3.9. FPKM scores of genes that significantly correlated with HvCSLC2, tissue by tissue

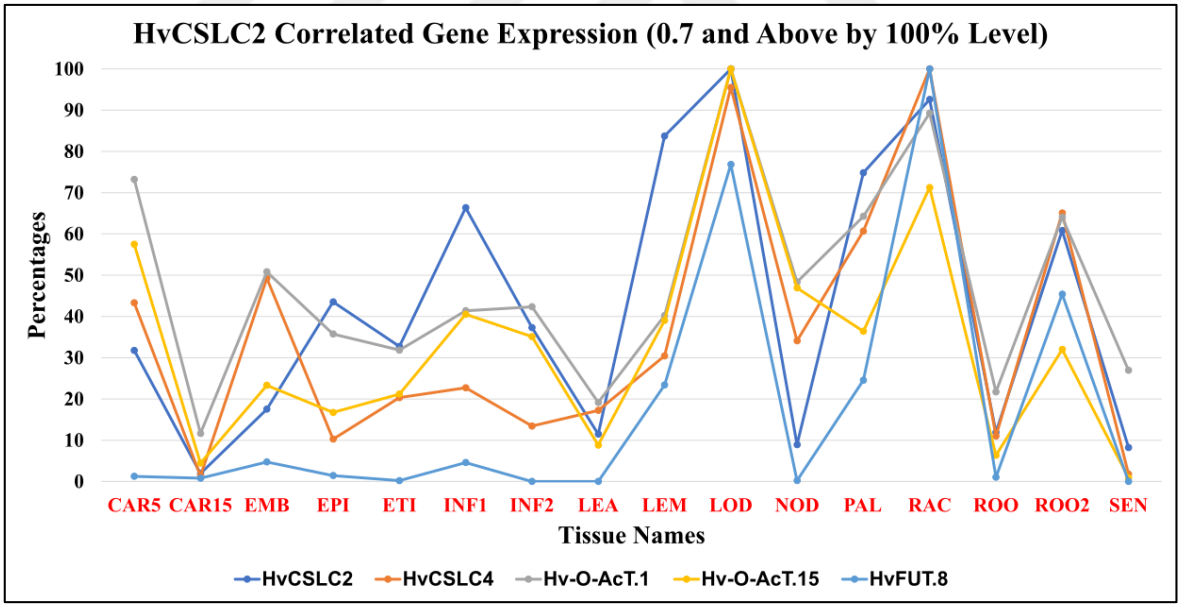


Figure 3.10. Expression level (as percentages) of genes that significantly correlated with HvCSLC2, tissue by tissue

In Figure 3.9. and 3.10., just like the previous case, HvCSLC2 might co-express with significantly correlated genes in Grain-2, Lodicule, Rachis, and Root-2 tissues.

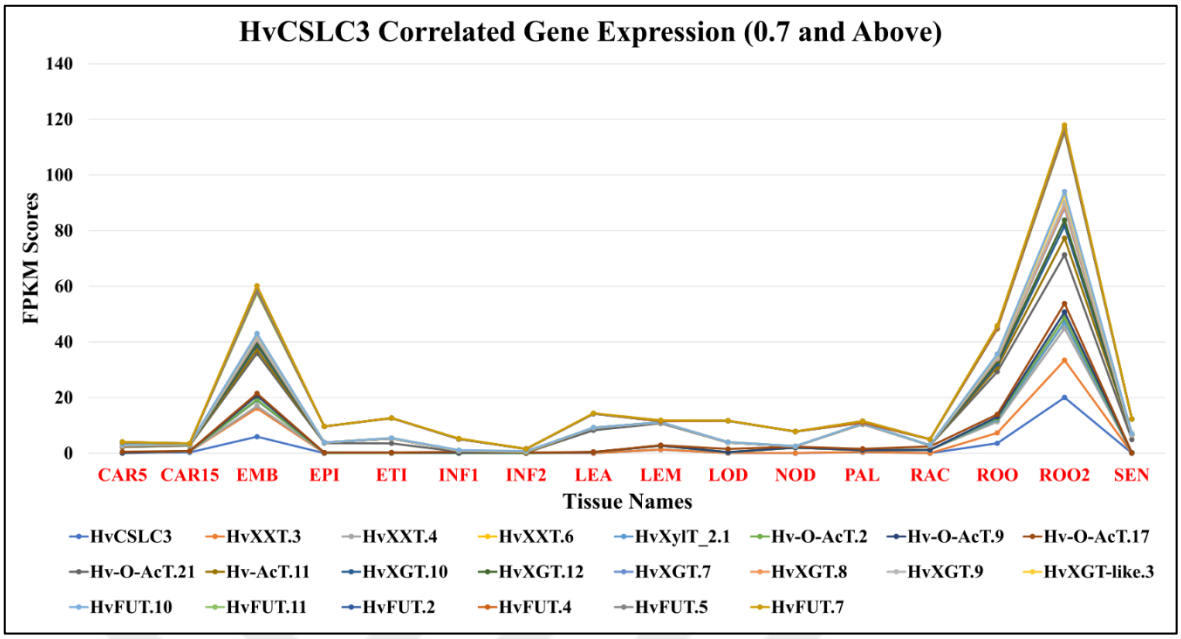


Figure 3.11. FPKM scores of genes that significantly correlated with HvCSLC3, tissue by tissue

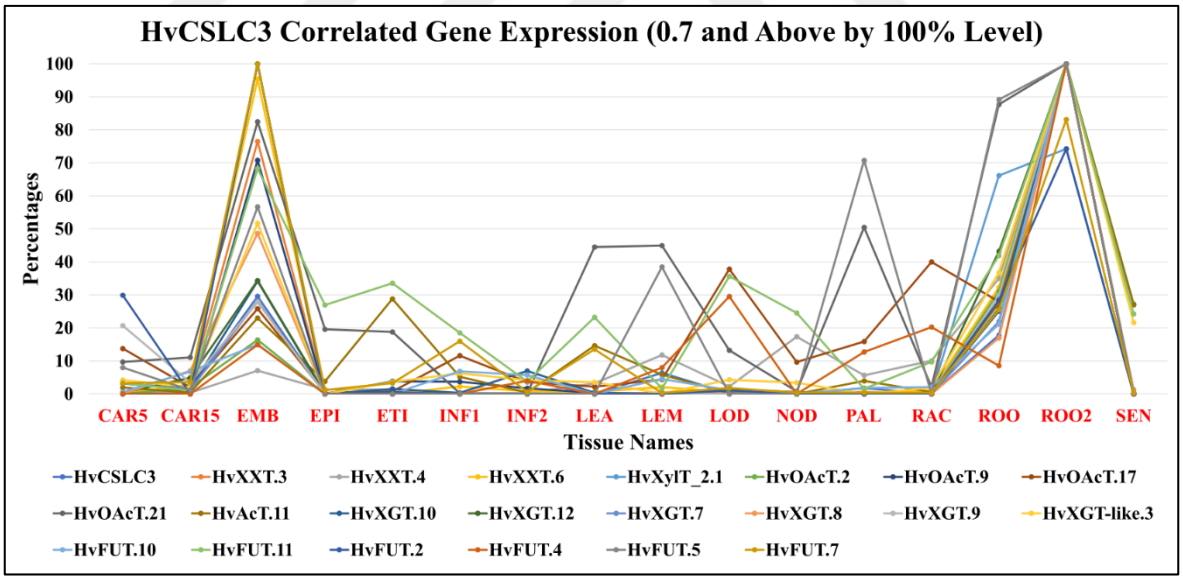


Figure 3.12. Expression level (as percentages) of genes that significantly correlated with HvCSLC3, tissue by tissue

According to Figure 3.11. and 3.12, HvCSLC3 might co-express with significantly correlated genes in Embryo, Root and Root-2 tissues.

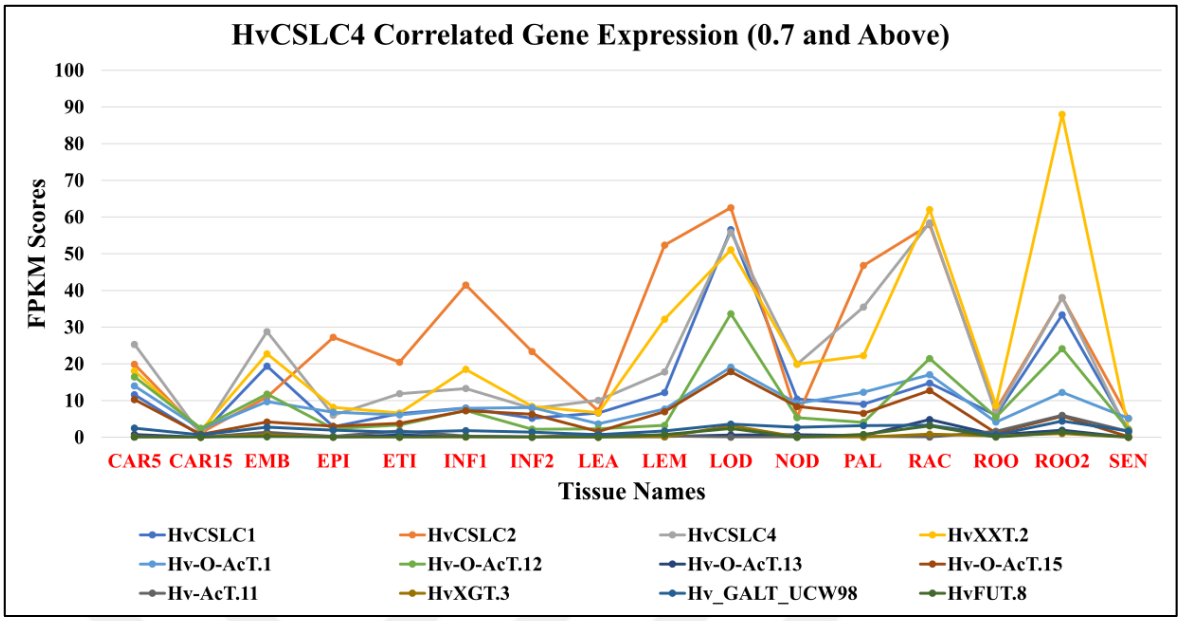


Figure 3.13. FPKM scores of genes that significantly correlated with HvCSLC4, tissue by tissue

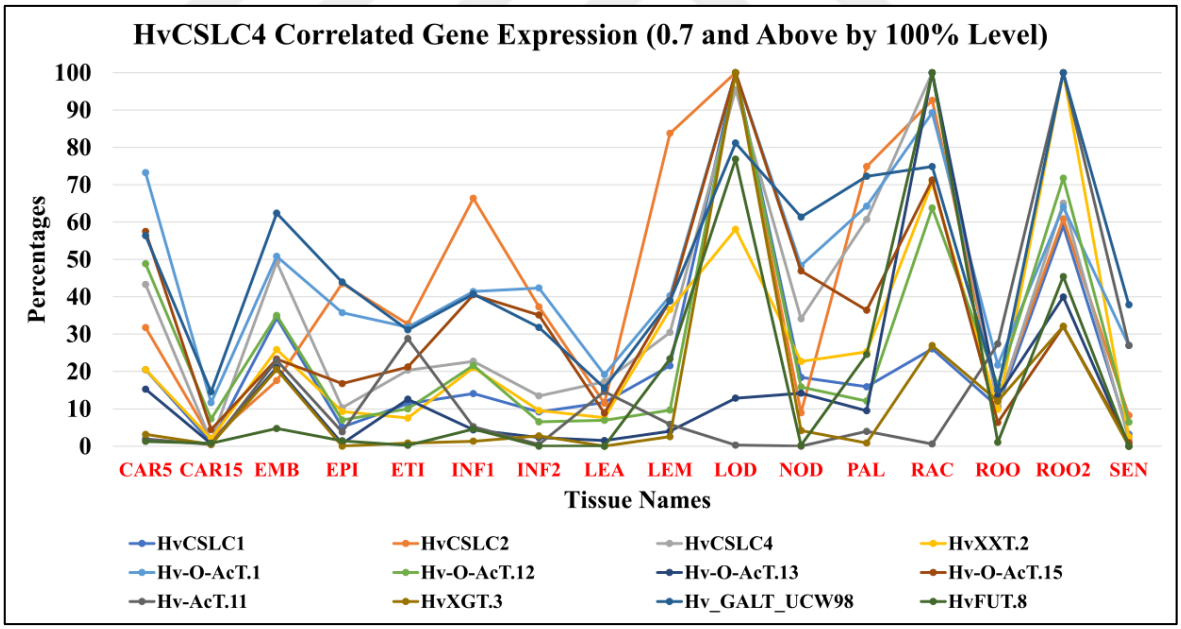


Figure 3.14. Expression level (as percentages) of genes that significantly correlated with HvCSLC4, tissue by tissue

In Figure 3.13. and 3.14., it is seemed that HvCSLC4 might co-express with significantly correlated genes in Embryo, Lodicule, Rachis, and Root-2 tissues in barley.

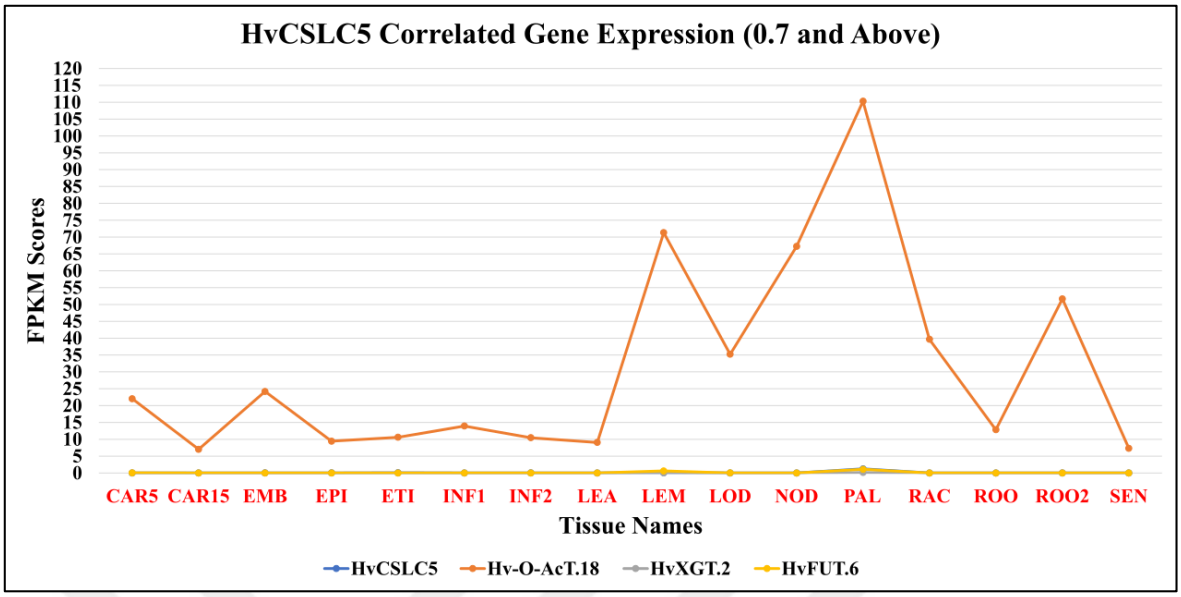


Figure 3.15. FPKM scores of genes that significantly correlated with HvCSLC4, tissue by tissue

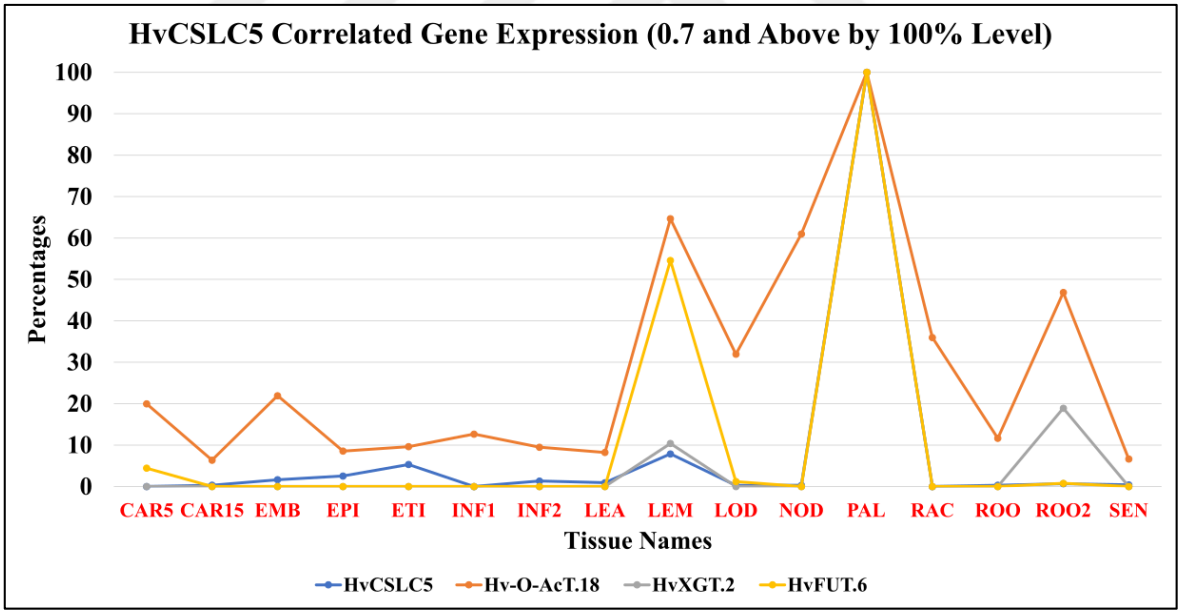


Figure 3.16. Expression level (as percentages) of genes that significantly correlated with HvCSLC4, tissue by tissue

In Figure 3.15. and 3.16., it is seemed that HvCSLC5 might co-express with significantly correlated genes in Lemma, Palea, and Root-2 tissues in barley.

4. DISCUSSION

In this study, a group of genes that could potentially be involved in xyloglucan synthesis process in barley were investigated, including sequences CSLC, XXT, XGT, and O-AcT genes. From this point, comparisons were made especially in between XyG synthesis related genes from *Arabidopsis thaliana* and *Oryza sativa*. To date, the cell walls of all known flowering plants contain xyloglucan, so XyG has many important roles in the wall structural such as structural organization, development and growth control. Because of this, XyG biosynthesis in barley was investigated and genes that are likely to be involved in this process were detected and analyzed. The main target was the CSLC gene family due to its essential role in synthesizing the 1,4-β-linked glucosyl backbone of XyG. In addition, other gene families and their enzymes were studied and analyzed, including XXT, XGT, O-AcT, and FUTs. When the genes were detected and investigated their protein's sequences determined, MSA were made for making comparisons between barley and other species such as *Arabidopsis thaliana*, rice, black cottonwood, and *Brachypodium distachyon*, for each gene family. In addition to making sequence alignments, phylogenetic trees were constructed to visualize the level of relationships between the sequences. Genes' location on chromosomes were identified, along with their sizes and intron/exon regions to aid in analysis of the RNA-seq data. Received RNA-seq data were analyzed to investigate co-expression patterns of the genes potentially involved in XyG biosynthesis in barley.

In identification of barley gene groups, after careful screening from available barley sequences, this current work revealed that there might be at least 5 CSLC genes in barley. Previous research had detected only 4 CSLC genes in barley [82]. Looking at the MSA and phylogenetic tree results, it can be claimed that there is a high similarity and relatedness in between all HvCSLCs including the probable fifth one. Just like in *Arabidopsis thaliana* case, it should be claimed that there is a fifth CSLC gene in barley because even monocotyledon/dicotyledon difference *Arabidopsis thaliana* has 5 different CSLC genes in its genome [83]. On the other hand, *Oryza sativa* has 6 different CSLC genes [82], as a monocot plant that there is a possibility about even barley might have sixth CSLC in its genome, because of high relatedness in between rice species. It could also support the idea that CSLC gene family could evolve from the ancient gene family that it could widespread among plant classes.

Looking for possible xylosyl transferases that located in barley genome, there is a difference in between “xylosyltransferase 2” and “xyloglucan 6-xylosyltransferase”. It is identified that there are 8 xyloglucan 6-xylosyltransferase and 2 xylosyltransferase 2. Normally it is known [84] and accepted that xyloglucan 6-xylosyltransferase have been taking a part in xyloglucan synthesis in plants because this enzyme can make α-1,6-linkages with glucose for making xyloglucan residues. In other words, these xyloglucan 6-xylosyltransferases could be barley form of *Arabidopsis thaliana*’s XXT1, XXT2, and XXT5 genes that involved for transferring xylosyls to glucose residues [73]. However, xylosyltransferase 2 can make β-1,2-linkage with glucose again, but this time for making glycan residues [85]. On the other hand, because of similar regions and phylogenetic relatedness, these xylosyl transferases could involve xyloglucan biosynthesis even it is not main purpose. Besides that, there might 10 different xylosyl transferases that could be involved xyloglucan biosynthesis.

For identification of possible galactosyl transferases that have been involved in barley, the number of possible gene that could affect in xyloglucan backbone is increasing, about 17 different gene. It is a fact that the well-known MUR3 gene that located in *Arabidopsis thaliana* genome can involve galactosyl transferase activity in xyloglucan biosynthesis, on the third xylose residue especially [86]. On the other hand, there might be more than 10 possible coding regions that highly similar with MUR3’ sequence in *Arabidopsis thaliana* [87]. For this reason, there might be more than 10 different genes that encode galactosyl transferases and these proteins could involve enzymatic activities for synthesis of XyG. Those different proteins might interact with different residues on the XyG backbone, like the second or first xylosyl residue. There is a research that tries to explain this possibility in different plant species [88].

In the acetyl transferase identification case, the searching could be done carefully and deeply because of there are many types of acetyl transferases that make additions to compounds like histone, serine, and threonine [89]. To analyze and discover the possible acetyl transferase genes that involved in XyG biosynthesis in barley, just like previous cases, BLAST investigations have been made and ENSEMBL protein predictions were used. Just like in xylosyl transferases, in this group, there are two subgroup that could be named as “acetyltransferases” and “xyloglucan o-acetyltransferases”. There is a known information that the group of o-acetyltransferases that can involve and catalyze xyloglucan backbone acetylation process in *Arabidopsis thaliana* and rice species [90], so many other ones could

interpret acetylation process in XyG biosynthesis in barley, though. The potential candidates could be two different protein family which are reduced wall acetylation (RWA) and the trichome birefringence-like (TBL), because the study claimed they could involve o-acetylation process with different polymers in *Arabidopsis* cell wall [91]. That's why, 21 potential xyloglucan o-acetyltransferases have been identified and investigated. Moreover, due to high similarity and relativeness, 12 acetyl transferases have been added to the list from barley genome for further investigations.

Identification of fucosyl transferases for XyG biosynthesis in barley should have done because even though there are some studies that point out that fucosylation is a rare process in monocotyledons in xyloglucan synthesis, even some of them have no fucosylated XyG backbone [56]. However, any possible contribution onto biosynthesis of XyG should be investigated, so possible fucosyl transferase genes were searched for and analyzed. There are 11 possible FUTs have been identified for catalyzation of galactosyl residues in the xyloglucan backbone in barley.

In barley's CSLC proteins, looking at their MSA results, first of all, it is clearly detected that there is a high number of conserved regions with the same patterns even among different species. Even those species belong different plant classes (monocotyledon/dicotyledon), this situation could be identified directly. In review article that recently published, they mentioned and pointed many conserved regions that located among many CSL and CESA proteins. In here, they mentioned that there are some motifs that located in CESA proteins, could be observed and identified in many CSL proteins.

Especially, they pointed out four important motifs, which are three "D" motifs and one "QXXRW" motif for CESA genes, but the first motif, described as "DDG" was not observed especially in OsCSLC3 [79]. However, in this thesis' study, it might be claimed that those motifs are not only located some of CSL, it could be located as conserved for among all CSLC proteins. For this thesis, 23 different CSLC protein were investigated from 4 different species, and all of them have two "D" motifs (second and third ones of CESAs) and one "QXXRW", just like CESA proteins. These "D" motifs showed as "DAD" and "DMD" among all 23 CSLC proteins. Furthermore, "QXXRW" motif is displayed as "QQHRW" among all proteins, except AtCSLC6, which it has "QQYRW". Those highly conserved motifs could be marked as great evolutionary background for CSLC genes. On the other hand, in review paper [79], they also mentioned that there are some other motifs that located

in CESA proteins, they or their variabilities could be located on CSLC proteins. In this point, there should do some further investigations about other conserved motifs on CSLC proteins, and many other CSLC proteins should be included from other species for investigation of motifs and plant-conserved regions.

About other protein families' MSA results, especially in xylosyl transferases, there are some similar regions identified and detected among different species' proteins. On the other hand, in galactosyl and acetyl transferases ones, the level similarity is low due to high number of proteins that be aligned. In addition, it could be happened because in this study could try to cover all possible proteins that intervene and involve in XyG biosynthesis in some level or way, so just described above, some proteins could express more specific for synthesis of xyloglucan, but other ones could be involved transfer onto totally different substrates. In future, when the group of possible enzymes could be narrowed with some other investigations such as dedicating proteins more precisely, there might be a possibility for detecting some conserved regions among those protein groups even among species like CSLC case. At this point, this study could not focus and investigate those MSA results deeply.

In phylogenetic tree of CSLC proteins, it is clearly understood that there is a high relatedness among CSLCs. Bootstrap values would also describe and explain relatedness level and this situation can also support the idea that CSLC gene family might evolve from very ancient gene family. That's why, those conserved regions, same motifs, and high bootstrap values could be received and detected even among other plant species. In short, generally monocotyledon species' CSLCs are close to each other. On the other hand, they are not separated apart from dicotyledon species ones in phylogenetic tree.

There is a same pattern that could be observed in phylogenetic tree of xylosyl transferases, as in the CSLC case. Same class species' proteins have been clustered to each other closely. Moreover, bootstrap values could prove to relativeness of each of them. The important part of this tree is the difference in between xylosyltransferase 2 and xyloglucan 6-xylosyltransferase is observable clearly. Even this slight difference can come out that figure and the closest proteins to them are *Brachypodium distachyon* ones. These 2 proteins might be *Brachypodium distachyon*'s "xylosyltranferase 2" ones but it should be investigated in further researches. Moreover, this difference can be detectable not only phylogenetic tree, MSA results can give this information, though.

Just like previous cases, in phylogenetic tree of galactosyl transferase proteins, there is a certain difference and diversion in between monocotyledon and dicotyledon class species' proteins. Besides, on the middle side of the tree, a group of proteins that looks very close to each other, according to their bootstrap values and position of them. These are HvXGT.1 and 2 (xyloglucan galactosyltransferase candidate from barley), PtXGT_GT19.1 and 2 (xyloglucan galactosyltransferase candidate from black cottonwood), OsXGT_GT19 (xyloglucan galactosyl transferase candidate from rice), and AtXGT_GT19 (xyloglucan galactosyl transferase candidate from *Arabidopsis thaliana*). It might be claimed that HvXGT1. and 2 are member of GT19 subfamily that they could be involved in the catalytic process of XyG biosynthesis. Furthermore, this situation could explain the idea that there could be an evolutionary pathway among these proteins, so if they had any conserved regions on their sequences, it should be investigated in any future studies. At the end, rice XLT2 protein did branching with HvXGT.3 on phylogenetic tree, what's more they are very close to poplar and *Arabidopsis thaliana*'s XLT2s, so this barley protein might XLT2 derivatives, Normally, XLT2s are responsible from galactosylation at second position of the XXGG [87], so HvXGT.3 could interact with second xylosyl group on xyloglucan composition.

There were 77 different acetyl transferases placed in their phylogenetic tree, it means that it is the broadest participation protein family that was investigated in this study. There is an important detail in this tree, actually a protein that named as "BdXyBAT1" on tree. It is a studied protein before, and explained in depth in this letter [80]. According to this study, this specific enzyme from *Brachypodium distachyon* can modulate catalytic reaction of xyloglucan. Normally this protein is a member of TBL family, but its ability of acetylation is quite high that it is not only XyG backbone acetylase in this plant. In short, this protein has clear proof of its role in the acetylation process in XyG synthesis. For this reason, on the phylogenetic tree, the closest protein to BdXyBAT1 is Hv-O-AcT.3 which could be related with acetylation process on XyG biosynthesis in barley at some level. In addition, there are 2 more barley proteins that are close to BdXyBAT1 on phylogenetic tree, which are Hv-O-AcT.2 and 6 could also be related acetylation process in XyG biosynthesis. At the end, class difference could be observed again on tree, and the difference in between "acetyltransferases" and "xyloglucan o-acetyltransferases" can be observed clearly on phylogenetic tree, and also alignment results, though.

For examining the expression patterns of the targeted barley genes to investigate XyG biosynthesis, RNA-seq data, from on barley cultivar Morex was obtained from James Hutton Institute, Dundee, Scotland. From this data the FPKM values for each of them studied genes was calculated. In this RNA-seq data, there are 16 different tissue; each of them has 3 replicates from 2 different batches, so in total 96 samples were sequenced, and the data could be investigated in a statistically robust manner. Besides that, there are 77 candidate barley genes were examined, plus reference genes. In this study, after receiving FPKM values from specialized software, every value for any gene has been averaged tissue by tissue. At last, they are 16 averaged scores for every gene, and they were correlated with each other with these averaged FPKM values. In this research, CSLC genes' correlation results were focused and tabled above, in Table 3.13.

In Pearson Correlation results of CSLC barley genes, they are many genes that can be interpreted as having positive correlation. These positive numbers could be counted as strong/high and significant correlations because of statistical approaches, especially when the values are equal 0.7 or above [81]. On the other hand, there is no negatively correlated result among all correlation results. First of all, it might be counted as first strong sign of co-expression of all of genes for XyG biosynthesis.

Looking for genes with significant correlation results to barley CSLCs, there are 9 different genes for CSLC1, 4 genes for CSLC2, 22 genes for CSLC3, 13 genes for CSLC4, and 3 genes for CSLC5 were found where the correlation was highly significant. These genes are the best candidates for co-expression analyses because their correlation results are equal 0.7 or above value.

Starting with HvCSLC1, especially in 4 tissues which are Grain-1 (5 DPA), Lodicule, Rachis and Root-2 tissue (28-day seedling) showed co-expression with 9 different genes that could be involved XyG biosynthesis (listed in Table 4.1.). In addition, the top 3 FPKM values for the majority of the highly correlated genes are in the Lodicule, Rachis, and Root-2 tissues. The interesting point is that one of these genes that is highly correlated to HvCSLC1 is HvCSLC4, so CSLCs might enhance their activities to each other, as well.

Table 4.1. HvCSLC1 co-expression pattern in LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow

Gene/Tissue	LOD	RAC	ROO2
HvCSLC1	56,5917	14,74898333	33,36406667
HvCSLC4	55,82945	58,4573	38,03121667
HvXXT.2	51,07791667	62,05865	87,96508333
HvXGT.3	3,1757	0,855295833	1,018858833
HvXGT.5	56,97591667	16,86605	8,326125
HvXGT-like.1	9,451386667	0,371187333	1,885266667
Hv_GALT_UCW98	3,576075	3,298956667	4,406601667
Hv-O-AcT.1	19,09201667	17,04236667	12,22596667
Hv-O-AcT.12	33,65391667	21,45271667	24,14755
Hv-O-AcT.15	17,8707	12,72942167	5,720926667

In the HvCSLC2 case, 4 genes were listed as significantly correlated. Just like the previous case, the same tissues (LOD, RAC, and ROO2) have the top three FPKM values among all tissues. Again, another CSLC gene, same as the previous one, HvCSLC4 has co-expressed with HvCSLC2. Furthermore, HvCSLC2 co-expressed with same O-AcTs and FUT that are also co-expressed with HvCSLC4, so HvCSLC2 might be leaded to responsible from XyG compositions that have acetyl and fucosyl subunits. According to result, it might be claimed that HvCSLC2 would be co-expressed with these listed in Table 4.2. genes in significant level, to involve XyG biosynthesis in barley at some level.

Table 4.2. HvCSLC2 co-expression pattern in CAR5, LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow

Gene/Tissue	CAR5	LOD	RAC	ROO2
HvCSLC2	19,86823333	62,54158333	57,90226667	38,04838333
HvCSLC4	25,30503333	55,82945	58,4573	38,03121667
Hv-O-AcT.1	13,97888333	19,09201667	17,04236667	12,22596667
Hv-O-AcT.15	10,271485	17,8707	12,72942167	5,720926667
HvFUT.8	0,0392094	2,390685667	3,112025	1,412712167

The strongest correlations of the co-expression analyses were found in the HvCSLC3 results, because it comprises 23 genes. This is the largest group of co-expressing genes among the HvCSLCs, so it is of great interest and might be an essential part of XyG biosynthesis in barley. Furthermore, these 23 genes are covering all groups that were investigated. That's why, HvCSLC3 might be a major gene responsible for affecting or triggering other groups

of genes for inclusive co-expression pattern in barley, especially in roots. It is observable and detectable that when the HvCSLC3 and other genes' (significant correlated one) correlation results, especially in both Root tissues (Root-1/2) and Embryonic tissue, the highest FPKM values for these genes (especially top three) were recorded and noted. In this way, HvCSLC3 might be one of the major gene that govern and orient XyG biosynthesis on these tissues. First of all, HvCSLC3 is highly co-expressed with HvXXT.3 and HvXXT.4, especially in two different root tissues, which means that at these tissues we might observe much xylosylated sidechain on XyG like XXGGG. This might be the most common XyG composition in these tissues, though. In addition, HvXXT.6 is also co-expressed with HvCSLC3, but in low level. However, this xylosyl transferase could lead to expect to observe synthesis of XXXGG subunits in root tissues. Secondly, HvXGT.7 and HvXGT.10 are co-expressed with HvCSLC3, especially in Root-2 tissue, so in there would be interesting to see if barley roots had any XLGGG or XXLGG compositions. Because of high co-expression level in between these HvXGTs and HvCSLC3, Hv-O-AcTs are also co-expressed, so these XLGGG and XXLGG compositions would be also acetylated. However, there is a high FKPM value in root tissues for Hv-O-AcT.21, comparing with others. It might be explained that Hv-O-AcT.21 could lead the acetylation process on galactosyl residues located on XyG, in root tissues. Furthermore, with contribution of FUTs, at these same compositions' galactosyl residues could be fucosylated, and this time maybe XFGGG or XXFGG subunits could be observable. Besides, FUT.11 has showed reasonable expression levels, comparing with other FUTs. On the other hand, this situation might be explained that FUT.11 was actually fucosylating something that not related with galactosyl residues on XyG. In Table 4.3., it is more understandable that the co-expression level of HvCSLC3 with other genes in EMB, ROO, and ROO2 tissues.

Table 4.3. HvCSLC3 co-expression pattern in EMB, ROO, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow

Gene/Tissue	EMB	ROO	ROO2
HvCSLC3	5,92175	3,568135	20,04356667
HvXXT.3	10,22028833	3,735768333	13,36308333
HvXXT.4	0,806579833	3,995753333	11,41826667
HvXXT.6	2,001138333	0,6766545	2,097391667
HvXylT_2.1	0,034310133	0,022696783	0,025479333
HvXGT.7	1,214341333	0,947234	4,315948333

HvXGT.8	0,328615333	0,1145767	0,675709333
HvXGT.9	0,369651	0,396196	1,328846
HvXGT.10	1,5341715	1,125944	4,490371667
HvXGT.12	0,696657167	0,876133833	2,027788333
HvXGT-like.3	1,400858333	0,997027833	2,710016667
Hv-AcT.11	1,375823333	1,6382365	5,986823333
Hv-O-AcT.2	0,295774	0,562080333	1,806188333
Hv-O-AcT.9	1,401194667	0,511977833	1,97946
Hv-O-AcT.17	0,7899875	0,855053	3,057946667
Hv-O-AcT.21	14,43533333	15,35701833	17,51186667
HvFUT.2	0,7883715	0,224351833	0,5851575
HvFUT.4	0,023748667	0,013639767	0,159345867
HvFUT.5	0,408946667	0,643602	0,721855833
HvFUT.7	1,330882	0,339201667	1,106563833
HvFUT.10	0,180820783	0,255062833	1,205374667
HvFUT.11	14,57788333	8,952145	21,35761667

For co-expression analysis of HvCSLC4 and other genes that correlated with HvCSLC4 in significant level, there are some similar co-expression patterns with HvCSLC1. First of all, their FPKM values showed highest ones especially in LOD, RAC and ROO2 tissues (On Table 4.4., it was listed), just happened in HvCSLC1. Furthermore, both CSLC have been found as high correlated with same genes which are HvXXT.2, Hv_GALT_UCW98, Hv-O-Act.1-12-15s. Even more, they are also showing strong correlation to each other, too. For these reasons, it should be assumed that HvCSLC1 and 4 could work together with for XyG biosynthesis, especially in specific barley tissue which are Lodicule, Rachis, and Root-2. Furthermore, they could co-express with same xylosyl transferase, galactosyl transferase, and acetyl transferases for some specific oligosaccharide compositions.

Table 4.4. HvCSLC4 co-expression pattern in LOD, RAC, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow

Gene/Tissue	**LOD**	**RAC**	**ROO2**
HvCSLC1	56,5917	14,74898333	33,36406667
HvCSLC2	62,54158333	57,90226667	38,04838333
HvCSLC4	55,82945	58,4573	38,03121667
HvXXT.2	51,07791667	62,05865	87,96508333
HvXGT.3	3,1757	0,855295833	1,018858833
Hv_GALT_UCW98	3,576075	3,298956667	4,406601667
Hv-AcT.11	0,016119483	0,034688917	5,986823333
Hv-O-AcT.1	19,09201667	17,04236667	12,22596667

Hv-O-AcT.12	33,65391667	21,45271667	24,14755
Hv-O-AcT.13	0,619165317	4,823753333	1,926411667
Hv-O-AcT.15	17,8707	12,72942167	5,720926667
HvFUT.8	2,390685667	3,112025	1,412712167

The last HvCSLC's result pool is quite few, so making a suitable co-expression analysis is harder than the previous ones. There are only 3 genes that could co-express with HvCSLC5, listed in Table 4.5. On the other hand, there is an interesting point that it is worth for discussion which is almost all highest FPKM among these genes were recorded different tissues which are Lemma, Palea, and Root-2. Except Root-2, other two tissues were collected from 6-week-old barleys, so HvCSLC5 might be taking a role especially in older tissue like these ones. In other words, this candidate CSLC may be increased its activity on older tissues, so it might be hard to detect that. Generally, barley studies contain much younger samples or tissues (in between 1-day and 28-day seedlings) so it might be missed that gene activity in some level. On the other hand, these FPKM values are very low, so contribution of HvCSLC5 in XyG biosynthesis might be negligible.

Table 4.5. HvCSLC5 co-expression pattern in LEM, PAL, and ROO2 tissues. Top three FPKM values for each gene, across all tissues, are highlighted in yellow

Gene/Tissue	**LEM**	**PAL**	**ROO2**
HvCSLC5	0,097154933	1,233805	0,0084587
HvXGT.2	0,016528017	0,159143783	0,030093333
Hv-O-AcT.18	71,30486667	110,2898833	51,67293333
HvFUT.6	0,60449155	1,10767	0,008507767

Comparing with CSLC co-expression results to each other, there is a point that xyloglucan synthesis might be increased or reached at high level in some specific tissues that were collected and isolated from around 4-6 weeks seedling which are Lodicule, Rachis, and Root-2 especially. At that time period, cell wall structure could be needed to be thick, so this time interval and also these tissues should be investigated in future studies. Besides, according to barley growth and development guideline [92] this time interval overlaps with the time period that when the jointing was started (4 weeks) and when the boot formation was observed (around 45th days). That's why, beginning of jointing and boot arousing time should be searched for its importance in XyG biosynthesis how could be increase or accelerate those targeted genes expression levels. On the other hand, there are some high

FKPM values and highly possible co-expression patterns that could be observed in tissue which is Embryo that in germinating period, due to increasing in number of cells (also number of cell wall contents) and also starting to differentiation, need for any kind of fundamental materials and molecules is urging. That's why, at that time period, XyG biosynthesis might be increased.

Another point is that which xyloglucan oligosaccharide subunit could be abundant in between all tissues or time periods in barley. According to research on barley coleoptiles [67], the most abundant xyloglucan oligosaccharide subunit is XXGGG (27 Mole percent). The research was conducted only in the coleoptiles, on the other hand, it might be expended among barley's all lifetime periods. That's why, there might be constructed a composition with the most possible outcome with its synthesis. Because the XXGGG is observed in coleoptile tissue as the most abundant, so this backbone part might be synthesized by HvCSLC3 and its the most significant co-expressed xylosyl transferases. Because of HvCSLC3's high activity in Embryo and also Root, tissue, it might be made this kind of assumption in this study for how to be demonstrated XyG biosynthesis. On the other hand, there is no enough information for abundancy of XyG motifs in barley, so future studies should be aimed for percentage of these compositions, either.

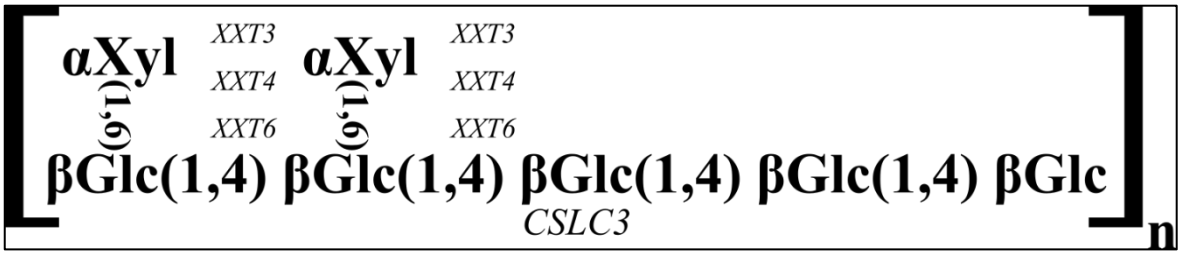


Figure 4.1. A demonstrated visual of "XXGGG" composition, could be the most abundant xyloglucan oligosaccharide composition one among all tissue and time intervals in barley, with name of genes that possibly involve and co-express for its biosynthesis, at highest level and possibility

5. CONCLUSION

In this study, to better understand the XyG biosynthesis in barley, a group of gene families were researched and their co-expression analyses were made. Assuming that these genes co-express to each other to provide xyloglucan substrate for the cell wall construction, so it might be expected that there should be a strong relation for each other's expression patterns. In addition, Pearson Correlation was examined on calculated FPKM values that showing expression levels from public RNA-seq data. Not only mRNA level expression analyses, also protein identification and multiple sequence alignment were done to analyze relatedness level of each other.

In this study, expression patterns of CSLC genes in barley were examined, to understand the process of XyG biosynthesis in barley. CSLCs are taking primary role in XyG biosynthesis, so their contribution is important. These HvCSLCs might co-express with a group of genes in significant and strong level, according to statistical analyses and expression calculations. Furthermore, HvCSLCs are also affected to each other in these expression patterns such as HvCSLC1 co-expresses with HvCSLC4, and HvCSLC4 could also co-express with HvCSLC2. On the other hand, these 3 genes were expressed in important level in different tissues, comparing with other tissues, but these tissues were collected from 4-6 weeks seedlings, so that time interval might be very important for analyzing XyG biosynthesis in barley.

In future studies, the time interval that in between fourth and sixth weeks should be investigated to understand the increasing in XyG biosynthesis in barley. Furthermore, due to inadequate number of studies that were worked on barley's genome and its proteins, researches were stayed at *in silico* level in this study, so more detailed protein analyzing and expanded genome sequencing should be done or researched for more. On the other hand, this study was covering the process of a "polysaccharide" analyzing with mRNA readings and protein sequences *in silico* way. That's why, also carbohydrate-based analyses should be performed in future for better understanding the biosynthesis of XyG.

REFERENCES

1. Baik B-K, Ullrich SE. Barley for food: Characteristics, improvement, and renewed interest. J Cereal Sci. 2008;48(2):233–42.

2. Newman CW, Newman RK. A brief history of barley foods. Cereal Foods World. 2006;51(1):4–7.

3. Behall KM, Scholfield DJ, Hallfrisch J. Diets containing barley significantly reduce lipids in mildly hypercholesterolemic men and women. Am J Clin Nutr. 2004;80(5):1185–93.

4. Qureshi AA, Qureshi N, Wright JJ, Shen Z, Kramer G, Gapor A, et al. Lowering of serum cholesterol in hypercholesterolemic humans by tocotrienols (palmvitee). Am J Clin Nutr. 1991;53(4 Suppl):1021S-1026S.

5. Crops and livestock products: FAOSTAT Visualize Data; [cited 2021 1 October]. Available from http://www.fao.org/faostat/en/#data/QCL/visualize

6. Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, et al. On the origin and domestication history of Barley (Hordeum vulgare). Mol Biol Evol. 2000;17(4):499–510.

7. Cosgrove DJ. Growth of the plant cell wall. Nat Rev Mol Cell Biol. 2005;6(11):850–61.

8. Carpita NC, Defernez M, Findlay K, Wells B, Shoue DA, Catchpole G, et al. Cell wall architecture of the elongating maize coleoptile. Plant Physiol. 2001;127(2):551–65.

9. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, et al. Toward a systems approach to understanding plant cell walls. Science. 2004;306(5705):2206–11.

10. Popper ZA. Evolution and diversity of green plant cell walls, Curr. Opin Plant Biol. 2008;11:286–292,.

11. Burton RA, Fincher GB. Plant cell wall engineering: applications in biofuel production and improved human health. Curr Opin Biotechnol. 2014;26:79–84.

12. Cosgrove DJ. Wall extensibility: its nature, measurement and relationship to plant cell growth. New Phytol. 1993;124(1):1–23.

13. Smith BG, Harris PJ. The Polysaccharide Composition of Poales Cell Walls: Poaceae Cell Walls Are Not Unique. Biochemical Systematics and Ecology. 1999;27:33–53,.

14. Cassab GI. Plant cell wall proteins. Annu Rev Plant Physiol Plant Mol Biol. 1998;49(1):281–309.

15. Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol. 2008;11(3):301–7.

16. Saxena IM, Brown RM Jr. Biosynthesis of Cellulose. In: Progress in Biotechnology. Elsevier; 2001. p. 69–76.

17. Römling U. Molecular biology of cellulose production in bacteria. Res Microbiol. 2002;153(4):205–12.

18. Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV. Arabidopsis - a powerful model system for plant cell wall research. Plant J. 2010;61(6):1107–21.

19. Taiz L, Zeiger E. Plant physiology and plant cells. 3rd ed. 2002. p. 319,.

20. Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB. The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 2004;134(1):224–36.

21. Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA. 2003;100(3):1450–5.

22. Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM Jr. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant vigna angularis. Plant Cell. 1999;11(11):2075–86.

23. Richmond TA, Somerville CR. The cellulose synthase superfamily. Plant Physiol. 2000;124(2):495–8.

24. Liepman AH, Wilkerson CG, Keegstra K. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc Natl Acad Sci U S A. 2005;102(6):2221–6.

25. Burton RA, Jobling SA, Harvey AJ, Shirley NJ, Mather DE, Bacic A, et al. The genetics and transcriptional profiles of the cellulose synthase-like HvCslF gene family in barley. Plant Physiol. 2008;146(4):1821–33.

26. Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, et al. A Gene from the Cellulose Synthase-Like C Family Encodes a β-(1,4) Glucan Synthase. Proceedings of the National Academy of Sciences. 2007;104:8550–8555.

27. Bernal AJ, Jensen JK, Harholt J, Sørensen S, Moller I, Blaukopf C, et al. Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis. Plant J. 2007;52(5):791–802.

28. Fincher GB. Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. Plant Physiol. 2009;149(1):27–37.

29. Buckeridge MS, Rayon C, Urbanowicz B, Tiné MAS, Carpita NC. Mixed Linkage (1→3),(1→4)-β- d -Glucans of Grasses. Cereal Chem. 2004;81:115–127,.

30. Simmons TJ, Uhrín D, Gregson T, Murray L, Sadler IH, Fry SC. An unexpectedly lichenase-stable hexasaccharide from cereal, horsetail and lichen mixed-linkage β-glucans (MLGs): implications for MLG subunit distribution. Phytochemistry. 2013;95:322–32.

31. Kofuji K, Aoki A, Tsubaki K, Konishi M, Isobe T, Murata Y. Antioxidant activity of β-glucan. ISRN Pharm. 2012;2012:125864.

32. Inouhe M, Nevins DJ. Auxin-enhanced glucan autohydrolysis in maize coleoptile cell walls. Plant Physiol. 1991;96(1):285–90.

33. Pettolino F, Sasaki I, Turbic A, Wilson SM, Bacic A, Hrmova M, et al. Hyphal Cell Walls from the Plant Pathogen Rhynchosporium Secalis Contain (1, 3/1, 6)-β-D-Glucans, Galacto- and Rhamnomannans, (1,3;1,4)-β-D-Glucans and Chitin. Febs Journal. 2009;276:3698–3709.

34. Carpita NC, Mccann MC. The Maize Mixed-Linkage (1→ 3),(1→ 4)-β-DGlucan Polysaccharide Is Synthesized at the Golgi Membrane. Plant Physiology. 2010;153(1362):1371,.

35. Fry SC, Mohler KE, Nesselrode BH, Frankova L. Mixed-Linkage Beta-Glucan :Xyloglucan Endotransglucosylase, a Novel Wall-Remodelling Enzyme From Equisetum (Horsetails) and Charophytic Algae. Plant Journal. 2008;55:240–252.

36. Sørensen I, Pettolino FA, Wilson SM, Doblin MS, Johansen B, Bacic A, et al. Mixed-linkage (1,3),(1,4)-b-D-glucan is not unique to the Poales and is an abundant component of Equisetum arvense cell walls. Plant J. 2008;54:510–521,.

37. Coomey JH, Sibout R, Hazen SP. Grass secondary cell walls, Brachypodium distachyon as a model for discovery. New Phytol. 2020;227(6):1649–67.

38. Kim S-J, Zemelis S, Keegstra K, Brandizzi F. The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan. Plant J. 2015;81(4):537–47.

39. Vega-Sánchez ME, Verhertbruggen Y, Christensen U, Chen X, Sharma V, Varanasi P, et al. Loss of Cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. Plant Physiol. 2012;159(1):56–69.

40. Rose JKC, Lee S-J. Straying off the highway: trafficking of secreted plant proteins and complexity in the plant cell wall proteome. Plant Physiol. 2010;153(2):433–6.

41. Viëtor RJ, Angelino SAGF, Voragen AGJ. Structural features of arabinoxylans from barley and malt cell wall material. J Cereal Sci. 1992;15(3):213–22.

42. Rennie EA, Scheller HV. Xylan biosynthesis. Curr Opin Biotechnol. 2014;26:100–7.

43. Wu A-M, Rihouey C, Seveno M, Hörnblad E, Singh SK, Matsunaga T, et al. The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. Plant J. 2009;57(4):718–31.

44. Jeoh T, Ishizawa CI, Davis MF, Himmel ME, Adney WS, Johnson DK. Cellulase digestibility of pretreated biomass is limited by cellulose accessibility, Biotechnol. Bioeng. 2007;98:112–122,.

45. Yang B, Wyman CE. Effect of Xylan and Lignin Removal by Batch and Flowthrough Pretreatment on the Enzymatic Digestibility of Corn Stover Cellulose, Biotechnol. Bioeng. 2004;86:88–95,.

46. Zhang H, Nishinari K, Williams MAK, Foster TJ, Norton IT. A molecular description of the Gelation Mechanism of Konjac Mannan. Int J Biol Macromol. 2002;30:7–16,.

47. Manna S, McAnalley BH. Determination of the position of the O-acetyl group in a β-(1 → 4)-mannan (acemannan) from Aloe barbardensis Miller, Carbohydr. Res. 1993;241:317–319,.

48. Moreira LRS, Filho EXF. An overview of mannan structure and mannan-degrading enzyme systems. Appl Microbiol Biotechnol. 2008;79(2):165–78.

49. Schröder R, Wegrzyn TF, Sharma NN, Atkinson RG. Leman4 Endo-B- Mannanase from Ripe Tomato Fruit Can Act as A Mannan Transglycosylase or Hydrolase. Planta. 2006;224:1091–1102.

50. Goubet F, Barton CJ, Mortimer JC, Yu X, Zhang Z, Miles GP, et al. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J. 2009;60(3):527–38.

51. da Cruz A. Mannan-Degrading Enzyme System. Fungal Enzymes. CRC Press; 2013.

52. O'neill MA, Ishii T, Albersheim P, Darvill AG. Rhamnogalacturonan II: Structure and Function of A Borate Cross-Linked Cell Wall Pectic Polysaccharide. Annual Review of Plant Biology. 2004;55:109–139.

53. Ridley BL, O'Neill MA, Mohnen D. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry. 2001;57(6):929–67.

54. Fraeye I, Colle I, Vandevenne E, Duvetter T, Van Buggenhout S, Moldenaers P, et al. Influence of pectin structure on texture of pectin–calcium gels. Innov Food Sci Emerg Technol. 2010;11(2):401–9.

55. Jackson CL, Dreaden TM, Theobald LK, Tran NM, Beal TL, Eid M, et al. Pectin induces apoptosis in human prostate cancer cells: correlation of apoptotic function with pectin structure. Glycobiology. 2007;17(8):805–19.

56. Hsieh YSY, Harris PJ. Xyloglucans of monocotyledons have diverse structures, Mol. Plant. 2009;2:943–965,.

57. Peña MJ, Darvill AG, Eberhard S, York WS, O'Neill MA. Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. Glycobiology. 2008;18(11):891–904.

58. Rose JKC, Braam J, Fry SC, Nishitani K. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol. 2002;43(12):1421–35.

59. Hoffman M, Jia Z, Peña MJ, Cash M, Harper A, Blackburn AR 2nd, et al. Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. Carbohydr Res. 2005;340(11):1826–40.

60. Lima DU, Loh W, Buckeridge MS. Xyloglucan-cellulose interaction depends on the sidechains and molecular weight of xyloglucan. Plant Physiol Biochem. 2004;42(5):389–94.

61. Lopez M, Bizot H, Chambat G, Marais M-F, Zykwinska A, Ralet M-C, et al. Enthalpic studies of xyloglucan-cellulose interactions. Biomacromolecules. 2010;11(6):1417–28.

62. Hayashi T, Ogawa K, Mitsuishi Y. Characterization of the adsorption of xyloglucan to cellulose. Plant Cell Physiol. 1994;35(8):1199–205.

63. Thompson JE, Fry SC. Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells: Detection of xyloglucan restructuring in vivo. Plant J. 2001;26(1):23–34.

64. Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau J-P, et al. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. Physiol Plant. 1993;89(1):1–3.

65. Tuomivaara ST, Yaoi K, O'Neill MA, York WS. Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. Carbohydr Res. 2015;402:56–66.

66. Koehler P, Wieser H. Chemistry of cereal grains. Handbook on Sourdough Biotechnology. Boston, MA: Springer US; 2013. p. 11–45.

67. Gibeaut DM, Pauly M, Bacic A, Fincher GB. Changes in cell wall polysaccharides in developing barley (Hordeum vulgare) coleoptiles. Planta. 2005;221(5):729–38.

68. Obel N, Erben V, Schwarz T, Kühnel S, Fodor A, Pauly M. Microanalysis of plant cell wall polysaccharides, Mol. Plant. 2009;2:922–932,.

69. Hrmova M, Farkas V, Harvey AJ, Lahnstein J, Wischmann B, Kaewthai N, et al. Substrate specificity and catalytic mechanism of a xyloglucan xyloglucosyl transferase HvXET6 from barley (Hordeum vulgare L.): Substrate specificity and catalytic mechanism of HvXET6. FEBS J. 2009;276(2):437–56.

70. Schantz L, Gullfot F, Scheer S, Filonova L, Gunnarsson LC, Flint J, et al. Affinity maturation generates greatly improved xyloglucan-specific carbohydrate binding modules. BMC Biotechnol. 2009;9:92,.

71. Keegstra K, Raikhel N. Plant glycosyltransferases. Curr Opin Plant Biol. 2001;4(3):219–24.

72. Zabotina O. Xyloglucan and Its Biosynthesis, Front. Plant Sci. 2012;3:2008–2012,.

73. Burton RA, Fincher GB. Current challenges in cell wall biology in the cereals and grasses. Front Plant Sci. 2012;3:130.

74. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

75. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012;61(6):1061–7.

76. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.

77. Grech V. WASP (write a scientific paper) using excel – 13: Correlation and regression. Early Hum Dev. 2018;122:60–3.

78. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006;7 Suppl 1(Suppl 1):S10.1-12.

79. Daras G, Templalexis D, Avgeri F, Tsitsekian D, Karamanou K, Rigas S. Updating insights into the catalytic domain properties of plant cellulose synthase (CesA) and cellulose synthase-like (csl) proteins. Molecules. 2021;26(14):4335.

80. Liu L, Hsia MM, Dama M, Vogel J, Pauly M. A xyloglucan backbone 6-O-acetyltransferase from Brachypodium distachyon modulates xyloglucan xylosylation. Mol Plant. 2016;9(4):615–7.

81. Moore DS, Notz WI. The basic practice of statistics. 9th ed. Macmillan International; 2021.

82. Dwivany FM, Yulia D, Burton RA, Shirley NJ, Wilson SM, Fincher GB, et al. The CELLULOSE-SYNTHASE LIKE C (CSLC) family of barley includes members that are integral membrane proteins targeted to the plasma membrane. Mol Plant. 2009;2(5):1025–39.

83. Kim S-J, Chandrasekar B, Rea AC, Danhof L, Zemelis-Durfee S, Thrower N, et al. The synthesis of xyloglucan, an abundant plant cell wall polysaccharide, requires CSLC function. Proc Natl Acad Sci USA. 2020;117(33):20316–24.

84. Culbertson AT, Chou Y-H, Smith AL, Young ZT, Tietze AA, Cottaz S, et al. Enzymatic activity of xyloglucan xylosyltransferase 5. Plant Physiol. 2016;171(3):1893–904.

85. Mercx S, Smargiasso N, Chaumont F, De Pauw E, Boutry M, Navarre C. Inactivation of the β(1,2)-xylosyltransferase and the α(1,3)-fucosyltransferase genes in Nicotiana tabacum BY-2 Cells by a Multiplex CRISPR/Cas9 Strategy Results in Glycoproteins without Plant-Specific Glycans. Front Plant Sci. 2017;8:403.

86. Madson M, Dunand C, Li X, Verma R, Vanzin GF, Caplan J, et al. The MUR3 gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. Plant Cell. 2003;15(7):1662–70.

87. Li X, Cordero I, Caplan J, Mølhøj M, Reiter W-D. Molecular analysis of 10 coding regions from Arabidopsis that are homologous to the MUR3 xyloglucan galactosyltransferase. Plant Physiol. 2004;134(3):940–50.

88. Jensen JK, Schultink A, Keegstra K, Wilkerson CG, Pauly M. RNA-Seq analysis of developing nasturtium seeds (Tropaeolum majus): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis. Mol Plant. 2012;5(5):984–92.

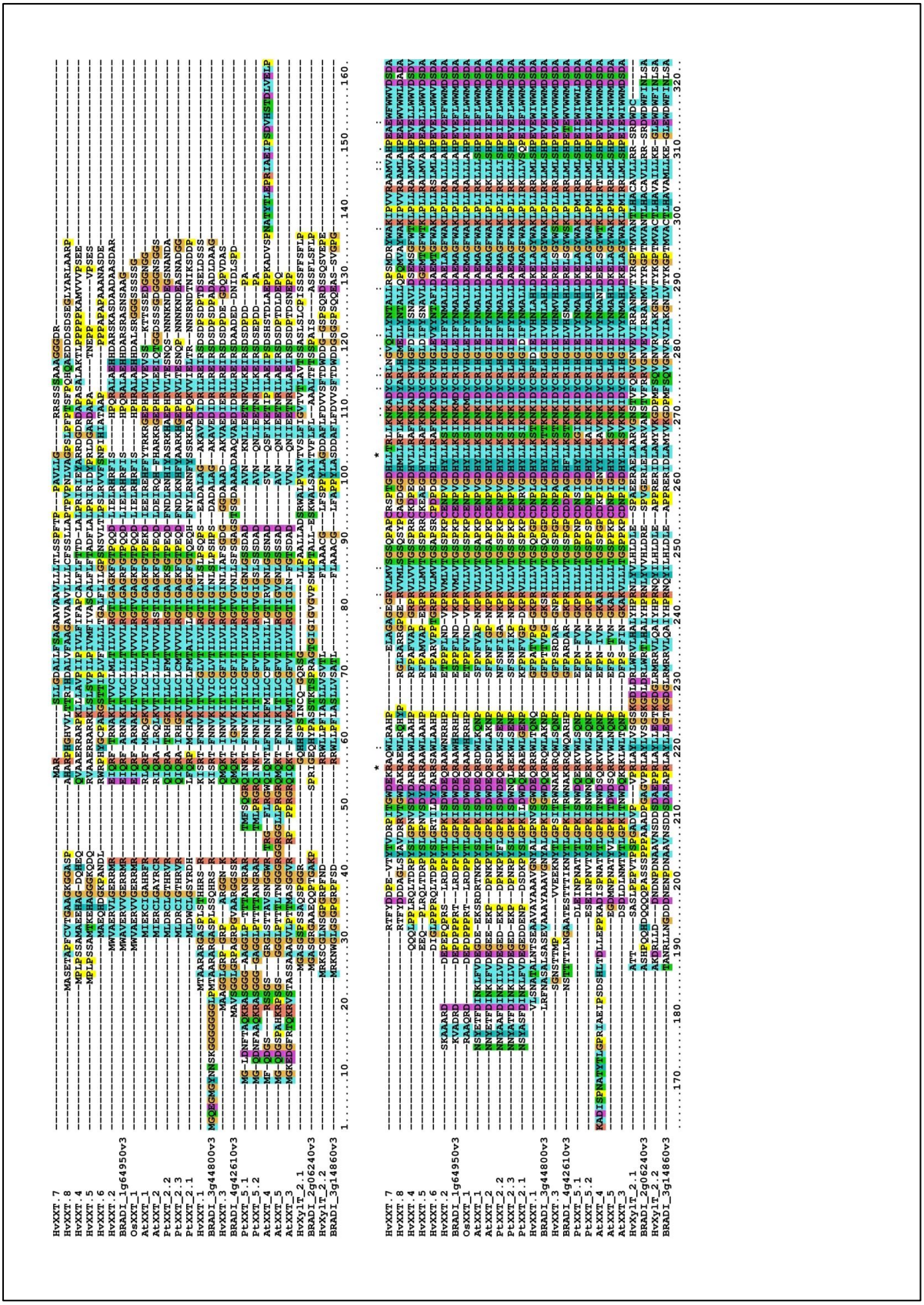
89. Mukherjee S, Hao Y-H, Orth K. A newly discovered post-translational modification--the acetylation of serine and threonine residues. Trends Biochem Sci. 2007;32(5):210–6.

90. Zhong R, Cui D, Phillips DR, Richardson EA, Ye Z-H. A group of O-acetyltransferases catalyze xyloglucan backbone acetylation and can alter xyloglucan xylosylation pattern and plant growth when expressed in Arabidopsis. Plant Cell Physiol. 2020;61(6):1064–79.

91. Gille S, Pauly M. O-acetylation of plant cell wall polysaccharides. Front Plant Sci. 2012;3:12.

92. Spring barley growth and development guide [Internet]. Umn.edu. [cited 2021 Dec 21]. Available from: https://extension.umn.edu/growing-small-grains/spring-barley-growth-and-development-guide

APPENDIX A: OTHER MSA RESULTS



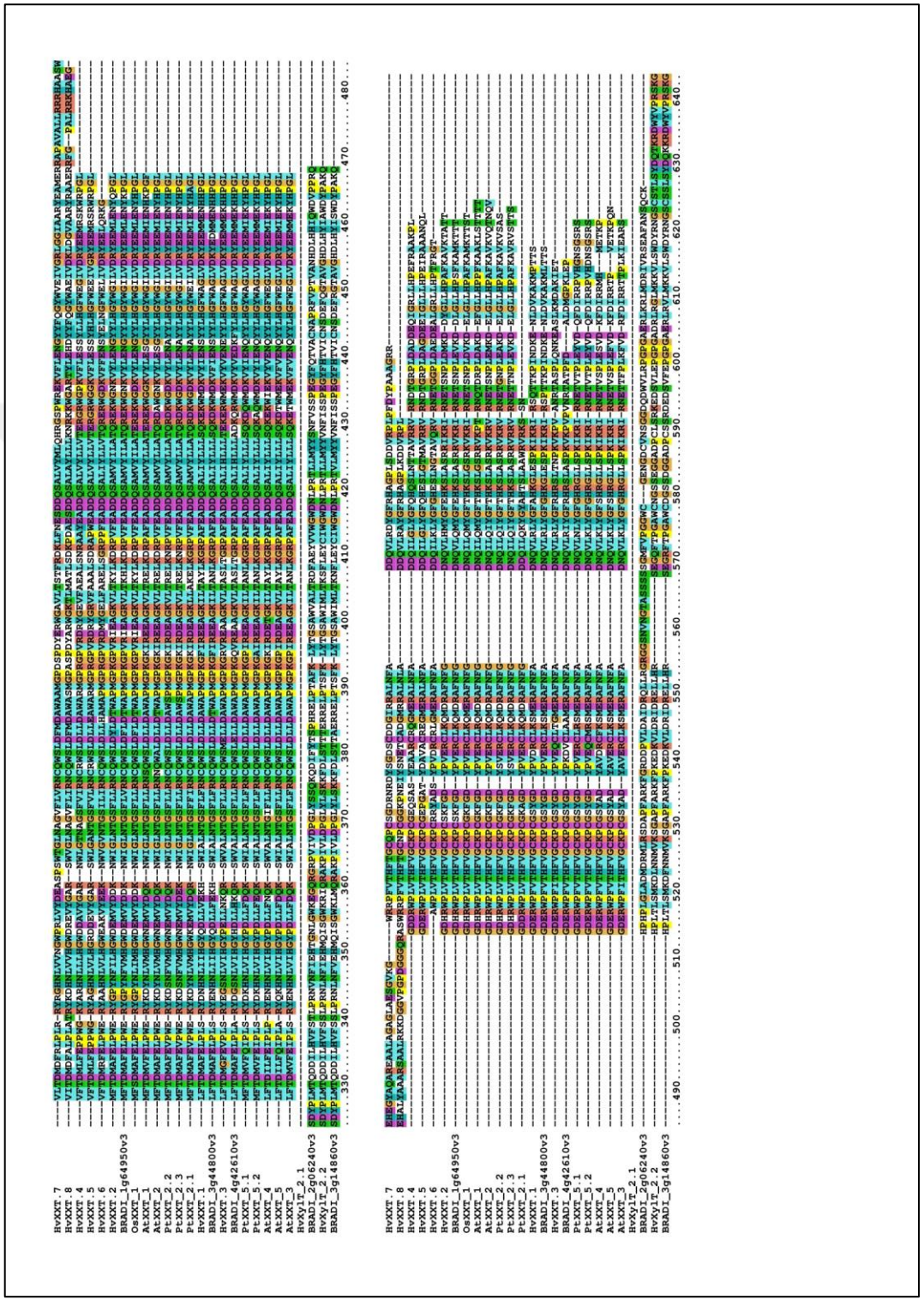
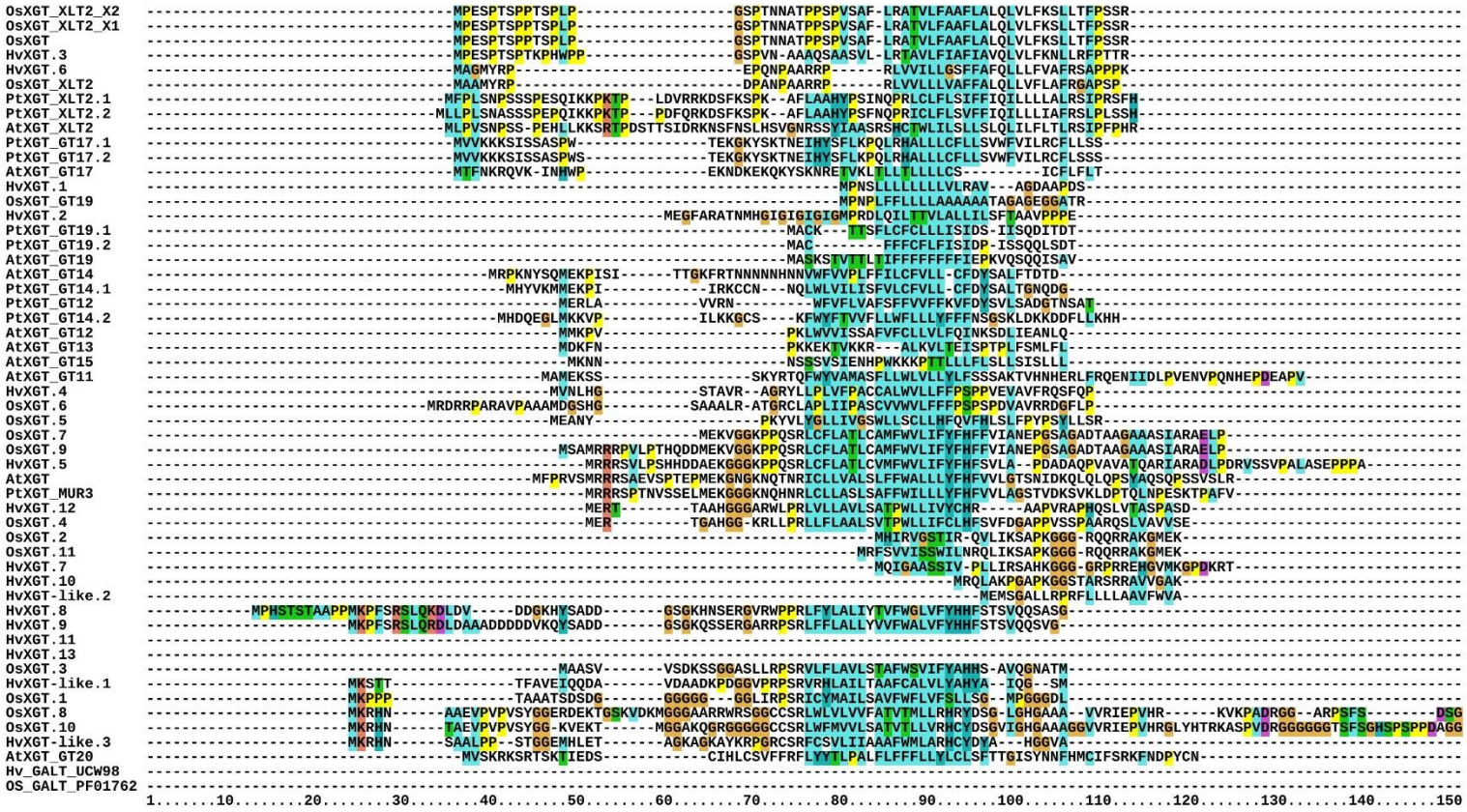
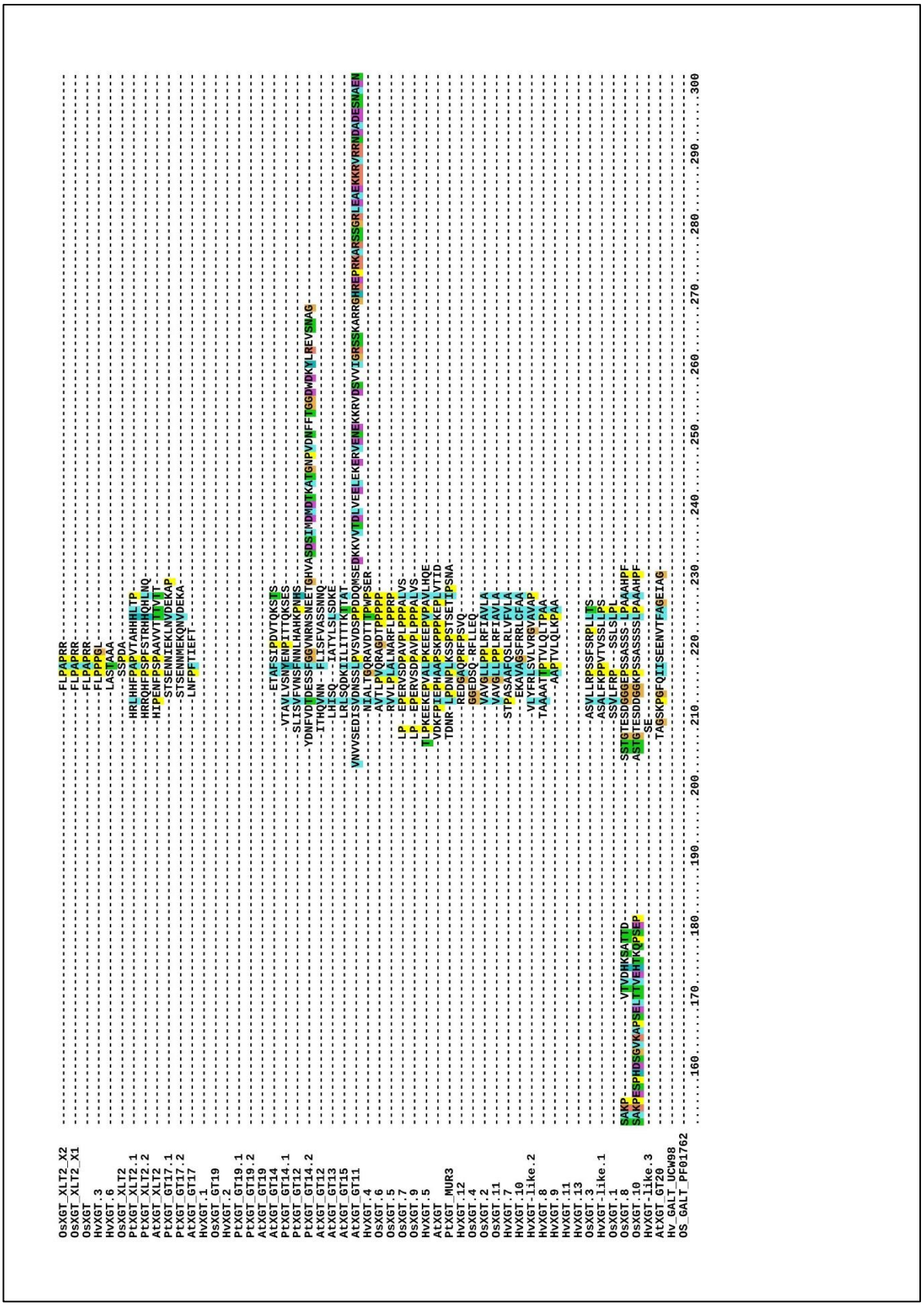
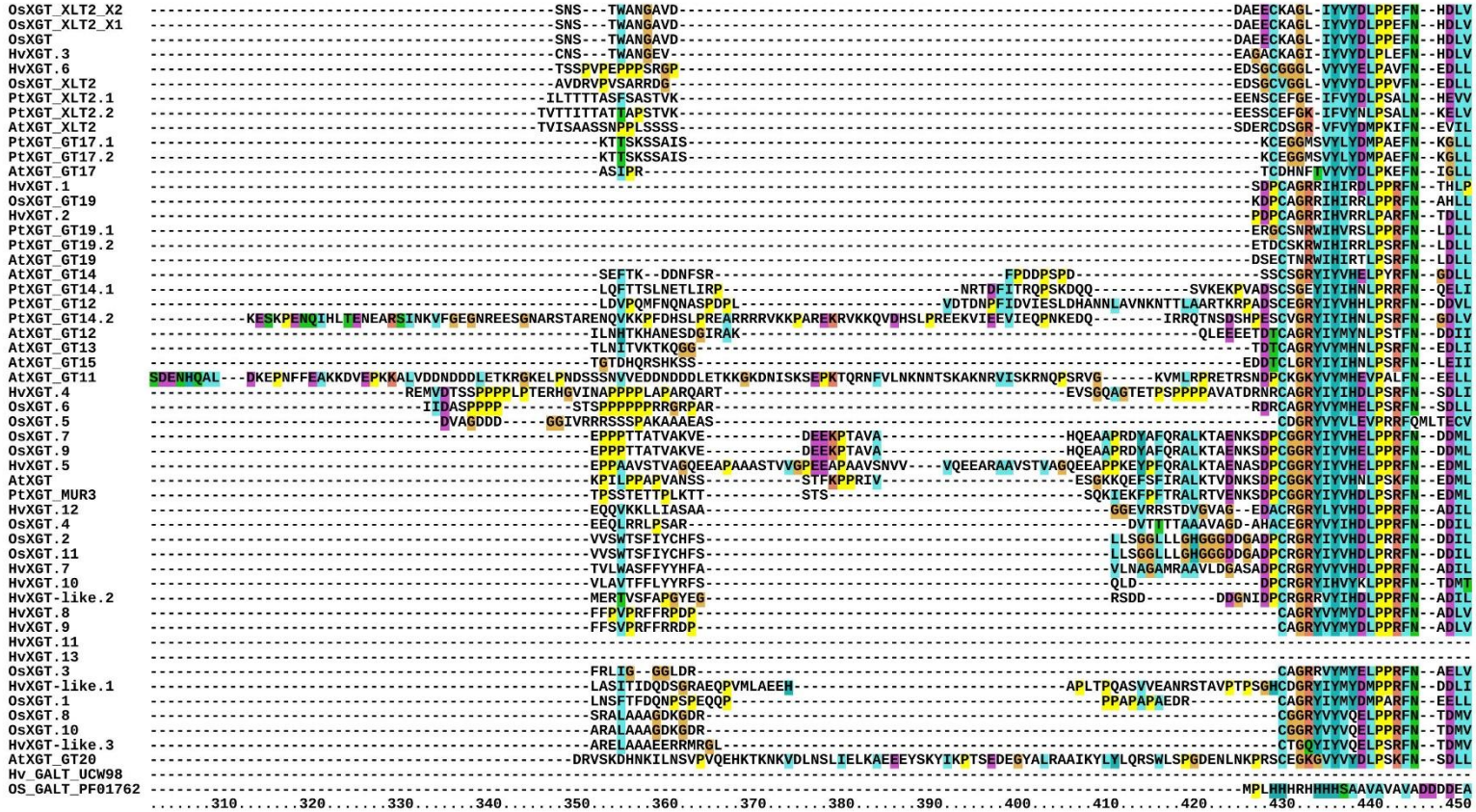
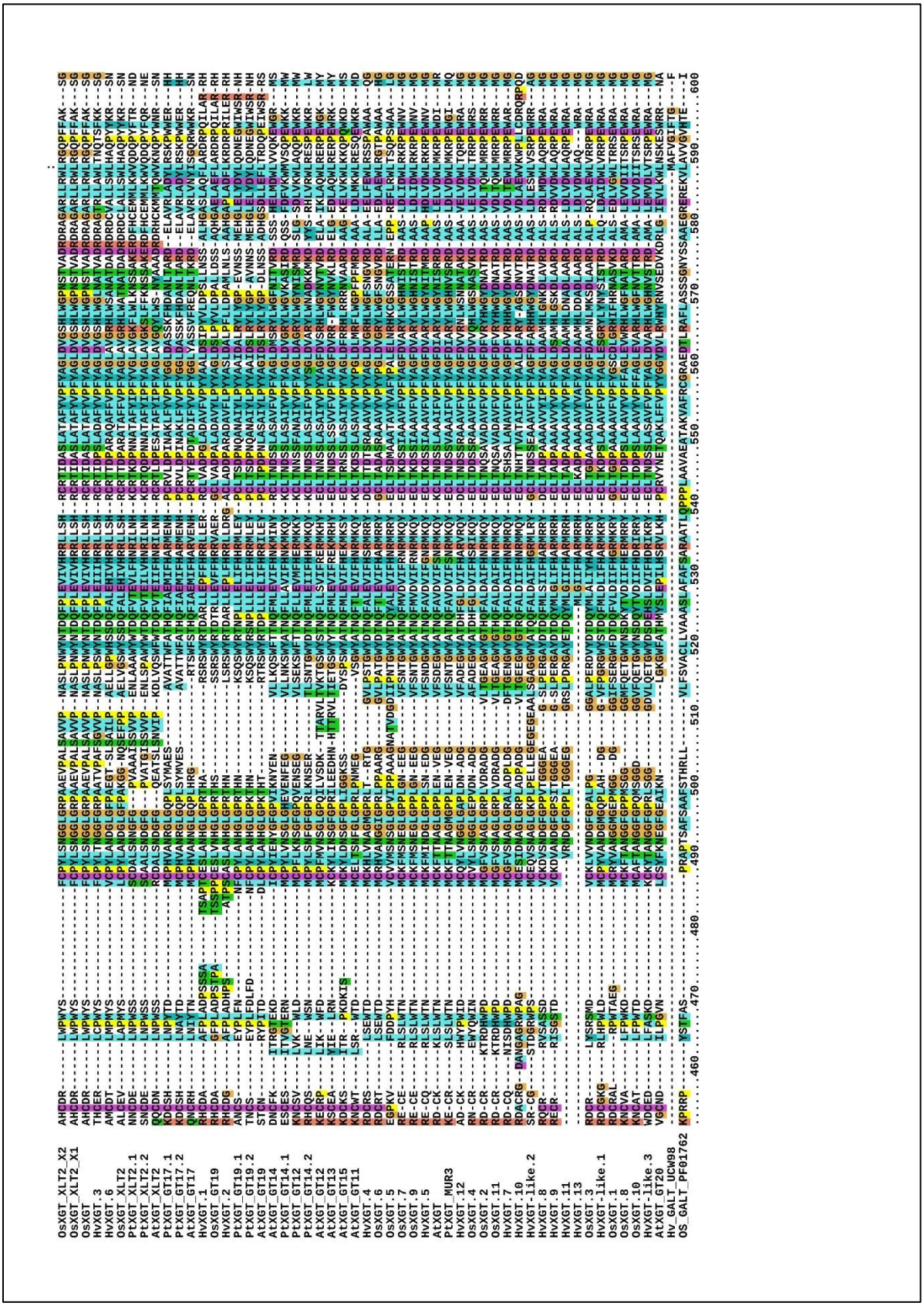


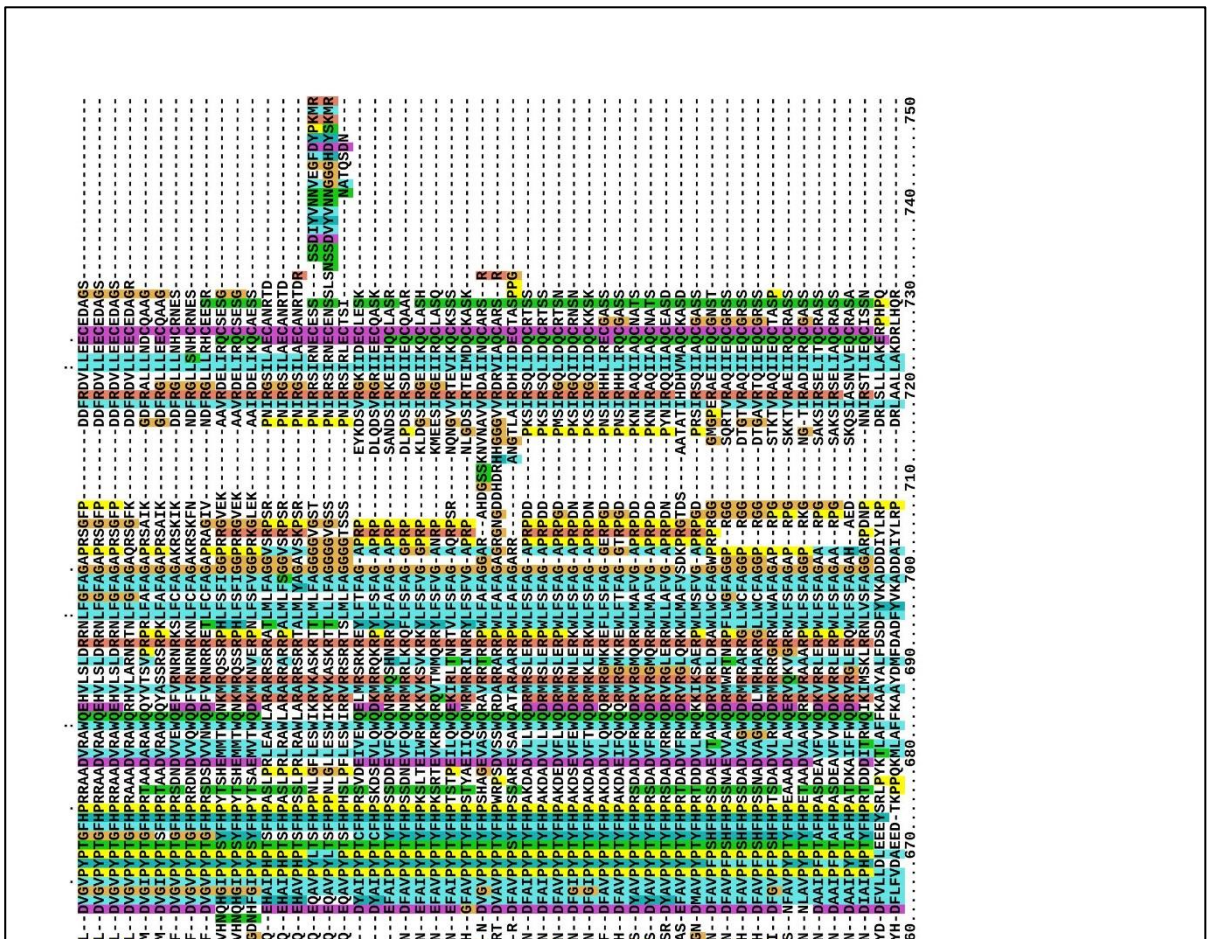
Figure A.1. Multiple sequence alignment results of xyloglucan xylosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon*

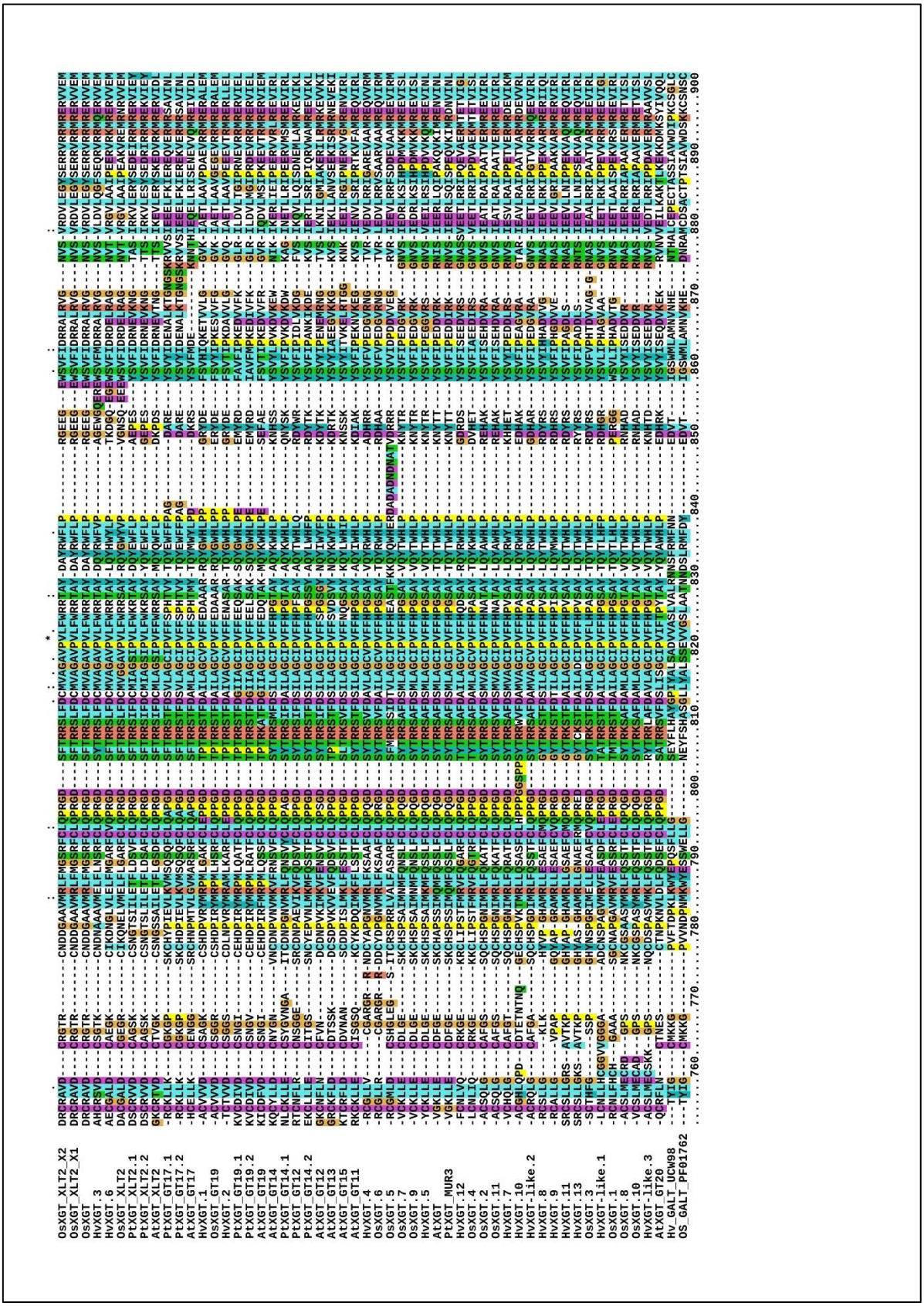

OsXGT_XLT2_X2
OsXGT_XLT2_X1
OsXGT
HvXGT.3
HvXGT.6
OsXGT_XLT2
PtXGT_XLT2.1
PtXGT_XLT2.2
AtXGT_XLT2
PtXGT_GT17.1
PtXGT_GT17.2
AtXGT_GT17
HvXGT.1
OsXGT_GT19
HvXGT.2
PtXGT_GT19.1
PtXGT_GT19.2
AtXGT_GT19
AtXGT_GT14
PtXGT_GT14.1
PtXGT_GT12
PtXGT_GT14.2
AtXGT_GT12
AtXGT_GT13
AtXGT_GT15
AtXGT_GT11
HvXGT.4
OsXGT.6
OsXGT.5
OsXGT.7
OsXGT.9
HvXGT.5
AtXGT
PtXGT_MUR3
HvXGT.12
OsXGT.4
OsXGT.2
OsXGT.11
HvXGT.7
HvXGT.10
HvXGT-like.2
HvXGT.8
HvXGT.9
HvXGT.11
HvXGT.13
OsXGT.3
HvXGT-like.1
OsXGT.1
OsXGT.8
OsXGT.10
HvXGT-like.3
AtXGT_GT20
Hv_GALT_UCW98
OS_GALT_PF01762
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150

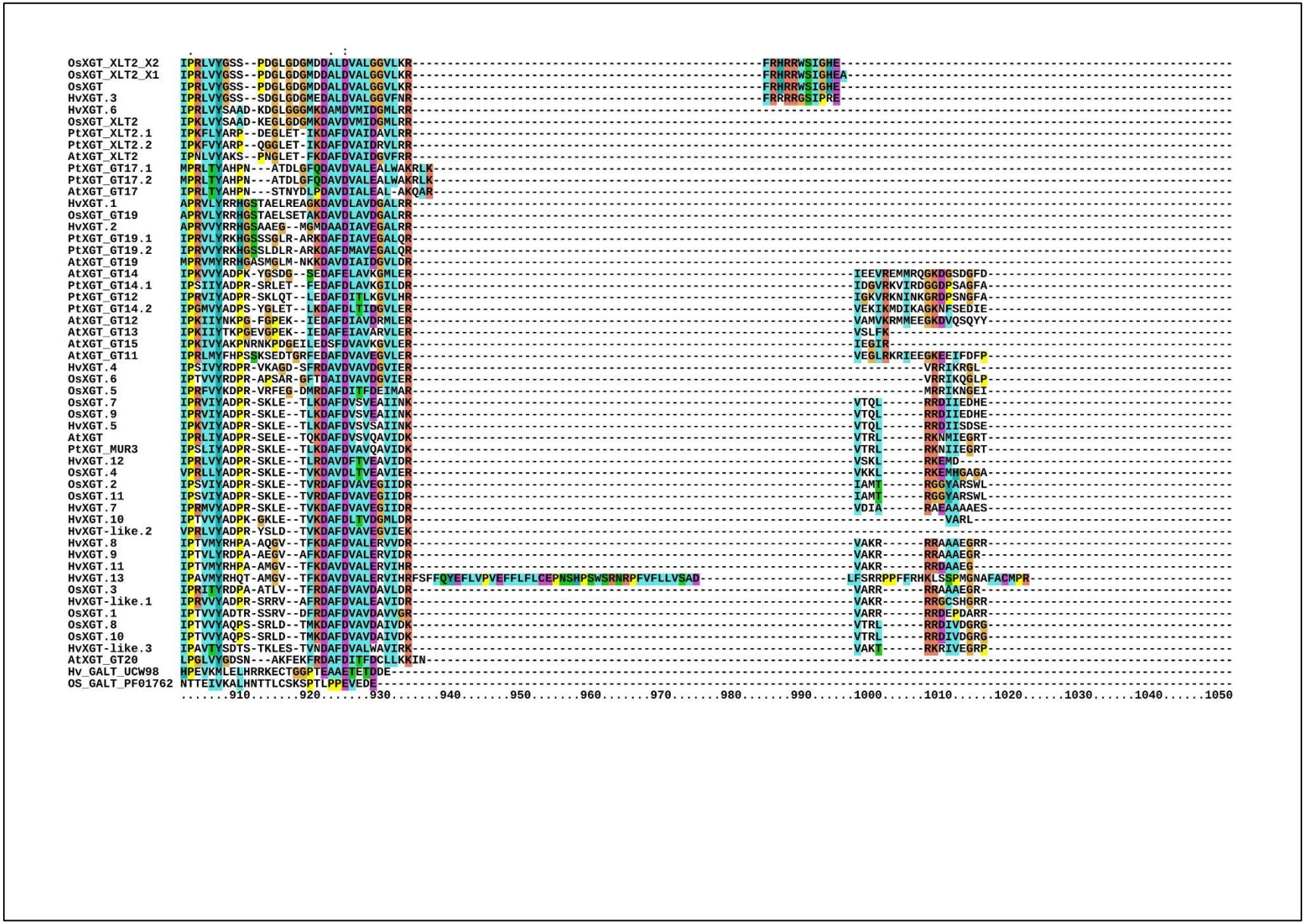

OsXGT_XLT2_X2
OsXGT_XLT2_X1
OsXGT
HvXGT.3
HvXGT.6
OsXGT_XLT2
PtXGT_XLT2.1
PtXGT_XLT2.2
AtXGT_XLT2
PtXGT_GT17.1
PtXGT_GT17.2
AtXGT_GT17
HvXGT.1
OsXGT_GT19
HvXGT.2
PtXGT_GT19.1
PtXGT_GT19.2
AtXGT_GT19
AtXGT_GT14
PtXGT_GT14.1
PtXGT_GT12
PtXGT_GT14.2
AtXGT_GT12
AtXGT_GT13
AtXGT_GT15
AtXGT_GT11
HvXGT.4
OsXGT.6
OsXGT.5
OsXGT.7
OsXGT.9
HvXGT.5
AtXGT
PtXGT_MUR3
HvXGT.12
OsXGT.4
OsXGT.2
OsXGT.11
HvXGT.7
HvXGT.10
HvXGT-like.2
HvXGT.8
HvXGT.9
HvXGT.11
HvXGT.13
OsXGT.3
HvXGT-like.1
OsXGT.1
OsXGT.8
OsXGT.10
HvXGT-like.3
AtXGT_GT20
Hv_GALT_UCW98
OS_GALT_PF01762


OsXGT_XLT2_X2
OsXGT_XLT2_X1
OsXGT
HvXGT.3
HvXGT.6
OsXGT_XLT2
PtXGT_XLT2.1
PtXGT_XLT2.2
AtXGT_XLT2
PtXGT_GT17.1
PtXGT_GT17.2
AtXGT_GT17
HvXGT.1
OsXGT_GT19
HvXGT.2
PtXGT_GT19.1
PtXGT_GT19.2
AtXGT_GT19
AtXGT_GT14
PtXGT_GT14.1
PtXGT_GT12
PtXGT_GT14.2
AtXGT_GT12
AtXGT_GT13
AtXGT_GT15
AtXGT_GT11
HvXGT.4
OsXGT.6
OsXGT.5
OsXGT.7
OsXGT.9
HvXGT.5
AtXGT
PtXGT_MUR3
HvXGT.12
OsXGT.4
OsXGT.2
OsXGT.11
HvXGT.7
HvXGT.10
HvXGT-like.2
HvXGT.8
HvXGT.9
HvXGT.11
HvXGT.13
OsXGT.3
HvXGT-like.1
OsXGT.1
OsXGT.8
OsXGT.10
HvXGT-like.3
AtXGT_GT20
Hv_GALT_UCW98
OS_GALT_PF01762
.....310.....320.....330.....340.....350.....360.....370.....380.....390.....400.....410.....420.....430.....440.....450








OsXGT_XLT2_X2
OsXGT_XLT2_X1
OsXGT
HvXGT.3
HvXGT.6
OsXGT_XLT2
PtXGT_XLT2.1
PtXGT_XLT2.2
AtXGT_XLT2
PtXGT_GT17.1
PtXGT_GT17.2
AtXGT_GT17
HvXGT.1
OsXGT_GT19
HvXGT.2
PtXGT_GT19.1
PtXGT_GT19.2
AtXGT_GT19
AtXGT_GT14
PtXGT_GT14.1
PtXGT_GT12
PtXGT_GT14.2
AtXGT_GT12
AtXGT_GT13
AtXGT_GT15
AtXGT_GT11
HvXGT.4
OsXGT.6
OsXGT.5
OsXGT.7
OsXGT.9
HvXGT.5
AtXGT
PtXGT_MUR3
HvXGT.12
OsXGT.4
OsXGT.2
OsXGT.11
HvXGT.7
HvXGT.10
HvXGT-like.2
HvXGT.8
HvXGT.9
HvXGT.11
HvXGT.13
OsXGT.3
HvXGT-like.1
OsXGT.1
OsXGT.8
OsXGT.10
HvXGT-like.3
AtXGT_GT20
Hv_GALT_UCW98
OS_GALT_PF01762
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050

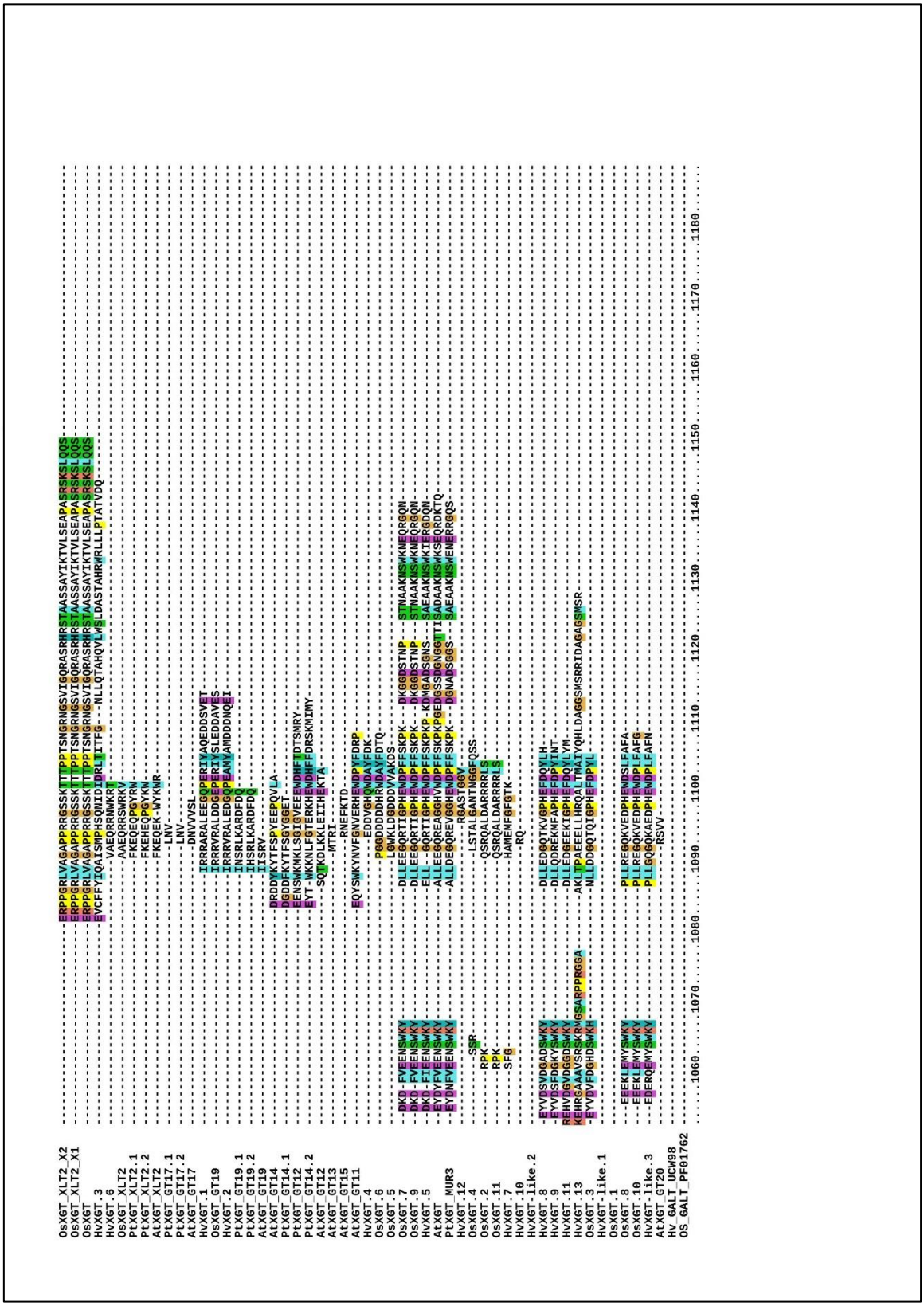
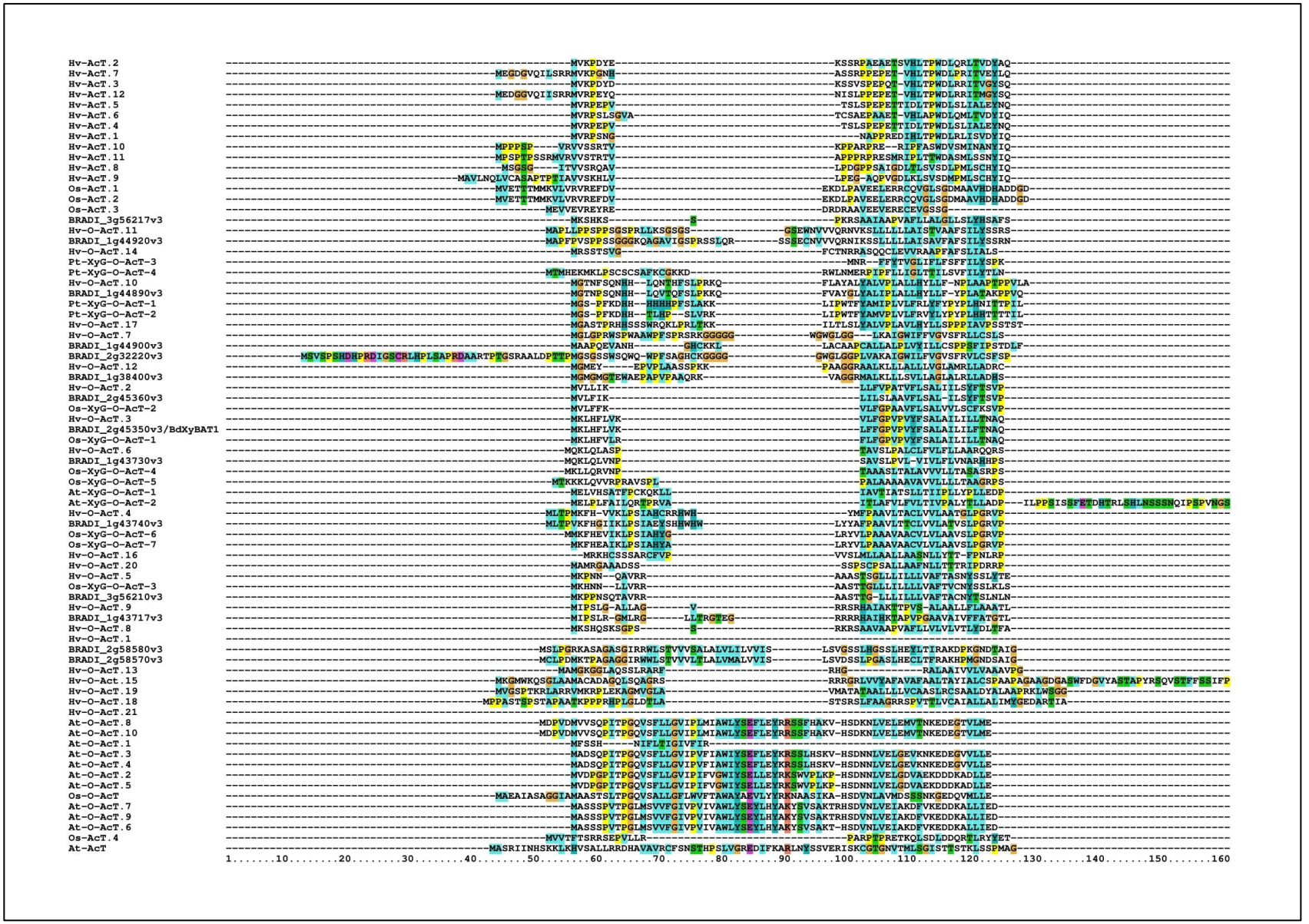
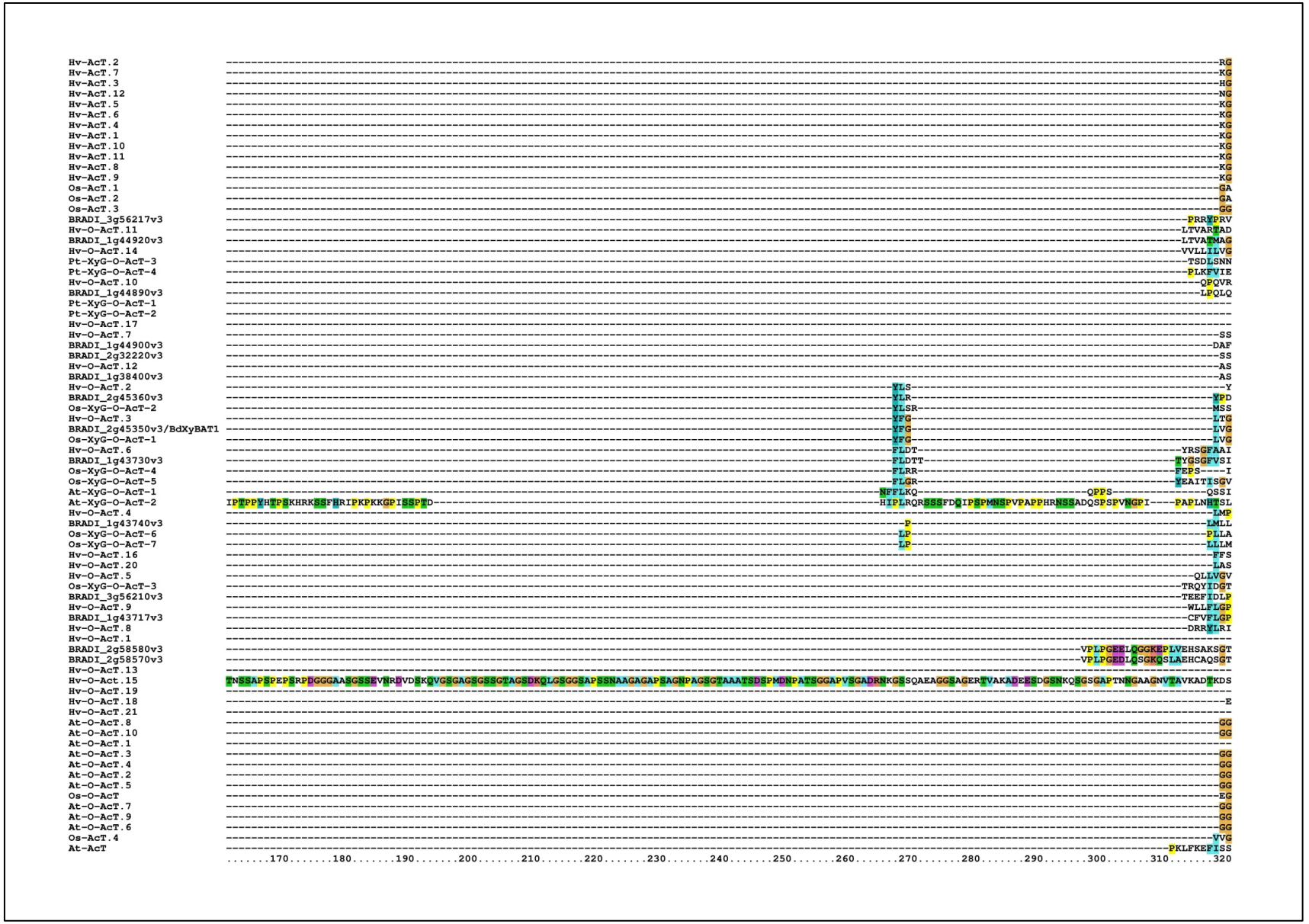
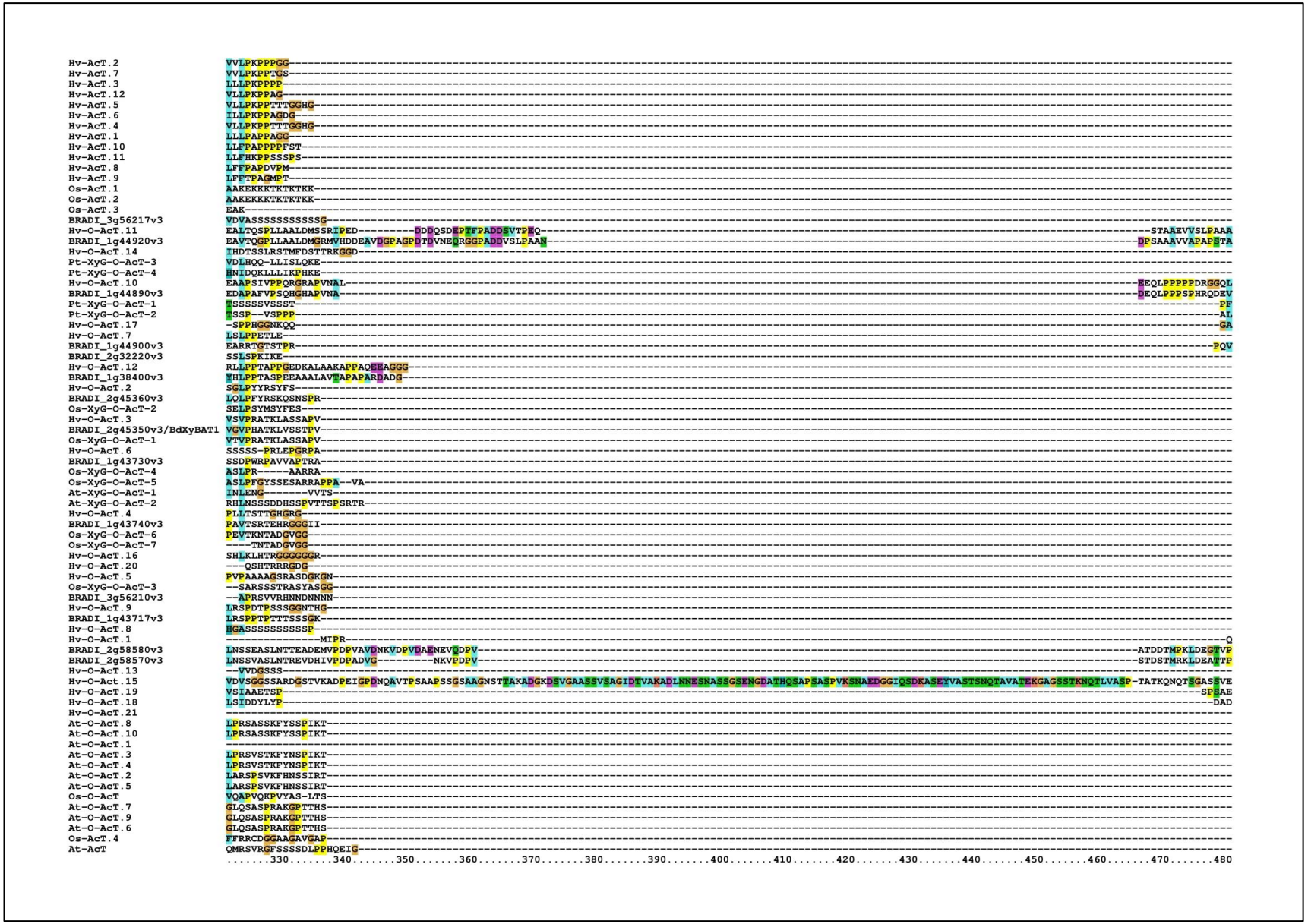
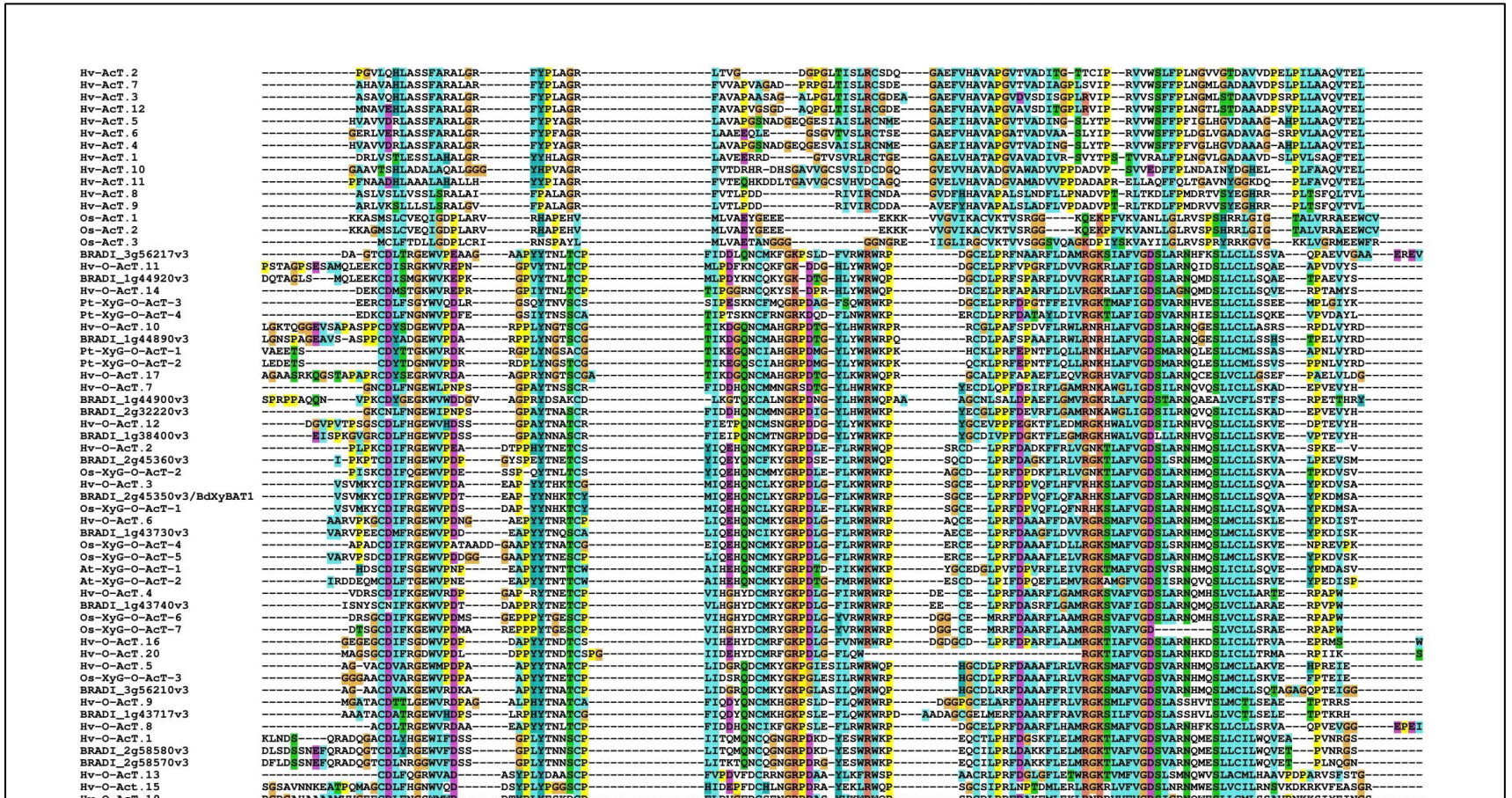


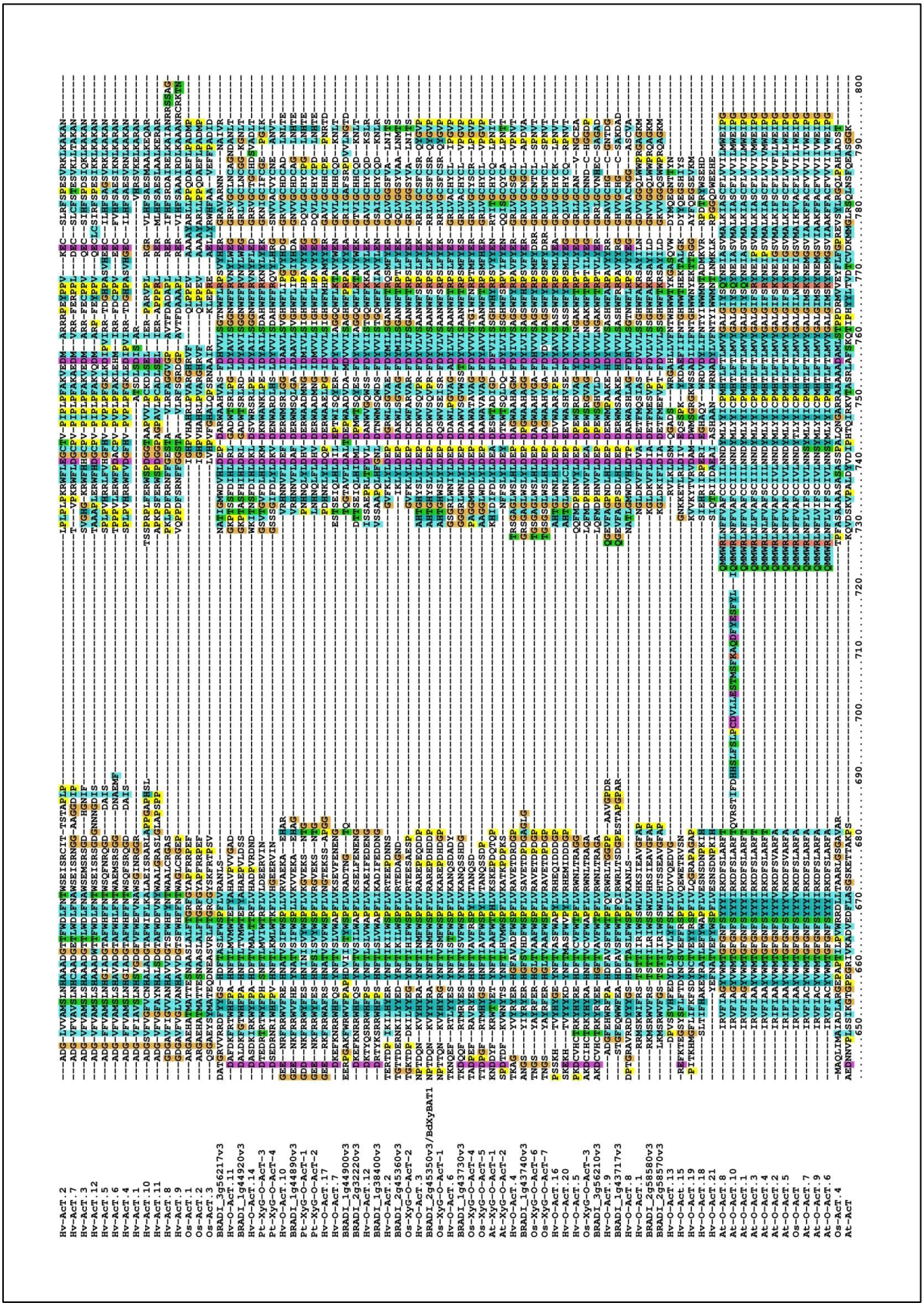
Figure A.2. Multiple sequence alignment results of xyloglucan galactosyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*

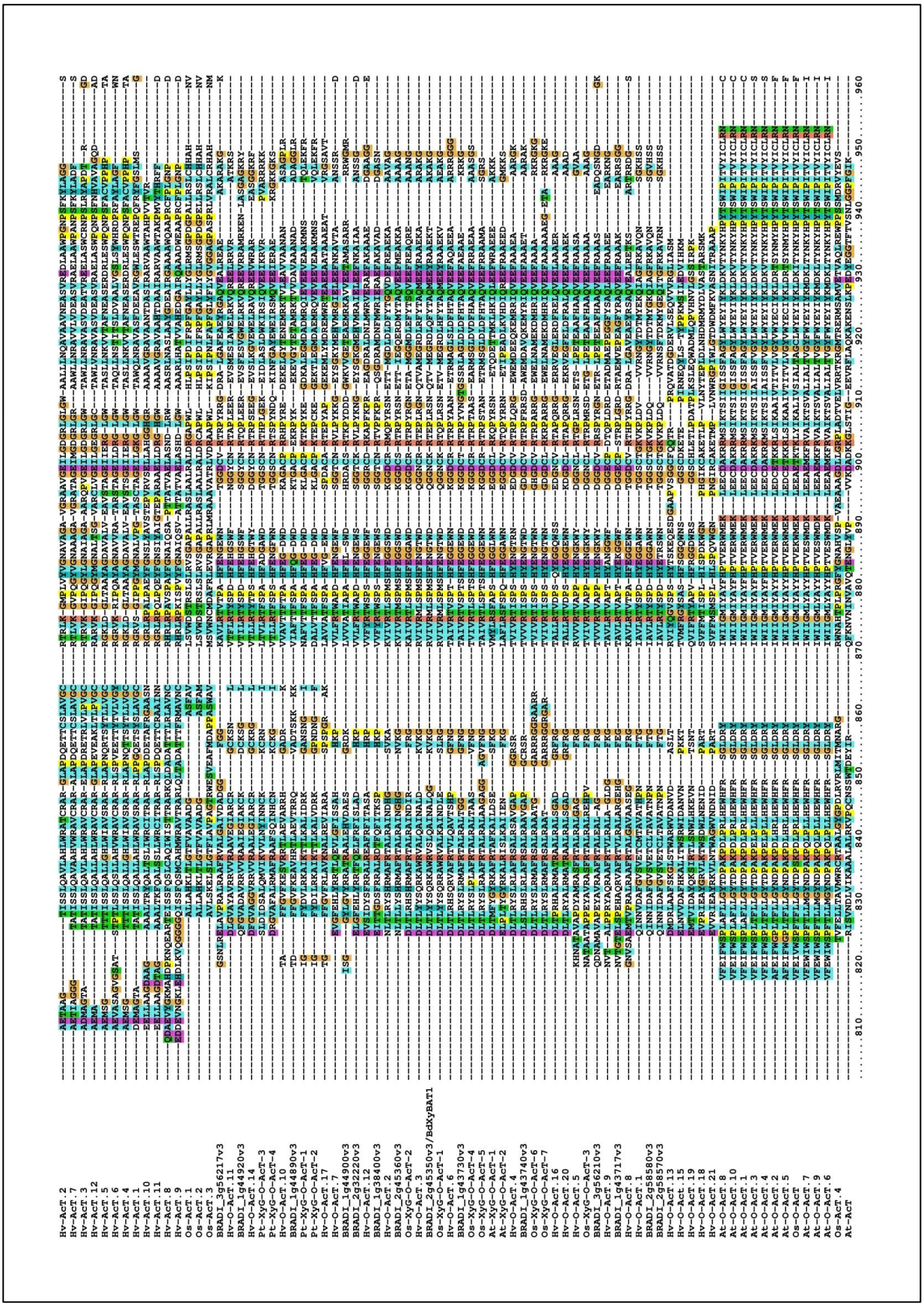

Hv-AcT.2
Hv-AcT.7
Hv-AcT.3
Hv-AcT.12
Hv-AcT.5
Hv-AcT.6
Hv-AcT.4
Hv-AcT.1
Hv-AcT.10
Hv-AcT.11
Hv-AcT.8
Hv-AcT.9
Os-AcT.1
Os-AcT.2
Os-AcT.3
BRADI_3g56217v3
Hv-O-AcT.11
BRADI_1g44920v3
Hv-O-AcT.14
Pt-XyG-O-AcT-3
Pt-XyG-O-AcT-4
Hv-O-AcT.10
BRADI_1g44890v3
Pt-XyG-O-AcT-1
Pt-XyG-O-AcT-2
Hv-O-AcT.17
Hv-O-AcT.7
BRADI_1g44900v3
BRADI_2g32220v3
Hv-O-AcT.12
BRADI_1g38400v3
Hv-O-AcT.2
BRADI_2g45360v3
Os-XyG-O-AcT-2
Hv-O-AcT.3
BRADI_2g45350v3/BdXyBAT1
Os-XyG-O-AcT-1
Hv-O-AcT.6
BRADI_1g43730v3
Os-XyG-O-AcT-4
Os-XyG-O-AcT-5
At-XyG-O-AcT-1
At-XyG-O-AcT-2
Hv-O-AcT.4
BRADI_1g43740v3
Os-XyG-O-AcT-6
Os-XyG-O-AcT-7
Hv-O-AcT.16
Hv-O-AcT.20
Hv-O-AcT.5
Os-XyG-O-AcT-3
BRADI_3g56210v3
Hv-O-AcT.9
BRADI_1g43717v3
Hv-O-AcT.8
Hv-O-AcT.1
BRADI_2g58580v3
BRADI_2g58570v3
Hv-O-AcT.13
Hv-O-Act.15
Hv-O-AcT.19
Hv-O-AcT.18
Hv-O-AcT.21
At-O-AcT.8
At-O-AcT.10
At-O-AcT.1
At-O-AcT.3
At-O-AcT.4
At-O-AcT.2
At-O-AcT.5
Os-O-AcT
At-O-AcT.7
At-O-AcT.9
At-O-AcT.6
Os-AcT.4
At-AcT
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150.....160


Hv-AcT.2
Hv-AcT.7
Hv-AcT.3
Hv-AcT.12
Hv-AcT.5
Hv-AcT.6
Hv-AcT.4
Hv-AcT.1
Hv-AcT.10
Hv-AcT.11
Hv-AcT.8
Hv-AcT.9
Os-AcT.1
Os-AcT.2
Os-AcT.3
BRADI_3g56217v3
Hv-O-AcT.11
BRADI_1g44920v3
Hv-O-AcT.14
Pt-XyG-O-AcT-3
Pt-XyG-O-AcT-4
Hv-O-AcT.10
BRADI_1g44890v3
Pt-XyG-O-AcT-1
Pt-XyG-O-AcT-2
Hv-O-AcT.17
Hv-O-AcT.7
BRADI_1g44900v3
BRADI_2g32220v3
Hv-O-AcT.12
BRADI_1g38400v3
Hv-O-AcT.2
BRADI_2g45360v3
Os-XyG-O-AcT-2
Hv-O-AcT.3
BRADI_2g45350v3/BdXyBAT1
Os-XyG-O-AcT-1
Hv-O-AcT.6
BRADI_1g43730v3
Os-XyG-O-AcT-4
Os-XyG-O-AcT-5
At-XyG-O-AcT-1
At-XyG-O-AcT-2
Hv-O-AcT.4
BRADI_1g43740v3
Os-XyG-O-AcT-6
Os-XyG-O-AcT-7
Hv-O-AcT.16
Hv-O-AcT.20
Hv-O-AcT.5
Os-XyG-O-AcT-3
BRADI_3g56210v3
Hv-O-AcT.9
BRADI_1g43717v3
Hv-O-AcT.8
Hv-O-AcT.1
BRADI_2g58580v3
BRADI_2g58570v3
Hv-O-AcT.13
Hv-O-Act.15
Hv-O-AcT.19
Hv-O-AcT.18
Hv-O-AcT.21
At-O-AcT.8
At-O-AcT.10
At-O-AcT.1
At-O-AcT.3
At-O-AcT.4
At-O-AcT.2
At-O-AcT.5
Os-O-AcT
At-O-AcT.7
At-O-AcT.9
At-O-AcT.6
Os-AcT.4
At-AcT
.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....290.....300.....310.....320


Hv-AcT.2
Hv-AcT.7
Hv-AcT.3
Hv-AcT.12
Hv-AcT.5
Hv-AcT.6
Hv-AcT.4
Hv-AcT.1
Hv-AcT.10
Hv-AcT.11
Hv-AcT.8
Hv-AcT.9
Os-AcT.1
Os-AcT.2
Os-AcT.3
BRADI_3g56217v3
Hv-O-AcT.11
BRADI_1g44920v3
Hv-O-AcT.14
Pt-XyG-O-AcT-3
Pt-XyG-O-AcT-4
Hv-O-AcT.10
BRADI_1g44890v3
Pt-XyG-O-AcT-1
Pt-XyG-O-AcT-2
Hv-O-AcT.17
Hv-O-AcT.7
BRADI_1g44900v3
BRADI_2g32220v3
Hv-O-AcT.12
BRADI_1g38400v3
Hv-O-AcT.2
BRADI_2g45360v3
Os-XyG-O-AcT-2
Hv-O-AcT.3
BRADI_2g45350v3/BdXyBAT1
Os-XyG-O-AcT-1
Hv-O-AcT.6
BRADI_1g43730v3
Os-XyG-O-AcT-4
Os-XyG-O-AcT-5
At-XyG-O-AcT-1
At-XyG-O-AcT-2
Hv-O-AcT.4
BRADI_1g43740v3
Os-XyG-O-AcT-6
Os-XyG-O-AcT-7
Hv-O-AcT.16
Hv-O-AcT.20
Hv-O-AcT.5
Os-XyG-O-AcT-3
BRADI_3g56210v3
Hv-O-AcT.9
BRADI_1g43717v3
Hv-O-AcT.8
Hv-O-AcT.1
BRADI_2g58580v3
BRADI_2g58570v3
Hv-O-AcT.13
Hv-O-Act.15
Hv-O-AcT.19
Hv-O-AcT.18
Hv-O-AcT.21
At-O-AcT.8
At-O-AcT.10
At-O-AcT.1
At-O-AcT.3
At-O-AcT.4
At-O-AcT.2
At-O-AcT.5
Os-O-AcT
At-O-AcT.7
At-O-AcT.9
At-O-AcT.6
Os-AcT.4
At-AcT
.....330.....340.....350.....360.....370.....380.....390.....400.....410.....420.....430.....440.....450.....460.....470.....480


Hv-AcT.2
Hv-AcT.7
Hv-AcT.3
Hv-AcT.12
Hv-AcT.5
Hv-AcT.6
Hv-AcT.4
Hv-AcT.1
Hv-AcT.10
Hv-AcT.11
Hv-AcT.8
Hv-AcT.9
Os-AcT.1
Os-AcT.2
Os-AcT.3
BRADI_3g56217v3
Hv-O-AcT.11
BRADI_1g44920v3
Hv-O-AcT.14
Pt-XyG-O-AcT-3
Pt-XyG-O-AcT-4
Hv-O-AcT.10
BRADI_1g44890v3
Pt-XyG-O-AcT-1
Pt-XyG-O-AcT-2
Hv-O-AcT.17
Hv-O-AcT.7
BRADI_1g44900v3
BRADI_2g32220v3
Hv-O-AcT.12
BRADI_1g38400v3
Hv-O-AcT.2
BRADI_2g45360v3
Os-XyG-O-AcT-2
Hv-O-AcT.3
BRADI_2g45350v3/BdXyBAT1
Os-XyG-O-AcT-1
Hv-O-AcT.6
BRADI_1g43730v3
Os-XyG-O-AcT-4
Os-XyG-O-AcT-5
At-XyG-O-AcT-1
At-XyG-O-AcT-2
Hv-O-AcT.4
BRADI_1g43740v3
Os-XyG-O-AcT-6
Os-XyG-O-AcT-7
Hv-O-AcT.16
Hv-O-AcT.20
Hv-O-AcT.5
Os-XyG-O-AcT-3
BRADI_3g56210v3
Hv-O-AcT.9
BRADI_1g43717v3
Hv-O-AcT.8
Hv-O-AcT.1
BRADI_2g58580v3
BRADI_2g58570v3
Hv-O-AcT.13
Hv-O-Act.15
Hv-O-AcT.19
Hv-O-AcT.18
Hv-O-AcT.21
At-O-AcT.8
At-O-AcT.10
At-O-AcT.1
At-O-AcT.3
At-O-AcT.4
At-O-AcT.2
At-O-AcT.5
Os-O-AcT
At-O-AcT.7
At-O-AcT.9
At-O-AcT.6
Os-AcT.4
At-AcT
.....490.....500.....510.....520.....530.....540.....550.....560.....570.....580.....590.....600.....610.....620.....630.....640





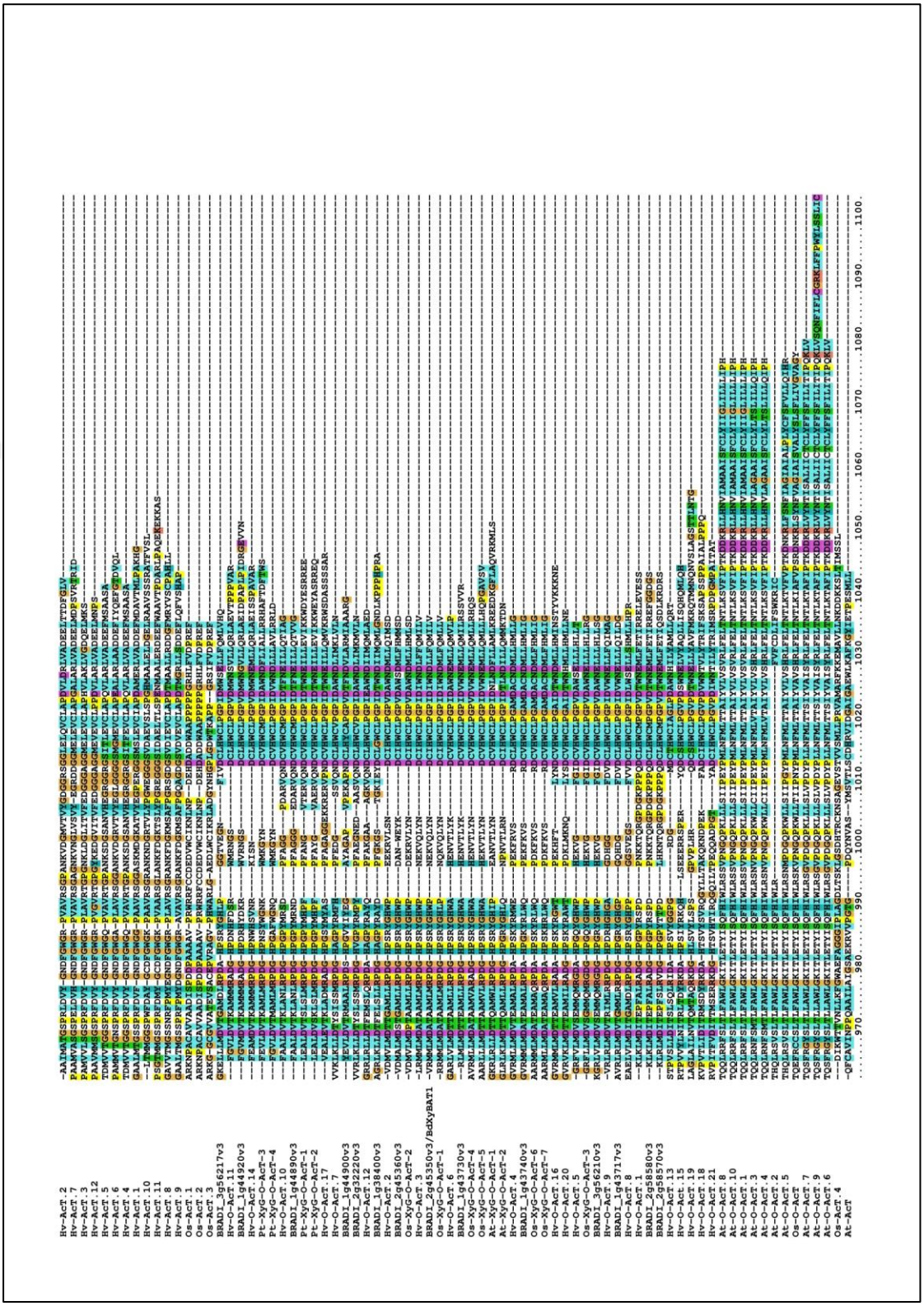


Figure A.3. Multiple sequence alignment results of xyloglucan O-acetyl transferase proteins of *Hordeum vulgare*, *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Brachypodium distachyon*