



Kandidatarbeten
i skogsvetenskap
Fakulteten för skogsvetenskap

2016:12

Cellulosasyntesgener i två Populusarter
- släktskap och uttrycksmönster under vedbildning

Cellulose synthase genes in two *Populus* species
- phylogeny and expression pattern during wood formation



Foto: Hilda Mikaelsson

Jenny Dahl & Hilda Mikaelsson

Sveriges Lantbruksuniversitet
Institutionen för skogens ekologi och skötsel
Kandidatarbete i skogsvetenskap, 15 hp,

Handledare: Totte Niittylä, SLU, Inst för skoglig genetik och växtfysiologi
Examinator: Tommy Mörling, SLU, Inst för skogens ekologi och skötsel

Program: Jägmästarprogrammet

Kurs: EX0592 Nivå: G2E

Umeå 2016



Kandidatarbeten i Skogsvetenskap

Fakulteten för skogsvetenskap,
Sveriges lantbruksuniversitet

Enhet/Unit	Institutionen för skogens ekologi och skötsel Department of Forest Ecology and Management
Författare/Author	Jenny Dahl & Hilda Mikaelsson
Titel, Sv	Cellulosasyntasgener i två Populusarter – släktskap och uttrycksmönster under vedbildning
Titel, Eng	Cellulose synthase genes in two <i>Populus</i> species – phylogeny and expression pattern during wood formation
Nyckelord/ Keywords	<i>Populus trichocarpa</i> , <i>Populus tremula</i> , genom, cellväggssyntes/ <i>Populus trichocarpa</i> , <i>Populus tremula</i> , genome, cell wall synthesis
Handledare/Supervisor	Totte Niittylä, institutionen för skoglig genetik och växtfysiologi
Examinator/Examiner	Tommy Mörling Institutionen för skogens ekologi och skötsel/ Department of Forest Ecology and Management
Kurstitel/Course	Kandidatarbete i skogsvetenskap Bachelor Degree in Forest Science
Kurskod	EX0592
Program	Jägmästarprogrammet
Omfattning på arbetet	15 hp
Nivå och fördjupning på arbetet	G2E
Utgivningsort	Umeå
Utgivningsår	2016

FÖRORD

Denna kandidatuppsats omfattar 15 högskolepoäng och är en del av jägmästarprogrammet. Vi vill tacka vår handledare Totte Niittylä vid Institutionen för skoglig genetik och växtfysiologi för bra vägledning, samt John Baison på Umeå plant science center för råd kring den fylogenetiska analysen.

Umeå, april 2016

Hilda Mikaelsson & Jenny Dahl

SAMMANFATTNING

Cellulosa är ett förnyelsebart material som är högaktuellt i dagens miljömedvetna samhälle. För att kunna nyttja den effektivt krävs kunskap om de processer som styr vedbildning och trädens tillväxt. Detta arbete syftar till att reda ut vilka gener som medverkar i cellulosasyntas (cesA-gener) i *Populus trichocarpa* och i *Populus tremula*, samt att kartlägga genernas uttrycksmönster under vedbildning. Båda arternas genom är fullständigt sekvenserade och finns tillgängliga i databasen PopgenIE. Som utgångspunkt användes en lista över cesA-gener från Kumar et al.(2009). En del av de funna generna visade sig vara CSL-gener, eller gener som inte uttrycks under vedbildningen. Dessa gener användes inte för vidare analyser. I *P. trichocarpa* hittades 19 cesA-gener som uttrycks under vedbildning, medan 18 gener hittades i *P. tremula*. Av dessa skapades ett fylogenetiskt träd. CesA-generna i *P. trichocarpa* har i de flesta fall en motsvarighet i *P. tremula* på samma gren. För att undersöka genernas uttrycksmönster användes databasen Aspwood. Uttrycksmönstren varierar mycket mellan generna och täcker både primär och sekundär cellväggssyntes samt den sista fasen med mognad och celldöd. De som är aktiva under den sista fasen är troligen inblandade i märkecellernas cellväggssyntes. Merparten av generna uttrycks under sekundär cellväggssyntes. Generna delades in i grupper efter uttrycksmönster, och i det fylogenetiska trädet är generna i hög utsträckning grupperade på ett liknande sätt. Våra resultat ger en översikt av cesA-generna och deras funktion.

Nyckelord: *Populus trichocarpa*, *Populus tremula*, genom, cellväggssyntes

ABSTRACT

Cellulose is a renewable material with great importance in today's society, where the concern of environmental issues is increasing. To be able to use it efficiently, there is a need for deeper knowledge concerning wood formation processes. This paper aims to unravel which genes are contributing in cellulose synthase (*cesA* genes) in *Populus trichocarpa* and *Populus tremula*, and to map the expression patterns of the genes during wood formation. The genomes of both species are sequenced and available in PopGenIE. As an outset a list of *cesA* genes from Kumar et al. (2009) was used. Some of the found genes turned out to be *cesa* like genes (CSL), or genes that are not expressed during wood formation. These genes were not further analysed. 19 *cesA* genes were found in *P. trichocarpa*, while 18 were found in *P. tremula*. From these, a phylogenetic tree was created. In most cases, *P. trichocarpa* has an equivalent in *P. tremula* on the same branch. To examine the expression patterns of the genes, the database Aspwood was used. The expression varies between the genes and covers primary and secondary cell wall synthesis, as well as the last phase with maturation and cell death. The genes active during the last phase are probably involved in ray cells. Most of the genes are expressed during secondary cell wall synthesis. The genes were grouped based on expression pattern, and are in the phylogenetic tree grouped in a similar way. Our results give an overview of the *cesA* genes and their function.

Keywords: *Populus trichocarpa*, *Populus tremula*, genome, cell wall synthesis

INTRODUKTION

30 % av jordens landyta täcks av skog som bidrar med råvara som timmer och biobränsle samt står för en stor del av biodiversiteten på jorden. Den står också för en stor del av kolinlagringen. Träden är därför mycket viktiga ur både ekologiskt och ekonomiskt perspektiv. Med den ökade miljömedvetenheten är förnyelsebara råvaror på uppgång, en arena där trä och cellulosa spelar en mycket stor roll. (Tuskan et al. 2006) Trädens uppbyggnad och biologiska processer är väl studerade men mycket finns kvar att utforska (Li et al 2005). Processerna i träd, till exempel vedbildning, drivs av gener. Hur och när olika gener fungerar och uttrycks under vedbildningen är ett område som intresset växer för och mycket forskning genomförs.

Gener och proteinsyntes

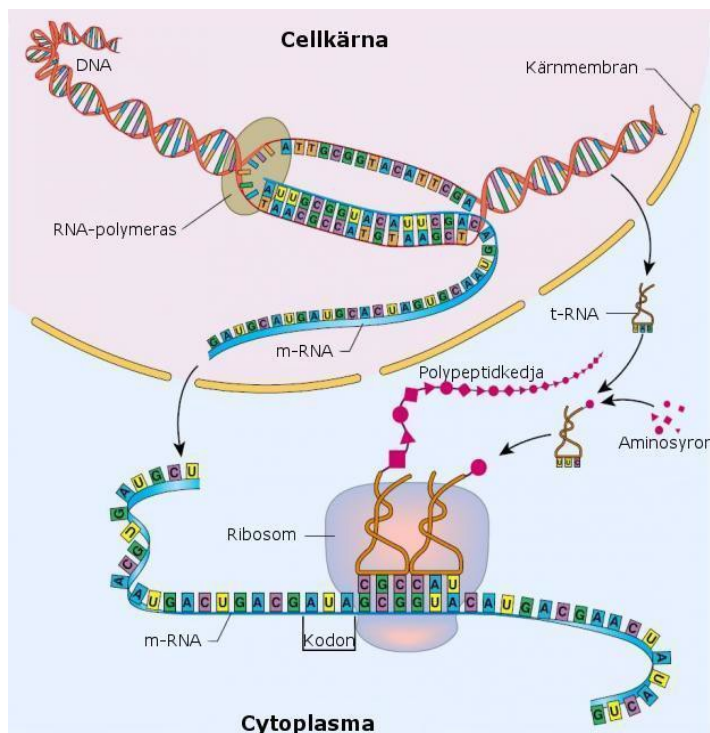
Genomet är det som avgör vilka egenskaper en växt får, det finns i cellkärnan och är uppbyggt av DNA (Deoxiribonucleic Acid). DNA-molekylen består av två strängar av sammanlänkade nukleotider som tillsammans bildar en spiralformad struktur, en dubbelhelix. Varje nukleotid innehåller en av fyra möjliga kvävebaser (A, T, G eller C), och det är deras ordning samt genomets längd som är avgörande. DNAt är uppdelat i segment som kallas kromosomer. Kromosomuppsättningen kan vara enkel (haploid), dubbel (diploid) eller flertalig (polyploid). Att kromosomuppsättningen är diploid innebär att den har två kopior av varje kromosom och gen, och att den är polyploid innebär att det finns tre eller fler kopior. (Alberts et al. 2010)

I varje kromosom finns ett antal gener, alltså sekvenser av DNA, som fungerar som instruktioner för uppbyggandet av specifika proteiner. I början av varje gen finns en startsignal (promoter) och i slutet en stoppsignal (terminator). Däremellan finns kodande sekvenser (exoner) och icke-kodande (introner). (Alberts et al. 2010)

Proteinerna som bildas från DNAt styr de flesta processer i cellerna. När ett nytt protein bildas börjar det med att den aktuella delen av DNAt kopieras till en annan sorts nukleinsyra, mRNA (messenger Ribonucleic Acid). Detta steg kallas transkription. mRNA består liksom DNA av nukleotider, den är dock enkelsträngad och inte dubbel som DNA. För varje kvävebas i DNA finns en motsvarighet i mRNA och vid kopieringen paras de rätta kvävebaserna ihop. Hela genen, med introner och exoner, kopieras. I ett senare skede klipps intronerna bort, i en process som kallas för RNA-splicing. När mRNA är färdigt transporteras det ut från cellkärnan till någon av ribosomerna i cytoplasman. (Alberts et al. 2010)

Tre kvävebaser i följd kallas för ett kodon, och varje kodon kodar för en aminosyra. Det finns $4^3=64$ olika tripletter av kvävebaser men bara 20 aminosyror, vissa av dem har alltså fler än ett kodon. Aminosyror sitter på bärarmolekyler som kallas tRNA, som också innehåller ett matchande kodon, så kallat antikodon. Ribosomen förflyttas längs mRNA-molekylen, läser av varje kodon, och fogar på så vis ihop aminosyror i rätt ordning. Detta steg kallas för translation. Kedjan av aminosyror veckas sedan till en 3D-struktur, vars utförande beror på

aminosyroras ordning. Veckningen utgör det sista steget i bildningen av ett protein. (Alberts et al. 2010) Proteinsyntesen illustreras i figur 1.



Figur 1. Proteinsyntesen börjar med att mRNA bildas från DNA via transkription. Kvävebaserna i mRNA kodar för aminosyror som sammanfogas till ett polypeptidkedja. I sista steget i proteinsyntesen veckas polypeptidkedjan för att bilda ett specifikt protein. (Bjorn Norheim 2013)

Figure 1. Protein synthesis starts with mRNA being formed from DNA during the transcription. The base pairs in mRNA are coding for specific amino acids, that are put together to a poly peptide chain. In the last step of the protein synthesis the poly peptide chain is folded to form a specific protein. (Bjorn Norheim 2013)

Gener kan uttryckas olika starkt och ger olika då mängd mRNA-kopior, beroende på vad cellen behöver för tillfället. Varje mRNA-kopia kan sedan medverka i syntes av flera likadana proteiner. Eftersom varje cell bara innehåller en eller ett fåtal kopior av varje gen ger denna “flerstegseffekt” möjlighet att snabbt producera stora mängder av det protein som behövs. (Alberts et al. 2010)

Forskning kring genetik hos växter

Människan har odlat växter i över 10 000 år. Jordbruket uppkom i olika delar av världen under stenåldern, och sedan dess har människor medvetet valt ut växter med önskade egenskaper. Växterna har anpassats på olika sätt för att bättre uppfylla våra behov. Den äldsta formen av växtförädling är urval vilket går ut på att leta efter intressanta egenskaper hos arter, som man sedan försöker föra vidare till nästa generation. Trots att kunskapen om genetik länge varit begränsad har många olika metoder för att förädla växter uppkommit. (Gentekniknämnden 2014) På senare tid har genforskningen gjort stora framsteg. 1977 utvecklade Fredrick Sanger en metod för att sekvensera hela gener, alltså åskådliggöra nukleotidernas ordning. Metoden har sedan automatiserats och förbättrats, och nu finns flera metoder för att sekvensera hela genom. (Kim et al. 2008)

Den första växten vars genom kartlades helt var *Arabidopsis thaliana*, år 2000. Det består av ca 125 miljoner baspar (Mb). (The Arabidopsis Genome Initiative 2000) *A. thaliana* är en ört och har därför ingen naturlig vedbildning, varför den främst har använts till att studera primär tillväxt. Under vissa tillväxtförhållanden kan dock en sorts tillväxt som liknar vedbildning hos träd induceras. Därför har *A. thaliana* även använts som modellväxt för att studera processer kopplade till vedbildning. (Li et al. 2006)

Även om den inducerade vedbildningen hos *A. thaliana* är betydelsefull för forskningen, finns det skillnader mot den naturliga vedbildningen i träd (Li et al., 2005). För att bättre förstå processerna bakom vedbildning sekvenserades det första trädet, *Populus trichocarpa*, år 2006 (Tuskan et al. 2006). Dess genom uppskattas bestå av cirka 485 Mb, det är alltså ungefär fyra gånger större än genomet hos *A. thaliana* (Winzell 2007) Även genomet hos *Populus tremula* har sekvenserats på Umeå Plant Science Centre, men har inte publicerats ännu ¹.

Populussläktets historia och egenskaper

Det finns flera anledningar till att en art från släktet *Populus* valdes som modellväxt för träden. Tack vare ett nära släktskap mellan *Populus* och *Arabidopsis* underlättas studier av olika funktioner, eftersom de går att jämföra med varandra (Djerbi et al. 2004). Släktet *Populus* har också ett litet genom som är relativt lätt att transformera, vilket gör dem lämpliga för genetiska studier (Li et al. 2006; Winzell 2007). Släktet har en stor naturlig utbredning och är betydelsefull både ekologiskt och ekonomiskt och är därför relevant att studera (Li et al 2005).

Fossila fynd indikerar att släktet *Populus* fanns redan under Paleocene, för 60 miljoner år sedan. Den spreds från antingen Nordamerika eller Asien till stora delar av norra hemisfären där den finns idag. *Populus* ingår i samma familj som *Salix*; *Salicaceae*. Antalet arter inom släktet *Populus* är omtvistat och förslagen varierar mellan 22 och 85 olika arter. Det mest vedertagna artantalet är 29. Den stora variationen beror på att släktet lätt bildar hybrider och att åsikterna går isär om vilka som ska klassas som egna arter. Hybridiseringen är en bidragande faktor i populussläktets evolution då den påskyndar bildningen av nya arter. (Jansson et al. 2010) Släktet är indelat i sex sektioner där *P. trichocarpa* tillhör sektionen *Tacamahaca* medan *P. tremula* tillhör *Leuce* (Cervera et al. 2005).

Populussläktets egenskaper varierar mycket mellan olika arter, men de flesta kännetecknas av att vara tåliga och ha en snabb tillväxt. *P. trichocarpa* har sin naturliga utbredning i Nordamerika och växer mest i våtmarker och översvåmningsområden, medan *P. tremula* växer i Europa och Asien, och tillhör arterna som förekommer mer i höglänta områden och berg. (Jansson et al. 2010)

¹ Totte Niittylä, Umeå Plant Science Center, 2016-03

Släktet *Populus* har en stor genetisk variation trots att genomen hos arterna är relativt små (Djerbi et al. 2004). Den sexuella förökningen medför att den är en "obligate outcrosser", vilket inför nytt genetiskt material och därmed ökar den genetiska variationen. Dessutom är populationerna relativt stora och har pollen och frön som sprids långt med hjälp av vinden, detta minskar risken för inavel. (Jansson et al. 2010) Ytterligare en anledning till den stora genetiska variationen är att stora delar av populussläktets genom har duplicerats, vilket medför att en del gener finns på flera olika ställen i genomet. Genduplicering är relativt vanligt och har förekommit hos över 50% av angiospermerna (Smith et al. 2010). Dupliceringen kan innefatta hela eller delar av genomet och ge upphov till ökat antal kromosomer. De duplicerade generna kan anta nya funktioner, vilket ökar den genetiska variationen. Chansen att detta sker ökar ju längre tid som gått sedan dupliceringen. (Smith et al. 2010) En annan möjlighet för de nya generna är att de är "tysta" och bara finns där utan vidare funktion, eller att de försvinner igen. De kan även vara aktiva båda två och fylla samma funktion, om en av generna då går förlorad kan den andra genen ta över. På så vis behöver ingen fenotypisk förändring ske. Dupliceringen kan därmed fungera som skydd mot mutationer. Gendupliceringen kan på sikt leda till bildning av nya arter. (Smith et al. 2010)

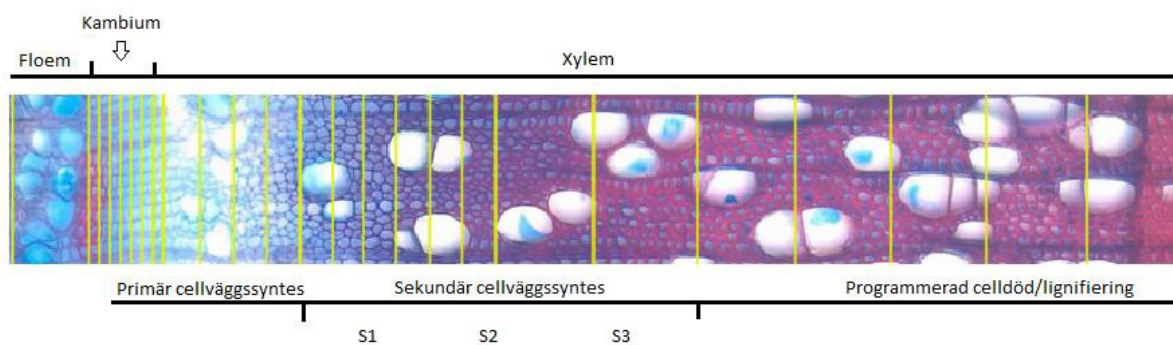
Det är känt att en duplicering skett hos *Populussläktet* och att den troligtvis påverkade ca 92% av genomet (Tuskan et al. 2006), däremot är tidpunkten oklar. Sterck et al. (2005) föreslår att en duplicering skedde för 8 till 13 miljoner år sedan, medan Tuskan et al. (2006) tror att det snarare handlar om 60-65 miljoner år sedan. De olika arternas genom kan därefter ha utvecklats olika.

Växters tillväxt och vedbildningsprocess

I forskningen kring växters tillväxt är det framför allt den primära tillväxten som studerats, medan den sekundära tillväxten där vedbildningen sker inte är kartlagd i samma utsträckning (Li et al. 2006). Den primära tillväxten innebär att rötter, grenar och stam tillväxer i spetsen och därmed förlängs. Sekundär tillväxt utgår istället från det vaskulära kambiet och korkkambiet, vilket leder till en radiell tillväxt av stammen. Det är alltså i den sekundära tillväxten som veden bildas, den delen av trädet som är mest betydelsefull för människans användande av träd som naturresurs. (Li et al. 2006)

Ved är uppbyggt av xylemceller som bildas i det vaskulära kambiet och sedan differentieras för att fylla olika funktioner. Veden består i huvudsak av cellulosa, hemicellulosa och lignin. (Li et al. 2006) De tre vanligaste celltyperna i xylemet i lövved är fibrer, kärl och märkeceller. Fibrerna fungerar som stöd för stammen och är därför stabila och har tjocka väggar. Kärlen bildar långa rör genom trädet och står för vattentransporten. Märkecellerna bygger upp märkestrålar som går radiellt genom xylemet och förenar floemet (innerbarken) med märkegen. Märkestrålarnas uppgift är att transportera näringsämnen, vatten och lösta gaser, samt att lagra stärkelse. (Smith et al. 2010) De märkeceller som finns i märkestrålarnas utkant är sammanlänkade med vedcellerna utanför genom porer medan de inre märkecellerna är mer isolerade från veden (Nakaba et al. 2012).

Vid vedbildning bildas först den primära cellväggen. Den består av cellulosededjor som binds samman av hemicellulosa. Mellan cellulosededjorna finns pektin som är ett tånjbart material och tillåter den primära cellväggen att expandera. När den primära cellväggen bildats skapas den sekundära cellväggen som består av cellulosa, hemicellulosa, strukturproteiner och lignin. Ligninet fyller ut utrymmet mellan cellulosan och hemicellulosan. Det är starkt och styvt och hindrar cellen från att expandera mer. (Winzéll 2007) Den sekundära cellväggen består av tre lager; S1, S2 och S3. Det som främst skiljer dem åt är cellulosededjornas (mikrofibrillernas) vinkel. Även märgcellerna bildar sekundär cellvägg, först de i anslutning till andra vedceller, men senare även märgcellerna i mitten av märgstrålarna (Nakaba et al. 2012). Efter att det tredje lagret bildats går cellen in i sin slutliga fas, celldöden. Då slutar cellen fungera och en ihålig vedcell är bildad (Winzéll 2007). Vedbildningen illustreras i figur 2. Cellerna i märgstrålarna lever i allmänhet längre än övriga celler. Nakaba et al. (2012) föreslår att märgcellerna är aktiva i upp till 5 år, medan t ex de celler som bildar kärldör direkt efter differentiering.



Figur 2. Tvärsnitt av ved under utveckling. Först bildas den primära cellväggen som består av cellulosa, hemicellulosa och pektin. Därefter bildas den sekundära cellväggen bestående av cellulosa, hemicellulosa, strukturproteiner och lignin. I sista fasen sker lignifiering av cellväggarna och programmerad celldöd. (Figur modifierad från Aspwood).

Figure 2. Cross section of developing wood. First the primary cell wall is formed, consisting of cellulose, hemicellulose and pectin. Then the secondary cell wall is formed. It consists of cellulose, hemicellulose, structural proteins and lignin. In the last phase, lignifying of the cell walls and programmed cell death takes place. (Figure modified from Aspwood)

Cellulosans betydelse och användning

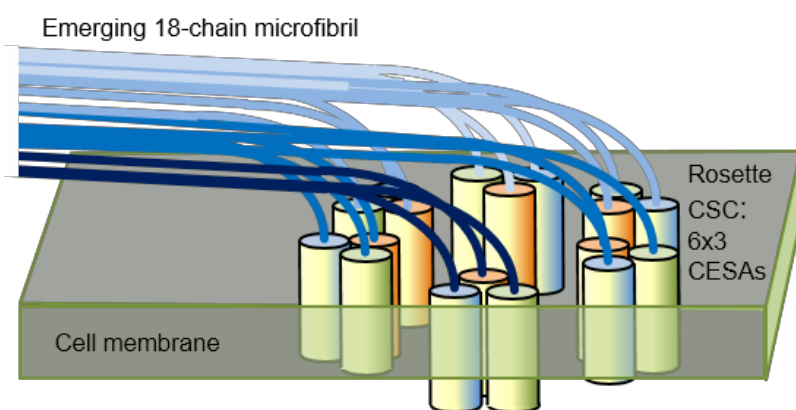
Den komponent i veden som har störst betydelse för människan är cellulosan. Cellulosa är naturens mest förekommande polymer och utgör cirka 40 % av ett trädets torrsvikt. Cellulosa är uppbyggt av $\beta(1,4)$ -glukosenheter som tillsammans bildar oöngade kedjor. Kedjorna bildar mikrofibriller, och de i sin tur bildar fibriller. (Li et al. 2006) Redan på 1800-talet framtoqs metoder för att frilägga cellulosan från vedens övriga ämnen och utvecklingen har sedan dess ständigt gått framåt. Från att tidigare främst ha använts till pappersmassa så finns idag många användningsomåden för cellulosan, till exempel i textilier eller inom farmakologin. (Ek &

Illergård 2013) Cellulosa används också för framställning av etanol, och är med nutidens miljömedvetenhet högst aktuell som förnyelsebar råvara. Förädling mot träd med högre andel cellulosa skulle ge en mer effektiv användning av råvaran i industrierna (Suzuki et al. 2006).

Cellulosasyntasgenerna och deras funktion

Cellulosan bildas av cellulosasyntasproteiner (cesA-proteiner), som i sin tur bildas genom att speciella gener, cesA-gener, uttrycks. Antalet cesA-gener skiljer sig mellan olika växter. *A. thaliana* har 10 olika cesA-gener medan det vid senaste kartläggningen av cesA-gener hos *P. trichocarpa* hittades 18 stycken (Djerbi et al. 2005). Kumar et al. (2009) föreslår dock att det finns 17 cesA-gener hos *P. trichocarpa* eftersom två av dem (PtiCesA8-A och PtiCesA8-B) är i det närmaste identiska. (Suzuki et al. 2006) Genomet hos *P. tremula* är sekvenserat men det är inte utrett hur cesA-familjen ser ut och om den skiljer sig från *P. trichocarpa*.

CesA-proteinerna sitter i cellens plasmamembran och bildar så kallade rosettkomplex eller cellulosasyntaskomplex (CSC). Varje CSC består av 6 subenheter med cesA-proteiner. Man har länge trott att varje subenhet i komplexet är uppbyggt av 6 cesA-proteiner som var och en bildar en cellulosapolymer, alltså att totalt 36 stycken bildas och förenas till en mikrofibrill. (Kumar & Turner 2015). Jarvis (2013) anser dock att mikrofibrillerna är för tunna för att bestå av 36 cellulosapolymerer, vilket indikerar att det snarare rör sig om 3-4 cesA-proteiner i varje subenhet. I sådana fall bildas 18-24 cellulosapolymerer från varje CSC (Se figur 3). CSC-komplexen förflyttas längs med plasmamembranet samtidigt som de syntetiserade cellulosapolymererna integreras i cellväggen (Smith et al. 2010).



Figur 3. Cellulosasyntaskomplex är uppbyggda av subenheter med cesA-proteiner, som bildar cellulosapolymerer. (Jarvis 2013)

Figure 3. Cellulose synthase complexes are made from subunits of cesA-proteins that form cellulose polymers. (Jarvis 2013)

En del cesA-gener uttrycks och bildar proteiner när den primära cellväggen bildas medan andra uttrycks i samband med syntesen av sekundär cellvägg (Djerbi et al. 2004). Song et al. (2010) har undersökt vilka CSC som är aktiva i vedceller under vedbildning. Genom att analysera i vilka celler som olika typer av CSC finns, och vilka CesA-proteiner som ingår i dem, har CSC delats in i två grupper med olika egenskaper. De proteiner som ingår i CSC av typ I överensstämmer till stor del med de cesA-gener som uttrycks under sekundär cellväggsbildning i både *Arabidopsis* och *Populus*. Typ II utgörs av proteiner som är

involverade i bildningen av primär cellvägg i *Arabidopsis*. I *Populus* uttrycks de motsvarande generna under differentiering av xylemceller och ännu mer i kambieceller, vilket indikerar att cesA-gener av typ II är involverade i bildning av både primär och sekundär cellvägg. (Song et al. 2010)

Syfte och mål

Virke och cellulosa spelar en allt större roll som förnyelsebar råvara och det behövs kunskap om dess syntes. De pågående klimatförändringarna är ytterligare en anledning till att öka kunskapen om genetiska processer bakom vedbildning. I och med dem kan abiotiska faktorer som till exempel temperatur och vattentillgång förändras, faktorer som påverkar genernas respons och därmed vedens sammansättning (Wang et al. 2016). Det kan påverka vedens egenskaper och människans möjligheter att använda den. Ett exempel på hur vedbildningen påverkas av yttre faktorer är reaktionsved, som bildas vid mekanisk belastning, till exempel vind. I lövträd kallas reaktionsveden för dragved och den får i jämförelse med vanliga vedceller ett tunnare S2-skikt och ett tjockt, geléartat S3-skikt som nästan uteslutande består av cellulosa. (Winzéll 2007) Vid bildning av dragved kan därför antas att generna inblandade i cellulossyntes uttrycks mer än vanligt.

Ökad förståelse för genetiska processer och cesA-genernas roll kan vara till stor nytta, både för att kunna möta klimatförändringar och för att kunna tillgodose framtida behov av träråvara. Med en ökad kunskap om i vilka processer de olika generna uttrycks och vilka gener som bidrar till olika komponenter i veden kan användandet av trä effektiviseras. Cellulosasyntasgenerna hos *P. trichocarpa* är kartlagda, men det har inte genomförts på ett systematiskt sätt hos *P. tremula*. Eftersom *P. tremula* är mer vanligt förekommande i Europa är det av intresse att undersöka cesA-generna och vedbildningsprocessen även hos den.

Arbetet har två mål:

- Undersöka vilka cesA-gener som finns i *P. tremula* och jämföra dem med de cesA-gener som finns i *P. trichocarpa*. Sedan genomet hos *P. trichocarpa* sekvenserades första gången har det kartlagts mer noggrant. Därför söks först *P. trichocarps* genom igenom efter eventuella nya cesA-gener, och därefter söks genomet hos *P. tremula* igenom efter cesA-gener. Hypotesen är att det finns cesA-gener i *P. tremula* som skiljer sig från de i *P. trichocarpa*.
- Analysera när under vedbildningen de olika cesA-generna hos *P. trichocarpa* uttrycks; om det är när den primära eller sekundära cellväggen bildas, senare under celldöden/lignifiering eller i mägstrålar.

MATERIAL & METODER

Arbetet påbörjades med litteraturstudier om tidigare och aktuell forskning inom områdena vedbildning, cesA-proteinernas funktion och *Populus*släktets historia och evolution.

Den första praktiska delen av arbetet gick ut på att leta efter cesA-gener i databasen PopGenIE (Populus Genome Integrative Explorer). Som utgångspunkt användes en lista över tidigare kända cesA-gener i *P. trichocarpa* från Kumar et al. (2009). PopGenIE är en databas uppbyggd för att utforska gener och geners funktion hos släktet *Populus*. Tanken är att den ska samla och tillgängliggöra det data kring *Populus*genomet som är känt. (Sjödén et al., 2009)

Sökning efter nya cesA-gener hos *P. trichocarpa*

Steg ett var att söka på varje cesA-gens namn (kolumnen *Populus trichocarpa genome assembly gene model*, tabell 1 i Kumar et al. (2009)) i databasen PopGenIE, för att få fram varje gens proteinsekvens. Därefter användes verktyget BLAST med inställningen protein query to protein db och *Populus trichocarpa* protein. Under fliken "advance search parameter" valdes att programmet skulle visa 100 träffar istället för 10. Sökningen resulterade i en lista över gener med olika likhet (similarity) i proteinsekvens. De med minst 60 % likhet valdes ut och deras proteinsekvens sparades i formatet FASTA för att användas till att skapa ett fylogenetiskt träd senare. Med namnet på generna med över 60 % likhet gjordes nya sökningar i PopGenIE. Under rubriken "synonyms" listas andra namn på den sökta genen. Synonymerna jämfördes med tabell 1 i Kumar et al. (2009) för att se om genen redan var känd som cesA-gen men under ett annat namn. Några av generna fanns inte att hitta i PopGenIE, då användes istället databasen Phytozome som komplement. För de gener som inte hittades i Phytozome heller gjordes sökningar på Google för att hitta genens proteinsekvens och på så vis kunna hitta den i Phytozome eller PopGenIE. Utifrån de funna generna skapades en ny lista över cesA-gener i *P. trichocarpa* (se tabell 1).

Sökning efter cesA-gener hos *P. tremula*

Genomet för *P. tremula* genomsöktes i popGenIE efter cesA-gener utifrån den nya listan med cesA-gener från *P. trichocarpa*. Proteinsekvensen användes i verktyget BLAST, med inställningen protein query to protein db och *Populus tremula* protein 1.1. Även här valdes inställningen 100 träffar under fliken "advance search parameter". Sökningen gav en lista över gener med olika likhet i proteinsekvens. Genomet för *P. tremula* har inte bearbetats lika mycket som det för *P. trichocarpa*, därför kom ofta flera träffar på samma gen upp. Dessa gener hade samma namn, men efter namnet var de numrerade med 1, 2, 3 osv. I dessa fall valdes endast en av dem eftersom genvarianterna är i det närmaste identiska. De med minst 60 % likhet (similarity) valdes ut även här och deras proteinsekvenser sparades i formatet

FASTA för att senare skapa ett fylogenetiskt träd. De funna generna sammanställdes i en lista över cesA-gener i *P. tremula* (se tabell 1).

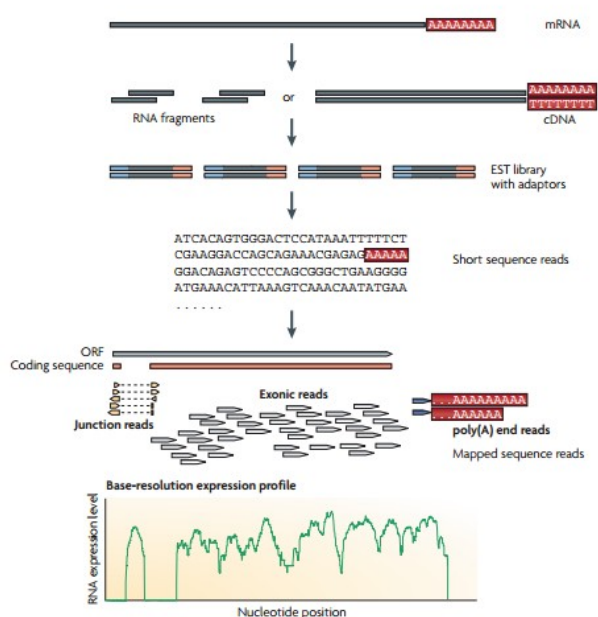
Kontroll av genernas funktion i Aspwood

När alla gener listats efter sökningarna i PopGenIE användes databasen Aspwood för att se genernas funktion, detta för att skilja ut cesA-generna från övriga. I Aspwood finns information om genernas motsvarigheter i *A. thaliana* och vilken funktion de har. För att klassificeras som cesA-gen krävdes att dess motsvarighet i *A. thaliana* är en cesA-gen samt att genen uttrycks under vedbildningen. Aspwood är uppbyggt av gener från *P. tremula* men sökning är endast möjlig för *P. trichocarpa*. Därför söktes information enbart om generna från *P. trichocarpa*.

Beskrivning av mRNA-analys och databasen Aspwood

Datan i Aspwood är framtagen genom RNA-analys av vedremсор från *P. tremula*. Vedremсорna innehåller vävnad från floem, kambium, nybildat xylem samt moget xylem. Cellerna i de olika vävnadstyperna har olika egenskaper, och därmed olika genuttryck som visualiseras i diagram. (Opublicerat material) För varje enskild cesA-gen visas om den uttrycks i bildande av primär cellvägg (Fas 1), i bildande av sekundär cellvägg (Fas 2), eller i den senare fasen med lignifiering/celldöd eller i mörgrårlarna (Fas 3).

När en gen uttrycks starkt bildas mycket RNA från den (Alberts et al. 2010). Att mäta mängden mRNA som bildas är därför en metod för att se när i vedsyntesen de olika cesA-generna uttrycks. RNA-sekvenseringen illustreras i figur 4.



Figur 4. Sekvensering av RNA: RNA konverteras till cDNA (komplementärt DNA som bildas vid transkriptionen av en RNA-sekvens till DNA), en sequencing adaptor tillförs och sedan sekvenseras cDNA från en eller båda ändar. Genom att jämföra sekvenserna med referenssekvenser kan uttrycksmönstret ses. Mycket RNA bildas när en gen uttrycks starkt. (Wang et al. 2009)

Figure 4. Sequencing of RNA: RNA is converted to cDNA (complementary DNA, formed in the transcription of RNA to DNA), a sequencing adaptor is added and then the cDNA is sequenced. By comparing the sequences with reference sequences, the expression pattern can be visualized. A lot of RNA is formed when a gene is expressed strongly. (Wang et al. 2009)

Fylogenetiskt träd samt bootstrapanalys av cesA-gener hos *P. trichocarpa* och *P. tremula*

Ett fylogenetiskt träd visar evolutionära släktskap mellan gener (Kumar et al. 2009). Uppdelningen i grenar visar evolutionära händelser där gener differentierats. Trädet skapas genom att leta efter likheter i sekvenser, så kallade alignments. Gener med många homologa sekvensavschnitt hamnar nära varandra i trädet (Savva 2003).

CesA-generna i *P. trichocarpa* och *P. tremula* användes för att skapa ett fylogenetiskt träd där släktskapet mellan dem visualiseras. Först användes hemsidan <http://www.ebi.ac.uk> som tillhandahålls av European Bioinformatics Institute. Den är till för att tillgängliggöra molekylära databaser och ge möjligheter att analysera data på olika sätt, däribland genom fylogenetiska träd. (European Bioinformatics Institute 2015) Verktöget Clustal Omega (multiple sequence alignment) användes, det radar upp sekvenserna för att möjliggöra jämförelser mellan dem. FASTA-sekvenserna för båda trädslagen kopierades in samtidigt. Resultatet exporterades i FASTA-format för att kunna användas i MEGA till ett fylogenetiskt träd med bootstrapanalys. Bootstrap är en vanlig statistisk metod som går ut på att datat testas flera gånger, i detta fall 500. Vid varje delning fås ett bootstrapvärde som visar vid hur stor andel av omkörningarna som delningen uppkommit. Ett högt värde innebär hög sannolikhet för att delningen är korrekt (Savva 2003) I MEGA konverterades först FASTA-filen till .meg-format. Sedan användes funktionen Construct/test neighbor joining tree med inställningen Protein sequences och Bootstrap test. Resultatet blev ett original tree samt ett bootstrap tree, vilket var det som användes i resultatet. Generna från både *P. trichocarpa* och *P. tremula* användes i samma träd för att lättare kunna jämföra cesA-generna i de båda trädslagen med varandra. I vissa fall gjordes därefter undersökningar i PopgenIE för att utröna om några av generna hade samma DNA-sekvens (gener som fick oväntade placeringar i trädet).

Undersökning av cesA-genernas uttrycksmönster i Aspwood

För att kartlägga när i vedbildningsprocessen olika gener uttrycks användes återigen databasen Aspwood.

En sökning på alla cesA-gener i *P. trichocarpa* gjordes och grafer över deras uttrycksmönster skapades. Därefter jämfördes de med varandra, och liknande gener placerades i samma grupp utifrån en visuell bedömning. Generna delades in i grupper utifrån vilken fas i vedbildningen de uttrycks mest under. Alla gener i samma grupp lades in i en gemensam graf för en bra överblick. Med hjälp av dem visades också om de cesA-gener som hittats verkligen var involverade i vedbildning.

RESULTAT

Gener hos *P. trichocarpa* och deras funktion enligt Aspwood

Vid sökning i PopGenIE efter cesA-gener hos *P. trichocarpa* hittades 37 stycken (Se tabell 1). Alla från listan enligt Kumar et al. (2009) återfanns i PopGenIE utom PtiCesA1-B och PtiCesA7-B. Deras proteinsekvenser hittades dock genom sökning i Phytozome och på Google. 19 nya gener hittades utöver de redan listade (Se tabell 1). I Aspwood framkom att 15 av de nyfunna generna i *P. trichocarpa* var CSL-gener (cesa-like genes, inblandade i syntes av hemicellulosa (Li et al. 2006) och tre var CesA-gener, men som inte uttrycks i veden. Dessa 18 visas inte i det fylogenetiska trädet. Den sista nyfunna genen, Potri.018G103900, visades vara en CesA-gen som uttrycks i veden.

Gener hos *P. tremula*

Vid sökningen efter cesA-gener hos *P. tremula* hittades totalt 27 gener med över 60% likhet i relation till de funna cesA-generna hos *P. trichocarpa*. Av dessa visade sig 9 stycken ha motsvarande gener i *P. trichocarpa* som är CSL-gener och inte cesA. Dessa redovisas i tabell 1 (figur 5) men är inte med i det fylogenetiska trädet.

Tabell 1. Sammanställning av gener med >60 % likhet i proteinsekvens med cesA-gener enligt Kumar et al. (2009), deras olika namn och motsvarande gen hos *P. tremula*. De redan tidigare kända cesA-generna är de som finns i kolumnen längst till vänster. Här visas även om genen är cesA eller CSL (cesA-like) samt under vilken fas i vedbildningen de uttrycks. Fas 1 = primär cellväggssyntes, fas 2 = sekundär cellväggssyntes och fas 3 = lignifiering/celldöd/i märkestrålar.

Table 1. Table of genes with >60 % similarity in protein sequence with cesA-genes according to Kumar et al. (2009), their names and corresponding gene in P. tremula. In the column to the left are the cesA-genes previously known. The genes functions (cesA or CSL) are also shown, as well as during which phase of wood formation they are expressed. Phase 1 = primary cell wall synthesis, phase 2 = secondary cell wall synthesis and phase 3 = lignification/programmed cell death/ in ray cells.

Populus trichocarpa genome assembly gene model	Potri name (P. trichocarpa)	Kumar et al. 2009 (P. trichocarpa)	Potra name (P. tremula)	Arabidopsis descendant (funktion)	Uttryck i vedbildning
eugene3.00002636	Potri.002G257900	PtiCesA4	Potra002707g1 9806	CesA	Fas 2
estExt_fgenes4_pg.C_LG_V1107	Potri.005G194200	PtiCesA6-C	Potra004085g2 4472	CesA	Fas 1, 2, 3
eugene3.00020623	Potri.002G066600	PtiCesA6-D	Potra001021g0 8534	CesA	Fas 2+3
gw1.XI.3218.1	Potri.011G069600	PtiCesA8-A	Potra004051g2 4387	CesA family protein	Fas 2
eugene3.00040363	Potri.004G059600	PtiCesA8-B	Potra000473g0 2869	CesA family protein	Fas 2
gw1.V.3193.1	Potri.005G087500	PtiCesA6-A	Potra000807g0 6409	CesA	Fas 1
estExt_fgenes4_pg.C_LG_VI10650	Potri.007G076500	PtiCesA6-B	Potra000488g0 3037	CesA	Fas 1
estExt_Genewise1_v1.C_LG_11792	Potri.001G266400	PtiCesA3-D	Potra003868g2 3243	CesA family protein	Fas 1, 2
estExt_fgenes4_pg.C_LG_IX0979	Potri.009G060800	PtiCesA3-C	Potra002409g1 8324	CesA family protein	Fas 2, 3
fgenes4_pm.C_LG_XIII000084	Potri.013G019800	PtiCesA6-E	Potra001230g1 0582	CesA	Fas 1
fgenes4_pg.C_scaffold_133000012	Potri.005G027600	PtiCesA6-F	Potra002888g2 0235	CesA	Fas 1, 2, 3
eugene3.00160483	Potri.016G054900	PtiCesA3-B	Potra000167g0 0627	CesA family protein	Fas 1
eugene3.00060479	Potri.006G052600	PtiCesA3-A	Potra002846g2 0119	CesA family protein	Fas 1
fgenes4_pg.C_LG_VI001789	-	PtiCesA1-B	-	-	Fas 1
estExt_fgenes4_pm.C_LG_XVIII0125	Potri.018G029400	PtiCesA1-A	Potra003411g2 1584	CesA	Fas 1
estExt_fgenes4_pg.C_13980001	Potri.006G251900	-	Potra002914g2 0296	CesA	Fas 1
estExt_Genewise1_v1.C_LG_VI2188	Potri.006G181900	PtiCesA7-A	Potra002910g2 0286	CesA family protein	Fas 2
gw1.XVIII.3152.1	-	PtiCesA7-B	-	-	Fas 2
	Potri.018G103900	-	-	CesA family protein	Fas 2
	-	-	<u>Potra003935g3 5108</u>	-	-

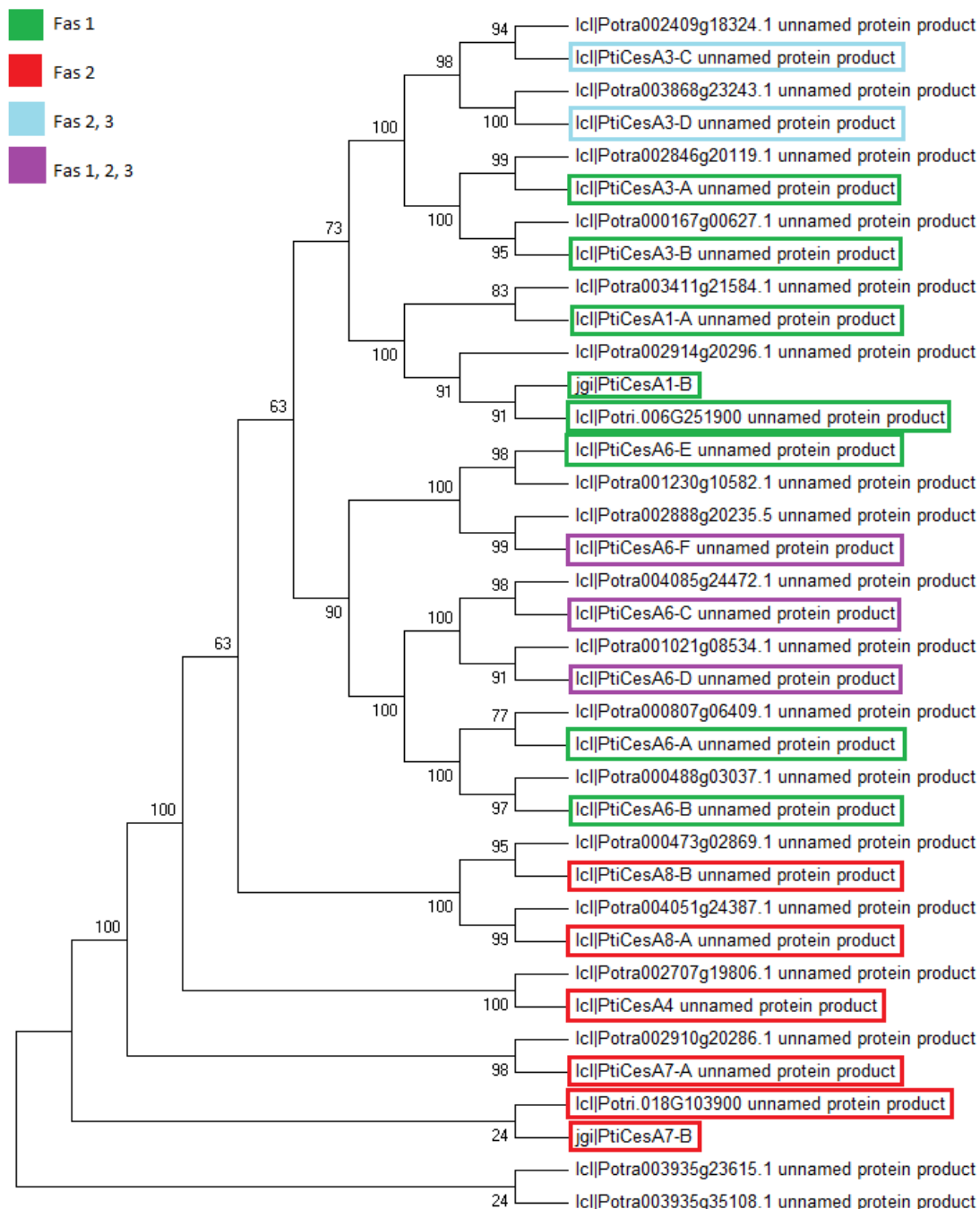
	-	Potra003935g2 3615	
	Potri.019G049700	Potra003265g2 1167	CSL
	Potri.001G136200	Potra002239g1 7201	CSL
	Potri.001G050200	Potra001665g1 3679	CSL
	Potri.003G177800	Potra000882g0 7128	CSL
	Potri.006G004200	Potra000490g0 3080	CSL
	Potri.002G200300	Potra001085g0 9396	CSL
	Potri.014G125100	Potra002583g1 9426	CSL
	Potri.004G208800	Potra001288g1 1102	CSL
	Potri.009G170000	Potra001034g0 8699	CSL
	Potri.001G187900	-	CesA men inget uttryck i veden
	Potri.003G142200	-	CSL
	Potri.013G082200	-	CSL
	Potri.003G097100	-	CSL
	Potri.001G449300	-	CSL
	Potri.012G041500	-	cesA men inget uttryck i veden
	Potri.011G152300	-	CSL
	Potri.003G142200	-	CSL
	Potri.005G154200	-	cesA men inget uttryck i veden

Fylogenetiskt träd

Det fylogenetiska trädet visar att de flesta cesA-generna från *P. trichocarpa* har en motsvarighet i *P. tremula* som finns på samma gren (Se figur 5). Undantagen är PtiCesA7-B som delar gren med Potri.018G103900 samt Potra003935g23615 som delar gren med Potra003935g35108. Båda dessa grenar finns längst ner i trädet och är de som tidigast skiljts från de övriga. CesA7-B har 97,84 % likhet med Potri.018G103900. Potra003935g23615 har 99,43 % likhet i aminosyrasekvens med Potra003935g35108. DNA-sekvensen mellan de båda Potragenerna är identisk.

Siffrorna är vid de flesta delningarna ganska höga (63-100%). Vid en del delningar är siffrorna lägre (det lägsta värdet är 24%). (Figur 5)

Trädet visar även relationen mellan de gener som Kumar et al. (2009) tyckte var samma gen (PtiCesA8-A och PtiCesA8-B). De är nära besläktade enligt trädet. CesA8-A har 58, 12 % likhet i proteinsekvens med cesA-8B och likheten åt andra hållet är 95, 07 %.

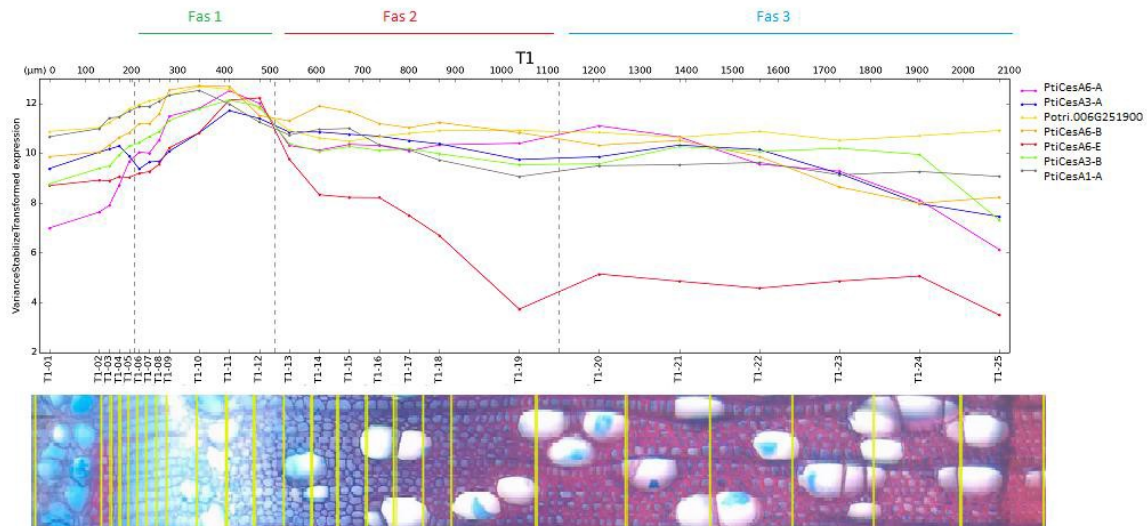


Figur 5. Fylogenetiskt träd över cesA-generna i *P. tremula* och *P. trichocarpa*. Färgmarkeringarna visar under vilka faser i vedbildningen generna uttrycks.

Figure 5. Phylogenetic tree of cesA-genes in *P. trichocarpa* and *P. tremula*. The colors show in which phases the genes are expressed.

Uttrycksmönster hos CesaA-gener under vedbildning

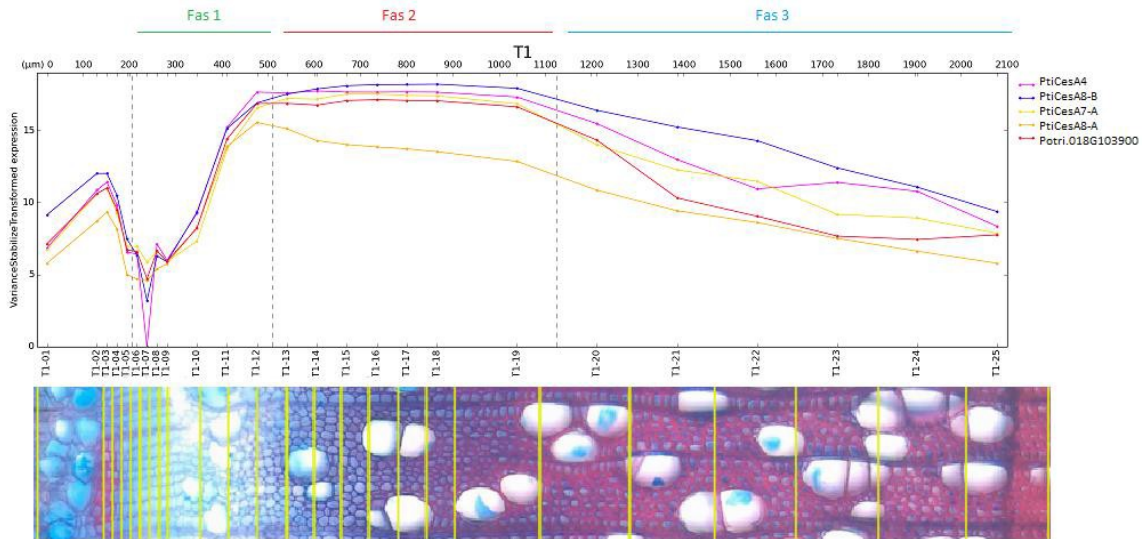
CesaA-generna uttrycks vid olika faser i vedbildningen. Cesa6-A, Cesa3-A, Potri.006G251900, cesA6-B, cesA-3-B och cesA1-A uppvisar alla liknande uttrycksmönster (figur 6). De uttrycks som mest vid bildning av den primära cellväggen i fas 1 men fortsätter att uttryckas, om än minskande, i såväl fas 2 som fas 3. Cesa6-E uttrycks också som mest i fas 1 men är till skillnad från de andra inte mycket uttryckt i fas 2 eller 3.



Figur 6. cesA-gener som uttrycks mest i fas 1. Alla gener i grafen utom Cesa6-E fortsätter att uttryckas även under fas 2 och delvis under fas 3.

Figure 6. cesA-genes that are mostly expressed during phase 1. All genes in the graph apart from Cesa6-E continue to be expressed during phase 2 and partly during phase 3.

De gener som uttrycks mest under fas 2 är Cesa4, cesA8-A, cesA7-A, cesA8-B och Potri. 018G103900. De fortsätter också att uttryckas svagt och minskande under fas 3. (Figur 7)

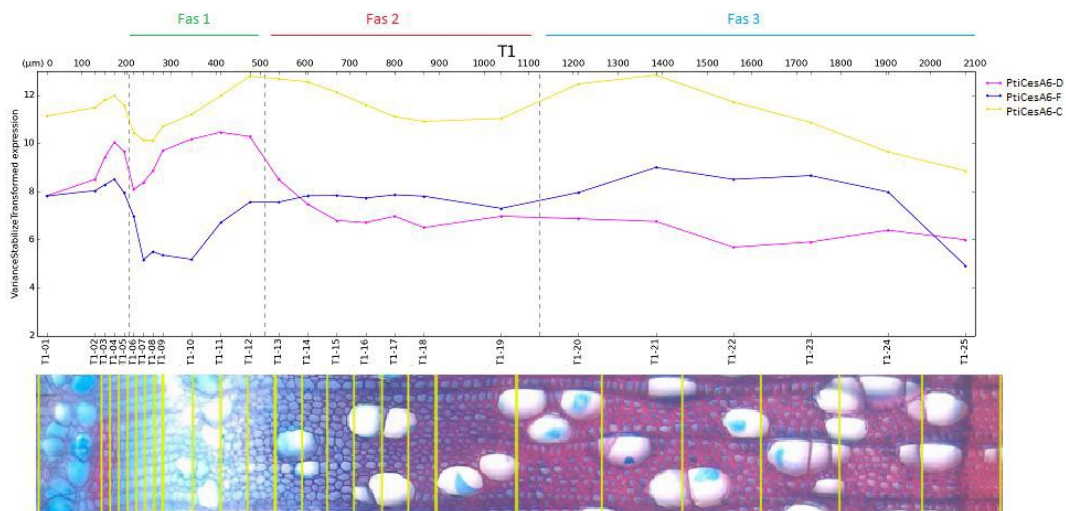


Figur 7. cesA-gener som uttrycks mest i fas 2. De visar i stort sett samma uttrycksmönster. Cesa8-B uttrycks lite mer än de andra under fas 3 och Cesa8-A uttrycks mindre än de övriga under både fas 2 och fas 3.

Figure 7. cesA-genes that are expressed mostly during phase 2. They show quite similar expression patterns. CesA8-B is expressed a bit more than the other genes during phase 3, and CesA8-A is expressed less than the others during phase 2 and 3.

Cesa6-C uttrycks under både fas 1, 2 och 3, med tyngdpunkt i fas 2 och 3 (figur 8).

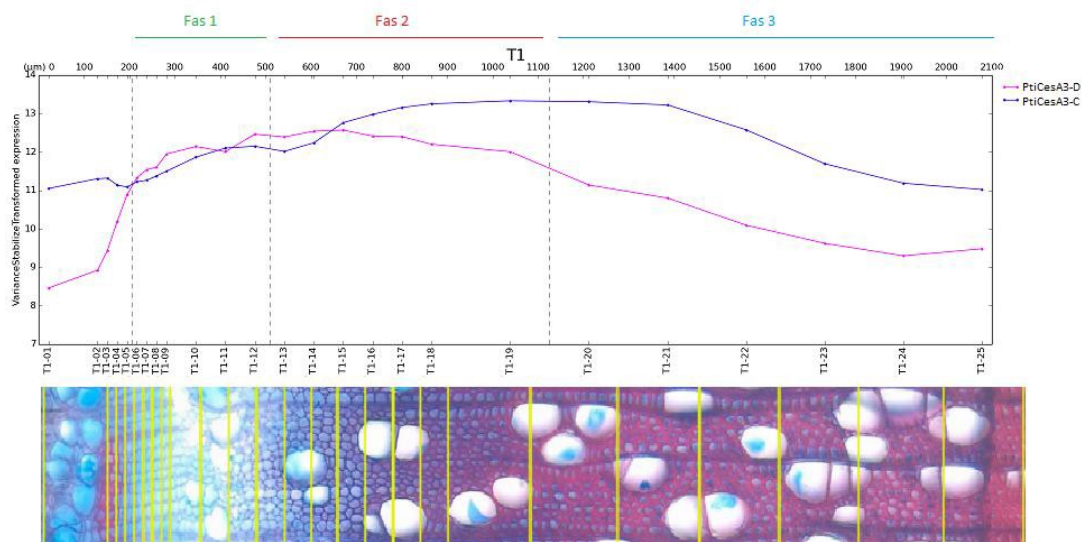
Detsamma gäller för cesA6-F med skillnaden att den uttrycks mest under 3:e fasen. Cesa6-D uttrycks som mest under fas 1, men även en del under fas 2. Generellt uttrycks de på en lägre nivå än de andra CesA-generna. (Notera skalan på y-axeln)



Figur 8. CesA6-D, CesA6-F och CesA6-C uttrycks under alla tre faser i varierande grad.

Figure 8. CesA6-D, CesA6-F and CesA6-C are expressed at different rates during all three phases.

CesA3-C har ett relativt jämnt uttrycksmönster. Den uttrycks mycket under alla tre faserna men är som allra mest uttryckt under fas 2 och 3. CesA3-D är som mest uttryckt under fas 1 och 2 och minskar sedan under fas 3. (Figur 9)



Figur 9. CesA3-D och CesA3-C uttrycks mest under fas 2 och 3, men även en del under fas 1.

Figure 9. CesA3-D and CesA3-C are expressed mostly during phase 2 and 3, but also slightly during phase 1.

De gener som uttrycks mest under den primära cellväggssyntesen återfinns i den övre delen av det fylogenetiska trädet (se figur 5). De gener som uttrycks mest under den sekundära cellväggssyntesen är mer spridda på olika grenar, men merparten av dem finns i den nedre delen av det fylogenetiska trädet. De två översta generna i trädet är mest aktiva i vedbildningens senaste fas men har ett uttrycksmönster som är spritt över flera faser.

Uttrycksmönstren för PtiCesA7-B och PtiCesA1-B har inte kunnat analyseras i Aspwood eftersom de saknar Potri-namn. Enligt Song et al. (2010) är cesA-7B verksam i sekundär cellväggssyntes (fas 2) och cesA-1B i primär (fas 1).

CesA8-A och CesA8-B som Kumar et al. (2009) klassade som samma gen på grund av stor likhet i proteinsekvens visar i Aspwood liknande uttrycksmönster, men på olika nivåer i grafen (se figur 7). CesA8-A uttrycks mindre än cesA8-B.

DISKUSSION

Gener hos *P. Trichocarpa* och deras funktion enligt Aspwood

Vid sökningen hittades 37 gener, varav 15 visade sig vara CSL-gener och tre cesA-gener utan uttryck i vedbildningen. Dessa räknades bort (men redovisas i tabell 1) eftersom vårt mål var att finna cesA-gener som uttrycks under vedbildning. De resterande 19 var sådana och de flesta av dem överensstämmer med de som Kumar et al. (2009) listat. En av dem hade dock inte listats som cesA-gen tidigare: Potri.018G103900. Eftersom den ligger på samma gren som cesA7-B (som inte hittades i PopGenIE och därför inte har något Potrinamn) är det rimligt att anta att cesA7-B har ersatts med Potri.018G103900 när genomet uppdaterats efter noggrannare kartläggning. PtiCesA7-B har även stor likhet med Potri.018G103900 (97,84 %) vilket stödjer den teorin. Eftersom det bara är möjligt att söka på Potrinamn i Aspwood går det tyvärr inte att se hur stora skillnader i uttrycksmönster det finns mellan dessa två gener. Om de haft liknande uttrycksmönster hade även det kunnat vara ett argument för att Potri.018G103900 ersatt cesA7-B.

Kumar et al. (2009) föreslår att cesa8-A och cesa8-B är samma gen på grund av att proteinsekvenserna för dessa är väldigt lika. Enligt vårt fylogenetiska träd är de nära besläktade, och efterforskningar i PopGenIE visade att cesA8-A har 58, 12 % likhet i proteinsekvens med cesA-8B och likheten åt andra hållet är 95, 07 %. Den nära fylogenin tillsammans med den höga likheten i proteinsekvens tyder på att de är nära besläktade, men inte nödvändigtvis samma gen. De har också olika uttrycksmönster under vedbildningen, vilket kan ses i Figur 7. Därför har vi valt att behandla dem som två olika gener.

Gener hos *P. Tremula*

I genomet hos *P. tremula* hittades 27 gener med över 60 % likhet i proteinsekvens med våra funna cesA-gener i *P. trichocarpa*. Av dem har vi bara tagit med 18 stycken i det fylogenetiska trädet. De resterande nio generna har motsvarigheter i *P. trichocarpa* som inte är cesA-gener utan CSL-gener. I Aspwood kan man bara söka på Potrinamn, och det finns därför ingen möjlighet att kontrollera uttrycksmönstren hos *P. tremula*-generna. Vid bildandet av Aspwood har data från *P. tremula* och inte från *P. trichocarpa* använts, och det har antagits att de är jämförbara med varandra. Därför har vi dragit slutsatsen att de nio generna hos *P. tremula* också är CSL-gener och att de inte bör vara med i det fylogenetiska trädet.

Fylogenetiskt träd

I det fylogenetiska trädet valde vi att bara ta med de cesA-gener som uttrycks i vedbildningen enligt Aspwood, alltså 19 gener från *P. trichocarpa* och 18 gener från *P. tremula* (figur 5). De flesta generna från *P. trichocarpa* har en motsvarande gen från *P. tremula* på samma gren,

vilket indikerar att *cesA*-generna har bibehållits i de båda arterna sedan *Populus*-släktets delning. Några undantag finns dock: *CesA1-B* och *Potri.006G251900* finns tillsammans med *Potra.0022914g20296* på en gren, men *cesA1-B* finns inte att hitta i PopGenIE längre. Vår tolkning är att *Potri.006G251900* är den korrekta genen och att *cesA-1B* är en gammal, inaktuell genmodell. Både *Potri.006G251900* och *cesA-1B* finns listade bland *cesA*-generna i Kumar et al. (2009). En möjlig förklaring kan vara att man funnit att *cesA1-B* har en annan funktion när genomet kartlagts mer noggrant, alltså att den i själva verket inte är en *cesA*-gen. Den ursprungliga genen kan också ha ersatts av en annan genmodell efter mer noggrann sekvensering.

Näst längst ner i trädet finns en gren endast innehållande två gener från *P. trichocarpa*: *cesA-7B* och *Potri.018G103900* (figur 1). Den sistnämnda finns inte med i listan av Kumar et al. (2009) och *cesA-7B* finns inte att hitta i PopGenIE, vilket kan innebära att *Potri.018G103900* ersatt *cesA-7B*. *CesA7-B* har 97,84 % likhet med *Potri.018G103900*, de är alltså väldigt lika, vilket stödjer teorin om att *Potri.018G103900* ersatt *cesA7-B*.

Längst ner finns också en gren med enbart två gener från *P. tremula*: *Potra003935g35108* och *Potra003935g23615*. *Potra003935g23615* har 99,43 % likhet i proteinsekvens med *Potra003935g35108* vilket innebär att även de skulle kunna vara samma gen. Om man istället tittar på DNA-sekvensen ser man att den är identisk för de båda generna. Det styrker teorin om att dessa två gener är varianter av samma gen. Här är värt att notera att *P. trichocarpagenomet* är mycket mer genomarbetat i PopGenIE än *P. tremulagenomet*, därför är det möjligt att flera varianter av vissa *tremulagen* finns i databasen.

Bootstrapvärdena i det fylogenetiska trädet är vid många av delningarna över 90 % (figur 5) vilket innebär att de delningarna med stor sannolikhet är korrekta. Ett par tidiga delningar har bootstrapvärden på 63 och 72, alltså en lägre sannolikhet. De är dock så pass höga att de kan räknas som trovärdiga (Baisson, J²). Två av delningarna i trädet har emellertid låga värden, de återfinns längst ner i trädet. Det är grenarna med enbart två gener från *P. trichocarpa* respektive *P. tremula*. Delningarna mellan både *tremulagen*erna och *trichocarpagen*erna har bootstrapvärdet 24, vilket innebär att de inte är så trovärdiga. Möjligheten att de två generna på varje gren egentligen är varianter av samma gen har tidigare diskuterats. Det skulle kunna förklara varför generna tycks vara närmare besläktade med varandra inom arten än med motsvarande gen från den andra arten. Om så är fallet hade det varit bättre att ha endast gen från varje art med i trädet.

Uttrycksmönster hos *CesA*-gener under vedbildning

De gener som är aktiva i fas 1 respektive fas 2 var lättast att urskilja och graferna över de grupperna visar tydliga samband (se figur 6 och 7). Alla gener som uttrycks mest i fas 1 utom *cesA6-E* fortsätter att visa höga, om än minskande värden, genom fas 2 och 3. Det innebär att

² John Baisson, PostDoc på Umeå Plant Science center, 23 mars 2016

de är verksamma genom hela bildnings- och mognadsprocessen av veden. Generna som uttrycks mest under fas 2 (figur 7) visar däremot en tydlig topp under just den sekundära cellväggsbildningen och uttrycks inte mycket under fas 1. De uttrycks fortfarande mycket i början av fas 3 men minskar en hel del. Notera även skalan på y-axeln, som visar att dessa gener uttrycks starkare än de andra gör i den mest aktiva fasen. Det verkar också som att flest gener är aktiva under fas 2. Den sekundära cellväggen, som bildas i fas 2, utgör en stor del av totala andelen cellvägg vilket kan vara orsaken till att många gener är inblandade och uttrycks starkt under den fasen.

De resterande generna visar mer otydliga uttrycksmönster, men vi har ändå valt att dela in dem i två grupper (figur 8 och 9). *CesA6-C* uttrycks relativt mycket i alla tre faser. Den har liknande uttrycksmönster med *cesA6-D*, men ligger på en högre nivå (se figur 8). *CesA3-C* ser ut att vara den gen som har mest avvikande uttrycksmönster då den uttrycks med ett högt värde under alla tre faserna (se figur 9). Det kan därför vara extra intressant att undersöka dess funktion vidare för att se om den har någon speciell roll under vedbildningen. De gener som uttrycks under fas 3 kan antas vara aktiva under märgcellernas cellväggssyntes. Märgcellerna är aktiva under flera år och utvecklar sin sekundära cellvägg senare än andra celler (Nakaba et al. 2012). Det är alltså främst de som är aktiva under vedbildningens sista fas.

Gener utan Potrinamn (*PtiCesA1-B* och *PtiCesA7-B*) kunde inte hittas i Aspwood. De finns dock med i en studie av Song et al. (2010), där proteinet från *PtiCesA1-B* beskrivs som typ II (primär och sekundär) och proteinet från *PtiCesA7-B* beskrivs som typ I (sekundär). Eftersom vi inte kunnat undersöka det själva har vi valt att utgå ifrån att deras analys stämmer.

De gener som är mest aktiva under fas 2 är samlade i den nedre delen av det fylogenetiska trädet (se figur 5) och är därmed nära besläktade med varandra. Likaså de gener som är mest aktiva under fas 1, de är samlade i den övre halvan av trädet. I den övre delen av trädet finns också generna som uttrycks under alla tre faser. Enligt Song et al. (2010) är dessa gener mest uttryckta under fas 1 och 2 men vår tolkning är att en del av dem även är aktiva under fas 3. Det är intressant att de ligger mitt bland de gener som tydligt uttrycks under fas 1. Eftersom uttrycksmönstren skiljer sig mycket vore det rimligt om de var grupperade för sig själva. *CesA3-C* och *cesA3-D* uttrycks mest under fas 2 men i varierande grad även under fas 1 och 3. I det fylogenetiska trädet är de grupperade längst upp, utanför fas 1-gruppen. Placeringen av dem känns logisk eftersom uttrycksmönstren skiljer sig så pass mycket från de som uttrycks mest under fas 1 (se figur 6 och 9).

Förslag på förbättringar och vidare forskning

För att avgränsa sökningen valdes 60 % likhet i proteinsekvens med redan kända *cesA*-gener enligt Kumar et al. (2009). Vilken gräns de använde i den studien anges inte, men de lägsta värdena är 77 %, i jämförelsen mellan *A. thaliana* och *P. trichocarpa*. För att bredda sökningen och öka chansen att hitta eventuella ytterligare *cesA*-gener valde vi att sänka gränsen till 60 %. I efterhand visade det sig dock att den var lite väl låg eftersom resultatet

förde med sig CSL-gener och de flesta av dem låg just över 60 %. En gräns på 65-70 % hade kanske varit bättre, men att sätta gränsen för högt medför ju samtidigt en risk att missa några gener som faktiskt är cesA-gener och inte CSL-gener.

PopGenIE är en databas som uppdateras kontinuerligt i takt med att tekniker för att sekvensera genom förbättras. Genomet hos *P. trichocarpa* har bearbetats och uppdaterats sedan sökningen som Kumar et al. (2009) gjorde. Därför är det möjligt att gener som de fann har ändrats/försvunnit/ersatts med andra gener som kan ha liknande men inte identisk sekvens. Detta är en möjlig förklaring till att våra resultat inte överensstämmer helt med de från Kumar et al. (2009). Sökningen efter cesA-gener i PopGenIE gjordes helt manuellt, vilket medför att den mänskliga faktorn är en eventuell felkälla. Det är dock inte särskilt troligt eftersom de flesta gener dök upp flera gånger under sökningen. För att missa en gen helt ska alltså samma gen missas vid upprepade tillfällen. Ett frågetecken kring datat i PopGenIE är proteinsekvensernas likhet med varandra. I vissa fall när vi sökt på en gen visar det sig att den har ned till 60 % likhet med sig själv, fast den rimligtvis borde vara 100 %. Den är då mer lik andra gener än sig själv vilket är ett mysterium som vi står utan svar på.

I Aspwood är det bara möjligt att söka efter gener hos *P. trichocarpa*, även om databanken är uppbyggd av data från *P. tremula*. När databasen bildades antogs att de båda arternas genom är så pass lika att deras gener är jämförbara med varandra. Det vore intressant att veta varför de drog den slutsatsen, och vilka belägg som finns för det. Om det finns osäkerheter angående detta vore det intressant att kunna se uttrycksmönster från cesA-gener i båda arterna och se om de verkligen är så lika.

I det här arbetet har vi valt att avgränsa oss till cesA-gener och inte tittat närmre på de CSL-gener vi hittat. De tros vara inblandade i syntesen av hemicellulosa som också används till vedbaserade förnyelsebara produkter (Innventia 2016). Därför vore CSL-generna också ett bra område att forska vidare på, för att kunna effektivisera och förbättra användandet av hemicellulosa.

Det finns ett behov av en större och mer noggrann analys av släktskapet mellan cesA-generna. De flesta bootstrappvärdena är höga så det är stor sannolikhet att det fylogenetiska trädet stämmer, med undantag av de nedersta grenarna. För att få bättre sannolikhet även för dem skulle det vara bra att göra en större analys. Ett alternativ är att inkludera en "outgroup", en grupp gener som inte är cesA-gener, för att få en bättre överblick av genernas placering i genomet (UW-Madison Department of Botany).

Uttrycksmönstren hos generna är också ett fascinerande område att forska vidare på, särskilt de gener som uttrycks under fas 3. Under den fasen dör cellerna, med undantag från cellerna i märkestrålarna. Därför är det troligt att de gener som uttrycks under den fasen är aktiva i märkestrålarnas cellväggsbildning. Det verkar inte finnas så mycket forskning gjord på cellväggsbildning hos just märkestråleceller. Därför vore det intressant att undersöka dessa mer ingående.

De tre olika lager som finns i den sekundära cellväggen (S1, S2 och S3) skiljs åt beroende på cellulosedjornas (mikrofibrillernas) vinkel, och mikrofibrillvinkeln är mycket viktig för vedegenskaperna (Smith et al., 2010). En intressant inriktning på vidare forskning vore därför att försöka ta reda på om det är olika gener som är aktiva under bildningen av S1, S2 och S3. Det skulle kunna underlätta förädlingen av trä för att få önskade vedegenskaper.

Framtida forskning kan också fokusera på andra träarter. För Sveriges del är det intressant med barrträd eftersom det till största delen är de som odlas här och används inom industrin. Gran är Sveriges vanligaste träslag (Skogsstyrelsen 2010), och dess genom är sekvenserat. Man skulle kunna söka efter och jämföra cesA-generna hos granen och se om de skiljer sig från aspens, samt titta på uttrycksmönster hos granens cesA-gener.

Vidare forskning kring genetiken bakom vedbildning känns mycket viktig. I och med att miljömedvetenheten ökar och det blir mer tydligt att vi människor måste ta bättre hand om jorden, är behovet av att utveckla och effektivisera användandet av förnyelsebara resurser högst aktuellt. Eftersom träråvaran redan används till mycket känns det som en bra väg att fortsätta på. Både cellulosa och hemicellulosa är på uppgång som komponent i allt från förpackningar till kläder. Genom att fortsätta forskningen kring den bakomliggande genetiken kan användandet effektiviseras och vara en del av lösningen för en mer hållbar värld.

REFERENSER

Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., (2010) *Essential Cell Biology*. 3rd edition. New York: Garland Science.

Cervera MT, Storme V, Soto A et al. (2005) Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theor Appl Genet* 111: 1440–1456.

Djerbi, S., Aspeborg, H, Nilsson, P, Sundberg, B., Mellerowicz, E., Blomqvist, K. and Teeri, T.T. (2004). Identification and expression analysis of genes encoding putative cellulose synthases (CesA) in the hybrid aspen, *Populus tremula* (L.) *P. tremuloides* (Michx.). *Cellulose* 11: 301–312.

Djerbi, S. et al. (2005). The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* 221: 739–746

Jansson, S., Bhalerao, R.P. & Groover, A. T. (2010). *Genetics and genomics of Populus*. Vol. 8. Springer science + business media: New York.

Jarvis, M.C. (2013). Cellulose Biosynthesis: Counting the Chains. *Plant Physiol.* 163, 1485–1486.

Kim, S., Mardis, E., Tang, H. (2008) *Genome Sequencing Technology and Algorithms*. 1st edition. Norwood:Artech House Books

Kumar et al., (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. *Cell press.* 14:5, pp 248-254.

Kumar, M., Turner, S., (2015) Plant cellulose synthesis: CESA proteins crossing kingdoms. *Phytochemistry*, Vol 112, pp 91-99

Li, L.G. et al. (2006) A genomic and molecular view of wood formation. *Crit. Rev. Plant Sci.* 25, 215–233

Nakaba, S., Begum, S., Yamagishi, Y., Jin, H-O., Kubo, T. and Funada, R. (2012) Differences in the timing of cell death, differentiation and function among three different types of ray parenchyma cells in the hardwood *Populus sieboldii* 3 *P. grandidentata*. *Trees*. Vol. 26, pp. 743-750

Sjödin A, Street NR, Sandberg G, Gustafsson P and Jansson S (2009). *New Phytologist*. 1013-1025.

Smith, A.M., Coupland, G., Dolan, L., Harberd, N., Jones, J., Martin, C., Sablowski, R. & Amey, A. (2010). *Plant biology*. 1st edition. New York: Garland Science.

Song, D. et al. (2010) Characterization of cellulose synthase complexes in *Populus* xylem differentiation. *New Phytologist* Vol 187. pp. 777–790

Sterck, L., Rombauts, S., Jansson, S., Sterky, F., Rouzé, P., Van De Peer, Y. EST data suggest

that poplar is an ancient polyploid. *New Phytologist*, 2005, Vol.167(1), pp.165-170

Suzuki, S. et al. (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthaselike genes in *Populus trichocarpa*. *Plant Physiol.* 142, 1233–1245

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, Vol 408. pp. 796-815

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray ex. Brayshaw). *Science* 313:1596–604

Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57-63

Wang, T., McFarlane, H.E., Persson, S. (2016). The impact of abiotic factors on cellulose synthesis. *Journal of Experimental Botany*, Vol. 67, No. 2 pp. 543–552

Winzell, A., (2007) *Analysis of Genes and Proteins Involved in Wood Formation*. Lic. -avh. Stockholm: Kungliga Tekniska Högskolan

Webbsidor

Ek, M., Illergård, J. Kungliga tekniska högskolan (2013-01-11) *Mars: Cellulosa- Från makro till mikro*. Tillgänglig: <https://www.kth.se/che/archive/arkiv/matmar-1.82408> [2016-02-24]

European Bioinformatics Institute (2015-10-15). *Clustal omega*. Tillgänglig: <http://www.ebi.ac.uk/Tools/msa/clustalo/> [2016-03-09]

Gentekniknämnden, Karolinska Institutet Science park. (2014-08-28). *Växtförädling*. Tillgänglig: <http://genteknik.nu/vaxtforadling/>. [2016-03-03]

Innventia. (2016) *Hemicelullosa*. Tillgänglig: <http://www.innventia.com/sv/Det-har-kan-vi-Nya-material/Barriarer-och-filmer/Hemicellulosaprodukter/>. [2016-03-31]

Skogsstyrelsen (2010-10-18). *Gran*. Tillgänglig: <http://www.skogsstyrelsen.se/Upptack-skogen/Upplev-skogen/Om-skogen/Tradslag/Gran/> [2016-03-30]

UW-Madison Department of Botany, University of Wisconsin-Madison. (2016-03-17). *Data for phylogenetic analysis*. Tillgänglig: http://www.botany.wisc.edu/courses/botany_563/563_readings/Data%20for%20Phylogenetics.pdf. [2016-03-19]

Opublicerat material

Savva, G. (2003) Introduction to Phylogenetic Analysis. Booklet for the course “Introduction to phylogenetic analysis”