Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Montes

**Respuesta transcripcional de *Populus* a
bifenilos policlorados: Aspectos aplicados a la
fitorremediación**

TESIS DOCTORAL

GABRIELA OROZCO GUTIÉRREZ

Ingeniera Bioquímica

Madrid, 2016

Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Montes

Autor:

Gabriela Orozco Gutiérrez

Ingeniera Bioquímica

Maestra en ciencias en Agrobiotecnología

Director:

Luis Gómez Fernández

Doctor en Biotecnología Agroforestal

Madrid, 2016

POLITÉCNICA

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad
Politécnica de Madrid, el día de de 20

Presidente:

Secretario:

Vocal:

Vocal:

Suplente:

Suplente:

Realizado el acto de defensa y lectura de la Tesis el día de de 20 en la E.T.S.I.

Montes.

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

A mi hija Gabrielita con amor incondicional

RECONOCIMIENTOS

El trabajo presentado en esta Tesis se ha desarrollado en el Centro de Biotecnología y Genómica de Plantas (CBGP) del Campus de Excelencia Internacional de Montegancedo, perteneciente a la UPM.

La Tesis ha sido realizada gracias a la financiación de diversas instituciones principalmente por la beca otorgada por el CONACYT (Consejo Nacional de Ciencia y Tecnología), la fuente de financiación mayoritaria. Además, es muy importante para mí mencionar la confianza y el apoyo económico otorgado por el INIFAP (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias) la cual es una Institución de investigación pública del gobierno federal mexicano, de la cual formo parte orgullosamente. Finalmente, se contó con el apoyo de entidades públicas y privadas Españolas para la compra y adquisición de materiales necesarios para mi proyecto, estos incluyen proyectos de I+D o convenios universidad-empresa.

Aparte de la financiación, me gustaría agradecer a las siguientes personas sin las cuales los resultados no hubiesen sido los mismos:

Gonzalo Claros y a su apreciable equipo de trabajo, en especial a Rocío, Hicham, Pedro e Isa todos del SCBI: supercomputación (RES y UMA) y Bioinformática (PAB), centro perteneciente a el departamento de Biología Molecular y Bioquímica de la Universidad de Málaga, por la oportunidad de aprender bioinformática. Tambien me es importante mencionar al Dr Luis Oñate y al Dr. Juan Florencio Gómez Leyva por su ayuda desinteresada.

AGRADECIMIENTOS

Quiero que estas líneas sirvan para expresar mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización del presente trabajo, en especial al Dr. Luis Fernández Gómez, director de esta investigación, por la orientación, el seguimiento y la supervisión de la misma, gracias.

Aemás, un especial agradecimiento por el interés mostrado por mi trabajo y las sugerencias recibidas de mis compañeras de laboratorio, con los que me encuentro en deuda por los conocimientos transmitidos. Les agradezco el excelente ambiente del laboratorio y su cariño sincero.

Ángela, gracias por tu ayuda y por lo mucho que me enseñaste, eres una gran mujer y amiga, tu tenacidad e intelecto contagia a cualquiera, nunca te rindas.

Julia, gracias eres un luz con estrella, gracias infinitas por todo lo que me apoyaste y por los conocimientos transmitidos, te admiro mucho.

Irene, gracias por la fuerza que transmites. Por enseñarme y compartir tus conocimientos. Tienes todo para ser una grande.

Álvaro, gracias por tu ayuda desinteresada, por tus ocurrencias, comida y chascarrillos que contagiaron de alegría al laboratorio.

A Felipe, por ser mi apoyo en todo momento. Pero sobre todo por enseñarme que no hay obstáculos grandes para un gran esfuerzo.

A mis padres Héctor y Teresa y a mis hermanos Juan Carlos y Héctor José, por quererme tanto. Nunca importó la distancia física porque siempre los he sentido cerca. Gracias por confiar en mí y por apoyarme sin límites y por levantarme cuando me he sentido derrotada.

A mis amigas y amigos que me esperan y que piden por mí todos los días en especial a Kathy, Alfa, Fabiola, Lucy, Gilda, gracias!!. También a Gabriel, por tu invaluable ayuda y compañía durante años pero sobre todo por haberme hecho la madre de la niña más maravillosa: Gabrielita, mi niña hermosa que me da fuerza y entereza todos los días.

Finalmente, ¡Gracias a España! Hermoso país que me ha acogido y me ha hecho sentir una más. Pero sobre todos y todo a Dios, por darme fuerza, paz y alegría en cada paso.

ÍNDICE

RECONOCIMIENTOS	XI
AGRADECIMIENTOS	XII
ÍNDICE DE FIGURAS	XVII
ÍNDICE DE TABLAS	XIX
ABREVIATURAS	XXI
RESUMEN	XXIII
SUMMARY	XXV
1. Introducción	17
1.1. El suelo como recurso natural amenazado	17
1.2. Contaminación por Bifenilos Policlorados (PCBs)	20
1.3. Remediación de suelos contaminados con PCBs	27
1.4. Fitorremediación: perspectivas moleculares	29
1.5. <i>Populus</i> : arbórea modelo con potencial remediador y en investigación molecular	31
1.6. Genes y factores reguladores relacionados con el metabolismo de xenobióticos	33
1.6.1. Factores de transcripción MYB	35
1.6.2. Factores de transcripción WRKY	36
1.6.3. Factores de transcripción NAC	37
1.7. Las tecnologías de secuenciación de nueva generación (NGS)	37
1.8. Proteómica	38
1.8.1. Transcriptómica	39
1.8.1.1. RNA-seq	39
1.7.1.1. Procesamiento de lecturas	40
2. Objetivos	45
3. Material y Métodos	49
3.1. Material vegetal	49
3.2. Cepas bacterianas	49
3.3. Plásmidos	50
3.4. Oligonucleótidos	50
3.5. Métodos de plantas	51
3.5.1. Cultivo de <i>Populus tremula</i> x <i>P. alba</i>	51
3.5.2. Cultivo de <i>Arabidopsis thaliana</i>	51
3.5.3. Tratamientos con Aroclor (PCBs)	52
3.6. Métodos de ácidos nucleicos	52

3.6.1.	Extracción de DNA genómico de <i>Populus tremula</i> x <i>P. alba</i>	52
3.6.2.	Extracción y secuenciación de DNA plasmídico de <i>E. Coli</i>	53
3.6.3.	Extracción de RNA total de <i>Populus tremula</i> x <i>P. alba</i> <i>in vitro</i>	53
3.7.	Métodos de amplificación	54
3.7.1.	PCR (Reacción en cadena de la polimerasa)	54
3.7.2.	Tratamiento con DNAsas y síntesis de cDNA	54
3.7.3.	Cuantificación en tiempo real (qRT-PCR)	55
3.8.	Análisis de RNA-seq	55
3.8.1.	Construcción y secuenciación de librerías de cDNA	55
3.9.	Métodos de plantas	56
3.9.1.	Cultivo de chopo	56
3.9.2.	Cultivo de <i>Arabidopsis thaliana</i>	56
3.9.3.	Tratamientos con Aroclor 1221 (PCBs)	57
3.10.	Métodos de ácidos nucleicos	58
3.10.1.	Extracción de DNA genómico de chopo	58
3.10.2.	Extracción y secuenciación de DNA plasmídico	58
3.10.3.	Extracción de RNA total de chopo	58
3.11.	Métodos de amplificación	59
3.11.1.	PCR (Reacción en cadena de la polimerasa)	59
3.11.2.	Tratamiento con DNAsas y síntesis de cDNA	60
3.11.3.	PCR cuantitativas en tiempo real (qRT-PCR)	60
3.12.	Análisis de RNA-seq	61
3.12.1.	Construcción y secuenciación de librerías de cDNA	61
3.12.2.	Pre-procesamiento de secuencias	62
3.12.3.	Alineamiento sobre el genoma de referencia <i>P. trichocarpa</i>	62
3.12.4.	Genes diferencialmente expresados (DEGs)	63
3.12.4.1.	Análisis de series temporales	64
3.12.5.	Anotación funcional	64
3.12.6.	Análisis de redes de co-expresión génica	65
3.12.7.	Validación de los datos de expresión génica	65
3.13.	Métodos de proteínas	66
3.13.1.	Extracción y precipitación de proteínas	66
3.13.1.1.	Cuantificación de proteínas	66
3.13.2.	Electroforesis unidimensional en geles de poliacrilamida con SDS	67
3.13.3.	Electroforesis bidimensional de alta resolución	68
3.13.3.1.	Primera dimensión: Isoelectroenfoque (IEF)	68
3.13.3.2.	Segunda dimensión: SDS-PAGE	69

3.13.4.	Tinción de proteínas	70
3.13.5.	Identificación de la proteína PtCSP4	70
3.14.	Análisis filogenéticos	71
3.14.1.	Base de datos para alineamientos y análisis filogenéticos	71
3.15.	Método de clonaje	71
3.15.1.	Preparación de células y transformación con <i>Agrobacterium</i>	72
3.15.2.	Transformación genética de <i>A. thaliana</i>	73
4.	Resultados	77
4.1.	Crecimiento vegetativo en presencia de Aroclor 1221	77
4.1.2.	Secuenciación de RNA de <i>Populus in vitro</i> tratados con Aroclor 1221	79
4.1.2.1.	Calidad y cantidad de RNA	79
4.1.2.2.	Calidad y robustez de la secuenciación	79
4.1.3.	Pre-procesamiento de secuencias	82
4.1.4.	Mapeo de lecturas en el transcriptoma y genoma de referencia	83
4.1.5.	Cuantificación de la expresión génica	84
4.1.5.1.	Genes que interaccionan en los cuatro tratamientos	85
4.1.5.2.	Análisis secuencial (<i>Time course</i>)	92
4.2.	Genes relacionados con el metabolismo de degradación de PCBs	98
4.3.	Fraccionamiento unidimensional de proteicos	107
4.3.1.	Detección de la proteína PCSP4 en mapas bidimensionales	108
4.3.2.	Gen PtCSP4 (Potri.004G172600)	111
4.3.	Identificación de proteínas homólogas en <i>P. trichocarpa</i>	111
4.4.	Alineamiento múltiple de proteínas CSP en plantas	114
4.4.1.	Análisis filogenético de la familia CSP en plantas	118
4.4.2.	Análisis de regiones conservadas	121
4.5.	Expresión heteróloga de PCSP4	123
4.6.	Análisis de expresión diferencial	125
4.7.	Selección de líneas transformadas	126
4.7.1.	Nivel de expresión del transgén	126
4.7.2.	Tolerancia de las líneas transgénicas a Aroclor 1221	128
5.	Discusión	132
5.1.	Efectos de la presencia de Aroclor 1221 en chopo	133
5.2.	Secuenciación por RNA seq	134
5.3.	Calidad de la secuenciación y pre-procesamiento de lecturas	135
5.4.	Cuantificación de la expresión diferencial	136

5.5. Anotación funcional de genes diferencialmente expresados	137
5.6. Selección de genes y factores de transcripción	139
5.7. Genes y factores de transcripción que participan en el metabolismo de PCBs	140
5.8. PCSP4 candidato fitorremediador	151
5.8.1. Análisis filogenético de la familia CSP en plantas	152
5.9. Sobre-expresión de PCSP4 en <i>Arabidopsis</i> aumenta la tolerancia a PCBs	155
6. Conclusiones	160
7. Bibliografía	164

ÍNDICE DE FIGURAS

Figura 1. Degradación del suelo a nivel mundial..	18
Figura 2. Estructura química y propiedades fisicoquímicas de los PCBs.	20
Figura 3. Nivel acumulativo de PCBs en el suelo a nivel mundial.	22
Figura 4. Concentraciones ($\pm$ DE) de Σ HCHs, Σ DDTs, y Σ PCBs en el Ártico	23
Figura 5. Bioacumulación de PCBs en la pirámide trófica marina.	24
Figura 6. Predicción de concentraciones del PCB-52 en Río Thames.	25
Figura 7. Cloracné causado por envenenamiento con organoclorados.	26
Figura 8. Importancia económica y potencial biotecnológico	32
Figura 9. Regulación de la transcripción.	35
Figura 10. Esquema general del alineamiento de lecturas sobre un genoma de referencia.	41
Figura 11. Representación del flujo de análisis del RNA-seq.	61
Figura 12. Efecto del Aroclor 1221 en plántulas de <i>Populus</i> .	78
Figura 13. Análisis de electroforesis nanocapilar de RNA.	79
Figura 14. Saturación de las lecturas obtenidas respecto a la referencia.	80
Figura 15. Cobertura de cada transcrito identificado.	81
Figura 16. Resumen de las lecturas generadas por dos programas bioinformáticos para el pre-procesamiento de las secuencias obtenidas para las 5 librerías.	82
Figura 17. Transcritos que se acumulan de forma diferencial en las comparaciones.	85
Figura 18. Patrón de expresión mostrado como “heat map”	86
Figura 19. Diagrama de Venn de todas las comparaciones.	87
Figura 20. Scatter plot de la distribución de transcritos que se acumulan de forma diferencial en alguna de las comparaciones llevadas a cabo.	88
Figura 21. MA-plots mostrando el cambio (fold change) en abundancia de los transcritos.	89
Figura 22. Perfiles de expresión normalizados por Z- score.	91
Figura 23. Patrones de expresión detectados por el programa maSIGpro.	93
Figura 24. Términos GO de función molecular para los genes expresados diferencialmente usando un modelo de distribución y una corrección de Bonferroni (p-value < 0,05).	95
Figura 25. Agrupamiento de los genes expresados diferencialmente en los cuatro tratamientos según su clasificación GO.	96
Figura 26. Análisis KEGG de enriquecimiento de rutas metabólicas de los transcritos expresados diferencialmente.	98
Figura 27. Análisis empírico de expresión diferencial mediante qRT-PCR.	102
Figura 28. Visualización de la co-expresión entre genes con el programa exNet (expression network).	106
Figura 29. Análisis mediante electroforesis unidimensional SDS-PAGE.	107
Figura 30. Análisis del proteoma de chopo mediante fraccionamiento IEF x SDS-PAGE.	108
Figura 31. Análisis MS a partir de las electroforesis 2D.	109
Figura 32. Cuantificación de la proteína PCSP4 inducida por PCB.	110
Figura 33. Secuencia de la región codificante del gen PtCSP4 <i>P. trichocarpa</i> y la traducción de la misma (624 pares de bases).	111
Figura 34. Ubicación de las secuencias que codifican CSPs en los cromosomas de chopo	112

Figura 35. Alineamiento múltiple de secuencias aminoacídicas mediante CLUSTAL W (1,82) deducidas para la familia CSP.	113
Figura 36. Alineamiento múltiple de secuencias aminoacídicas de la familia CSP con CLUSTAL W (1.82).	116
Figura 37. Dendrograma consenso inferido por el método Neighbor-Joining.....	119
Figura 38. Dominios conservados agrupados por familias.	120
Figura 39. Analisis de secuencias con MEME (Motif-based sequence analysis tools).	122
Figura 40. Comprobación del clonaje de PtCSP4 en geles de agarosa con enzimas de restricción.....	124
Figura 41. Clonaje de la región codificante de PtCSP4 en el vector pB2GW7.	124
Figura 42. Expresión de PtCSP4 en plantas de Arabidopsis.....	125
Figura 43. Expresión de PtCSP4 en plantas transgénicas de Arabidopsis.	128
Figura 44. Líneas transgénicas con mayor tolerancia al contaminante.....	129
Figura 45. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs.	143
Figura 46. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs..	145
Figura 47. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs..	149

ÍNDICE DE TABLAS

Tabla 1. Las propiedades principales de las mezclas comerciales de PCBs y sus congéneres	21
Tabla 2. Clasificación de las tecnologías de remediación de suelos contaminados con PCBs	28
Tabla 3. Ejemplos selectos de fitorremediación de PCB y organoclorados en el periodo 2010-2015.	29
Tabla 4. Identificación de factores de transcripción asociados a la respuesta a xenobiótico	34
Tabla 5. Programa de fraccionamiento IEF de proteínas. Voltaje y tiempo aplicados a la fuente de alimentación.	68
Tabla 6. Resumen estadístico de genes alineados provenientes de las 5 librerías, teniendo como referencia el genoma completo de <i>P.trichocarpa</i>	83
Tabla 7. Resumen estadístico de genes alineados provenientes de las 5 librerías, teniendo como referencia el transcriptoma.	83
Tabla 8. Lista de posibles genes que podrían participar en el metabolismo degradativo de Aroclor.	99
Tabla 9. Oligonucleótidos diseñados para la validación	101
Tabla 10. Correlación entre valores de expresión obtenidos con RNA-seq y con qRT-PCR.	103
Tabla 11. Analisis de Pfam (familia de proteínas) mediante el programa TopGO. El color rojo/naranja (alta significancia) a amarillo (baja significancia).	104
Tabla 12. Analisis de enriquecimiento biológico por el programa TopGO.	105
Tabla 13. Listado de las secuencias CSP identificadas en el genoma de chopo	112
Tabla 14. Proteínas CSP. Se muestra el nombre del locus, la longitud de la proteína (en aminoácidos, AA), las repeticiones RGG y los dominios dedos de zinc	115
Tabla 15. Los oligonucleótidos se diseñaron como se ha indicado previamente, incluyendo sitios attB en sus extremos.	123

ABREVIATURAS

AA: Aminoácido

ABRC: Arabidopsis Biological Resource Center

ATSDR: Agencia de sustancias toxicas y el registro de enfermedades

BASTA: 4-[hidroxi-(metil)-fosfinol]-DL-Homoalanina sal de amonio. (Marca de herbicida)

BGI: Instituto de Genómica de Beijing

CERCLA: Acta de Responsabilidad, Compensación y Recuperación Ambiental

CHAPS: (3-[(3-colamidopropil) dimetilamonio]-1-propanosulfonato)

Cl: Cloro

CNIO: Centro Nacional de Investigaciones Oncológicas

COPs: Contaminantes orgánicos persistentes

CSD: Cold Shock Domain

CSPs: cold shock proteins

Ct: treshold cycle

DHHS: Departamento de Salud y Servicios humanos de USA

DMSO: dimetil sulfóxido

DMSO: dimetilsulfóxido

DTT: Dicloro-difenil-tricloroetano

EDTA: ácido etildiaminotetraacético

EPA: Agencia de protección ambientales

FAO: Organización de las Naciones Unidas para la alimentación y la Agricultura

FDR: False discovery rate

GO: ontología génica

GRPs: Glycine Rich proteins

GSEA: análisis funcional de enriquecimiento de genes

IBA: ácido indol-3-butírico

IEF: Isoenfoque

IPG: gradiente de inmovilización de pH (buffer)

IRIS: Integrated Risk Information System

JTT: Jones Taylor Thornton

KEGGs: Kyoto Enciclopedia de Genes y Genomas

kt: kilo tonelada

LB: Luria- Bertani

LiCl: Cloruro de litio

MALDI-TOF/TOF: Matrix-Assisted Laser Desorption/Ionization-Time Of Flight

MEME: Múltiple EM for motif elicitation

MS/MS: Espectrofotometría de masas

MS: Murashige y Skoog

NBI: National Center for Biotechnology Information

NJ: Neighbor Joining

O/N: Over night

OD: Densidad óptica

pb: pares de bases

PCBs: Bifenilos policlorados

PCR: Reacción en cadena de la polimerasa

PEN: PCBs Elimination network

pH: potencial de hidrogeno

PSA: persulfato

qRT-PCR: Reacción en cadena de la polimerasa cuantitativa en tiempo real

RB: Rehydration Buffer

RIN: número de integridad del RNA

RPKM: lecturas por kilobase por millones de lecturas mapeadas

SCBI: Supercomputing and Bioinnovation Center of the University of Malaga

SDS: dodecilsulfato sódico

SDS-PAGE: Electroforesis en gel de poliacrilamida con dodecilsulfato sódico

SMART: Simple Modular Architecture Research Tool

SOAP: Short Oligonucleotide Analysis Package

TCA: ácido tricloro acético

TE: Tris-EDTA

TEMED: N, N, N, N tetrametilnediamina

TMM: Trimmed Mean of M-value

UNEP: Programa de Naciones Unidas para el Medio Ambiente

La degradación del suelo ha adquirido una magnitud preocupante. Los métodos tradicionales de descontaminación, son costosos e insuficientes. La fitorremediación representa una alternativa eficaz, de bajo coste, respetuosa con el medio ambiente, que además mejora las propiedades del suelo, si bien ha habido desarrollos relevantes en la última década. Desde el punto de vista científico, el reto principal es descifrar las rutas metabólicas implicadas en respuesta a contaminantes y comprender su regulación. Esta información es imprescindible si aspiramos a mejorar las capacidades naturales de algunas especies vegetales para remediar los suelos contaminados.

Los estudios de esta Tesis se han centrado en *Populus*, el mejor modelo forestal disponible a raíz de la secuenciación de su genoma completo. Por otra parte, *Populus* tiene una gran capacidad natural para la degradación de contaminantes orgánicos, lo que explica su predominio en los programas forestales de fitorremediación que se desarrollan actualmente. Hemos elegido en concreto al híbrido *Populus tremula* x *P. alba*, por la facilidad con que se cultiva y su particular interés biotecnológico. La presente Tesis plantea un estudio comprehensivo de la respuesta molecular a bifenilos policlorados (PCBs), una familia de contaminantes orgánicos persistentes de particular relevancia a escala mundial. Se ha utilizado para ello una aproximación transcriptómica, basada en tecnología RNA-seq, para identificar los genes implicados en el metabolismo de los compuestos *in planta* y cuantificar sus niveles de activación en distintas situaciones controladas. La tesis pretende asimismo definir el control transcripcional subyacente a la respuesta bioquímica frente a este tipo de contaminantes. Resulta sorprendente que dicha respuesta sea prácticamente desconocida a nivel molecular, a pesar de su gran potencial aplicado en el contexto de la tecnología fitorremediadora.

Para desarrollar este proyecto aplicamos a nuestros cultivos de chopo híbridos concentraciones diferentes de Aroclor 1221, una mezcla de PCBs muy utilizada a nivel comercial durante décadas, su uso está prohibido hoy internacionalmente. Y tomamos muestras de RNA a dos concentraciones y dos momentos distintos de exposición al contaminante, generando así una matriz de cuatro elementos con sus controles correspondientes. Con el fin de incrementar la especificidad de nuestro análisis, consideramos sobre todo los genes diferencialmente expresados más significativos según cuatro algoritmos estadísticos distintos. Por otra parte, realizamos análisis funcionales con herramientas bioinformáticas basadas en comparaciones de secuencias y en redes de co-expresión génica. La respuesta de los genes de particular interés fue validada mediante tecnología qRT-PCR (reacción de la polimerasa en cadena cuantitativa en tiempo real). Se trata del primer estudio comprehensivo de la respuesta de un organismo vegetal ante la presencia de PCBs.

Este estudio nos ha permitido identificar una cantidad considerable de genes estructurales y reguladores, definiendo nuevos factores de transcripción cuya expresión es proporcional a la concentración de contaminante en el medio o al tiempo de exposición al mismo. Los análisis de correlación nos permiten afirmar en que la respuesta metabólica a PCBs, incluyendo posibles rutas degradadoras, participan en al menos quince factores de transcripción y unas cuarenta proteínas o enzimas que resultan particularmente inducidas. Entre las familias implicadas destacan los citocromos P450, la glutatión transferasas, las deshidrogenasas reductasas (*short-chain dehydrogenase reductase*) y las proteínas MDR (*multi-drug resistance*). Mientras que los factores de transcripción encontrados pertenecen a la familia de ZF-TF, MYBs, WRKYs entre otros.

También identificamos proteínas de función desconocida que no se habían vinculado previamente a este tipo de respuestas en plantas, como la CSP (cold-shock domain proteins). Para estudiar su posible relación con la presencia de PCBs, se caracterizó un gen de esta familia detectado mediante espectrometría de masas en tándem (MS/MS) a partir de mapas IEF x SDS-PAGE (*isoelectro focusing x sodium dodecyl sulphate- polyacrylamide gel electrophoresis*) de alta resolución. Mediante qRT-PCR pudimos confirmar la inducción del gen correspondiente, ortólogo a *PtCSP4* de *P. trichocarpa* (Potri.004g172600), en respuesta a Aroclor 1221. El análisis fenotípico de las líneas transgénicas de *Arabidopsis thaliana* que sobre-expresaba la proteína CSP de chopo híbrido confirmó un papel para la misma tolerancia a PCBs, posiblemente a través de mecanismos reguladores que activan proteínas MDR.

Este trabajo, además de aportar datos novedosos sobre los mecanismos moleculares desencadenados por la presencia de un PCB en *Populus*, utilizado aquí como sistema modelo. Con ello se demuestra el potencial de las especies arbóreas no solo como agentes descontaminantes, ya explotado comercialmente, sino también como fuente potencial de genes interesantes. Entre los genes identificados en esta Tesis hay candidatos evidentes a participar en mecanismos de tolerancia al estrés inducido por la contaminación y también rutas metabólicas degradadores de PCBs. Precisamente la posibilidad de degradar al contaminante confiere particular interés a este tipo de estudios frente a la fitorremediación de metales pesados y otros contaminantes elementales. La comparación de los datos generados en este estudio con estudios análogos que se realicen en el futuro con otras especies y xenobióticos, contribuirán a definir mejor la respuesta de las plantas ante la contaminación orgánica y mejorar su potencial descontaminante.

Soil degradation has acquired a disturbing magnitude. Traditional methods of decontamination are expensive and insufficient. Phytoremediation represent an effective alternative, low cost, respectful of the environment, that also improves soil properties, although there have been relevant developments in the last decade. From a life scientist, the challenge is to decipher the major metabolic pathways involved in response to pollutants and understand their regulation. This information is essential if we desire to enhance the natural abilities of some plant species to remediate contaminated soils.

This thesis studies have focused on *Populus*, the best available forestry model following the sequencing of the entire genome. Moreover, *Populus* has a natural ability to degrade organic pollutants, which explains its predominance in phytoremediation forestry programs currently being developed. We have chosen specifically to hybrid *Populus tremula* x *P. alba*, the ease with which it is grown and its particular biotechnological interest. This thesis presents a comprehensive study of the molecular response to polychlorinated biphenyls (PCBs), a family of persistent organic pollutants of particular relevance worldwide. It has been used for a transcriptomic approach using RNA-seq technology, to identify genes involved in the metabolism of compounds *in plant* and quantify their levels of activation in different controlled situations. The thesis also aims to define the underlying transcriptional control the biochemical response to these pollutants. It is surprising that the response is virtually unknown at the molecular level, despite its great potential applied in the context of phytoremediation technology.

To develop this project we applied our hybrid poplar crops different concentrations of Aroclor 1221, a mixture of PCBs widely used commercially for decades, its use is now banned internationally. And we RNA samples at two different concentrations and times of exposure to the pollutant, generating an array of four elements with their corresponding controls. In order to increase the specificity of our analysis, we consider mainly the most significant differentially expressed genes in four different statistical algorithms. Moreover, functional analyzes conducted with bioinformatics tools based on sequence comparisons and networks gene co-expression. The response of genes of particular interest was validated by qRT-PCR (polymerase reaction chain in real-time quantitative. This is the first comprehensive study of the response of a plant organism in the presence of PCBs.

This study allowed us to identify a considerable amount of structural and regulatory genes, defining new transcription factors whose expression is proportional to the concentration of contaminant in the middle or at the time of exposure. Correlation analyzes allow us to affirm that the metabolic response to PCBs, including possible degradative pathways, at least fifteen involved in transcription factors and forty proteins or enzymes which are particularly induced. Among the families involved include cytochromes P450, the glutathione transferases, dehydrogenases reductases (short -chain dehydrogenase reductase) and MDR proteins (multi - drug resistance). While transcription factors belong to the family found ZF-TF, MYBs, WRKYs among others.

We also identify proteins of unknown function that had not been previously linked to such responses in plants such as CSP (cold- shock domain proteins). To study their possible relationship with the presence of PCBs, a gene in this family was characterized and was detected by tandem mass spectrometry (MS/MS) from maps IEF x SDS -PAGE (*sodium dodecyl isoelectro x sulphate- polyacrylamide gel electrophoresis*) of high resolution. By qRT -PCR could confirm the induction of the corresponding gene, ortholog to PtCSP4 of *P. trichocarpa* (Potri.004g172600), in response to Aroclor 1221. Phenotypic analysis of transgenic *Arabidopsis thaliana* lines over- expressing the protein CSP poplar hybrid confirmed a role for PCBs same tolerance, possibly through regulatory mechanisms activated MDR proteins.

This work, in addition to providing new data on the molecular mechanisms triggered by the presence of PCBs in *Populus*, used here as a model system. Thus the potential of tree species not only as decontamination agents, and commercially exploited, but also as a potential source of interesting genes is shown. Among the genes identified in this thesis there are evident candidates to participate in tolerance mechanisms to stress induced by pollution and degrading metabolic pathways of PCBs. Precisely the possibility of degrading the pollutant attaches particular interest to this type of study off the phytoremediation of heavy metals and other elemental pollutants. The comparison of the data generated in this study with similar studies carried out in the future with other species and xenobiotics contribute to better define the response of plants to organic pollution and improve their decontamination potential.

1. INTRODUCCIÓN

1. Introducción

«La naturaleza utiliza solamente las hebras más largas para tejer sus formas, de manera que cada pequeño rincón de su tela revela la organización de la totalidad del tapiz»

Richard P. Feynman (Premio Nobel de Física 1965)

1.1. El suelo como recurso natural amenazado

El suelo tiene una importancia central para el funcionamiento de los ecosistemas y para el sector agroforestal. La influencia del mismo en la producción de alimentos, piensos, combustibles, fibras o madera, entre otros servicios, le confiere un estatus clave desde el punto de vista de la seguridad económica y de la explotación sostenible de los recursos naturales. Así lo ha reconocido la comunidad internacional (FAO, 2015), que también ha puesto de manifiesto la necesidad de mejorar significativamente la gestión actual del suelo e incrementar sus niveles de protección ante las amenazas derivadas del cambio climático y de la creciente presión demográfica.

La degradación del suelo ha adquirido ya una magnitud preocupante. Por otra parte, es improbable que la superficie productiva vaya a crecer. El problema de los usos del suelo es muy complejo y responde a un conjunto de factores sociales, económicos y ambientales. Un aspecto importante es la confrontación entre usos alternativos, de modo que el urbanismo, el transporte o las actividades recreativas compiten con la producción de energía o la obtención de alimentos y materias primas. Otros aspectos importantes son la falta de planes globales de actuación, la sobreexplotación de los recursos actuales, la ausencia de legislación y, en general, un escaso aprecio de la importancia del suelo. Como señalan los expertos de la FAO (2015) en el informe antes citado, “*la tasa actual de degradación del suelo amenaza la capacidad de las generaciones futuras para satisfacer sus necesidades*” (FAO, 2015). Este informe estima que la mayoría de los suelos del planeta (un 77%, en superficie) presenta un nivel significativo de degradación, siendo ésta elevada en un 43% de los casos y moderada en el 57% restante. Las principales causas de degradación edáfica son la contaminación química, la erosión, la salinización y la compactación. Durante los últimos cincuenta años, la

superficie utilizada para la agricultura se ha ampliado sensiblemente, sacrificando en gran medida masa forestal, sobre todo en las regiones tropicales. Se ha deforestado en el periodo 2010-2015 una media de 5,2 millones de ha cada año, lo que repercute a su vez en el clima global y la biodiversidad. En la Figura 1 se muestran los países donde el suelo está muy degradado (rojo), degradado (naranja), medianamente degradado (amarillo), poco degradado (verde) y sin degradar (gris) (UNEP, 2014).

Figura 1. Degradación del suelo a nivel mundial. Los colores representan el nivel de degradación del suelo (UNEP, 2014).

La tendencia a la expansión de las tierras de cultivo, derivada del estancamiento de la productividad agrícola, así como la creciente degradación del suelo fértil debida a la erosión, la contaminación y la climatología extrema, señalan la necesidad urgente de mejorar los modelos actuales de gestión. Los aumentos de productividad que se esperan en las próximas décadas no serán suficientes para mantener un uso sostenible del suelo. Aun así, se estima que las mejoras en la gestión y restauración de tierras degradadas, junto con cambios en los hábitos de consumo, podrían permitir recuperar entre 160 y 320 millones de hectáreas hasta el año 2050 (UNEP, 2014).

Este esfuerzo será en vano, sin embargo, si no va acompañado de una auténtica mejora en la calidad global del suelo. Uno de los problemas más relevantes en este sentido es su

nivel actual de contaminación, que tiene un impacto creciente sobre la salud humana y el funcionamiento de los ecosistemas. Si nos centramos en el problema concreto de la contaminación química, el programa de las Naciones Unidas para el Medio Ambiente (UNEP, 2014) estima que la mayor parte del suelo mundial está contaminado de forma significativa.

La contaminación del suelo disminuye su fertilidad, biodiversidad y capacidad de regeneración, amén de perturbar los ciclos biogeoquímicos naturales y alterar la calidad del agua y del aire. La naturaleza poco renovable del suelo agrava considerablemente el problema. Por otra parte, en la mayoría de emplazamientos catalogados concurren contaminantes orgánicos e inorgánicos, lo que dificulta considerablemente su remediación (Meuser, 2014). Ateniéndonos tan solo al ámbito europeo, la Agencia Europea para el Medio Ambiente (www.eea.europa.eu) estima que las actividades contaminantes afectan a unos tres millones de emplazamientos. Aproximadamente medio millón supera los límites máximos establecidos legalmente y deberían ser objeto de remediación a la mayor brevedad. Es evidente, en suma, que el problema de la contaminación del suelo es prioritario y requiere actuaciones urgentes a todos los niveles, desde la investigación científica hasta la gobernanza (FAO, 2015). Ante la gravedad creciente del problema de la contaminación, las Naciones Unidas impulsaron en los años noventa un tratado internacional que abordase las cuestiones más urgentes. Tras más de una década de negociaciones, este esfuerzo fructificó en el establecimiento de la Convención de Estocolmo, que entró en vigor en mayo de 2004. Este tratado requiere que los países participantes tomen medidas para eliminar o reducir la liberación de contaminantes orgánicos persistentes (COPs) al medio ambiente.

Los COPs son compuestos particularmente tóxicos y estables, entre los que se encuentran insecticidas, herbicidas, bifenilos policlorados (PCBs), dioxinas y retardantes de llama (puede accederse a la lista completa actualizada en www.pops.int). Este tratado es administrado por la UNEP, con base en Ginebra, y cuenta actualmente con 152 países signatarios (<http://chm.pops.int>, Status of Ratifications). Existen otros tratados afines de menor entidad geográfica, como el de Basilea y el de Rotterdam, con el objetivo común de proteger la salud humana y el medio ambiente de las sustancias químicas tóxicas y los residuos peligrosos. Todos ellos contribuyen a controlar y establecer acciones concretas para remediar la contaminación mundial.

1.2. Contaminación por Bifenilos Policlorados (PCBs)

1.2.1. Estructura, usos y localización

Los PCBs están formados por dos anillos bencénicos conectados por un enlace C1-C1', en los que los átomos de hidrógeno pueden estar sustituidos o no por átomos de cloro (entre 1 y 10) (Figura 2). Esto origina un total de 209 posibles congéneres. Los congéneres mono y no- orto-sustituídos (coplanares) presentan similitud estructural con la 2, 3, 7, 8- tetraclorodibenzo-p-dioxina (2, 3, 7, 8-TCDD o dioxina) y por lo tanto son clasificados como “dioxin-like” (DL-PCBs) (Carpenter, 2006; Srogi 2008). Los congéneres se diferencian en sus propiedades químicas, físicas y toxicológicas (Tabla 1).

Figura 2. Estructura química de los PCBs. Estructura molecular general de un bifenilo policlorado. Se muestra así mismo la estructura de un bifenilo monoclorado y uno decaclorado (ATSDR, 2014).

En general se trata de compuestos bastante hidrofóbicos, si bien la polaridad de cada congénere depende de su patrón concreto de cloración. Estos compuestos se caracterizan por su gran estabilidad química, lo que ha contribuido en gran medida a su dispersión por los ecosistemas del planeta. Debido a dicha estabilidad y baja solubilidad en agua, los PCBs son extremadamente difíciles de eliminar de las matrices del suelo. Como regla general, cuanto más clorada esté la molécula es menos soluble en agua y menos volátil, si bien la posición concreta de los cloros influye en ambas propiedades. El coeficiente de partición octanol-agua ($\log K_{ow}$) permite predecir hasta cierto punto la movilidad de los PCBs en el medio ambiente. Los PCBs con $\log K_{ow}$ superior a 6 se

asocian a partículas en la atmósfera, suelos y sedimentos; los menos clorados se encuentran típicamente en la fase gaseosa. Como consecuencia, los suelos generalmente contienen mayor proporción de congéneres más clorados, mientras que el aire está dominado por fracciones menos cloradas (Van Aken et al., 2009; Passatore et al., 2014).

Tabla 1. Propiedades principales de las mezclas comerciales de PCBs y sus congéneres (Van den Berg et al., 1998; Beyer et al., 2009; Passatore et al., 2014).

Características	
Propiedades físicas	Aceite, líquido viscoso, resina pegajosa
Solubilidad en el Agua (mg/L)	0,0027 (Aroclor 1260) 0,59 (Aroclor 1221)
Coeficiente de partición Octanol-Agua (Log Ko/w)	4,7 (Aroclor 1221) 6,8 (Aroclor 1260) 60 (penta-y-hexa-CBs en Ciliate protozoa)
Factores de bioconcentración en organismos acuáticos (BCFs)	274.000 (tri-y-tetra-CBs en el pez común de agua dulce)
Factor de equivalencia tóxica en animales (TEF)	0,00001 (2,3',4,4'5,5'- HexaCB)-0,1 (3,3',4,4',5-PentaCB)

Los PCBs se sintetizaron por primera vez a mediados del siglo XIX a escala de laboratorio. Su fabricación comercial comenzó en 1930, mediante la cloración progresiva de bifenilo en presencia de un catalizador, normalmente FeCl_3 . Los PCBs han sido ampliamente producidos y utilizados como lubricantes, refrigerantes en transformadores, condensadores, intercambiadores de calor y aislantes; también se incluyen en tubos fluorescentes, artefactos eléctricos con condensadores (motores monofásicos, heladeras, televisores, radios, equipos de audio, etc.), aceite para microscopios, fluidos hidráulicos, aceites de recubrimiento, tintas de impresión, retardantes de fuego, asfalto, plastificantes, adhesivos, componentes de resinas y gomas sintéticas como el caucho, baldosas, ceras, agentes anti-polvo, pesticidas, papel carbónico (Passatore et al., 2014; EPA, 2015).

A finales de 1970 la mayoría de los gobiernos cesaron la producción de PCBs, sin embargo persiste una contaminación ambiental como consecuencia de derrames accidentales, fugas ocasionadas en el transporte, almacenamiento indebido y una disposición inapropiada de los residuos contaminados (Passatore et al., 2014).

Los PCBs normalmente se presentaron como mezclas comerciales (por ejemplo: Delor, Askarel, Aroclor, Fenclor), la más común fue la serie de Aroclor (utilizado en esta tesis). El Aroclor (Producido por Monsanto Chemical Industry, Illinois) es una mezcla de PCBs. Estas mezclas tienen un número de cuatro dígitos, los dos primeros se refieren al número de átomos de carbono y los otros dos indican el porcentaje de clorinación en la mezcla (Beyer et al., 2009; EPA, 2014).

Debido a lo anteriormente expuesto, los PCBs siguen representando un problema global hasta la fecha. Países como Australia, Canadá, Alemania, Suecia, Reino Unido y USA, consideran que concentraciones de 50 ppm de PCBs son el umbral de referencia para que se aplique la legislación. De modo que los materiales con concentraciones superiores a este valor deben manejarse de acuerdo con regulaciones específicas. Estos compuestos se han utilizado principalmente en los países más industrializados. La EPA ha determinado que los países responsables de la mayor parte del consumo mundial de PCBs son Francia, Alemania, Italia, Japón, España, Reino Unido y USA. El resultado es que la contaminación del suelo es más preocupante en estos países (Figura 3).

Figura 3. Nivel acumulativo de PCBs en el suelo a nivel mundial. El mapa tiene una resolución de 1° x 1° (longitud x latitud). La coloración en azul más intenso significa una concentración mayor de PCBs (EPA, 2014).

La problemática es global, ya que los PCBs circulan por el planeta a través de un proceso conocido como “efecto saltamontes”, que consiste en un ciclo repetido de evaporación y depósito, por el cual los PCBs que se emiten en una parte del mundo pueden ser transportados a través de la atmósfera a regiones muy distantes de la fuente original (Boer y Fiedler, 2013). Un ejemplo de este efecto es lo que ocurre en el Ártico, donde la problemática con diversos contaminantes va en aumento. Un ejemplo se muestra en la Figura 4, que muestra las concentraciones de tres tipos de contaminantes orgánicos en dicha región. Los más acumulados con diferencia son los PCBs.

Figura 4. Concentraciones ($\pm$ DE) de Σ HCHs, Σ DDTs, y Σ PCBs en el Ártico recogidas en la Ciudad de Barrow (Alaska), Bahía Norte de Baffin, Isla Noruega de Jan Mayen, y el mar de Barents (Gabrielsen et al., 1997, Borgå et al., 2001, Fisk et al., 2001; Hoekstra, 2002)

1.2.2. Efectos sobre el medioambiente y la salud

Los PCBs poseen la capacidad de mantenerse inalterados en el medio ambiente durante largos periodos de tiempo, de ahí que una vez liberados puedan propagarse ampliamente por el planeta como resultado de procesos naturales en los que intervienen el suelo, el agua y el aire. En el Convenio de Estocolmo se considera que un compuesto es persistente si su vida media es superior a 2 meses en agua, o superior a 6 meses en el suelo, o en sedimentos si persiste más de 6 meses. Debido a su capacidad derivada para bioacumularse, su estabilidad física e hidrofobicidad, los PCBs están presentes en toda la cadena alimentaria del planeta. La biomagnificación es inevitable: su concentración aumenta a medida que se asciende en las pirámides tróficas, asociadas generalmente a tejidos grasos. Esto significa que las presas tienen menor concentración de sustancias tóxicas que los predadores (Figura 5). Respecto a la cadena trófica humana, los PCBs predominan en alimentos grasos de origen animal, como huevos, leche o pescado azul. Los alimentos de origen vegetal presentan concentraciones mucho menores (Fernández-González et al., 2013; Malisch y Kotz, 2014).

Figura 5. Ejemplo de bioacumulación de PCBs en una pirámide trófica acuática. Se han descrito factores de biomagnificación superiores a 62×10^6 en los Grandes Lagos, USA (Secher et al., 2013)

En las aguas superficiales los PCBs están presentes en cantidades sustanciales. De hecho, estos compuestos tienden a adherirse a la superficie de la materia orgánica que está en suspensión en el agua y pueden permanecer en los sedimentos durante mucho tiempo, liberándose lentamente en el agua y evaporándose en el aire. Regresan de nuevo a la atmósfera, sobre todo cuando hace calor, por medio de la evaporación y también cuando las concentraciones de PCBs en los sedimentos secos son muy altas. Los PCBs presentes en el aire se pueden depositar en la tierra cuando llueve o nieva, o cuando se adhieren a partículas, como el polvo u hollín, que se instalan en el suelo. En tierra, la acumulación de PCBs se produce a través del suelo o de las plantas, de las cuales se alimentan muchos invertebrados que finalmente pasan a las aves y los mamíferos. La luz solar puede desempeñar un papel en la degradación de los PCBs presentes en el aire, el agua y el suelo superficial. En la atmósfera, estos contaminantes pueden reaccionar con el ozono y las partículas de agua bajo el efecto de la luz solar. Como resultado de estas reacciones se eliminaron átomos de cloro (Lu et al., 2015). En la Figura 6 podemos observar como parte de los PCBs se mantienen en movimiento, pudiendo ser este aleatorio dependiendo del medio y los efectos degradadores mencionados anteriormente, como por ejemplo la luz solar. También podemos ver que parte de esos PCBs (pero un porcentaje muy bajo) termina por degradarse definitivamente, desapareciendo del medio.

Figura 6. Predicción de concentraciones del PCB-52 en el Río Támesis (Reino Unido) para el año 2020 (Lu et al., 2015).

La exposición y consumo de estos compuestos trae consigo consecuencias importantes para la salud. Se ha demostrado que el contacto prolongado con PCBs causa daños neurológicos, reproductivos, endocrinos y dermatológicos, entre otros. También promueve las infecciones, debido a la reducción de la respuesta inmune. Los estudios epidemiológicos sugieren que la mayoría de estos efectos son el resultado de una exposición crónica (Taylor et al., 2013; Zeliger, 2013). La Agencia Internacional para la Investigación del Cáncer (IARC) ha clasificado a los PCBs como cancerígenos para seres humanos (Lauby-Secretan et al., 2013).

La enfermedad de la piel conocida como cloracné es uno de los efectos típicos de la ingesta de altas concentraciones de PCBs y dioxinas, las cuales están muy relacionadas estructuralmente y se incluyen también en el listado inicial del Convenio de Estocolmo (Passatore et al., 2014). Uno de los casos más mediáticos fue protagonizado por Viktor Yuschenko, presidente de Ucrania, que sufrió un intento de envenenamiento en 2006 (Figura 7).

Figura 7. Cloracné causado por envenenamiento con organoclorados. Viktor Yuschenko, 2006.

Numerosas investigaciones sugieren una correlación entre exposición a PCBs y diversas enfermedades metabólicas, como diabetes tipo 2 (Taylor et al., 2013; Zeeliger et al., 2013), obesidad (Tang-Péronard et al., 2011), síndrome metabólico y enfermedad del hígado graso no-alcohólico (Wahlang et al., 2013).

1.3. Remediación de suelos contaminados con PCBs

La cantidad de suelos contaminados por PCBs en el mundo es desconocida. En USA, 350 de los 1290 sitios contaminados incluidos en el programa “Superfund” de la EPA están contaminados con PCBs (27%) (<http://www3.epa.gov/>). En Canadá hay al menos 148 sitios análogos de acuerdo al Federal Contaminated Sites Inventory (TBS-SCT). En Europa se estima que hay 242.000 sitios contaminados, de los cuales un 2,4% lo está con hidrocarburos clorados (EEA). Es muy importante el uso de tecnologías de remediación eficientes para tratar de paliar el problema de los PCBs.

Las tecnologías de remediación de suelos pueden aplicarse *in situ* o *ex situ* y se vienen usando desde hace décadas para recuperar suelos que sufren la acumulación de PCBs y otros contaminantes persistentes.

La mayoría de las técnicas tradicionales de remediación de PCBs en suelos y sedimentos, tales como la extracción con solventes, la desorción térmica o la degradación alcalina, no son apropiadas para grandes superficies. Estos métodos requieren costosas excavaciones y el transporte del material contaminado en grandes cantidades necesita un gasto de energía excesivo. Se calcula que limpiar una tonelada de tierra puede exceder de los 1000 US\$. Pensemos en un ejemplo para apreciar el costo: 1 ha de suelo, cavando a solo 1m de profundidad, tendrá aproximadamente 13.500 tm de tierra, a 1000 US\$ la tonelada (Cano-Camacho, 2015).

Por otro lado, las alternativas *in situ*, como los tratamientos térmicos usando energía de microondas, están limitada a las características del sitio contaminado y requieren también inversiones elevadas (Gomes et al., 2013). Es por eso que la aplicación de tecnologías como la fitorremediación y la biorremediación se expanden rápidamente. Sin embargo, una de las limitaciones de la biorremediación es la dificultad de verificar la toxicidad de los intermediarios o productos que se encuentran en el suelo (Passatore et al., 2014; Doty, 2008). Sin embargo, la importancia de las tecnologías limpias ha ido creciendo cada día más.

La fitorremediación es una alternativa que ha despertado un interés considerable en los últimos años (Campos et al., 2008 y Ansari et al., 2015). En comparación con los métodos antes expuestos, es una tecnología significativamente más barata, funciona con energía solar, sirve para grandes extensiones y puede llegar a ser bastante más eficaz.

Por otra parte, en regiones poco desarrolladas donde los recursos de ingeniería necesarios para la descontaminación pueden no estar disponibles, la fitorremediación puede llegar a ser la única tecnología disponible capaz de solucionar el problema (Erakhrumen, 2007).

Las tecnologías más comunes para la remediación de suelos contaminados con PCBs se resumen en la Tabla 2 (Gomes et al., 2013).

Tabla 2. Clasificación de las tecnologías de remediación de suelos contaminados con PCBs. En color naranja la Fitorremediación utilizada en esta Tesis (Gomes et al., 2013).

1.4. Fitorremediación: perspectivas moleculares

La fitorremediación está basada en el uso de plantas para extraer, secuestrar y/o destoxificar contaminantes del suelo. Esta tecnología reduce significativamente los costes de la descontaminación tradicional (Gomes et al., 2013), por otra parte la fitorremediación ha ganado popularidad por ser ecológica que utiliza energía solar esencialmente y contribuir a la lucha contra la erosión y ayudar a la restauración de espacios degradados (Dickinson, 2009). Los grandes países y las empresas gustan de colaborar en estos proyectos ganando imagen y con ello ayudan al medio ambiente, dado su gran potencial como técnica sostenible de descontaminación, por lo que ha sido catalogada como una de las ‘tecnologías *top-ten* que mejoran la vida humana’ (Daar, 2002). La EPA norteamericana promueve decididamente su utilización (www.epa.gov).

La fitorremediación utiliza al menos uno de los siguiente procesos: fitoextracción, fitodegradación, rizofiltración, fitoestabilización o fitovolatilización (Ayoub et al., 2010; Liu et al., 2013; Passatore et al., 2014 Furini et al., 2015; Truua et al., 2015). Su efectividad ha sido demostrada con todo tipo de contaminantes que van desde moléculas inorgánicas hasta las orgánicas. La Tabla 3 presenta algunos ejemplos relativos a fitorremediación de PCBs.

Tabla 3. Ejemplos selectos de fitorremediación de PCB y organoclorados en el periodo 2010-2015.

Contaminante	Planta	Referencia
PCBs	<i>Medicago sativa</i> L. <i>Cucurbita</i>	Teng et al., 2010
PCBs	<i>Medicago sativa</i> , <i>Lolium perenne</i> L, <i>Festuca arundinacea</i> y <i>Oryza sativa</i>	Chen et al., 2010
Aroclor 1254/ 1260 Aroclor 1248	<i>Cucurbita pepo</i>	Ficko et al., 2010 y 2011
PCB	<i>Medicago sativa</i> L.	Tu et al., 2011
PCB	<i>Howden pumpkins</i>	Ficko et al., 2011
PCB (Delor 106 o Aroclor 1260) y PAH	<i>Populus nigra</i> x <i>P. maximowiczii</i> y <i>Salix x smithiana</i>	Kacálková y Tlustos, 2011
Mezcla de PCBs, PAH y metales pesados	<i>Salix miyabeana</i>	Guidi et al., 2011
Mezcla de PCBs congeneres	<i>Salix alaxensis</i> y <i>Picea glauca</i>	Slater et al., 2011
PCB, hidrocarburos y metales pesados	<i>Populus nigra</i>	Doni et al., 2012
PCBs	<i>P. deltoides</i> × <i>nigra</i>	Meggo y Schnoor, 2013
HCH, DDT, heptacloro, aldrin	<i>Ricinus communis</i>	Rissato et al., 2015

Una limitación importante es el tamaño y profundidad de las raíces. En este punto, las herbáceas logran alcanzar hasta unos 50 cm, mientras que las arbóreas pueden incrementar considerablemente este valor, dependiendo de la especie y circunstancias concretas del emplazamiento. Además, el tiempo de actuación es elevado en comparación con los métodos convencionales donde normalmente se requieren varios años para descontaminar un terreno (Pilon-Smits, 2005).

La efectividad de la fitorremediación puede incrementarse identificando genes que participen en el metabolismo de PCBs. El concepto de que las plantas pueden degradar xenobióticos surgió en 1940, en relación con la descontaminación plaguicidas (Sanderman, 1994). Desde entonces, el desarrollo de la biología molecular ha contribuido a dilucidar el metabolismo de algunos contaminantes en plantas, sobre todo metales pesados (Eapen et al., 2007; Aken, 2008; Wang et al., 2012; Bansal et al., 2014).

Se han generado algunas variedades transgénicas para la fitorremediación de xenobióticos, pero su comercialización plantea problemas. En la Unión Europea el cultivo de plantas transgénicas todavía se asocia a riesgos para los ecosistemas (Maestri y Marmiroli, 2011). La biomasa producida durante la fitorremediación de sitios contaminados podría ser valorizada económicamente en forma de bioenergía (biogás, biocombustibles y biomasa para la producción de energía y calefacción), lo que representa un importante co-beneficio ambiental. Gomes (2012) identificó los retos y oportunidades asociados con el uso de la fitorremediación para la producción de bioenergía.

Aunque gran parte de la mejora genética se ha experimentado en plantas modelo de laboratorio, como *N. tabaccum* y *A. thaliana*, su poca biomasa y el corto tiempo de vida podría no ser adecuado para la remediación *in situ*. Por lo tanto, hay un interés particular en la mejora genética de árboles que mantengan un crecimiento rápido, sistemas de radicales extensos y abundante producción de biomasa, como chopos y sauces.

1.5. *Populus*: arbórea modelo con potencial remediador y en investigación molecular

El uso de especies arbóreas permite soslayar una de las limitaciones de la fitorremediación: el hecho de que el volumen tratado de suelo se restringe al que está en contacto directo con las raíces. Particularmente los chopos (género *Populus*; familia Salicáceas) son candidatos excepcionales para su empleo en la fitorremediación y, de hecho, dominan los programas actuales de fitorremediación (Doty, 2015). El género *Populus* comprende unas 40 especies de árboles originarias de las zonas templadas y frías septentrionales. Tiene altas tasas de crecimiento (hasta 3 m/año) y de transpiración (hasta 100-200 L/día, en árboles de ≥ 5 años de edad) así como un gran sistema radicular que asegura la absorción eficiente de agua subterránea (Isebrands y Richardson, 2014). En esta tesis se ha elegido el híbrido *P. tremula* $\times$ *P. alba* por la facilidad con que se cultiva y su particular interés biotecnológico. Este híbrido, conocido como álamo cano, es un árbol con buen crecimiento y tolerancia a estreses abióticos como salinidad o sequía (Isebrands y Richardson, 2014).

La idoneidad del chopo no solo radica en sus características fisiológicas. *Populus* es hoy el mejor sistema modelo forestal disponible, a raíz de la secuenciación de su genoma completo. El tamaño del genoma de *P. trichocarpa* es moderado: 423 millones de pb (pares de bases) y su transcriptoma cuenta con 73,013 transcritos. Aunque es más grande que el de otras especies modelo, como *A. thaliana*, es bastante reducido en comparación con otras arbóreas. El nivel medio de redundancia de secuenciación es de 7.5x. El genoma corresponde al clon femenino “Nisqually-1” y está organizado en 19 cromosomas. Hasta hoy se encuentra disponible la versión 3.1. (<http://phytozome.jgi.doe.gov>). Su tamaño es solo cuatro veces mayor que el de la especie vegetal modelo *Arabidopsis thaliana*, pero cincuenta veces menor que el genoma estimado de *Pinus* (Campos *et al.*, 2007). Conocer el genoma completo del chopo tiene una importancia inmensa desde el punto de vista científico. La proximidad filogenética del chopo (*Dicotyledonea Rosida*) a varias especies herbáceas cuyos genomas están también secuenciados (*Arabidopsis*, *Lotus*, *Medicago*) permite, mediante la comparativa de secuencias genómicas, establecer posibles similitudes funcionales y evolutivas entre sus genes. De este modo, el chopo puede ser empleado como especie

punto para estudios de la eco-fisiología molecular así como para asociaciones gen-función en otras arbóreas (Brunner *et al.*, 2004).

Populus fue la primera arbórea que fue transformada genéticamente de forma estable (Fillatti *et al.*, 1987). Varios estudios han comprobado el potencial molecular y fitorremediador de *Populus*. Por ejemplo, Gullner *et al.*, (2001) utilizó a *P.trichocarpa* para sobre-expresar γ -ECS (γ -Glutamylcisteína sintetasa) de chopo para incrementar la tolerancia al herbicida cloroacetanilida. También Doty *et al.*, (2007) sobre-expresaron la proteína CYP450 de conejo en *P. tremula x P. alba* incrementando la remoción TCE, benceno y cloroformo entre otros. Van Dillewijn *et al.*, (2008) utilizaron *P. tremula x P. tremuloides* para sobre-expresar *pnrA* (nitroreductasa) de *Pseudomonas putida*, con el fin de degradar el explosivo TNT del suelo. *Populus* posee un metabolismo secundario muy desarrollado que concentra múltiples mecanismos de defensa frente al estrés, a través de la síntesis de compuestos defensivos (Isebrands y Richardson, 2014).

Figura 8. Importancia económica y potencial biotecnológico del chopo (genero *Populus*).

1.6. Genes y factores reguladores relacionados con el metabolismo de xenobióticos

El conocimiento actual de la degradación de contaminantes orgánicos por parte de las plantas es mucho menor que en animales o bacterias. Sin embargo algunas plantas parecen transformar y/o mineralizar eficientemente una amplia variedad de compuestos orgánicos tóxicos. Los procesos que subyacen a la fitorremediación de estos contaminantes orgánicos prácticamente no se conocen (Passatore et al., 2014).

El proceso de degradación de contaminantes se puede resumir en tres fases: (1) transformación del compuesto, (2) conjugación y (3) compartimentación. En la fase de transformación, los contaminantes se modifican químicamente (oxidación, reducción e hidrólisis) y se transforman en moléculas más polares. Esto se debe a la acción de enzimas como citocromos P450, liasas, transferasas, dioxigenasas y esterasas, entre otras. La conjugación a moléculas endógenas como azúcares y péptidos también aumenta la solubilidad de los contaminantes. Mediante glucosil transferasas y glutatión S transferasas. Dicha comparación facilita además el ataque enzimático de los contaminantes y su compartimentación celular (almacén en vacuolas, integración en la membrana celular o secreción al apoplasto (Pilon-Smits, 2005; Abhilash et al., 2009; Truua et al., 2015). En esta última etapa (compartimentación) existe una subclase de transportadores ABC, denominados “multidrug resistance” (MDR), que están implicados en la regulación de la afluencia y la salida celular de xenobióticos (Rea, 2007; Xi et al., 2011). Sorprendentemente, se sabe muy poco sobre el papel de las MDRs de plantas en la resistencia a la xenobióticos, ya que se han hecho pocos estudios (Conte et al., 2011). Los mecanismos que rigen estos procesos bioquímicos están finamente regulados por una gran cantidad de genes que codifican proteínas reguladoras. Los factores de transcripción ocupan un papel central, alterando la expresión de numerosos genes. Muchos de estos factores son co-inducidos en respuesta a diferentes estreses, lo que sugiere la existencia de interacciones complejas (Yuan y Perry, 2011; Golldack et al 2014).

La transcripción de un gen se inicia cuando el complejo de iniciación se une a secuencias específicas del promotor, como la TATA box. Este complejo es regulado por factores de transcripción que interaccionan con otras secuencias específicas del promotor, denominados elementos de respuesta o reguladores. Estos a su vez son

activados o reprimidos por estímulos ambientales u hormonales (Figura 9) (Phillips, 2008). Numerosos estudios han demostrado que las familias de factores de transcripción MYB, bZIP y WRKY están implicados en la regulación de respuestas a estrés (Ramel et al., 2012). La respuesta a contaminantes tampoco parece depender de una sola familia de factores transcripcionales (Ekman et al., 2005; Ramel et al., 2007, 2012; Golldack, 2014) (Tabla 4).

Tabla 4. Identificación de factores de transcripción asociados a la respuesta a xenobióticos. Los datos de expresión génica fueron extraídos del perfil transcriptómico de *Arabidopsis* en presencia de atrazina 2-cloro-4-(etilamina)-6-(isopropilamina)-s-triazina (Ramel et al., 2007) registrado como E-MEXP-411 en ArrayExpress bajo las siguientes condiciones: control (M= Manitol), estrés por atrazina (MA= Manitol + Atrazina) y en presencia de sacarosa exógena (S=sacarosa) y en presencia de atrazina bajo condiciones de tolerancia a estrés (SA= sacarosa + Atrazina). Las anotaciones fueron actualizadas de (Arabidopsis Information Resource) (Lamesch et al., 2012).

Número de accesión	Descripción del gen	MA/M	S/M	SA/M
		Tratamiento	log ₂ (ratio)	
At2g30250	Factor de transcripción WRKY25	0.75	1.14	2.10
At2g23320	Factor de transcripción WRKY15	nde	1.05	2.33
At5g13080	Factor de transcripción WRKY75	0.79	2.00	2.89
At1g42990	Factor de transcripción TbZIP (AtbZIP60)	nde	0.78	1.96
At2g47890	Proteína Zinc finger (B-box type)	1.19	1.37	3.05
At5g59820	Proteína Zinc finger (C2H2 type) (Zat12)	nde	1.15	2.08
At1g08050	Proteína Zinc finger (C3HC4-type RING finger)	nde	0.87	1.77
At3g28210	Proteína Zinc finger (AN1-like) (PMZ)	nde	2.75	3.19

Muchos factores de transcripción presentan una estructura modular formada por ciertos tipos de dominios o regiones, entre los que se encuentran los de unión al DNA y los que regulan la transcripción (región reguladora). La región de unión al DNA permite que el factor se fije específicamente a unos motivos o secuencias breves del DNA gracias a distintos mecanismos moleculares (homeodominios, dedos de zinc y otros). La región reguladora permite que la proteína aumente la transcripción del gen (dominio de activación) o la disminuya (dominio de represión) (Phillips, 2008).

Figura 9. Regulación de la transcripción. El complejo de iniciación es regulado por factores de transcripción (representados por A, B y C) que son activados o inhibidos por señales ambientales y hormonales.

1.6.1. Factores de transcripción MYB

La familia de proteínas MYB (mieloblastosis) se encuentra tanto en plantas como en animales y tiene funciones muy diversas. Esta familia tiene un número variable de dominios MYB. La región C-terminal puede contener elementos para la activación o represión transcripcional (Ambawat et al., 2013). El criterio de clasificación de estas proteínas está basado en el número de dominios MYB presentes en sus secuencias, variando de uno a cuatro. Cada repetición consiste en aproximadamente 52 residuos aminocídicos y tres elementos α hélix-turn-helix (HTH).

En plantas, algunos factores de transcripción MYB participan en la activación de genes de respuesta a estrés biótico (Hirayama y Shinozaki, 2010), tales como infección por patógenos (Vailleau et al., 2002); en la regulación del metabolismo primario y secundario, como el control de la acumulación de flavonoles (Stracke et al., 2007), biosíntesis de antocianinas (González et al., 2008), el desarrollo del polen (Preston et al., 2004), la biosíntesis de lignina (Newman et al., 2004; Zhou et al., 2009; Zhong et al., 2007, 2008) y la apertura estomática (Liang et al., 2005), entre otras. Sin embargo también se ha encontrado su participación en mecanismos de respuesta a estrés abiótico (Alves et al., 2015; Li et al., 2015). Ejemplos: estrés por sequía (Davey et al., 2009; Gollack et al., 2011; Pereira et al., 2011; Ranjan et al., 2012; Yan et al., 2012;

Bhardwaj et al., 2013; Cao et al., 2013); estrés por temperatura, la cual incluye altas temperaturas, bajas temperaturas y congelación (Rensink et al., 2005; Lim et al., 2006; Bedon et al., 2010; Yun et al., 2010; Jiang et al., 2013; Tian et al., 2013); estrés por metales, como el hierro (Colangelo y Guerinot, 2004; Sperotto et al., 2008; Palmer et al., 2013) y el aluminio (Zeng et al., 2012; Yang et al., 2012b); y estrés por luz (Stracke et al., 2010; Wang et al., 2012; Rius et al., 2012), entre otros.

1.6.2. Factores de transcripción WRKY

Los genes WRKY están restringidos principalmente a las plantas, con la excepción de algunos ejemplos en protozoos. La mayoría de las proteínas de esta familia contiene al menos un dominio WRKY de ~ 60 aminoácidos. El extremo N-terminal de este dominio se caracteriza por contener un motivo altamente conservado: WRKYGQK, mientras que la región C-terminal contiene un motivo “dedo de zinc” C2H2 o C2HXC. Las secuencias WRKYGQK están implicadas en la unión a los elementos W-box (TTGACY) del DNA, un elemento que se encuentran en los promotores de muchos genes relacionados con el estrés (Yamasaki et al., 2013).

Están implicadas en diversos eventos fisiológicos en plantas, tales como el desarrollo de las semillas, la senescencia, la dormancia y la germinación, si bien la respuesta a patógenos parece ser su función más relevante (Rushton et al., 2010). Algunos estudios recientes han revelado la enorme significancia de estos factores de transcripción en la respuesta a estrés abiótico (Banerjee y Roychoudhury, 2015), principalmente a sequía y salinidad (Golldack et al., 2011; Niu et al., 2012; Babitha et al., 2013; Wang et al., 2013;; Ding et al., 2014; Agarwal et al., 2014; Yan et al., 2014; Das y Roychoudhury, 2014; Jiang et al., 2014); estrés oxidativo (Zhang et al., 2012; Vanderauwera et al 2012; Van Aken et al., 2013; Wang et al., 2013; Han et al., 2014); estrés por temperaturas (Li et al., 2010; Zou et al., 2010; Shekhawat et al., 2011; Yokotani et al., 2013; Wang et al., 2014); estrés por déficit de nutrientes (Devala et al., 2007; Chen et al., 2009; Kasajima et al., 2010; Bakshi y Oelmuller, 2014) y estrés por radiación (Wang et al., 2007).

1.6.3. Factores de transcripción NAC

En los últimos años se ha reportado el papel de factores de transcripción NAC como reguladores clave en respuesta a estreses abióticos. Los factores de transcripción NAC comprenden una de las mayores familias conocidas de genes, y hasta la fecha sólo se encuentran en las plantas.

El dominio NAC se encuentra en la región N-terminal, mientras que la región reguladora de la transcripción está en la región C-terminal. Dicha región es generalmente diversa en cuanto la composición de aminoácidos y su longitud. Los genes NAC se han implicado en el desarrollo de órganos, la división celular, la síntesis de la pared secundaria, la senescencia, la homeostasis del hierro y la defensa contra patógenos, y también actúan como reguladores maestros en respuestas al estrés abiótico (Guo y Gan, 2006; Kim et al., 2006; Uauy et al., 2006; Zhong et al., 2006; Puranik et al., 2012).

1.7. Las tecnologías de secuenciación de nueva generación (NGS)

Desde su aparición, las tecnologías NGS (Next Generation Sequencing) han dado lugar a una explosión en resultados que ayudan a comprender mejor el funcionamiento de las plantas.

Esto ha propiciado a su vez un gran desarrollo de otras tecnologías moleculares de alto rendimiento, como la proteómica, la transcriptómica y la metabolómica, que están contribuyendo a dilucidar, entre otros aspectos, la compleja respuesta al estrés producido por los compuestos contaminantes (Thapa et al., 2012, Claire-Lise y Nathalie, 2012, Bell et al., 2014, Ilyas et al., 2015). Además, han permitido abrir nuevos horizontes para el entendimiento detallado y global de procesos de expresión génica (Mochida y Shinozaki, 2011; Schneeberger y Weigel, 2011; Ward et al., 2012).

1.7.1. Proteómica

La proteómica, término acuñado por Marc Wilkins (1994) permite estudios detallados de las proteínas celulares en respuesta a estímulos definidos. El inicio de la proteómica (finales de los 70) fue posible gracias al desarrollo de la electroforesis bidimensional IEF x SDS-PAGE (del inglés *Sodium Dodecyl Sulfate Polyacrilamide Gel Electrophoresis*). Esta técnica permite la visualización de proteínas individuales a partir de extractos, mediante su separación en dos dimensiones perpendiculares entre sí. La primera dimensión consiste en una separación basada en el punto isoelectrico (pH al cual la carga neta de la proteína es cero), y una segunda separación basada en la masa molecular (Thelen, 2007).

En los años 90 se aplicó a la proteómica una técnica analítica sensible y precisa conocida como espectrometría de masas (MS, *Mass Spectrometry*), con gran potencial para identificar proteínas a partir de los fraccionamientos bidimensionales antes mencionados. Este hecho, unido al gran desarrollo experimentado por las bases de datos genómicos, ha dado un impulso sin precedentes al estudio del proteoma (Santucci et al., 2015). Las técnicas de análisis de extractos complejos de proteínas suelen combinar fraccionamientos bidimensionales IEF x SDS-PAGE y análisis mediante MS. Para ello han sido necesarios varios avances tecnológicos, entre los que destaca el desarrollo de las técnicas de ionización de proteínas para MS.

La espectrometría de masas MALDI-TOF es uno de los métodos más empleados para la identificación de proteínas a partir de fraccionamientos IEF x SDS-PAGE. Con este fin, las proteínas individualizadas en los mapas (*spots*) se recortan manualmente y se digieren con una proteasa, típicamente tripsina. Dependiendo de la estructura primaria de la proteína analizada se generan diversos péptidos, cuyas masas constituyen una *huella* denominada “huella peptídica” que permite identificar la proteína original. La idea subyacente es que cada proteína celular genera una huella diferente, lo que permite una identificación teóricamente inequívoca (Velez et al., 2015).

Para la determinación precisa de masas moleculares, los péptidos se adsorben a una matriz especial que permite su ionización con pulsos láser. Una vez que ésta ha ocurrido, las especies resultantes se separan físicamente de la matriz bajo la influencia de un campo electromagnético (MALDI proviene de *Matrix-Assisted Laser*

Desorption/Ionization). Los péptidos ionizados son acelerados a lo largo de un tubo, lo que permite determinar sus masas moleculares en función del “tiempo de vuelo” (TOF viene de *Time Of Flight*). Este tiempo está directamente relacionado con la relación m/z (masa/carga). Determinados péptidos (iones), con una relación m/z concreta, pueden ser sometidos a la acción de un segundo pulso láser, que provoca ahora la rotura de sus enlaces peptídicos. El tiempo de vuelo de los fragmentos resultantes (*espectro de fragmentación*) proporciona información valiosa sobre la secuencia de aminoácidos del péptido en cuestión. Así, la espectrometría MALDI-TOF permite la obtención de huellas peptídicas (masas de los distintos péptidos generados por la digestión enzimática) y también la secuenciación de pequeños péptidos cuando se aplica una segunda vez (MALDI-TOF/TOF). Esta información permite generalmente la identificación de la proteína comparando con bases de datos apropiadas (Aebersold y Mann, 2003).

1.7.2. Transcriptómica

1.7.2.1. RNA-seq

El RNA-seq está fundamentado en la secuenciación de cDNA mediante tecnología NGS (Ward et al., 2012). Primeramente, se captura el RNA total o mRNA, el cual se convierte en una librería de cDNA. Uno de los pasos fundamentales es la obtención de un RNA de buena calidad que incluya idealmente todos los transcritos que se producen en la condición y tejido en estudio. Para el aislamiento del mRNA se emplea con frecuencia kits de extracción que basan la captura en la presencia de colas poly(A) (Ward et al., 2012). Tras la secuenciación de cDNA, las zonas comunes permiten reconstruir secuencias mayores mediante el procedimiento llamado ensamblaje (Miller et al., 2010). Este consiste en la búsqueda de regiones solapantes mediante criterios heurísticos apropiados, para ir formando una cadena de nucleótidos cada vez mayor hasta conseguir formar la cadena completa. La tecnología de Illumina, adquirida por Solexa, se basa en la secuenciación por síntesis (Sequencing by Synthesis, SBS), y está implementada en la plataforma Hi Seq 2000 utilizada en esta tesis (Minoche et al., 2011).

1.7.1.1. Procesamiento de lecturas

El estudio y análisis de una vasta cantidad de datos genéticos es posible gracias a la bioinformática. Esta disciplina utiliza la tecnología informática para organizar, analizar y presentar datos genéticos y moleculares, con la finalidad de responder a preguntas complejas. Surgió por la necesidad de dar sentido a la gran cantidad de datos biológicos generados en los últimos años, yendo en paralelo al incremento de información generada por las nuevas técnicas de análisis molecular (genes, proteínas, metabolitos, etc) (Guerrero-Fernández et al., 2015). No sería concebible un análisis de RNA-seq como el realizado en esta tesis sin el procesamiento de las lecturas obtenidas con ayuda de herramientas bioinformáticas.

Dicho procesamiento se puede dividir en las siguientes etapas: pre-procesamiento, ensamblaje y anotación o mapeo, además de los análisis estadísticos y funcionales. A continuación se describen con más detalle las etapas del procesamiento de lecturas:

A) Pre-procesamiento

Todas las lecturas de secuenciación necesitan este paso, que incrementa la calidad de los resultados y facilita su posterior análisis. Por ejemplo en la etapa de preparación de muestras se añadieron adaptadores artificiales, que deben eliminarse después de las lecturas. En esta etapa también es necesario eliminar cualquier información ajena al fenómeno que se pretende estudiar, como secuencias contaminantes de otros organismos presentes en las preparaciones (Falgueras et al., 2010).

B) Ensamblaje

Una vez eliminada la información artefactual, la segunda fase se basa en comparar todas las secuencias obtenidas en el experimento contra el genoma o transcriptoma de referencia. La finalidad es identificar secuencias solapadas que permitan eliminar redundancias y mejorar la calidad del mapeo (Figura 8). Cada tecnología de secuenciación tiene sus propios programas comerciales de mapeo, como “Eland” de Illumina, “Corona” de SOLiDTM o “Reference Mapper” de 454/Roche. Sin embargo existe una gran cantidad de programas de código abierto que son alternativas perfectamente viables. El tipo de algoritmo utilizado para los ensamblajes de esta tesis

fue “Bowtie2 (Langmead y Salzber, 2012), que utiliza la transformada de Burrows–Wheeler (Salson et al., 2008) para encontrar coincidencias exactas entre las lecturas y el genoma de referencia. Los resultados se complementan posteriormente con un algoritmo de búsqueda inversa (backtracking) para aumentar las coincidencias, permitiendo una pequeña cantidad de errores en las lecturas. Los resultados de mapeo se proporcionaban tradicionalmente en un fichero de texto tabulado, en un formato llamado Sequence Alignment Map (SAM) (The SAM and Format Specifications, 2011). A causa del incremento en el volumen de datos, producido por las nuevas técnicas de secuenciación, se hizo necesario el desarrollo de su equivalente binario, denominado Binary Alignment Map (BAM), que ocupa menos espacio. Para manipular ficheros en ambos formatos existen las herramientas SAMtools (Li et al., 2009) y BAMtools (Barnett et al., 2011), que están disponibles gratuitamente. Las anotaciones más comunes son las que siguen una ontología (Sauro y Bergmann, 2008), como Gene Ontology (GO) (Harris et al., 2006) o InterPro (Mulder y Apweiler, 2008). Para determinar rutas metabólicas suele utilizarse la Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2014).

Figura 10. Esquema general del alineamiento y cuantificación de lecturas sobre un genoma de referencia

2. OBJETIVOS

2. Objetivos

«Quien nunca ha cometido un error, nunca ha probado algo nuevo»

Albert Einstein (Físico y científico Alemán)

Teniendo en cuenta los antecedentes expuestos en el apartado anterior, el propósito principal de este trabajo es estudiar el potencial del chopo como sistema modelo para identificar nuevos genes (nuevas rutas metabólicas) implicados en la degradación de contaminantes orgánicos relevantes. Se han elegido en concreto los bifenilos policlorados (PCBs), compuestos de particular interés a escala mundial por su toxicidad, persistencia y capacidad de bioacumularse. Los programas actuales de bioingeniería ambiental tienen un déficit importante de información en este sentido, particularmente los programas de fitorremediación. Bajo esta premisa, se han abordado los siguientes objetivos concretos:

1. Estudiar los cambios producidos en el transcriptoma de chopo en respuesta a tratamientos controlados con PCBs. Identificar genes y factores de transcripción cuya expresión se modifica significativamente a lo largo de los tratamientos.
2. Realizar análisis funcionales con herramientas bioinformáticas basadas en comparaciones de secuencias, y redes de co-expresión génica. Validar la respuesta de los genes de particular interés mediante tecnología qRT-PCR.
3. Evaluar el posible potencial descontaminante de alguna de las proteínas o factores de transcripción identificados anteriormente. Para ello, se analizará su estructura primaria, sus relaciones filogenéticas o cualquier otro aspecto que se considere interesante.
4. Evaluar el potencial biotecnológico de algún gen seleccionado en el objetivo anterior mediante la obtención de líneas sobre-expresoras y análisis fenotípicos de tolerancia a PCBs.

3. MATERIAL Y MÉTODOS

3. Material y Métodos

« *La ciencia se construye a partir de aproximaciones que gradualmente se acercan a la verdad*».

Isaac Asimov (Escritor, historiador y bioquímico ruso)

3.1. Material vegetal

- A. *Populus tremula* L. × *P. alba* L.** En los experimentos con chopos se utilizó el clon INRA 7171-B4 de *Populus tremula* × *P. alba* L. Este clon fue elegido debido a su excelente comportamiento *in vitro* y *ex vitro*, así como su facilidad para ser transformado genéticamente y además es un clon con mucho vigor en campo. Cabe destacar que prácticamente la totalidad de los trabajos publicados sobre biotecnología de chopo utilizan este clon.
- B. *Arabidopsis thaliana* L.** El ecotipo Columbia de *Arabidopsis thaliana* (L.) Heynh fue utilizado para la transformación genética. Las semillas se obtuvieron en el Arabidopsis Biological Resource Center de Ohio, USA.

3.2. Cepas bacterianas

- A. *Escherichia coli*.** La cepa DH5α de *E. coli* (hsdR17 [r_K^- , m_K^+] supE44 thi-1 recA1 gyrA [NaI^r + relA1 [lacZYA-argF] U_{169} [m80lacZM15]) (Hanahan, 1983) se empleó para clonar y manipular fragmentos de DNA. Los cultivos se realizaron en medio LB (Luria-Bertani; 10 g/L triptona, 5 g/L de extracto de levadura, 10 g/L de NaCl), suplementado o no con antibiótico. La incubación tuvo lugar a 37°C y 250 rpm. Para cultivos en medio sólido añadimos 8 g/L de bacto-agar al medio LB e incubamos a 37°C. Para el almacenamiento de cepas bacterianas a -80°C se utilizó DMSO (dimetil sulfóxido) a una concentración final del 9% (v/v) como agente estabilizante.

B. *Agrobacterium tumefaciens*. Para la transformación de *A. thaliana* se utilizó la cepa C58C1 (pMP90) de *A. tumefaciens*, que contiene el plásmido pMP90 derivado de pTiC58. Esta cepa fue cedida por la Dra. Pamela Green (Michigan State University). Los cultivos se realizaron en medio LB suplementado con los antibióticos apropiados (rifampicina a 50 mg/L; gentamicina a 25 mg/L y espectinomicina a 100 mg/L). La incubación tuvo lugar durante 72 h a 28°C y 200 rpm. Para cultivos en medio sólido se añadió bacto-agar (8 g/L).

3.3. Plásmidos

Para el clonaje con el sistema Gateway se utilizaron los siguientes plásmidos:

pDONRTM 221 Gateway® (Invitrogen) es un plásmido de 4262 pb, usado para diseñar clones de entrada con extremos flanqueantes attL, lo cual permite recombinar el gen de interés en una gran variedad de vectores de expresión adaptados a la tecnología Gateway®. Contiene sitios de M13 forward y reverse para secuenciación del inserto y un gen de resistencia a kanamicina en *E. coli*.

pB2GW7 Gateway® (Invitrogen) es un vector binario, de expresión en plantas controlada por el promotor 35S CAMV. El vector pB2GW7 confiere resistencia al herbicida glufosinato de amonio (BASTA) en planta y a espectinomicina en bacteria.

3.4. Oligonucleótidos

Los oligonucleótidos se diseñaron con ayuda del software Oligo Analyzer 1.0.3 y manualmente. Fueron sintetizados por Sigma Aldrich y se mantuvieron en frío a una concentración de 100 µM en buffer TE (Tris-EDTA; 10 mM Tris-HCl, pH 8,0; 1 mM EDTA).

3.5. Métodos de plantas

3.5.1. Cultivo de *Populus tremula* x *P. alba*

Cultivo *in vitro*. El cultivo *in vitro* de chopo se realizó en tarros cilíndricos de 150 mL de capacidad. El medio contenía sales MS (Murashige y Skoog) 1x (Duchefa®), 20 g/L sacarosa, 7 g/L de fitoagar y 104 mg/L de vitaminas Mc Cown's. El pH se ajustó a 5,8 con KOH antes de esterilizar en autoclave. Después se añadió como regulador de crecimiento la auxina IBA (ácido indol-3-butírico) a una concentración final de 0,5 mg/L. Para evitar contaminaciones, la solución de IBA se pasó por un filtro Millipore® estéril con un tamaño de poro de 0,22 μm . Los explantos se sub-cultivaron cada 6 a 8 semanas y se mantuvieron en cámaras de cultivo SGC 970/C/RO-HFL (Sanyo®) en condiciones controladas de temperatura (22 a 24°C) y fotoperiodo (16 h de luz y 8 h de oscuridad). Para la iluminación se emplearon tubos de luz fluorescente blanca Sylvania Cool White F72T12/CW/VHO, con una densidad de flujo fotónico de $120 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

3.5.2. Cultivo de *Arabidopsis thaliana*

Cultivo *in vitro*. Para el cultivo *in vitro* las semillas se esterilizaron 5 min con una solución de etanol al 70% (v/v) y Tritón-X100 al 2% (v/v) en agua. Después se lavaron tres veces con 1 mL de etanol absoluto. Tras secarlas con papel de filtro estéril, se sembraron alrededor de 20 semillas por placa Petri de 90 mm de diámetro, colocándolas manualmente. Las placas contenían medio MS (sales MS 1x, 20 g/L sacarosa y se ajustó a pH 5,7) solidificado con fitoagar (Duchefa®) al 0,8% (p/v). Tras permanecer dos días en oscuridad a 4°C, las placas se mantuvieron en cámaras de cultivo a una temperatura de 22°C, fotoperiodo corto (10 h de luz y 14 h de oscuridad) y una densidad de flujo fotónico entre 100 y 125 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

Cultivo en tiesto. Para el cultivo en tiesto, las semillas se sembraron en alveolos de plástico de 5 x 5 cm, a razón de 40 alveolos por bandeja. Se utilizó como sustrato una mezcla de turba (Kekkila Ibérica®) y vermiculita (Projar®) previamente humedecida y esterilizada. Tras un periodo de 2-3 días en oscuridad y a 4°C, los alvéolos con 8-10 semillas se colocaron en cámaras Conviron CMP 4030. La

temperatura se mantuvo a 22°C y la humedad relativa al 70%. Utilizamos una densidad de flujo fotónico en torno a 100 a 125 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ y un fotoperiodo corto (10 h de luz y 14 h de oscuridad).

3.5.3. Tratamientos con Aroclor (PCBs)

Para los tratamientos con estos xenobióticos se evaluó previamente el crecimiento de los chopos *in vitro* en presencia de distintas concentraciones de Aroclor 1221 (mezcla de PCBs comercializada por Sigma Aldrich): 0 (control), 50, 100 y 200 mg/L. Estos experimentos se realizaron en condiciones *in vitro*, utilizando en cada tarro 20 mL de MS-fitoagar suplementado con DMSO 1000x (controles) o Aroclor 1221 diluido en el mismo volumen de DMSO. Tras determinar el efecto de los PCBs sobre el crecimiento vegetativo se monitorizaron la biomasa foliar y la longitud radicular decidimos realizar los tratamientos definitivos a 50 y 200 mg/L de Aroclor 1221. A partir de dichos tratamientos se obtuvieron las muestras de proteína y RNA analizadas en esta tesis.

3.6. Métodos de ácidos nucleicos

3.6.1. Extracción de DNA genómico de *Populus tremula* x *P. alba*

Para la extracción de DNA genómico de planta (chopo) se partió de 100 mg de tejido de chopo *in vitro* pulverizado en nitrógeno líquido. La extracción se realizó mediante el Kit *DNesay Plant Mini* (Qiagen). Se cuantificó con un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies) a un ratio de $280/260 \geq 1,8$. Para visualizar el DNA se utilizaron geles de agarosa al 0,8 % (p/v), preparados con buffer TAE (Tris HCl 40 mM, pH 8,0, acetato sódico 20 mM, EDTA 1 mM) suplementado con bromuro de etidio a 0,1 mg/L. Las muestras se mezclaron con buffer de carga (azul de bromofenol al 0,05% (p/v) y glicerol 5% (p/v)). Una vez terminada la electroforesis el DNA se visualizó en un Molecular Imager Gel Doc XR System. Para la cuantificación virtual se utilizó el software Quantity One 4.6.5 de Bio Rad.

3.6.2. Extracción y secuenciación de DNA plasmídico de *E. Coli*

El aislamiento de plásmidos se llevó a cabo con el Kit *QIAprep Spin Miniprep* (Qiagen). Este método está diseñado para la purificación de más 20 µg de DNA plasmídico partiendo de cultivo de *E. Coli*. El kit se basa en una lisis alcalina y posterior absorción de DNA plasmídico a una resina. Para comprobar que la calidad se cuantificó el DNA plasmídico con un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies) a un ratio de $280/260 \geq 1,8$. Posteriormente, se seleccionaron enzimas de restricción que realizaron cortes generando fragmentos que fueron visualizados en un gel de agarosa al 0,8%. Finalmente, comprobados los fragmentos en gel, se enviaron a secuenciar el DNA en el CNIO (Centro Nacional de Investigaciones Oncológicas) donde se cuenta con un sistema de secuenciación ABI 3730 (Applied Biosystems).

3.6.3. Extracción de RNA total de *Populus tremula x P. alba in vitro*

La extracción de RNA se realizó a partir de 200 mg de pulverizado de planta *in vitro* completa. Se utilizó el método de Fenol-Cloroformo y posteriormente se purificó utilizando el kit *RNeasy Plus Mini* (Qiagen). Para este fin se añadió una relación 3:1 de Buffer de extracción (0,2 M Tris-HCl pH 9,0; 0,4 M LiCl, 25 mM EDTA y 1% SDS) y se mezcló durante 60 seg. Una vez mezclado, se agregó un volumen igual de fenol y se mezcló vigorosamente durante 10 seg hasta homogeneizar y se centrifugó a 13.000 rpm durante 2 min. Se recuperó el sobrenadante. Para hacer un lavado se agregó un volumen de fenol y se centrifugar a 13.000 rpm durante 2 min y se transfirió el sobrenadante a un tubo nuevo para hacer un lavado con un volumen de cloroformo y centrifugar nuevamente. Se cogió el sobrenadante midiendo el volumen recuperado, para añadir 1/3 de LiCl 8 M. Mezclar bien y dejar en hielo al menos 2 h en reposo. Después, se centrifugaron las muestras a 13.000 rpm durante 30 min a 4°C. Se formó un pellet translucido, por lo que hay que eliminar el sobrenadante y disolver el pellet en 0,5 volúmenes de H₂O. Posteriormente, se añadió 1/10 de Acetato de Sodio 3 M, pH 5,2. Se mezcló y se lavó con un volumen de etanol 100% frío. Se dejó precipitar a -20°C toda la noche.

Al día siguiente, se centrifugó durante 30 min hasta eliminar residuos acuosos. De aquí en adelante se utilizó el Kit *RNeasy Plus Mini* según las indicaciones del fabricante. Finalmente se eluyó en 60 µL de agua estéril libre de RNasa. Se cuantificó el RNA en un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies) a un ratio de absorbancia de 260 nm/280 nm ≥ 2 . Para fines de RNA-seq, se requieren de 200 a 400 ng de RNA y valores de RIN (RNA Integrity Number) $\geq 7,0$ y relación RNA 28S/ RNA 18S $< 0,7$ (Masotti y Preckel, 2006). Con esta metodología de extracción se obtuvieron 200 ng de RNA total. Para validar la calidad y cantidad del RNA se enviaron al CNIO para realizarles una electroforesis por nanocapilaridades con el equipo Bioanalyzer 2100 (Agilent). Este método es más fiable que la electroforesis de poliacrilamida o agarosa.

3.7. Métodos de amplificación

3.7.1. PCR (Reacción en cadena de la polimerasa)

La amplificación de DNA genómico mediante PCR se llevó a cabo en termocicladores Thermal Cycler 2720 (Applied Biosystems). La mezcla para una reacción incluyó: Oligo *forward* (sentido) y oligo *reverse* (antisentido) a una concentración de 0,5 µM cada uno; mezcla de dNTPs (dATP, dTTP, dCTP y dGTP) a una concentración 200 µM; 1 µL de MgCl₂ 2 mM; 0,4 mg/L de BSA; buffer suministrado con la polimerasa a una concentración 1x y 0,8 U de la enzima Long range DNA polimerase de Qiagen. Los oligonucleótidos y las condiciones utilizadas para la reacción en cada caso concreto se detallan en la Tabla 6 de Resultados. Los productos de amplificación resultantes de la PCR se analizaron mediante electroforesis en geles de agarosa tal como en el punto número 10 de la extracción de DNA genómico apartado 3.6.1.

3.7.2. Tratamiento con DNAsas y síntesis de cDNA

Antes de realizar la síntesis de cDNA, el RNA tiene que ser tratado con DNAsas para evitar contaminación con DNA genómico en las muestras. Con este

fin, se utilizó el Kit *Turbo DNA-Free* (Ambion, Applied Biosystems). Se utilizó siguiendo las instrucciones recomendadas por el fabricante. El cDNA se sintetizó a partir de 1,5 µg de RNA en 30 µL de reacción del kit *PrimeScriptTM RT reagent (Perfect Real Time)* de Takara en presencia de los cebadores “Oligo dT” y “Random 6 mers”, en las cantidades recomendadas por el fabricante, con una incubación de 90 minutos a 37°C.

3.7.3. Cuantificación en tiempo real (qRT-PCR)

Las PCRs en tiempo real se llevaron a cabo en el termociclador “ABI PRISM 7300” (Applied Biosystems). Todas las reacciones fueron realizadas en un volumen final de 20 µL conteniendo 10 µL de la mezcla SYBR[®] Green 2x (Applied Biosystems), 0,2 µM de cada oligonucleótido, y 1 µL de cDNA. Para los genes control se utilizó el cDNA a una dilución 1:100. Los ciclos de amplificación consistieron en un ciclo inicial de 2 min a 50°C, un ciclo de desnaturalización de 10 min a 95°C, seguido por 40 ciclos de amplificación de 15 seg a 95°C y 1 min a 60°C y un ciclo de disociación de 15 seg a 95°C para asegurar la presencia de un solo amplicón. Los datos fueron exportados desde el software 7300 Real-Time PCR System. Para el análisis de los niveles de expresión se realizó una cuantificación relativa del gen de estudio, usando el método Ct ($2^{-\Delta\Delta C_t}$) expuesto por Livak y Schmittgen, 2001.

3.8. Análisis de RNA-seq

3.8.1. Construcción y secuenciación de librerías de cDNA

El RNA total fue extraído tal como se describe en el apartado 3.6.3 La secuenciación se realizó en el Instituto de Genómica de Beijing, China (BGI). Se contó con tres replicas biológicas por cada punto de tiempo y tratamiento. Las colas Poly(A) del mRNA fueron extraídas del RNA total usando esferas magnéticas Dynabeads (Life Technologies). Fue agregado un buffer de fragmentación para romper el mRNA purificado en fragmentos cortos de 200 pb. La primera cadena de cDNA fue sintetizada con hexámeros (sintetizados al azar) y la segunda cadena fue sintetizada

agregando buffer, dNTPs, RNAase H y DNA polimerasa I. La doble cadena de cDNA fue purificada usando esferas magnéticas y se realiza la reparación final. Finalmente los adaptadores de secuencias son ligados a los fragmentos y son enriquecidos con amplificaciones de PCR. Se construyeron en total cinco librerías, a partir de las cuales se secuenció el correspondiente transcriptoma utilizando una plataforma High-Seq 2000 de Illumina. Después de obtener las lecturas crudas no pareadas de aproximadamente 50 pb. Se procede al alineamiento y análisis de las mismas. En la Figura 3, se muestra un diagrama del flujo del procesamiento de las lecturas de RNA- seq.

3.9. Métodos de plantas

3.9.1. Cultivo de chopo

Cultivo *in vitro*. Se realizó en tarros cilíndricos de 150 mL de capacidad. El medio contenía sales MS (Murashige y Skoog) 1x (Duchefa), 20 g/L sacarosa, 7 g/L de fitoagar y 104 mg/L de vitaminas Mc Cown's. El pH se ajustó a 5,8 con KOH antes de esterilizar en autoclave. Después se añadió como regulador de crecimiento la auxina IBA (ácido indol-3-butírico) a una concentración final de 0,5 mg/L. Para evitar contaminaciones, la solución de IBA se pasó por un filtro Millipore estéril con un tamaño de poro de 0,22 μm . Los explantos se sub-cultivaron cada 6 - 8 semanas y se mantuvieron en cámaras de cultivo SGC 970/C/RO-HFL (Sanyo) en condiciones controladas de temperatura (22°C) y fotoperiodo (16 h de luz y 8 h de oscuridad). Para la iluminación se emplearon tubos de luz fluorescente blanca Sylvania Cool White F72T12/CW/VHO, con una densidad de flujo fotónico de $120 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

3.9.2. Cultivo de *Arabidopsis thaliana*

Cultivo *in vitro*. Para el cultivo *in vitro* las semillas se esterilizaron 5 min con una solución de etanol al 70% (v/v) y Tritón-X100 al 2% (v/v) en agua. Después se lavaron tres veces con 1 mL de etanol absoluto. Tras secarlas con papel de filtro estéril, se sembraron alrededor de 20 semillas por placa Petri de 90 mm de diámetro,

colocándolas manualmente. Las placas contenían medio MS (sales MS 1x, 20 g/L sacarosa y se ajustó a pH 5,7) solidificado con fitoagar (Duchefa) al 0,8% (p/v). Tras permanecer dos días en oscuridad a 4°C, las placas se mantuvieron en cámaras de cultivo a una temperatura de 22°C, fotoperiodo corto (10 h de luz y 14 h de oscuridad) y una densidad de flujo fotónico entre 100 y 125 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

Cultivo en maceta. Las semillas se sembraron en alveolos de plástico de 5 x 5 cm, a razón de 40 alveolos por bandeja. Se utilizó como sustrato una mezcla de turba (Kekkila Ibérica) y vermiculita (Projar) previamente humedecida y esterilizada. Tras un periodo de 2-3 días en oscuridad y a 4°C, los alvéolos con 8-10 semillas se colocaron en cámaras Conviron CMP 4030. La temperatura se mantuvo a 22°C y la humedad relativa al 70%. Utilizamos una densidad de flujo fotónico en torno a 100 y 125 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ y un fotoperiodo corto (10 h de luz y 14 h de oscuridad).

3.9.3. Tratamientos con Aroclor 1221 (PCBs)

Para los tratamientos con estos xenobióticos se evaluó previamente el crecimiento de los chopos *in vitro* en presencia de distintas concentraciones de Aroclor 1221 (mezcla de PCBs comercializada por Sigma Aldrich) diluido en DMSO estéril: 0 (control), 50, 100 y 200 mg/L. Estos experimentos se realizaron en condiciones *in vitro*, utilizando en cada tarro 20 mL de MS-fitoagar suplementado con DMSO (controles) o Aroclor 1221 diluido en el mismo volumen de DMSO.

Tras determinar el efecto de los PCBs sobre el crecimiento vegetativo se monitorizaron la biomasa foliar y la longitud radicular decidimos realizar los tratamientos definitivos a 50, 100 y 200 mg/L de Aroclor 1221. A partir de dichos tratamientos se obtuvieron las muestras de proteína y RNA analizadas en esta tesis.

3.10. Métodos de ácidos nucleicos

3.10.1. Extracción de DNA genómico de chopo

Para la extracción de DNA genómico de chopo se partió de 100 mg de tejido de chopo *in vitro* pulverizadas en nitrógeno líquido. La extracción se realizó con el Kit *DNesay Plant Mini* (Qiagen). Se verificó con un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies) a un ratio de $260/280 \geq 1,8$. Para visualizar el DNA se utilizaron geles de agarosa al 0,8 % (p/v), preparados en tampón TAE (Tris HCl 40 mM, pH 8,0, acetato sódico 20 mM, EDTA 1 mM) suplementado con bromuro de etidio a 0,1 mg/L. Las muestras se mezclaron con tampón de carga: azul de bromofenol al 0,05% (p/v) y glicerol 5% (p/v). Una vez terminada la electroforesis el DNA se visualizó en un equipo Molecular Imager Gel Doc XR. Para la cuantificación se utilizó el software Quantity One 4.6.5 de Bio Rad.

3.10.2. Extracción y secuenciación de DNA plasmídico

El aislamiento de plásmidos se llevó a cabo con el Kit *QIAprep Spin Miniprep* (Qiagen). El protocolo se basa en una lisis alcalina y posterior adsorción de DNA plasmídico a una resina específica. La calidad del DNA se verificó con un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies) como se indicó en el apartado anterior (3.6.1).

Para comprobar el tamaño del inserto se utilizaron se utilizaron enzimas de restricción apropiadas. Los fragmentos fueron visualizados en geles de agarosa al 0,8%. La secuenciación se realizó en el Centro Nacional de Investigaciones Oncológicas, que cuenta con un sistema ABI 3730 (Applied Biosystems).

3.10.3. Extracción de RNA total de chopo

La extracción de RNA se realizó a partir de 200 mg de pulverizado (de planta *in vitro* completa). Se utilizó el método de fenol-cloroformo y posteriormente se purificó utilizando el kit *RNeasy Plus Mini* (Qiagen).

Para este fin se añadió una relación 3:1 de tampón de extracción (0,2 M Tris-HCl pH 9,0, 0,4 M LiCl, 25 mM EDTA y 1% SDS) y se mezcló durante 60 s. Una vez mezclado, se agregó un volumen igual de fenol y se mezcló vigorosamente durante 10 seg hasta homogeneizar. Se centrifugó entonces a 13.000 rpm durante 2 min y se recuperó el sobrenadante. Se agregó a este 1 volumen de fenol, se centrifugó a 13.000 rpm durante 2 min y se transfirió el sobrenadante a un tubo nuevo. Este proceso se repitió una segunda vez con cloroformo. Finalmente, se añadió 1/3 volúmenes de LiCl 8 M, se mezcló bien y se dejó en hielo al menos 2 h. Después, se centrifugaron las muestras a 13.000 rpm durante 30 min a 4°C y se decantó el sobrenadante. Se formó un pellet translucido que se disolvió en 0,5 volúmenes de H₂O. Posteriormente, se añadió 1/10 de acetato de sódico 3 M, pH 5,2. Se añadieron 2 volúmenes de etanol absoluto frío dejando precipitar de 16-18 h a -20°C. Tras centrifugar 30 min a 13.000 rpm, el RNA se purificó con el Kit *RNeasy Plus Mini* según las indicaciones del fabricante. Finalmente se resuspendió en 60 µL de agua estéril libre de RNasa. El RNA Se cuantificó en un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies). Para fines de RNA-seq, se requieren de 200 a 400 ng de RNA, valores de RIN (RNA Integrity Number) $\geq 7,0$ y una relación RNA 28S/ RNA 18S $< 0,7$ (Masotti y Preckel, 2006). Con esta metodología de extracción se obtuvieron al menos 200 ng de RNA de buena calidad a partir de 200 mg de material fresco. La calidad y cantidad de RNA se evaluaron mediante electroforesis por nanocapilar en un equipo Bioanalyzer 2100 (Agilent). Este método es mucho más fiable que la electroforesis.

3.11. Métodos de amplificación

3.11.1. PCR (Reacción en cadena de la polimerasa)

La amplificación de DNA genómico mediante PCR se llevó a cabo en termocicladores Thermal Cycler 2720 (Applied Biosystems). La mezcla para una reacción incluyó: oligonucleótido *forward* (sentido) y oligonucleótido *reverse* (antisentido) a una concentración de 0,5 µM cada uno; una mezcla de dNTPs (dATP, dTTP, dCTP y dGTP) a una concentración 200 µM; 1 µL de MgCl₂ 2 mM; 0,4 mg/L de BSA; tampón suministrado con la polimerasa a una concentración 1x y 0,8 U de

la enzima Long- range DNA polimerase de Qiagen. Los oligonucleótidos y las condiciones utilizadas para la reacción en cada caso concreto se detallan en la Tabla 6 de Resultados. Los productos de amplificación resultantes de la PCR se analizaron mediante electroforesis en geles de agarosa según como se describe en el apartado 3.6.1.

3.11.2. Tratamiento con DNAsas y síntesis de cDNA

Antes de realizar la síntesis de cDNA, el RNA fue tratado con DNAsas para evitar contaminación con DNA genómico en las muestras. Con este fin se utilizó el Kit *Turbo DNA-Free* (Ambion, Applied Biosystems) siguiendo las instrucciones recomendadas por el fabricante. El cDNA se sintetizó a partir de 1,5 µg de RNA en 30 µL de reacción del kit *PrimeScriptTM RT reagent (Perfect Real Time)* de Takara en presencia de los cebadores “Oligo dT” y “Random 6 mers”, en las cantidades recomendadas por el fabricante, con una incubación de 90 minutos a 37°C.

3.11.3. PCR cuantitativas en tiempo real (qRT-PCR)

Las recientes PCR en tiempo real se llevaron a cabo en un termociclador ABI PRISM 7300 (Applied Biosystems). Todas las reacciones se realizaron en un volumen final de 20 µL, conteniendo 10 µL de la mezcla SYBR[®] Green 2x (Applied Biosystems), 0,2 µM de cada oligonucleótido, y 1 µL de cDNA (apartado 3.7.2.). Para los genes control se utilizó el cDNA a una dilución 1:100. Los ciclos de amplificación consistieron en un paso inicial de 2 min a 50°C, un paso de desnaturalización de 10 min a 95°C y por 40 ciclos de amplificación de 15 seg a 95°C y 1 min a 60°C, para asegurar la presencia de un solo amplicón. Los datos fueron analizados por un software 7300 Real-Time PCR System. Para el análisis de los niveles de expresión se realizó una cuantificación relativa del gen en estudio, usando el método Ct ($2^{-\Delta\Delta C_t}$) expuesto por Livak y Schmittgen, 2001.

3.12. Análisis de RNA-seq

3.12.1. Construcción y secuenciación de librerías de cDNA

El RNA total fue extraído tal como se describe en el apartado 3.6.3 La construcción de las librerías y la tecnología de secuenciación fueron realizadas en el Instituto de Genómica de Beijing, China (BGI). (<http://bgi-international.com>). Se contó con tres replicas biológicas por cada punto de tiempo y tratamiento. Se construyeron en total cinco librerías, a partir de las cuales se secuenció el correspondiente transcriptoma utilizando una plataforma High-Seq 2000 de Illumina. Después de obtener las lecturas crudas no pareadas de aproximadamente 50 pb. Se procede al alineamiento y análisis de las mismas. En la Figura 11, se muestra un diagrama del flujo del procesamiento de las lecturas de RNA- seq.

Figura 11. Representación del flujo de análisis del RNA-seq. Primero se crea un índice de la referencia (Índice ref), después se hace el alineamiento con cada una de las condiciones y el control usando Bowtie2 (Ej: Alineamiento_C1_5d x Alineamiento_Control) para crear las estadísticas a través de matrices de conteo mediante el programa Phyton script sam2counts.py. Finalmente se hace el análisis de expresión diferencial con el programa DEGen Hunter. Las flechas en color rojo indican los programas utilizados.

3.12.2. Pre-procesamiento de secuencias

Para obtener lecturas fiables a partir de las lecturas crudas generadas durante la secuenciación directa de las librerías cDNA utilizamos el programa SeqTrimNext (<http://www.scbi.uma.es/seqtrim>, Falgueras et al., 2010) disponible en la Plataforma Andaluza de Bioinformática de la Universidad de Málaga.

El software está diseñado para filtrar las secuencias demasiado cortas (menos de 40 nucleótidos), así como secuencias mal determinadas o secuencias inapropiadas (secuencias bacterianas, secuencias de plásmidos, colas poli A o poli T, etc.). Para poder hacer el pre-procesamiento se requiere que las lecturas crudas se encuentren en formato fastQ el cual es un formato de texto como el formato fasta convencional, sin embargo en el formato fastQ la "Q" se refiere a "quality" e implica que cada posición, además de tener una letra, tiene asociado un valor de calidad (Cock et al., 2010).

3.12.3. Alineamiento sobre el genoma de referencia *P. trichocarpa*

Alineamiento en el genoma. Fue realizado con el programa SOAP (Short Oligonucleotide Analysis Package) (Li et al., 2008), diseñado para el manejo de secuencias cortas generadas en plataformas Illumina. Las lecturas fueron alineadas sobre la última versión del genoma de *P. trichocarpa* disponible en la base de datos Phytozome (www.phytozome.net, versión 3.0). La calidad de la secuenciación fue estimada mediante un análisis de saturación y también mediante evaluación aleatoria. Para el mapeo se utilizó como referencia el transcriptoma.

Alineamiento en el transcriptoma. Los análisis fueron realizados en el Centro de Supercomputación y Bio-innovación de la Universidad de Málaga, lugar donde realicé una estancia. Las lecturas se alinearon sobre la última revisión disponible del transcriptoma de *P. trichocarpa* en la base de datos Phytozome (www.phytozome.net, versión 3.0). Para lo análisis utilizamos programación en lenguaje Bash (Bourne again Shell), basado en la Shell de Unix, un intérprete de datos del sistema operativo Linux. Además se requiere de un editor de texto simple que no introduzca caracteres innecesarios en los ficheros de Linux. En nuestro caso se utilizó el Sublime Text3 (<http://www.sublimetext.com/3>). El alineamiento de las

secuencias teniendo como referencia el transcriptoma de *P. trichocarpa* se llevó a cabo con el programa Bowtie2 (Langmead y Salzberg, 2012). Se trata de un sistema muy eficiente para alinear lecturas solapantes en formato SAM (Sequence Alignment Map), un formato genérico para almacenar largas secuencias de nucleótidos. Para transformar las lecturas crudas a formato SAM, utilizamos el paquete SAMtools.

Una vez descargado el transcriptoma entero de la base del Phytozome, se usó *Ptrichocarpa_210_c3.0.transcript.fa* para hacer el índice de la referencia (Index Ref). El índice creado no cambiará y se realizará el mapeo de cada secuencia única haciendo referencia a este índice. Posteriormente, se crearon matrices de conteo (Count) para la detección de genes expresados diferencialmente alineando una condición con respecto al control. Estas instrucciones se ejecutan en el programa Bowtie2 tal como se muestra en la Figura 9.

3.12.4. Genes diferencialmente expresados (DEGs)

DEGs del genoma: Para la cuantificación de la expresión génica se utilizó el método RPKM (Reads per Kilobase per Million mapped Reads) (Mortazavi et al., 2008). Este sistema normaliza el conteo de cada transcrito según su longitud y el número total de lecturas del mismo en la muestra. Para estimar los niveles de expresión solo utilizamos lecturas mapeadas. Se obtuvo así una lista de genes expresados diferencialmente, utilizando el algoritmo desarrollado por Audic y Claverie (1997). Como en el caso anterior, los valores p fueron corregidos, usando en este caso un FDR (False Discovery Rate) ≤ 0.001 como umbral (Benjamini y Yekutieli, 2001). Para su visualización y agrupamiento se realizaron mapas de calor con el programa Tree viewer 3.0. A su vez se realizaron diagramas de Venn para encontrar los genes que interaccionaban en los cuatro tratamientos.

DEGs del transcriptoma: El análisis de expresión diferencial se llevó a cabo con el programa DEgene HUNTER (González-Gayte et al., 2015), un programa que combina cuatro algoritmos diferentes para detectar DEGs: DESeq2 (Love et al., 2014), limma (Ritchie, 2015), edgeR (Robinson et al., 2010) y NOISeq (Tarazona et al., 2011). Estos algoritmos utilizan el mismo método de normalización TMM

(Trimmed Mean of M-values) (Robinson y Oshlack, 2010) para la cuantificación de la expresión. Los valores p fueron corregidos con el método FDR (False Discovery Rate) descrito por Benjamini y Yekutieli (2001). La expresión diferencial se cuantificó como cambio absoluto (*fold change*) utilizando como umbral un $FDR \leq 0,05$. Para aumentar la significancia estadística solo aquellos genes resultantes de la intersección de los cuatro algoritmos fueron utilizados en los análisis posteriores. Se realizaron también diagramas de Venn para observar la interacción de los cuatro tratamientos. Realizamos las siguientes comparaciones a partir de la correspondiente tabla de conteo: 1) Control vs. C1_5d, 2) Control vs. C1_15d; 3) Control vs. C2_5d; y 4) Control vs. C2_15d.) C1 se refiere a la concentración más baja (50 mg/L) y C2 la concentración más alta (200 mg/L). 5d (5 días), indica el tiempo temprano y 15d (15 días) indica tiempo tardío.

3.12.4.1. Análisis de series temporales

Para analizar series temporales de datos transcriptómicos hemos utilizado el programa MaSigPro (Microarray Significan Profile; Conesa et al., 2006), que utiliza un modelo de regresión polinómica para identificar cambios de expresión a lo largo del tiempo. El paquete incluye herramientas para agrupar y visualizar los genes que muestran los mismos patrones de expresión. Esta metodología se ha utilizado con éxito para identificar funciones celulares o procesos biológicos (Nueda et al., 2014). El programa corrige el p-value para múltiples comparaciones aplicando regresión lineal y el FDR (Benjamini y Hochberg, 1995). Analizamos en concreto dos concentraciones de PCBs (Aroclor 1221), a los cuales se denominaron: "trat1" de 50 mg/L y "trat2" de 200 mg/L, ambos en dos puntos temporales (días 5 y 15), tras el inicio del experimento.

3.12.5. Anotación funcional

Los transcritos que mostraron abundancia diferencial se anotaron a partir de alineamientos BLAST (2.2.25) con valores de $E \leq 1E-05$.

Para el enriquecimiento funcional GO (Ontología génica) (Ashburner et al., 2000) de los transcritos diferencialmente expresados, se usó el programa Blast2GO 2.2.5 (Conesa et al., 2005). Este usa un modelo de distribución hipergeométrica y desarrolla una corrección con el método de Bonferroni para refinar el valor p . Se realizó una distribución de funciones moleculares, procesos biológicos y componentes celulares. La interpretación biológica de los DEGs en respuesta a Aroclor 1221 se hizo mediante análisis de familias de proteínas pfam y asignando las rutas metabólicas a través de la anotación de KEGGs (Enciclopedia de Genes y Genomas de Kioto).

3.12.6. Análisis de redes de co-expresión génica

Se realizaron estos análisis para detectar posibles candidatos a participar en rutas de degradación de PCBs. Para ello se utilizó la herramienta exNet (<http://popgenie.org/exnet>), que visualiza las redes de co-expresión detectadas. Las medidas de correlación vienen dadas por un valor CLR (Context Likelihood of Relatedness), basado a su vez en un valor de información mutua (MI), esto es, una probabilidad de relación entre dos genes (Shiquan Wu y Jing Li, 2007). El programa exNet (<http://popgenie.org/exnet>) genera una imagen donde los nodos que se muestran como cuadros amarillos son los factores transcripcionales y los círculos en color verde los genes. Para la visualización de imágenes de co-expresión se utilizó como plataforma el programa Cytoscape 3.2.1 (Shannon et al., 2003).

3.12.7. Validación de los datos de expresión génica

Para validar los resultados de RNA-seq, se cuantificaron quince genes mediante qRT-PCR con oligonucleótidos específicos. La metodología se detalla en el apartado 3.7.3. Los genes fueron seleccionados por su elevada expresión diferencial, por la fiabilidad de los datos de cuantificación ($p < 0,05$) y por el interés de su hipotética función en el contexto de la respuesta a PCBs.

3.13. Métodos de proteínas

3.13.1. Extracción y precipitación de proteínas

Para la extracción de proteínas se siguió el protocolo descrito por Demerval (1986) con algunas modificaciones. Se extrajeron 500 mg de tejido pulverizado añadiendo una solución fría de TCA (10 % (v/v), 2-mercaptoetanol 0,07 % (v/v) y acetona 90 % (v/v), durante 60 min a 4°C, mezclando ocasionalmente. Se centrifugaron las muestras durante 15 min a 15.000 rpm y 4°C. Los precipitados se lavaron con 1,5 mL de acetona fría con 2-mercaptoetanol 0,07 % (v/v) y se incubaron durante 30 min a 4°C (dos veces). Se eliminaron los restos de acetona en un *speed-vac* durante 20 min.

El precipitado se incubó en un agitador orbital durante 10 min en 1 mL de tampón RB (Rehydration buffer): urea 7 M, tiourea 2 M, CHAPS (3-[(3-colamidopropil) dimetilamonio]-1-propanosulfonato) 4 % (p/v), Triton X-100 0,2 % (v/v) y DTT (ditiotreitól) 10 mM. Se centrifugó durante 5 min a 14.000 rpm todo el extracto proteico, con el fin de eliminar cualquier partícula insoluble. El extracto fue re-suspendido en 200 µL de tampón RB y se cuantificó tal como se relata en el apartado siguiente.

3.13.1.1. Cuantificación de proteínas

Para determinar la concentración de proteínas en las muestras se realizó una cuantificación por un método espectroscópico basado en el de Bradford (1976). Se empleó el Kit Coomassie Plus Protein Assay Reagent (Pierce) que ofrece una respuesta óptima para concentraciones de proteína entre 1 y 1.500 µg/mL y para proteínas con peso molecular superior a 3 KDa. Para el diseño de la recta patrón, que relaciona absorbancia con concentración, se usaron muestras de albumina de suero bovino (BSA) de concentraciones conocidas (50, 125, 250, 500 y 750 µg/mL).

Tras la extracción, cada muestra se diluyó 6 veces con agua milli-Q destilada (10 μ L de muestra y 50 μ L de agua). Esta solución se dividió a su vez en dos de 30 μ L y se añadieron 900 μ L de Coomassie Plus Protein Assay a cada una. Se mezclaron bien y se dejaron incubar 10 min. Posteriormente, se procedió con la calibración del espectrofotómetro Ultraspec 3.300 pro UV/visible (Amersham Biosciences) y la medición de la absorbancia de los patrones y las muestras a una longitud de onda de 595 nm. Se midieron dos replicas por cada punto de muestreo. Una vez medidas las absorbancias se estimó la cantidad de proteína presente en cada muestra interpolando en la recta de regresión diseñada con los patrones de concentración conocida.

3.13.2. Electroforesis unidimensional en geles de poliacrilamida con SDS

Los análisis electroforéticos se realizaron en unos mini-geles SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis). La electroforesis se realizó según el método Laemmli, (1970) utilizando el sistema *Miniprotean II* (Bio-Rad). El gel separador contenía un 15% (p/v) de poliacrilamida en tampón Tris-HCl 0,375 M, pH 8,8, SDS 0,1 % (p/v), persulfato de amonio 0,1 % (p/v) y TEMED 0,04 % (p/v). El gel espaciador contenía un 4 % (p/v) de poliacrilamida en Tris-HCl 0,125 M, pH 6,8, SDS 0,1 % (p/v), persulfato amónico 0,1 % (p/v) y TEMED (Tetramethylethylenediamine) 0,1 % (p/v). Como tampón de electroforesis se utilizó Tris-HCl 25 mM, pH 8,3, glicerol 192 mM y SDS 0,1% (p/v). Las proteínas se re-suspendieron en Tris-HCl 0,0625 M; pH 6,8, SDS 0,2 % (p/v), urea 8 M y 2-mercaptoetanol 5 % (v/v). Antes de la electroforesis las muestras se incubaron 5 min a 100°C. Para eliminar el material insoluble se centrifugaron 5 min a 14.000 rpm. Como colorante se utilizó azul de bromofenol 0,001 % (p/v). Para la determinación de la masa molecular se utilizaron mezclas comerciales de proteínas marcadoras (Bio-Rad) que cubren el intervalo 14-98 kDa. La electroforesis se realizó a 150 Voltios durante 90 min, empleándose una fuente de alimentación *Power Pac Basic* (Bio-Rad).

3.13.3. Electroforesis bidimensional de alta resolución

3.13.3.1. Primera dimensión: Isoelectroenfoque (IEF)

El isoelectroenfoque permite separar las proteínas en función de su punto isoelectrónico. Una vez resuspendidas las muestras, éstas se repartieron por los geles de IEF, montados en tirillas de acetato (*Immobiline DryStrips*) dentro de *sarcófagos Ettan IPGphor Strip Holder* siguiendo las instrucciones del fabricante (Amersham Biosciences). Se utilizaron geles de IEF de 18 cm de longitud con unos gradientes lineales de pI 3-10. Posteriormente se colocó cada soporte cerámico, con su respectiva muestra, sobre la fuente de alimentación y se aplicó el siguiente programa para fraccionar las proteínas (Tabla 55):

Tabla 5. Programa de fraccionamiento IEF de proteínas. Voltaje y tiempo aplicados a la fuente de alimentación.

Paso	Cambio de voltaje	Voltaje (V)	Tiempo (h)	V x h
1	Constante	30	6	180
2	Constante	60	6	360
3	Constante	500	1	500
4	Constante	1.000	1	1.000
5	Gradiente	8.000	3	13.500
6	Constante	8.000	6	48.000
TOTAL			24	63.540

Una vez terminada la primera dimensión, se retiraron los soportes y se extrajeron las tiras con las proteínas teóricamente “enfocadas”. Estas tiras se lavaron con agua destilada para eliminar el tampón IPG y se almacenaron a -20°C hasta la realización de la segunda dimensión.

3.13.3.2. Segunda dimensión: SDS-PAGE

Se necesitan 3 etapas para la segunda dimensión: Polimerización de geles, equilibrado e inserción de los geles de primera dimensión y electroforesis SDS-PAGE. Se utilizaron geles de poliacrilamida de 26 x 20 cm. Antes de pulverizar se limpiaron los cristales minuciosamente con alcohol al 70%. Posteriormente se preparó la mezcla de polimerización: 11,5% (p/v) de poliacrilamida (2,6% de N, N'-metilen-bis-acrilamida respecto al total), Tris-HCl 375 mM pH 8,8, SDS 0,1%, TEMED 0,08% (v/v), persulfato amónico 0,08% (v/v) y agua destilada 0,3% (v/v). Tras remover la mezcla durante unos min para asegurar una buena distribución de los componentes, se vertió cuidadosamente entre los cristales. Después de tres a cuatro horas los geles se cubrieron con tampón de almacenamiento para prevenir su deshidratación (Tris-HCl 375 mM; pH 8,8 y SDS al 0,1% en agua destilada). Se mantuvieron en la cámara fría (4°C) dos días, para asegurar una polimerización completa de la acrilamida. Antes de comenzar la segunda dimensión se equilibraron las tiras de isoelectroenfoque con tampón Tris-HCl 50 mM, pH 8,8, que contenía urea 6 M, glicerol 30% (v/v) y SDS 2% (p/v). Este proceso constó de dos fases:

En la primera fase (20 min) se añadió DTT al 2% (p/v) a la solución y en la segunda fase se añadió iodoacetamida al 2,5% (p/v) y una traza de azul de bromofenol. Transcurrida la incubación anterior, se eliminó el tampón de almacenamiento y se introdujeron las tiras entre los cristales con ayuda de unas pinzas, hasta que entraron en contacto directo con los geles. Se quitaron las burbujas entre ambos geles con una espátula y se inmovilizaron las tiras con agarosa al 1% en agua con una traza de azul de bromofenol. Los geles se ensamblaron en el sistema de Amersham Biosciences llamado Ettan Dalt-Six, que con capacidad para 6 geles. Se utilizó un tampón de cubetas 1X (Tris 250 mM, glicina 1,92 M y SDS 1% (p/v)). Una vez ensamblados todos los elementos, se conectó la cubeta a un refrigerador de circulación que mantuvo el sistema de electroforesis a unos 15°C, impidiendo sobrecalentamientos. El sistema se conectó a una fuente de alimentación Power Pac™ Universal 164-5070. La electroforesis se programó en dos fases, una primera de 30 min de duración a una potencia de 3W por gel y una segunda a 18W por gel durante unas 4 h, finalizando ésta en el momento en el que el azul de bromofenol alcanzó la parte inferior de los geles.

3.13.4. Tinción de proteínas

Una vez finalizadas las electroforesis (uni y bidimensionales), se procedió a la tinción de las proteínas con Coomassie Brilliant Blue R-250 (BioRad). Los geles se introdujeron en una solución acuosa que contenía Coomassie al 0,05% (p/v), metanol al 40% (v/v) y ácido acético al 7% (v/v), durante 4 h para minigeles y 48 h para geles bidimensionales. Los geles se mantuvieron bajo agitación constante en la cámara fría (4°C). Trascurrido el tiempo antes dicho, se procedió a la des-tinción en una solución de metanol al 40% (v/v) y ácido acético al 10% (v/v) durante 4 h para geles unidimensionales y 24 h para el gel bidimensional, en agitación constante y en la cámara fría, cambiándose la solución de desteñido cuando se saturaba de colorante. Finalmente, los geles se embolsaron en una solución para almacenamiento a largo plazo compuesta por un 7% de ácido acético y un 5% de metanol.

3.13.5. Identificación de la proteína PtCSP4

Las proteínas de interés se recortaron de los geles con un bisturí y unas pinzas estériles. Cada pieza de acrilamida se cubrió con una pequeña cantidad de agua milli-Q estéril para su traslado al Servicio de Proteómica del Centro Nacional de Biotecnología (CNB). Cada proteína se digirió con tripsina y los péptidos resultantes se purificaron. Las masas precisas de los péptidos (huellas peptídicas) se determinaron mediante espectrometría de masas (MS). MALDI-TOF/TOF (Matrix-Assisted Laser Desorption/Ionization-Time of Flight). Las huellas peptídicas son características de cada proteína y suelen permitir su identificación usando un software apropiado y bases de datos de proteínas o DNA traducido. La cantidad de proteína en los geles bidimensionales es importante para una identificación exitosa. Los resultados del servicio de espectrofotometría se usaron para comparar con predicciones teóricas resultantes de digerir virtualmente con tripsina proteínas de una base de datos. Todas las secuencias obtenidas se caracterizaron mediante el software alojado en el servidor ExPASy (Expert Protein Analysis System) del Swiss Institute of Bioinformatics. Una vez obtenida la secuencia de un péptido mediante MS/MS, se buscaron proteínas/genes homólogos utilizando bases de datos apropiados.

El análisis semicuantitativo de los cambios de expresión proteica se realizó con los programas *Quantity One* v4.6.3 (Bio-Rad) e *Image Lab 5.0* (Bio-Rad). El valor obtenido para cada *spot* fue expresado como porcentaje de volumen relativo en relación a los controles.

3.14. Análisis filogenéticos

3.14.1. Base de datos para alineamientos y análisis filogenéticos

Las búsquedas de secuencias nucleótídicas y aminoacídicas se realizaron en la base de datos Phytozome V.10.3 (<http://phytozome.jgi.doe.gov/pz/portal.html>) alojada en el Joint Genome Institute (JGI). Los alineamientos múltiples de secuencias se llevaron a cabo con el servidor ClustalW (<http://www.ch.embnet.org/software/ClustalW.html>). Para la identificación de dominios conservados se utilizó BoxShade 3.21. (http://www.ch.embnet.org/software/BOX_form.html). Estos dominios fueron analizados mediante un (Novo motif analysis) con el servidor MEME suite 4.10.1 (<http://meme-suite.org/tools/meme>).

Los árboles filogenéticos basados en secuencias nucleótídicas se elaboraron con el programa Mega 5.0. (Molecular Evolutionary Genetics Analysis), (Tamura et al., 2007) utilizando el método Neighbor-Joining y con 1.000 repeticiones bootstrap.

3.15. Método de clonaje

Todos los clonajes de la tesis se realizaron con el sistema Gateway (Invitrogen Life technologies), basado en recombinaciones homólogas (Hartley et al., 2000). Este sistema es una buena alternativa a los métodos de clonaje tradicional basados en digestiones y ligaciones. Este sistema necesita la conversión de los tradicionales plásmidos binarios de plantas en los llamados “plásmidos destino”, mediante la introducción de los llamados “cassetes Gateway”, que incorporan los sitios de recombinación attR1 y attR2, así como el gen letal *ccd4* para selección

negativa. Los oligos para amplificar los fragmentos a clonar contenían las siguientes secuencias: forward: AAAAAGCAGGCT seguido de unos 15 pares de bases del fragmento a clonar, comenzando por el triplete ATG (inicio de la traducción). Reverse: AGAAAGCTGGGT seguido de unos 15 pares de bases antisentido del fragmento a clonar, comenzando por el codón de fin de la traducción, pero cambiando una de las bases por otra distinta, con la finalidad de eliminarlo y así poder agregar más fragmentos de DNA y que éstos puedan transcribirse. En el sistema Gateway se realizan dos reacciones: Las amplificaciones fueron realizadas mediante PCR. Una vez realizada la ligación, se transformaron células competentes, generalmente *E. coli* TOP 10 para comprobar clonajes en pDONR, BL21 (DE3) para la expresión de la proteína en bacterias y *Agrobacterium* para los vectores binarios y posterior transformación de plantas. Las colonias obtenidas fueron analizadas mediante su cultivo y posterior aislamiento de plásmidos, comprobando la presencia del inserto por digestión y secuenciación.

3.15.1. Preparación de células y transformación con *Agrobacterium*

La preparación de células competentes y su transformación se realizó según el protocolo descrito por Hofgen y Willmitzer (1988), con algunas modificaciones. A partir de una colonia individualizada, se preparó un cultivo en 10 mL de medio LB con rifampicina (50 mg/L), gentamicina (25 mg/L) y kanamicina (50 mg/L). El cultivo se incubó a 28°C y 200 rpm durante 16 h. Al día siguiente, 1 mL se inoculó en 100 mL del mismo medio (dilución 1:100), incubando a 28°C y 200 rpm hasta que alcanzó una OD₆₀₀ ~ 0,8. Posteriormente el cultivo se centrifugó a 5.000 rpm y 4°C durante 5 min, se descartó el sobrenadante y se lavaron las células con 20 mL de NaCl 150 mM (estéril) previamente enfriado. Se centrifugaron de nuevo las células a 4°C y 5.000 rpm durante 5 min. Se eliminó el sobrenadante y el precipitado se resuspendió en 2 mL de CaCl₂ 20 mM (estéril y frío). Finalmente, el cultivo se alícuotó (50 µL/vial) en tubos estériles de 1,5 mL y se congelaron en nitrógeno líquido inmediatamente. El stock de células de *Agrobacterium* competentes se almacenó a -80°C. Para la transformación de *Agrobacterium*, se añadió 1 ng del plásmido pB2GW7 a 10 µL de células competentes (Agro 58C1). Se incubaron 5 min a 37°C con agitación suave, después, se mantuvieron en hielo durante 30 min.

Posteriormente se añadieron 500 µL de medio LB y se incubaron durante 3 h a 28°C y 150 rpm. Tras una centrifugación de 1 min a baja velocidad (2.000 rpm), el pellet de células se re-suspendió en 150 mL de medio LB y se plaqueó en LB-agar suplementado con rifampicina (50 mg/L) y espectinomicina (100 mg/L). Las colonias fueron visibles tras incubar dos días a 28°C.

3.15.2. Transformación genética de *A. thaliana*

Se utilizaron plantas sembradas en alveolos de 5x5 cm (9 semillas/alveolo) con inflorescencias de unos 10 cm de largo (3 semanas). La transformación se llevó a cabo mediante el método de infiltración al vacío de Clough y Bent (1998), partiendo de cultivos apropiados de *A. tumefaciens*. Células recombinantes de la cepa C58C1 (pMP90), transformada con pB2GW7, se inocularon en placas LB suplementadas con rifampicina (50 mg/L) y espectinomicina (50 mg/L). Tras dos días a 200 rpm y 28°C, se seleccionaron colonias individuales y se prepararon pre-inóculos en 3 mL de medio LB con los mismos antibióticos que en el medio sólido. A partir de éstos se inocularon 200 mL de medio LB con antibióticos, manteniendo el cultivo en las mismas condiciones de antes, hasta alcanzar una OD₆₀₀ de 1,2 aproximadamente. Las células se aislaron entonces por centrifugación (15 min a 4°C y 5.500 rpm en un rotor Beckman JA20) y se suspendieron en medio de infiltración (MI). Este contiene sales MS 1 X pH 5,7; Sacarosa 5 % (p/v), MES 0,05 % (p/v), bencilaminopurina (BAP) 0,044 µM, Silwet L 77 0,02 % (v/v) (Lehle Seeds), que se agrega inmediatamente antes de infiltrar. Las infiltraciones se realizaron durante 15 minutos aplicando vacío y se repitieron dos veces. Después de secar las plantas con papel absorbente, éstas se colocaron de nuevo en las cámaras de cultivo en condiciones normales de crecimiento (véase más arriba). Las líneas transformantes se seleccionaron mediante sembrado de semillas en medio MS suplementado con el herbicida BASTA (fosfinotricina) (10 µg/mL). Después de haber pasado dos semanas se transfirieron las *Arabidopsis* vivas a una nueva placa con herbicida para descartar falsos positivos. Para los análisis se obtuvieron las plantas T3. El éxito de la transformación se verificó mediante amplificaciones q-PCR.

4. RESULTADOS

4. Resultados

« *Las ciencias aplicadas no existen, solo las aplicaciones a la ciencia*».

Louis Pasteur (Químico y Bacteriólogo Francés)

4.1. Crecimiento vegetativo en presencia de Aroclor 1221

A pesar de que se han hecho grandes avances en el estudio de *Populus* poco se sabe de su respuesta a la contaminación, sobre todo a nivel molecular. En esta tesis hemos elegido los PCBs (Aroclor 1221) por razones que ya se han expuesto en la introducción (apartado 1.2.1.). Destacan en particular su peligrosidad, persistencia y su elevada toxicidad.

Para elucidar los mecanismos que se activan el contacto con los PCBs, se han comparado plantas de chopo sin tratar (control), con a plantas tratadas con Aroclor 1221, utilizando condiciones mencionadas en el apartado 3.5.3. de Material y Métodos. Posteriormente se ha llevado a cabo un análisis transcriptómico utilizando tecnología RNA-seq. Previamente optimizamos las condiciones de tratamiento, evaluando la sintomatología resultante de la exposición a PCBs. Esta sintomatología incluyó una disminución clara del crecimiento radical y menor producción de biomasa aérea. En la parte aérea se observó una coloración blanquecina conforme la concentración aumentaba. Por otra parte, la reducción de biomasa fue más pronunciada en la parte aérea que en las raíces. Las plántulas *in vitro* exhibían características de estrés, sin embargo, a las concentraciones ensayadas las plántulas continuaron sin crecimiento sin llegar a morir.

La longitud promedio de las raíces de las plantas control fue de 3,58 cm, mientras que a 5 días ésta fue de 4,0 cm a 50 mg/L de PCB y 3,1 cm a 200 mg/L de PCB. A los 15 días se obtuvo una medición de 2,54 cm a 50 mg/L y de 1,3 cm a 200 mg/L. Como observamos, a 200 mg/L la longitud de las raíces fue significativamente reducida a aproximadamente el 50% comparado con el tratamiento a 100 mg/L y 65% comparado con el control. El mismo patrón se observa en la evaluación de peso

seco foliar. La cantidad de biomasa fue en decremento de 0,37 g en el control a 0,13 g al tiempo de exposición y concentración más altos. Para los análisis transcriptómicos y proteómicos posteriores (Figura 12).

Figura 12. Efecto del Aroclor 1221 en plántulas de *Populus*. A) Crecimiento de plántulas tratadas con 50, 100 y 200 mg/L a 5 y 15 días. B) Peso seco foliar (biomasa). C) Crecimiento de raíces. Se muestran las medidas y desviaciones estandar para cada concentración, C1= 50 mg/L y C2= 200 mg/L (n=20).

4.1.2. Secuenciación de RNA de *Populus in vitro* tratados con Aroclor 1221

4.1.2.1. Calidad y cantidad de RNA

El RNA fue extraído tal como se describe en el apartado 3.6.3 de Material y Métodos. Se obtuvieron un total de 5 muestras de RNA según se describe en el apartado 3.5.3 de Material y Métodos. La calidad de dichas muestras se comprobó mediante electroforesis nanocapilar (Figura 13). Los resultados del análisis realizado por el Bioanalyzer nos señalaron que las concentraciones oscilaban entre 214 y 272 ng/ μ L, la relación promedio 28S/18S fue de 2,5 y el RIN promedio de 8,5. Estos resultados superaban lo requerido para un análisis de RNA-seq.

Figura 13. Análisis de electroforesis nanocapilar de RNA. 1) Control; 2) 50 mg/L a 5 días; 3) 200 mg/L a 5 días; 4) 50 mg/L a 15 días y 200 mg/L a 15 días.

4.1.2.2. Calidad y robustez de la secuenciación

La secuenciación por RNA-seq fue utilizada para generar números genes (transcritos) expresados por efecto del tratamiento con PCBs. Las cinco librerías fueron secuenciadas usando un sistema Illumina Hi-Seq 2000. Puesto que *P. trichocarpa* es un sistema modelo con genoma y transcriptoma secuenciados, las lecturas obtenidas fueron pre-procesadas y mapeadas sobre ellos. Para el control de calidad de la secuenciación, se estimó el nivel de saturación y el nivel de aleatoriedad. En RNA-seq la saturación puede alcanzarse cuando un incremento en el número de lecturas no supone una diferencia notable y la expresión permanece esencialmente contante (Tarazona et al., 2011). En este estudio, el número de genes

expresados muestra porcentajes superiores al 30% en los análisis de saturación para todas las librerías. Esto significa que la cantidad de lecturas obtenidas es suficiente para detectar los transcritos expresados diferencialmente. Los gráficos acumulativos de lecturas mapeadas indican que se alcanza prácticamente la saturación en los cinco experimentos analizados (Figura 14).

Figura 14. Saturación de las lecturas obtenidas respecto a la referencia. La saturación representa el porcentaje de genes identificados (eje Y) sobre la cantidad de lecturas (eje X), 1) control; 2) 50 mg/L a 5 días; 3) 200 mg/L a 5 días; 4) 50 mg/L a 15 días y 5) 200 mg/L a 15 días.

La cobertura de las lecturas mapeadas indica cómo fue la dinámica de la transcriptómica. Se detectaron 31.418 transcritos en la muestra Control, 28.792 a 50 mg/L de Aroclor a 5 días (baja concentración / tiempo corto), 31.122 a 200 mg/L de Aroclor a 5 días (alta concentración / tiempo corto), 31.037 a 50 mg/L de Aroclor a 15 días (baja concentración / tiempo largo), y 30.582 a 200 mg/L de Aroclor a 15 días (alta concentración / tiempo largo) (Figura 15).

Figura 15. Cobertura de cada transcrito identificado. Los porcentajes indican la proporción de transcrito cuya secuencia queda cubierta en un determinado porcentaje, que se indica con un color, por las lecturas que se alinean con el transcrito. Cada grafica corresponde a una librería. 1) Control; 2) 50 mg/L a 5 días; 3) 200 mg/L a 5 días; 4) 50 mg/L a 15 días y 5) 200 mg/L a 15 días.

4.1.3. Pre-procesamiento de secuencias

Un paso previo muy importante para el análisis de datos es la limpieza de las secuencias, que pasan de lecturas crudas a lecturas limpias. Este paso se conoce como pre-procesamiento. En él se lograron remover artefactos de los datos obtenidos mediante RNA-seq, que ayudaron a mejorar la calidad de las lecturas y con ello la eficiencia del ensamblaje. El 97% de las lecturas crudas pueden ser consideradas lecturas limpias según el sistema de pre-procesamiento utilizado por BGI (ver apartado 3.8.2 de Material y Métodos). Sin embargo, con el programa SeqTrimNext los porcentajes son menores, debido a que el filtrado es más estricto. Nuestros resultados con este programa sugieren un porcentaje promedio de lecturas limpias del 83% y un 13 % de lecturas rechazadas (Figura 16).

Figura 16. Resumen de las lecturas generadas por dos programas bioinformáticos para el pre-procesamiento de las secuencias obtenidas para las 5 librerías. 1) Control; 2) 50 mg/L a 5 días; 3) 200 mg/L a 5 días; 4) 50 mg/L a 15 días y 5) 200 mg/L a 15 días.

4.1.4. Mapeo de lecturas en el transcriptoma y genoma de referencia

Se obtuvieron 1.785.315.539 pb totales (36.435.011 lecturas). Se mapearon un promedio de 585.570.000 nucleótidos por muestra. La secuenciación generó 11.95 millones de lecturas mínimas y 12.61 millones de lecturas máximas (Tablas 6 y 7). Más del 60% de las lecturas pudieron ser mapeadas sobre el genoma. En concreto, un mínimo del 61,8% y un máximo del 78,78% (Tabla 6).

Tabla 6. Resumen estadístico de genes alineados provenientes de las 5 librerías, teniendo como referencia el genoma completo de *P.trichocarpa*.

Tratamiento	Total de lecturas mapeadas	Total de lecturas no mapeadas
Control	9.360.532 (78,78%)	2.520.702 (21,22%)
50 mg/L/5d	7.207.791 (61,85%)	4.445.027 (38,15%)
200 mg/L/5d	9.120.351 (78,09%)	2.559.146 (21,91%)
50 mg/L/15d	9.445.171 (74,91%)	3.164.100 (25,09%)
200 mg/L/15d	9.074.427 (75,97%)	2.870.079 (24,03%)

En el caso del mapeo sobre el transcriptoma, más del 60% de las lecturas pudieron también ser mapeadas: un mínimo del 62,1% de lecturas y un máximo del 82,5% (Tabla 7).

Tabla 7. Resumen estadístico de genes alineados provenientes de las 5 librerías, teniendo como referencia el transcriptoma.

Tratamiento	Total de lecturas mapeadas	Total de lecturas no mapeadas
Control	8.744.029 (82,21%)	1.892.305 (17,79%)
50 mg/L/5d	7.235.826 (62,10%)	4.416.992 (37,90%)
200 mg/L/5d	8.555.081 (82,56%)	1.804.573 (17,44%)
50 mg/L/15d	8.744.029 (82,20%)	1.892.305 (17,79%)
200 mg/L/15d	8.442.127 (79,61%)	2.831.056 (23,70%)

4.1.5. Cuantificación de la expresión génica

Se seleccionaron para los análisis subsecuentes los transcritos que presentaban una abundancia relativa mayor de 2 con respecto al control. Las comparaciones se hicieron dos a dos (control vs. tratamiento). Para evaluar el efecto global del contaminante Aroclor se comparó el control contra cada uno de los cuatro tratamientos. La comparación en donde se obtuvo menor cantidad de transcritos fue con las muestras tomadas a 5 días.

A. Genóma de referencia:

La cuantificación de la expresión génica en BGI se llevó a cabo con el método RPKM. Se evaluaron los niveles de expresión diferencial entre las distintas condiciones analizadas. El proceso de filtrado generó una buena cantidad de genes expresados y no expresados o reprimidos. Las comparaciones mostraron 144 genes sobre-expresados y 40 reprimidos (Control vs. 50 mg/L a s 5 d); 189 sobre-expresados y 62 reprimidos (Control vs. 200 mg/L a 5 d); 300 sobre-expresados y 288 reprimidos (Control vs. 50 mg/L a 15 d), 715 sobre-expresados y 579 reprimidos (Control vs. 200 mg/L a 15 d) (Figura 17A).

B. Transcriptoma de referencia:

La cuantificación de la expresión génica también se realizó en el SCBI, utilizando el método TMM (Trimmed Mean of M-values). Este método estima la escala de factores entre muestras que pueden ser incorporadas en el método estadístico usado para determinar la expresión diferencial. Este parámetro estima los niveles de producción de RNA relativo a los datos de RNA-seq. Se evaluaron los niveles de expresión diferencial entre las diferentes condiciones analizadas. El proceso de filtrado generó una buena cantidad de genes expresados y no expresados o reprimidos. Las comparaciones realizadas mostraron 87 genes sobre-expresados y 517 reprimidos (Control vs. 50 mg/L a s 5 d); 737 sobre-expresados y 1001 reprimidos (Control vs. 200 mg/L a 5 d); 1319 sobre-expresados y 1683 reprimidos (Control vs. 50 mg/L a 15 d), 3366 sobre-expresados y 3352 reprimidos (Control vs. 200 mg/L a 15 d) (Figura 17B).

Figura 17. Transcritos que se acumulan de forma diferencial en las comparaciones llevadas a cabo. A) Número de genes sobre-expresados y reprimidos con referencia al genoma de *Populus*. B) Número de genes sobre-expresados y reprimidos con referencia al transcriptoma de *Populus*. La distribución de genes de las 4 comparaciones: 1 vs. 2= (Control vs. 50 mg/L a s 5 d) ; 1 vs. 3= (Control vs. 200 mg/L a 5 d) ; 1 vs. 4= (Control vs. 50 mg/L a 15 d) ; 1 vs. 5= (Control vs. 200 mg/L a 15 d). Se utilizó un valor de FDR de 0,001 y un log ratio ≥ 1 . Colores : rojo, genes sobre-expresados ; verde, genes reprimidos.

4.1.5.1. Genes que interaccionan en los cuatro tratamientos

Tras la obtención de los genes diferencialmente expresados, se obtuvieron diversos listados de genes. Así es que, una buena forma de seleccionar aquellos genes que participan cuando se ha provocado un proceso a nivel experimental es prestando atención a aquellos genes que participan en la respuesta a los contaminantes. En el análisis de BGI se obtuvieron 22 genes que se encuentran diferenciados positiva y negativamente. De estos genes, solo 16 se agruparon

diferencialmente en los heat maps (Figura 18A y B). También se muestran en un diagrama de Venn los 22 genes inducidos por las cuatro condiciones experimentales.

Figura 18. Patrón de expresión mostrado como “heat map” A) Agrupación de 16 genes diferencialmente expresados en un heat map de 22 genes que interaccionan en las cuatro condiciones experimentales. En rojo, los sobreexpresados y en verde, los reprimidos. Al lado derecho del heat map se incluye el nombre del gen para valores p menores que 0.01. B) Diagrama de Venn mostrando la interacción de las cuatro condiciones experimentales.

En el caso de los genes que se expresan diferencialmente, tomando como referencia el transcriptoma analizado con el programa DEGen Hunter, nos generó una lista de genes mayor que en análisis anterior, esto significa que la cantidad de genes que interaccionan en los cuatro tratamientos es diferente según el algoritmo utilizado. En la comparación 1 vs. 2 interactúan 66 genes, en la comparación 1 vs. 3 interactúan 137 genes, en la comparación 1 vs. 4 interactúan 403 genes y en la comparación 1 vs. 5 interactúan 1.106 genes (Figura 19).

Figura 19. Diagrama de Venn de todas las comparaciones. En el centro se muestra la interacción de los DEGs que son iguales en los cuatro paquetes del programa DEGen Hunter (limma, DESeq2, NOISeq y edgeR), 1 vs.2 tiene 66 interacciones. 1 vs.2 137 interacciones, 1 vs.3 403 interacciones y 1 vs.5 1.106 interacciones.

Figura 20. Scatter plot de la distribución de transcritos que se acumulan de forma diferencial en alguna de las comparaciones llevadas a cabo. 1 vs. 2= (Control vs. 50 mg/L a s 5 d); 1 vs. 3 = (Control vs. 200 mg/L a 5 d); 1 vs. 4= (Control vs. 50 mg/L a 15 d); 1 vs. 5= (Control vs. 200 mg/L a 15 d).

Los resultados obtenidos con el programa DEGene Hunter se representaron como gráficas de tipo MA. Se muestran los niveles de cambio (fold change) de los transcritos de RNA en función de la abundancia de los mismos para los cuatro paquetes estadísticos (DESeq2, EdgeR, Limma y NOISeq) (Figura 21A y B). Se realizó además un perfil de expresión normalizado por Z-Score, donde se puede observar la tendencia de los genes control con respecto a los tratamientos (Figura 22).

A)

B)

Figura 21A. MA-plots mostrando el cambio (fold change) en abundancia de los transcritos. Los DEGs ($p < 0,05$) se muestran en rojo. Los genes sobre-expresados se muestran en el eje $y > 0$; los genes reprimidos en el eje $y < 0$. La abundancia alta de los transcritos aparece a la derecha del eje x y la baja al lado izquierdo. Se utilizaron cuatro paquetes del programa: A) DESeq2, B) EdgeR, C) Limma y D) NOISeq.

Figura 21B. MA-plots mostrando el cambio (fold change) en abundancia de los transcritos. Los DEGs ($p < 0,05$) se muestran en rojo. Los genes sobre-expresados se muestran en el eje $y > 0$; los genes reprimidos en el eje $y < 0$. La abundancia alta de los transcritos aparece a la derecha del eje x y la baja al lado izquierdo. Se utilizaron cuatro paquetes del programa: A) DESeq2, B) EdgeR, C) Limma y D) NOISeq.

Figura 22. Perfiles de expresión normalizados por Z- score. Grupo 1: genes que se inhiben. Grupo 2: genes que no presentan cambios y Grupo 3: genes que se sobreexpresan. En el eje de las X del número 1 al 3 se representan las tres réplicas de las muestras control y del 4 al 6 se representa el correspondiente tratamiento (50 y 200 mg/L).

4.1.5.2. Análisis secuencial (*Time course*)

Este análisis se realiza sobre los datos de genes normalizados y se agruparon en base a su perfil de comportamiento con el programa maSIGpro, que usa un análisis de regresión lineal para modelar la expresión génica.

Se realizó un análisis secuencial para todos los tratamientos (1: concentración baja y 2: concentración alta) a tiempo temprano y tardío (5 y 15 días). Se obtuvo así una clasificación de los genes según 4 patrones de expresión. No todos los genes están incluidos en los cuatro grupos, ya que este programa clasifica los genes más representativos en cada caso. Generamos en concreto 4 grupos o clusters, que muestran 4 comportamientos distintos (Figura 23):

En el **grupo 1** se incluyen 595 genes en el tratamiento 2 con un valor promedio de expresión génica que aumenta más de 100 veces a los 5 días y 78 veces de expresión en el tratamiento 1. Luego decaen a los 15 días.

En el **grupo 2**, se incluyen 3.125 genes que se encuentran reprimidos en ambos tratamientos.

En el **grupo 3**, se agrupan 2.229 cuya expresión aumenta con el tiempo. En este grupo se encuentran por tanto los genes más expresados en ambos tiempos y tratamientos. Destaca el aumento considerable de expresión en el tratamiento 2.

En el **grupo 4**, se incluyen 2.054 genes cuya expresión aumenta inicialmente respecto al control pero luego disminuye. En el caso del tratamiento 2 pasan de 85 a 75, y en el tratamiento 1 de 55 a 50.

Figura 23. Patrones de expresión detectados por el programa maSIGpro. El color rojo representa el tratamiento 1 (50 mg/L de Aroclor) y el verde el tratamiento 2 (200 mg/L de Aroclor).

La clasificación funcional de los transcritos que presentaron un patrón de regulación positivo, ubicados en el grupo 3, nos ayudó a entender el conjunto de genes implicados en la respuesta a PCBs. Las funciones moleculares más significativas aparecen en color rojo. Le siguen el color naranja y el amarillo. Las más interesantes para nuestros objetivos serían: actividad catalítica, actividad oxido-reductasa, actividad hidrolasa, actividad fosfatasa, actividad ATPasa, unión a iones, transporte y función de regulación transcripcional. Así, cuando comparamos el efecto de la concentración y el tiempo de tratamiento con su correspondiente control, vemos que los procesos catabólicos, el transporte y óxido-reducción son las anotaciones más significativas. Las categorías que están más sobre-expresadas se observan en la Figura 24.

A su vez se generaron heat maps representando los niveles de expresión génica clasificada por sus categorías funcionales más expresadas. Fueron elegidas las

categorías más relevantes del análisis de RNA-seq. Dicha relevancia se debía a su alta expresión y posible candidatura en la ruta de degradación de PCBs. Los genes comprendidos en estas categorías están listados y organizados de acuerdo al tratamiento y valor de inducción. Las categorías de la anotación correspondiente a función molecular son: transporte, compuestos orgánicos cíclicos, actividad monooxigenasa y actividad catalítica. El heat map incluye el nombre y función del gen. Las escalas se encuentran al costado izquierdo de cada mapa. El color depende del nivel de expresión y sigue una escala bi o tricolor, verde (baja expresión) (datos no mostrados), negro (expresión media) y rojo (alta expresión) (Figura 25 A y B).

En el caso de la actividad de transporte demuestra que el tratamiento donde se realizó de manera notable una sobreexpresión mayor que el resto de los tratamientos es el tratamiento a 200 mg/L 15 días (1 vs 5), donde se hacen notar transportadores como ABC_MDR, transportadores de aminoácidos, transportadores de cationes y canales de cloro.

En el caso de la actividad de orgánicos cíclicos la participación del citocromo P450 (Potri.008G099100.1), del factor de transcripción bZIP (Potri.008G106700.1), y varios factores de transcripción que mantienen una expresión mayor a los 15 días.

La actividad monooxigenasa está muy bien representada en los cuatro tratamientos. Sin embargo, la actividad funcional más representada y con un valor de expresión mayor es la actividad catalítica. En esta categoría destacan las actividades glutatión transferasa, short chain deshidrogenasa, UDP glucosil tansferasa, peroxidasa, quitinasa y tioredoxina, entre otras.

Figura 24. Términos GO de función molecular para los genes expresados diferencialmente usando un modelo de distribución y una corrección de Bonferroni (p-value < 0,05). Un E-value $\leq 1E-05$ fue considerado como significativo. Control vs. 200 mg/L a 15 días.

Figura 25A. Agrupamiento de los genes expresados diferencialmente en los cuatro tratamientos según su clasificación GO: Funciones moleculares: A) Transporte, B) Compuestos orgánicos cíclicos, C) Monooxigenasa y D) Actividad catalítica. Heat map incluye el nombre y la función del gen. Las escalas se encuentran en la parte superior de cada mapa. El color depende del nivel de expresión y sigue una escala bi o tricolor : verde (baja expresión) (datos no mostrados), negro (expresión media) y rojo (alta expresión).

Figura 25 B. Agrupamiento de los genes expresados diferencialmente en los cuatro tratamientos según su clasificación GO: Funciones moleculares: A) Transporte, B) Compuestos orgánicos cíclicos, C) Monooxygenasa y D) Actividad catalítica. Heat map incluye el nombre y la función del gen. Las escalas se encuentran en la parte superior de cada mapa. El color depende del nivel de expresión y sigue una escala bi o tricolor : verde (baja expresión) (datos no mostrados), negro (expresión media) y rojo (alta expresión).

La interpretación biológica de los genes expresados diferencialmente se completó aún más mediante el análisis de las rutas metabólicas KEGG (Figura 26). Se muestran las 11 anotaciones más representativas: biosíntesis de flavonoides, estilbenoides, diarilheptanoides y gingerol; benzoxazinoides; flavonas y flavonoles; transportadores ABC; antocianinas; carotenoides; isoflavonoides, tirosina, ubiquinona y otros terpenoides.

Figura 26. Análisis KEGG de enriquecimiento de rutas metabólicas de los transcritos expresados diferencialmente. Se representan once categorías con un valor de $p < 0,05$. Se representan dos comparaciones Control vs. 50 mg/L y Control vs. 200 mg/L a los 15 días.

4.2. Genes relacionados con el metabolismo de degradación de PCBs

La tabla 8 muestra los genes candidatos a participar en la respuesta a Aroclor en plantas de *Populus*. Solo se consideraron las expresiones que presentaban una diferencia de 2 veces con la referencia. (Valor absoluto de Relación entre los valores log2 de las muestras para comparar fueron superiores a 1).

Tabla 8. Lista de posibles genes que podrían participar en el metabolismo degradativo de Aroclor. Se proponen tres fases de desintoxicación según los datos de RNA- seq (Log2 ratio).

Nombre del gen	Nombre del Locus	Orthologo Arabidopsis	log2 (200_15d/Co)	FDR
Fase I (Transformación: Reacciones de Oxidación, hidrólisis o reducción)				
Polifenol oxidasa	Potri.001G388900.1	--	2,45	2,52E-295
Glioxal oxidase	Potri.005G235500.1	At3G57620	4,62	1,07E-05
Tioredoxina	Potri.001G416500.1	--	3,62	1,33E-14
	Potri.010G059400.2	--	4,15	0,00070933
1-aminociclopropano-1-carboxilato	Potri.002G163700.1	At3G61510	5,98	1,61E-16
Cinamil-alcohol deshidrogenasa	Potri.009G063100.1	At3G03980	2,41	5,00E-28
Short-chain deshidrogenasa	Potri.013G059100.2	At3g03980	1,51	1,36E-05
D-Lactato deshidrogenasa E	Potri.001G462200.1	At1g30700	4,19	2,92E-32
Inosina-5-Monofosfato deshidrogenasa	Potri.009G099200.1	--	1,65	0,00031137
Xantoxina deshidrogenasa	Potri.004G199900.1	--	4,37	6,49E-138
Naringenin/Flavanona 3-dioxigenasa	Potri.011G150100.1	--	2,94	2,43E-06
2,4-dihidroxi-1,4-benzoxazin-3-1-glucosido Dioxigenasa	Potri.010G107500.1	--	1,26	3,32E-91
Flavin monooxigenasa	Potri.002G254200.1	At4g28720	2,08	8,10E-05
4,5-DOPA dioxigenasa extradiol (ring opening)	Potri.004G135300.1	--	2,06	4,52E-228
Citocromo P450 monooxigenasa CYP82G1	Potri.004G106600.1	At3G25180	2,78	2,73E-57
	Potri.003G173500.1	--	3,08	2,68E-07
	Potri.008G099100.1	--	4,24	2,90E-196
Ubiquitina ligasa E3 Zinc finger C3HC4	Potri.011G063100.1	--	2,48	0,00010991
	Potri.004G042700.1	--	2,62	2,98E-22
Nitrato reductasa / NADH-citocromo B5 reductasa	Potri.002G088600.1	At1g77760	2,93	0
Glutation reductasa	Potri.003G178200.1	At3g24170	1,34	3,67E-16
Isocitrato liasa	Potri.007G122900.1	At3g21720	11,41	8,42E-49
Proteín quinasa activada por AMP	Potri.011G122000.1	At4g27460	4,22	3,04E-08
Quitinasa	Potri.012G033900.1	--	3,49	1,49E-119
	Potri.006G188400.1	--	4,51	3,34E-176
Zeta-caroteno desaturasa	Potri.005G177700.1	--	2,03	2,46E-27
	Potri.003G213700.1	--	3,62	4,92E-232
Poteina Cold Shock	Potri.004G172600.1	At4G38680	1,28	0,000121511
Fase II (Conjugación) Desactivación por la formación de enlaces covalentes con moléculas hidrófilas endógenas tales como glucosa, malonato, glutatión, o ácidos carboxílicos (glucosil/glutation transferasa)				
Glucosil transferasa	Potri.010G075400.1	At2g38870	10,25	0,00053911
UDP-glucosa:(indol-3-yl) acetato beta-D-Glucosiltransferasa	Potri.002G236500.1	--	2,30	1,52E-42
	Potri.014G146000.1	--	3,27	8,51E-298
UDP-glucosil transferasa subfamilia 73C	Potri.001G303000.1	--	2,19	1,30E-08
Homocisteina S-metiltransferasa	Potri.008G155900.1	At3G22740	3,68	1,53E-20
Omega-hidroxipalmitato O-feruloil transferasa	Potri.008G180400.1	--	2,15	6,71E-60
	Potri.010G186300.1	--	2,08	8,09E-05
UDP-glucotrasferasa	Potri.006G055600.1	--	4,54	2,87E-241
Glutation transferasa GST25	Potri.012G050100.1	--	9,39	0,00015295
	Potri.010G060900.2	--	3,02	3,00E-12
Xiloglucano:xiloglucosil transferasa	Potri.018G095200.1	--	3,22	1,00E-06
	Potri.005G007200.1	--	4,68	5,84E-06
2-amino-2-carboxietil transferasa	Potri.005G048400.1	--	2,11	5,16E-05
1 amino ciclo propano carboxilato	Potri.002G163700.1	At3G61510	5,97	3,07E-14

Fase III				
(Compartimentación) Exportación de los conjugados , ya sea a la vacuola al apoplasto (a través de transportadores)				
Canal de Cloro 7	Potri.018G124100.1	At5g33280	1,04	0,00078673
Transportador MFS, Familia OPA	Potri.018G115000.1	At2g13100	1,14	2,92E-05
Transportador MFS, Familia SP	Potri.010G089800.1	--	1,82	1,49E-107
Transportador MDR (Multidrug resistance) Familia ABC	Potri.012G033400.1	--	1,45	0,00033333
	Potri.008G003300.1	At3g54540	1,09	2,76E-28
Ácido biliar:Na+ symporter, Familia BASS	Potri.003G054500.1	--	8,25	0,00053958
Factores de Transcripción				
Factor de transcripción dedos de zinc del tipo C2H2	Potri.008G051200.1	At2g28710	3,21	2,90E-08
	Potri.010G209400.1	--	3,76	8,41E-28
Factor de transcripción Homeobox-leucine zipper	Potri.012G023700.1	--	2,04	1,23E-50
	Potri.014G103000.1	At2g46680	3,24	7,08E-43
Factor de transcripción NAC	Potri.001G404100.1	At3g15500	2,30	4,26E-58
	Potri.013G054200.1	--	3,56	6,46E-180
Factor de transcripción TGA, STIP1.	Potri.006G107600.1	At5g01830	2,79	5,18E-40
Bzip factor de transcripción	Potri.009G164300.1	At1G08320	2,01	0.0115341
Zinc Finger , PMZ, SAP12	Potri.011G138500.1	At3g28210	1,58	1,83E+00
Factor de transcripción WRKY 75	Potri.015G099200.1	At5G13080	1,54	0.00365706
	Potri.003G169100.1	At5G13080	1,20	4,83E-04
	Potri.T043800.1	At5G13080	2,55	8,47E-10
Basic helix-loop-helix BHLH071	Potri.001G141100.1	At5G46690	2,17	0.0285048
Factor de transcripción GRAS2, SCL14	Potri.009G032800.1	At1G07530	1,32	0.00838648
Factor de transcripción MYB3	Potri.013G109300.1	At1G22640	1,50	3,64E-119
Factor de transcripción MYB Proto oncogen	Potri.019G081500.1	AT1G22640	1,38	1,41E-10
	Potri.008G122100.1	--	3,66	2,41E-58

Para la validación de los datos de RNA seq se seleccionaron 15 genes, teniendo como criterio a elegir las más sobre-expresados y que se consideraran posibles candidatos de mecanismos de degradación. La expresión diferencial fue comprobada mediante una validación por medio de qRT-PCR. Los resultados finales se reportan como cambio de expresión (fold change) (Figura 27). Para esta validación diseñamos previamente oligos específicos para cada gen en estudio (Tabla 9).

Tabla 9. Oligonucleotidos diseñados para la validación

Símbolo	AGI Gene ID	Tamaño del oligo (pb)	Sentido (Forward)	Antisentido (Reverse)
18s	---	18 subunit	5'-GTAACCCGTTGAACCCATT-3'	5'-CCATCCAATCGGTAGTAGCG-3'
IMPDH	Potri.009G099200.1	126	5'-CCCAATCCTCTAGTCGTTTA-3'	5'-CCCAGAGTAATGATTCCAACC-3'
FMO	Potri.002G254200.1	126	5'-CGGTCTCTATGCTGTTGG-3'	5'-GTTGCCCTCTTGGTTTGC-3'
PPO	Potri.001g388900.1	130	5'-CCCAATCCTCTAGTCGTTTA-3'	5'-GTTGCCCTCTTGGTTTGC-3'
XANT	Potri.004G199900.1	135	5'-GCAGGAATAGTAGGCACGC-3'	5'-CCGAACAGGGATCATTACGC-3'
DHF	Potri.009G155800.1	105	5'-CCTGGTAGTAATGGGAGGTA-3'	5'-CCTTCATACTCAACATACTGG-3'
CAD	Potri.009G063100.1	149	5'-CGATGGATTATGTGAACACAG-3'	5'-ACGAATCTCACTTTACATGCG-3';
NAR	Potri.011G150100.1	156	5'-CCGAGCAGTTGTGAACAGT-3'	5'-AACTTCTCGTAGTATTACCG-3'
DIOXI	Potri.004G135300.1	132	5'-GGGTGCAGCAGGTGAAGAT-3'	5'-CTCCAGCAGAACTTGTAACG-3'
P450	Potri.004G106600.1	111	5'-GCTTGCCTGTGGACATGATT-3'	5'-ACACTGATACCTAGTAGAGC-3'
UBIE3	Potri.011G063100.1	120	5'-GGGAGAGATCATCAAGGAC-3'	5'-ATAATGATACAGAACGAGTGG-3'
CHITI	Potri.006G188400.1	148	5'-CCTATGCCAAGAACAGAGGA-3'	5'-ACAGATACTGATACCTATATGC-3'
WRKY	Potri.001g328000.1	146	5'-GGTATAGATGGAGGAAGTAT-3'	5'-TAAGTCGCTACTACAACACC-3'
UDP	Potri.010G060900.1	94	5'-GGAGAACGAGGAAATGAGATG-3'	5'-GATTATTGTAAGAAGAGCCACC-3'
ABC	Potri.012G033400.1	142	5'-GCAGGGTATTGGTGCTTGAT-3'	5'-CAAGTACCATTGAGCAATCC-3'
C ₂ H ₂	Potri.010G209400.1	255	5'-CCCTCCTAGCACTGATGAT-3'	5'-CAAGATGTATAGCGAGTGTTC-3'
CSP4	Potri.004G172600.1	180	5'-GAAAACGGCATGGGTGAGAGG-3'	5'-CCTCACACTAAAATCAACGTCC-3'

Los valores de correlación entre la información obtenida con RNA-seq y las mediciones experimentales mediante qRT-PCR se encuentran entre 0,865 a 0,992 (Tabla 1010). Esto ratifica la validez de nuestro experimento de RNA-seq para detectar genes expresados diferencialmente.

Figura 27. Análisis empírico de expresión diferencial mediante qRT-PCR. Se analizaron 15 genes identificados previamente mediante RNA-seq. Los cambios en la expresión génica se estimaron respecto a un control (fold change). Los genes validados se indican con un acrónimo del gen. Las barras indican el error estándar del fold change promedio. Inosina-5-monofosfato deshidrogenasa (IMPDH); flavin monooxigenasa (FMO); polifenol oxidasa (PPO) Potri.001G388900.1; xantoxina deshidrogenasa (XAN) Potri.004G199900.1; 2-hidroxi isoflavanona deshidratasa (DHF) Potri.009G155800.1, cinamil-alcohol deshidrogenasa (CAD) Potri.009G063100.1, naringenina 3-dioxigenasa (NAR) Potri.011G150100.1, 4,5-DOPA dioxigenasa extradiol (DIOXI) Potri.004G135300.1, citocromo P450 monooxigenasa (P450) Potri.004G106600.1; cold shock (CSP4) Potri.004G172600.1; quitinasa (CHITI) Potri.006G188400.1; WRKY factor de transcripción Potri.001G328000.1; UDP-glucosiltransferasa Potri.006G055600.1; transportador MDR (multidrug resistance), transportador ABC (ABC) Potri.012G033400.1 y factor de transcripción C2H2 Potri.010G209400.1.

Tabla 10. Correlación entre valores de expresión obtenidos con RNA-seq y con qRT-PCR. La primera fila es la abreviatura del nombre del transcrito. El coeficiente de correlación entre ambos valores se encuentra en la última fila.

	IMPDH	FMO	PPO	XANT	DHF	CAD	NAR	DIOXI	P450	CHITI	WRKY	UDP	ABC	C2H2	CSP4
Muestra	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR	qPCR
50_5d	2,17	0,29	0,79	5,97	1,91	3,21	1,66	1,33	3,26	1,44	9,25	2,56	0,84	2,30	1,20
200_5d	0,98	0,11	2,24	6,15	2,21	5,55	1,92	1,83	4,66	2,90	8,03	4,27	1,99	3,57	1,89
50_15d	0,91	2,21	1,78	7,50	2,89	2,23	3,42	1,64	3,57	3,73	4,60	2,92	0,86	2,42	1,99
200_15d	3,01	3,34	3,46	10,17	6,19	5,49	4,14	3,39	6,39	7,81	11,93	10,35	2,10	6,47	5,21
Muestra	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq
50_5d	0,77	- 0,46	0,38	2,415	1,15	1,54	1,12	0,26	1,24	1,20	7,52	1,22	0,33	1,12	1,12
200_5d	0,66	- 0,14	1,44	2,829	1,58	2,93	1,50	0,53	2,08	1,74	7,26	2,04	1,39	1,59	1,50
50_15d	0,45	1,93	0,67	3,502	1,62	1,00	2,82	0,48	1,27	1,83	6,17	1,42	0,42	1,13	2,82
200_15d	1,65	2,08	2,45	4,370	2,65	2,41	2,94	2,06	2,78	4,51	9,80	4,54	1,45	3,76	2,94
Correlación	0,91	0,96	0,97	0,98	0,98	0,99	0,98	0,99	0,99	0,984	0,951	0,991	0,92	0,99	0,87

Se realizaron dos tipos de análisis de enriquecimiento biológico funcional a los 90 genes enlistados en la Tabla 8.

A) Análisis Pfam: Este análisis nos permitió observar los dominios proteicos y familias comunes de los genes seleccionados

Los resultados obtenidos para este análisis nos indican las anotaciones funcionales más significativas: UDP-glucoronosil y UDP-glucosil transferasa, Dominio de unión FAD, Isocitrato liasa, Dominio de unión a DNA (WRKY), Familia Glicosil hidrolasas 18, Dominio CBS, Familia Glicosil hidrolasas 16, POT familia, Deshidrogenasa de unión a Zinc, Glutación S-transferasa Dominio C-terminal, Dominios de unión a DNA (Myb), Short chain deshidrogenasas, azúcar y otros transportadores y Dominio Homeobox. La coloración indica el grado de significancia (Tabla 11).

Tabla 11. Análisis de Pfam (familia de proteínas) mediante el programa TopGO. El color rojo/naranja (alta significancia) a amarillo (baja significancia). En la columna de estadísticas se indica la relación de genes encontrados / la relación de genes totales asociados a esa familia de proteínas (PFAM).

Pval (corregido)	Estadísticas	Descripción	PFAM id
6,12E-01	6/86 224/29014	UDP-glucoronosil y UDP-glucosil transferasa	PF00201
1,02E+00	4/86 88/29014	Dominio de unión FAD	PF01565
1,14E+00	2/86 4/29014	Isocitrato liasa	PF00463
1,35E+00	4/86 102/29014	Dominio de unión a DNA (WRKY)	PF03106
8,49E+00	2/86 22/29014	Familia Glicosil hidrolasa 18	PF00704
2,20E+01	2/86 39/29014	Dominio CBS	PF00571
2,28E+01	2/86 43/29014	Familia Glicosil hidrolasa 16	PF00722
5,29E+01	2/86 85/29014	Familia POT	PF00854
5,57E+01	2/86 83/29014	Deshidrogenasa de unión a Zinc	PF00107
5,78E+01	2/86 80/29014	Glutación S-transferasa, Dominio C-terminal	PF00043
5,87E+01	4/86 355/29014	Dominios de unión a DNA (Myb)	PF00249
8,38E+01	2/86 120/29014	Short chain deshidrogenasas	PF00106
8,54E+01	2/86 116/29014	Azúcar y otros transportadores	PF00083
8,85E+01	2/86 129/29014	Dominio Homeobox	PF00046

C) Análisis de enriquecimiento biológico (función molecular): Este análisis nos permite obtener las anotaciones de ontología génica (GO)

Los resultados de este análisis nos mostraron las siguientes anotaciones: actividad isocitrato liasa, actividad transferasa, transferencia de grupos hexosil, actividad oxo-acido-liasa, FAD, actividad oxidoreductasa, actividad UDP-N-acetilmuramato deshidrogenasa, actividad transferasa, transferencia de grupos glicosil, unión a cofactor, unión a ácidos nucleicos, actividad factor de transcripción y unión a una secuencia específica de DNA con actividad de factor de transcripción (Tabla 12).

Tabla 12. Análisis de enriquecimiento biológico por el programa TopGO. Los términos son anotación GO (ontología génica). El color rojo/naranja (alta significancia) a amarillo (baja significancia). En la columna de estadísticas se indica la relación de genes encontrados / la relación de genes totales con anotación GO (Ontología génica).

Pval (corregido)	Estadísticas	Descripción	GO id
0.000232264698755	2/78 2/19622	Actividad isocitrato liasa	GO:0004451
0.00202861652037	9/78 528/19622	Actividad transferasa, transferencia de grupos hexosil	GO:0016758
0.00206888216204	2/78 6/19622	Actividad oxo-acid-liase	GO:0016833
0.0021066237305	6/78 213/19622	Flavin adenina unión a di nucleótido	GO:0050660
0.0021293516543	21/78 2275/19622	Actividad oxidoreductasa	GO:0016491
0.00344188908023	4/78 89/19622	Actividad UDP-N-acetilmuramate deshidrogenasa	GO:0008762
0.00753744947179	9/78 644/19622	Actividad transferasa, transferencia de grupos glicosil	GO:0016757
0.0146806339821	9/78 715/19622	Unión a cofactor	GO:0048037
0.0189701530851	9/78 754/19622	Unión a ácidos nucleicos actividad factor de transcripción	GO:0001071
0.0189701530851	9/78 754/19622	Unión a una secuencia -específica de DNA con actividad de factor de transcripción	GO:0003700

A partir de los 90 genes de la Tabla 8 se realizó un análisis de co-expresión genética. El análisis reveló 162 co-expresiones entre genes y factores de transcripción (Figura 28).

Figura 28. Visualización de la co-expresión entre genes con el programa exNet (expression network). Se muestran 162 correlaciones en forma de nodos y líneas conectadas. Los genes están indicados como círculos verdes y los factores de transcripción como cuadrados amarillos.

4.3. Fraccionamiento unidimensional de proteicos

Análisis previos realizados en nuestro laboratorio reportaron la posible implicación de la proteína CSP en respuesta a tratamientos con contaminantes orgánicos (Campos, 2011). En el análisis transcriptómico realizado durante el desarrollo de esta tesis (apartado 4.1.2 de Resultados), observamos en efecto su inducción diferencial a nivel de transcrito. Con el fin de comprobar dicha inducción a nivel proteico se analizaron las modificaciones en el proteoma mediante técnicas de electroforesis bi-dimensional tal como se describe en el apartado 3.9 de Material y Métodos.

Los tratamientos evaluados fueron los mismos utilizados para el análisis transcriptómico mediante RNA-seq. Todas las muestras se analizaron inicialmente mediante electroforesis en geles unidimensionales SDS-PAGE, para corroborar la concentración de proteína y la calidad de los extractos (Figura 29)

Figura 29. Análisis mediante electroforesis unidimensional SDS-PAGE. Se fraccionó la misma cantidad de proteína en todas las muestras (15 μ g) muestras de chopo : 1) Control; 2) 50 mg/L 5 días ; 3) 200 mg/L 5 días ; 4) 50 mg/L 15 días y 200 mg/L 15 días). Las masas moleculares de proteínas estandar (M) están indicadas a la izquierda.

4.3.1. Detección de la proteína PCSP4 en mapas bidimensionales

Se obtuvieron geles bidimensionales con tres repeticiones por tratamiento con la misma cantidad de proteína. Se realizaron las electroforesis IEF x SDS-PAGE según se describe en el apartado 3.5 de Material y Métodos.

Una vez obtenidos los geles, se realizó la identificación de posibles proteínas CSP buscando *spots* que aparecieran alrededor de su pI teórico (6,25) y peso molecular teórico (19 kDa). Se localizó solo un spot que cumplía con esa ubicación y que además no se encontraba en los controles (Figura 30). Aparte de esta proteína, se encontraron alrededor de 20 spots diferenciales, si bien solo nos enfocamos en la posible proteína CSP.

Figura 30. Análisis del proteoma de chocho mediante fraccionamiento IEF x SDS-PAGE. Los geles corresponden a proteínas de chocho tras un tratamiento con Aroclor 1221 teñidas con Coomassie Brilliant Blue R-250. El intervalo de pI fué de 3-10.

Después de ser detectado el *spot* de interés, se recortó y sometió a digestión con tripsina *in situ*. Los péptidos resultantes se analizaron mediante espectrometría MS. Una vez obtenida la huella peptídica y los datos espectrométricos MALDI-TOF/TOF (Figura 31), se realizó una búsqueda de secuencias homólogas en bases de datos. Para ello, se utilizaron herramientas del ExPASy Proteomics Server del Swiss Institute of Bioinformatics (www.expasy.ch). Se empleó la base de datos PHYTOZOME v10.3 (www.phytozome.net), entre otras. La comparación de nuestras huellas peptídicas con el genoma traducido de *Populus trichocarpa* nos permitió identificar inequívocamente la proteína PtCSP4 (Potri.004G172600).

A

Inicio – Fin	Masa (obs)	Mr(exp)	ppm	P	Secuencia de péptidos
44 - 62	2066.94	2065.94	14	0	R.SLGDGEEVEFVIENSDDGR.T
44 - 64	2296.10	2295.09	16	1	R.SLGDGEEVEFVIENSDDGRTK.A
44 - 64	2296.10	2295.09	16	1	R.SLGDGEEVEFVIENSDDGRTK.A
63 - 80	1780.00	1779.00	18	1	R.TKAVDVTAPGGNPVQGIR.S
65 - 80	1550.86	1549.85	18	0	K.AVDVTPAGGNPVQGIR.S
65 - 80	1550.86	1549.85	18	0	K.AVDVTPAGGNPVQGIR.S

B

Figura 31. Análisis MS a partir de las electroforesis 2D. El *spot* indicado en la Figura 24 se digirió con tripsina y los péptidos resultantes se identificaron por MALDI TOF/TOF. Se identificó inequívocamente la proteína PtCSP4 (Potri.004G172600). Se indican el tamaño y secuencia de los péptidos obtenidos, la masa esperada y la observada, y el número de cortes fallidos.

La comparación de muestras control y los tratados con Aroclor 1221 a diferentes concentraciones (50 y 200 mg/L) y tiempos de exposición (5 y 15 días) mostró diferencias cuantitativas en la acumulación de PCSP4. Las mayores diferencias se identificaron a los 15 días de tratamiento. La intensidad de las proteínas seleccionadas mostró que el patrón de acumulación es gradual, observándose diferencias a partir de los cinco días de tratamiento a la dosis más baja. En la gráfica se pueden observar los patrones de acumulación de PtCSP4 a los 15 días de exposición al tratamiento (Figura 32A y B).

Figura 32. Cuantificación de la proteína PCSP4 inducida por PCB. Se muestran los detalles de las zonas recuadradas de la Figura 24. A) El gel de la izquierda corresponde al control y los de la derecha al del tratamiento con Aroclor transcurridos 15 días desde su aplicación. Aparece señalada la proteína CSP, que está en el gel control. B) El gráfico de barras muestra la cuantificación relativa de la proteína PCSP4 en plantas tratadas, a utilizando el programa Image Lab 5.0. A la cantidad máxima de proteína detectada se le asignó el valor de 1.

4.3.2. Gen PtCSP4 (Potri.004G172600)

La secuencia completa del gen PtCSP4 (1065 pb) presenta un marco de lectura de 624 pb y codifica una proteína de 207 aminoácidos. La región 5' no traducida comprende 48 pb, mientras que la región 3' no traducida de es 328 pb. Este factor transcripcional tiene un punto isoeléctrico de 6,28 y un peso molecular de 19 kDa. El gen está en el cromosoma 4 de *P. trichocarpa* (código Potri.004G172600.1) de (posición 17.886.390-17.887.454).

El peso molecular teórico es de, 19.113 Da; y el pI 6,28, destaca la presencia de numerosas (aproximadamente 45%), un rasgo típico de esta familia. La proteína tiene un índice de inestabilidad alto, de 56,14. La secuencia de la región que codifica la proteína PCSP4 y la traducción de la misma se muestran en la Figura 3333.

```

Atgggtgagaggcaaaactgggaaggtcaagtgggttagtgaccaaagggttttggtttc 60
M G E R Q T G K V K W F S D Q K G F G F
Ataagcccagatgatggcagcgacgatctttttgtccaccaatcttccatcaaattccgaa 120
I S P D D G S D D L F V H Q S S I K S E
Gggtatcgagtcctcgagatggtgaagaggttgagtttgattgaaaactctgatgat 180
G Y R S L G D G E E V E F V I E N S D D
Ggcgtactaaggctgttgatgtcactgcccctgggtggaacccagttcagggtatcaga 240
G R T K A V D V T A P G G N P V Q G I R
Tccggaggcggctttggcgggtggctctgggggcccggaggaggccggcggtggga 300
S G G G F G G G S G G R G G G R G G G G
Tatggaggtggaggaggatgaggaggaggtggtgggtatggaggtggaggttatggt 360
Y G G G G G Y G G G G G Y G G G G Y G
Ggaggtaggaggggtggtggtggtggcggtacggaggaggcggcggtatggcgaggc 420
G G R R G G G G G G Y G G G G G Y G G G
Ggagggcggtggaggggtgttatagttgtggggagtcggggcatatggctagagattgt 480
G G G G G C Y S C G E S G H M A R D C
Ccacaaggtggtagtggtggtggaggtggcggtggaaggtatggtggtggaacggtggt 540
P Q G G S G G G G G G G R Y G G G N G G
Gggggaggtggaggaggtggaggggtgtataattgtggtgggtctggtcattttgctagg 600
G G G G G G G C Y N C G G S G H F A R
Gactgtcctaatagtggacgttga 624
D C P N S G R -

```

Figura 33. Secuencia de la región codificante del gen PtCSP4 *P. trichocarpa* y la traducción de la misma (624 pares de bases). En la fila superior se muestra la secuencia de nucleótidos y en la inferior los aminoácidos traducidos.

4.3. Identificación de proteínas homólogas en *P. trichocarpa*

Tras un escrutinio exhaustivo del genoma de chopo encontramos que existen seis genes CSP (Tabla 13).

Tabla 13. Listado de las secuencias CSP identificadas en el genoma de chopo, versión 3.0. Se muestran los nombres de las proteínas, el locus y el alias en *Populus* y finalmente la localización cromosómica y orientación.

Nombre de la proteína	(EMBL-EBI)	Locus en <i>Populus trichocarpa</i>	Alias	Localización en el cromosoma y orientación
PtCSP1	B9HR86	Potri.009G132000	POPTR_0009s13460	Chr09 (+) 10709796 - 10711081
PtCSP2	B9GUE2	Potri.002G086900	POPTR_0002s08740	Chr02 (+) 6126306 - 6127650
PtCSP3	B9H7B9	Potri.005G174300	POPTR_0005s15990	Chr05(-) 18991872 - 18992809
PtCSP4	B9H173	Potri.004G172600	POPTR_0004s17840	Chr04 (+) 19135102 - 19136332
PtCSP5	No disponible	Potri.005G112800	POPTR_0005s11495	Chr05 (-) 8650547 - 8651583
PtCSP6	No disponible	Potri.007G057100	POPTR_0007s09635	Chr07 (-) 6038103 - 6039280

Dichas secuencias fueron analizadas una vez traducidas a proteína. Las CSPs se encuentran distribuidas en los cromosomas 2, 4, 5, 7 y 9, (Figura 34) representando solo 5 de los 19 cromosomas del chopo. No hay intrones en ninguna de las secuencias.

Figura 34. Ubicación de las secuencias que codifican CSPs en los cromosomas de chopo. El diagrama fue realizado utilizando el servidor PopGenIE v3 (popgenie.org/). Las CSPs se distribuyen en cinco de los 19 cromosomas de chopo.

El alineamiento de las secuencias aminoacídicas de la familia CSP en chopo se observó en la Figura 35.

Figura 35. Alineamiento múltiple de secuencias aminoacídicas mediante CLUSTAL W (1,82) deducidas para la familia CSP. Los residuos idénticos entre secuencias se indican con (*), las sustituciones conservadoras con (:) y las sustituciones semi-conservadoras con (.) . Los motivos conservados identificados se representan mediante cuadros de colores. La gama de colores y el significado se muestra en la gráfica.

4.4. Alineamiento múltiple de proteínas CSP en plantas

La familia CSP de plantas tiene un dominio muy conservado, conocido como “Cold shock domain” (CSD) y cuenta con dos o más dedos de zinc “Zinc Knuckle domain” del tipo CCHC (CX₂CX₄HX₄C), dos sitios de reconocimiento de RNA (RRM) y cajas RGG repetidas. Se realizó un alineamiento múltiple con el programa Clustal W y el resultado se coloreó con Box shade. Se encontró que las secuencias completas codifican proteínas de 152- 301 aminoácidos (Figura 36). En la Tabla 144, se detallan todos los genes de plantas nombrados como Cold shock domain en el banco de genes NCBI, mencionándose el nombre del locus o ID, el número de aminoácidos, la cantidad de secuencias RGG repetidas y el número de dedos de zinc.

La lista incluye 45 genes de la familia CSP: 1 de *Picea glauca*, 6 de *Populus trichocarpa*, 3 de *Glycine max*, 1 de *Vitis vinifera*, 2 de *Ricinus communis*, 3 de *Physcomitrella patens*, 3 de *Carica papaya*, 2 de *Zea mays*, 4 de *Hordeum vulgare* var. *Distichum*, 4 de *Brachypodium distachyon*, 2 de *Oryza sativa* subsp. *Japónica*, 4 de *Arabidopsis thaliana*, 2 de *Thellungiella salsuginea*, 3 de *Triticum aestivum*, 1 de *Picea sitchensis*, 1 de *Lotus japonicus*, 1 de *Nicotiana glauca*, 1 de *Sorghum bicolor* y 1 de *Cucumis melo* subsp.

Tabla 14. Proteínas CSP. Se muestra el nombre del locus, la longitud de la proteína (en aminoácidos, AA), las repeticiones RGG y los dominios dedos de zinc

Organismo	Locus	AA	CSD	RGG	Dedos de zinc
<i>Picea glauca</i>	PGQ3HRT	156	1	2	2
<i>Populus thrichocarpa</i>	PT13460	168	1	2	2
<i>Glycine max</i>	Glyma06g11560	176	1	4	2
<i>Vitis vinifera</i>	GSVIVT01024055001	177	1	5	2
<i>Populus thrichocarpa</i>	PT08740	184	1	4	2
<i>Ricinus communis</i>	RCOM1031420	184	1	3	2
<i>Physcomitrella patens</i>	PP1S4168V6	187	1	1	2
<i>Carica papaya</i>	CPevmTU3618	188	1	3	2
<i>Zea mays</i>	Grnmzm2g389768	188	1	1	3
<i>Hordeum vulgare</i> var. distichum	HVF2DGP2	191	1	3	2
<i>Hordeum vulgare</i> var. distichum	HVF2CX17	192	1	3	2
<i>Brachypodium distachyon</i>	Bradi1g02650	193	1	2	2
<i>Brachypodium distachyon</i>	Bradi2g07090	193	1	2	2
<i>Populus thrichocarpa</i>	PT15990	193	1	4	2
<i>Glycine max</i>	Glyma04g00660	194	1	4	2
<i>Physcomitrella patens</i>	PP1S10365V6	195	1	1	2
<i>Oryza sativa</i> subsp. japonica	Os08g0129200	197	1	1	2
<i>Physcomitrella patens</i>	PP1S644V6	198	1	3	2
<i>Glycine max</i>	Glyma12g03470	200	1	3	2
<i>Arabidopsis thaliana</i>	At2g21060	201	1	4	2
<i>Thellungiella salsuginea</i>	TSE5CYD4	201	1	5	2
<i>Brachypodium distachyon</i>	Bradi1g02640	203	1	2	2
<i>Arabidopsis thaliana</i>	At4g38680	203	1	5	2
<i>Triticum aestivum</i>	TAQ75QN9	205	1	3	2
<i>Picea sitchensis</i>	PSA9NNT8	205	1	5	2
<i>Populus thrichocarpa</i>	PT17840	207	1	3	2
<i>Lotus japonicus</i>	chr3CM0112180r2d	207	1	5	2
<i>Zea mays</i>	Grnmzm5g895313	208	1	1	2
<i>Carica papaya</i>	CPevmTU37164	208	1	5	2
<i>Nicotiana glauca</i>	NSP27484	213	1	3	2
<i>Triticum aestivum</i>	TAQ8LPA7	228	1	2	2
<i>Hordeum vulgare</i> var. distichum	HVF2DD34	231	1	1	3
<i>Triticum aestivum</i>	TAQ75QN8	231	1	2	3
<i>Populus thrichocarpa</i>	PT11495	235	1	0	5
<i>Carica papaya</i>	CPevmTU6256	238	1	3	5
<i>Oryza sativa</i> subsp. japonica	Os02g0121100	241	1	1	4
<i>Populus thrichocarpa</i>	PT09635	243	1	0	5
<i>Sorghum bicolor</i>	Sb04g001720	251	1	1	4
<i>Brachypodium distachyon</i>	Bradi3g01960	257	1	3	4
<i>Hordeum vulgare</i> var. distichum	HVF2DJM5	264	1	3	4
<i>Ricinus communis</i>	RCOM0431020	266	1	2	5
<i>Cucumis melo</i> subsp.	CME5RDD6	272	1	3	5
<i>Thellungiella salsuginea</i>	TSE5CYD5	295	1	0	6
<i>Arabidopsis thaliana</i>	At4g36020	299	1	0	7
<i>Arabidopsis thaliana</i>	At2g17870	301	1	0	7

Figura 36. Alineamiento múltiple de secuencias aminoacídicas de la familia CSP con CLUSTAL W (1.82). Los residuos idénticos entre secuencias se indican con (*), las sustituciones conservadoras con (:) y las sustituciones semi-conservadoras con (.). Los motivos conservados identificados se representan mediante cuadros de colores. Sombreado en gris claro, dominios cold shock ; en gris oscuro láminas β ; en verde, motivos de unión a RNA ; en rosa, sitios de unión a DNA ; en azul ; dominios “zinc knuckle” ; en rojo, dedos de zinc tipo CCHC; y en amarillo, secuencias repetidas RGG (continúa en la página siguiente).

Figura 36. (Continuación)

4.4.1. Análisis filogenético de la familia CSP en plantas

El árbol filogenético fue realizado a partir de un alineamiento de secuencias de aminoácidos mediante el método de Neighbour joining (NJ) implementado el programa Mega 5.0. Dicho árbol mostró la agrupación de todas las secuencias aminoacídicas de la familia CSP.

La estimación de la distribución de secuencias generó una matriz (datos no mostrados) con una distancia media entre ellas de 0,47 por el modelo de Jones Taylor Thornton (JTT). El bootstrap del árbol consenso nos ayudó a verificar la solidez del muestreo, otorgándonos valores de confianza mayores del 50%. El porcentaje de árboles repetidos en los cuales los taxones aparecen juntos en el test bootstrap se indica en las ramificaciones de la Figura 37. La topología de las 45 accesiones de secuencias proteínicas consideradas para la filogenia genera dos grandes subgrupos, los cuales se subdividen en 3 grupos mayoritarios (Grupos I, II y III).

Figura 37. Dendrograma consenso inferido por el método Neighbor-Joining. Las ramas indican el valor de bootstrap para 500 iteraciones. El número sobre la barra representa el porcentaje de identidad. Los nombres representan el locus: *Populus thrichocarpa*: PT, *Vitis vinifera*: GSVIVT, *Nicotiana sylvestris*: NS, *Triticum aestivum*: TA, *Picea sitchensis*: PS, *Sorghum bicolor*: Sb, *Oryza sativa* subsp. Japonica: Os, *Hordeum vulgare* var. *distichum*: HV, *Arabidopsis thaliana*: At, *Cucumis melo* subsp.: CM, *Carica papaya*: CP, *Picea glauca*: PG, *Ricinus communis*: RC, *Thellungiella salsuginea*: TS, *Zea mays*: Grnzm, *Brachypodium distachyon*: Bradi, *Glycine max*: Glyma, *Lotus japonicus*: chr, *Physcomitrella patens*: PP. El número al lado del locus es el tamaño (en aminoácidos) y el número después del guión son los znf (zinc finger).

Por otro lado, se realizó un diagrama de dominios conservados, tomando como referencia el árbol filogenético anterior (Figura 38).

Figura 38. Dominios conservados agrupados por familias, derivado del mismo alineamiento del dendrograma anterior. Se muestran en color rojo el dominio cold shock, en rojo intenso: znf1, verde azul: znf2, verde limón: znf3, morado: znf4, amarillo: znf5, mostaza: znf6 y rosa: znf7.

4.4.2. Análisis de regiones conservadas

El análisis de los alineamientos reveló la presencia de al menos ocho regiones conservadas o consenso. Para su detección se utilizó el software MEME (multiple EM for motif elicitation), versión 4.9.0 <http://meme.nbcr.net/meme/cgi-bin/meme.cgi>. La altura total de cada grupo de símbolos en indican la conservación de secuencia en esa posición, mientras que la altura de cada uno de los símbolos dentro del motivo indica la frecuencia relativa de cada aminoácido en esa posición. Para las proteínas, las categorías se basan en las propiedades bioquímicas de los aminoácidos, que se muestran en una tabla anexa a la Figura 39.

El dominio Cold Shock es el dominio más conservado de todos. Este dominio comprende las láminas $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ y $\beta 5$ (Figura 39 A). Los elementos RGG se encuentran intercalados después del dominio Cold Shock y antes de los dominios dedos de zinc o entre ellas. La gran cantidad de glicina (se pueden repetir de 1 a 5 veces) (Figura 39 B) explica que a estas proteínas también se les conozca como proteínas ricas en glicina o GRPs. Los dominios Zinc knuckle o dedos de Zinc más abundantes son del tipo CCHC ($CX_2CX_4HX_4C$). Se han detectado seis dominios de este tipo en el alineamiento (Figura 39 C).

Figura 39. Análisis de secuencias con MEME (Motif-based sequence analysis tools). A) Motivo 1 correspondiente al dominio Cold Shock (E- value 3,0e-1534, 97 sitios). B) Motivo 2 correspondiente a elementos RGG (E- value 7,5e-005, 40 sitios). C) Motivos 3 (E- value 1,2e-857), 4, 5, 6, 7 y 8 correspondiente al dominio de dedos de Zinc tipo CCHC. La altura de las letras representa la frecuencia del aminoácido en una posición determinada y la anchura la frecuencia pero sin considerar las deleciones. Se indican también los aminoácidos hidrofóbicos (azul), polares (verde), básicos (rojo) y ácidos (magenta).

4.5. Expresión heteróloga de PCSP4

La secuencia codificante de Pt CSP4 (Potri.004G172600.1) se amplificó por PCR a partir de DNA genómico, se introdujeron sitios *aatb* completos utilizando la enzima de alta fidelidad Long Range DNA Polimerasa. El fragmento obtenido (624 pb) se clonó mediante tecnología Gateway.

Los oligonucleótidos utilizados para el clonaje se diseñaron a partir de las secuencias genómicas, teniendo en cuenta que la temperatura de fusión se sitúa alrededor de 60°C. Para clonar el gen PtCSP se utilizaron oligonucleótidos *forward* y *reverse* los cuales se diseñaron en la zona más divergente del extremo 5' de la región codificante del extremo amino-terminal (Tabla 15).

Tabla 15. Los oligonucleótidos se diseñaron como se ha indicado previamente, incluyendo sitios *attB* en sus extremos. Los sitios *attB* para el clonaje Gateway se muestran sin subrayar. Se indican la T_M , el porcentaje de GC y la longitud en nucleótidos (nt)

Nombre	Secuencia	T_M (°C)	%GC	nt
Fw_Yb	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTT <u>AGAAGG</u> <u>AGAAAACGGCATGGGTGAGAGG</u> 3'	87,2	48,2	58
Rv_Yb	5'GGGGACCACTTTGTACAAGAAAGCTGGGT <u>CCTCACAC</u> <u>TAAATCAACGTCC</u> 3'	84,9	49,0	51

Primero se clonó en el vector de expresión pDONR y posteriormente para que el transgén quedase en bajo el control del promotor 35s se procedió a pasarlo al vector binario PB2GW7 (Figura 40).

Figura 40. Comprobación del clonaje de PtCSP4 en geles de agarosa con enzimas de restricción

Con la construcción resultante, se transformaron células de la cepa C58C1 (pMP90) de *Agrobacterium*, que se utilizó para infiltrar plantas de *Arabidopsis* (Figura 41). Como control se transformaron plantas de *Arabidopsis* sin inserto.

Figura 41. Clonaje de la región codificante de PtCSP4 en el vector pB2GW7. Diseño de clonaje para la expresión ectópica de PtCSP4 en *A. thaliana* bajo el control del promotor CaMV35S. La zona comprendida entre las regiones LB y RB (T-DNA) incluye el gen BAR, que confiere resistencia al herbicida glufosinato de amonio.

4.6. Análisis de expresión diferencial

Se cuantificó la expresión relativa del gen endógeno a los 5, 10, 15 y 20 días y se normalizó con respecto a los valores de expresión en el tratamiento control. Aparentemente la inducción provocó un efecto progresivo conforme la concentración de Aroclor 1221 se incrementaba. No obstante, teniendo en cuenta el error de precisión de esta tecnología se observa que el gen está inducido después del análisis cuantitativo, cuatro veces más con respecto al control, por lo que el incremento es significativo, tomando en cuenta que el disolvente DMSO también muestra un efecto en la expresión pero sin embargo dicha expresión es menor que cuando las plantas tienen Aroclor. El disolvente no parece afectar a la inducción progresiva del gen con respecto al tiempo (Figura 42).

Figura 42. Expresión de PtCSP4 en plantas *in vitro* de chopo. Niveles promedio de transcrito detectado mediante qRT-PCR. Se representa la media de 3 réplicas técnicas independientes. Los valores numéricos se normalizaron a los niveles de un control interno (rRNA 18S), asignándose valor de 1 a las plantas Control.

Los resultados señalan que la mayor expresión se observa entre los 10 y 15 días a concentraciones superiores a 100 mg/L de Aroclor triplicando y cuadruplicando a la expresión del control. La curva de disociación del gen PCSP nos genera un solo producto de PCR.

4.7. Selección de líneas transformadas

Para seleccionar plantas transformadas, germinamos las semillas en medio MS suplementado con glufosinato de amonio (10 µg/mL). Se realizaron cuatro rondas de selección con herbicida hasta obtener plantas homocigotas para el transgen, con las que se abordaron los análisis posteriores. Seleccionamos 8 líneas transgénicas para los ensayos posteriores.

4.7.1. Nivel de expresión del transgén

Una vez que se contaba con suficientes plantas transgénicas establecidas en tierra, se procedió a recolectar material de cada línea para obtener RNA. Se siguió el procedimiento de extracción detallado en Material y Métodos, obteniéndose concentraciones y calidades muy aceptables para posteriormente sintetizar cDNA con la enzima transcriptasa reversa. Todas las reacciones de qRT-PCR llevadas a cabo usando como molde este cDNA, dieron lugar a amplificaciones que fueron analizadas posteriormente.

Se utilizaron oligonucleotidos diseñados sobre el transcrito que mostraron elevada especificidad, de modo que no se amplificaron productos inespecíficos en ningún caso. La eficiencia de las reacciones se mantuvo en todos los casos entre el 85 y el 95%, aunque para los cálculos de cuantificación relativa se consideró óptima e idéntica, según lo establecido por Livaka y Schmittgen (2001). Los resultados de las reacciones de qRT-PCR se obtuvieron en forma de valores Ct (Valor umbral del ciclo), que fueron utilizados para calcular los niveles de expresión génica. Los niveles de expresión relativa del gen estudiado fueron calculados por el método $\Delta\Delta C_t$ desarrollado por PE Applied Biosystems en base a la fórmula $2^{-\Delta\Delta C_t}$. Los valores obtenidos representan el número de veces que se expresa más el

gen diana que el calibrador, que es el mismo gen diana en las condiciones del tratamiento control. Es decir, valores iguales a uno representan una expresión idéntica entre el gen diana y el calibrador. Valores menores que uno, una expresión menor del gen diana frente al calibrador y valores mayores que uno, una expresión mayor del gen diana frente al calibrador.

La cuantificación normalizada proporciona información sobre la cantidad de expresión de un gen en términos relativos. Es conocido que para una eficiencia de la reacción del 100% el valor absoluto de expresión guarda una correlación positiva con el valor $2^{-\Delta\Delta Ct}$ (Livaka y Schmittgen, 2001). Mediante la transformación $2^{-\Delta\Delta Ct}$ eliminamos las variaciones debidas al efecto “tubo” (diferencias producidas por la degradación parcial al RNA en alguna muestra, distinta eficiencia en la transcripción inversa, errores de pipeteo, etc) al incluir un control endógeno. Obtenemos además un valor directamente proporcional a la cantidad total de transcrito de cada gen, pudiendo comparar así niveles de expresión.

La expresión en las plantas transgénicas es muy clara respecto a los controles, en promedio unas 400 veces más. Se evaluaron ocho líneas transgénicas con un alto nivel de expresión. Las líneas 1, 2, 3, 4, 5, 6, 7 y 8 expresan el gen PtCSP4 entre 4.200 y 35 veces más controles (Figura 43 A y B).

Figura 43. Expresión de PtCSP4 en plantas transgénicas de *Arabidopsis*. A) Niveles promedio de transcrito detectado en las plantas transgénicas mediante qRT-PCR. Se representa la media de 3 réplicas técnicas independientes. Los valores numéricos se normalizaron a los niveles de un control interno (rRNA 18S), asignándose un valor de 1 a las plantas Col0 (Control). B) Foto de las líneas transgénicas de *A. thaliana* analizadas en esta Tesis (L1 a L8) y *Arabidopsis* Col 0.

4.7.2. Tolerancia de las líneas transgénicas a Aroclor 1221

Los ensayos de tolerancia a la presencia de PCBs se llevaron a cabo con plantas germinadas y crecidas *in vitro* en condiciones controladas. Se eligieron tres líneas transgénicas de acuerdo a su nivel de expresión (altas y bajas expresoras), las líneas 8, 6 y 3, considerando también su vigor con respecto a la Col 0. En las placas se observa una tolerancia mayor en las plantas transgénicas que en el control (Figura

44 A y B). Al evaluar las plantas a la semana de tratamiento, se registró un 30% más de supervivencia en la línea menos expresora (L4) con relación a las plantas wt. En las líneas L8, L6 y L3 el valor fue un 70% superior. En las plantas transgénicas, el valor de peso fresco fue unas 3 veces superior al de los controles, sin diferencias significativas entre líneas. Todas las líneas fueron capaces de reanudar su crecimiento normal y producir semillas.

Figura 44. Líneas transgénicas con mayor tolerancia al contaminante. A) Tolerancia a 200 ppm de Aroclor 1221. B) Las mismas líneas sin contaminante.

5. DISCUSIÓN

5. Discusión

« *La ciencia viene, la sabiduría se queda*».

Alfred Tennyson (Poeta y Dramaturgo Inglés)

Al igual que otros organismos vivos, las plantas muestran sensibilidad a numerosos contaminantes y pueden sufrir daños significativos causados por ellos. El problema de la contaminación va en aumento y esto afectará cada vez más a las producciones agrícolas y forestales. También causará más problemas de salud. El impacto socio-económico puede ser ingente, aparte del daño ambiental. Este escenario pone de manifiesto la necesidad de ir hacia una economía más sostenible y respetuosa con el medio. También la necesidad de explorar al máximo el potencial de las nuevas tecnologías de descontaminación, como la fitorremediación, una alternativa de bajo coste, ecológica y de probada eficiencia (Gomes et al., 2013; Passatore et al., 2014). Para esta Tesis se seleccionó *Populus*, cuyo genoma está secuenciado completamente y constituye hoy el mejor sistema modelo forestal. Además, es una arbórea de rápido crecimiento, genera grandes cantidades de biomasa, muestra altas tasas de transpiración y cuenta con un sistema radical profundo. Por otra parte, *Populus* tiene una gran capacidad natural para la degradación de contaminantes orgánoclorados, lo que explica su predominio en programas de fitorremediación que se desarrollan actualmente. Hemos elegido en concreto el híbrido *P. tremula* x *P. alba*, por la facilidad con que se cultiva *in vitro* y *ex vitro* y su particular interés biotecnológico.

Entre los innumerables contaminantes que existen, destacan por su peligrosidad los PCBs, contemplados en el Convenio de Estocolmo sobre contaminantes orgánicos persistentes. Son compuestos muy estables (persistentes) y ubicuos, se bioacumulan en el tejido adiposo, y se encuentran en concentraciones significativas en numerosos alimentos. Estas características, unidas a su toxicidad y a la falta de datos sobre su metabolismo en plantas, nos llevaron a elegir a los PCBs para nuestro estudio. Su fabricación está prohibida en muchos países donde se comercializaron durante

décadas. Se eligió para esta tesis la mezcla comercial Aroclor 1221 debido a que, de las trece mezclas registradas, fue la más utilizada industrialmente.

5.1. Efectos de la presencia de Aroclor 1221 en chopo

Una vez elegida la especie y el contaminante, se realizaron pruebas con plantas *in vitro* en condiciones controladas de luz y temperatura. Se eligió el cultivo *in vitro* para garantizar que no se producirían interferencias microbianas en los resultados. Las pruebas consistieron en la evaluación del crecimiento foliar y radicular de los chopos híbridos en presencia de distintas concentraciones de Aroclor 1221 diluido en DMSO estéril, desde 0 (control), hasta 200 mg/L y a dos tiempos, 5 y 15 días. Era difícil establecer las condiciones óptimas de tratamiento a priori, ya que la tolerancia varía mucho según la especie y el estado fisiológico de cada ejemplar, por no mencionar las condiciones del medio (Yang et al., 2013).

En nuestros análisis utilizamos chopos enraizados durante 2 meses, para evitar expresiones génicas asociadas a estadios muy juveniles y también para minimizar los posibles efectos de los transplantes. El cultivo *in vitro* de leñosas, se considera generalmente muy difícil, si bien hay protocolos de regeneración más o menos eficaces para algunas especies. El híbrido elegido para esta Tesis es particularmente viable y, de hecho, lidera la literatura sobre biotecnología forestal (Harfouche et al., 2012; Isebrands y Richardson, 2014).

Las comparaciones entre plantas tratadas y plantas control revelaron diferencias significativas en algunas variables cuantitativas (ver apartado 4.1.1. de Resultados). El crecimiento de la planta en general y el sistema radical disminuyó conforme aumentaban la concentración de PCB y el tiempo de exposición. También aparecieron síntomas de clorosis y necrosis foliar (Figura 12). Estos efectos concuerdan con respuestas típicas de estrés celular (Proudfoot et al., 2002; Yang et al., 2013). En nuestro caso, los análisis de crecimiento y morfológicos han servido para valorar la respuesta del chopo a distintos niveles de PCBs y seleccionar las condiciones experimentales óptimas para los tratamientos.

5.2. Secuenciación por RNA seq

Desde el punto de vista científico, el reto principal de esta Tesis era identificar mecanismos relevantes que pudieran participar en el metabolismo de PCBs y comprender su regulación. No se conoce hasta la fecha ninguna ruta degradativa en plantas para este tipo de compuestos. Esta información es imprescindible, sin embargo, si aspiramos a mejorar las capacidades naturales de algunas especies vegetales para remediar suelos contaminados. En este sentido, las tecnologías de alto rendimiento, como la transcriptómica utilizada aquí, ofrecen una serie de ventajas a la hora de procesar y analizar datos biológicos complejos como los mecanismos moleculares que subyacen a la respuesta a PCBs. Se sabe que las modificaciones en la expresión génica desempeñan un rol importante en la respuesta al estrés ambiental y que estas modificaciones siguen patrones específicos de co-expresión y co-regulación (Ramel et al., 2007; Weisman et al., 2010).

Decidimos utilizar en concreto tecnología RNA-seq, porque proporciona una visión comprehensiva del transcriptoma y permite detectar un rango amplio dinámico de niveles de expresión (Mader et al., 2011). Esta estrategia presenta ventajas frente al uso de microarrays. En primer lugar, tiene mayor sensibilidad y reproducibilidad, además de requerir menor cantidad de RNA de partida (Wang et al., 2009; Oshlack et al., 2010, Chen et al., 2011). En segundo lugar, permite capturar un mayor número de transcritos expresados. De hecho, los mientras que los análisis basados en microarrays dependen de anotación genética a priori y no son capaces de detectar nuevos genes o transcritos. En tercer lugar, genera información no solo de la secuencia, sino también de la estructura de exones y posibles eventos de splicing alternativo (Lister et al., 2009; Gullledge et al., 2012). Por si fuera por los datos arrojados por RNA-seq tienen en general una precisión alta, comparable a los estudios de expresión génica que se obtienen a través de qRT-PCR (Nagalakshmi et al., 2008; Wang et al., 2009; Ward et al., 2012).

El procedimiento general en este tipo de análisis consiste en la fragmentación del RNA y la secuenciación de fragmentos cortos mediante alguna de las tecnologías disponibles en el mercado, como la de Solexa/Illumina, SOLiD o Roche, generando secuencias de entre 35 y 500 pb (Martin y Wang 2011). Para esta Tesis se seleccionó a Solexa/Illumina como plataforma de secuenciación, ya que en comparación con las otras es más económica, y fiable. Además, ya contábamos con un genoma de

referencia (*P. trichocarpa*) por lo que era innecesaria la mayor capacidad de lecturas proporcionadas por Roche/454. Está demostrada la aplicabilidad de esta tecnología a la hora de caracterizar y cuantificar transcriptomas, así como de analizar variaciones de la expresión génica. En cuanto estudios relacionados con contaminación podemos citar a Halimaa et al. (2014), quienes emplearon RNA-seq para analizar la expresión génica diferencial entre ecotipos de *Noccaea carulescen* e identificar genes candidatos para fitorremediar metales pesados. También a Uren et al. (2013) que utilizaron RNA-seq para analizar la respuesta global a metales en trucha. Sin embargo nunca se han estudiado ni identificado genes que participan en el mecanismo de respuesta a PCBs con RNA-seq u otras tecnologías de secuenciación similares.

El primer paso del proceso experimental hacia la secuenciación consistió en la extracción de RNA de alta calidad (ver Figura 13 de Resultados) para dos concentraciones de PCBs y en dos momentos distintos de exposición. Esto nos permitió tener información para un nivel bajo y otro alto de contaminante, además de un tiempo temprano y otro tardío. Obtuvimos así cinco librerías de cDNA conteniendo numerosos transcritos que fueron analizados en esta Tesis. Tras la purificación del RNA, procedimos a su fragmentación, transcripción reversa, síntesis de cDNA y ligación de adaptadores para secuenciación en un sistema RNA-seq Illumina 2000.

5.3. Calidad de la secuenciación y pre-procesamiento de lecturas

Para evaluar la calidad de las lecturas obtenidas se estimó el porcentaje de saturación y la cobertura de secuenciación. Nuestros resultados indicaron un porcentaje de saturación del 30% de genes identificados con las lecturas de la secuenciación (Ver Figura 14 de Resultados). Esto significa que la cantidad de lecturas obtenidas es suficiente para detectar los transcritos diferencialmente expresados según lo establecido por Tarazona et al., (2011). La calidad de nuestras librerías fue también corroborada por la distribución de las lecturas sobre los genes de referencia referida también como cobertura (Wang et al., 2009; Sims et al., 2014). Dicha cobertura fue mayor del 45% en todas nuestras librerías (Ver Figura 15 de

Resultados), lo que supera con creces el mínimo sugerido por Haas et al., (2013) que establecen un valor umbral de 30% para RNA-seq.

Una vez analizada la calidad de la secuenciación, se procedió a un pre-procesamiento de las lecturas de la secuenciación inicial para incrementar su calidad y eliminar elementos artefactuales (Falgueras et al., 2010). Existen varias herramientas bioinformáticas para este fin tales como Trim-Seq, TrimEST, VectorStrip, Lucy y SeqClean (Scheetz et al., 2003; White et al., 2008; Chou y Holmes, 2001; Li y Chou, 2004). En comparación con estos programas, el utilizado aquí (SeqtrimNEXT) es capaz de mantener un alto rendimiento de análisis de la información, incluso con secuencias cortas, siendo flexible en los formatos de entrada y salida. Además es accesible por cualquier usuario a través de una interfaz web y proporciona resultados más precisos que otros programas (Falgueras et al., 2010). En nuestro caso SeqtrimNEXT resultó ser muy eficiente para la eliminación de lecturas de baja calidad, adaptadores, lecturas muy complejas, colas PolyA/T, secuencias contaminadas etc. El porcentaje promedio de lecturas fué de un 13%, mientras que el porcentaje de lecturas descartadas por el algoritmo de BGI fue solo del 3%. Esto es debido a que como hemos mencionado anteriormente, SeqtrimNEXT es un programa más complejo y estricto (Ver Tablas 6 y 7 de Resultados).

5.4. Cuantificación de la expresión diferencial

El siguiente paso fué el alineamiento de las lecturas, para convertir los datos en una medida cuantitativa de la expresión génica. Se alinearon un promedio de 585.570.000 nucleótidos por librería. La secuenciación generó entre 11,95 y 12,61 millones de lecturas. En conjunto, más del 60% de las lecturas pudieron ser alineadas sobre el genoma. En concreto, un mínimo del 61,8% y un máximo del 78,78%, en el caso del transcriptoma, más del 60% de las lecturas pudieron también ser alineadas: un mínimo del 62,1% y un máximo del 82,5%. Resultados similares fueron obtenidos por Zhang et al., (2014) quienes realizaron un análisis transcriptómico de *P. x euroamericana*, obteniendo un 73% de secuencias alineadas en el genoma de *P. trichocarpa*. Nuestros datos de mapeo son también comparables a los publicados por Chen et al., (2015) y Ariani et al., (2015), quienes realizaron un perfil

transcriptómico con *Populus tomentosa* y *Populus x canadensis*, respectivamente, revelando porcentajes de lecturas alineadas en torno al 65- 70%.

Los análisis de expresión diferencial realizados en esta tesis se basaron en estadísticas de recuento de lecturas que mapean sobre el transcriptoma y el genoma de *P. trichocarpa*. Los chopos utilizados en nuestro estudio son híbridos *P.tremula x P.alba*, del mismo género y por tanto cercanamente emparentados. Es por eso que el radio de mapeo es menor que si la especie utilizada hubiese sido *P. trichocarpa*. Además, la validación de 15 genes por qRT-PCR y los resultados de correlación del RNA-seq apoyan fuertemente la viabilidad de nuestro estudio y avalan su aplicabilidad a genotipos híbridos de chopo. Como parte del análisis de expresión diferencial fue necesario generar de una matriz de cuatro elementos para todas las comparaciones (control vs. tratamientos). Con el fin de incrementar la especificidad y fiabilidad de nuestro análisis, consideramos además los genes diferencialmente expresados que interaccionan en los cuatro tratamientos, según cuatro algoritmos estadísticos distintos (González-Gayte et al., 2014). Se seleccionaron para los análisis subsecuentes los transcritos que presentaban una abundancia relativa mayor de 2 con respecto al control. El proceso generó una buena cantidad de genes expresados y no expresados o reprimidos, como se puede observar en el apartado 4.1.5 de Resultados. Se observa claramente que la cantidad de genes expresados diferencialmente va en aumento cuando la concentración y el tiempo también aumentan. Esto es debido probablemente a que existe un mayor estrés celular y se activa una mayor cantidad de genes. Para lograr nuestro objetivo de encontrar genes candidatos, se decidió analizar únicamente los DEGs sobreexpresados. Además, nos hemos enfocado en los genes que se interconectan en los cuatro tratamientos para tener una mayor fiabilidad.

5.5. Anotación funcional de genes diferencialmente expresados

Para identificar cuáles son los genes diferencialmente expresados que participan en metabolismo de PCBs, es necesario descifrar que está ocurriendo con la célula a un nivel global después de un estímulo (Aroclor). Con el fin de conocer la función molecular esto se realizó un análisis de anotación por ontología génica. En este estudio un 80% de las anotaciones (Ver Figura 24 de Resultados) de función

molecular encontradas son de actividades relacionadas con el metabolismo de xenobióticos son relacionadas en procesos con la presencia de un agente externo (contaminante), entre ellas tenemos:

La **actividad de transporte** representa un 35% de los transcritos. Según la EMBL-EBI (<http://www.ebi.ac.uk/>) los genes pertenecientes a esta anotación son capaces de transferir sustancias de un lado de la membrana a otro. En este grupo se incluyen la actividad ATPasa de movimiento transmembrana de sustancias. Esta actividad ayuda a que estas enzimas importen muchos metabolitos necesarios para el metabolismo celular y exportación de toxinas o desechos que puedan entorpecer el proceso celular, lo cual resulta normal para el metabolismo de PCBs. Resultados similares fueron encontrados por Wang et al., 2011, quién realizó un análisis transcriptómico en *Arabidopsis thaliana* en presencia de Ocratoxina-A encontró que un 50% de los genes se encuentran en el tipo de anotación de transporte transmembrana. Otro estudio transcriptómico donde se dilucidó el efecto de algunos xenobióticos presentes en *Crassostrea brasiliana* realizado por Lüchmann et al., (2015) encontró que un 31% de los genes expresados tienen una actividad transferasa.

La **actividad catalítica** (GO:0003824) engloba los transcritos que ayudan a aumentar la velocidad de reacción y la actividad oxidoreductasa donde se encuentran transcritos de especies reactivas de oxígeno (ROS) (<http://www.ebi.ac.uk/>).

La **actividad oxidoreductasa** (GO:001670) fue una de las más representativas con un 52%, resultados similares fueron obtenidos en el perfil transcriptómico de *Arabidopsis* de Wang et al (2011) con un 60% de esta actividad en presencia de Ocratoxina-A. Resultados similares fueron obtenidos por Ariani et al., (2015) quien realizó un perfil transcriptómico para dilucidar el mecanismo de exceso de Zinc en las raíces de *Populus x canadensis* encontrando un 23 % de genes que se agrupan en esta anotación.

El tercer grupo de funciones moleculares fue la de **unión a DNA con actividad de factor de transcripción** con un 12%. Di Baccio et al (2011) han encontrado un porcentaje equivalente al nuestro en el análisis funcional correspondiente al perfil transcriptómico realizado, utilizando el clon *Populus x euroamericana* expuesto a un exceso de zinc.

5.6. Selección de genes y factores de transcripción

El análisis de interacción génica en los cuatro tratamientos se realizó mediante cuatro paquetes estadísticos (edgeR, limma, NOISeq y DESeq2) con el fin de tener una mayor validez estadística. Los genes que se encuentran en la interacción son, en su mayoría, genes con función de oxidación-reducción, hidrólisis, conjugación, transporte, compartimentación y factores de transcripción cuya participación serán analizados más adelante. Resultando gráficamente útil el uso de Diagramas de Venn y heat maps. Esta estrategia también ha sido utilizada por Ariani et al., (2015) quien ha utilizado los genes que se agrupan en la intersección, sin embargo solo utilizó un paquete estadístico (DESeq) por lo que nuestro estudio esta mas representado estadísticamente. Por otro lado para entender los cambios transcripcionales a través del tiempo y los genes que agrupan ante un evento dado (contaminación con PCBs) se realizó un análisis secuencial. Los resultados obtenidos nos han ayudado a entender mejor el mecanismo de descontaminación. Esta estrategia fue propuesta por Nueda et al (2009) y ha sido aplicada este por varios autores (Prado-López et al., 2010 y Rizza et al., 2012). Nosotros identificamos cuatro grupos o perfiles de expresión génica los cuales señalan que el agrupamiento que nos da mayor y mejor información fue el Grupo 3 debido al comportamiento creciente (baja-alta concentración) de los genes sobre-expresados. En este grupo se han encontrado más de 18 transportadores ABC, 10 transportadores MDR, 5 lacasas, 4 catalasas, 4 tioredoxinas, 7 glutaciones y más de 26 citocromos P450 esto sugiere que estos procesos están muy representados, sobre todo en el tiempo más largo (15 días). Además, el análisis secuencial nos indica que en el tiempo corto (5 días) solo se activan genes de señalización, receptores de quinasas y factores que podrían participar en la detección y señalización primaria precedentes de la decodificación de PCBs. Ciertamente los genes en el tiempo mas largo pueden resultar genes candidatos clasificados en la etapa I: transformación como P450 etc, etapa II: conjugación GST y UGT, y etapa III: compartimentación transportadores ABC, MDR etc. Finalmente, los genes expresados y normalizados fueron sometidos a un análisis de co-expresión. Al ser integrados pueden proveer modelos de redes transcripcionales que controlan las redes de expresión génica de un mecanismo en particular, en este caso buscamos encontrar el mecanismo de metabolismo de PCBs (Ver Figura 49). Los resultados de este análisis se discuten en el siguiente apartado.

5.7. Genes y factores de transcripción que participan en el metabolismo de PCBs

La correlación potencial de la expresión génica que envuelve el metabolismo de PCBs en chopo y su regulación, parece estar coordinada por 35 genes y 15 factores de transcripción. Se proponen los siguientes genes claves o posibles llaves transcripcionales que sugieren una candidatura fitorremediadora. También nos permite observar una cadena de genes unidos por un coeficiente de co-expresión que interrelaciona e induce el comportamiento de un sistema (Bhattacharyya, 2009).

En la Figura 45, se observa un mapa de co-expresión génica en la que el papel de la Glutación transferasa esto es debido a que la síntesis de glutatión es regulada por la familia de enzimas conocidas como glutatión-S transferasas (GSTs). Esta comprobado que en presencia de un contaminante se forma un complejo con la molécula de GST, la cual interactúa con enzimas que facilitan su degradación, como la glutaminil-transpeptidasa (Brazier et al., 2005; Wang et al., 2012). En el análisis de co-expresión se encontró un GST (GSTU25) (Potri.010G60900) que está relacionado con xenobióticos, por lo que es muy probable que esté relacionada con la detoxificación.

En un estudio del año pasado se menciona que el tripeptido tiol glutatión (g-L-glutamyl-L-cysteinyl-glycine) es el más importante antioxidante en plantas el cual está relacionado con la detoxificación de especies (ROS), señalización REDOX entre otras (Zechmann, 2014). Otros estudios reafirman la relación del estrés oxidativo (Mecanismos REDOX) en presencia de xenobióticos (Mylona et al., 2007).

Consecuentemente la inducción y coordinación de ambas repuestas (detoxificación y antioxidante) son de gran importancia para una defensa eficiente. La detoxificación de xenobióticos típicamente muestra activación a través de hidrólisis u oxidación catalizada por la monooxigenasa citocromo P50, seguida de una unión covalente de moléculas hidrofílicas endógenas como Glutathione (GSH), Glucosa y Malonato para formar compuestos más solubles y conjugados menos tóxicos (Zechmann, 2014). En nuestros resultados la presencia del Cyt P450 (Potri.004G106600) es notable y según los estudios representa un gen clave para el metabolismo de xenobióticos. El sistema P450 comprende un total de 90 genes que se codifican en el genoma de *Populus*

trichocarpa v.3. Las proteínas del citocromo P450 comprende una superfamilia de proteínas que contienen el cofactor heme que cataliza la oxidación de sustancias orgánicas (Olsen et al., 2015).

Además el sistema de coe-expresión está unido a flavin monooxigenasa (Potri.002G254200), UDP-glucosiltransferasa (Potri.001G303000) y cinamil alcohol deshidrogenasa (Potri009G063100), entre otros genes.

Las flavín-monooxigenasas ocupan el segundo lugar en importancia en cuanto a las oxidaciones metabólicas. Estas enzimas están implicadas en la oxigenación de compuestos nitrogenados (con la formación de N-óxidos), organofosforados y organosulfurados. A diferencia de los citocromos, son enzimas que poseen como coenzima flavina-adenosina dinucleótido (FAD) y son capaces de utilizar directamente NADPH como cofactor para reducir uno de los dos átomos de oxígeno. Su mecanismo de acción catalítico es singular: a diferencia de los citocromos P450, la interacción del xenobiótico con la enzima es mucho más laxa. La oxidación se produce a través de un intermedio reactivo oxidante generado en el mismo centro catalítico de la enzima (hidroperoxi-flavin monooxigenasa) que oxida a un contaminante.

En esa misma línea hay una enzima clave denominada Short Chain deshidrogenasa (Potri.013G059100), que mantiene co-expresión con Glutatión transferasa (Potri.010G60900) y a una proteína de transporte MDR (Potri.012G033400) la cual es por ortología una MRP14. Esta última tiene relación con el mecanismo de transporte para la exclusión de xenobióticos y de sus metabolitos (Fekete et al., 2015). Lo cual sugiere que este transporte es parte integral del mecanismo de detoxificación de plantas. Las plantas no tienen sistema eficiente para excretar compuestos para el medio ambiente, tienen que almacenar los compuestos potencialmente tóxicos en vacuolas que tienen una actividad metabólica muy limitada. Además existen numerosas investigaciones que revelan que las MDR no sólo están implicados en la desintoxicación, sino también en la regulación celular de almacenaje (Gaedeke et al., 2001; Fekete et al., 2015; Cummins et al., 2015). Se ha encontrado que AT3G59140 (AtMRP14; MRP14) está inducido más de 2,5 veces (fold change) en respuesta a Benzoxazolinona en *Arabidopsis* (Baerson et al., 2005). Lo mencionado anteriormente confirma que el transportador MDR

Potri.012G033400 (34) tiene relación con el proceso de contaminación ya que tiene como ortólogo a AT3G59140 (MRP14).

SDRs constituyen una superfamilia de NADPH - dependiente deshidrogenasas. Estas enzimas tienen una baja homología de secuencia, pero una buena conservación de la estructura en 3D, lo que sugiere similitudes también en mecanismo catalizador (Llenar et al., 2002). La alta variabilidad del centro activo explica la amplia gama de sustratos que pueden procesar SDRs. Los principales caracteres funcionales de estas proteínas incluyen la capacidad de interconvertir grupos carboxilo e hidroxilo y la reducción de los dobles enlaces C -C y N -N (Jörnvall et al., 1995). Otra característica interesante de las SDRs es su alta similitud en estructura y posiblemente en función con deshalogenasas bacterianas de halohidridos. Las deshalogenasas catalizan la eliminación nucleofílica de un halógeno, sustituido por un hidroxilo (Van Hylckama Vlieg et al., 2001). Estas enzimas son muy interesantes para la bioremedation de haluro alifático en el suelo. Estos genes están recibiendo cada vez más atención desde el punto de vista farmacológico y biotecnológico (Kavangh et al., 2008). Hasta la fecha ningún mecanismo deshalogenación se ha descrito en las plantas

Figura 45. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs. Los genes candidatos están marcados en rojo junto al acrónimo del gen. MDR (multidrug resistance), SDR (Short chain deshidrogenasa), GST (Glutation transferasa), CAD (Cinamil alcohol deshidrogenasa), FMO (Flavin monooxygenasa), P450 (Citocromo P450), CSP (Cold shock protein).

En la Figura 46 se observa que el coeficiente de co-expresión es alto, pudiéndose visualizar por lo ancho de las líneas de unión. El transportador MFS (11) (Potri.010G089800) pertenece a una de las cinco familias de transportadores MDR que mantiene redes de co-expresión con cuatro genes TRX (5) (Potri.001G416500), leucine zipper EF hand (8) (Potri.001G187100) y Quinasa (13) (Potri.006G138400), mostrando con ello su papel central en el proceso (Figura 44) Esto es debido a que además de la familia ABC, la superfamilia principal de transportadores facilitadores (MFS) también previene la acumulación de compuestos tóxicos (Pao et al., 1998; Reedy et al., 2012).

Anteriormente, se consideraba que los transportadores MFS se ocupaban solamente de la secreción de toxinas endógenas (Reedy et al., 2012). Sin embargo, existen estudios que han demostrado que los transportadores MFS de *Candida albican* y *S. cerevisiae* también están involucrados en la protección contra compuestos exógenos, tales como los inhibidores de la desmetilación de esteroides (Calabrese et al., 2000; Hayashi et al., 2002). Otros estudios en *Botrytis cinerea* revelaron la existencia de un transportador MFS, denominado Bcmfs1. La función principal de esta enzima es metabolizar xenobióticos. Son flavín monooxigenasas de la clase A dependiente de NADPH, que intervienen en la ruta de desulfuración de compuestos orgánicos, acopla la oxidación de NADH con la oxidación del sustrato en cada caso, mediante la reducción de FMN a FMNH₂. Catalizan reacciones de oxigenación a partir de oxígeno molecular, que usa como sustrato para oxidar un compuesto orgánico (Hayashi et al., 2002).

En esta misma red (Figura 46) se encuentra una proteína quinasas AMP (20) (Potri.011G122000), éstas son activadas por estrés metabólico, drogas o xenobióticos, que además, interfieren con la generación de ATP catabólico y acelera la consumición de ATP (Hardie, 2007).

Figura 46. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs. Los genes candidatos están marcados en rojo junto al acrónimo del gen. AMP (proteína quinasa), GSR (Glutación reductasa), CHIA (Quitinasa), LETM (leucine zipper EF hand), MFS (Transportador MFS), UDP_GT (UDP glucotransferasa), TRX (Tioredoxina).

La respuesta a xenobióticos en las plantas envuelve una larga escala de redes de genes, grandes complejos de regulación, factores de transcripción y proteínas quinasas, además de efectos de señalización (Ramel et al., 2007; Weisman et al., 2010). La respuesta a xenobióticos no solo depende de un solo factor de transcripción (Ekman et al., 2005; Ramel et al., 2007; Wang et al., 2012). Se ha encontrado que los MYBs y WRKYs están envueltos en la regulación de respuesta al estrés (Singh et al., 2002).

Ramel et al., (2012) señala a At5G13080 (WRKY75) como un factor de transcripción asociado a tolerancia a estrés por xenobióticos. Este factor de transcripción es el ortólogo de Potri.003G169100 (E), Potri.T04800 (K), Potri.015G099200 (F) confirmando la participación de estos tres factores de transcripción en el proceso de tolerancia a PCBs (Figura 45). Está comprobado que el WRKY75-codifica genes que son inducidos por varios efectores ROS como el metilo viológeno diclorado (paraquat), la toxina del hongo *Alternaria alternata* (ácido tenuazonico), y 3-amino- triazol conocido (amitrol) (Gadjev et al., 2006). A su vez la Ocratoxina A induce WRKY75 (Wang et al., 2012), el cual está asociado a respuesta de defensa a antrazina.

Varios elementos cis-reguladores (CAREs) han sido identificados en las regiones promotoras de genes que responden a antrazina (Ekman et al., 2005; Ramel et al., 2007; Wang et al., 2012). Se ha encontrado que, los sitios motivos (secuencia consenso TAACTAAC) de unión de MYB3 han sido relacionados en repuesta a estrés (Bang et al., 2008). En nuestro análisis de co-expresión se encontró un factor de transcripción MYB3 Potri.003G109300 (G) por lo que es muy probable su participación en respuesta a estrés. Resulta interesante que mantenga co-expresión con el gen PtCSP4 (Potri.004G172600) (12), un gen nunca antes relacionado con respuesta a xenobióticos (Figura 47).

Por otro lado, Baerson et al., (2005) y Fode et al., (2008) enfatizaron el rol central de los factores TGA a la respuesta a xenobióticos. El factor de transcripción TGA del tipo STP1 (Potri.009G164300) (N) del análisis de co-expresión podría participar en la ruta de degradación de xenobióticos. Además, en nuestro análisis se encontró un factor de transcripción GRAS (SCL14) Potri.009G032800 (I). SCARECROW-like

14 (SCL14) es un miembro de la familia GRAS la cual responde a herbicidas (Behringer et al., 2011).

Los factores de transcripción Helix-loop-helix (bHLH) regula la distribución de ROS (Tsukagoshi et al., 2010). Estos genes en animales son receptores nucleares con actividad transcripcional o son parte de complejos de factores de transcripción (Ulrich, 2003; Baker, 2005), los cuales están vinculadas estrechamente a señales ambientales. Ciertamente, los receptores AhRs (aril hidrocarbon), reconocen contaminantes orgánicos. Los AhRs son miembros de la familia de factores de transcripción bHLH-Per-ARNT-Sim (bHLH-PAS) (Ulrich, 2003; McMillan y Bradfield, 2007). En nuestro análisis de co-expresión hemos encontrado un bHLH07 (Potri.001G141100) (H). Este hallazgo confirme que este bHLH07 podría formar parte en la regulación en respuesta a xenobióticos.

El AhR es un factor de transcripción activado por unión de ligando, que media los efectos biológicos de PCBs a través de su heterodimerización con una segunda proteína denominada translocador nuclear del AhR (ARNT, *Aryl hydrocarbon Receptor*). La formación de este heterodímero transcripcionalmente activo induce la expresión de genes diana que codifican para enzimas implicadas en destoxificación, tanto de la fase I (citocromos P450, CYP450) como de la fase II (transferasas).

La unión del AhR a sus ligandos produce la activación del receptor, lo que conlleva cambios en su compartimentalización celular. En ausencia de ligando, la forma latente del AhR se encuentra asociada a dos moléculas del chaperón HSP90 (*Heat Shock Protein 90*), a una molécula de la chaperonina p2 y a una de la proteína XAP2 (*Hepatitis B virus X-Associated Protein 2*), también conocida como ARA9 (*AhR-Associated protein 9*) o AIP (*AhR-Interacting Protein*) (Bessede et al., 2014).

Tras la unión del ligando, el AhR se acumula rápidamente en el núcleo celular donde forma un heterodímero transcripcionalmente activo con ARNT (*AhR Nuclear Translocator*). Esta heterodimerización disocia el complejo HSP90-XAP2-p23 del AhR. El heterodímero activo AhR/ARNT se une a elementos reguladores denominados XREs o DREs (*Xenobiotics o Dioxin Response Elements*) situados en regiones potenciadoras/promotoras de genes diana, incrementando así la transcripción de los mismos (Petersen et al., 2003).

La secuencia consenso establecida para el XRE es 5' GCGTG 3' (Huang y Elferink, 2012). Mediante el análisis de bases de datos ha sido posible identificar esta secuencia consenso de los genes: PtCyp_P450 (Potri.004G106600), PtSDR (Potri.013G059100), 4,5- DOPA dioxigenasa extradiol (Potri.004G135300) y la transferasa PtXGT (Potri.005G007200) (ver Figura 48) potencialmente regulados por el AhR esta interacción se realiza a través de AhRT (Kumar et al., 2001).

En la Figura 48 se muestra un esquema del mecanismo de acción propuesto para la activación transcripcional de genes dependientes del AhR. Los genes diana para el AhR identificados codifican para enzimas implicadas en la destoxificación metabólica, tanto de la Fase I (p450, flavin monoxigenasa, cinamil alcohol deshidrogenasa etc) como de la Fase II (UDP-glucosiltransferasa, glutathione-S transferasa).

Figura 47. Red de co-expresión de los DEGs envueltos en el presunto proceso de estrés por PCBs. Los factores de transcripción candidatos se representan con cuadrados azules y seguidos del acrónimo del gen señalado.

Figura 48. Modelo hipotético de la red reguladora de genes y factores de transcripción que responden a PCBs en *Populus*. Transducción de señales mediado por el heterodímero AhR/ARNT mostrando los cambios en la distribución celular de los componentes tras la activación por unión de ligando.

5. 8. PCSP4 candidato fitorremediador

Los datos presentados en este trabajo aportan información novedosa sobre las características y funcionalidad de una proteína CSP. Las proteínas pertenecientes a la familia Cold-Shock Domain (CSD) se ha descrito en bacterias y animales, pero muy poco en plantas. Las proteínas CSP en animales (conocidas como Yb) se ha encontrado que también se translocan al núcleo cuando las células se exponen a agentes tóxicos o estímulos medioambientales. Las proteínas Yb de células cancerígenas han sido muy estudiadas debido a ser las responsables de la resistencia de éstas a compuestos antitumorales, como antraciclinas, epipodofilotoxinas o taxoles. Las proteínas Yb-1 se unen a la secuencia Yb-1 invertida CCATT (ATTGG) y activan el gen MDR1 que codifica un transportador ABC (ATP-Binding Cassette), que actúa expulsando de las células los compuestos que les resultan tóxicos (por ejemplo, los compuestos antitumorales). En pacientes de cáncer siempre se comprueba si se está sintetizando esta proteína ya que su expresión en células cancerígenas es un indicador de resistencia a tratamientos antitumorales (Nakaminami, 2006). Sin embargo, en herbáceas, esta familia nunca ha sido asociada en respuesta a xenobióticos. Casi todos los trabajos de investigación se han desarrollado en especies agrícolas o bien especies modelo como *Arabidopsis thaliana*. La escasez de estudios confiere un interés añadido a este trabajo, ya que es la primera vez que se relaciona a estas proteínas con procesos de contaminación orgánica.

Como se mencionó anteriormente no se ha descrito prácticamente nada sobre la participación de las proteínas CSP en plantas, pero es muy probable que su acción sea la misma que en mamíferos, como la activación de genes MDR que sintetizen transportadores ABC de membrana para expulsar a la matriz vacuolar, compuestos potencialmente tóxicos para la planta.

Por último, los PCBs podrían alterar las membranas por ser compuestos muy hidrofóbicos o a través del estrés oxidativo. Esto abre la opción, aún sin explorar, de que los factores transcripcionales CSP estén implicados también en la activación de genes estabilizantes de membranas. Por lo tanto se sugiere que la proteína PtCSP4 tiene un papel relevante en la descontaminación de PCBs. Se propone como una llave transcripcional que podría tener, entre una de sus funciones, activar genes

implicados en la degradación de contaminantes e incrementar la tolerancia a estrés oxidativo en las plantas.

El estudio de preguntas biológicas a través de la proteómica, pueden ser aclaradas. Esta tecnología no ha dejado de ser un herramienta importante en el estudio de la biología de plantas (Thelen, 2007, Santucci et al., 2015). La aproximación proteómica realizada en esta Tesis ha permitido la identificación puntual de la proteína PCSP4.

Se observaron claras diferencias entre los chopos *in vitro* tratados con contaminantes (50 y 200 mg/L) y los chopos *in vitro* control. Se identificó inequívocamente a la proteína PtCSP que se encontraban fuertemente inducidas por los tratamientos. La huella peptídica de la proteína, así como sus espectros de fragmentación nos permitieron su identificación utilizando herramientas bioinformáticas y bases de datos apropiadas.

5.8.1. Análisis filogenético de la familia CSP en plantas

Las proteínas CSP se encuentran ampliamente distribuidas en plantas inferiores y superiores incluyendo mono y dicotiledóneas (Karlson y Imai, 2003). Las CSP de plantas hasta ahora caracterizadas, contienen un dominio Cold shock, con regiones ricas en glicina y secuencias repetidas RGG, además una región grande de dedos de zinc del tipo CCHC intercalados en la zona C-Terminal conocida como (Zinc knuckle domain) (Karlson et al., 2002). Una de las características de estas proteínas es que todas las secuencias de esta familia no contienen intrones.

El número de especies miembros de esta familia contienen de dos a siete genes por especie, además de dos o cuatro dedos de zinc en monocotiledóneas y de uno a siete dedos de zinc en dicotiledóneas. Las regiones ricas en glicina y los dedos de zinc CCHC están implicados en la unión a ácidos nucleicos y a la unión a otras proteínas (Karlson et al., 2002; Nakaminami et al., 2006; Sasaki et al., 2007). La mayoría de los estudios de esta familia se han centrado a la caracterización de proteínas CSP de bacterias y animales, existiendo en la actualidad poca información sobre de las proteínas CSP en plantas.

La primera proteína CSD caracterizada funcionalmente fue WCSP1 de trigo. La proteína WCSP1 contiene regiones ricas en glicina intercaladas con tres regiones de dedos de zinc CHCC en la C- Terminal. WCSP1 es regulada en respuesta al frío y la proteína correspondiente es sustancialmente acumulada en tejido de la corona durante la aclimatación al frío prolongado. Los niveles de transcripción de WCSP1 no se inducen por otros factores ambientales tales como la salinidad, la sequía y el calor, o tratamientos con ácido abscísico (Karlson et al., 2002).

En el arroz hay dos proteínas CDP, OsCSP1 y OsCSP2, las cuales exhiben actividad de unión a ácido nucleicos (Chaikam y Karlson, 2008). La expresión de OsCSPs se incrementó por bajas temperaturas en brotes y raíces. Sin embargo, en el mismo estudio se observó que los niveles de expresión no aumentaron en el tejido de la corona durante los 10 días de tratamiento a baja temperatura. Estos datos contrastan con la expresión observada para WCSP1. En el mismo estudio los patrones de expresión de OsCSPs revelaron que las OsCSP son altamente acumuladas en el desarrollo de la panícula, en la flor y en la semilla (Chaikam y Karlson, 2008). Por lo tanto, las funciones de OsCSPs son más asociadas a los procesos de desarrollo que a la tolerancia al frío, a diferencia de las WCSP1.

En *Arabidopsis* se han identificado cuatro proteínas CSP (AtCSP1, AtCSP2, AtCSP3 y AtCSP4) con las que se han realizado varios estudios. Park et al., (2009) describieron que la sobreexpresión de AtCSP1 (At4g36020), retrasa la germinación de semillas bajo deshidratación o condiciones salinas de estrés. Por otro lado se ha encontrado que AtCSP2 regula negativamente el tiempo de floración, y regula positivamente desarrollo del embrión de la semilla (Fusaro et al., 2007). Además, la sobreexpresión de AtCSP3 confiere una mayor tolerancia frente congelación en *Arabidopsis* sin efectos notorios en el desarrollo (Kim et al., 2009). AtCSP3 no afecta a la expresión de genes CBFS y COR (genes de aclimatación al frío), pero regula la expresión de genes relacionados con el estrés (Kim et al., 2009, 2013). Recientemente, se demostró que AtCSP4 (A2g21060), el parólogo más cercano de AtCSP2, también juega un papel importante en el desarrollo. Dando como resultado silicuas de menor longitud y mayor letalidad embrionaria (Yang y Karlson, 2011).

Los estudios funcionales anteriores ponen de manifiesto que la expresión de proteínas vegetales de que son diferencialmente reguladas por señales de desarrollo y

de estrés. Es evidente dada la información anterior que las diferencias funcionales entre las proteínas CSP son muchas. Es importante conocer a fondo la filogenia. Existen estudios filogenéticos que mayormente están enfocados a bacterias y mamíferos, donde solo se hace referencia a las proteínas en plantas de forma escasa.

En el árbol filogenético obtenido en la Figura 29 de Resultados se crearon 3 grupos:

El grupo I contiene secuencias que mantienen alta similitud en el dominio Cold Shock, al igual que en todas las proteínas evaluadas. Sin embargo comparten los mismos un residuo R (arginina) en lugar de W (triptófano) siendo este el sitio de unión a DNA compartido en el resto de las proteínas. Otra característica de este grupo es que prácticamente no muestran residuos RGG en dos de ellas. En este grupo se encuentran dos proteínas “Cold shock” de *Populus trichocarpa* y una de *Ricinus communis*, las cuales tienen cinco dominios zinc fingers (znf).

El grupo II se encuentran cuatro proteínas con los dominios “Cold shock” altamente conservados, con escasos residuos RGG. En este grupo se encuentran dos proteínas de *Arabidopsis thaliana*, una de *Carica Papaya* y una de *Thellungiella salsuginea*. Todas ellas con cinco dedos de zinc.

En el grupo III se encuentran todas las proteínas cold shock que presentan dos, tres y cuatro dedos de Zinc. Representa el grupo mayoritario de proteínas Cold Shock y esta subdividido en seis subgrupos:

III.1. Seis miembros entre ellos dos de *Populus trichocarpa* entre otros. III.2. Dos proteínas de *Populus trichocarpa*, III.3. Es el segundo subgrupo mas grande que mantiene a nueve CSP. III. 4. Tres proteínas de *Physcomitrella patens* y dos de *Picea glauca* y *sitchensis*. III.5 Seis proteínas CSP y finalmente el subgrupo mas grande, III.6, con diez CSPs. Las especies pertenecientes a estos subgrupos mantienen muy conservado el dominio Cold Shock y los sitios de unión a RNA y DNA. Las diferencias radican en la similitud de los grupos conservados y en la cantidad de secuencias RGG repetidas.

5.9. Sobre-expresión de PCSP4 en *Arabidopsis* aumenta la tolerancia a PCBs

No es sencillo conseguir plantas tolerantes a contaminantes. Varias proteínas CSP (como hemos visto en el apartado anterior) están relacionadas con estrés abiótico entre otros procesos tales como el crecimiento, floración etc, pero nunca han sido relacionadas a el estrés por xenobióticos. Las plantas sobre-expresoras de PCSP4 han mostrado mayor tolerancia al estrés por PCBs (Aroclor 1221). Esto se comprueba la menor severidad del daño visible causado en las plantas transgénicas en comparación con las plantas control. La tolerancia observada puede deberse a los elevados niveles de expresión de PCSP4 y de una cascada de genes que presentan las plantas transgénicas y que participan en la defensa de *Arabidopsis* frente al estrés por PCBs y a estrés oxidativo. De forma que podemos concluir que las líneas sobre-expresoras se comportan como si hubieran sufrido una aclimatación frente al estrés.

La eficacia de las plantas transgénicas para la fitorremediación de contaminantes orgánicos han sido probada en diferentes especies (James y Strand, 2009). Un gen (LAC1) de lacasa de algodón (*Gossypium arboreum*) se sobreexpresa en *Arabidopsis thaliana* dando como resultado en aumento de la actividad secretora de lacasa y mantiene una mejor resistencia a triclorofenol en suelos. Las plantas LAC1 mostraron una mejor resistencia a una variedad de compuestos fenólicos en comparación con los tipos salvajes. Las plantas de tipo silvestre mostraron clorosis severa mientras que el daño a las plantas LAC1 fue menos severa. Los datos sugirieron que LAC1 fue secretada a partir de las plantas transgénicas que metabolizan el triclorofenol *ex situ* resultando en un ambiente menos tóxico para el crecimiento vegetal (Wang et al., 2004). Nuestras plantas fueron capaces de tolerar hasta 200 mg/L de Aroclor 1221. En un estudio similar, la lacasa de *Coriolus versicolor* se expresó en tabaco que resulta en la secreción de lacasa en la rizosfera y una mayor degradación de bisfenol A y pentaclorofenol (PCP) en cultivo hidropónico (Sonoki et al., 2005). Sin embargo, la degradación en suelos no se examinó. Oller et al. (2005) desarrollado un tomate transgénico (*Lycopersicon esculentum* Mill. cv. Pera) que sobreexpresa TPX1, una peroxidasa nativa, resultando en una mayor actividad de la peroxidasa (Oller et al., 2005). Un gen manganeso peroxidasa (MnP) de *Coriolus versicolor* se expresó en plantas de tabaco (Iimura et al., 2002). La actividad MnP en el medio que contenía esquejes de raíz

transgénicas era 50 veces mayor en comparación con los controles que sugieren la secreción de la enzima MnP. Una incubación similar en la presencia de PCP resultó en una reducción de aproximadamente 2 veces en la concentración PCP con la línea transgénica en comparación con los controles. El mismo grupo de laboratorio transformado chopo (*Populus seiboldii* × *Populus gradientata*) con MnP de *Trametes versicolor* (Iimura et al., 2007). Varias líneas transgénicas tenían mayor actividad MnP que el control y contribuyeron a una eliminación más rápida-de bisfenol A que en los medios hidropónicos.

6. CONCLUSIONES

6. Conclusiones

«La ciencia no es sino una perversión de sí misma, a menos que tenga como objetivo final el mejoramiento de la humanidad»

Nikola Tesla (Inventor e Ingeniero Serbio)

Las conclusiones más significativas de esta tesis son las siguientes:

1. Es el primer estudio que describe un análisis transcriptómico en *Populus* con el fin de dilucidar el mecanismo de acción de uno de los contaminantes orgánicos mas peligrosos del mundo, los PCBs
2. Los resultados transcriptomicos se suman a los conocimientos actuales sobre la respuesta al estrés en *Populus*. Además, los datos generados pueden ser utilizados como transcriptoma de referencia para estudios posteriores en *Populus* u otras especies forestales para hacer frente a los mecanismos moleculares que subyacen a la susceptibilidad a contaminantes y resistencia al estrés.
3. Se han identificado genes que presumen una importante participación en el metabolismo de PCBs. Los cuáles serán útiles en muchos aspectos de la biotecnología vegetal y del medio ambiente. Estos genes puede dar lugar a nuevos biomarcadores que estén relacionados con las redes de genes que responden a estrés por PCBs. En particular un Citocromo P450 (Potri.004G106600), GST (Potri.010G060900), SDR (Potri.013G059100), 4,5- DOPA dioxigenasa extradiol (Potri.004G135300), transferasa XGT (Potri.005G007200), 1-amino ciclo propano carboxilato (Potri.002G163700), Transportador MDR (Potri.012G033400), Transportador MFS (Potri.010G089800) y CSP4 (Potri.004G172600).
4. Se han identificado factores de trascrición que participan en el metabolismo de PCBs. Pertenecientes a WRKY75 (Potri.015G099200, Potri.003G169100

y Potri.T043800), MYB3 (Potri.013G109300), GRAS2 (SCL14) (Potri.009G032800), ZF-TF (Potri.008G051200) y bHLH071 Potri.001G141100, NAC Potri.001G404100 y Potri.013G054200, TGA Potri.009G164300, Zinc Finger C2H2 Potri.008G051200 y Potri.010G209400, Zinc Finger ZAT12 Potri.011G138500.

5. Se ha propuesto una ruta metabólica posible de la acción y contribución de los genes y factores de transcripción identificados.
6. Se confirmó la inducción del gen correspondiente, a *PtCSP4* de *P. trichocarpa* (Potri.004G172600), en respuesta a Aroclor 1221, tanto por proteómica como por transcriptómica.
7. El análisis fenotípico de las líneas transgénicas de *Arabidopsis thaliana* que sobre-expresaba la proteína CSP de chopo híbrido confirmó un papel para la misma tolerancia a PCBs.
8. Se tienen ocho líneas transgénicas potencialmente más tolerantes a Aroclor 1221. Sin embargo se requerirá de una caracterización más profunda de sus funciones y comportamiento en campo con fines de fitorremediación.

9. BIBLIOGRAFÍA

10. Bibliografía

«Libros, caminos y días dan al hombre sabiduría».

Proverbio árabe

- Abhilash C, Jamil S and Singh N.** 2009. Transgenic plants for enhanced biodegradation and phytoremediation of organic xenobiotics. P *Biotechnology Advances*. 27 474–488.
- Alves MS, Dadalto SP, Gonçalves AB, de Souza GB, Barros VA, Fietto LG.** 2015. Transcription Factor Functional Protein-Protein Interactions in Plant Defense Responses proteomes
- Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB.** 2013. Transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants*. 19, 307–321.
- Ansari AA, Gill SS, Gill R, Lanza GR, Newman L.** 2015. Phytoremediation. Management of Environmental Contaminants, Volume 1. Phytoremediation.
- Ariani A, Di Baccio D, Romeo S, Lombardi L, Andreucci A, Lux A, Stephen D, Luca H.** 2015. RNA Sequencing of *Populus x canadensis* roots identifies key molecular mechanisms Underlying Physiological Adaption to Excess Zinc. *PLoS ONE*. 10(2)
- Aryal UK, Xiong Y, McBride Z, Kihara D, Xie J, Hall MC, Szymanski DB.** 2014. A Proteomic Strategy for Global Analysis of Plant Protein Complexes. *Plant Cell*. 26(10): 3867–3882.
- Ashburner M, Ball CA, et al.** 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000 May. 25(1):25-9.
- ATSDR (Agency for Toxic Substances and Disease Registry).** 2014. Polychlorinated Biphenyls (PCBs) Toxicity. <http://www.atsdr.cdc.gov/csem/pcb/docs/pcb.pdf>
- Ayoub K, Van Hullebusch ED, Cassir M, Bermond A.** 2010. Application of advanced oxidation processes for TNT removal: A review. *J Hazard Mater*. 178(1-3):10-28.
- Baerson SR, Sanchez-Moreiras A, Pedrol-Bonjoch N, Schulz M, Kagan IA, Agarwal AK, Reigosa MJ, Duke SO.** 2005. Detoxification and transcriptome response in *Arabidopsis* seedlings exposed to the allelochemical benzoxazolin-2(3H)-one. *Journal of Biological Chemistry* 280, 21867–21881.

- Baker ME.** 2005. Xenobiotics and the evolution of multicellular animals: emergence and diversification of ligand-activated transcription factors. *Integrative and Comparative Biology* 45, 172–178.
- Bandyopadhyay S and Bhattacharyya M.** 2009. Analyzing miRNA co-expression networks to explore TF-miRNA regulation. *BMC bioinformatics* 10 (1), 163.
- Bang WY, Kim SW, Jeong IS, Koiwa H, Bahk JD.** 2008. The C-terminal region (640–967) of *Arabidopsis* CPL1 interacts with the abiotic stress- and ABA-responsive transcription factors. *Biochemical and Biophysical Research Communications* 372, 907–912.
- Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT.** 2011. Marth. Bamtools: A C++ API and toolkit for analysing and managing BAM files. *Bioinformatics*, 27(12):1691–1692.
- Baron E, Gimenez J, Verborgh P, Gauffier P, De Stephanis R, Eljarrat E, Barcelo D.** 2015. Biomagnification of classical flame retardants, related halogenated natural compounds and alternative flame retardants in three delphinids from Southern European waters. *Environmental Pollution*. 203: 107-115.
- Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Bhardwaj J, Chauhan R, Swarnkar MK, Chahota RK, Singh AK, Shankar R, Yadav SK.** 2013. Comprehensive transcriptomic study on horse gram (*Macrotyloma uniflorum*): de novo assembly, functional characterization and comparative analysis in relation to drought stress. *BMC Genomics*. 14.pp. 647–663.
- Behringer C, Bartsch K and Schaller A.** 2011. Safeners recruit multiple signalling pathways for the orchestrated induction of the cellular xenobiotic detoxification machinery in *Arabidopsis*. *Plant Cell Environment* 34, 1970–1985.
- Bell TH, Joly S, Pitre FE, Yergeau E.** 2014. Increasing phytoremediation efficiency and reliability using novel omics approaches. *Trends in Biotechnology*. 32:5
- Benjamini Y and Hochberg Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57:289–300.
- Benjamini Y and Yekutieli D.** 2001. Control of the False Discovery Rate in multiple testing under dependency. *The Annals of Statistics*. (29):4, 1165–1188.
- Beyer A and Biziuk M.** 2009. Environmental fate and global distribution of polychlorinated biphenyls. *Reviews of Environmental Contamination and Toxicology*. 201: 137-58.
- Beyer A and Biziuk M.** 2009. Environmental fate and global distribution of polychlorinated biphenyls *Reviews of Environmental Contamination and Toxicology*. vol. 201. pp. 137–158.
- Bhardwaj J, Chauhan R, Swarnkar MK, Chahota RK, Singh AK, Shankar R, Yadav SK.** 2013. Comprehensive transcriptomic study on horse gram (*Macrotyloma*

- uniflorum): de novo assembly, functional characterization and comparative analysis in relation to drought stress. *BMC Genomics*. 14,pp. 647–663.
- Borgå K, Gabrielsen GW and Skaare JU.** 2001. Biomagnification of organochlorines along a Barents Sea food chain. *Environmental Pollution*, 113:187-198.
- Bradford MM.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*. (7) 72:248-54.
- Brazier-Hicks M and Edwards R.** 2005. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant Journal*. 42, 556-566.
- Breivik K, Sweetman A, Pacyna JM, Jones KC.** 2002. Towards a global historical emission inventory for selected PCB congeners – a mass balance approach. 1. Global production and consumption. *The Science of the Total Environment*. 290: 181-198.
- Brentner LB, Mukherji ST, Walsh SA, Schnoor JL.** 2010. Localization of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) and 2,4,6-trinitrotoluene (TNT) in poplar and switchgrass plants using phosphor imager autoradiography. *Environmental Pollution*. 158(2):470-5.
- Calabrese D, Bille J and Sanglard D.** 2000. A novel multidrug efflux transporter gene of the major facilitator superfamily from *Candida albicans* (*FLU1*) conferring resistance to fluconazole. *Microbiology*. 146:2743–2754.
- Campos VM.** 2011. Tesis Doctoral: Fitorremediación de contaminantes persistentes. Una aproximación biotecnológica utilizando chopo (*Populus*) como sistema modelo. Universidad Politécnica de Madrid.
- Cao ZH, Zhang SZ, Wang RK, Zhang RF, Hao YJ.** 2013. Genome wide analysis of the apple MYB transcription factor family allows the identification of MdoMYB121 gene conferring abiotic stress tolerance in plants. *PLoS ONE*. 8. p. e69955.
- Carpenter DO.** 2006. Polychlorinated biphenyls (PCBs): routes of exposure and effects on human health. *Rev Environ Health*. 21(1):1-23.
- Casado-Vela J, Cebrián A, Gómez Del Pulgar MT, Lacal JC.** 2011. Approaches for the study of cancer: towards the integration of genomics, proteomics and metabolomics. *Clinical and Translational Oncology*. 13(9):617-28.
- Chaikam V and Karlson D.** 2008. Functional characterization of two cold shock domain proteins from *Oryza sativa*. *Plant Cell Environ*. 31(7):995-1006.
- Chen J, Chen B and Zhang D.** 2015. Transcript profiling of *Populus tomentosa* genes in normal, tension, and opposite wood by RNA-seq. *BMC Genomics*. 16:164
- Chibuike GU and Obiora SC.** 2014. Heavy Metal Polluted Soils: Effect on Plants and Bioremediation Methods. *Applied and Environmental Soil Science*. 12

- Chou HH and Holmes MH.** 2001. DNA sequence quality trimming and vector removal. *Bioinformatics*. 17:194-198.
- Claire-Lise M, Nathalie V.** 2012. The use of the model species *Arabidopsis halleri* towards phytoextraction of cadmium polluted soils. *New Biotechnology*. 30: 9-14.
- Clough SJ and Bent AF.** 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*. 16(6):735-43.
- Cock PJA, Fields CJ, Goto N, Heuer ML, and Rice PM.** 2010. The sanger fastq file format for sequences with quality scores, and the solexa/illumina fastq variants. *Nucleic Acids Res*, 38(6):1767–71
- Colangelo EP and Gueriot ML.** 2004. The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *Plant Cell*. 16. pp. 3400–3412.
- Cole SPC, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, Stewart AJ, Kurz EU, Duncan AMV, Deeley RG.** 1992. Overexpression of a transporter gene in a multidrug-resistant human lung-cancer cell-line. *Science*. 258: 1650–1654.
- Conesa A, Nueda MJ, Ferrer A, Talón M.** 2006. maSigPro: a Method to Identify Significantly Differential Expression Profiles in Time-Course Microarray Experiments. *Bioinformatics*. 22 (9), 1096-1102.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M.** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 15;21(18):3674-6.
- Conte SS and Lloyd AM.** 2011. Exploring multiple drug and herbicide resistance in plants- Spotlight on transporter proteins. *Plant Science*. 180 :196–203
- Cummins I, Wortley DJ, Sabbadin F, He Z, Coxon CR, Straker HE, Sellars JD, Knight K, Edwards L.** 2015. Key Role for a Glutathione Transferase in Multiple-Herbicide Resistance in Grass Weeds. *Proceedings of the National Academy of Sciences of the United States of America*. 110.15(2013): 5812–5817.
- Damerval C, De Vienne D, Zivy M and Thiellement H.** 1986. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat seedling proteins. *Electrophoresis* 7, 52–54.
- Davey MW, Graham NS, Vanholme B, Swennen R, May ST, Keulemans J.** 2009. Heterologous oligonucleotide microarrays for transcriptomics in a non-model species: a proof-of-concept study of drought stress in *Musa*. *BMC Genomics*.10. pp. 436–455.
- Davletova S, Schlauch K, Coutu J, Mittler R.** 2005. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Journal of Plant Physiology*. 139:847-856.

- Dejardin A, Leple JC, Lesage-Descauses MC et al Right.** 2004. Expressed sequence tags from poplar wood tissues--a comparative analysis from multiple libraries. *Plant Biol (Stuttg)* 6. 55-64.
- Doni S, Macci C, Peruzzi E, Arenella M, Ceccanti B, Masciandaro G.** 2012. In situ phytoremediation of a soil historically contaminated by metals, hydrocarbons and polychlorobiphenyls. *Journal of Environmental Monitoring*. 14:1383–90.
- Doty S.** 2015. Improving crop growth, biomass production and phytoremediation using endophytes of poplar. *Australasian Biotechnology*. 25: 46-63.
- Eapen S, Singh S, D'Souza SF.** 2007. Advances in development of transgenic plants for remediation of xenobiotic pollutants. *BiotechnolAdv*. 25: 442-51.
- Edwards R, and Dixon DP.** 2005. Plant glutathione transferases. *Methods in Enzymology*. 401: 169–186.
- Ekman DR, Wolfe NL, Dean JF.** 2005. Gene expression changes in *Arabidopsis thaliana* seedling roots exposed to the munition hexahydro-1,3,5-trinitro-1,3,5-triazine. *Environmental Science & Technology*. 39, 6313-6320.
- Emerson RO and Thomas JH.** 2009. Adaptive Evolution in Zinc Finger Transcription Factors. *Plos one*.
- EPA (Agencia de Protección Ambiental de Estados Unidos).** 2014. Evaluation of health hazards by exposure to Polychlorinated biphenyls (PCB) and proposal of a health-based quality criterion for soil. The Danish Environmental Protection Agency, Copenhagen, Denmark.
- Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, Claros MG.** 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics*. 20: 11:38.
- Falgueras A, Lara J, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, and Claros MG.** 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics*. 11:38, 2010.
- FAO (Food and Agriculture Organization of the United Nations).** 2015. Soil is a non-renewable resource: Its preservation is essential for food security and our sustainable future. <http://www.fao.org/3/a-i4373e.pdf>
- Fekete Z, Rajnai Z, Nagy T, Jakab KT, Kurunczi A, Gémes K, Herédi-Szabó K, Fülöp F, Tóth GK, Czerwinski M, Loewen G, Krajcsi P.** 2015. Membrane Assays to characterize interaction of drugs with ABCB1. *J Membr Biol*. 248(6):967-77
- Fernández-González R., Yebra-Pimentel I, Martínez-Carballo E, Regueiro J, Simal-Gándara J.** 2013. Inputs of polychlorinated biphenyl residues in animal feeds. *Food Chemistry*. 140: 296–304.

- Ficko SA, Rutter A, Zeeb BA.** 2011. Effect of pumpkin root exudates on ex situ polychlorinated biphenyl (PCB) phytoextraction by pumpkin and weed species. *Environmental Science and Pollution Research*. 18:1536–43.
- Ficko SA, Rutter A, Zeeb BA.** 2011. Phytoextraction and uptake patterns of weathered polychlorinated biphenyl-contaminated soils using three perennial weed species. *J Environ Qual* 40: 1870-7.
- Ficko SA, Rutter AB, Zeeb BA.** 2010. Potential for phytoextraction of PCBs from contaminated soils using weeds. *Science of the Total Environment*. 408.3469–3476.
- Fillatti JJ, Selmer J, McCown B, Hassig B and Comai L.** 1987 Agrobacterium mediated transformation and regeneration of Populus. *Molecular and General Genetics*, 206:192-199.
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A.** 2010. The Pfam protein families database. *Nucleic Acids Research*. 36(1)281-288.
- Fode B, Siemsen T, Thurow C, Weigel R, Gatz C.** 2008. The Arabidopsis GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. *The Plant Cell* 20, 3122–3135.
- Furring A, Manara A and Dal Corso G.** 2015. Editorial: Environmental phytoremediation: plants and microorganisms at work. *Frontiers in plant Science*. 6:520.
- Fusaro AF, Bocca SN, Ramos RL, Barroco RM, Magioli C, Jorge VC, Coutinho TC, Rangel-Lima CM, De Rycke R, Inze D.** 2007. AtGRP2, a coldinduced nucleocytoplasmic RNA-binding protein, has a role in flower and seed development. *Planta* 225: 1339–1351.
- Gabrielsen GW, Alsos IG and Brekke B.** 1997. Undersøgelser av PCB i jord, fisk og sjøfugl i området rundt avfallsfyllingen på Jan Mayen (FBT lokalitet 1805 001). Norsk Polarinstitut Rap- port nr. 104, Oslo, Norway, 31 pp (in Norwegian).
- Gadjev I, Vanderauwera S, Gechev TS, Laloi C, Minkov IN, Shulaev V, Apel K, Inze D, Mittler R. and Van Breusegem F.** 2006. Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis. *Plant Physiology* 141, 436–445.
- Gaedeke N, Klein M, Kolukisaoglu U, Forestier C, Muller A, Ansong M, Becker D, Mamnun Y, Kuchler K, Schulz B, Mueller-Roeber B, and Martinoia E.** 2001. The Arabidopsis thaliana ABC transporter AtMRP5 controls root development and stomata movement. *EMBO J*. 20. 1875 –1887.
- Garbisu C and Alkorta I.** 2001. Phytoextraction: a cost-effective plant-based technology for the removal of metals from the environment. *Bioresource Technology*. 77(3):229-36.

- Gasperi CD, Colton J and Greyell C.** 2014. Science and Technical Support Section, Water and Land Resources Division, Department of Natural Resources and Parks. Seattle, Washington. King County. Development of Lake Washington PCB Fate and Bioaccumulation Models.
- Gerhardt KE, Huang XD, Glick BR., Greenberg BM.** 2009. Phytoremediation and Rhizoremediation of Organic Soil Contaminants: Potential and Challenges. *Plant Science*. 176(1) 20–30,
- Gershenson J, Grima-Pettenati J, Seguin A, MacKay J.** 2010. Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *J. Exp. Bot.* 61.pp. 3847–3864.
- Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin, JZ, & Gnirke A.** 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol*, 13 (3) R23.
- Golldack D, Li C, Mohan H, Probst N.** 2014. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front Plant Sci.* 5, 151.
- Golldack D, Luking I, Yang O.** 2011. Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep.* 30. pp. 1383–1391.
- Gomes HI, Dias-Ferreira C, Ribeiro AB.** 2013. Overview of in situ and ex situ remediation technologies for PCB-contaminated soils and sediments and obstacles for full-scale application. *Science of the Total Environment*. 237–260.
- Gomes HI.** 2012. Phytoremediation for bioenergy: challenges and opportunities. *Environmental Science and Technology*. 1–8.
- Gonzalez A, et al.** 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 53. pp. 814–827.
- González-Gayte I, Bautista-Moreno R, Claros MG.** 2015. Capítulo de libro. DEgenes Hunter A self-customised gene expression analysis workflow for non-model organisms. *Bioinformatics and Biomedical Engineering. Tools for Next Generation Sequencing Data Analysis. Part II.* 313- 321.
- Guerrero- Fernández D.** 2015. Plataforma de supercomputación para Bioinformática. Tesis doctoral. Universidad de Málaga.
- Hanahan D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*. 166(4):557-80.
- Hardie DG.** 2007. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nat Rev Mol Cell Biol.* 8(10):774-85.

- Harris MA, JI Clark, A Ireland, J Lomax, M Ashburner, R Collins, K Eilbeck, S Lewis, et al.** 2006. The Gene Ontology (GO) project in 2006. *Nucleic Acids Research*. 34:D322–D326.
- Hartley JL, Temple GF and Brasch MA.** 2000. DNA cloning using in vitro site-specific recombination. *Genome Res*. 10(11):1788-95
- Hayashi K, Schoonbeek H. and De Waard MA.** 2002. Bcmfs 1, a novel major facilitator superfamily transporter from *Botrytis cinerea*, provides tolerance towards the natural toxic compounds camptothecin and cercosporin and towards fungicides. *Applied and Environmental Microbiology* 68:4996-5004.
- Herbert M Sauro and Frank T Bergmann.** 2008. Standards and ontologies in computational systems biology. *Essays in biochemistry*. 45:211–222.
- Higgins CF.** 1992. ABC transporters: from microorganisms to man. *Annual review of cell biology*. 8:67-113.
- Hirayama T, and Shinozaki K.** 2010. Research on plant abiotic stress responses in the postgenome era: past, present and future. *Plant J*. 61:1041-1052.
- Hoekstra PF.** 2002. Organochlorine data on anadromous and marine fish from Alaska and the Western Canadian Arctic. Un- published data. National Water Research Institute, Burlington, ON, Canada
- Höfgen R and Willmitzer L.** 1988. Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Research*. 16(20): 9877.
- Howorka S and Siwy Z.** 2009. Nanopore analytics: sensing of single molecules. *Chemical Society reviews*. 38(8):2360–2384.
- Huang J, Yang X, Wang MM, Tang HJ, Ding LY, Shen Y, Zhang HS.** 2007 A novel rice C2H2-type zinc finger protein lacking DLN-box/EAR-motif plays a role in salt tolerance *Biochim Biophys. Acta*. 1769 pp. 220–227
- Iimura Y, Ikeda S, Sonoki T, Hayakawa T, Kajita S, Kimbara K, Tatsumi K, Katayama Y.** 2002. Expression of a gene for Mn-peroxidase from *Coriolus versicolor* in transgenic tobacco generates potential tools for phytoremediation. *Appl Microbiol Biotechnol*. 59:246–251.
- Iimura Y, Yoshizumi M, Sonoki T, Uesugi M, Tatsumi K, Horiuchi KI, Kajita S, Katayama Y.** 2007. Hybrid aspen with a transgene for fungal manganese peroxidase is a potential contributor to phytoremediation of the environment contaminated with bisphenol A. *J Wood Sci*. 53:541– 544.
- Ilyas M, Ahmad S, Faheem M, Naseer MI, Kumosani TA, Al-Qahtani MH, Gari M, Ahmed F.** 2015. Next Generation Sequencing of Acute Myeloid Leukemia: Influencing Prognosis. *BMC Genomics*. 16(1):5.

- IRIS (Integrated Risk Information System).** 2012. Polychlorinated Biphenyls (PCBs). U.S. Environmental Protection Agency. <http://www.epa.gov/IRIS/subst/0294.htm>.
- Isebrands JG and Richardson J.** 2015. Poplars and Willows: Trees for Society and the Environment. FAO. 1-634.
- Ito H and Gray WM.** 2006. A gain-of-function mutation in the Arabidopsis pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides. *Plant Physiology*. 142, 63–74.
- James CA, Xin G, Doty SL, Strand SE.** 2008. Degradation of low molecular weight volatile organic compounds by plants genetically modified with mammalian cytochrome P450 2E1. *Environmental Science & Technology*. 42, 289–293.
- Jenkins Y, Markovtsov V, Lang W, Sharma P, Pearsall D, Warner J, Franci C, Huang B, Huang J, Yam GC, Vistan JP, Pali E, Vialard J, Janicot M, Lorens JB, Payan DG, Hitoshi Y.** 2005. Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Molecular Biology of the Cell* Vol. 16, 5621–5629.
- Jia J, Fu J, Zheng J, Zhou X, Huai J, Wang J, Wang M, Zhang Y, Chen X, Zhang J, Zhao J, Su Z, Lv Y, Wang G.** 2006. Annotation and expression profile analysis of full-length cDNAs from stress-induced maize (*Zea mays*L.) seedlings. *Plant J.* 48. pp. 710–727.
- Jiang Y, Peng D, Bai LP, Ma H, Chen LJ, Zhao MH, Xu ZJ, Guo ZF.** 2013. Molecular switch for cold acclimation—atomy of the cold-inducible promoter in plants. *Biochemistry (Mosc.)* 78. pp. 342–354.
- Kacálková L and Tlustoš P.** 2011. The uptake of persistent organic pollutants by plants. *Central European Journal of Biology*. 6:223–35.
- Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tsihchlis PN, Makris AM.** 2000. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. *Journal of Biological Chemistry*. 275: 29207–29216.
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, and Tanabe M.** 2014. Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*. 42(D1).
- Karlson D and Imai R.** 2003. Conservation of the cold shock domain protein family in plants. *Plant Physiology*. 131, 12–15.
- Karlson D, Nakaminami K, Toyomasu T, Imai R.** 2002. A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. *Journal of Biological Chemistry*. 277, 35248–35256.
- Kielbowicz-Matuk A.** 2011. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. *Plant Science*. 185–186. 78–85.

- Kim HR, Chae KS, Han KH, Han DM.** 2009. The nsdC gene encoding a putative C2H2-type transcription factor is a key activator of sexual development in *Aspergillus nidulans* Genetics, 182. pp. 771–783.
- Kim MH, Sasaki K and Imai R.** 2009. Cold shock domain protein 3 regulates freezing tolerance in *Arabidopsis thaliana*. Journal of Biological Chemistry 284: 23454–23460.
- Kim MH, Sonoda Y, Sasaki K, Kaminaka H, Imai R.** 2013. Interactome analysis reveals versatile functions of *Arabidopsis* COLD SHOCK DOMAIN PROTEIN 3 in RNA processing within the nucleus and cytoplasm. Cell Stress and Chaperones, in press.
- Kircher M and Kelso J.** 2010. High through put DNA sequencing—concepts and limitations. BioEssays : news and reviews in molecular, cellular and developmental biology. 32(6):524–536.
- Kohno K, Izumi H, Uchiumi T, Ashizuka M, Kuwano M.** 2003. The pleiotropic functions of the Y-box-binding protein, YB-1. Bioessays. 25, 691–698.
- Komives T, and Gullner G.** 2005. Phase I xenobiotic metabolic systems in plants. Zeitschrift für Naturforschung. 60: 179–185.
- Kyte J and Doolittle R.** 1982. A Simple Method for Displaying the Hydropathic Character of a Protein. J. Mol Biol. 157. 105-132.
- Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 15: 227(5259):680-5.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E.** 2012. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research. 40, D1202–D1210.
- Landy A.** 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annual Review of Biochemistry. 58:913-49.
- Langmead B and Salzberg S.** 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357- 359.
- Lauby-Secretan B, Loomis D, Grosse Y, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Baan R, Mattock H, Straif K.** 2013. Carcinogenicity of polychlorinated biphenyls and polybrominated biphenyls. Lancet Oncology. 14:287.
- Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN.** 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 26(4), 493-500
- Li C, Ng CK, Fan YL. MYB.** 2015. Transcription factors, active players in abiotic stress signaling. Environ. Exp. Bot. 114, 80–91.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map (SAM) Format and SAMtools. *Bioinformatics*.
- Li R, Li Y, Kristiansen K, and Wang J.** 2008. SOAP: Short oligonucleotide alignment program. *Bioinformatics*. 24(5):713–714.
- Li S and Chou HH.** 2004. LUCY2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics*. 20:3657-3665.
- Liang YK, et al.** 2005. AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana*. *Current Biology*, 15. pp. 1201–1206.
- Lim CJ, Yang KA, Hong JK, Choi JS, Yun DJ, Hong JC, Chung WS, Lee SY, Cho MJ, Lim CO.** 2006. Gene expression profiles during heat acclimation in *Arabidopsis thaliana* suspension-culture cells. *J. Plant Res.* 119. pp. 373–383.
- Liu K and Dong LJ.** 2012. Research on Cloud Data Storage Technology and Its Architecture Implementation.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, and Law M.** 2012. Comparison of next generation sequencing systems.
- Liu X, Li X, Chermaine Ong SM, Chu Z.** 2013. Progress of Phytoremediation: Focus on New Plant and Molecular Mechanism. *J Plant Biol Soil Health*. 1(1): 5.
- Livak KJ and Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402–408.
- Lockhart D and Winzeler E.** 2000. Genomics, Gene Expression and DNA Arrays. *Nature* 415:827-836.
- Loughlin FE, Gebert LF, Towbin H, Brunschweiler A, Hall J, Allain FH.** 2012. Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nature Structural & Molecular Biology*. 19:84–89.
- Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15,550.
- Lu YP, Li ZS, Drozdowicz YM, Hortensteiner S, Martinoia E, Rea PA.** 1998. AtMRP2, an *Arabidopsis* ATP binding cassette transporter able to transport glutathione S-conjugates and chlorophyll catabolites: Functional comparisons with AtMRP1. *Plant Cell*. 10:267-282.
- Maestri E and Marmioli N.** 2011. Transgenic plants for phytoremediation. *International Journal of Phytoremediation*. 13:264–79.
- Maitra RD, Kim J and Dunbar WB.** 2012. Recent advances in nanopore sequencing. *Electrophoresis*, 33: 3418–3428.

- Malisch R and Kotz A.** 2014. Dioxins and PCBs in feed and food-review from European perspective. *Science of the Total Environment*. 1:491-492.
- Marmiroli N and McCutcheon SC.** 2003. Making phytoremediation a successful technology. In: Mc-Cutcheon SC, Schnoor JL. *Phytoremediation: transformation and control of contaminants*. Hoboken (NJ): Wiley-Interscience, Inc.75-107.
- Martinoia E, Grill E, Tommasini R, Kreuz K, Amrhein N.** 1993. ATP-dependent glutathione *S*-conjugate export pump in the vacuolar membrane of plants. *Nature*. 364: 247–249.
- Masotti A and Preckel T.** 2006. Analysis of small RNAs with the Agilent 2100 Bioanalyzer. *Nature Methods*. 1-3.
- McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Ye ZH.** 2010. The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiol*. 51:1084-1090.
- McMillan BJ and Bradfield CA.** 2007. The aryl hydrocarbon receptor sans xenobiotics: endogenous function in genetic model systems. *Molecular Pharmacology* 72, 487–498.
- Meggo RE and Schnoor JL.** 2013. Cleaning Polychlorinated Biphenyl (PCB) Contaminated Garden Soil by Phytoremediation. *Environmental Sciences*. 1(1), 33–52.
- Metzker ML.** 2010. Sequencing technologies the next generation. *Nature reviews genetics*. 11(1). 31-46.
- Meuser H.** 2014. Soil remediation and rehabilitation. Treatment of contaminated and disturbed land. Series: Environmental Pollution, vol. 23, Springer Netherlands
- Miller JR, Koren S, and Sutton G.** 2010. Assembly algorithms for next-generation sequencing data. *Genomics*. 95(6):315–327.
- Minoche AE, Dohm JC and Himmelbauer H.** 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems.
- Mochida K and Shinozaki K.** 2011. Advances in Omics and Bioinformatics Tools for System Analyses of Plants Functions *Plants Cell Physiol*. 52(12): 2017-2038
- Moore M, and Ullman C.** 2003. Recent developments in the engineering of zinc finger proteins. *Brief Funct Genomic Proteomic*. 1(4):342-55.
- Morant M, Baky S, Lindberg B, Werck- Reichhart D.** 2003. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current Opinion in Biotechnology*. 14, 151-162.

- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 5(7):621-8.
- Mulder NJ, and Apweiler R.** 2008. The InterPro database and tools for protein domain analysis.
- Mylona PV, Polidoros AN and Scandalios JG.** 2007. Antioxidant gene responses to ROS-generating xenobiotics in developing and germinated scutella of maize. 58(6): 1301-1312.
- Nakaminami K, Karlson DT, Imai R.** 2006. Functional conservation of cold shock domains in bacteria and higher plants. *Proceedings of the National Academy of Sciences of the United States of America*. 103, 10122–10127.
- Nevado B., Ramos-Onsins SE, and Perez-Enciso M.** 2014. Resequencing studies of nonmodel organisms using closely related reference genomes: Optimal experimental designs and bioinformatics approaches for population genomics. *Molecular Ecology*. 23(7):1764–1779.
- Newman LJ, et al.** 2004. Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. *Plant J*. 37. pp. 239–250.
- Nueda MJ, Tarazona S, Conesa A.** 2014. Next maSigPro: updating maSigPro Bioconductor package for RNA-seq time series. *Bioinformatics*. 30; 18:1-5-
- Ockenden WA, Breivik K, Meijer SN, Steinnes E, Sweetman AJ, Jones KC.** 2002. The global re-cycling of persistent organic pollutants is strongly retarded by soils. *Environmental Pollution*. 121:71-80.
- Ohkama-Ohtsu N, Zhao P, Xiang CB, Oliver DJ.** 2007. Glutathione conjugates in the vacuole are degraded by gamma-glutamyl transpeptidase GGT3 in *Arabidopsis*. *Plant Journal*. 49, 878-888.
- Oller ALW, Agostini E, Talano MA, Capozucca C, Milrad SR, Tigier HA, Medina MI.** 2005. Overexpression of a basic peroxidase in transgenic tomato (*Lycopersicon esculentum* Mill. cv. Pera) hairy roots increases phytoremediation of phenol. *Plant Sci*. 169:1102–1111.
- Olsen L, Oostenbrink C, Jørgensen FS.** 2015. Prediction of cytochrome P450 mediated metabolism. *Adv Drug Deliv Rev*. 23; 86:61-71.
- Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML.** 2013. MYB10 and MYB72 are required for growth under iron-limiting conditions. *PLoS Genet*. 9.pp. e1003953.
- Panz K and Miksch.** 2012. K. Phytoremediation of explosives (TNT, RDX, HMX) by wild-type and transgenic plants. *J Environ Manage*. 113: 85-92.

- Pao SS, Paulsen IT and Saier MH.** 1998. Major facilitator superfamily, Microbiol. Mol. Biol. 62. 1–34[63]
- Park SJ, Kwak KJ, Oh TR, Kim YO, Kang H.** 2009. Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. Plant and Cell Physiology 50: 869–878.
- Passatore L, Rossetti S, Juwarkar AA, Massacci A.** 2014. Phytoremediation and bioremediation of polychlorinated biphenyls (PCBs): state of knowledge and research perspectives. Journal of Hazardous Materials. 278 189–202.
- Pereira SS, Guimaraes FC, Carvalho JF, Stolf-Moreira R, Oliveira MC, Rolla AA, Farias JR, Neumaier N, Nepomuceno AL.** 2011. Transcription factors expressed in soybean roots under drought stress. Genet. Mol. Res. 10. pp. 3689–3701.
- Pereira SS, Guimaraes FC, Carvalho JF, Stolf-Moreira R, Oliveira MC, Rolla AA, Farias JR, Neumaier N, Nepomuceno AL.** 2011. Transcription factors expressed in soybean roots under drought stress. Genet. Mol. Res. 10. pp. 3689–3701.
- Pilon-Smits E.** 2005. Phytoremediation.. Annual Review of Plant Biology. 56 :15–39.
- Prado-Lopez S, Conesa A, Armiñán A, Martínez-Losa M, Escobedo-Lucea C, Gandia C, Tarazona S, Melguizo D, Blesa D, Montaner D, Sanz-González S, Sepúlveda P, Götz S, O'Connor J E, Moreno R, Dopazo J, Burks DJ and Stojkovic M.** 2010. Hypoxia Promotes Efficient Differentiation of Human Embryonic Stem Cells to Functional Endothelium. STEM CELLS. 28: 407–418.
- Preston J, et al.** 2004. AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. Plant J. 40. pp. 979–995.
- Proudfoot N, Furger A and Dye M.** 2002. Integrating mRNA Processing with Transcription. Cell 108: 501-512.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP and Gu Y.** 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics. 13(1):341.
- Quick J, Quinlan AR, and Loman NJ.** January 2014. A reference bacterial genome dataset generated on the MinION (TM) portable single-molecule nanopore sequencer. GigaScience. 3(1):22.
- Raj D. Maitra, Jungsuk Kim, and William B. Dunbar.** 2012. Recent advances in nanopore sequencing.
- Rakesh Y, Pooja A, Sandeep K, Ashok C.** 2010. Perspectives for genetic engineering of poplars for enhanced phytoremediation abilities. Ecotoxicology 19:1574-1588.
- Ramel F, Sulmon C, Cabekki-Hurtadi F, Magniette ML, Renou JP, Amrani A, Couée I, Gouesbet G.** 2007. Genome-wide interacting effects of sucrose and

herbicide-mediated stress in *Arabidopsis thaliana*: novel insights into atrazine toxicity and sucrose-induced tolerance. BMC Genomics. 8:450..

Ramel F, Sulmon C, Serra AA, Gouesbet G and Coué I. 2012. Xenobiotic sensing and signalling in higher plants. Journal of Experimental Botany. Page 1 of 16.

Ranjan, Pandey N, Lakhwani D, Dubey NK, Pathre UV, Sawant SV. 2012. Comparative transcriptomic analysis of roots of contrasting *Gossypium herbaceum* genotypes revealing adaptation to drought. BMC Genomics. 13. pp. 680–702.

Rao MR, Halfhill MD, Abercrombie LG, Ranjan P, Abercrombie JM, Gouffon JS, Saxton AM, Stewart C. 2009. Phytoremediation and phytosensing of chemical contaminants, RDX and TNT: identification of the required target genes. Functional and Integrative Genomics. 9:537-547.

Rea PA. 2007. Plant ATP binding cassette transporters. Annual Review of Plant Physiology. 58:347–375.

Reddy VS, Shlykov MA, Castillo R, Sun EI, and Saier MH. 2012. The major facilitator superfamily (MFS) revisited. The FEBS journal. 279:2022-2035. type of sequence read. BMC Bioinformatics, 11:38.

Reichenauer TG, Germida JJ. 2008. Phytoremediation of organic contaminants in soil and groundwater. ChemSusChem. 1:708-17.

Rensink WA, Iobst S, Hart A, Stegalkina S, Liu J, Buell CR. 2005. Gene expression profiling of potato responses to cold, heat, and salt stress. Funct. Integr. Gen. 5. pp. 201–207.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 43(7).

Rius SP, Grotewold E, Casati P. 2012. Analysis of the P1 promoter in response to UV-B radiation in allelic variants of high-altitude maize. BMC Plant Biol. 12.pp. 92–106.

Rizza S, Conesa A, Juarez J, Catara A, Navarro L, Duran-Vila N, Ancillo G. 2012. Microarray analysis of Etrog citron (*Citrus medica* L.) reveals changes in chloroplast, cell wall, peroxidase and symporter activities in response to viroid infection. Mol Plant Pathol. Oct. 13(8):852-64.

Robinson MD and Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology. 11-25.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26,1.

Rushton PJ, Somssich IE, Ringler P, Shen QJ. 2010. WRKY transcription factors. Trends Plant Sci.15(5):247-58.

- Saiyood S, Inthorn D, Vangnai AS, Thiravetyan P.** 2013. Phytoremediation of bisphenol A and total dissolved solids by the mangrove plant, *Bruguiera gymnorhiza*. *Int J Phytoremediation*. 15: 427-38.
- Salson M, Lecroq T, Leonard M, and Mouchard L.** 2008. Burrows-Wheeler Transform. *Theoretical Computer Science* accepted 2009, pages 13–25.
- Santucci L, Bruschi M, Ghiggeri GM, Candiano G.** 2015. The Latest Advancements in Proteomic Two-dimensional Gel Electrophoresis Analysis Applied to Biological Samples. *Clinical Proteomics*. 1243: 103-125
- Sasaki K, Kim MH, Imai R.** 2007. Arabidopsis Cold Shock Domain Protein2 is a RNA chaperone that is regulated by cold and developmental signals. *Biochemical and Biophysical Research Communications*. 364, 633–638.
- Sauro HM and Bergmann FT.** 2008. Standards and ontologies in computational systems biology. *Essays in biochemistry*. 45: 211–222.
- Schadt EE, Turner S, and Kasarskis A.** 2010. A Window into Third Generation Sequencing. *Human molecular genetics*. 19(2):227–240.
- Schäffer A, Messener F, Langebartels C, Sanderman H.** 2002. Genes and enzymes for in- plant phytoremediation of air waste and soils. *Acta Biotechnologica*. 22, 141-152.
- Scheetz TE, Trivedi N, Roberts CA, Kucaba T, Berger B, Robinson NL, Birkett CL, Gavin AJ, O'Leary B, et al.** 2003. ESTprep: preprocessing cDNA sequence reads. *Bioinformatics*. 19:1318-1324.
- Schneeberger K1, Ossowski S, Ott F, Klein JD, Wang X, Lanz C, Smith LM, Cao J, Fitz J, Warthmann N, Henz SR, Huson DH, Weigel D.** 2011. Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc Natl Acad Sci U S A*. 108(25):10249-54.
- Schumann U, Prestele J, O'Geen H, Brueggeman R, Wanner G, Gietl C.** 2007. Requirement of the C3HC4 zinc RING finger of the Arabidopsis PEX10 for photorespiration and leaf peroxisome contact with chloroplasts. *Proceedings of the National Academy of Sciences*. 104(3), 1069-1074.
- Secher C, Lollier M, Jézéquel K, Cornu JY, Amalric L, Lebeau T.** 2013. Decontamination of a Polychlorinated Biphenyls-Contaminated Soil by Phytoremediation-Assisted Bioaugmentation. *Biodegradation*. 24 (4)549– 562.
- Shannon P, Markiel A and Ozier O.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13(11): 2498-504
- Shiquan Wu and Jing Li.** 2007. Comparative Analysis of Gene-Coexpression Networks Across Species a Springer-Verlag Berlin Heidelberg. ISBRA. LNBI 4463, pp. 615–626, 2007.

- Singh A, Kuhad RC, Ward OP.** 2009. Biological remediation of soil: an overview of global market and available technologies, in: *Advances in Applied Bioremediation*, Springer, and Berlin Heidelberg. 1–19.
- Singh K, Foley RC, Oñate-Sanchez L.** 2002. Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology*. 5:430–436.
- Singh OV and Jain RK.** 2003. Phytoremediation of toxic aromatic pollutants from soil. *Applied Microbiology and Biotechnology*. 63:128–135.
- Sonoki T, Kajita S, Ikeda S, Uesugi M, Tatsumi K, Katayama Y, Iimura Y.** 2005. Transgenic tobacco expressing fungal laccase promotes the detoxification of environmental pollutants. *Appl Microbiol Biotechnol*. 67:138–142.
- Sperotto RA, Boff T, Duarte GL, Fett JP.** 2008. Increased senescence-associated gene expression and lipid peroxidation induced by iron deficiency in rice roots. *Plant Cell Rep*. 27. pp. 183–195.
- Sperotto RA, Boff T, Duarte GL, Fett JP.** 2008. Increased senescence-associated gene expression and lipid peroxidation induced by iron deficiency in rice roots. *Plant Cell Rep*. 27. pp. 183–195.
- Srogi K.** 2008. Levels and congener distributions of PCDDs, PCDFs and dioxin-like PCBs in environmental and human samples: a review. *Environ Chem Lett* 6:1–28.
- Stein L.** 2001. Genome annotation: from sequence to biology. *Nature reviews. Genetics*. 2(7):493–503.
- Stracke R, et al.** 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J*. 50. pp. 660–677.
- Stracke R.** 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol*. 4. pp. 447–456.
- Stuart JM, Segal E, Koller D, Kim SK.** 2003. A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science*. 302: 249–255.
- Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP.** 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 25; 102(43): 15545–15550.
- Sun SJ, Guo SQ, Yang X, Bao YM, Tang HJ, Sun H, Huang J, Zhang HS.** 2010. Functional analysis of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice *J. Exp. Bot.*, 61. pp. 2807–2818.
- Sun Y, Zhou Q, Xu Y, Wang L, Liang X.** 2011. Phytoremediation for cocontaminated soils of benzo[a]pyrene (B[a]P) and heavy metals using ornamental plant *Tagetes patula*. *J Hazard Mater* 28: 2075–82.

- Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin YC, Sjodin A, Van de Peer Y, Stefan J, Torgeir R. Hvidsten, Nathaniel R. Street.** 2015. The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytologist*. 1-8.
- Sylvestre M, Macek T, Mackova M.** 2009. Transgenic plants to improve rhizoremediation of polychlorinated biphenyls (PCB's). *Current Opinion in Biotechnology*. 20, 242–247.
- Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24(8):1596-9.
- Tang-Péronard JL, Heitmann BL, Andersen HR, Steuerwald U, Grandjean P, Weihe P, Jensen TK.** 2014. Association between prenatal polychlorinated biphenyl exposure and obesity development at ages 5 and 7 y: a prospective cohort study of 656 children from the Faroe Islands. *American Journal of Clinical Nutrition*. 99(1):5-13
- Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A.** 2011. Differential expression in RNA-seq: a matter of depth. *Genome Research*. 21(12),4436.
- Taylor NF, Dodd KJ, Baker RJ, Willoughby K, Thomason P, Graham HK.** 2013. Progressive resistance training and mobility-related function in young people with cerebral palsy: a randomized controlled trial. *Developmental Medicine & Child Neurology*. 55(9):806-12.
- Thapa G, Sadhukhan A, Panda SK, Sahoo L.** 2012. Molecular mechanistic model of plant heavy metal tolerance. *Biometals*. 25: 489-505. 26.
- The Sam and Format Specification.** 2011. The SAM Format Specification (v1.4-r985). Read, pages 1–11.
- Thelen JJ and Peck SC.** 2007. Quantitative Proteomics in Plants: Choices in Abundance. *Plant Cell*. 19(11): 3339–3346.
- Tian DQ, Pan XY, Yu YM, Wang WY, Zhang F, Ge YY, Shen XL, Shen FQ, Liu XJ.** 2013. De novo characterization of the *Anthurium* transcriptome and analysis of its digital gene expression under cold stress. *BMC Genomics*. 14.pp. 827–841.
- Tian DQ, Pan XY, Yu YM, Wang WY, Zhang F, Ge YY, Shen XL, Shen FQ, Liu XJ.** 2013. De novo characterization of the *Anthurium* transcriptome and analysis of its digital gene expression under cold stress. *BMC Genomics*. 14.pp. 827–841.
- Tian ZD, Zhang Y, Liu J, Xie CH.** 2010. Novel potato C2H2-type zinc finger protein gene, StZFP1, which responds to biotic and abiotic stress, plays role in salt tolerance *Plant Biol*. 12. pp. 689–697.
- Truua J, Truu M, Espenberg M, Nölvak H, Juhanson J.** 2015. Phytoremediation and Plant-Assisted Bioremediation in Soil and Treatment Wetlands: A Review. *The Open Biotechnology Journal*, Volume 9.

- Tsukagoshi H, Busch W and Benfey PN.** 2010. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. *Cell* 143, 606–616.
- Tu C, Teng Y, Luo YM, Pan C, Sun XH, Li ZG.** 2010. Phytoremediation of polychlorinated biphenyls contaminated soil by leguminosae-gramineae intercropping: a field trial. 31(12):3062-6.
- Ulrich RG.** 2003. The toxicogenomics of nuclear receptor agonists. *Current Opinion in Chemical Biology* 7, 505–510.
- UNEP (Programa de Naciones Unidas para el Medio Ambiente).** 2014. Chemicals, IOMC, Guidelines for the identification of PCBs and Materials Containing PCBs, First Issue.
- Van Aken B, Correa PA, Schnoor JL.** 2009. Phytoremediation of polychlorinatedbiphenyls: new trends and promises. *Environmental Science and Technology*. 44:2767–2776.
- Van Aken B, Yoon JM, Schnoor JL.** 2004. Biodegradation of nitro-substituted explosives 2,4,6-trinitrotoluene, hexahydro- 1,3,5-trinitro-1,3,5-triazine, and octahydro- 1,3,5,7-tetranitro-1,3,5-tetrazocine by a phytosymbiotic *Methylobacterium* sp. (1):508-17.
- Van den Berg M, Birnbaum L, Bosveld A, Brunström B, Cook P, Feeley M, Giesy JP, Hanberg A, Hasegawa R, Kennedy SW.** 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife, *Environ Health Perspect.* 106, 775.
- Vélez P, Izquierdo I, Rosa I, García A.** 2015. A 2D-DIGE-based proteomic analysis reveals differences in the platelet releasate composition when comparing thrombin and collagen stimulations *Scientifics reports*. 5: 8198.
- Verbruggen N, Hermans C, Schat H.** 2009. Molecular mechanisms of metal hyperaccumulation in plants. *New Phytologist*. 181: 759-76.
- Wagner U, Edwards R, Dixon DP, Mauch F.** 2002. Probing the diversity of the *Arabidopsis* glutathione-S-transferase gene family. *Plant Molecular Biology*. 49:515-532.
- Wang GD, Li QJ, Luo B, Chen XY.** 2004. Ex planta phytoremediation of trichlorophenol and phenolic allelochemicals via an engineered secretory laccase. *Nat Biotechnol*. 22:893–897.
- Wang W, Huang MJ, Zheng JS, Cheung KC, Wong MH.** 2013. Exposure assessment and distribution of polychlorinated biphenyls (PCBs) contained in indoor and outdoor dusts and the impacts of particle size and bioaccessibility. *Science Of The Total Environment*. 463:1201–1209.
- Wang Y, Dou D, Wang X, Li A, Sheng Y, Hua C, Cheng B, Chen X, Zheng X, Wang Y.** 2009. The PsCZF1 gene encoding a C2H2 zinc finger protein is required for

growth, development and pathogenesis in *Phytophthora sojae* Microb. Pathog. 47. pp. 78–86.

- Wang Y, Peng X, Xu W, Luo YB, Zhao W, Hao J, Liang Z, Zhang Y, Huang KL.** 2012. Transcript and protein profiling analysis of OTA-induced cell death reveals the regulation of the toxicity response process in *Arabidopsis thaliana*. Journal of Experimental Botany. 63, 2171–2187.
- Wang Y, Zhou B, Sun M, Li Y, Kawabata S.** 2012. UV-A light induces anthocyanin biosynthesis in a manner distinct from synergistic blue+ UV-B light and UV-A/blue light responses in different parts of the hypocotyls in turnip seedlings. Plant Cell Physiol. 53. pp. 1470–1480.
- Weisman D, Alkio M, Colón-Carmona A.** 2010. Transcriptional responses to polycyclic aromatic hydrocarbon-induced stress in *Arabidopsis thaliana* reveal the involvement of hormone and defense signalling pathways. BMC Plant Biology. 10:59.
- White JR, Roberts M, Yorke JA, MP.** 2008. Figaro: a novel statistical method for vector sequence removal. Bioinformatics. 24:462-467.
- Windsor B, Roux SJ, Lloyd A.** 2003. Multiherbicide tolerance conferred by AtPgp1 andapyrase overexpression in *Arabidopsis thaliana*. Nature Biotechnology. 21: 428–433.
- Wolfe NL and Hoehamer CF.** 2003. Enzymes used by plants and microorgan- isms to detoxify organic compounds. In SC McCutcheon, JL Schnoor, eds, Phytoremediation: Transformation and Control of Contaminants. John Wiley & Sons, Hoboken, NJ, 159–187.
- Wu Shiquan and Jing Li.** 2007. Comparative Analysis of Gene-Coexpression Networks Across Species. I. Mndoiu and A. Zelikovs.ky (Eds.) 4463: 615–626.
- Xi J, Xu P, Xiang CB.** 2011. Loss of AtPDR11, a plasma membrane- localized ABC transporter, confers paraquat tolerance in *Arabidopsis thaliana*. The Plant Journal. 69, 782–791.
- Xiao H, Tang J, Li Y, Wang W, Li X, Jin L, Xie R, Luo H, Zhao X, Meng Z, He GL.** 2009. ZhuSTAMENLESS 1, encoding a single C2H2 zinc finger protein, regulates floral organ identity in rice Plant J., 59. pp. 789–801.
- Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S.** 2013. DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. Trends Plant Sci. 18, 267–276.
- Yan DH, Fenning T, Tang S, Xia X, Yin W.** 2012. Genome-wide transcriptional response of *Populus euphratica* to long-term drought stress. Plant Sci. 195. pp. 24–35.
- Yang Y and Karlson DT.** 2011. Overexpression of AtCSP4 affects late stages of embryo development in *Arabidopsis*. Journal of Experimental Botany 62: 2079–2091.

- Yang ZB, Eticha D, Albacete A, Rao IM, Roitsch T, WJ.** 2012. Horst Physiological and molecular analysis of the interaction between aluminium toxicity and drought stress in common bean (*Phaseolus vulgaris*)
- Yuan L, Perry SE.** 2011. Plant Transcription Factors: Methods and Protocols. SPRINGER SCIENCE+BUSINESS MEDIA, LLC, New York, USA, pp 307-321.
- Yun KY, Park MR, Mohanty B, Herath V, Xu F, Mauleon R, Wijaya E, Bajic VB, Bruskiewich R, de Los Reyes BG.** 2010. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biol.* 10. pp. 16–45.
- Zechmann B.** 2014. Compartment-specific importance of glutathione during abiotic and biotic stress. *Frontiers in Plant Science.* 5:566.
- Zeeb BA, Amphlett JS, Rutter A, Reimer KJ.** 2006. Potential for phytoremediation of polychlorinated biphenyl-(PCB)-contaminated soil. *International Journal of Phytoremediation.* 8:199–221.
- Zeliger HI.** 2013. Lipophilic chemical exposure as a cause of cardiovascular disease. *Interdisciplinary Toxicology.* 6(2): 55–62.
- Zeng QY, Yang CY, Ma QB, Li XP, Dong WW, Nian H.** 2012. Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. *BMC Plant Biol.* 12.pp. 182.
- Zhan X, Yuan J, Yue L, Xu G, Hu B, and Xu R.** 2015. Response of uptake and translocation of phenanthrene to nitrogen form in lettuce and wheat seedlings, *Environmental Science and Pollution Research.* vol. 22, no. 8. pp. 6280–6287.
- Zhong R, et al.** 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. *Plant Cell.* 19. pp. 2776–2792.
- Zhong R, et al.** 2008. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell,* 20. pp. 2763–2782.
- Ziarati and Alaedini.** 2014. The phytoremediation technique for cleaning up contaminated soil by *Amaranthus sp.* Parisa. *Journal of Environmental & Analytical Toxicology.* 4:2.