Molecular and Comparative Phylogenetic Analysis of the Polyphenol Oxidase Gene Family in Poplar (*Populus* spp.)

by

Lan T. Tran
BSc, University of Victoria, 2004

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

Supervisory Committee

Molecular and Comparative Phylogenetic Analysis of the
Polyphenol Oxidase Gene Family in Poplar (*Populus* spp.)

by

Lan T. Tran
BSc, University of Victoria, 2004

Supervisory Committee

Dr. C. Peter Constabel, Department of Biology
Supervisor

Dr. Robert L. Chow, Department of Biology
Departmental Member

Dr. Jürgen Ehlting, Department of Biology
Departmental Member

Abstract

Supervisory Committee

Dr. C. Peter Constabel, Department of Biology
Supervisor

Dr. Robert L. Chow, Department of Biology
Departmental Member

Dr. Jürgen Ehlting, Department of Biology
Departmental Member

Polyphenol oxidases (PPOs) are ubiquitous enzymes that oxidize phenols to quinones in the presence of molecular oxygen, often leading to tissue discolouration. They are sometimes considered as defense proteins but other functions, for example in phenolic compound biosynthesis, have also been found. In this thesis, bioinformatic searches were conducted to identify putative *PPO* genes from available genomes representing five Viridiplantae lineages: chlorophytes, bryophytes, lycophytes, monocotyledonous anthophytes and eudicotyledonous anthophytes. Duplicated *PPO* genes were found in most land plant genomes. A detailed investigation of the poplar (*Populus trichocarpa*) *PPO* gene family found nine genes that exhibit differential expression profiles during development and following stress, of which *PtrPPO1* was the only significant wound-inducible *PPO* gene. A phylogenetic reconstruction of the poplar PPOs identified PtrPPO13 to be an unusual PPO homolog and it was studied in detail. Experimental evidence indicated that *PtrPPO13* is expressed in most organs, and unlike most PPOs, is localized to the vacuole. Together, the phylogeny, gene expression and subcellular localization studies suggest that PPOs are likely to have variable physiological functions in plants and that PtrPPO13 is distinct from most typical PPOs.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AmAS1 Aureusidin Synthase

BLAST Basic Local Alignment Search Tool

CIWOG Common Introns Within Orthologous Genes

CTAB Cetyltrimethylammonium Bromide

cTP Chloroplast Transit Peptide

DOPA Dihydroxyphenylalanine

GFP Green Fluorescent Protein

Hc Hemocyanin

HCD Hydroxycinnamic Acid Derivatives

JGI Joint Genome Institute

LPI Leaf Plastochron Index

MeJA Methyl Jasmonate

NCBI National Center for Biotechnology Information

NJ Neighbour-Joining

ORF Open Reading Frame

PA Proanthocyanidins

PG Phenolic Glycosides

PPO Polyphenol Oxidase

RT-PCR Reverse Transcriptase Polymerase Chain Reaction

SMART Simple Modular Architecture Research Tool

TaT Twin arginine-dependent Translocation

TI Trypsin Inhibitor

TTD Thylakoid Transfer Domain

WGD Whole Genome Duplication

Acknowledgments

First, I would like thank Dr. C. Peter Constabel for giving me the opportunity to explore the field of plant molecular biology. Next, I would like to thank my committee, Dr. Bob Chow and Dr. Jürgen Ehlting for their commitment to this research project. Thank you to Dr. John Taylor for his enthusiasm and input into my project. I would also like to thank the past and present members of the Constabel lab for their friendship, guidance and continued support: Dr. Nicole Dafoe, Dr. Andreas Gesell, Dr. Ian Major, Dr. Robin Mellway, Dr. Manoela Miranda, Dr. Lynn Yip, Russell Chedgy, Vasko Veljanovski, Alpha Wong, Michael Zifkin, Yavor Denchev, Amy Franklin, Laura Rix, Kyle Slydell, Kevin Tam, Laura Wallace and anyone else I may have missed. Thank you to Roderick Haesevoets and his team at the Centre for Biomedical Research DNA Sequencing Facility and Shawn Salsiccioli for the biolistics assistance. To everyone in the Centre for Forest Biology and the Department of Biology, thank you for always being friendly faces. Last but not least, I would like to thank my family and friends for their encouragement and helping me see the light at the end of the tunnel.

Chapter 1. Introduction

1.1 General Introduction

Plant ecological interactions are strongly influenced by the diversity of compounds produced from plant secondary metabolism. The sessile nature of plants exposes them to numerous abiotic and biotic factors, especially in long-lived perennials, such as trees. Thus, plants have evolved a number of protein-based defenses and secondary metabolites for environmental adaptation. *Populus* (aspens, cottonwoods and poplars) is a dominant genus with multiple species, referred to here as poplars, and produces an array of phenolic compounds that impact ecosystem dynamics (Lindroth and Hwang, 1996; Tuskan et al., 2006; Whitham et al., 2006). Furthermore, it contains several highly expressed polyphenol oxidases (PPOs) (Constabel et al., 2000; Wang and Constabel, 2004a) which make poplar and its complex phenolic chemistry an interesting model for the analysis of PPOs and their potential interactions with carbon-based metabolites.

PPOs are ubiquitous copper-binding enzymes that catalyze the oxidation of phenols to quinones in the presence of molecular oxygen. The reactive quinone products spontaneously form polymers with other molecules such as proteins, and often cause the undesired discolouration of food products (Vámos-Vigyázó, 1981; Matheis and Whitaker, 1984). The first report of PPO surfaced in 1895 (Bertrand, 1896). However, a method for its isolation was not described until the late 1930s (Keilin and Mann, 1938; Kubowitz, 1938; Mayer, 2006). Since then PPOs have been purified from various

sources including peach (Wong et al., 1971), avocado (Kahn, 1976), spinach (Golbeck and Cammarata, 1981), pear (Wissemann and Montgomery, 1985), mung bean (Shin et al., 1997) and potato (Marri et al., 2003). Thus in the food and agricultural sciences, this biochemical reaction is of significant concern as it can be detrimental to the post-harvest physiology of most crop plants and can decrease the nutritional value and consumer acceptance. This has motivated researchers to find potential mechanisms to either inhibit or reduce PPO activity. For example, decreased browning has been observed in transgenic apple calli and potato tubers expressing antisense *PPO* genes (Bachem et al., 1994; Coetzer et al., 2001; Murata et al., 2000; Murata et al., 2001). However in certain processed foods such as cocoa, coffee and tea, the effects of PPO are thought to enhance organoleptic properties (van Gelder et al., 1997; Yoruk and Marshall, 2003).

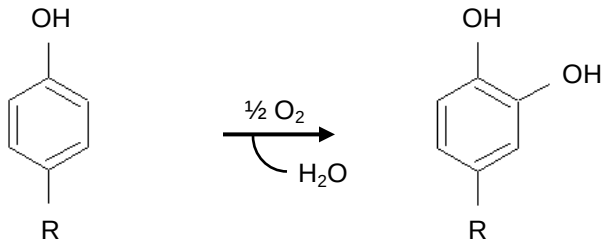
Apart from its effects in foods, the physiological functions of PPOs have also been studied, but in most cases remain enigmatic. PPOs are often considered to be defense-related proteins against insects and pathogens but other functions, in particular in biosynthesis, have also been found. Together, this suggests that the roles associated with this enzyme are more diverse than what is often reported. In the following sections, the current literature on PPOs will be presented.

1.2 Structure and Reaction Catalyzed by PPOs

In general, PPOs catalyze (i) the conversion of monophenols to *ortho*-diphenols (monophenolase, tyrosinase; EC 1.14.18.1) and/or (ii) the oxidation of *ortho*-diphenols to *ortho*-diquinones (diphenolase, catechol oxidase; EC 1.10.3.1) in the presence of molecular oxygen (Figure 1.1). In the former reaction, as typical of bacteria, fungi and

Figure 1.1 The reaction catalyzed by PPO. In the presence of molecular oxygen, PPO catalyzes (i) the conversion of monophenols to *ortho*-diphenols (monophenolase, tyrosinase; EC 1.14.18.1) and/or (ii) the oxidation of *ortho*-diphenols to *ortho*-diquinones (diphenolase, catechol oxidase; EC 1.10.3.1), with the concomitant release of water. *Ortho*-diquinones are reactive and can polymerize with other compounds to produce brown polymers (melanin). In plants, the latter reaction (ii) is typical.

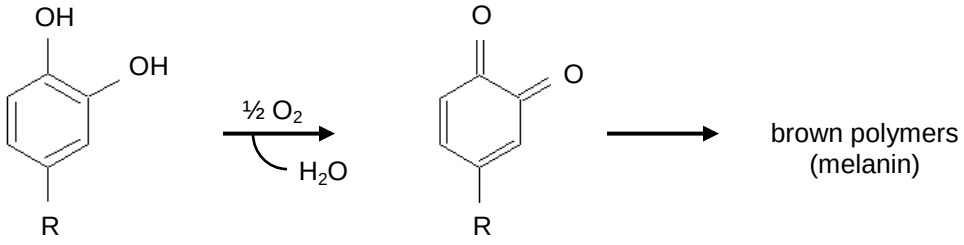
(i)



monophenol
(eg. tyrosine)

ortho-diphenol

(ii)



ortho-diphenol
(eg. catechol, DOPA)

ortho-diquinone

mammal tyrosinases, the conversion of monophenols often results in the formation of melanin (Gerdemann et al., 2002). In molluscs and some arthropods, another distantly related class of PPO-like proteins exists, but these are non-catalytic and function as oxygen transporters called hemocyanins (Hcs) (Decker et al., 2007). Collectively, PPOs and Hcs are type-3 copper proteins that share a homologous copper-binding domain (van Gelder et al., 1997; Decker et al., 2007).

In most organisms including plants, the core of the PPO polypeptide consists of two conserved copper-binding domains, designated CuA and CuB. Each domain contains three active site histidine residues that bind a copper ion. In turn, the copper ion facilitates the coordination of molecular oxygen and a phenolic substrate (van Gelder et al., 1997; Klabunde et al., 1998; Gerdemann et al., 2002). Most purified plant PPOs exhibit broad substrate specificities and oxidize common diphenolic compounds such as catechol, chlorogenic acid and L-3,4-dihydroxyphenylalanine (DOPA) to their respective quinones (Mazzafera and Robinson, 2000; Li and Steffens, 2002; Wang and Constabel, 2003). PPOs have also been found to oxidize flavonoid compounds such as catechin and epicatechin, the monomers that form proanthocyanidins (PAs), or condensed tannins (Liu et al., 2007; Munoz-Munoz et al., 2008). In rare instances, plant PPOs with monophenolase activity have been identified which oxidize tyrosine (Steiner et al., 1999; Yamamoto et al., 2001). However, substrate preferences are variable between different PPO isoforms, and depend on the plant species (Constabel and Barbehenn, 2008).

Analysis of N-terminal sequences from different plant PPOs have delineated the common presence of a chloroplast transit peptide, consisting of an N-terminal stromal domain enriched in hydroxylated amino acids, a hydrophobic thylakoid transfer domain

and a polar C-terminal domain that ends in an alanine motif (von Heijne et al., 1989; Keegstra and Cline, 1999), which suggests that most PPOs localize to the thylakoid lumen of the chloroplast. PPOs are nuclear-encoded proteins that are translated in the cytoplasm and imported across the outer and inner membranes of the chloroplast, and into the stroma. A stromal processing peptidase cleaves the transit peptide and exposes the thylakoid transfer domain (Koussevitzky et al., 1998) for subsequent import into the thylakoid lumen via translocation across the twin arginine ΔpH-dependent pathway (Robinson and Mant, 1997; Koussevitzky et al., 2008). Upon import, the thylakoid domain is cleaved at the C-terminal alanine motif.

Direct evidence of the chloroplast localization of PPOs has been demonstrated by several independent studies. An experiment by Sommer et al. (1994) traced the import of a [^{35}S]methionine-labeled 67 kDa tomato PPO precursor to the stroma, where it was processed to a 62 kDa intermediate. Upon translocation into the thylakoid lumen, it accumulated as a 59 kDa mature protein. More recently, studies fusing a dandelion PPO to a fluorescent protein revealed its plastid localization in dandelion protoplasts (Wahler et al., 2009). Chloroplast localization has been shown to be abolished in the presence of tentoxin, a cyclic peptide that affects the function of the F_1-ATPase (Groth, 2002). In mung bean seedlings, tentoxin treatment inhibited the import of PPO into the plastids and found it to accumulate outside the plastid envelope (Vaughn and Duke, 1981; Vaughn and Duke, 1984). An exception to the chloroplast targeting of PPOs however, is the vacuolar localization of aureusidin synthase (AmAS1), a PPO homolog with a biosynthetic hydroxylase function found in snapdragon (Nakayama et al., 2000; Ono et al., 2006). To date, this is the only reported plant PPO homolog with an non-plastid

localization; this presents the possibility that other PPOs may also be targeted to other cellular locales.

1.3 Potential Functions of PPOs in Abiotic and Biotic Stress Responses

1.3.1 PPOs as Insect Defense Proteins

The compartmentalization of PPOs in the chloroplast has suggested a role in photosynthesis, but so far the evidence is inconclusive (Lax and Vaughn, 1991). Instead, it is believed that in some cases the PPO-generated quinones are involved in defense against insects and pathogens. To date, the best evidence for a role of PPOs as defense proteins is from the glandular trichomes of *Solanum berthaultii* (Kowalski et al., 1992). The disruption of glandular trichomes releases their contents containing PPO and their respective substrates. Upon exposure to atmospheric oxygen, PPO activity causes the oxidative polymerization mediated by the quinones and hardening of the exudate which entraps small-bodied insects such as aphids (Kowalski et al., 1992; Yu et al., 1992).

However, to a large extent, the idea that PPO acts as an herbivore defense protein is due to a number of reports of inducible *PPO*s, such as apple *pAPO5* and hybrid poplar *PtdPPO1* (Boss et al., 1995; Constabel et al., 2000). Furthermore, PPOs have been found to be induced simultaneously with other protein-based defenses, including lipoxygenases, peroxidases and trypsin inhibitors (TIs) (Constabel et al., 1996). For example, TIs are well known anti-nutritive defensive proteins that are co-induced with PPOs in poplar (Christopher et al., 2004; Major and Constabel, 2006). In tomato plants overexpressing the mobile wound signal systemin, high levels of *PPO* transcripts were correlated with

the induction of the octadecanoid pathway (Constabel et al., 1995). Taken together, these findings suggest that in some plants, PPOs are involved in herbivore defense.

The hypothesized mechanism for PPOs in defense is as anti-nutritive proteins. Consumption of foliage by insects results in the release of PPOs and their vacuolar phenolic substrates. Consequently, quinones may be produced in the gut and modify dietary proteins by alkylation of the amino ($-NH_2$) and sulfhydryl (-SH) groups of cysteine, histidine, lysine and methionine residues. This prevents their absorption and assimilation, causing the loss of essential amino acids (Felton et al., 1992). This has been demonstrated by studies monitoring the oxidation of chlorogenic acid in the gut of beet armyworm (*Spodoptera exigua*) and other leaf-chewing insects. Protein content decreased as more phenolics were oxidized, which led to decreased larval growth (Felton et al., 1989; Felton et al., 1992). However subsequent studies have suggested that the anoxic environment of the lepidopteran midgut may limit the effectiveness of PPO (Barbehenn et al., 2007).

Studies involving tomato plants overexpressing *PPO* have further demonstrated the efficacy of PPO against insects. Common tomato pests such as cotton bollworm (*Helicoverpa armigera*), beet armyworm (*S. exigua*) and common cutworm (*S. litura*) were found to have decreased growth while feeding on foliage with increased PPO activity (Mahanil et al., 2008; Bhonwong et al., 2009). In poplar, the defense role of PPO was tested using hybrid aspen overexpressing *PPO* fed to forest tent caterpillars (*Malacosoma disstria*). Larvae from older egg masses reared on poplar foliage with increased PPO activity showed decreased growth rates (Wang and Constabel, 2004b). However, additional studies in the same hybrid aspen overexpressing *PPO* lines with

gypsy moth (*Lymantria dispar*) and white-marked tussock (*Orgyia leucostigma*) did not show a clear correlation between increased levels of PPO and insect performance (Barbehenn et al., 2007). Therefore, the efficacy of PPO as a defensive protein may depend on specific plant-insect interactions (Schmidt et al., 2005), and the strongest evidence comes from studies in tomato.

1.3.2 PPOs as Pathogen Defense Proteins

The conspicuous browning caused by quinone polymerization has also been thought to prevent the spread of pathogen infection. The previous identification of a wound-inducible *PPO* in tomato led to studies with *Alternaria solani* in which infection was found to decrease. Again, most of the work demonstrating the efficacy of PPOs in pathogen infection comes from studies using transgenic tomato. For example, *PPO* suppressed tomato lines were found to be more susceptible to *Pseudomonas syringae* infection (Thipyapong et al., 2004). In contrast, tomato plants overexpressing *PPO* showed decreased symptoms of infection (Li and Steffens, 2002).

The effect of PPO on the coffee leaf rust pathogen *Hemileia vastatrix* has also been investigated (Melo et al., 2006). Fifteen coffee genotypes with different levels of pathogen resistance were analyzed to determine the association between pathogen infection and PPO activity. The results were found to be variable and PPO activity did not correlate with pathogen resistance. Thus, it was concluded that the preexisting levels of phenolic compounds may be more of a contributing factor as a defense mechanism, rather than increased levels of PPO activity.

1.3.3 PPOs as Wound-Sealing Proteins

It has been speculated that the melanin produced by the cross-linking of PPO-generated quinones provides protection during wounding. In some latex-producing plants, latex has been identified to contain defensive proteins and thus has been postulated to be involved in a wound response (Wititsuwannakul et al., 2002). PPOs have been speculated to be present in latex from the discolouration that occurs over time (Roberts, 1971). Recent studies have identified PPOs in various latex producing plants such as opium poppy and the rubber tree, *Hevea brasiliensis*. In the latter, two PPOs were purified (Wititsuwannakul et al., 2002). In latex vesicles, PPOs and pathogenesis-related proteins are found which suggests that these proteins are constituents of a defense response. The recent characterization of dandelion PPO in laticifers suggests that they are involved in wound-sealing (Wahler et al., 2009). In two dandelion species, the expression of *PPO* was silenced using RNA interference (RNAi) in transgenic plants. Upon wounding, more latex flowed from the RNAi lines compared to the control lines. Thus in dandelion, PPO is involved in latex coagulation to repair wounds. Coincidently, PPOs in some arthropods have been described to function in wound-sealing and help polymerize the exoskeleton during wounding or molting.

1.3.4 PPOs as Phytoremediation Proteins

An intriguing example of PPOs involved in stress tolerance is the accumulation of cadmium (Cd) in aquatic plants. Cadmium is a toxic heavy metal that has severe consequences in the environment and thus, the use of aquatic plants for phytoremediation is being explored as some of these plants appear to be capable of surviving in wastewater.

Heavy metals have been found to accumulate in the epidermal glands of aquatic plants, such as water lily (Lavid et al., 2001b). Coincidently, phenolic compounds have also been found in the epidermis of these plants, which has led to the speculation of the presence of PPO in these tissues. In the semiaquatic plant *Nymphoides peltata*, PPO was found to be induced upon the accumulation of Cd in the hydropotes, or water-holding cells, of the epidermis. However, after four days of exposure to Cd, *N. peltata* was found to incur conspicuous leaf damage compared to non-exposed plants (Lavid et al., 2001a). Although the function of PPO in *N. peltata* is inconclusive, evidence suggests that PPOs in some plants may be involved in tolerance of various abiotic stresses.

1.4 PPOs as Biosynthetic Enzymes

In a few species, biosynthetic roles for PPO-like enzymes have been described in secondary metabolism. The phenylpropanoid pathway creates a diversity of phenolic compounds important, for example, in UV stress responses and as fruit and floral pigments. In snapdragon, aureusidin synthase functions in a branch of the phenylpropanoid pathway to catalyze the hydroxylation and oxidative cyclization of chalcones to aurones (Nakayama et al., 2000; Nakayama et al., 2001). Similarly, the PPO-like hydroxylase larreatricin hydroxylase (LtH) was purified from creosote bush and was shown to specifically convert the (+)-larreatricin enantiomer to (+)-3'-hydroxylarreatricin, a precursor for the potent antioxidant compound nordihydroguaiaretic acid (Cho et al., 2003).

PPOs have also been shown to be involved in the synthesis of betacyanin and betaxanthin precursors. Here, PPOs catalyze the oxidation of tyrosine in the betalain

pathway. Early studies in pokeweed showed that the accumulation of *PPO* transcripts is correlated with the appearance of betalains in ripening fruit (Joy et al., 1995). This led to speculations in other betalain-accumulating plants for the involvement of PPO in pigment synthesis. In betacyanin-accumulating portulaca, a 53 kDa PPO partially purified from callus cultures was found to hydroxylate tyrosine to produce DOPA, an early intermediate of the betalain pathway (Steiner et al., 1999; Yamamoto et al., 2001). Studies by Gandia-Herrero et al. (2005) postulated that PPO is also involved in the alternate betalain pathway that creates the yellow/orange betaxanthin pigments.

1.5 Poplar as a Model Tree

The recent availability of plant genomes now facilitates the study of PPOs as an entire gene family within a species. Poplar, as the third plant with a sequenced genome (Tuskan et al., 2006) after Arabidopsis (Arabidopsis Genome Initiative, 2000) and rice (Goff et al., 2002; Yu et al., 2002), contains a large diversity of phenolic-based secondary metabolism compounds and is suitable for studies of PPO, especially at the level of gene families. The *Populus trichocarpa* (Torr. & Gray) genome was chosen to be sequenced due to its relatively small genome size, and other attractive traits which include rapid growth, ease of genetic manipulation and propagation and amenability to *Agrobacterium*-mediated plant transformations. It is also an ideal candidate for studying physiological processes that are absent in annuals such as Arabidopsis, in particular perennial growth and wood formation.

A growing collection of genomics resources including cDNA clones, expressed sequence tags (ESTs) and microarray data sets have been established and now facilitate

the identification and characterization of genes, their gene families and their biological functions (Sterky et al., 2004; Ralph et al., 2006; Miranda et al., 2007; Nanjo et al., 2007; Rinaldi et al., 2007). The availability of the poplar genome complemented with an abundance of genomics resources has also provided the possibility of plant comparative genomics between monocotyledonous and contrasting eudicotyledonous plants. For instance, genome-wide studies have been conducted to identify homologs of phenylpropanoid genes in Arabidopsis, rice and poplar (Tsai et al., 2006; Hamberger et al., 2007; Souza et al., 2008). Together, this will also facilitate studies related to understanding the evolutionary relationships of gene families of interest.

1.6 Phenylpropanoid Metabolism in Poplar

The genus *Populus* is a dominant component of various ecosystems throughout the Northern Hemisphere. Like other long-lived perennials, poplars face many challenges for their survival, which include herbivory and pathogen infection. The diversity of phenolic and phenylpropanoid compounds in poplar, many of which are related to its adaptations to environmental conditions and stresses, provide an important rationale for studies of PPOs in this model plant. First, the carbon-based phenolic compounds are potential substrates for PPO and can lead to a diversity of quinones. Second, these compounds are potential products of PPO-catalyzed reactions, if PPO has a biosynthetic role, as in some plants. Many of the later steps in the phenylpropanoid pathway are unknown which leaves potential for PPOs to be involved.

Poplar accumulates large amounts of three major classes of phenolic compounds: salicin-based phenolic glycosides (PGs), flavonoids and hydroxycinnamic acids and their

derivatives (HCDs) (Constabel and Lindroth, 2010). Salicin-based PGs are a large class of compounds exclusive to the Salicaceae, some of which are considered to be potent antiherbivore chemicals and a constituent of poplar defense (Lindroth and Hwang, 1996). The breakdown products of the PG tremulacin, for example, have been identified as a potential PPO substrate (Haruta et al., 2001b). Flavonoids are a diverse group of compounds that include anthocyanins and polymers such as PAs. The HCDs are another group of compounds that include caffeic acid and chlorogenic acid, most of which are suitable substrates for PPOs. Collectively, PGs and PAs can constitute nearly one-third of poplar leaf dry mass and are therefore ecologically significant compounds (Lindroth and Hwang, 1996; Constabel and Lindroth, 2010).

1.7 Research Objectives

The diversity of phenolic chemicals, their potential ecological significance and the abundance of genomics resources make poplar an ideal system to study PPOs. Previous to this analysis, three hybrid poplar *PPO* cDNAs (*PtdPPO1*, *PtdPPO2* and *PtdPPO3*) were cloned and characterized (Constabel et al., 2000; Wang and Constabel, 2003; Wang and Constabel, 2004a, b). With the availability of the *P. trichocarpa* genome, there is now the possibility to study the entire *PPO* gene family. To date, there are no reports of an analysis on an entire *PPO* gene family from a sequenced genome. Therefore, the goal of this study was to investigate the *PPO* gene family in poplar at multiple levels. In Chapter Two, the objective was to compare the *PPO* gene families from sequenced genomes of the green plants, including poplar. A genome-wide analysis of *PPO* genes from available genomes representing five Viridiplantae lineages was

completed and their sequence features and phylogenetic relationships were compared. In Chapter Three, the objective was to compare the *PPO* genes within poplar. Here, the phylogenetic relationship and gene expression profile were analyzed, leading to a more detailed analysis of poplar PtrPPO13, a unique PPO. Note: Both chapters are presented as stand-alone manuscripts.

Chapter 2. Comparative Phylogenetic Analysis of the Polyphenol Oxidase Gene Family in Poplar

2.1 Abstract

Polyphenol oxidases (PPOs) are ubiquitous copper-binding enzymes in land plants that catalyze the oxidation of *ortho*-diphenols to *ortho*-diquinones, in the presence of molecular oxygen. PPO functions appear to be diverse and species dependent. Some reports suggest a role in defense, but other roles including PPOs as hydroxylases in biosynthetic pathways have also been described. In this study, bioinformatic searches to identify putative *PPO* genes were completed in 23 genomes representing five Viridiplantae lineages (chlorophytes, bryophytes, lycophytes, monocotyledonous anthophytes and eudicotyledonous anthophytes). *Physcomitrella patens*, *Selaginella moellendorffii* and soybean (*Glycine max*) were found to contain the largest *PPO* gene families with more than 10 genes. Poplar (*Populus trichocarpa*) was found to contain a highly diversified *PPO* gene family of nine genes. By contrast, in green algae and Arabidopsis (*Arabidopsis lyrata* and *A. thaliana*), no *PPO*-like sequences were identified. Intron-containing *PPO* genes appear to be widespread in land plants and are not restricted to monocotyledonous plants as previously thought. Most PPOs are also targeted to the chloroplast. However, N-terminal amino acid sequence analysis using TargetP predicted a number of putative PPOs to localize to the secretory pathway. Phylogenetic analysis showed that gene duplications have produced expanded *PPO* gene families within many species clades, which appear to be related by certain residues found in the CuA domain.

The variation in *PPO* gene family size and structure is likely consistent with a diversity of ecological and physiological functions for PPOs.

2.2 Introduction

Since their transition onto land, plants have evolved secondary metabolic pathways that play an important role in their survival, adaptation and ecological interactions. Some of these pathways generate an array of phenolic compounds that are potential substrates for polyphenol oxidases (PPOs). PPOs are ubiquitous in nature and have been studied in bacteria, fungi, mammals and plants. These dicopper enzymes exhibit oxygenase/oxidase activity and convert monophenols to *ortho*-diphenols (monophenolase activity; EC 1.14.18.1) and/or *ortho*-diphenols to *ortho*-diquinones (diphenolase activity; EC 1.10.3.1) in the presence of molecular oxygen (Gerdemann et al., 2002). In plants, PPOs are often considered to be defense proteins due to their herbivore-, pathogen- and wound-induced expression (Thipyapong et al., 1995; Thipyapong and Steffens, 1997; Constabel and Ryan, 1998; Constabel et al., 2000; Stewart et al., 2001; Raj et al., 2006; Pinto et al., 2008). However, their physiological functions extend beyond plant defense since roles as hydroxylases in phenolic compound biosynthesis have also been demonstrated (Steiner et al., 1999; Nakayama et al., 2000; Cho et al., 2003; Gandía-Herrero et al., 2005).

PPOs are nuclear-encoded proteins that consist of three components: an N-terminal chloroplast transit peptide (cTP), a dicopper centre, and a C-terminal region (Gerdemann et al., 2002). An 8-12 kDa bipartite cTP (Bucheli et al., 1996) is usually found at the N-terminus for import into the thylakoid lumen via the twin arginine-dependent translocation (Tat) pathway (Koussevitzky et al., 2008). Isotope labelling studies by Sommer et al. (1994) demonstrated the sequential removal of the cTP from a 67 kDa precursor PPO, from which a 62 kDa intermediate was processed to a 59 kDa

protein. Surprisingly, PPO proteins are not exclusively found in the chloroplast as a snapdragon (*Antirrhinum majus*) homolog has been found to be targeted to the vacuole (Ono et al., 2006).

Mature PPO proteins have a molecular mass of approximately 56-62 kDa (Bucheli et al., 1996) and contain a dicopper centre that consists of two conserved copper-binding domains (CuA and CuB), each with three histidine residues that coordinate a copper ion and comprise the active site (Klabunde et al., 1998; Virador et al., 2010). Each domain is approximately 50 amino acids in length, separated by a linker segment of approximately 100 residues (van Gelder et al., 1997). Although both domains are conserved, the CuA domain is more variable in sequence when compared to the CuB domain and may affect substrate preferences. At the C-terminal end of the PPO polypeptide is a region that has been identified to be susceptible to proteolytic cleavage in some plants, such as in hybrid poplar (*Populus trichocarpa* x *P. deltoides*; Wang and Constabel, 2004b), broad bean (*Vicia faba*; Robinson and Dry, 1992) and grape berry (*Vitis vinifera*; Dry and Robinson, 1994), and has been linked to enzyme activation.

With few exceptions, notably Arabidopsis, PPOs are ubiquitous among plants. However most of the studies have been focused on angiosperms. To date, only one bryophyte (*Physcomitrella patens*) *PPO* cDNA has been isolated (Richter et al., 2005). PPO activity and the generated quinone products are highly reactive and are responsible for the undesirable appearance and reduced nutrition of fruits and vegetables. This has motivated numerous studies in agricultural crop and forage plants to identify PPO sequences (Demeke and Morris, 2002; Sullivan et al., 2004; Yu et al., 2008; Parveen et al., 2010; Taketa et al., 2010). The high level of conservation of the PPO Cu-binding

domain has facilitated the successful isolation of *PPO* cDNAs from apple (*Malus domestica*; Boss et al., 1995), tomato (*Solanum lycopersicum*; Newman et al., 1993) and potato (*S. tuberosum*; Hunt et al., 1993). These and other eudicotyledonous species contain multiple, intronless *PPO* genes. For instance, tomato contains seven single exon *PPO* genes (Newman et al., 1993), and potato (Thygesen et al., 1995) and red clover (*Trifolium pratense*; Winters et al., 2009) contain five single exon *PPO* genes. Interestingly, four *PPO* clones were isolated from banana (*Musa cavendishii*) including one sequence with a 94 bp intron, the first identified for a *PPO* gene (Gooding et al., 2001). Subsequent studies of other monocotyledonous *PPO* genes revealed one intron in pineapple (*Ananas comosus*) *PINPPO1* and *PINPPO2* and two introns in wheat (*Triticum* spp.) *PPO* genes (Zhou et al., 2003; Sun et al., 2005; Massa et al., 2007).

Many gene families have been elucidated from the genomes of Arabidopsis (*Arabidopsis thaliana*; Arabidopsis Genome Initiative, 2000), rice (*Oryza sativa*; Goff et al., 2002; Yu et al., 2002) and poplar (*P. trichocarpa*; Tuskan et al., 2006) and analyzed from a comparative genomics perspective. Perhaps because *PPO* genes are not found in the Arabidopsis genome (Van der Hoeven et al., 2002), *PPO* gene families have not yet been studied extensively at this level. Instead, plant and fungal PPOs have often been compared as they share structural and reactivity characteristics (van Gelder et al., 1997; Marusek et al., 2006; Mayer, 2006; Selinheimo et al., 2007; Flurkey and Inlow, 2008).

Several recently sequenced plant genomes are now available from multiple lineages of green plants. With the recent release of genomes of evolutionary significance such as *Physcomitrella patens* (Rensing et al., 2008) and *Selaginella mollendorffii*, a comparative genomics approach should give insight into *PPO* gene families and their

evolution. The current work describes the first genome-wide analysis of *PPO* gene families from multiple land plants and their relatives. Genomes from five Viridiplantae lineages, including chlorophytes (*Chlamydomonas reinhardtii*, *Micromonas pullisa*, *Osterococcus lucimarinus*, *Osterococcus tauri* and *Volvox carteri*), bryophytes (*Physcomitrella patens*), lycophytes (*Selaginella mollendorffii*), monocotyledonous anthophytes (*Brachypodium distachyon*, *Oryza sativa*, *Sorghum bicolor* and *Zea mays*) and eudicotyledonous anthophytes (*Arabidopsis lyrata*, *Arabidopsis thaliana*, *Carica papaya*, *Cucumis sativus*, *Glycine max*, *Manihot esculenta*, *Medicago truncatula*, *Mimulus guttatus*, *Populus trichocarpa*, *Prunus persica*, *Ricinus communis* and *Vitis vinifera*) were surveyed for *PPO* genes. *PPO* gene and protein structures were compared, and a phylogenetic analysis was performed, in order to identify the relationships of plant PPOs. The results show that a variable number of *PPO* genes are present in the genomes analyzed here, and that some of these genomes have retained recently duplicated *PPO* genes.

2.3 Materials and Methods

Between June 2009 to June 2010, TBLASTX searches using default parameters were completed for 23 (masked) genomes, available from the United States Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/): *Chlamydomonas reinhardtii* (JGI v2.0; Merchant et al., 2007), *Micromonas pullisia* (JGI v2.0), *Osterococcus lucimarinus* (JGI v2.0), *Ostreococcus tauri* (JGI v2.0; Derelle et al., 2006), *Volvox carteri* (JGI v1.0; Prochnik et al., 2010), *Arabidopsis lyrata* (JGI v1.0), *Arabidopsis thaliana* (TAIR 9), *Brachypodium distachyon* (JGI v1.0; Vogel et al., 2010), *Carica*

papaya (ASGPB v0.4; Ming et al., 2008), *Cucumis sativus* (Roche/JGI v1.0; Huang et al., 2009), *Glycine max* (JGI Glyma 1.0; Schmutz et al., 2010), *Manihot esculenta* (JGI v1.0), *Medicago truncatula* (JGI v3.0), *Mimulus guttatus* (JGI v1.0), *Oryza sativa* L. ssp. *japonica* (MSU 6.0; Goff et al., 2002), *Physcomitrella patens* (JGI v1.1; Rensing et al., 2008), *Populus trichocarpa* (JGI v1.1; Tuskan et al., 2006), *Prunus persica* (IPGI v1.0), *Ricinus communis* (TIGR v0.1), *Selaginella mollendorffii* (JGI v1.0), *Sorghum bicolor* (MIPS/JGI 1.4; Paterson et al., 2009), *Vitis vinifera* (Genoscope v1.0; Jaillon et al., 2007), and *Zea mays* (maizesequence.org v4a53; Schnable et al., 2009). Hybrid poplar (*P. trichocarpa* x *P. deltoides*) *PtdPPO1*, *PtdPPO2* and *PtdPPO3* nucleotide sequences (GenBank Accessions: AF263611, AY665681 and AY665682, respectively) were used as queries.

Genome database BLAST hits were conceptually translated, manually inspected and run through NCBI BLASTP and SMART (Schultz et al., 1998) (**S**imple **M**odular **A**rchitecture **R**esearch **T**ool; http://smart.embl-heidelberg.de/) to confirm the presence of the conserved CuA and CuB domains. Putative N-terminal transit peptide sequences were predicted using ChloroP 1.1 (Emanuelsson et al., 1999) and TargetP 1.1 (Emanuelsson et al., 2007). The genomic sequences of the identified gene models were inspected for intron annotations. For some of the monocotyledonous *PPO* genes that were predicted to contain introns, the online tool CIWOG (Wilkerson et al., 2009) (**C**ommon **I**ntrons **W**ithin **O**rthologous **G**enes; http://ciwog.gdcb.iastate.edu/) was used to compare their relative positions. Altogether, 94 putative full-length, or near full-length *PPO* sequences of at least 1351 bp were identified and retained for this analysis.

All sequences were aligned using MUSCLE (Edgar, 2004) (**Mu**ltiple **S**equence **C**omparison by **L**og **E**xpectation; http://www.ebi.ac.uk/Tools/muscle/index.html) set on default parameters to further verify the presence and positions of the conserved histidine residues in both the CuA and CuB domains. For the purpose of this analysis, it was expected that the conserved histidine residues lined up in each domain. The N- and C-termini were removed, leaving the core PPO protein containing the CuA and CuB domains, and the PPO1_DWL domain (Supplemental Figure 2.1). Additional alignment manipulations were completed in BioEdit (Hall, 1999).

A neighbour-joining phylogenetic tree based on the alignment described above was generated using MEGA 4.0 (Tamura et al., 2007). Genetic distances were estimated using the Dayhoff amino acid substitution matrix. Positions in the alignment lacking amino acid residues were excluded from the pairwise distance estimates. Bootstrap replicates (1000) were used to indicate the level of support for the data in each node of the tree.

2.4 Results

2.4.1 Genome-Wide Identification of *PPO* Genes in Land Plants

This analysis identified putative *PPO* genes in 16 genomes representing four lineages of land plants (bryophytes, lycophytes, monocotyledonous anthophytes and eudicotyledonous anthophytes). A total of 94 *PPO* genes of at least 1351 bp in coding sequence length were found (Supplemental Table 2.1). Only six of these full-length sequences had been previously cloned and characterized (Dry and Robinson, 1994; Constabel et al., 2000; Wang and Constabel, 2004a; Richter et al., 2005; Yu et al., 2008).

In addition, two maize *PPO* cDNAs have been isolated recently (Alexandrov et al., 2009). All gene models were confirmed to encode PPOs based on the presence of a tyrosinase (Pfam00264) domain as detected by SMART (Schultz et al., 1998). The tyrosinase (CuA and CuB) domain, referred to as the Cu-binding domain in plants, is homologous in eukaryotes and prokaryotes (van Gelder et al., 1997). The amino acid sequences were aligned and the positions of the conserved histidine residues were inspected. Sequences without both domains were excluded from further analysis (Supplemental Table 2.2). Truncated *PPO* sequences that were less than 1200 bp and contained premature termination codons, and gene models with annotation discrepancies were also discarded.

Previous studies, such as in potato and wheat, had suggested that angiosperm genomes contain fewer than 10 *PPO* genes (Thygesen et al., 1995; Massa et al., 2007). Prior to this analysis, the *PPO* gene family in tomato was the largest known with seven genes (Newman et al., 1993). The present work found several plants with *PPO* gene families of more than 10 genes (Table 2.1). Surprisingly, the largest family was in *Physcomitrella,* with 13 *PPO* genes. *Selaginella,* whose genome is one of the smallest plant genomes reported (Banks, 2009), contains 11 *PPO* genes. Among the flowering plants, soybean has the largest *PPO* gene family, also with 11 genes. Nine *PPO* genes were identified in *Mimulus* and poplar. Interestingly, cassava and *Ricinus,* which like poplar belong to the order Malpighiales (Wurdack and Davis, 2009), appear to each have only one *PPO* gene. In the monocotyledonous plants, *Brachypodium* and *Sorghum* contain eight *PPO* genes, whereas maize contains six genes and only two genes were identified in rice. No *PPO* genes were detected in the genome of Arabidopsis, an

Table 2.1 Number of putative functional *PPO* genes identified from BLAST analysis of publically available Viridiplantae genomes.

Genome		Estimated Genome Size (Mb)[a]	*PPO* Genes[b]
Chlorophytes			
green algae (unicellular)	*Chlamydomonas reinhardtii**	120	0
	Micromonas pullisia	15	0
	Ostreococcus lucimarinus	13	0
	*Ostreococcus tauri**	12	0
(multicellular)	*Volvox carteri**	120	0
Bryophytes			
moss	*Physcomitrella patens**	500	13
Lycophytes			
spike moss	*Selaginella moellendorffii*	100	11
Monocotyledonous Anthophytes			
purple false brome	*Brachypodium distachyon**	355	6
rice	*Oryza sativa**	466	2
cereal grass	*Sorghum bicolor**	760	8
corn	*Zea mays**	2400	6
Eudicotyledonous Anthophytes			
lyrate rockcress	*Arabidopsis lyrata*	230	0
thale cress	*Arabidopsis thaliana**	125	0
papaya	*Carica papaya**	372	4
cucumber	*Cucumis sativus**	367	1
soybean	*Glycine max**	1200	11
cassava	*Manihot esculenta*	760	1
barrel medic	*Medicago truncatula*	500	4
monkey flower	*Mimulus guttatus*	430	9
poplar	*Populus trichocarpa**	480	9
peach	*Prunus persica*	290	4
castor bean	*Ricinus communis*	400	1
grape	*Vitis vinifera**	500	4
			94

[a] Estimated genome sizes as indicated in NCBI (http://www.ncbi.nlm.nih.gov/genomeprj).

[b] Denotes minimum number of *PPO* genes as identified from this analysis. Additional putative functional *PPO* gene models with discrepancies were identified for some genomes, but were excluded from this analysis and are listed in Supplemental Table 2.2.

* Denotes genomes that have been described in a publication: *Chlamydomonas reinhardtii* (Merchant et al., 2007), *Ostreococcus tauri* (Derelle et al., 2006), *Volvox carteri* (Prochnik et al., 2010), *Physcomitrella patens* (Rensing et al., 2008), *Brachypodium distachyon* (Vogel et al., 2010), *Oryza sativa* (Goff et al., 2002), *Sorghum bicolor* (Paterson et al., 2009), *Zea mays* (Schnable et al., 2009), *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), *Carica papaya* (Ming et al., 2008), *Cucumis sativus* (Huang et al., 2009), *Glycine max* (Schmutz et al., 2010), *Populus trichocarpa* (Tuskan et al., 2006) and *Vitis vinifera* (Jaillon et al., 2007).

observation consistent with the analysis performed by Van der Hoeven et al. (2002).

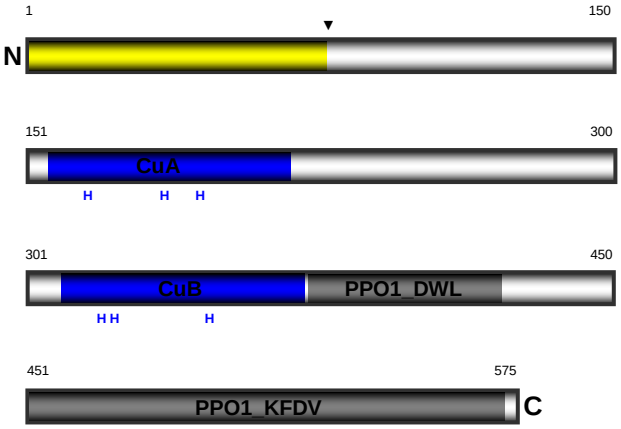
The above numbers represent minimum estimates since several gene models predicted to encode functional proteins were identified, but discarded since they were either incomplete or had annotation discrepancies (Supplemental Table 2.2). For instance, the soybean gene model Glyma15g07700.1 sequence is incomplete at the CuB domain. Another soybean gene model Glyma07g31290.1 is predicted to be a five-exon gene that encodes an 1100 amino acid protein, which is much too large for plant PPOs. Manual inspection of exons one, two, four and five revealed that these encode a 615 amino acid PPO protein, suggesting inaccuracies in the gene model annotation. Similarly, the annotated initiation codon for the *Mimulus* gene model mgf021284m is suspected to be incorrect as a 689 amino acid protein is predicted. Comparison with the other *Mimulus PPO* sequences suggests an alternate ATG initiation codon in exon two which would encode a PPO protein of expected size. Nevertheless, since I could not independently verify these sequences, they were excluded from further analyses.

2.4.2 Functional Domains of PPOs are Conserved in Land Plants

PPO proteins consist of three distinct regions (Figure 2.1a), most of which have been described in other bioinformatic studies (van Gelder et al., 1997; Marusek et al., 2006; Flurkey and Inlow, 2008). Protein structure comparisons of plant PPOs with *Neurospora crassa* tyrosinase (Lerch, 1987) and giant octopus (*Octopus dofleini*) hemocyanin (Cuff et al., 1998) has led to the identification of the histidine residues that are essential for enzymatic catalysis (Klabunde et al., 1998; Virador et al., 2010). Therefore, consensus sequences for the N-terminus, the copper-binding region and the

Figure 2.1 PPOs contain three distinct regions. (A) Schematic diagram of PPO proteins. Most PPOs are targeted to the chloroplast via an N-terminal transit peptide (yellow) that is cleaved at the alanine motif (inverted triangle) upon import into the thylakoid lumen. Removal of the transit peptide leaves the core PPO polypeptide that consists of the CuA and CuB domains (blue) and the C-terminal end (grey). At the C-terminal end are the PPO1_DWL (Pfam12142) and the PPO1_KFDV (Pfam12143) domains. (B) WebLogo (Crooks et al., 2004) PPO amino acid consensus sequences for the N-terminal transit peptide, CuA and CuB domains, and the PPO1_DWL and PPO1_KFDV domains. For the N-terminal transit peptide, the first 35 amino acids of the stromal domain are presented. The underlined sequences are known regions, including the thylakoid transfer domain, the alanine (AxA) cleavage motif, the DWL motif, the tyrosine (YxY) motif and the KFDV motif. The three conserved histidine residues (blue) in the CuA and CuB domains are numbered 1, 2 and 3, respectively. Black stars indicate 100% conserved residues in each of the domains. Boxed sequences in the PPO1_KFDV domain represent conserved regions that have not been discussed in literature.

A



B

N-terminal Transit Peptide

CuA Domain

CuB Domain

PPO1_DWL Domain

PPO1_KFDV Domain

C-terminus were generated to confirm the presence of these domains, as well as to identify new sequence features within these domains.

Manual inspection of the WebLogo (Crooks et al., 1994) consensus sequence for the first 35 residues of the stromal domain of the cTP found a high proportion of serine residues (Figure 2.1b). Adjacent to the stromal domain, a thylakoid transfer domain (TTD) and an alanine cleavage motif were evident. Together the presence of these features suggests that most of the PPOs are chloroplast proteins. For approximately 75% of the identified PPOs, this was also supported by predictions from ChloroP 1.1 (Emanuelsson et al., 1999). In some instances, a cTP was not predicted despite the presence of a putative TTD (Supplemental Table 2.1).

Surprisingly, PPOs from *Physcomitrella* and a number of monocotyledonous and eudicotyledonous anthophytes were conspicuous in the absence of a cTP. N-terminal sequence analysis using TargetP 1.1 (Emanuelsson et al., 2007) predicted with high specificity (> 0.95) that most of these novel PPOs localize to the secretory pathway. To date, aureusidin synthase from snapdragon is the only PPO homolog that has been experimentally confirmed to localize to the vacuole (Ono et al., 2006).

PPOs contain a copper-binding region that consists of a CuA and CuB domain, each with three conserved histidine residues. In the CuA domain, the first conserved histidine residue is part of the sequence motif HxxxC, as described by Klabunde et al. (1998). Analysis of the CuA domain consensus sequence found HCAYC to be the most common sequence motif, in which the second cysteine residue is 100% conserved (Figure 2.1b). This conserved cysteine forms a thioether bond with the second conserved histidine residue in the CuA domain (Klabunde et al., 1998). Based on this analysis,

other residues in the CuA domain that were identified to also be conserved are arginine, glutamic acid, phenylalanine, tryptophan and aspartic acid, located C-terminal to the third histidine.

Analysis of the CuB domain found the first two conserved histidine residues to be contained in a previously unidentified HxxxH sequence motif, where the three middle residues are variable (Figure 2.1b). However at the fourth position in the motif, a hydrophobic residue, either alanine, valine, leucine, isoleucine or methionine, is present. Other amino acids C-terminal to the second conserved histidine in the CuB domain, specifically the aspartic acid and phenylalanine residues, were found to be 100% conserved. C-terminal to phenylalanine is a histidine residue that is present in most PPO sequences, but is not 100% conserved. Asparagine, aspartic acid, and two tryptophan residues are invariable residues C-terminal to the third conserved histidine. Therefore, in addition to the three essential histidine residues in the CuA and CuB domains, other conserved amino acids are also present in plant PPOs.

The PPO C-terminal end consists of a 50 amino acid PPO1_DWL domain (Pfam12142), and a 140-150 amino acid PPO1_KFDV domain (Pfam12143). The significance of this end has not been studied in detail. However, in PPOs where proteolytic processing of this end has been observed, cleavage occurs immediately C-terminal to the tyrosine (YxY) motif (Marusek et al., 2006; Flurkey and Inlow, 2008) contained in the PPO1_DWL domain. As a result, a fragment of approximately 16-18 kDa containing the PPO1_KFDV domain is removed (Robinson and Dry, 1992; Dry and Robinson, 1994). The PPO1_KFDV domain has not been described in previous bioinformatic studies of PPOs. Here in the latter, two uncharacterized sequence motifs

were identified (Figure 2.1b). One motif is enriched in glutamic acid residues. Although the glutamic acid-rich motif EEEEEVLVI is conserved, the first three glutamic acid residues of the motif are not present in the *Physcomitrella* and *Selaginella* PPOs. However in the sequence EEEEEVLVI, the EVLVI motif is evident in most of the identified land plant PPO sequences. C-terminal to this sequence motif is the KFDV motif, which was found only in the flowering plant PPOs and three *Selaginella* PPOs, SmoPPO1, SmoPPO2 and SmoPPO3. C-terminal to the KFDV motif is another newly identified sequence motif, EFAGSF that is present in most of the identified PPOs in which a glutamic acid and glycine residue appear to be common (Figure 2.1b). In some PPO sequences, immediately C-terminal to the histidine in the EFAGSF sequence motif (Figure 2.1b) are up to four additional histidines residues that have been postulated by Steffens et al. (1994) to form a potential third copper-binding (CuC) domain. Therefore, the C-terminal end of the PPO protein contains regions of unknown function that appear to be conserved, especially within the higher plants.

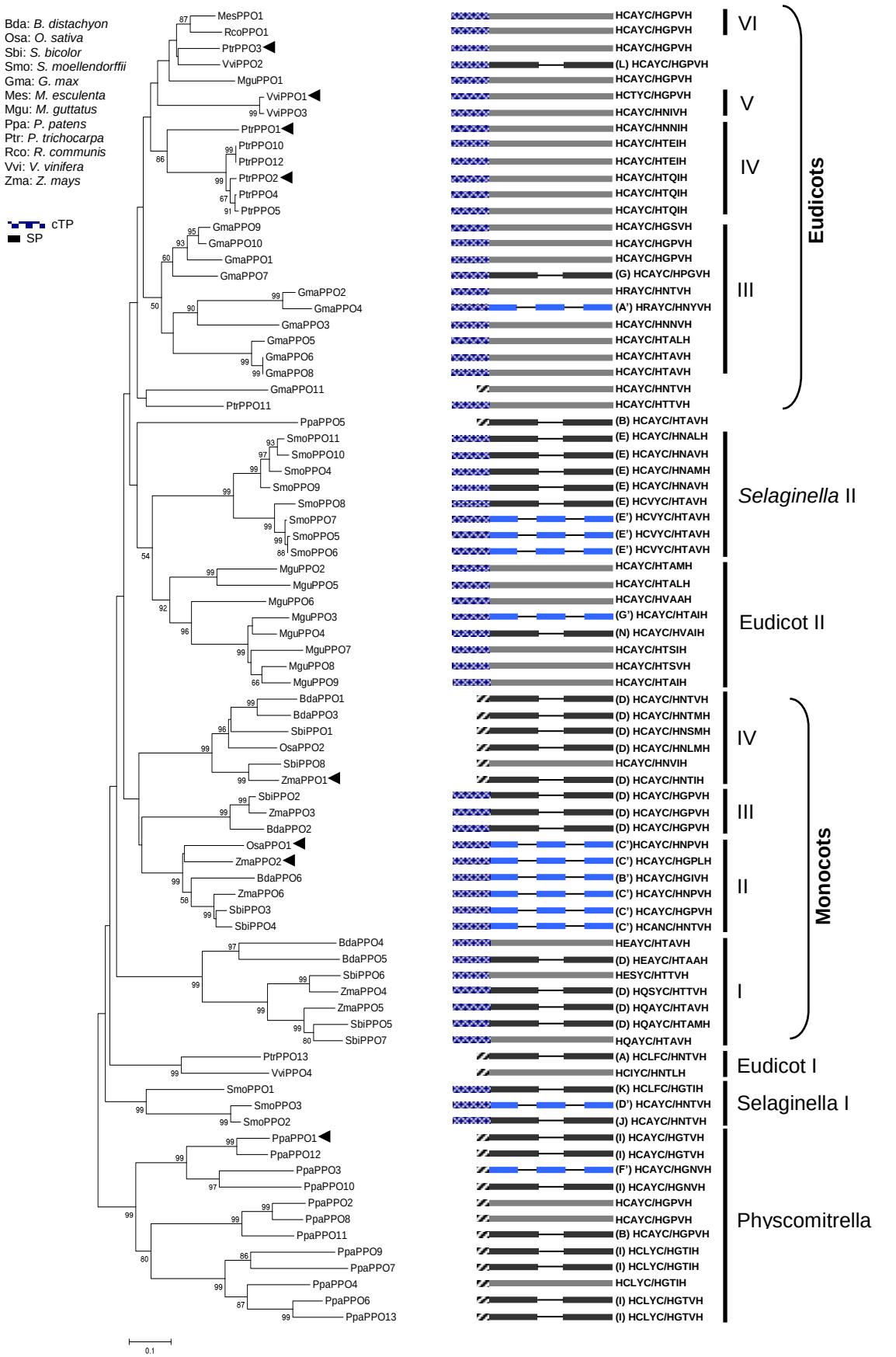
2.4.3 Phylogenetic Reconstruction of PPOs Reveals a Tree with Several Well Supported Groups

A neighbour-joining phylogenetic analysis was generated from the region containing the conserved copper-binding domains (excluding the linker region) and the PPO1_DWL domain, described above (Supplemental Figure 2.1). Amino acid PPO sequences from the genomes of *Physcomitrella*, *Selaginella*, *Brachypodium*, rice, *Sorghum*, maize, soybean, *Mimulus,* poplar and grape were selected for the phylogenetic analysis based on their evolutionary significance and their quality of genome annotations.

PPOs from cassava (*M. esculenta*) and *Ricinus* were also included as they are the closest relatives of poplar with sequenced genomes.

With *Physcomitrella* PPOs as the outgroup, this analysis shows the separation of land plant PPOs into a number of distinct clades (Figure 2.2). Firstly, most of the *Physcomitrella* PPOs are found in the most basal clade, with the exception of PpaPPO5, which appears to be more related to the *Selaginella* and *Mimulus* sequences. The *Selaginella* PPOs form two distinct clades, Selaginella I and Selaginella II. *Selaginella* SmoPPO1, SmoPPO2 and SmoPPO3 appear to form Selaginella I, a small clade near the base of the tree. In the clade Selaginella II, the other *Selaginella* PPOs cluster with most of the homologs from *Mimulus*. Secondly, the monocotyledonous PPOs form four supported clades, designated Monocot I-IV. Lastly, with the exception of the *Mimulus* sequences in Eudicot II, most of the eudicotyledonous PPOs (soybean, cassava, *Ricinus*, poplar and grape) form the clades Eudicot III-VI. Unfortunately the ancestral relationships of plant PPOs, in particular between the monocotyledonous and eudicotyledonous PPOs, cannot be resolved as indicated by bootstrap support of less than 50% at many of the nodes (Figure 2.2) and may be correlated with the rapid emergence and speciation of the angiosperms, especially among the rosids (Wang et al., 2009). However, recent relationships within species-specific clades are well supported. Many of the large clades show a consistent pattern in which a common ancestor underwent an initial duplication event to give rise to two sister sequences, which underwent subsequent duplication events after speciation occurred; this is most apparent in PPO families of *Physcomitrella*, *Selaginella*, *Mimulus*, soybean and poplar (Figure 2.2). Phylogenetic analyses were also tried with the maximum likelihood and maximum parsimony methods

Figure 2.2 A neighbour-joining phylogenetic reconstruction of land plant PPO from four lineages: bryophytes, lycophytes, monocotyledonous anthophytes and eudicotyledonous anthophytes. Amino acid sequences of PPOs from *Physcomitrella* (Ppa), *Selaginella* (Smo), *Brachypodium* (Bda), rice (Osa), *Sorghum* (Sbi), maize (Zma), soybean (Gma), cassava (Mes), *Mimulus* (Mgu), poplar (Ptr), *Ricinus* (Rco) and grape (Vvi) form the following clades: Physcomitrella, Selaginella I and II, Monocot I-IV and Eudicot I-VI. Genetic distances were estimated using the Dayhoff amino acid substitution matrix. Positions in the alignment lacking amino acid residues were excluded from the pairwise distance estimates. Bootstrap replicates (1000) were used to indicate the level of support for the data in each node of the tree and only values of greater than 50% are indicated. Illustrated on the right are diversified *PPO* gene structures with arbitrary intron positions indicated. Actual intron positions are illustrated in Figure 2.3 and correspond to the information provided in Supplemental Table 2.3. Intron groups are also indicated in brackets and correspond to the information provided in Supplemental Table 2.3. Light grey lines represent single exon *PPO* genes, dark grey lines represent one intron (two exon) *PPO* genes and blue lines represent two intron (three exon) *PPO* genes. Also indicated on the gene is the presence of an encoded chloroplast transit peptide (blue hatch lines) or putative signal peptide (black hatch lines) as predicted by ChloroP 1.1 (Emanuelsson et al., 1999) and TargetP 1.1 (Emanuelsson et al., 2007). The respective CuA HxxxC and CuB HxxxH sequence motifs for each of the PPO sequences are also shown. The black arrows indicate *PPO* cDNAs that have been characterized: *OsaPPO1* (*Phr1*; Yu et al., 2008), *PpaPPO1* (*Pp_ppo1*; Richter et al., 2005), *PtrPPO1* (*PtdPPO1*; Constabel et al., 2000), *PtrPPO2* (*PtdPPO2*; Wang and Constabel, 2004a), *PtrPPO3* (*PtdPPO3*; Wang and Constabel, 2004a), *VviPPO1* (*GPO1*; Dry and Robinson, 1994), *ZmaPPO1* (GenBank Accession ACG28948; Alexandrov et al., 2009) and *ZmaPPO2* (GenBank Accession ACG35817; Alexandrov et al., 2009).


Bda: B. distachyon
Osa: O. sativa
Sbi: S. bicolor
Smo: S. moellendorffii
Gma: G. max
Mes: M. esculenta
Mgu: M. guttatus
Ppa: P. patens
Ptr: P. trichocarpa
Rco: R. communis
Vvi: V. vinifera
Zma: Z. mays
cTP
SP
MesPPO1
RcoPPO1
PtrPPO3
VviPPO2
MguPPO1
VviPPO1
VviPPO3
PtrPPO1
PtrPPO10
PtrPPO12
PtrPPO2
PtrPPO4
PtrPPO5
GmaPPO9
GmaPPO10
GmaPPO1
GmaPPO7
GmaPPO2
GmaPPO4
GmaPPO3
GmaPPO5
GmaPPO6
GmaPPO8
GmaPPO11
PtrPPO11
PpaPPO5
SmoPPO11
SmoPPO10
SmoPPO4
SmoPPO9
SmoPPO8
SmoPPO7
SmoPPO5
SmoPPO6
MguPPO2
MguPPO5
MguPPO6
MguPPO3
MguPPO4
MguPPO7
MguPPO8
MguPPO9
BdaPPO1
BdaPPO3
SbiPPO1
OsaPPO2
SbiPPO8
ZmaPPO1
SbiPPO2
ZmaPPO3
BdaPPO2
OsaPPO1
ZmaPPO2
BdaPPO6
ZmaPPO6
SbiPPO3
SbiPPO4
BdaPPO4
BdaPPO5
SbiPPO6
ZmaPPO4
ZmaPPO5
SbiPPO5
SbiPPO7
PtrPPO13
VviPPO4
SmoPPO1
SmoPPO3
SmoPPO2
PpaPPO1
PpaPPO12
PpaPPO3
PpaPPO10
PpaPPO2
PpaPPO8
PpaPPO11
PpaPPO9
PpaPPO7
PpaPPO4
PpaPPO6
PpaPPO13
0.1
HCAYC/HGPVH
HCAYC/HGPVH
HCAYC/HGPVH
(L) HCAYC/HGPVH
HCAYC/HGPVH
HCTYC/HGPVH
HCAYC/HNIVH
HCAYC/HNNIH
HCAYC/HTEIH
HCAYC/HTEIH
HCAYC/HTQIH
HCAYC/HTQIH
HCAYC/HTQIH
HCAYC/HGSVH
HCAYC/HGPVH
HCAYC/HGPVH
(G) HCAYC/HPGVH
HRAYC/HNTVH
(A') HRAYC/HNYVH
HCAYC/HNNVH
HCAYC/HTALH
HCAYC/HTAVH
HCAYC/HTAVH
HCAYC/HNTVH
HCAYC/HTTVH
(B) HCAYC/HTAVH
(E) HCAYC/HNALH
(E) HCAYC/HNAVH
(E) HCAYC/HNAMH
(E) HCAYC/HNAVH
(E) HCVYC/HTAVH
(E') HCVYC/HTAVH
(E') HCVYC/HTAVH
(E') HCVYC/HTAVH
HCAYC/HTAMH
HCAYC/HTALH
HCAYC/HVAAH
(G') HCAYC/HTAIH
(N) HCAYC/HVAIH
HCAYC/HTSIH
HCAYC/HTSVH
HCAYC/HTAIH
(D) HCAYC/HNTVH
(D) HCAYC/HNTMH
(D) HCAYC/HNSMH
(D) HCAYC/HNLMH
HCAYC/HNVIH
(D) HCAYC/HNTIH
(D) HCAYC/HGPVH
(D) HCAYC/HGPVH
(D) HCAYC/HGPVH
(C')HCAYC/HNPVH
(C') HCAYC/HGPLH
(B') HCAYC/HGIVH
(C') HCAYC/HNPVH
(C') HCAYC/HGPVH
(C') HCANC/HNTVH
HEAYC/HTAVH
(D) HEAYC/HTAAH
HESYC/HTTVH
(D) HQSYC/HTTVH
(D) HQAYC/HTAVH
(D) HQAYC/HTAMH
HQAYC/HTAVH
(A) HCLFC/HNTVH
HCIYC/HNTLH
(K) HCLFC/HGTIH
(D') HCAYC/HNTVH
(J) HCAYC/HNTVH
(I) HCAYC/HGTVH
(I) HCAYC/HGTVH
(F') HCAYC/HGNVH
(I) HCAYC/HGNVH
HCAYC/HGPVH
HCAYC/HGPVH
(B) HCAYC/HGPVH
(I) HCLYC/HGTIH
(I) HCLYC/HGTIH
HCLYC/HGTIH
(I) HCLYC/HGTVH
(I) HCLYC/HGTVH
VI
V
IV
III
Eudicots
Selaginella II
Eudicot II
IV
III
II
I
Monocots
Eudicot I
Selaginella I
Physcomitrella

from which similar tree topologies to the neighbour-joining method were obtained (data not shown).

Within the group containing the monocotyledonous PPOs, a different duplication pattern is observed. Monocot clades I-IV consist of PPOs that are suspected to be derived from at least three ancestral Poaceae *PPO* genes that existed before the speciation of modern cereals and grasses. This is because most of the species examined contain the same set of PPO orthologs. One exception is rice which appears to have lost one of the orthologs from Monocot I (Figure 2.2). An interesting characteristic of Monocot II and Monocot IV is that both of these clades contain two distinct groups of PPO sequences. Monocot II consists of all six of the two intron monocotyledonous *PPO* genes identified from this analysis (*BdaPPO6, OsaPPO1, SbiPPO3, SbiPPO4, ZmaPPO2* and *ZmaPPO6*). The other clade, Monocot IV, consists of uncharacterized PPOs that are predicted to contain a signal peptide for localization to the secretory pathway (Supplemental Table 2.1). This is an unusual feature for most plant PPOs.

Inspection of the Eudicot group found several supported clades containing PPOs from soybean and poplar, as well as PPOs from cassava and *Ricinus* (Figure 2.2). The soybean PPOs form one clade and include PPOs from other legume species, such as the ones from *Medicago* (data not shown). Next to the soybean PPO clade are PPO sequences from cassava, poplar, *Ricinus* and grape. Analysis of these PPOs revealed that some of these sequences do not show the same species-specific clustering, such as seen for PPOs from *Physcomitrella, Selaginella*, soybean and *Mimulus*.

In poplar, there appears to have been four ancestral *PPO* sequences, from which one duplicated to produce *PtrPPO1* and the ancestral *PtrPPO2* gene. Multiple gene

duplications of the latter resulted in the expansion of the *PtrPPO2* subgroup (Figure 2.2). Interestingly, PtrPPO3 represents a distinct PPO that clusters with grape VviPPO2, cassava MesPPO1 and *Ricinus* RcoPPO1. By contrast, PtrPPO11 and PtrPPO13 group with soybean GmaPPO11 and grape VviPPO4, respectively. However, note that the phylogenetic relationships of PtrPPO3 and PtrPPO11 are not supported here.

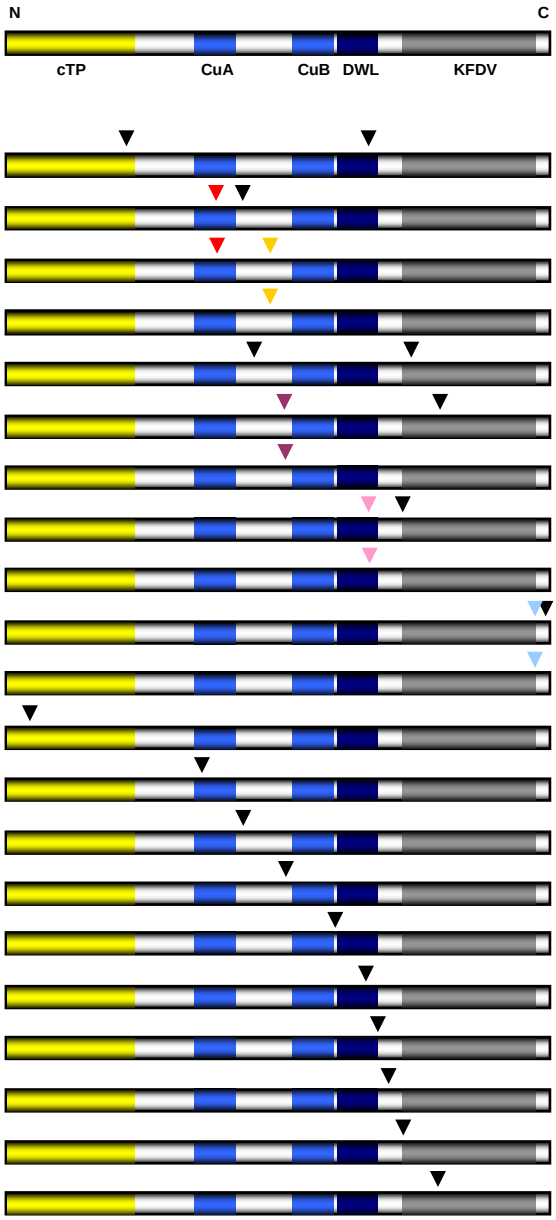
One of the most intriguing clades is Eudicot I which consists of poplar PtrPPO13 and grape VviPPO4 (Figure 2.2). This clade is of interest as these homologs are unusual in both sequence and subcellular targeting, compared to the most of the other identified PPOs. The basal position of this clade suggests that it may be an ancestral PPO. However, PPO sequences from *Annona cherimola* (Prieto et al., 2007) and *Argemone mexicana* (GenBank Accession ACJ76786) also group in Eudicot I and further suggests that the sequences in this clade represent ancestral PPOs that were present in early eudicotyledons (data not shown).

2.4.4 *PPO* Gene Structure and the Presence of Introns

While introns have not been detected in most of the well studied *PPO* genes, previous studies indicate that they are common in *PPO* sequences from monocotyledonous plants (Gooding et al., 2001; Zhou et al., 2003; Sun et al., 2005; Taketa et al., 2010). Therefore, the genomic sequences were inspected for the presence of introns. A number of intron-containing *PPO* genes were predicted for monocotyledonous sequences. Introns were also predicted in *PPO* genes of *Physcomitrella*, *Selaginella* and some eudicotyledonous plants (Figure 2.2, Figure 2.3, Supplemental Table 2.1). Eudicotyledonous *PPO* genes were previously considered to

Figure 2.3 *PPO* genes are diverse in their gene structures and contain multiple exons. Intron insertion positions with respect to the encoded PPO protein are shown. The different intron positions found in the *PPO* genes identified here are divided into arbitrary groups. Groups A'-G' represent two intron (three exon) *PPO* genes and groups A-N represent one intron (two exon) *PPO* genes. Black arrows indicate unique intron positions. Same coloured arrows indicate identical intron positions. Red indicates an identical intron position for groups B' and C'. Orange indicates an identical intron position for groups C' and D. Purple indicates an identical intron position for groups E' and E. Pink indicates an identical intron position for groups F' and I. Blue indicates an identical intron position for groups G' and N. The number of *PPO* genes associated with each intron position (designated as an intron group) is indicated on the left, and the total number of *PPO* genes with (bold) and without (in brackets) introns are indicated below. For additional details, refer to Supplemental Tables 2.1 and 2.3. P-*Physcomitrella*, S-*Selaginella*, M-monocotyledonous anthophytes and E-eudicotyledonous anthophytes.

Intron Group	P	S	M	E
A’	0	0	0	**1**
B’	0	0	**1**	0
C’	0	0	**5**	0
D	0	0	**12**	0
D’	0	**1**	0	0
E’	0	**3**	0	0
E	0	**5**	0	0
F’	**1**	0	0	0
I	**7**	0	0	0
G’	0	0	0	**1**
N	0	0	0	**1**
A	0	0	0	**1**
B	**1**	0	0	0
C	**1**	0	0	0
F	0	0	0	**1**
G	0	0	0	**1**
H	0	0	0	**1**
J	0	**1**	0	0
K	0	**1**	0	0
L	0	0	0	**1**
M	0	0	0	**1**
	10	**11**	**18**	**9**
	(13)	(11)	(22)	(47)



be generally intronless. However in *A. cherimola AcPPO*, an 83 bp intron is present at the 5' terminus (Prieto et al., 2007).

This analysis predicted introns in 48 of the 94 PPO-encoding genes (Figure 2.3, Supplemental Table 2.1). Of these, most contain a single intron although some are interrupted by two introns. To date, the latter has only been reported for *PPO* genes in barley (*Hordeum vulgare*; Taketa et al., 2010) and wheat (Sun et al., 2005; Massa et al., 2007). Nearly all of the *PPO* genes in *Physcomitrella* contain at least one intron except *PpaPPO2*, *PpaPPO4* and *PpaPPO8* which are single exon genes (Supplemental Table 2.1). In *Selaginella*, all of the *PPO* genes are predicted to contain intron sequences; *SmoPPO3*, *SmoPPO5*, *SmoPPO6* and *SmoPPO7* are predicted to contain two introns and three exons. Most of the monocotyledonous *PPO* genes also contain at least one intron. All of the *PPO* genes in rice and maize are interrupted by introns, whereas in *Brachypodium* and *Sorghum*, intronless *PPO* genes are also present. Conversely, very few eudicotyledonous *PPO* genes are predicted to be intronic (Figure 2.3, Supplemental Table 2.1). Notably, cucumber *CsaPPO1*, soybean *GmaPPO7*, *Medicago MtrPPO4*, *Mimulus MguPPO4*, poplar *PtrPPO13*, peach *PpePPO4* and grape *VviPPO2* are predicted to each contain an intron. Soybean *GmaPPO4* and *Mimulus MguPPO3* are the only eudicotyledonous *PPO* genes predicted to consist of two introns (Supplemental Table 2.3).

Among the intron-containing *PPO* genes, the intron lengths were found to be variable and ranged from 39 to 2203 nucleotides (Supplemental Table 2.3). Intron lengths in *PPO* genes from *Physcomitrella* and *Selaginella* ranged from 45 to 988 nucleotides. Monocotyledonous *PPO* introns were found to be 50 to 2203 bp in length

and in most instances, were larger than ones present in eudicotyledonous *PPO* genes. Inspection of the predicted intron sequences in all of the intron-containing *PPO* genes found a 5' GT-AG 3' terminal dinucleotide consensus sequence except for rice *OsaPPO1* (Yu et al., 2008) which has a 5' GC-AG 3' sequence. The GT-AG terminal dinucleotide sequence is typical of eukaryotic U2-type introns, a major class of spliceosomal introns (Rogozin and Milanesi, 1997; Sheth et al., 2006; Basu et al., 2008) which were also predicted for the monocotyledonous intron-containing *PPO* genes using the online tool CIWOG (Wilkerson et al., 2009).

The positions of the introns are most often within the linker region that separates the CuA and CuB domains, and within the PPO_DWL domain (Figure 2.3, Supplemental Table 2.3). For example, introns in the linker region were common at the 3' terminus, immediately after the encoded glutamine in the xMYK/RQ sequence. In the PPO_DWL domain, introns were often inserted before the arginine residue in the RYxYx sequence. Only rarely was an intron predicted to interrupt one of the functional domains. However a subset of sequences, mainly the two intron-containing *PPO* genes, was predicted to have an intron within the CuA domain. *PPO* genes from *Brachypodium* (*BdaPPO6*), rice (*OsaPPO1*), *Physcomitrella* (*PpaPPO5*), *Sorghum* (*SbiPPO3* and *SbiPPO4*) and maize (*ZmaPPO2* and *ZmaPPO6*) are interrupted by introns immediately after the conserved histidine residue in the FFPWH sequence motif (Supplemental Table 2.3), which suggests that intervening sequences within the encoded functional domain do not interfere with expression. Few *PPOs* were found to contain introns in other positions, but some were found to be inserted at either the 5' or 3' terminus. For instance, poplar

PtrPPO13 is known to have a 164 bp intron at the 5' terminus, similar to the one found in the *A. cherimola PPO* gene (Prieto et al., 2007).

2.5 Discussion

Large scale sequencing projects have facilitated the elucidation of entire genomes and the identification of gene families and their encoded proteins. Here, *PPO* genes from 16 land plant genomes were identified, most of which are being analyzed and reported for the first time. Gene duplication and diversification has led to the formation of large *PPO* gene families in *Physcomitrella*, *Selaginella*, soybean, *Mimulus* and poplar. Phylogenetic analysis revealed that many of the *PPO* genes evolved within a family or clade.

2.5.1 Multiple *PPO* Genes Appear to Correlate with Whole Genome Duplication Events

Large *PPO* gene families of greater than 10 genes were found in *Physcomitrella*, *Selaginella* and soybean (Table 2.1). Surprisingly, the largest number of *PPO* genes was identified in *Physcomitrella*, from which only *PpaPPO1* (Richter et al., 2005) has been previously described. It is suspected that some secondary metabolism genes, such as those associated with the phenylpropanoid pathway, are overrepresented in the *Physcomitrella* genome. For instance, at least 17 putative chalcone synthase (*CHS*) genes have been identified from this non-vascular plant (Jiang et al., 2006; Koduri et al., 2010). Transcriptomics resources are available for *Selaginella* (Weng et al., 2005) and soybean (*Glycine* spp.) (Ji et al., 2006; Umezawa et al., 2008), yet *PPO* sequences from neither

plant have been reported. However, *PPO* genes are known to be present and expressed in soybean as high levels of PPO activity have been detected (Felton et al., 1994; Constabel and Ryan, 1998).

Whole genome duplications (WGDs) are a common phenomenon in plants and have a number of implications such as ploidy, gene duplication and gene diversity. Often after a WGD event, most duplicated genes either become pseudogenes or are degraded (De Bodt et al., 2005). However, some studies suggest that secondary metabolism genes are retained if they are essential for environmental adaptation. Here, the identification of large *PPO* gene families correlates with recent whole genome duplication events. Poplar (Tuskan et al., 2006) and soybean (Schmutz et al., 2010) genomes are known to have undergone recent genome duplications, whereas a similar duplication event has not been detected in the genomes of cucumber (Huang et al., 2009) and papaya (Ming et al., 2008), two species with few *PPO* genes present. Based on this analysis, it appears that most plant species have retained at least two ancestral *PPO* genes, some of which have undergone subsequent duplication and expansion of *PPO* gene families within a clade (Figure 2.2).

Duplicated secondary metabolism genes have sometimes been found to be arranged in a tandem pattern (Hanada et al., 2005). Although the physical organization of *PPO* genes on their respective chromosomes has not been studied for any of the recently sequenced genomes, it can be inferred that in some species, *PPO* genes are tandemly arrayed. For instance in the soybean genome, multiple *PPO* genes appear to be arranged in such a manner (Supplemental Table 2.1). In soybean, nearly three quarters of the genes exist as duplicate or multiple copies, some of which are organized in a tandem

pattern (Nelson and Shoemaker, 2006; Schmutz et al., 2010). In poplar, where the genome has undergone a recent duplication, the *PtrPPO2* subgroup (*PtrPPO2*, *PtrPPO4*, *PtrPPO5*, *PtrPPO10* and *PtrPPO12*) has expanded through additional gene duplications. Although the chromosomal location and exact number of *PtrPPO2* subgroup genes has not been resolved, this analysis suggests that these genes could be arranged in close proximity to each other. Tandem duplications have been documented for the tomato *PPO* gene family, in which all seven *PPO* genes were mapped to chromosome eight (Newman et al., 1993).

2.5.2 The Presence of Introns Contributes to Structural Diversity of *PPO* Genes

PPO genes are often described as intronless (Newman et al., 1993; Dry and Robinson, 1994; Chevalier et al., 1999; Winters et al., 2009) since most of the reported sequences are from eudicotyledonous plants. Surprisingly, this study found a significant number of *PPO* genes in *Physcomitrella*, *Selaginella* and monocotyledonous plants with different intron/exon structures. Introns were primarily found in the sequences encoding the copper-binding domains and the PPO1_DWL domain (Figure 2.3). In contrast, most of the *PPO* genes in the eudicotyledonous plants were found not to contain introns. However, in some eudicotyledonous *PPO*s with introns, the sequences were found at the 5' terminus.

At least one *PPO* gene with two introns and three exons was identified in each of the land plant lineages represented here and suggests a similar ancestral *PPO* gene structure. This intron/exon structure is rare since most of the other intron-containing

PPO genes contain only one intron; this may be a consequence of intron loss during evolution. This is most evident in the highly conserved *PPO* gene structures in monocotyledonous plants. For instance in intron groups C' and D, the intron in the encoded CuA domain is lost in the latter while the 3' intron is retained (Figure 2.3). Interestingly, all of the *PPO* genes in intron group C' are found in the clade Monocot II (Figure 2.2). Meanwhile, the rest of the monocotyledonous *PPO* genes appear to have lost both introns entirely. Conversely in the eudicotyledonous plants, it is suspected that intron gain has occurred. For example, only eudicotyledonous *PPO* genes contain introns at the 5' terminus. A variety of intron patterns have been found in other gene families such as the peroxidases (Passardi et al., 2004; Mathe et al., 2010). It has been debated whether intron loss is more common than intron insertion. However, from this analysis, it can be concluded that both processes occurred in the *PPO* gene family. The ancestral *PPO* genes likely contained two introns that were subsequently lost in most of the recent eudicotyledonous plants. However, intron gain is also evident, such as in poplar *PtrPPO13*.

2.5.3 The Distribution of PPOs in Land Plants

Green algae (chlorophytes) are considered the closest algal relatives of land plants. No evidence of *PPO* genes was found in four unicellular species of green algae (*C. reinhardtii*, *M. pusilla*, *O. lucimarinus* and *O. tauri*). Similarly, the multicellular species *V. carteri* does not contain *PPO*s. Therefore with the current genomics resources, the studies carried out here suggest that *PPO* genes emerged concurrently with the first land plants. Biochemical studies by Sherman et al. (1991) found significant levels of

PPO activity in the liverwort *Marchantia polymorpha*, an older bryophyte lineage than *Physcomitrella* (Qiu et al., 2006). Interestingly, class III peroxidases (EC 1.11.1.7), a highly diversified group of heme-binding oxidative proteins, are also absent in green algae but have been isolated from *Marchantia* (Passardi et al., 2004).

The *Physcomitrella patens* genome is a pivotal resource for understanding land plant evolution as it is a connection between chlorophytes and tracheophytes. The phylogenetic analysis showed that most of the *Physcomitrella* PPOs form a monophyletic clade (Figure 2.2). On the other hand, PPOs from the seedless tracheophyte *Selaginella* are separated into two clades. In the clade Selaginella I, there are three PPOs whereas in Selaginella II/Eudicot II, eight PPOs cluster with the *Mimulus* PPOs. Therefore, this suggests that features of present day PPOs may have arose from an ancestor of the seedless, vascular plants.

It is evident from the phylogenetic reconstruction that there are several well supported clades (above 80%). From these results, the monocotyledonous PPOs evolved from a common ancestor of the Poaceae that contained at least three *PPO* genes. Further support for the phylogenetic relationship depicted in clades Monocot II and Monocot IV comes from a recent phylogenetic analysis of a barley *PPO* gene that contains three exons; one clade consisted of the two intron and three exon *PPO* genes including the barley *PPO* sequence, and the other clade consisted of the putative signal peptide encoding *PPO* genes (Taketa et al., 2010).

The structure of the eudicotyledonous PPO clades, however, is not as defined as the monocotyledonous clades. Although there are obvious clades in which gene duplications have contributed to the expansion of PPO families, the relationship of PPOs

between different species is not clear. One explanation for this is the effects of the rapid emergence and radiation of the angiosperms; the phylogeny of the rosids, the largest group of eudicots, has been difficult to elucidate as they have been documented to diverge rapidly under a short period of time (Wang et al., 2009). Therefore, PPOs may have also rapidly evolved during the speciation events to facilitate plant adaptation.

This analysis found a subset of unique PPOs that are not retained within species-specific clades. *Physcomitrella* PpaPPO5, soybean GmaPPO11, *Mimulus* MguPPO1, poplar PtrPPO11 and PtrPPO13 and grape VviPPO4 are all sequences that do not cluster with their respective clades and do not appear to have undergone gene duplication. This may suggest a specialized function. In comparison to the other *PPO* gene families identified here, the poplar *PPO* gene family is unique as it contains four single-copy *PPO* genes of which one, *PtrPPO13*, has a close homolog in other plants. One interesting observation is that a homolog of the wound-inducible *PtrPPO1* is not found in any other species and thus is specific to poplar. Together, this finding suggests that poplar has retained its PPO repertoire for specialized purposes, which may relate to its diverse suite of phenolic metabolites.

From this, it is clear that *PPO* genes are present in nearly all land plants, except Arabidopsis. The speculation is that the common ancestor between Brassicales and Malvales contained *PPO* genes, as sequences were identified in papaya (Table 2.1). In addition, a recent study reported the isolation of PPOs from cotton (*Gossypium hirsutum*) cell suspensions (Kouakou et al., 2009). Consequently, after the divergence of Brassicales and Malvales and the most recent genome duplication in Arabidopsis, it is speculated that Arabidopsis successively lost *PPO* genes. In part, this may be attributed

to the phenolic chemistry that exists in Arabidopsis. For instance, the composition of phenolic compounds derived from the Arabidopsis phenylpropanoid pathway is less complex than in other species such as poplar, and therefore may not contain suitable substrates for PPO. This suggests that PPO is not an essential enzyme in Arabidopsis.

2.5.4 PPO Functions are Likely Influenced by Various Ecological Interactions

Multiple gene duplications and the structural diversity of PPOs suggest different physiological functions. The retention of multiple secondary metabolism genes is considered to be beneficial to plant adaptation, survival and ecological interactions (De Bodt et al., 2005). Many studies often implicate PPOs as defense proteins and plant PPOs may have evolved such a role to protect against insects and pathogens. However, this does not seem to be the case in all plants. For example, previous studies by Constabel and Ryan (1998) showed that in maize and other monocotyledonous plants, PPOs are not significantly wound-induced which suggests that other physiological functions for PPOs are likely.

Most of the plants studied here have diverse secondary metabolism pathways, which could relate to other physiological roles for PPOs. For instance in *Physcomitrella*, various phenolic compounds are considered to function as UV-B sunscreens (Rausher, 2006). In monocotyledonous plants such as *Sorghum*, many phenolic compounds are also present and include hydroxycinnamic acids, flavonoids and proanthocyanidins, similar to poplar (Dicko et al., 2006). *Mimulus* is a plant of ecological significance due to its flower morpohogy and colour variation and is suspected to contain diverse phenolic

compounds that are potential substrates for PPOs. Therefore, the evolution of phenolics and phenylpropanoids in land plants appears to be correlated with the emergence of PPOs.

PPOs as biosynthetic enzymes (ie. hydroxylases) have received less attention and only a handful of studies have been published. A unique example of a PPO homolog involved in flavonoid synthesis in the phenylpropanoid pathway is aureusidin synthase (AmAS1) in snapdragon, which catalyzes the hydroxylation and oxidative cyclization of chalcones to aurones (Nakayama et al., 2000; Nakayama et al., 2001). Interestingly, AmAS1 was found to localize to the vacuole, an unusual localization for PPOs (Ono et al., 2006). Another example of PPO involved in the phenylpropanoid pathway is in creosote bush (*Larrea tridentata*), in which the hydroxylase LtH is involved in the synthesis of 8-8' linked lignans (Cho et al., 2003). Therefore, it appears that PPOs are important for the enzymatic conversions in the phenylpropanoid pathway, but are not limited to this as a role in the biosynthesis of betalains has been described (Steiner et al., 1999; Yamamoto et al., 2001; Strack et al., 2003; Gandía-Herrero et al., 2005).

With the recent identification of a novel group of putative secretory pathway PPOs, this lends further support to the idea that there are previously unsuspected functions for PPOs in phenolic metabolism. As a first step to defining such a function, a more detailed analysis of the poplar *PPO* gene family was undertaken (see Chapter Three). In particular, work was focused on poplar PtrPPO13, which is predicted to be non-plastidic and thus, is an ideal candidate for having a novel poplar biosynthetic function.

Chapter 3. The Polyphenol Oxidase Gene Family in Poplar: Phylogeny, Differential Expression and Identification of a Novel, Vacuolar Isoform

3.1 Abstract

Polyphenol oxidases (PPOs) are oxidative enzymes that convert monophenols and *ortho*-diphenols to *ortho*-diquinones in the presence of molecular oxygen. The quinone products are highly reactive and can modify cellular constituents, such as dietary proteins. The induction of PPO in some plants as a result of wounding and pathogen infection has implicated a role in defense. However, PPO-like hydroxylases involved in lignan and pigment biosynthesis have also been demonstrated. Here, the first genome-based *PPO* gene family investigation is presented. Analysis of the *Populus trichocarpa* genome identified a *PPO* gene family of nine genes, including both recently duplicated and divergent sequences that are 36-98% identical at the amino acid level. RT-PCR gene expression profiling of poplar (*Populus* spp.) tissues and organs revealed that the gene family is differentially expressed during normal development, and only a small subset of genes is responsive to wounding, methyl jasmonate or leaf rust (*Melampsora medusae*) infection. TargetP analysis of the PtrPPO13 N-terminal sequence predicted its localization to the secretory pathway. Subcellular localization studies of PPO-GFP fusion proteins revealed that PtrPPO13 is localized to the vacuole. Together, these findings suggest that the poplar PPO family is diversified in their physiological functions.

3.2 Introduction

In plants, polyphenol oxidases (PPOs) are copper-binding enzymes that typically catalyze the oxidation of *o*-diphenols to *o*-quinones in the presence of molecular oxygen (diphenolase activity; EC 1.10.3.1). In some plants however, PPOs catalyze the hydroxylation and a subsequent oxidation of specific monophenolic substrates (monophenolase activity; EC 1.14.18.1). The quinone end products are highly reactive and spontaneously cross-link with amino acids, proteins and other phenolic compounds to form brown polymers that appear in plant extracts and wounded tissues (Parveen et al., 2010). Thus, PPO activity is often undesirable and its impact on food appearance and nutrition has motivated studies in numerous species in which PPOs have been identified. cDNA clones have been isolated from monocotyledonous and eudicotyledonous food plants such as cherimoya (*Annona cherimola*; Prieto et al., 2007), banana (*Musa cavendishii*; Gooding et al., 2001), sugarcane (*Saccharum* spp.; Bucheli et al., 1996) and broad bean (*Vicia faba*; Cary et al., 1992). Apart from normal development, PPO activity and *PPO* gene expression are up-regulated in wounded tissues and are thus often considered as plant defense proteins (Boss et al., 1995; Thipyapong et al., 1995; Stewart et al., 2001; Pinto et al., 2008). However, the physiological functions of PPOs have been shown to be more diverse.

PPO gene families have been described in a number of crop plants, and individual genes typically show distinct expression patterns. The *PPO* gene families from potato (*Solanum tuberosum*) and tomato (*S. lycopersicum*) have been studied most extensively. These plants contain five and seven genes, respectively, which are primarily expressed in fruit, flowers and leaves (Newman et al., 1993; Thygesen et al., 1995; Thipyapong et al.,

1997). In tomato, the identification of a wound-inducible *PPO* has motivated multiple studies that support an anti-herbivore role (Thipyapong and Steffens, 1997). Insect feeding trials with transgenic tomato lines and noctuid lepidopteran pests cotton bollworm (*Helicoverpa armigera*), beet armyworm (*Spodoptera exigua*) and common cutworm (*S. litura*) showed a negative correlation between PPO activity, and larvae consumption and growth (Mahanil et al., 2008; Bhonwong et al., 2009). Leaf material with increased PPO activity was not preferred by young larvae and led to decreased growth rates. Similarly, forest tent caterpillar (*Malacosoma disstria*) larvae reared on transgenic hybrid aspen (*Populus tremula* x *P. alba*) foliage with increased PPO activity also showed decreased growth rates, but only when derived from older egg masses (Wang and Constabel, 2004b). However, additional studies in the same transgenic hybrid aspen lines with lymantriid lepidopterans gypsy moth (*Lymantria dispar*) and white-marked tussock (*Orgyia leucostigma*), did not show a clear correlation between increased levels of PPO and insect performance (Barbehenn et al., 2007). Therefore, the efficacy of PPO as a defensive protein may depend on the specific plant-insect interaction and other as yet unknown factors.

In addition to multiple studies supporting an anti-herbivore role for PPOs, in some plants, biosynthetic roles for PPO-like enzymes have been described. In snapdragon (*Antirrhinum majus*), aureusidin synthase (AmAS1) catalyzes the hydroxylation and oxidative cyclization of chalcones (2’,4’,6’,4-tetrahydroxychalcone) to aurones (aureusidin) in the phenylpropanoid pathway, and unexpectedly was found to be a 39 kDa copper-binding PPO-like enzyme (Nakayama et al., 2000). Similarly a PPO-like hydroxylase, larreatricin hydroxylase (LtH), was purified from creosote bush (*Larrea*

tridentata) (Cho et al., 2003). This 43 kDa enzyme has been shown to specifically convert the (+)-larreatricin enantiomer to (+)-3'-hydroxylarreatricin, a precursor for 8-8' linked lignans. PPOs are also thought to be involved in the synthesis of precursors in the betalain pathway. In pokeweed (*Phytolacca americana*), early work by Joy et al. (1995) showed that the accumulation *PPO* transcripts is correlated with the appearance of betalains in ripening fruit. This led to investigations in other betalain-accumulating plants such as portulaca (*Portulaca grandiflora*) in which a 53 kDa PPO partially purified from callus cultures was found to hydroxylate tyrosine to produce 3,4-dihydroxyphenylalanine, an early intermediate of the betalain pathway (Steiner et al., 1999; Yamamoto et al., 2001).

PPOs are nuclear-encoded proteins and nearly all known PPOs in leaves are directed into the thylakoid lumen, via a bipartite N-terminal chloroplast transit peptide (cTP). This has been demonstrated experimentally by monitoring the import of a [^{35}S]methionine-labeled 67 kDa PPO precursor from tomato; after importing and processing, the mature 59 kDa protein accumulated in the thylakoid lumen of isolated chloroplasts (Sommer et al., 1994; Koussevitzky et al., 1998). Bipartite cTPs are variable in length and sequence but consist of a stromal-targeting domain enriched in serine and threonine residues, followed by a thylakoid-targeting domain (TTD) (Keegstra and Cline, 1999). An exception to the plastidic compartmentalization of PPO proteins is the vacuolar localization of AmAS1. The first 53 N-terminal amino acids of AmAS1 target the enzyme through the secretory pathway, and ultimately the vacuole lumen via the golgi apparatus (Ono et al., 2006). This was the first demonstration of a PPO-like

enzyme localized to the vacuole, and suggests that other PPO homologs may also localize to an organelle other than the chloroplast.

The genus *Populus* (aspens, cottonwoods and poplars) has been a focus for research on PPOs from a chemical ecology perspective, and the herbivore-induced leaf-specific *PPO1* has been extensively characterized in aspen (*P. tremuloides*; Haruta et al., 2001b) and hybrid poplar (*P. trichocarpa* x *P. deltoides*; Constabel et al., 2000; Wang and Constabel, 2004a, b). Two additional differentially expressed *PPO* genes, *PtdPPO2* and *PtdPPO3*, were later characterized in hybrid poplar and found to be expressed in stems and roots (Wang and Constabel, 2004a) and thus, *Populu*s is an excellent system for research on PPOs. The genus contains a broad array of phenolic compounds which could potentially interact with PPOs (Constabel and Lindroth, 2010). Salicin-based phenolic glycosides (PGs) and proanthocyanidins (PAs, or condensed tannins) together can constitute nearly one-third of poplar leaf dry mass (Lindroth and Hwang, 1996), and both are considered ecologically significant compounds. Hydroxycinnamate derivatives (HCDs) such as chlorogenic acid are also commonly found in poplar and other plants, and are common substrates for PPOs (Haruta et al., 1999; Mazzafera and Robinson, 2000; Wang and Constabel, 2003). Furthermore, *Populus* has been developed as a model plant and genomics resources including cDNA clones, ESTs and microarray data sets are abundant (Sterky et al., 2004; Ralph et al., 2006; Miranda et al., 2007). These tools, complemented with a sequenced genome (Tuskan et al., 2006), has facilitated the identification and characterization of various gene families and their biological significance.

The availability of the *P. trichocarpa* (Torr. & Gray) genome sequence provided an opportunity to study PPOs from genome-wide perspective. The current work represents the first analysis of a *PPO* gene family from a sequenced genome. Phylogenetic analysis and gene expression profiling of *PPO* genes in various tissues and organs during normal development and following stress are described. In addition, a novel PPO was discovered and subsequent subcellular localization studies using PPO-GFP fusion proteins demonstrated its vacuolar targeting. Together, these findings suggest that PPOs may have diverse biological functions in poplar.

3.3 Materials and Methods

3.3.1 Plant Growth Conditions

Hybrid poplar (*P. trichocarpa* x *P. deltoides*) clone H11-11 was propagated from greenwood cuttings and maintained under a 16 h photoperiod at the University of Victoria Bev Glover Greenhouse as described previously (Major and Constabel, 2006). Rooted cuttings were grown in peat (Sunshine Mix 4) supplemented with the following slow release nutrients: dolomite lime ($11.4 gl^{-1}$), micronax ($2.9 gl^{-1}$), osmocote ($21.4 gl^{-1}$) and superphosphate ($1.1 gl^{-1}$). Irrigation supplemented with 20-20-20 fertilizer ($0.1 gl^{-1}$) was provided daily. Greenhouse-grown trees were used for all experiments at an age of 10-12 weeks old. Leaves were designated according to the Larson and Isebrands (1971) leaf plastochron index (LPI) numbering system. The index leaf LPI 0, was defined as the first developing leaf with a lamina length of at least 20 mm, and subsequent leaves were designated basipetally thereafter. Midveins were removed from all experimental leaves. After collection, all samples were immediately frozen in liquid nitrogen and stored at

-80^{o}C until ready for analysis.

For the tissue and organ series, apical and young leaves refer to the apical meristem and LPI 0-4, respectively. Mature leaves refer to LPI 12-14, and petioles and midveins were harvested from LPI 10-14. The periderm and stem were collected by separating the green tissues from the stem. Young roots were collected from root tips and old roots were collected from main roots extending from the base of the stem. Dormant buds were collected from outdoor-grown *P. trichocarpa* clone Nisqually-1 at the University of Victoria Centre for Forest Biology compound. Catkins and immature fruit were collected from wild *P. trichocarpa* at the University of Victoria.

Mechanical wounding with pliers and MeJA application were performed as described previously (Wang and Constabel, 2004a). For the wounding experiments, leaf margins of LPI 9-11 were crushed with pliers at hourly intervals. For the MeJA experiments, leaf laminae LPI 7-9 were sprayed twice with 10µM MeJA in 0.1% (v/v) Triton X-100. Wounded and MeJA-sprayed leaves were harvested after 24 h. *Melampsora medusae* infection was performed as described previously (Miranda et al., 2007). *M. medusae* urediospores were diluted in water and sprayed onto the abaxial surface of the leaf laminae. Inoculated trees were covered for 48 h with clear, lightweight plastic bags to maintain a humid environment during the infection. Leaves corresponding to LPI 5-7 were collected at 1, 3, 6 and 9 days post inoculation (dpi).

3.3.2 *PPO* Gene Identification and Phylogenetic Analysis

TBLASTX searches were performed on the US Department of Energy (DOE) Joint Genome Institute (JGI) *Populus trichocarpa* genome version 1.1 (http://genome.jgi

psf.org/Poptr1_1/Poptr1_1.home.html) using hybrid poplar *PtdPPO1*, *PtdPPO2* and *PtdPPO3* cDNA sequences as queries (GenBank Accessions AF263611, AY665681 and AY665682, respectively). All retrieved gene models were conceptually translated and run through NCBI BLASTP to verify the presence of the copper-binding domains. *PPO* gene model annotations identified from genome version 1.1 were compared to previous annotations in version 1.0. In some instances, genome version 1.0 annotations were more accurate and used for analysis.

The amino acid alignment was generated using ClustalW (Thompson et al., 1994) on default parameters, followed by manual adjustments using BioEdit (Hall, 1999). A neighbour-joining phylogenetic tree based on this edited alignment was generated using MEGA version 4.0 (Tamura et al., 2007). Pairwise deletion was applied to exclude missing amino acid residues. Distances were estimated using the Dayhoff matrix and 1000 bootstrap replicates were applied.

3.3.3 Molecular Methods and RT-PCR

Genomic DNA was isolated from *P. trichocarpa* leaves, as described by Doyle and Doyle (1990), and subsequently used for PCR amplification with gene-specific primers (Supplemental Table 3.1) to verify incomplete gene models. PCR products were cloned into pGEM-T Easy (Promega) and sequenced for product verification.

Total RNA was isolated from tissues and organs isolated from *P. trichocarpa* and hybrid poplar clone H11-11 using the CTAB phenol/chloroform procedure as described previously (Haruta et al., 2001a). 25µg total RNA was treated with DNase I (Invitrogen) according to the manufacturer's recommendations. 1µg DNase-treated RNA was used

for first strand cDNA synthesis with an oligo (dT_{20}) primer and Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's recommendations.

RT-PCR analysis was performed with 100ng of first strand cDNA template, gene-specific primers (Supplemental Table 3.1) and *Taq* DNA polymerase (Invitrogen) using the following amplification protocols: 93^{o}C 30 s, 54^{o}C 30 s, 73^{o}C 1.30 min (*PtrPPO4/5*, 35 cycles); 93^{o}C 30 s, 55^{o}C 30 s, 73^{o}C 1.30 min (*PtrPPO1*, 35 cycles; *PtrPPO2*, 35 cycles; *PtrPPO3*, 35 cycles; *PtrPPO10*, 35 cycles; *PtrPPO12*, 35 cycles); 93^{o}C 30 s, 58^{o}C 30 s, 73^{o}C 1.30 min (*PtrPPO11*, 27 cycles; *PtrPPO13*, 35 cycles; actin, 25 cycles). RT-PCR products were electrophoresed on 2.0% agarose gels and detected with Sybrsafe (Invitrogen) and cloned into pGEM-T Easy (Promega) and sequenced for product verification.

3.3.4 Gene Model Verification for *PtrPPO11* and *PtrPPO13* and Construction of PPO-GFP Fusion Constructs

The predicted gene model fgenesh4_pm.C_LG_IV000465 for *PtrPPO13* from genome version 1.0 gene was confirmed by amplifying the entire sequence from *P. trichocarpa* genomic DNA, and from a hybrid poplar H11-11 cDNA library (Constabel et al., 2000). The predicted gene model gw1.182.27.1 for *PtrPPO11* from genome version 1.1 was also verified by amplifying the entire sequence from the H11-11 cDNA library. For localization construct cloning, primers with NotI restriction sites (Supplemental Table 3.1) were used to amplify the 5' termini sequences of *PtrPPO11* and *PtrPPO13* from the H11-11 cDNA library. PCR amplification of the fragments was performed with *Pfx* DNA polymerase (Invitrogen) as follows: 94^{o}C 30 s, 55^{o}C 30 s, 68^{o}C 1.30 min (35

cycles). Amplified fragments were first cloned into pGEM-T Easy vector (Promega) and then subcloned into a modifed Gateway donor vector pENTR-T (Invitrogen) via the NotI restriction sites. After sequence verification, the cloned inserts were recombined into the plant transformation Gateway destination vector pH7FWG2 (Karimi et al., 2002) using LR Clonase II (Invitrogen), according to the manufacturer's recommendations.

3.3.5 Particle Bombardment and Confocal Imaging

1μg plasmid DNA was coated onto 1μM diameter gold beads (BioRad) according to the manufacturer's recommendations. Bombardment was performed using the Biorad He/PSD 1000 at 1100psi and a 6cm shooting distance onto onion epidermal cells (Scott et al., 1999). After bombardment, cells were maintained on Murashige and Skoog media (Sigma) at room temperature for 48 h, in darkness. Wet mounts of onion epidermal sections were prepared and imaged using a Nikon Eclipse TE2000-U inverted confocal laser scanning microscope. GFP was detected at an excitation wavelength of 488 nm and an emission wavelength of 515 nm. Transformed cells were viewed using a 20x/0.75 objective lens (Plan Apo). Optical z-stack sections were collected at 1.50μm increments, with an average of 25 sections. Images were processed using EZ-C1 viewer version 3.90 and Adobe Photoshop 7.0.

3.4 Results

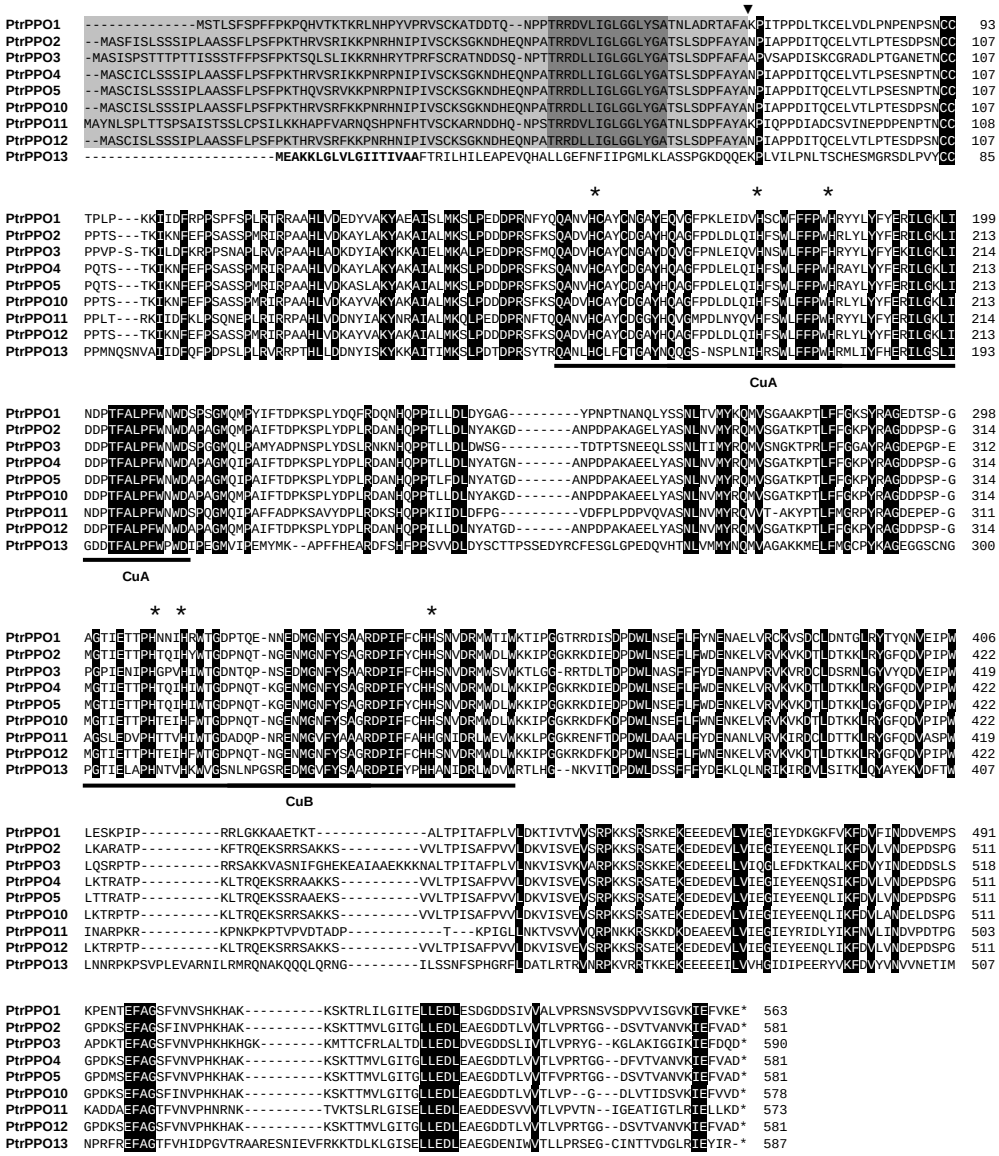
3.4.1 Bioinformatic Identification and Validation of *PPO* Gene Models in the *Populus* Genome

To identify new PPO-encoding genes in poplar, TBLASTX searches were carried out on the *Populus trichocarpa* genome at the Joint Genome Institute (JGI) portal (version 1.1; http://genome.jgi psf.org/Poptr1_1/Poptr1_1.home.html). Hybrid poplar *PtdPPO1*, *PtdPPO2* and *PtdPPO3* cDNA sequences were used as queries (Constabel et al., 2000; Wang and Constabel, 2004a). Initial BLAST results retrieved 23 gene models, including the *PPO1*, *PPO2* and *PPO3* genes. Inspection of the predicted gene models found several truncated genes and other incomplete sequences which appeared to represent genome assembly artifacts, and were eliminated from further analysis. Conceptual translation of the useful gene models resulted in six novel full-length PPO sequences which were retained for further analysis.

Previously, the *P. trichocarpa PPO* genes named *PtrPPO4*, *PtrPPO5*, *PtrPPO6*, *PtrPPO7* and *PtrPPO8* had been annotated (Tuskan et al., 2006). However, the annotated *PtrPPO6* gene spanned only 414 bp, while *PtrPPO7* contained multiple internal termination codons and *PtrPPO8* is a truncated, 1274 bp sequence. Therefore, these sequences appear not to encode functional proteins and were not included in further analysis. Collectively, nine *PPO* genes were identified including the previously described orthologs of *PPO1*, *PPO2* and *PPO3* (Figure 3.1). The remaining unannotated putative full-length *PPO* genes were named *PtrPPO9* through *PtrPPO13*. The complete list of names and correspondence to gene models is listed in Supplemental Table 3.2.

Among the remaining gene models, some ambiguity could not be resolved. The

Figure 3.1 ClustalW alignment of PPO amino acid sequences identified from the poplar (*P. trichocarpa*) genome. The N-terminal chloroplast transit peptide (light grey) including the thylakoid transfer domain (dark grey) is shown. This sequence is predicted to be cleaved after the alanine (AxA) motif (von Heijne et al., 1998), represented by the inverted triangle. The putative 18 amino acid signal peptide for PtrPPO13 is shown in bold font. The two copper-binding domains (CuA and CuB) as indicated by van Gelder et al. (1997) are underlined and the conserved histidine residues are indicated by asterisks. Identical amino acids are shaded in black. Actual amino acid positions are indicated on the right end of each sequence. For the complete list of *PPO* gene models, refer to Supplemental Table 3.2.


PtrPPO1
PtrPPO2
PtrPPO3
PtrPPO4
PtrPPO5
PtrPPO10
PtrPPO11
PtrPPO12
PtrPPO13
CuA
CuA
CuB

predicted gene model for *PtrPPO9* is problematic since the annotated 1573 bp open reading frame (ORF) predicted a non-functional protein with multiple internal termination codons. Manual inspection of the sequence facilitated the identification of a 1698 bp ORF, but due to a frame shift it did not encode the expected plastidic transit peptide. The latter predicted gene model for *PtrPPO9* is consistent with the poplar genome version 1.0 gene model eugene3.01780023. Thus the 5' nucleotide sequence was experimentally verified by PCR amplification and sequencing of the product from *P. trichocarpa* genomic DNA. However, subsequent RT-PCR gene expression experiments failed to detect *PtrPPO9* (see below); due to this result and the sequence ambiguity at the 5' terminus, the gene model for *PtrPPO9* still needs to be confirmed but was excluded from this analysis.

The gene models for *PtrPPO10* and *PtrPPO12* were also found to be incomplete and lack internal portions of coding sequence. *PtrPPO10* was incomplete at the 3' terminus and the available ESTs do not cover this area. Therefore PCR using primers flanking the region of interest was carried out, and an amplified fragment was used to complete the *PtrPPO10* gene model. Similarly, the *PtrPPO12* gene model was missing a significant portion of the internal coding sequence within the encoded CuB domain, which could be determined experimentally by PCR amplification of *P. trichocarpa* genomic DNA and sequencing. Thus an additional 633 bp of nucleotide sequence was used to complete the *PtrPPO12* gene model. However, since *PtrPPO10* and *PtrPPO12* are 99% identical at the nucleotide level (Table 3.1) it is therefore possible that these sequences represent allelic variants of the same gene.

Here, BLAST searches of the poplar genome identified two other unannotated

Table 3.1 Percent identity of poplar PPO nucleotide (italicized) and protein sequences.

	PtrPPO1	PtrPPO2	PtrPPO3	PtrPPO4	PtrPPO5	PtrPPO10	PtrPPO11	PtrPPO12	PtrPPO13
PtrPPO1	**100**	65	60	65	65	65	57	66	39
PtrPPO2	*72*	**100**	59	96	95	96	59	97	37
PtrPPO3	*69*	*70*	**100**	60	59	60	56	60	37
PtrPPO4	*73*	*98*	*70*	**100**	98	94	59	96	37
PtrPPO5	*72*	*98*	*70*	*99*	**100**	93	59	95	37
PtrPPO10	*72*	*98*	*70*	*97*	*96*	**100**	59	98	36
PtrPPO11	*67*	*69*	*68*	*69*	*69*	*69*	**100**	59	38
PtrPPO12	*73*	*98*	*70*	*98*	*98*	*99*	*69*	**100**	37
PtrPPO13	*49*	*48*	*47*	*48*	*48*	*48*	*47*	*48*	**100**

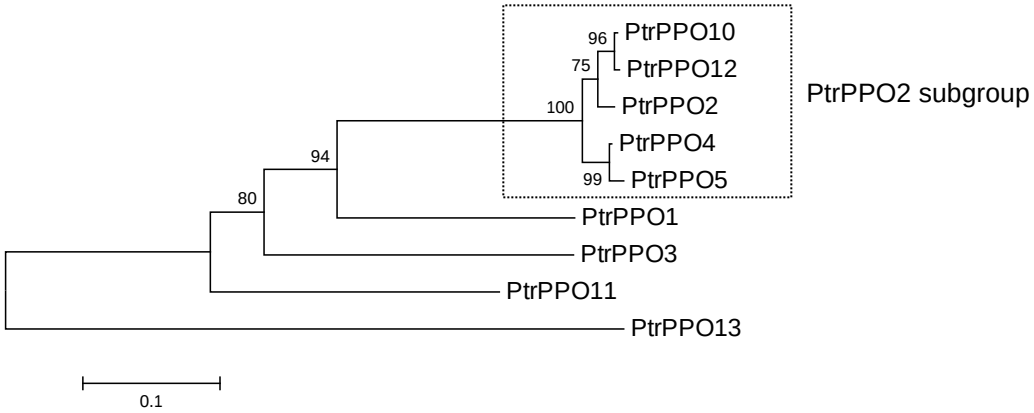
gene models, *PtrPPO11* and *PtrPPO13*. The 1536 bp ORF for *PtrPPO11* was found to be incomplete and my analysis identified a 1722 bp ORF. This was experimentally verified by cloning the entire *PtrPPO11* coding sequence from a hybrid poplar cDNA library (Constabel et al., 2000). Poplar genome versions 1.0 and 1.1 predicted two conflicting gene models for *PtrPPO13*. The former predicted a 1927 bp sequence with a 5' 164 bp intron, while the latter predicted a 1503 bp sequence with a 3' 42 bp intron. Introns are unusual features of plant *PPO* genes and have only been described from banana (Gooding et al., 2001), pineapple (*Ananas comosus*; Zhou et al., 2003), wheat (*Triticum* spp.; Sun et al., 2005; Massa et al., 2007), *Physcomitrella* (Richter et al., 2005) and cherimoya (Prieto et al., 2007). To confirm the presence of the intron and to determine the correct gene model for *PtrPPO13*, the complete genomic sequence for *PtrPPO13* was amplified from *P. trichocarpa* genomic DNA, and compared to the predicted 1764 bp coding sequence in genome version 1.0. Sequence analysis of the cloned fragments revealed that the annotation in poplar genome version 1.0 corresponds to the correct gene model (Figure 3.2), which was further supported by both 5' (DB892816) and 3' (CV239341 and DT480740) GenBank ESTs. Thus, the experimentally verified gene model was taken to be correct and used for further analysis.

The analysis of the *PPO* genes described above indicates that the *PPO* gene family is diverse, with protein identity ranging as low as 36-39% (Table 3.1). However, all the poplar PPOs are characterized by conserved Cu-binding regions (CuA and CuB). Predicted molecular weights of the mature proteins are predicted to be 64.0 to 67.5 kDa, typical of most PPOs (van Gelder et al., 1997). Inspection of the sequence identity values (Table 3.1) and a PPO phylogeny (Figure 3.3) suggests recent duplication for members of

Figure 3.2 The confirmed genomic sequence for *PtrPPO13*. Partial 5' and 3' UTR sequences are shaded in grey. The 164 bp intron is underlined. The CuA domain HCLFC and CuB domain HNTVH sequence motifs are boxed. Numbers on the left indicate the relative amino acid position.

```
  1                                                   M   E   A   K   K   L   G   L   V   L   G   I   I   T   I   V   A   A   F   T   R
     TAG AGC ATA CTG GGA GGA GTG TAG GAT TTT CCG AGC ATG GAA GCA AAG AAA TTG GGC CTG GTC CTC GGC ATA ATC ACT ATT GTG GCA GCT TTT ACT CGG

 22   I   L   H   I   L   E   A   P   E   V   Q   H   A   L   L
     ATT CTT CAT ATC CTT GAA GCA CCG GAA GTC CAG CAT GCG CTT TTG GG- GTA GGT TTT TCT TCT TCT TTT TTA ATG GTG CTT GTC ACC TCA AGC CAC TTT

 37                                                                               G   E   F   N   F   I   I   P   G   M   L   K   L
     CTC TCA ATT AGC TAA TGC AAT GTC ACT CTT CCT TTA TTT TCC TTT TAT TGT TGG TGG TT  --G GAA TTC AAT TTC ATA ATC CCA GGC ATG CTA AAA CTA

 50   A   S   S   P   G   K   D   Q   Q   E   K   P   L   V   I   L   P   N   L   T   S   C   H   E   S   M   G   R   S   D   L   P   V
     GCC TCG TCT CCG GGG AAG GAT CAA CAG GAA AAA CCT CTA GTT ATT TTG CCA AAT TTA ACA TCG TGC CAT GAA TCT ATG GGT CGC TCT GAT CTT CCT GTT

 83   Y   C   C   P   P   M   N   Q   S   N   V   A   I   I   D   F   Q   F   P   D   P   S   L   P   L   R   V   R   R   P   T   H   L
     TAC TGC TGC CCA CCA ATG AAT CAA TCA AAT GTT GCT ATT ATT GAT TTC CAA TTT CCT GAT CCT TCA TTG CCT TTG CGT GTA CGA CGG CCA ACC CAT CTC

116   L   D   D   N   Y   I   S   K   Y   K   K   A   I   T   I   M   K   S   L   P   D   T   D   P   R   S   Y   T   R   Q   A   N   L
     CTC GAC GAC AAC TAT ATT TCC AAG TAC AAG AAG GCT ATA ACC ATA ATG AAG TCT CTT CCA GAC ACT GAT CCT CGG AGT TAT ACT CGT CAA GCC AAT CTG

149   H   C   L   F   C   T   G   A   Y   N   Q   Q   G   S   N   S   P   L   N   I   H   R   S   W   L   F   F   P   W   H   R   M   L
     CAC TGT CTG TTC TGC ACA GGA GCA TAT AAC CAG CAG GGC TCC AAT TCT CCA CTT AAC ATC CAC AGA TCA TGG CTG TTC TTT CCC TGG CAC CGT ATG TTG

182   I   Y   F   H   E   R   I   L   G   S   L   I   G   D   D   T   F   A   L   P   F   W   P   W   D   I   P   E   G   M   V   I   P
     ATC TAT TTC CAC GAA CGT ATC CTT GGA AGT CTC ATA GGC GAT GAC ACA TTT GCT TTG CCA TTC TGG CCT TGG GAC ATT CCT GAA GGA ATG GTG ATT CCT

215   E   M   Y   M   K   A   P   F   F   H   E   A   R   D   F   S   H   F   P   P   S   V   V   D   L   D   Y   S   C   T   T   P   S
     GAA ATG TAC ATG AAG GCT CCA TTT TTC CAT GAA GCA CGT GAC TTT TCC CAC TTT CCA CCA AGT GTG GTT GAT TTA GAC TAC AGT TGT ACT ACG CCT AGC

248   S   E   D   Y   R   C   F   E   S   G   L   G   P   E   D   Q   V   H   T   N   L   V   M   M   Y   N   Q   M   V   A   G   A   K
     AGT GAA GAC TAC AGA TGC TTC GAG AGT GGG TTA GGA CCT GAA GAC CAA GTT CAC ACC AAC TTG GTG ATG ATG TAT AAC CAA ATG GTA GCA GGT GCA AAG

281   K   M   E   L   F   M   G   C   P   Y   K   A   G   E   G   G   S   C   N   G   P   G   T   I   E   L   A   P   H   N   T   V   H
     AAG ATG GAA CTC TTC ATG GGC TGC CCT TAT AAG GCA GGG GAG GGA GGG TCT TGC AAT GGG CCT GGC ACC ATA GAA CTA GCC CCT CAC AAC ACT GTG CAC

314   K   W   V   G   S   N   L   N   P   G   S   R   E   D   M   G   V   F   Y   S   A   A   R   D   P   I   F   Y   P   H   H   A   N
     AAG TGG GTA GGG AGC AAT TTG AAT CCG GGA AGC AGG GAG GAT ATG GGT GTC TTT TAC TCT GCT GCA AGA GAC CCT ATT TTC TAT CCA CAC CAT GCC AAT

347   I   D   R   L   W   D   V   W   R   T   L   H   G   N   K   V   I   T   D   P   D   W   L   D   S   S   F   F   F   Y   D   E   K
     ATC GAT CGT CTT TGG GAC GTT TGG AGG ACG CTC CAT GGG AAC AAG GTT ATC ACT GAT CCT GAT TGG CTA GAC TCT TCC TTT TTC TTT TAT GAC GAG AAA

380   L   Q   L   N   R   I   K   I   R   D   V   L   S   I   T   K   L   Q   Y   A   Y   E   K   V   D   F   T   W   L   N   N   R   P
     TTG CAG CTC AAT AGA ATC AAA ATT CGC GAT GTG CTG AGT ATC ACA AAG CTG CAA TAT GCG TAT GAG AAG GTC GAT TTC ACA TGG CTT AAT AAC CGT CCT

413   K   P   S   V   P   L   E   V   A   R   N   I   L   R   M   R   Q   N   A   K   Q   Q   Q   L   Q   R   N   G   I   L   S   S   N
     AAG CCT TCT GTT CCA CTA GAA GTC GCC CGT AAC ATA CTA AGG ATG AGG CAA AAT GCG AAG CAG CAG CAA CTG CAA CGG AAC GGC ATT CTA TCC TCT AAC

446   F   S   P   H   G   R   F   L   D   A   T   L   R   T   R   V   N   R   P   K   V   R   R   T   K   K   E   K   E   E   E   E   E
     TTC AGT CCC CAT GGC CGA TTT CTA GAC GCT ACC TTA AGA ACA AGG GTC AAT AGG CCA AAA GTC CGA AGA ACC AAA AAA GAG AAG GAA GAA GAG GAA GAG
479   I   L   V   V   H   G   I   D   I   P   E   E   R   Y   V   K   F   D   V   Y   V   N   V   V   N   E   T   I   M   N   P   R   F
     ATC CTG GTT GTG CAT GGA ATT GAT ATC CCA GAG GAA AGA TAT GTT AAG TTT GAT GTT TAT GTT AAT GTC GTT AAT GAG ACG ATT ATG AAT CCG AGG TTC

512   R   E   F   A   G   T   F   V   H   I   D   P   G   V   T   R   A   A   R   E   S   N   I   E   V   F   R   K   K   T   D   L   K
     AGG GAA TTT GCA GGA ACC TTT GTT CAC ATT GAT CCG GGC GTG ACC AGG GCA GCA AGG GAG AGC AAT ATT GAA GTT TTC AGA AAG AAG ACT GAT CTG AAG

545   L   G   I   S   E   L   L   E   D   L   E   A   E   G   D   E   N   I   W   V   T   L   L   P   R   S   E   G   C   I   N   T   T
     TTG GGG ATC TCG GAA CTG TTA GAA GAT TTG GAG GCA GAA GGA GAT GAA AAC ATA TGG GTG ACA TTG TTG CCG AGG AGT GAA GGC TGC ATC AAC ACG ACA

578   V   D   G   L   R   I   E   Y   I   R   *
     GTT GAT GGG TTA AGG ATT GAG TAT ATT AGA TAA ATA AAG CTT CCT TTC TTT CTT TTC TTT CTT CCT TGT
```

Figure 3.3 The poplar PPO family is comprised of both divergent and duplicated sequences. A neighbour-joining phylogenetic tree representing the nine poplar PPOs generated by MEGA 4.0 (Tamura et al., 2007). Pairwise deletion, Dayhoff distance matrix and 1000 bootstrap replications were applied.

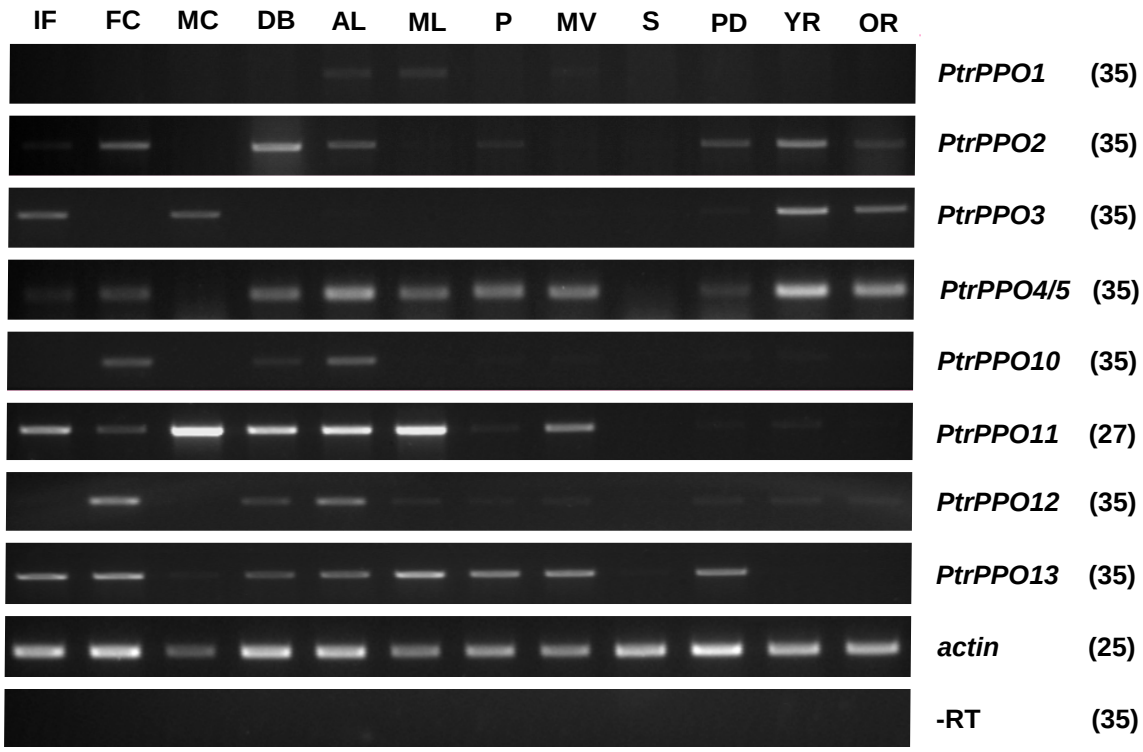

96
PtrPPO10
75
PtrPPO12
100
PtrPPO2
PtrPPO2 subgroup
PtrPPO4
94
99
PtrPPO5
80
PtrPPO1
PtrPPO3
PtrPPO11
PtrPPO13
0.1

the *PPO* gene family, and in particular the PtrPPO2 subgroup. PtrPPO4, PtrPPO5, and PtrPPO10 and PtrPPO12 are at least 93% identical at the protein level, in comparison to PtrPPO2 (Table 3.1). Although these sequences are suspected alleles, nucleotide differences in their respective 5' and 3' UTRs suggests that these are perhaps independent genes. By contrast, PtrPPO3, PtrPPO11 and PtrPPO13 are more distinct and have diverged from other PPOs (Figure 3.3).

3.4.2 The Poplar *PPO* Gene Family is Differentially Expressed during Development, and in Response to Wounding, MeJA and Pathogen Infection

To test the hypothesis that the *PPO* genes identified in poplar are expressed and independently regulated, their expression was profiled by RT-PCR (Figure 3.4). Total RNA was extracted from a series of tissues and organs harvested from young hybrid poplar saplings, and for reproductive tissues, local *P. trichocarpa* trees. Using gene-specific primers (Supplemental Table 3.1), *PPO*-specific products ranging from 253-1145 bp were amplified for each gene and analyzed by gel electrophoresis. Due to the high nucleotide sequence similarity between *PtrPPO4* and *PtrPPO5*, it was not possible to reliably amplify distinct products, and thus cannot distinguish the expression of these two genes. Moreover, transcripts could not be detected for *PtrPPO9* and perhaps is a pseudogene. Overall, the patterns of expression were diverse and not clearly linked to any one tissue or organ. Young apical leaves, female catkins and dormant buds appear to have the greatest number of *PPO* genes expressed. The visible diversity of expression patterns supports the conclusion that each of the *PPO* genes profiled are in fact expressed and regulated independently.

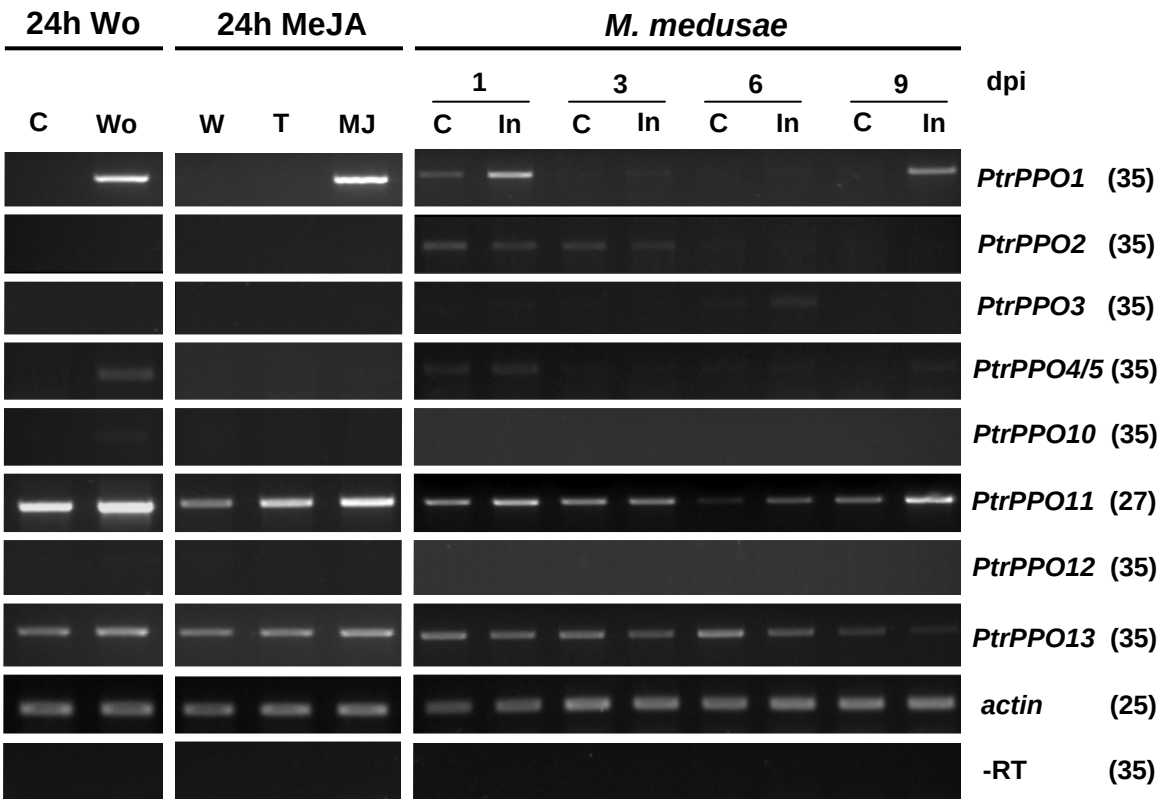
Figure 3.4 Poplar *PPO* genes are differentially expressed during development as profiled by RT-PCR. Expression of tissues and organs from outdoor-grown *P. trichocarpa* (IF-immature fruit, FC-female catkin, MC-male catkin, DB-dormant bud) and greenhouse grown hybrid poplar (*P. trichocarpa* x *P. deltoides*) clone H11-11 (AL-apical leaves, ML-mature leaves, P-petioles, MV-midvein, S-stem, PD-periderm, YR-young root, OR-old root). The numbers enclosed in brackets on the right indicate the number of cycles for each experiment. Expression of all genes is relative to actin. No RT control represents typical results.


IF
FC
MC
DB
AL
ML
P
MV
S
PD
YR
OR
PtrPPO1 (35)
PtrPPO2 (35)
PtrPPO3 (35)
PtrPPO4/5 (35)
PtrPPO10 (35)
PtrPPO11 (27)
PtrPPO12 (35)
PtrPPO13 (35)
actin (25)
-RT (35)

Transcripts for *PtrPPO1*, the previously studied herbivore-induced *PPO*, were detected only at low levels, in both apical and mature leaves (Figure 3.4). *PtrPPO2,* known to be expressed in stems and roots, was found to be expressed in female catkins, dormant buds and apical leaves. *PtrPPO3*, previously found to be expressed primarily in roots (Wang and Constabel, 2004a), was additionally expressed in immature fruit and male catkins. Transcripts of *PtrPPO4/5* are present in most of the organs surveyed, showing a broad pattern of expression. Although *PtrPPO10* and *PtrPPO12* are 99% identical at the nucleotide level, their expression profile differed in that *PtrPPO12* appears more strongly expressed in roots and the stem periderm. However both genes were expressed in three organs: female catkins, dormant buds and apical leaves. Significant accumulation of *PtrPPO11* transcripts were observed in male catkins, dormant buds and in both apical and mature leaves. Compared to all the other *PPO* genes, expression of the divergent *PtrPPO13* was particularly broad, and included most of the organs except roots.

To determine which *PPO* genes are induced by pest or pathogen stresses, leaves of healthy hybrid poplar saplings were subjected to wounding, methyl jasmonate (MeJA) treatment or inoculation with the leaf rust *Melampsora medusae* (Figure 3.5). As expected from previous work, *PtrPPO1* is strongly wound- and MeJA-induced (Constabel et al., 2000). This response also confirmed the efficacy of the induction treatments. *PtrPPO1* was also slightly induced after 24 hour infection by *M. medusae* but subsequently repressed, as observed previously (Miranda et al., 2007). *PtrPPO4/5* were somewhat up-regulated following wounding but not during *M. medusae* infection,

Figure 3.5 Poplar *PPO* genes are differentially expressed following abiotic and biotic stress. RT-PCR expression profile of hybrid poplar (*P. trichocarpa* x *P. deltoides*) clone H11-11 leaves in response to 24 hour mechanical wounding (24h Wo), 24h methyl jasmonate (24h MeJA) treatment and *M. medusae* 1, 3, 6 and 9 days post infection (dpi). Abbreviations are as follows: C-control, Wo-wound, W-water, T-tween, MJ-MeJA, In-infection. The numbers enclosed in brackets on the right indicate the number of cycles for each experiment. Expression of all genes is relative to actin. No RT control represents typical results.


24h Wo
24h MeJA
M. medusae
1
3
6
9
dpi
C
Wo
W
T
MJ
C
In
C
In
C
In
C
In
PtrPPO1 (35)
PtrPPO2 (35)
PtrPPO3 (35)
PtrPPO4/5 (35)
PtrPPO10 (35)
PtrPPO11 (27)
PtrPPO12 (35)
PtrPPO13 (35)
actin (25)
-RT (35)

and the constitutive expression of *PtrPO11* appeared to be slightly enhanced by wounding or MeJA. Surprisingly, *PtrPPO1* was the only *PPO* gene with significant wound and pathogen stress induction. Overall, the distinct pattern of gene expression of each member of the poplar *PPO* gene family in these experiments confirms that at least eight of the nine *PPO* genes identified in the poplar genome show independent expression patterns and supports the hypothesis they represent independent genes.

3.4.3 Poplar PtrPPO13 is a Unique PPO that is Targeted to the Vacuole

Analysis by TargetP 1.1 predicted a signal peptide for PtrPPO13, suggesting that this protein enters the secretory pathway. Furthermore, SignalP 3.0 (Emanuelsson et al., 2007) predicted PtrPPO13 to possess an 18 amino acid N-terminal signal sequence, cleaved after position 18, for entry into the secretory pathway. By contrast, all of the other PPOs are synthesized with a canonical N-terminal plastid transit peptide and are therefore predicted to be targeted to plastids (Table 3.2). Therefore, I sought to experimentally determine the cellular localization of PtrPPO13, specifically to determine if this protein is localized to an organelle other than the chloroplast. A series of PPO-GFP fusion proteins in plant expression vectors under the control of the cauliflower mosaic virus 35S promoter were generated, and expressed in onion epidermal cells following delivery by particle bombardment. After 48 hours, fluorescence due to the expression of the GFP fusion proteins was monitored using confocal laser scanning microscopy.

Preliminary experiments using the complete *PPO* coding sequence failed to produce sufficient GFP expression in the onion epidermal cells. Therefore, the

Table 3.2 TargetP 1.1 predictions for the subcellular localization for poplar PPOs.

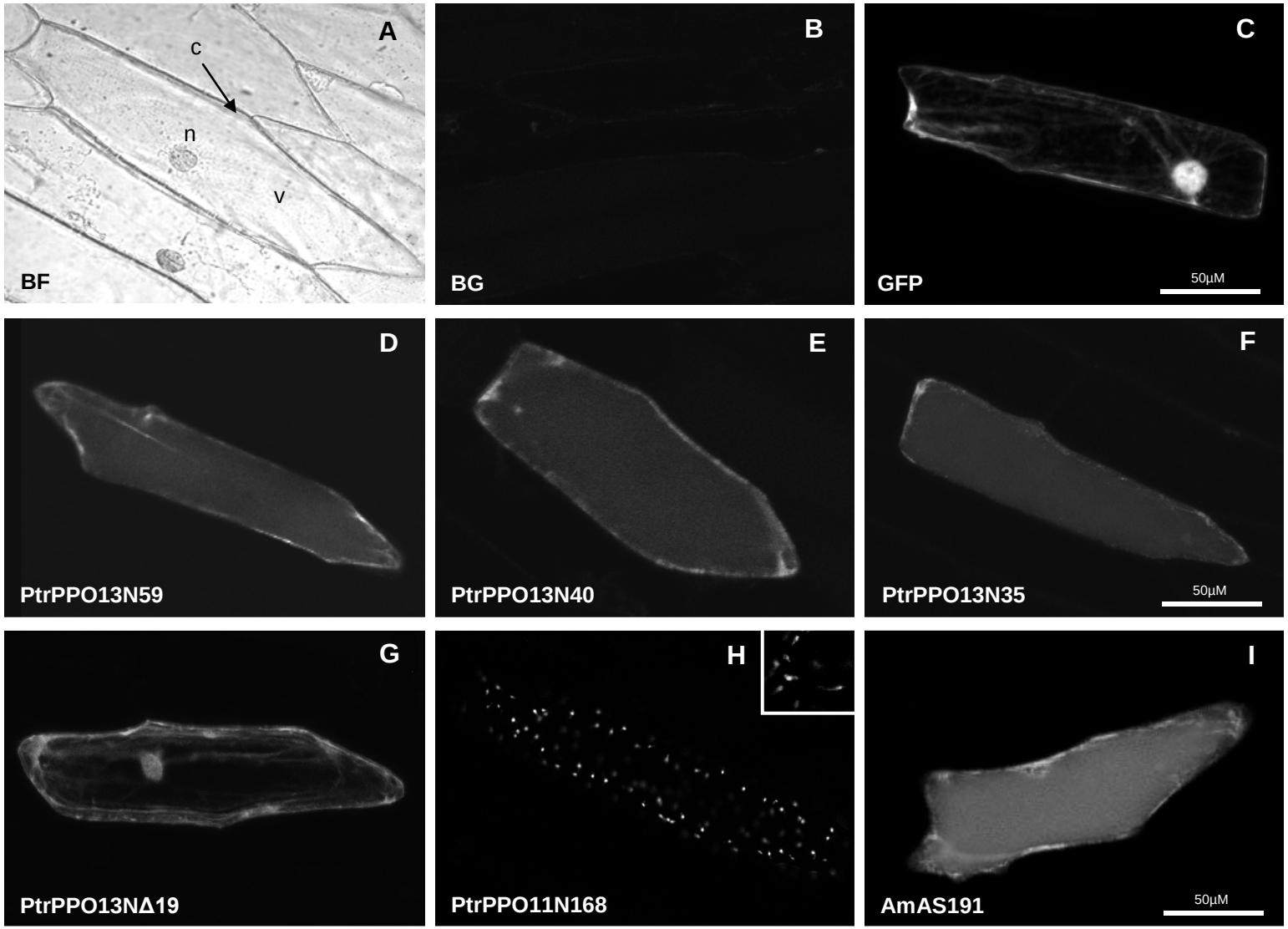
PPO	cTP	mTP	SP	Other	Predicted Location
PtrPPO1	0.677	0.149	0.012	0.262	chloroplast
PtrPPO2	0.918	0.057	0.079	0.022	chloroplast
PtrPPO3	0.956	0.061	0.005	0.160	chloroplast
PtrPPO4	0.914	0.056	0.076	0.023	chloroplast
PtrPPO5	0.893	0.052	0.083	0.040	chloroplast
PtrPPO10	0.916	0.062	0.069	0.061	chloroplast
PtrPPO11	0.933	0.155	0.027	0.061	chloroplast
PtrPPO12	0.916	0.062	0.060	0.030	chloroplast
PtrPPO13	0.008	0.019	**0.949**	0.078	secretory pathway

cTP chloroplast
mTP mitochondria
SP secretory pathway

constructs were modified and only N-terminal regions of the PPO polypeptide were used, similar to the experimental strategy used by Ono at el. (2006) to demonstrate the localization of AmAS1.

Ultimately, a series of GFP fusion constructs containing the N-terminal amino acids was successfully expressed in onion cells (Figure 3.6). Control constructs consisting only of GFP led to fluorescence throughout the cytoplasm surrounding the large central vacuole, and visible as thin strands bisecting the cell (Figure 3.6c). Fluorescence was also detected in the nucleus, the result of passive diffusion of GFP through the nuclear pores (Görlich and Mattaj, 1996). By contrast, PtrPPO13N59 (containing the N-terminal 59 amino acids including the signal peptide) fused to the N-terminus of GFP led to fluorescence throughout the vacuole, seen as an even signal throughout the cell (Figure 3.6d). Expression of the first 40 (Figure 3.6e) and 35 amino acids (Figure 3.6f) of PtrPPO13 also showed fluorescence throughout the vacuole, similar to the expression pattern seen for the first 91 N-terminal amino acids of AmAS1 (Figure 3.6i). Cells expressing PtrPPO13NΔ19-GFP, which lacks the first 19 amino acids at the N-terminus but retains the remainder of the mature polypeptide up to amino acid 59, fluoresced in the cytoplasm rather than in the vacuole (Figure 3.6g) which suggests that PtrPPO13 is synthesized on the ER prior to being transported to the vacuole. Targeting of PtrPPO13 to vacuoles is in clear contrast to the plastidic localization of PtrPPO11N168-GFP, a fusion protein that consists of the transit peptide within first 168 N-terminal amino acids fused to the N-terminus of GFP. Clearly, PtrPPO11 is localized to the plastids, likely leucoplasts, seen as small punctuate structures (Figure 3.6h). Together, these findings indicate that PtrPPO13 is localized to the vacuole and that

Figure 3.6 Poplar PtrPPO13 is a novel PPO targeted to the vacuole. PPO-GFP fusion proteins were expressed in onion epidermal cells 48h after transient transformation by particle bombardment. (A) Bright field image of a typical onion epidermal cell. The cytoplasm (c), nucleus (n) and vacuole (v) are apparent. (B) Typical background fluorescence observed for onion epidermal cells. (C) GFP only. (D) PtrPPO13N59-GFP. (E) PtrPPO13N40-GFP. (F) PtrPPO13N35-GFP. (G) PtrPPO13ΔN19-GFP. (H) PtrPPO11N168-GFP. Inset shows the presence of plastids with stromules. (I) aureusidin synthase AmAS191 (Ono et al., 2006). Each image represents an average compression of 25, 1.50µM optical z-sections. Scale bar represents 50µM.


A
c
n
v
BF
B
BG
C
GFP
50µM
D
PtrPPO13N59
E
PtrPPO13N40
F
PtrPPO13N35
50µM
G
PtrPPO13NΔ19
H
PtrPPO11N168
I
AmAS191
50µM

amino acids 1-18 are necessary and the first 35 amino acids are sufficient to facilitate this localization.

3.5 Discussion

This work describes the *P. trichocarpa PPO* gene family, which contains an estimated nine genes, most with distinct developmental and stress responsive expression profiles. Female flowers and young apical leaves expressed the greatest number of *PPO* genes, and roots the least. *PtrPPO1*, the previously studied herbivore-induced gene, was the only strongly stress-induced *PPO*. One gene, *PtrPPO13*, was investigated in greater detail as its structure, low sequence similarity and vacuolar localization indicates it is a unique PPO form with potentially novel functions.

3.5.1 The Poplar *PPO* Gene Family

PPOs have been characterized in numerous species and are known to exist as multigene families. However, to date a comprehensive analysis of the *PPO* gene family within a sequenced genome has not been reported. Interestingly, Arabidopsis (*Arabidopsis thaliana*), whose genome was sequenced over a decade ago, appears not to contain any genes with similarity to *PPO*. In the *P. trichocarpa* genome, nine genes that comprise the poplar *PPO* gene family were identified. The most populated branch of the family includes *PtrPPO*2 which has also undergone recent duplications, and this *PPO* subgroup appears to contain at least five highly similar genes (Figure 3.3) and is consistent with the prediction of five to seven genes from earlier Southern data (Wang

and Constabel, 2004a). However, since the *PtrPPO2* subgroup sequences are highly identical, and with the current genome annotations, it cannot be excluded that some of the identified genes represent allelic variants. Among the more distantly related PPOs, *PtrPPO11* and *PtrPPO13* represent previously undescribed isoforms, while *PtrPPO3* was previously described as a root-specific form (Wang and Constabel, 2004a).

The poplar PPO family is unusual compared to other angiosperms since it contains divergent PPO sequences. As indicated in the previous chapter, duplicated *PPO* genes are common. Unlike most of the eudicotyledonous *PPO* genes, *PtrPPO13* contains an intron at the 5' terminus and is similar to the one found in cherimoya *AcPPO* (Prieto et al., 2007). Coincidentally, AcPPO is also thought to have a putative N-terminal signal sequence rather than a chloroplast transit peptide. In other plants with characterized *PPO* gene families such as tomato and potato, all of the genes are intronless and PPO proteins are targeted to the chloroplast. In tomato, pioneering studies identified seven fairly similar *PPO* genes (A, A', B, C, D, E and F) clustered within a 165 kb region on chromosome eight, which suggests that tandem duplications gave rise to this family (Newman et al., 1995). *PPO* gene families from potato and red clover (*Trifolium pratense*) are also comprised of genes of high identity (Thygesen et al., 1995; Winters et al., 2009).

3.5.2 Differential Expression of Poplar *PPO*s during Development and Following Stress

The diverse expression profiles of the poplar *PPO* gene family suggest that most of these genes encode functional PPOs. The greatest number of *PPO* genes was expressed

in female catkins, dormant buds and young apical leaves (Figure 3.4). Three *PPO* genes, *PtrPPO4/5*, *PtrPPO11* and *PtrPPO13* were found to be constitutively expressed in leaves. *PtrPPO1* expression is specific to leaves and was found to be the only significant wound-induced *PPO* (Figure 3.5). Strong expression of *PPO* in flowers and leaves is typical, as observed for several Solanaceae species. For example, tobacco *TobP1* is specific to flowers (Goldman et al., 1998), and in potato and tomato, *PPO* transcripts accumulate in both young leaves and flowers (Thygesen et al., 1995; Thipyapong et al., 1997). *In situ* hybridization studies of *PPO* expression in tomato floral components showed complex differential expression of the isoforms during development of the stamens, style, sepals and petals (Shahar et al., 1992; Thipyapong et al., 1997).

A number of reports have shown that wounding, herbivory, and defense signaling molecules can induce *PPO* transcripts and activity, which has been used to infer a function in defense (Constabel and Barbehenn, 2008). For example, it was shown that *PPO* accumulates as part of the tomato wound response modulated by the defense signal systemin (Constabel et al., 1995). However, only one member of the tomato *PPO* family, *PPOF*, was found to be wound-inducible in young leaves (Thipyapong and Steffens, 1997). Apple *pAPO5* (Boss et al., 1995) and pineapple (*Ananas comosus*) *PINPPO1* and *PINPPO2* (Stewart et al., 2001) are two other *PPO* genes that encode wound-inducible PPOs. Interestingly, those PPOs that are wound-induced do not appear to exhibit significant levels of sequence homology.

The experiments here confirm that *PtrPPO1* is wound- and pathogen-inducible (Figure 3.5). However, ESTs and microarray data suggest that *PtrPPO11* may also be involved in pathogen responses. Although the induction of *PtrPPO11* from *M. medusae*

infection was weak under these experimental conditions, microarray expression analysis by Rinaldi et al. (2007) showed that in an incompatible interaction, *M. larici-populina* induced *PtrPPO11* transcript levels approximately nine-fold. *PtrPPO11* is highly expressed in male catkins and leaves. Recent studies comparing *M. larici-populina* infection in male and female trees showed that males were less susceptible to pathogen invasion, which would be consistent with a pathogen defense role for *PtrPPO11* in such trees (Zhang et al., 2010). Experiments with transgenic tomato plants directly demonstrate a role for *PPO* in pathogen resistance, as *Pseudomonas syringae* was observed not to colonize tomato *PPO* overexpressing lines (Li and Steffens, 2002). By contrast, *PPO* suppressed lines resulted in increased susceptibility (Thipyapong et al., 2004). Thus although a wound- and pathogen-inducible *PPO* exists in both poplar and tomato, it is possible that poplar also contains a specific *PPO* involved in pathogen defense.

3.5.3 A Vacuolar PPO Homolog in Poplar

A key finding from this genomics analysis is the identification of PtrPPO13, a novel poplar PPO sequence that shares less than 40% similarity with the other poplar PPOs and has an N-terminal signal sequence typical of proteins directed to the secretory pathway. The results further showed that a 35 amino acid N-terminal fragment, which included a putative 18 amino acid signal peptide, was sufficient to target PtrPPO13 to the vacuole in onion epidermal cells (Figure 3.6). Similar findings were demonstrated in snapdragon in which a 53 amino acid N-terminal fragment of AmAS1 fused to a GFP epitope directed the protein to the vacuole (Ono et al., 2006). AmAS1 is a PPO with

hydroxylase activity, and acts on its substrate in the vacuole (Nakayama et al., 2000; Nakayama et al., 2001). Although PtrPPO13 and AmAS1 are both targeted to the vacuole, BLASTP analysis indicates that these sequences are not similar and are only 39% identical.

In contrast, one of the closest matches to PtrPPO13 retrieved by BLASTP searches in public databases is cherimoya AcPPO (Prieto et al., 2007), which shares 51% similarity with PtrPPO13. AcPPO is expressed mainly in leaves, and to a lesser extent in fruit and flowers. Although originally thought to be localized to the chloroplast, reanalysis of the published AcPPO sequence indicates that AcPPO also contains a putative 33 amino acid signal peptide for co-translational export into the secretory pathway. From this study, it was also revealed that other poplar PtrPPO13 homologs exist, such as grape VviPPO4, discussed in Chapter Two. Both appear to contain a short signal peptide and thus likely do not localize to the chloroplast (Table 3.2). Therefore, it appears that in several species, individual PPO isoforms with non-plastid localization have evolved, and the speculation is that these novel localization patterns are linked to more specialized functions, such as hydroxylases.

Potential functions for these non-plastidic PPOs might include biosynthetic roles, as elegantly established for AmAS1. In snapdragon buds, this enzyme catalyzes the hydroxylation and oxidative cyclization of chalcones (2',4',6',4-tetrahydroxychalcone) to aurones (aureusidin), the compounds responsible for the yellow colouration of the petals (Nakayama et al., 2000; Nakayama et al., 2001). Notably, this enzyme was demonstrated to be localized to the vacuole (Ono et al., 2006). Similarly, a role for PPO in the biosynthesis of betalains, the red/purple (betacyanins) and yellow/orange (betaxanthins)

pigments in Caryophyllales has also been postulated (Yamamoto et al., 2001; Strack et al., 2003; Gandía-Herrero et al., 2005). In the betalain pathway, tyrosine is first hydroxylated to 3,4-dihydroxyphenylalanine (DOPA) and subsequently oxidized to dopa-quinone. In pokeweed, *PPO* transcripts accumulate exclusively in ripening fruit prior to colour formation (Joy et al., 1995), and the partially purified PPO from portulaca showed monophenolase activity only with tyrosine (Steiner et al., 1999; Yamamoto et al., 2001). Work by Gandía-Herrero et al. (2005) also suggested an alternative pathway in which PPO is also involved in the generation of betaxanthins from tyrosine. PPOs have also been shown to be involved as hydroxylases in the synthesis of lignans. Cho et al. (2003) determined that larreatricin hydroxylase (LtH) from creosote bush (*L. tridentata*), a highly stereospecific enzyme that converts (+)-larreatricin to (+)-3'-hydroxylarreatricin, is also a specialized PPO. This enzyme has retained its plastidic transit peptide, however.

Overall, it appears PPOs have been adapted to a variety of functions including biosynthesis of phenolic compounds, and it is speculated that in poplar, PtrPPO13 may also carry out such a biosynthetic role. Poplar is rich in phenolic phytochemicals including hydroxycinnamic acid derivatives and a variety of flavonoids (Constabel and Lindroth, 2010), most of which are believed to accumulate in the vacuole. Future work will involve functional analysis of poplar PtrPPO13 in the context of potential biosynthetic roles by assaying recombinant protein.

Chapter 4. Discussion

4.1 General Discussion

Since the first report of PPO in 1895, the range of physiological functions in plants is still not completely resolved. The literature suggests that the roles associated with this enzyme are diverse. However, PPOs are most often considered as defense-related proteins. Up until recently, advances in plants genomics have accelerated the identification and characterization of genes, their encoded proteins and their biological functions. The availability of the poplar genome complemented with genomics resources such as cDNA clones, ESTs and other plant genomes of evolutionary importance has facilitated the expansion of studies in plant comparative genomics. In this thesis, bioinformatics was used to identify the *PPO* genes in land plants. *PPO* genes were found to be duplicated within most of the genomes surveyed and a phylogenetic reconstruction showed that many PPO families have expanded within a species. Analysis of the N-terminal sequences found that, unexpectedly, not all land plant PPOs are likely directed to the chloroplast. In poplar, PtrPPO13 was experimentally confirmed to posses a functional signal peptide and targeting signal for localization into the vacuole, and is expressed at a basal level in most tissues and organs, except roots. Together, the presented findings are consistent with diverse functions of PPOs.

4.2 Differences in *PPO* Gene Numbers in Land Plant Genomes

PPOs have been previously characterized from a number of angiosperms and multiple genes found, often based on Southern blot analysis. The present, genome-wide analysis confirms that most land plants contain multiple, but a variable number, of *PPO* genes. For example in the non-vascular plant *Physcomitrella*, 13 *PPO* genes were identified, the most thus far (Table 2.1). Other plants such as papaya have four *PPO* genes. Grape also contains four *PPO* genes. However, previous Southern blot analysis identified only a single *PPO* gene, *GPO1* (Dry and Robinson, 1994). Cassava and *Ricinus* appear to have only one *PPO* gene, whereas poplar contains nine *PPO* genes. Phylogenetic analysis revealed that in most plant genomes, two ancestral *PPO* genes underwent subsequent duplications to produce multiple genes and thus, expanded *PPO* gene families (Figure 2.2).

In poplar, it appears that over 90% of the genome has been affected by a recent whole genome duplication event and as a result, genes associated with secondary metabolism appear to be within expanded families (Tuskan et al., 2006). For example, most of the phenylpropanoid genes are duplicated in poplar, relative to Arabidopsis (Tsai et al., 2006). Similarly, here it is found that *PPO* genes in poplar have been duplicated, especially the poplar *PtrPPO2* subgroup which has undergone multiple recent gene duplications. Evidence of this is also apparent in the genome as pseudogenes and additional truncated sequences with premature stop codons of the *PtrPPO2* subgroup were identified, such as *PtrPPO6*, *PtrPPO7* and *PtrPPO9* (Supplemental Table 3.2). Multiple *PPO* genes arranged along the same chromosome suggest recent tandem duplication events (Newman et al., 1993) and genes associated

with environmental adaptations and biological interactions appear to be more likely to be retained after gene duplication events (De Bodt et al., 2005). Thus, the number of genes in a species may reflect the significance of PPO-catalyzed reactions.

4.3 Variation in PPO Targeting Suggests Diverse Functions

In this analysis, PPOs were found to have diverse sequences at the N-terminus. A number of PPOs from *Physcomitrella*, monocotyledonous plants and eudicotyledonous plants were found to lack an N-terminal transit peptide (Supplemental Table 2.1). Bioinformatic subcellular localization prediction programs predicted most of these unique PPO homologs to localize to the secretory pathway. Experimental evidence from this study showed that poplar PtrPPO13 is a vacuolar PPO (Figure 3.6). Furthermore, the subcellular localization prediction of its related homologs including cherimoya AcPPO (Prieto et al., 2007), grape VviPPO4, and *Argemone mexicana* PPO also suggest a secretory localization. This hypothesis is strengthened by experiments with the snapdragon PPO homolog aureusidin synthase, which was shown to localize to the vacuole (Ono et al., 2006). Therefore, the bioinformatics analysis from this study suggests that the diverse localization of PPOs could reflect potentially diverse physiological functions.

AmAS1 was identified to be a PPO homolog involved in the biosynthesis of aurones, a phenylpropanoid compound (Nakayama et al., 2000; Nakayama et al., 2001). The identification of multiple, non-plastidic PPOs in *Physcomitrella* suggest the possibility of their involvement in biosynthesis. Bryophytes contain a diverse composition of phenolic compounds that are believed to be protective against desiccation

and UV-B radiation. Thus, the presence of flavonoid compounds and putative secretory pathway PPOs in this ancient lineage suggests a possible role in the synthesis of phenolic chemicals. Furthermore, these compounds are often synthesized in the cytoplasm (or associated with the endoplasmic reticulum) and stored in the vacuole, which is consistent with the predicted localization for the *Physcomitrella* PPOs. Similarly for other land plant PPOs such as poplar PtrPPO13, a biosynthesis function may also be possible. Many of the biosynthetic steps in the synthesis of specialized phenolic compounds are not known, and could involve PPOs as hydroxylases, for example. Recent studies have shown another group of oxidative enzymes, the class III peroxidases, to also be localized in the vacuole and are involved in alkaloid biosynthesis (Costa et al., 2008). Previous to this, functions associated with peroxidases included cell wall cross-linking and pathogen defense. Together, this suggests that oxidative enzymes share a multiplicity of functions, and some have evolved to accommodate a specific role for a particular species.

4.4 Differential Gene Expression and the Distribution of Land Plant PPOs Also Suggests Multiple Physiological Roles

The induction of *PPO* genes as a consequence of wounding has been reported in some plants, which has led to the assumption that PPOs are defense related proteins. Analysis of the poplar *PPO* gene family has confirmed that *PtrPPO1* is the most significant wound-inducible gene; most of the other poplar *PPO* genes are not expressed in the leaves, in response to abiotic or biotic stresses. This situation is similar to that in tomato where of seven genes, only *PPOF* is significantly wound-inducible (Thipyapong and Steffens, 1997). However, the gene expression studies suggest that poplar also

contains a *PPO* gene involved in pathogen defense in *PtrPPO11*, which is constitutively expressed in the leaves (Figure 3.4, Figure 3.5).

In other plants, a protective role for PPOs may also exist. For instance in grape berry, the *GPO1* homolog identified by Dry and Robinson (1994) is expressed in young leaves and developing berries. Proteomics-based studies of ripening grape berries identified GPO1 to also be co-expressed with other oxidative-stress and pathogenesis-related proteins, which suggests PPOs to be involved in fruit protection during development (Negri et al., 2008). Furthermore, the identification of multiple PPOs in papaya also suggests a protective function. Papaya secretes latex and may contain PPOs that are involved in wound-sealing, similar to the wound-sealing properties of the PPOs in dandelion (Wahler et al., 2009).

The large number of PPOs in *Mimulus* also suggests an ecological role for PPOs. *Mimulus* is a model species for ecological interactions, in part due to its interesting floral traits such as morphology and colouration. Thus, the presence of an expanded PPO family in this species suggests a possible function in pigment synthesis. For example, anthocyanins are found in these plants, which could involve PPOs. Therefore, the phenolic composition of this plant can be studied to elaborate on this hypothesis.

4.5 Conclusions and Future Directions

Plant secondary metabolism is a significant component of plant-environment interactions and their functions are diverse in different species. PPOs share features that are similar to secondary metabolism enzymes that appear to contribute to both ecological adaptation and various physiological processes. The work presented in this thesis shows

that the distribution of PPOs correlates with the emergence of land plants, and rapidly evolved with the appearance of angiosperms. However, their respective functions cannot be inferred from the phylogenetic relationships alone. It was demonstrated that land plant PPOs are diverse in their structure, gene expression profile and subcellular localization which further reflects their diversified functions. Thus, PPOs are not exclusively plant defense proteins and should be considered as enzymes with diverse potential roles, much like the class III peroxidases. It is hoped these findings will lead to a new perspective on PPOs and motivate their characterization in other plants to elucidate their functions.

Here, the presented phylogenetic analysis of the land plant PPOs provides a basis for a more in-depth bioinformatics analysis. As part of a future analysis, one possibility would be to determine which residues in the CuA and CuB domains have undergone positive selection. Another aspect would be to analyze the synteny of *PPO* genes on chromosomes to further elucidate the evolutionary relationships of these genes.

The phylogenetic analysis presented here is also basis for identifying candidates of interest for functional studies. One of the most surprising findings was the number of putative PPOs predicted to localize to the secretory pathway. One aspect of future work could involve the cloning and characterization of these unusual PPO sequences from species besides poplar. Furthermore, the orphan PPO sequences that do not group with their species-specific clades, such as PpaPPO5, MguPPO1 and PtrPPO11 would also be interesting candidates for further characterization.

Most of the identified *PPO* genes to date have not been functionally characterized with respect to enzymatic activity. Thus, an outstanding question that could not be addressed in this study is the biochemical function of poplar PtrPPO13. Although most

of the characterized PPOs convert diphenolic substrates to quinones, there are clear preferences, and it is hypothesized that PtrPPO13 is an exception and is involved in a biosynthetic role.

The primary limitation for biochemical studies related to PPOs is the difficulty of obtaining active recombinant protein. In some studies, recombinant protein expression of PPOs has been successful (Sullivan et al., 2004). In preliminary experiments, recombinant protein expression of PtrPPO13 in *Escherichia coli* was achieved. However, after the protein was purified it was difficult to determine whether the protein was active in the heterologous system. Therefore, a plant-based overexpression system is more likely to be successful.

The absence of PPOs in Arabidopsis presents it as an optimal system to overexpress PPOs via *Agrobacterium*-mediated transformation. This was attempted for several poplar PPOs including PtrPPO13, but the protein was concluded to be incorrectly processed. Thus, plant overexpression systems that are likely more suitable include *Agrobacterium*-mediated transformation of tobacco (*Nicotiana benthamiana*) and poplar hairy root cultures. In the latter, an epitope-tagged PtrPPO13 protein would facilitate purification from its native system and thus be used for biochemical assays. Plant hairy root cultures are mediated by *Agrobacterium rhizogenes* which induce a root phenotype. Its advantage over *Agrobacterium tumefaciens*-mediated plant transformation is the shortened regeneration time. Therefore, it is hoped that expression of PtrPPO13 in the hairy root cultures will provide a foundation for biochemical assays of PtrPPO13 and its potential substrates.

Bibliography

Alexandrov NN, Brover VV, Freidin S, Troukhan ME, Tatarinova TV, Zhang H, Swaller TJ, Lu YP, Bouck J, Flavell RB, Feldmann KA (2009) Insights into corn genes derived from large-scale cDNA sequencing. Plant Molecular Biology **69:** 179-194

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

Bachem CWB, Speckmann GJ, van der Linde PCG, Verheggen FTM, Hunt MD, Steffens JC, Zabeau M (1994) Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. Nature Biotechnology **12:** 1101-1105

Banks JA (2009) *Selaginella* and 400 million years of separation. Annual Review of Plant Biology **60:** 223-238

Barbehenn RV, Jones CP, Yip L, Tran LT, Constabel CP (2007) Limited impact of elevated levels of polyphenol oxidase on tree-feeding caterpillars? Assessing individual plant defenses with transgenic poplar. Oecologia **154:** 129-140

Basu MK, Rogozin IB, Koonin EV (2008) Primordial spliceosomal introns were probably U2-type. Trends in Genetics **24:** 525-528

Bertrand G (1896) Sur une nouvelle oxydase, ou ferment soluble oxidant, d'origine ve´ge´tale. Comptes Rendus de l'Académie des sciences **122:** 1215-1217

Bhonwong A, Stout MJ, Attajarusit J, Tantasawat P (2009) Defensive role of tomato polyphenol oxidases against cotton bollworm (*Helicoverpa armigera*) and beet armyworm (*Spodoptera exigua*). Journal of Chemical Ecology **35:** 28-38

Boss PK, Gardner RC, Janssen BJ, Ross GS (1995) An apple polyphenol oxidase is up-regulated in wounded tissues. Plant Molecular Biology **27:** 429-433

Bucheli CS, Dry IB, Robinson SP (1996) Purification of polyphenol oxidase and isolation of a full length cDNA from sugarcane, a C4 grass. Plant Molecular Biology **31:** 1233-1238

Cary JW, Lax AR, Flurkey WH (1992) Cloning and characterization of cDNAs coding for *Vicia faba* polyphenol oxidase. Plant Molecular Biology **20:** 245-253

Chevalier T, de Rigal D, Mbéguié-A-Mbéuié D, Gauillard F, Richard-Forget F, Fils-Lycaon BR (1999) Molecular cloning and characterization of apricot fruit polyphenol oxidase. Plant Physiology **119:** 1261-1270

Cho M, Moinuddin SGA, Helms GL, Hishiyama S, Eichinger D, Davin LB, Lewis ND (2003) (+)-Larreatricin hydroxylase, an enantio-specific polyphenol oxidase from the creosote bush (*Larrea tridentata*). Proceedings of the National Academy of Sciences of the United States of America **100:** 10641-10646

Christopher ME, Miranda M, Major IT, Constabel CP (2004) Gene expression profiling of systemically wound-induced defenses in hybrid poplar. Planta **219:** 936-947

Coetzer C, Corsini D, Love S, Pavek J, Tumer N (2001) Control of enzymatic browning in potato (*Solanum tuberosum* L.) by sense and antisense RNA from tomato polyphenol oxidase. Journal of Agricultural and Food Chemistry **49:** 652-657

Constabel CP, Barbehenn RV (2008) Defensive roles of polyphenol oxidase in plants. *In* A Schaller, eds, Induced Plant Resistance to Herbivory. Springer, pp 253-269

Constabel CP, Bergey DR, Ryan CA (1995) Systemin activates synthesis of wound-inducible tomato leaf polyphenol oxidase via the octadecanoid defense signaling pathway. Proceedings of the National Academy of Sciences of the United States of America **92:** 407-411

Constabel CP, Bergey DR, Ryan CA (1996) Polyphenol oxidase as a component of the inducible defense response in tomato against herbivores. *In* JT Romeo, JA Saunders, P Barbosa, eds, Phytochemical Diversity and Redundancy in Ecological Interactions. Plenum Press, pp 231-252

Constabel CP, Lindroth R (2010) The impact of genomics on advances in herbivore d efense and secondary metabolism in poplar. *In* S Jansson, RP Bhalerao, AT Groover, eds, Genetics and Genomics of *Populus*. Springer, pp 279-305

Constabel CP, Ryan CA (1998) A survey of wound- and methyl jasmonate-induced leaf polyphenol oxidase in crop plants. Phytochemistry **47:** 507-511

Constabel CP, Yip L, Patton JJ, Christopher ME (2000) Polyphenol oxidase from hybrid poplar. Cloning and expression in response to wounding and herbivory. Plant Physiology **124:** 285-295

Costa MMR, Hilliou F, Duarte P, Pereira GL, Almeida I, Leech M, Memelink J, Ros Barceló A, Sottomayor M (2008) Molecular cloning and characterization of a vacuolar class III peroxidase involved in the metabolism of anticancer alkaloids in *Catharanthus roseus*. Plant Physiology **146:** 403-407

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Research **14:** 1188-1190

Cuff Miller KI, van Holde KE, Hendrickson WA (1998) Crystal structure of a functional unit from octopus hemocyanin. Journal of Molecular Biology **278:** 855-870

De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends in Ecology and Evolution **20:** 2114-2117

Decker H, Schweikardt T, Nillius D, Salzbrunn U, Jaenicke E, Tuczek F (2007) Similar enzyme activation and catalysis in hemocyanins and tyrosinases. Gene **398:** 183-191

Demeke T, Morris CF (2002) Molecular characterization of wheat polyphenol oxidase (PPO). Theoretical and Applied Genetics **104:** 813-818

Derelle E, **Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H** (2006) Genome analysis of the smallest free living-eukaryote *Ostreococcus tauri* unveils many unique features. Proceedings of the National Academy of Sciences of the United States of America **103:** 11647-11652

Dicko MH, Hilhorst R, Gruppen H, Traore AS, Laane C, van Berkel WJH, Voragen AGJ (2002) Comparison of content in phenolic compounds, polyphenol oxidase, and peroxidase in grains of fifty sorghum varieties from Burkina Faso. Journal of Agricultural and Food Chemistry **50:** 3780-3788

Doyle J, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Bethesda Research Laboratories Focus **12:** 13–15

Dry IB, Robinson SP (1994) Molecular cloning and characterization of grape berry polyphenol oxidase. Plant Molecular Biology **26:** 495-502

Duffey SS, Felton GW (1991) Enzymatic antinutritive defenses of the tomato plant against insects. *In* PA Hedin, ed, Naturally Occurring Pest Bioregulators. ACS Press, pp 167-197

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792-1797

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols **2**: 953-971

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science **8:** 978-984

Felton GW, Donato KK, Broadway RM, Duffey SS (1992) Impact of oxidized plant phenolics on the nutritional quality of dietary protein to a noctuid herbivore, *Spodoptera exigua*. Journal of Insect Physiology **38:** 277-285

Felton GW, Donato KK, Del Vecchio RJ, Duffey SS (1989) Activation of plant foliar oxidases by insect feeding reduces nutritive quality of foliage for noctuid herbivores. Journal of Chemical Ecology **15:** 2667-2694

Felton GW, Summers CB, Mueller AJ (1994) Oxidative responses in soybean foliage to herbivory by bean leaf beetle and three-cornered alfalfa hopper. Journal of Chemical Ecology **20:** 639-650

Flurkey WH, Inlow JK (2008) Proteolytic processing of polyphenol oxidase from plants and fungi. Journal of Inorganic Biochemistry **102:** 2160-2170

Gandía-Herrero F, Escribano J, García-Carmona F (2005) Betaxanthins as substrates for tyrosinase. An approach to the role of tyrosinase in the biosynthetic pathway of betalains. Plant Physiology **138:** 421-432

Gerdemann C, Eicken C, Krebs B (2002) The crystal structure of catechol oxidase: New insight into the function of type-3 copper proteins. Accounts of Chemical Research **35:** 183-191

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Nature **296:** 92-100

Golbeck JH, Cammarata KV (1981) Spinach thylakoid polyphenol oxidase: Isolation, activation and properties of the native chloroplast enzyme. Plant Physiology **67:** 977-984

Goldman MH, Seurinck J, Marins M, Goldman GH, Mariani C (1998) A tobacco flower-specific gene encodes a polyphenol oxidase. Plant Molecular Biology **36:** 479-485

Gooding PS, Bird C, Robinson SP (2001) Molecular cloning and characterization of banana fruit polyphenol oxidase. Planta **213**: 748–757

Görlich D, Mattaj IW (1996) Nucleocytoplasmic transport. Science **271:** 1513-1518

Groth G (2002) Structure of spinach chloroplast F_1-ATPase complexed with the phytopathogenic inhibitor tentoxin. Proceedings of the National Academy of Sciences of the United States of America. **6:** 3464-3468

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid Symposium Series **41:** 95-98

Hamberger B, Ellis M, Friedmann M, Souza CDA, Barbazuk B, Douglas CJ (2007) Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: The *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. Canadian Journal of Botany **85:** 1182-1201

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology **148:** 993-1003

Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP (2001a) A Kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides* Michx.): Cloning, functional expression, and induction by wounding and herbivory. Plant Molecular Biology **46:** 347-359

Haruta M, Murata M, Kadokura H, Homma S (1999) Immunological and molecular comparison of polyphenol oxidase in Rosaceae fruit trees. Phytochemistry **50:** 1021-1025

Haruta M, Pedersen JA, Constabel CP (2001b) Polyphenol oxidase and herbivore defense in trembling aspen (*Populus tremuloides*): cDNA cloning, expression and potential substrates. Physiologia Plantarum **112:** 552-558

Huang, S, Li RQ, Zhang ZH, Li L, Gu XF, Fan W, Lucas WJ, Wang XW, Xie BY, Ni PX, Ren YY, Zhu HM, Li J, Lin K, Jin WW, Fei ZJ, Li GC, Staub J, Kilian A, van der Vossen EAG, Wu Y, Guo J, He J, Jia ZQ, Ren Y, Tian G, Lu Y, Ruan J, Qian WB, Wang MW, Huang, Q, Li B, Xuan ZL, Cao JJ, Asan, Wu ZG, Zhang JB, Cai QL, Bai YQ, Zhao BW, Han YH, Li Y, Li XF, Wang SH, Shi QX, Liu SQ, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng ZC, Zhang SP, Wu J, Yang YH, Kang HX, Li M, Liang HQ, Ren XL, Shi ZB, Wen M, Jian M, Yang HL, Zhang GJ, Yang ZT, Chen R, Liu SF, Li JW, Ma LJ, Liu H, Zhou Y, Zhao J, Fang XD, Li GQ, Fang L, Li YR, Liu DY, Zheng HK, Zhang Y, Qin N, Li Z, Yang GH, Yang S, Bolund L,

Kristiansen K, Zheng HC, Li SC, Zhang XQ, Yang HM, Wang J, Sun RF, Zhang BX, Jiang SZ, Wang J, Du YC, Li SG (2009) The genome of the cucumber, *Cucumis sativus* L.. Nature Genetics **41:** 1275-1281

Hunt MD, Eannetta NT, Yu H, Newman SM, Steffens JC (1993) cDNA cloning and expression of potato polyphenol oxidase. Plant Molecular Biology **21:** 59-68

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449:** 463-467

Ji W, Li Y, Li J, Dai CH, Wang X, Bai X, Cai H, Yang L, Zhu YM (2006) Generation and analysis of expressed sequence tags from NaCl-treated *Glycine soja*. BMC Plant Biology **6:** 4

Jiang C, Schommer CK, Kim SY, Suh DY (2006) Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochemistry **67:** 2531–2540

Joy IV RW, Sugiyama M, Fukuda H, Komamine A (1995) Cloning and characterization of polyphenol oxidase cDNAs of *Phytolacca americana*. Plant Physiology **107:** 1083-1089

Kahn V (1976) Polyphenol oxidase isoenzymes in avocado. Phytochemistry **15:** 267-272

Karimi M, Inze D, Depicker A (2002) Gateway vectors for *Agrobacterium*-mediated plant transformation. Trends in Plant Science **7:** 193-195

Keegstra K, Cline K (1999) Protein import and routing systems of chloroplasts. Plant Cell **11:** 557-570

Keilin D, Mann T (1938) Polyphenol oxidase: Purification, nature and properties. Proceedings of the Royal Society B **125:** 187–205

Klabunde T, Eicken C, Sacchettini JC, Krebs B (1998) Crystal structure of a plant catechol oxidase containing a dicopper center. Nature Structural and Molecular Biology **5:** 1084-1090

Koduri PKH, Gordon GS, Barker EI, Colpitts CC, Ashton NW, Suh DY (2010) Genome-wide analysis of the chalcone synthase superfamily genes of *Physcomitrella patens*. Plant Molecular Biology **72:** 247-263

Kouakou, T, Kouadio Y, Kouame P, Waffo-Teguo P, Decendit A, Merillon JM (2010) Purification and biochemical characterization of polyphenol oxidases from embryogenic and nonembryogenic cotton (*Gossypium hirsutum* L.) cells. Applied Biochemistry and Biotechnology **158:** 285-301

Koussevitzky S, Ne'eman E, Peleg S, Harel E (2008) Polyphenol oxidase can cross thylakoids by both the Tat and the Sec-dependent pathways: a putative role for two stromal processing sites. Physiologia Plantarum **133:** 266-277

Koussevitzky S, Ne'eman E, Sommer A, Steffens JC, Harel E (1998) Purification and properties of a novel chloroplast stromal peptidase. Processing of polyphenol oxidase and other imported precursors. Journal of Biological Chemistry **273:** 27064-27069

Kowalski SP, Eannetta NT, Hirzel AT, Steffens JC (1992) Purification and characterization of polyphenol oxidase from glandular trichomes of *Solanum berthaultii*. Plant Physiology **100:** 677-684

Kubowitz F (1938) Spaltung und resynthese der polyphenoloxydase und des haemocyanin. Biochemische Zeitschrift **299:** 32–57

Lang WH (1998) cDNA cloning of the *Octopus dofleini* hemocyanin: Sequence of the carboxyl-terminal domain. Biochemistry **27:** 7276-7282

Larson PR, Isebrands JG (1971) The plastochron index as applied to developmental studies of cottonwood. Canadian Journal of Forest Research **1:** 1–11

Lavid N, Schwartz A, Lewinsohn E, Tel-Or E (2001a) Phenols and phenol oxidases are involved in cadmium accumulation in the water plants *Nymphoides peltata* (Menyanthaceae) and *Nymphaeae* (Nymphaeaceae). Planta **214:** 323-331

Lavid N, Schwartz A, Yarden O, Tel-Or E (2001b) The involvement of polyphenols and peroxidase activities in heavy-metal accumulation by epidermal glands of the waterlily (Nymphaeaceae). Planta **212:** 323-331

Lax AR, Vaughn KC (1991) Colocalization of polyphenol oxidase and photosystem II proteins. Plant Physiology **96:** 26-31

Lerch K (1987) Molecular and active site structure of tyrosinase. Life Chemistry Reports **5:** 221-234

Li L, Steffens JC (2002) Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance. Planta **215:** 239-247

Lindroth RL, Hwang SY (1996) Diversity, redundancy, and multiplicity in chemical defense systems of aspen. *In* JT Romeo, JA Saunders, P Barbosa, eds, Phytochemical Diversity and Redundancy in Ecological Interactions. Plenum Press, pp 25-56

Liu L, Cao S, Xie B, Sun Z, Li X, Miao W (2007) Characterization of polyphenol oxidase from litchi pericarp using (-)-epicatechin as substrate. Journal of Agricultural and Food Chemistry **55:** 7140-7143

Mahanil S, Attajarusit J, Stout MJ, Thipyapong P (2008) Overexpression of tomato polyphenol oxidase increases resistance to common cutworm. Plant Science **174:** 456-466

Major IT, Constabel CP (2006) Molecular analysis of poplar defense against herbivory. Comparison of wound- and insect elicitor-induced gene expression. New Phytologist **172:** 617-635

Major IT, Constabel CP (2008) Functional analysis of the Kunitz trypsin inhibitor family in poplar reveals biochemical diversity and multiplicity in defense against herbivores. Plant Physiology **146:** 888-903

Marri C, Frazzoli A, Hochkoeppler A, Poggi V (2003) Purification of a polyphenol oxidase isoform from potato (*Solanum tuberosum*) tubers. Phytochemistry **63:** 745-752

Marusek CM, Trobaugh NM, Flurkey WM, Inlow JK (2006) Comparative analysis of polyphenol oxidase from plant and fungal species. Journal of Inorganic Biochemistry **100:** 108-123

Massa AN, Beecher B, Morris CF (2007) Polyphenol oxidase (PPO) in wheat and wild relatives: Molecular evidence for a multigene family. Theoretical Applied Genetics **114:** 1239–1247

Mathe C, Barre A, Jourda C, Dunand C (2010) Evolution and expression of class III peroxidases. Archives of Biochemistry and Biophysics **500:** 58-65

Matheis G, Whitaker JR (1984) Modification of proteins by polyphenol oxidase and peroxidase and their products. Journal of Food Biochemistry **8:** 137-162

Mayer AM (2006) Polyphenol oxidases in plants and fungi: Going places? A review. Phytochemistry **67:** 2318-2331

Mazzafera P, Robinson SP (2000) Characterization of polyphenol oxidase in coffee. Phytochemistry **55:** 285-296

Melo GA, Shimizu MM, Mazzafera P (2006) Polyphenoloxidase activity in coffee leaves and its role in resistance against the coffee leaf miner and coffee leaf rust. Phytochemistry **67:** 277-285

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren QH, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Balle D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meir I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan JM, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang PF, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo YG, Martinez D, Ngau WCA, Otillar B, Poliakov A, Porter A , Szajkowski L, Werner G, Zhou KM, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318:** 245-251

Ming R Hou SB, Yu QY, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen CX, Qian WB, Shen JG, Du P, Eustice M, Tong E, Tang HB, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan PZ, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang JM, Wang JP, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang XY, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei HR, Irikura B, Paidi M, Jiang N, Zhang WL, Presting G, Windsor A, Navajas-Perez R, Torres MJ, Feltus FA, Porter B, Li YJ, Burroughs AM, Luo MC, Liu L, Christopher DA Mount SM, Moore PH, Sugimura T, Jiang JM, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature **452:** 991-996

Miranda M, Ralph SG, Mellway R, White R, Heath MC, Bohlmann J, Constabel CP (2007) The transcriptional response of hybrid poplar (*Populus trichocarpa* x *P. deltoides*) to infection by *Melampsora medusae* leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. Molecular Plant-Microbe Interactions **20:** 816-831

Munoz-Munoz JL, **García-Molina F**, **Molina-Alarcón M**, **Tudela J**, **García-Cánovas F**, **Rodríguez-López JN** (2008) Kinetic characterization of the enzymatic and chemical oxidation of the catechins in green tea. Journal of Agricultural and Food Chemistry **56:** 9215-9224

Murata M, Haruta M, Murai N, Tanikawa N, Nishimura M, Homma S, Itoh Y (2000) Transgenic apple (*Malus* x *domestica*) shoot showing low browning potential. Journal of Agricultural and Food Chemistry **48:** 5243-5248

Murata M, Nishimura M, Murai N, Haruta M, Homma S, Itoh Y (2001) A transgenic apple callus showing reduced polyphenol oxidase activity and lower browning potential. Bioscience, Biotechnology and Biochemistry **65:** 383-388

Nakayama T, Sato T, Fukui Y, Yonekura-Sakakibara K, Hayashi H, Tanaka Y, Kusumi T, Nishino T (2001) Specificity analysis and mechanism of aurone synthesis catalyzed by aureusidin synthase, a polyphenol oxidase homolog responsible for flower colouration. Federation of European Biochemical Sciences Letters **499:** 107-111

Nakayama T, Yonekura-Sakakibara K, Sato T, Kikuchi S, Fukui Y, Fukuchi-Mizutani M, Ueda T, Nakao M, Tanaka Y, Kusumi T, Nishino T (2000) Aureusidin synthase: A polyphenol oxidase homolog responsible for flower colouration. Science **290:** 1163-1166

Nanjo T, Sakurai T, Totoki Y, Toyoda A, Nishiguchi M, Kado T, Igasaki T, Futamura N, Seki M, Sakaki Y, Shinozaki K, Shinohara K (2007) Functional annotation of 19,841 *Populus nigra* full-length enriched cDNA clones. BMC Genomics **8:** 448

Negri AS, Prinsi B, Rossoni M, Failla O, Scienza A, Cocucci M, Espen L (2008) Proteome changes in the skin of the grape cultivar Barbera among different stages of ripening. BMC Genomics **9:** 378

Nelson RT, Shoemaker R (2006) Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. BMC Genomics **7:** 204

Newman SM, Eannetta NT, Yu H, Prince JP, de Vicente MC, Tanksley SD, Steffens JC (1993) Organisation of the tomato polyphenol oxidase gene family. Plant Molecular Biology **21:** 1035-1051

Ono E, Hatayama M, Isono Y, Sato T, Watanabe R, Yonekura-Sakakibara K, Fukuchi-Mizutani M, Tanaka Y, Kusumi T, Nishino T, Nakayama T (2006) Localization of a flavonoid biosynthetic polyphenol oxidase in vacuoles. Plant Journal **45:** 133-143

Parveen I, Threadgill MD, Moorby JM, Winters A (2010) Oxidative phenols in forage crops containing polyphenol oxidase enzymes. Journal of Agricultural and Food Chemistry **58:** 1371-1382

Passardi F, Longet D, Penel C, Dunand C (2004) The class III peroxidase multigenic family in rice and its evolution in land plants. Phytochemistry **65:** 1879-1893

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang HB, Wang XY, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang LF, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551-556

Pinto MST, Siqueira FP, Oliveira AEA, Fernandes KVS (2008) A wouding-induced PPO from cowpea (*Vigna unguiculata*) seedlings. Phytochemistry **69:** 2297-2302

Prieto H, Utz D, Castro Á, Aguirre C, González-Agüero M, Valdés H, Cifuentes N, Defilippi BG., Zamora P, Zúniga G, Campos-Vargas R (2007) Browning in *Annona cherimola* fruit: Role of polyphenol oxidase and characterization of a coding sequence of the enzyme. Journal of Agricultural and Food Chemistry **55:** 9208-9218

Prochnik SE, **Umen J**, **Nedelcu AM**, **Hallmann A**, **Miller SM**, **Nishii I**, **Ferris P**, **Kuo A**, **Mitros T**, **Fritz-Laylin LK**, **Hellsten U**, **Chapman J**, **Simakov O**, **Rensing SA**, **Terry A**, **Pangilinan J**, **Kapitonov V**, **Jurka J**, **Salamov A**, **Shapiro H**, **Schmutz J**, **Grimwood J**, **Lindquist E**, **Lucas S**, **Grigoriev IV**, **Schmitt R**, **Kirk D**, **Rokhsar DS** (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science **329:** 223-226

Qiu YL, **Li LB**, **Wang B**, **Chen ZD**, **Knoop V**, **Groth-Malonek M**, **Dombrovska O**, **Lee J**, **Kent L**, **Rest J**, **Estabrook GF**, **Hendry TA**, **Taylor DW**, **Testa CM**, **Ambros M**, **Crandall-Stotler B**, **Duff RJ**, **Stech M**, **Frey W**, **Quandt D**, **Davis CC** (2006) The deepest divergences in land plants inferred from phylogenetic analysis. Proceedings of the National Academy of Sciences of the United States of America 103: 15511-15516

Raj SN, Sarosh BR, Shetty HS (2006) Induction and accumulation of polyphenol oxidase activities as implicated in development of resistance against pearl millet downy mildew disease. Functional Plant Biology **33:** 563-571

Ralph S, Oddy C, Cooper D, Yueh H, Jancsik S, Kolosova N, Philippe RN, Aeschliman D, White R, Huber D, Ritland CE, Benoit F, Rigby T, Nantel A, Butterfield YS, Kirkpatrick R, Chun E, Liu J, Palmquist D, Wynhoven B, Stott J, Yang G, Barber S, Holt RA, Siddiqui A, Jones SJM, Marra MA, Ellis BE, Douglas CJ, Ritland K, Bohlmann J (2006) Genomics of hybrid poplar (*Populus trichocarpa*× *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): Normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. Molecular Ecology **15:** 1275-1297

Rausher MD (2006) The evolution of flavonoids and their genes. *In* E Grotewold, ed, The Science of Flavonoids. Springer, pp 175-211

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang LX, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals the evolutionary insights into the conquest of land by plants. Science **319:** 64-69

Richter H, Lieberei R, von Schwartzenberg K (2005) Identification and characterisation of a bryophyte polyphenol oxidase encoding gene from *Physcomitrella patens*. Plant Biology **7:** 283-291

Rinaldi C, Kohler A, Frey P, Duchaussoy F, Ningre N, Couloux A, Wincker P, Le Thiec D, Fluch S, Martin F, Duplessis S (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. Plant Physiology **144:** 347-366

Roberts, MF (1971) Polyphenolases in the 1000*g* fraction of *Papaver somniferum* latex. Phytochemistry **10:** 3021-3027

Robinson C, Mant A (1997) Targeting of proteins into and across the thylakoid membrane. Trends in Plant Science **2:** 431-437

Robinson SP, Dry IB (1992) Broad bean leaf polyphenol oxidase is a 60-kilodalton protein susceptible to proteolytic cleavage. Plant Physiology **99:** 317-323

Rogozin IB, Milanesi L (1997) Analysis of donor splice sites in different eukaryotic organisms. Journal of Molecular Evolution **45:** 50-59

Schmidt DD, Voelckel C, Hartl M, Schmidt S, Baldwin IT (2005) Specificity in ecological interactions. Attack from the same lepidopteran herbivore results in species-specific transcriptional responses in two Solanaceous host plants. Plant Physiology **138:** 1763-1773

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu YS, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature **463:** 178-183

Schnable, PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du FY, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen WZ, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He RF, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin JK, Dujmic Z, Kim W, Talag J, Zuccolo A , Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren LY, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han YJ, Lee H, Li PH, Lisch DR, Liu SZ, Liu ZJ, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang LX, Yu Y, Zhang LF, Zhou SG, Zhu Q, Bennetzen JL, Dawe RK, Jiang JM, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: Complexity, Diversity and Dynamics. Science **326:** 1112-1115

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America **95:** 5857-5864

Scott A, Wyatt S, Tsou PL, Robertson D, Allen NS (1999) Model System for Plant Cell Biology: GFP Imaging in Living Onion Epidermal Cells. Biotechniques **26:** 1125-1132

Selinheimo E, NiEidhin D, Steffensen C, Nielsen J, Lomascolo A, Halaouli S, Record E, O' Beirne D, Buchert J, Kruus K (2007) Comparison of the characteristics of fungal and plant tyrosinases. Journal of Biotechnology **130:** 471-480

Shahar T, Hennig N, Gutfinger T, Hareven D, Lifschitz E (1992) The tomato 66.3-kD polyphenoloxidase gene: Molecular identification and developmental expression. Plant Cell **4:** 135-147

Sherman TD, Vaughn KC, Duke SO (1991) A limited survey of the phylogenetic distribution of polyphenol oxidase. Phytochemistry **30:** 2499–2506

Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R (2006) Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Research **34:** 3955-3967

Shin R, Froderman T, Flurkey WH (1997) Isolation and characterization of a mung bean leaf polyphenol oxidase. Phytochemistry **45:** 15-21

Sommer A, Ne'eman E, Steffens JC, Mayer AM, Harel E (1994) Import, targeting and processing of a plant polyphenol oxidase. Plant Physiology **105:** 1301-1311

Souza CD, Barbazuk B, Ralph SG, Bohlmann J, Hamberger B, Douglas CJ (2008) Genome-wide analysis of a land plant-specific acyl:coenzymeA synthetase (ACS) gene family in *Arabidopsis*, poplar, rice and *Physcomitrella*. New Phytologist **179:** 987-1003

Steffens JC, Harel E, Hunt MD (1994) Polyphenol oxidase *In* BE Ellis, GW Kuroki, HA Stafford, eds, Genetic Engineering of Plant Secondary Metabolism. Plenum Press, pp 275-312

Steiner U, Schliemann W, Böhm H, Strack D (1999) Tyrosinase involved in betalain biosynthesis of higher plants. Planta **208:** 114-124

Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Jonsson Lindvall J, Tandre K, Strauss SH, Sundberg B, Gustafsson P, Uhlén M, Bhalerao RP, Nilsson O, Sandberg G, Karlsson J, Lundeberg J, Jansson S (2004) A *Populus* EST resource for plant functional genomics. Proceedings of the National Academy of Science of the United States of America **101:** 13951-13956

Stewart RJ, Sawyer BJB, Bucheli CS, Robinson SP (2001) Polyphenol oxidase is induced by chilling and wounding in pineapple. Australian Journal of Plant Physiology **28:** 181-191

Strack D, Vogt T, Schliemann W (2003) Recent advances in betalain research. Phytochemistry **62:** 247-269

Sullivan ML, Hatfield RD, Thoma SL, Samac DA (2004) Cloning and characterization of red clover polyphenol oxidase cDNAs and expression of active protein in *Escherichia coli* and transgenic alfalfa. Plant Physiology **136:** 3234-3244

Sun DJ, He ZH, Xia XC, Zhang LP, Morris CF, Appels R, Ma WJ, Wang H (2005) A novel STS marker for polyphenol oxidase activity in bread wheat. Molecular Breeding **16:** 209–218

Taketa S, Matsuki K, Amano S, Saisho D, Himi E, Shitsukawa N, Yuo T, Noda K, Takeda K (2010) Duplicate polyphenol oxidase genes on barley chromosome 2H and their functional differentiation in the phenol reaction of spikes and grains. Journal of Experimental Botany **61:** 3983-3993

Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24:** 1596-1599

Thipyapong P, Hunt MD, Steffens JC (1995) Systemic wound induction of potato (*Solanum tuberosum*) polyphenol oxidase. Phytochemistry **40:** 673-676

Thipyapong P, Hunt MD, Steffens JC (2004) Antisense downregulation of polyphenol oxidase results in enhanced disease susceptibility. Planta **220:** 105-117

Thipyapong P, Joel DM, Steffens JC (1997) Differential expression and turnover of the tomato polyphenol oxidase gene family during vegetative and reproductive development. Plant Physiology **113:** 707-718

Thipyapong P, Steffens JC (1997) Tomato polyphenol oxidase. Differential response of the polyphenol oxidase F promoter to injuries and wound signals. Plant Physiology **115:** 409-418

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research **22:** 4673-4680

Thygesen PW, Dry IB, Robinson SP (1995) Polyphenol oxidase in potato. A multigene family that exhibits differential expression patterns. Plant Physiology **109:** 525-531

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytologist **172:** 47-62

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604

Umezawa T, Sakurai T, Totoki Y, Toyoda A, Seki M, Ishiwata A, Akiyama K, Kurotani A, Yoshida T, Mochida K, Kasuga M, Todaka D, Maruyama K, Nakashima K, Enju A, Mizukado S, Ahmed S, Yoshiwara K, Harada K, Tsubokura Y, Hayashi M, Sato S, Anai T, Ishimoto M, Funatsuki H, Teraishi M, Osaki M, Shinano T, Akashi R, Sakaki Y, Yamaguchi-Shinozaki K, Shinozaki K (2008) Sequencing and analysis of approximately 40 000 soybean cDNA clones from a full-length-enriched cDNA library. DNA Research **15:** 333-346

Vámos-Vigyázó L (1981) Polyphenol oxidase and peroxidase in fruits and vegetables. Critical Reviews in Food Science and Nutrition **15:** 49-127

Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S (2002) Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell **14:** 1441-1456

van Gelder CWG, Flurkey WH, Wichers HJ (1997) Sequence and structural features of plant and fungal tyrosinases. Phytochemistry **45:** 1309-1323

Vaughn KC, Duke SO (1981) Tentoxin-induced loss of plastidic polyphenol oxidase. Physiologia Plantarum **53:** 421-428

Vaughn KC, Duke SO (1984) Tentoxin stops the processing of polyphenol oxidase into an active protein. Physiologia Plantarum **60:** 257-261

Virador VM, Reyes Grajeda JP, Blanco-Labra A, Mendiola-Olaya E, Smith GM, Moreno A, Whitaker JR (2010) Cloning, sequencing, purification and crystal structure of Grenache (*Vitis vinifera*) polyphenol oxidase. Journal of Agricultural Food Chemistry **58:** 1189-1201

Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H, Grimwood J, McKenzie N, Huo NX, Gu YQ, Lazo GR, Anderson OD, You FM, Luo MC, Dvorak J, Wright J, Febrer M, Idziak D, Hasterok R, Lindquist E, Wang M, Fox SE, Priest HD, Filichkin SA (, Givan SA, Bryant DW, Chang JH, Wu HY, Wu W, Hsia AP, Schnable PS, Kalyanaraman A, Barbazuk B, Michael TP, Hazen SP, , Laudencia-Chingcuanco D, Weng YQ, Haberer G, Spannagl M, Mayer K, Rattei T, Mitros T, Lee SJ, Rose JKC, Mueller LA, York TL, Wicker T, Buchmann JP, Tanskanen J, Schulman AH, Gundlach H, de Oliveira AC, Maia LD, Belknap W, Jiang N, Lai JS, Zhu LC, Ma JX, Sun C, Pritham E, Salse J, Murat F, Abrouk M, Bruggmann R, Messing J, Fahlgren N, Sullivan CM, Carrington JC, Chapman EJ, May GD, Zhai JX, Ganssmann M, Gurazada SGR, German M, Meyers BC, Green PJ, Tyler L, Wu JJ, Thomson J, Chen S, Scheller HV, Harholt J, Ulvskov P, Kimbrel JA, Bartley LE, Cao PJ, Jung KH, Sharma MK, Vega-Sanchez M, Ronald P, Dardick CD, De Bodt S, Verelst W, Inze D, Heese M, Schnittger A, Yang XH, Kalluri UC, Tuskan GA, Hua ZH, Vierstra RD, Cui Y, Ouyang SH, Sun QX, Liu ZY, Yilmaz A, Grotewold E, Sibout R, Hematy K, Mouille G, Hofte H, Pelloux J, O'Connor D, Schnable J, Rowe S, Harmon F, Cass CL, Sedbrook JC, Byrne ME, Walsh S, Higgins J, Li PH, Brutnell T, Unver T, Budak H, Belcram H, Charles M, Chalhoub B, Baxter I (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763-768

von Heijne G, Steppuhn J, Herrmann RG (1989) Domain structure of mitochondrial and chloroplast targeting peptides. European Journal of Biochemistry **180:** 535-545

Wahler D, Schulze Gronover C, Richter C, Foucu F, Twyman RM, Moerschbacher BM, Fischer R, Muth J, Prüfer D (2009) Polyphenoloxidase silencing affects latex coagulation in *Taraxacum* species. Plant Physiology **151:** 334-346

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences of the United States of America **106:** 3853-3858

Wang J, Constabel CP (2003) Biochemical characterization of two differentially expressed polyphenol oxidases from hybrid poplar. Phytochemistry **64:** 115-121

Wang J, Constabel CP (2004a) Three polyphenol oxidases from hybrid poplar are differentially expressed during development and after wounding and elicitor treatment. Physiologia Plantarum **122:** 344-353

Wang J, Constabel CP (2004b) Polyphenol oxidase in transgenic *Populus* enhances resistance to forest tent caterpillar (*Malacosoma disstria*) herbivory and is activated in the insect gut. Planta **220:** 87-96

Weng JK, Tanurdzic M, Chapple C (2005) Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. BMC Genomics **6:** 85

Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf EV, Allan GJ, DiFazio SP, Potts BM, Fischer DG, Gehring CA, Lindroth RL, Marks JC, Hart SC, Wimp GM, Wooley SC (2006) A framework for community and ecosystem genetics: From genes to ecosystems. Nature **7:** 510-523

Wilkerson MD, Ru Y, Brendel VP (2009) Common introns within orthologous genes: Software and application to plants. Briefings in Bioinformatics **10:** 631-644

Winters A, Heywood S, Farrar K, Donnison I, Thomas A, Webb KJ (2009) Identification of an extensive gene cluster among a family of PPOs in *Trifolium pratense* L. (red clover) using a large insert BAC library. BMC Plant Biology **9:** 94

Wissemann KW, Montgomery MW (1985) Purification of d'Anjou pear (*Pyrus communis* L.) polyphenol oxidase. Plant Physiology **78:** 256-262

Wititsuwannakul D, Chareonthiphakorn N, Pace M, Wititsuwannakul R (2002) Polyphenol oxidases from latex of *Hevea brasiliensis*: Purification and characterization. Phytochemistry **61:** 115-121

Wong TC, Luh BS, Whitaker JR (1971) Isolation and characterization of polyphenol oxidase isozymes of clingstone peach. Plant Physiology **48:** 19-23

Wurdack KJ, Davis CC (2009) Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. American Journal of Botany **96:** 1551-1570

Yamamoto K, Kobayashi N, Yoshitama K, Teramoto S, Komamine A (2001) Isolation and purification of tyrosine hydroxylase from callus cultures of *Portulaca grandiflora*. Plant and Cell Physiology **42:** 969-975

Yoruk R, Marshall MR (2003) Physiochemical properties and functions of plant polyphenol oxidase: A review. Journal of Food Biochemistry **27:** 361-422

Yu H, Kowalski SP, Steffens JC (1992) Comparison of polyphenol oxidase expression in glandular trichomes of *Solanum* and *Lycopersicon* species. Plant Physiology **100:** 1885-1890

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ, Huang XB, Lin W, Ye C, Tong W, Cong LJ, Geng JN, Han YJ, Li L, Li W, Hu GQ, Huang XG, Li WJ, Li J, Liu ZW, Li L, Liu JP, Qi QH, Liu JS, Li L, Li T, Wang XG, Lu H, Wu TT, Zhu M, Ni PX, Han H, Dong W, Ren XY, Feng XL, Cui P, Li XR, Wang H, Xu X, Zhai WX, Xu Z, Zhang JS, He SJ, Zhang JG, Xu JC, Zhang KL, Zheng XW, Dong JH, Zeng WY, Tao L, Ye J, Tan J, Ren XD, Chen XW, He J, Liu DF, Tian W, Tian CG, Xia HG, Bao QY, Li G, Gao H, Cao T, Wang J, Zhao WM, Li P, Chen W, Wang XD, Zhang Y, Hu JF, Wang J, Liu S, Yang J, Zhang GY, Xiong YQ, Li ZJ, Mao L, Zhou CS, Zhu Z, Chen RS, Hao BL, Zheng WM, Chen SY, Guo W, Li GJ, Liu SQ, Tao M, Wang J, Zhu LH, Yuan LP, Yang HM (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science **296:** 79-92

Yu Y, Tang T, Qian Q, Wang Y, Yan M, Zeng D, Han B, Wu CI, Shi S, Li J (2008) Independent losses of function in a polyphenol oxidase in rice: Differentiation in grain discoloration between subspecies and the role of positive selection under domestication. Plant Cell **20:** 2946-2959

Zhang S, Lu S, Xu X, Korpelainen H, Li C (2010) Changes in antioxidant enzyme activities and isozyme profiles in leaves of male and female *Populus cathayana* infected with *Melampsora larici-populina*. Tree Physiology **30:** 116-128

Zhou Y, O'Hare TJ, Jobin-Décor M, Underhill SJR, Wills RBH, Graham MW (2003) Transcriptional regulation of a pineapple polyphenol oxidase gene and its relationship to blackheart. Plant Biotechnology Journal **1:** 463–478

Appendix

Supplemental Table 2.1 *PPO* gene models identified from BLAST analysis of land plant genomes.

Genome	Transcript Name	Location and Position	Intron(s)	ORF[a] (AA)	MW[b] (kDa)	cTP[c]	Reference[d]
Bryophytes							
P. patens							
	estExt_Genewise1.C_1210016	scaffold121: 167271-169785	1	536	60.1	N	PpaPPO1[§]
	e_gw1.3.8.1	scaffold3: 3693204-3694295	0	551*	62.4	N	PpaPPO2
	estExt_fgenesh1_pg.C_30144	scaffold3: 1801505-1803375	2	557	62.8	N	PpaPPO3
	gw1.16.4.1	scaffold16: 1780654-1781592	0	546*	64.3	N	PpaPPO4
	e_gw1.41.2.1	scaffold41: 343237-345110	1	535*	60.6	N	PpaPPO5
	gw1.83.3.1	scaffold83: 1328446-1329996	1	541*	61.5	N	PpaPPO6
	estExt_Genewise1.C_830181	scaffold83: 1345012-1347619	1	550*	61.2	N	PpaPPO7
	e_gw1.85.3.1	scaffold85: 1247599-1249239	0	546	61.9	N	PpaPPO8
	estExt_gwp_gw1.C_900109	scaffold90: 965980-968091	1	544*	61.8	N	PpaPPO9
	e_gw1.167.6.1	scaffold167: 513054-514871	1	559	62.7	N	PpaPPO10
	e_gw1.455.5.1	scaffold455: 71482-72630	1	534*	60.3	N	PpaPPO11
	estExt_fgenesh1_pg.C_4910010	scaffold491: 122370-124362	1	537	60.7	Y[α]	PpaPPO12
	fgenesh1_pg.scaffold_559000003	scaffold559: 43493-45125	1	514	58.3	N	PpaPPO13
Lycophytes							
S. moellendorffii							
	estExt_Genewise1Plus.C_10273	scaffold1: 1316498-1318475	1	580	64.1	N	SmoPPO1
	estExt_Genewise1Plus.C_120580	scaffold12: 2055719-2057711	1	578	67.9	N	SmoPPO2
	e_gw1.15.74.1	scaffold15: 883190-885980	2	558*	63.4	Y	SmoPPO3
	e_gw1.24.113.1	scaffold24: 1545570-1547393	1	589*	66.3	N	SmoPPO4
	estExt_Genewise1Plus.C_320067	scaffold32: 269009-271034	2	586	64.9	Y	SmoPPO5
	estExt_Genewise1Plus.C_320071	scaffold32: 275210-277243	2	586	65.1	Y	SmoPPO6
	estExt_Genewise1Plus.C_450130	scaffold45: 381546-383570	2	586	65.1	Y	SmoPPO7
	e_gw1.53.125.1	scaffold53: 349756-351594	1	596	66.4	Y	SmoPPO8
	e_gw1.112.34.1	scaffold112: 487019-488752	1	559*	62.7	N	SmoPPO9
	fgenesh2_pg.C_scaffold_112000076	scaffold112: 489352-491088	1	560*	63.2	N	SmoPPO10
	gw1.112.43.1	scaffold112: 508254-509647	1	564*	64.1	N	SmoPPO11

Supplemental Table 2.1 (continued) *PPO* gene models identified from BLAST analysis of land plant genomes.

Genome	Transcript Name	Location and Position	Intron(s)	ORF[a] (AA)	MW[b] (kDa)	cTP[c]	Reference[d]
Monocotyledonous Anthophytes							
B. distachyon							
	Bradi2g52090.1	Bd2: 51593048-51595095	1	577	64.3	Y[α]	BdaPPO1
	Bradi2g52210.1	Bd2: 51659222-51661501	1	592	65.0	Y	BdaPPO2
	Bradi2g52260.1	Bd2: 51683195-51685325	1	578	64.4	N	BdaPPO3
	Bradi3g14990.1	Bd3: 13289596-13291487	0	596	65.2	Y	BdaPPO4
	Bradi3g15040.1	Bd3: 13321324-13323373	1	579	64.3	Y	BdaPPO5
	Bradi5g22300.1	Bd5: 24716588-24718430	2	562	61.7	Y	BdaPPO6
O. sativa							
	LOC_Os04g53300.1	chr4: 31563995-31567166	2	570	62.7	Y	OsaPPO1[§]
	LOC_Os01g58100.1	chr1: 33600286-33604216	1	575	64.8	N	OsaPPO2
S. bicolor							
	Sb03g035850.1	chr3: 63865741-63867982	1	591	64.9	N	SbiPPO1
	Sb03g036870.1	chr3: 64925652-64927888	1	598	64.3	Y	SbiPPO2
	Sb06g025570.1	chr6: 54550106-54552846	2	566	62.3	Y	SbiPPO3
	Sb06g025580.1	chr6: 54566588-54568579	2	553	61.1	N	SbiPPO4
	Sb07g005410.1	chr7: 7535432-7537979	1	616	67.7	Y	SbiPPO5
	Sb07g005420.1	chr7: 7547514-7549843	0	659	72.8	Y	SbiPPO6
	Sb07g005430.1	chr7: 7583505-7585517	0	623	67.6	Y	SbiPPO7
	Sb10g024220.1	chr10: 53190460-53192373	0	578	63.2	N	SbiPPO8
Z. mays							
	GRMZM2G108103_T01	9: 97908078-97910247	1	569	63.0	N	ZmaPPO1[§]
	GRMZM2G121605_T01	10: 142617301-142619768	2	629	68.3	Y	ZmaPPO2[§]
	AC233851.1_FGT017	3: 187787528-187789449	1	597	65.9	Y	ZmaPPO3
	GRMZM2G319062_T01	4: 27917059-27919466	1	645	71.0	Y	ZmaPPO4
	GRMZM2G137909_T01	10: 43742633-43744950	1	527	57.9	Y	ZmaPPO5
	AC209206.3_FGT014	10: 137090376-137092756	2	546	60.1	Y	ZmaPPO6

Supplemental Table 2.1 (continued) *PPO* gene models identified from BLAST analysis of land plant genomes.

Genome	Transcript Name	Location and Position	Intron(s)	ORF[a] (AA)	MW[b] (kDa)	cTP[c]	Reference[d]
Eudicotyledonous Anthophytes							
C. papaya							
	evm.model.supercontig_27.43	supercontig27: 439746-441593	0	615	69.8	Y	CpaPPO1
	evm.model.supercontig_27.78	supercontig27: 867525-869333	0	602	68.4	Y	CpaPPO2
	evm.model.supercontig_530.1	supercontig530: 222-2015	0	597	68.4	Y	CpaPPO3
	evm.TU.contig_40494.1	contig40494: 543-1893	0	451*	51.6	N	CpaPPO4
C. sativus							
	cucsa.126490.1	scaffold01017: 399649-403116	1	577*	64.6	N	CsaPPO1
G. max							
	Glyma06g42170.1	Gm06: 45393709-45395668	0	596	67.2	Y	GmaPPO1
	Glyma07g31270.1	Gm07: 36264785-36266793	0	605	69.2	Y	GmaPPO2
	Glyma07g31280.1	Gm07: 36277825-36279786	0	578	65.5	Y	GmaPPO3
	Glyma07g31300.1	Gm07: 36308160-36309591	2	399	45.9	Y	GmaPPO4
	Glyma07g31310.1	Gm07: 36320692-36322806	0	626	71.4	Y	GmaPPO5
	Glyma13g25150.1	Gm13: 28423043-28424626	0	574*	64.6	Y	GmaPPO6
	Glyma13g25180.1	Gm13: 28448239-28450023	1	657*	74.0	N[•]	GmaPPO7
	Glyma13g25260.1	Gm13: 28493023-28494849	0	608	68.2	Y	GmaPPO8
	Glyma13g31590.1	Gm13: 33991444-33993249	0	601	67.7	Y	GmaPPO9
	Glyma15g07710.1	Gm15: 5435993-5437974	0	602	67.8	Y	GmaPPO10
	Glyma18g45900.1	Gm18: 55592071-55594259	0	584	66.8	Y[α]	GmaPPO11
M. esculenta							
	cassava41592.m1	scaffold03545: 5543-7192	0	587*	65.7	Y	MesPPO1
M. truncatula							
	Medtr2g014720.1	Mtchr2: 3850473-3852510	0	607	68.6	Y	MtrPPO1
	Medtr2g014730.1	Mtchr2: 3856838-3858664	0	608	68.6	Y	MtrPPO2
	Medtr4g042530.1	Mtchr4: 10195707-10199929	0	460	52.8	Q	MtrPPO3
	Medtr4g042630.1	Mtchr4: 10231884-10233396	1	479	54.5	Y	MtrPPO4

Supplemental Table 2.1 (continued) *PPO* gene models identified from BLAST analysis of land plant genomes.

Genome	Transcript Name	Location and Position	Intron(s)	ORF[a] (AA)	MW[b] (kDa)	cTP[c] (AA)	Reference[d]
Eudicotyledonous Anthophytes (continued)							
M. guttatus							
	mgf014786m	scaffold4: 2926196-2928216	0	587	65.4	N•	MguPPO1
	mgf025859m	scaffold45: 197253-199004	0	583	65.1	Y	MguPPO2
	mgf026307m	scaffold45: 217425-219744	2	618	69.4	Y	MguPPO3
	mgf019188m	scaffold45: 222771-225349	1	595*	67.0	Y	MguPPO4
	mgf002040m	scaffold230: 51671-53622	0	585	65.6	Y	MguPPO5
	mgf016367m	scaffold230: 85191-87364	0	581	64.8	Y	MguPPO6
	mgf002320m	scaffold230: 220000-222122	0	598	67.0	Y	MguPPO7
	mgf002161m	scaffold230: 313696-315923	0	597	67.2	Y	MguPPO8
	mgf014760m	scaffold230: 326522-328641	0	594	65.8	Y	MguPPO9
P. trichocarpa							
	gw1.XI.3509.1	LG_XI: 10322630-10324129	0	563	64.1	Y	PtrPPO1[§]
	gw1.178.49.1	scaffold_178: 208866-210437	0	581	65.0	Y	PtrPPO2[§]
	eugene3.00110271	LG_XI: 2759244-2761308	0	590	66.3	Y	PtrPPO3[§]
	eugene3.01780010	scaffold_178: 78111-80038	0	581	64.9	Y	PtrPPO4
	grail3.0182000101	scaffold_182: 69909-71874	0	581	64.6	Y	PtrPPO5
	gw1.182.19.1	scaffold_182: 115255-116727	0	533	60.0	Y	PtrPPO10
	gw1.182.27.1	scaffold_182: 128349-129884	0	573	64.8	Y	PtrPPO11
	eugene3.06370001	scaffold_637: 8299-9642	0	581	65.0	Y	PtrPPO12
	fgenesh4_pm.C_LG_IV000465	LG_IV: 14615253-14616755	1	587	67.5	N	PtrPPO13

Supplemental Table 2.1 (continued) *PPO* gene models identified from BLAST analysis of land plant genomes.

Genome	Transcript Name	Location and Position	Intron(s)	ORF[a] (AA)	MW[b] (kDa)	cTP[c] (AA)	Reference[d]
Eudicotyledonous Anthophytes (continued)							
P. persica							
	ppa019877m	scaffold4: 1961997-1963856	0	619	69.3	Y	PpePPO1
	ppa003190m	scaffold4: 1966978-1968762	0	594	66.0	Y	PpePPO2
	ppa003257m	scaffold4: 1953702-1955640	0	589	65.8	Y	PpePPO3
	ppa003608m	scaffold4: 1958822-1960549	1	562	62.4	Y	PpePPO4
R. communis							
	29851.m002367	29851: 55177-56946	0	589	66.2	Y	RcoPPO1
V. vinifera							
	GSVIVT00011782001	chrUnrandom: 74412583-74414406	0	607	67.4	Y	VviPPO1[§]
	GSVIVT00011776001	chrUnrandom: 74366302-74368107	1	584	65.1	Y	VviPPO2
	GSVIVT00011780001	chrUnrandom: 74392477-74394300	0	607	67.4	Y	VviPPO3
	GSVIVT00036366001	chr3: 3449151-3450767	0	538	62.2	N	VviPPO4

[a] Open reading frame (ORF) as annotated in the respective genome.
* Denotes manual ORF length annotation based on conceptual translation of available genomic DNA sequence.
[b] Preprocessed molecular weights (MW) calculated using the ExPASy proteomics server (http://ca.expasy.org/tools/pi_tool.html).
[c] cTP as predicted by ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/).
• Denotes disagreement with ChloroP 1.1 prediction based on the identification of a putative thylakoid transfer domain in the N-terminal cTP. cTP cannot be predicted for MtrPPO3 since the N-terminus annotation is incomplete.
[α] Denotes putative signal peptide predicted by TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/).
[d] Names pertaining to this paper.
[§] denotes characterized PPO homologs: OsaPPO1 (Phr1, Yu et al., 2008), PpaPPO1 (Pp_ppo1, Richter et al., 2005), PtrPPO1 (PtdPPO1, Constabel et al., 2000), PtrPPO2 (PtdPPO2, Wang and Constabel, 2004a), PtrPPO3 (PtdPPO3, Wang and Constabel, 2004a) and VviPPO1 (GPO1, Dry and Robinson, 1994). cDNA sequences for ZmaPPO1 (GenBank Accession ACG28948) and ZmaPPO2 (GenBank Accession ACG35817) have been deposited into NCBI by Alexandrov et al., 2009.

Supplemental Table 2.2 Discarded *PPO* gene models not used in this analysis.

Genome	Transcript Name	Location and Position	ORF (AA)	Comments
Bryophytes				
P. patens				
	fgenesh1_pg.scaffold_147000058	scaffold147: 625749-626220	136	no functional domains, truncated protein
	estExt_fgenesh1_pg.C_2220002	scaffold222: 7055-8088	234	no functional domains, truncated protein
Lycophytes				
S. moellendorffii				
	fgenesh1_pm.C_scaffold_7000039	scaffold7: 982575-983981	417	CuB domain incomplete, truncated protein
	fgenesh2_pg.C_scaffold_24000208	scaffold24: 1544208-1544931	211	no functional domains, truncated protein
	fgenesh2_pg.C_scaffold_32000040	scaffold32: 278040-278814	247	no functional domains, truncated protein
	fgenesh2_pg.C_scaffold_45000069	scaffold45: 384298-385016	202	no functional domains, truncated protein
Monocotyledonous Anthophytes				
O. sativa				
	LOC_Os01g58070.1	chr1: 33589592-33590617	341	no CuB domain, truncated protein
	LOC_Os04g53250.1	chr4: 31536610-31537997	199	no functional domains, truncated protein
	LOC_Os04g53260.1	chr4: 31542662-31544240	276	no CuB domain, truncated protein
	LOC_Os04g53290.1	chr4: 31555706-31556341	211	no functional domains, truncated protein
S. bicolor				
	Sb06g029040.1	chr6: 57745312-57747928	268	no functional domains, truncated protein
Eudicotyledonous Anthophytes				
C. papaya				
	evm.model.supercontig_27.77	supercontig27: 863290-864240	321	no CuB domain, truncated protein
	evm.model.supercontig_62.163	supercontig62: 999163-1000621	365	CuB domain incomplete, truncated protein
	evm.model.supercontig_62.169	supercontig62: 999163-1000621	395	CuB domain incomplete, truncated protein
C. sativus				
	cucsa.126480.1	scaffold01017: 394592-396265	172	no functional domains, truncated protein

Supplemental Table 2.2 (continued) Discarded *PPO* gene models not used in this analysis.

Genome	Transcript Name	Location and Position	ORF (AA)	Comments
Eudicotyledonous Anthophytes (continued)				
G. max				
	Glyma04g14360.1	Gm04: 14306524-14307851	357	no CuA domain, truncated protein
	Glyma07g31290.1	Gm07: 36292993-36298291	1100	suspicious gene model annotation*
	Glyma013g25170.1	Gm13: 28441074-28441464	130	no functional domains, truncated protein
	Glyma013g25190.1	Gm13: 28451796-28452112	105	no functional domains, truncated protein
	Glyma013g31570.1	Gm13: 33975600-33976774	217	no functional domains, truncated protein
	Glyma15g07700.1	Gm15: 5423442-5425678	607	CuB domain incomplete*
M. gattatus				
	mg013820m	scaffold230: 184101-185213	198	no functional domains, truncated protein
	mgf021284m	scaffold230: 149153-151939	689	suspicious gene model annotation*
P. persica				
	ppa022943m	scaffold4: 1972544-1974342	539	CuA domain incomplete*
V. vinifera				
	GSVIVT00006168001	chrUnrandom: 36468611-36483470	189	no functional domains, truncated protein
	GSVIVT00011778001	chrUnrandom: 74388714-74389397	200	no functional domains, truncated protein
	GSVIVT00011779001	chrUnrandom: 74389866-74390546	226	no CuB domain, truncated protein
	GSVIVT00012632001	chrUnrandom: 148074989-148075525	178	no functional domains, truncated protein
	GSVIVT00012737001	chrUnrandom: 151687167-151688972	584	allele of Vvi1*

* Denotes potential functional *PPO* genes, but due to annotation discrepancies the gene model was discarded from this analysis.

Supplemental Table 2.3 Intron/exon gene structures identified in *PPO* genes.

PPO Group and Location		Gene Model Name	Intron Position and Sequence		Sequence Motif
Class A'	N-terminus, PPO1_DWL	*GmaPPO4* (Glyma07g31300.1)	248-418 (171)	TG/GT-AG/CC	TGSAF▼*
			1275-1335 (61)	GG/GT-AG/AA	DSKKM▼*
Class B'	CuA, 5' Linker	*BdaPPO6* (Bradi5g22300.1)	654-697 (53)	AG/GT-AG/CT	FFPWH▼*
			762-811 (50)	AC/GT-AG/CC	PSIYN▼*
Class C'	CuA, 3' Linker	*OsaPPO1* (LOC_Os04g53300.1)	606-1086 (481)	AG/GT-AG/GA	FFPWH▼RMYLY
			1349-1457 (109)	AG/GC-AG/AT	IMYRQ▼MISGA
		SbiPPO3 (Sb06g025570.1)	579-730 (152)	AG/GT-AG/GT	FFPWH▼RLYLY
			993-1208 (216)	AG/GT-AG/AT	IMYRQ▼MISGA
		SbiPPO4 (Sb06g025580.1)	540-713 (174)	AG/GT-AG/GT	FFPWH▼RLYLY
			976-1131 (156)	AG/GT-AG/AT	IMYRQ▼MISGA
		ZmaPPO6 (AC209206.3_FGT014)	522-698 (177)	AG/GT-AG/GT	FFQWH▼RFYLY
			961-1523 (563)	AG/GT-AG/AT	IMYRQ▼MISGA
		ZmaPPO2 (GRMZM2G121605_T01)	750-877 (128)	AG/GT-AG/GC	FFPWH▼RLYVY
			1143-1240 (98)	AG/GT-AG/AT	VMYRQ▼MISGA
Class D'	5' Linker, PPO1_KFDV	*SmoPPO3* (e_gw1.15.74.1)	447-1434 (988)	TT/GT-AG/AG	CDGAF▼LETET
			2306-2482 (177)	CA/GT-AG/GA	EFGEA▼EILVI
Class E'	3' Linker, PPO1_KFDV	*SmoPPO5* (estExt_Genewise1Plus.C_320067)	957-1004 (48)	AG/GT-AG/GA	FGAPI▼RKGNN
			1423-1488 (66)	TG/GT-AG/CC	IVASL▼PVKVG
		SmoPPO6 (estExt_Genewise1Plus.C_320071)	957-1004 (48)	AG/GT-AG/GA	FGAPI▼RKGNN
			1423-1488 (66)	TG/GT-AG/CC	IVASL▼PVKVG
		SmoPPO7 (estExt_Genewise1Plus.C_450130)	957-1003 (47)	AG/GT-AG/GA	FGAPI▼RKGNN
			1422-1487 (66)	TG/GT-AG/CC	IVASL▼PVKVG

Supplemental Table 2.3 (continued) Intron/exon gene structures identified in *PPO* genes.

PPO Group and Location		Gene Model Name	Intron Position and Sequence[a]		Sequence Motif[b]
Class F'	PPO1_DWL, PPO1_KFDV	*PpaPPO3* (estExt_fgenesh1_pg.C_30144)	1083-1225 (143)	AG/GT-AG/AT	DLDKL▼RYKYD
			1310-1363 (54)	AT/GT-AG/AA	APSPS▼ITIAE
Class G'	PPO1_KFDV, C-terminus	*MguPPO3* (mgf026307m)	1718-1843 (126)	TG/GT-AG/GT	RHQGP▼GEAAS
			1914-2250 (337)	TT/GT-AG/AA	YVLCL▼LIIII
Class A	N-terminus	*PtrPPO13* (fgenesh4_pm.C_LG_IV000465)	111- 273 (163)	GG/GT-AG/GG	QHALL ▼GEFNF
Class B	CuA Domain	*PpaPPO5* (e_gw1.41.2.1)	336-1180 (845)	AG/GT-AG/GT	TFPFH▼RWYIY
Class C	5' Linker Region	*PpaPPOC* (e_gw1.455.5.1)	547-591 (45)	AA/GT-AG/AC	NIFPK▼TSPLF
Class D	3' Linker Region	*BdaPPO1* (Bradi2g52090.1)	853-973 (121)	AG/GT-AG/AT	VMYKQ▼MIGNA
		BdaPPO2 (Bradi2g52210.1)	886-989 (104)	AG/GT-AG/AT	VMYRQ▼MVSLS
		BdaPPO3 (Bradi2g52260.1)	856-958 (103)	AG/GT-AG/AT	VMYKQ▼MISNA
		BdaPPO5 (Bradi3g15040.1)	907-1017 (111)	AG/GT-AG/AT	AMYRQ▼MNVKE
		OsaPPO2 (LOC_Os01g58100.1)	865-3067 (2203)	AG/GT-AG/AT	IMYKQ▼MISSA
		SbiPPO1 (Sb03g035850.1)	904-1035 (132)	CC/GT-AG/AT	LDPAS▼MISGA
		*SbiPPO2 (*Sb03g036870.1)	874-1004 (131)	AG/GT-AG/AT	VMYRQ▼MVSLS
		SbiPPO5 (Sb07g005410.1)	925-1283 (359)	AG/GT-AG/CA	TIYTQ▼QVRKG
		ZmaPPO3 (AC233851.1_FGT017)	877-1004 (128)	AG/GT-AG/AT	VMYRQ▼MVSLS
		ZmaPPO4 (GRMZM2G319062_T01)	946-1020 (75)	AG/GT-AG/CA	TIYLQ▼QVRDG
		ZmaPPO1 (GRMZM2G108103_T01)	868-966 (99)	AG/GT-AG/AT	TMYKQ▼MVTNA
		ZmaPPO5 (GRMZM2G137909_T01)	892-996 (105)	AG/GT-AG/CA	TVYTQ▼QIRSG
Class E	3' Linker Region	*SmoPPO4* (e_gw1.24.113.1)	951-1004 (54)	AG/GT-AG/GA	FGYPL▼RKGSS
		SmoPPO8 (e_gw1.53.125.1)	933-980 (48)	AG/GT-AG/GA	FGAPV▼RKGNN
		SmoPPO9 (e_gw1.112.34.1)	864-917 (54)	AG/GT-AG/GA	FGHPI▼RKGSG
		SmoPPO10 (fgenesh2_pg.C_scaffold_112000076)	864-917 (54)	AG/GT-AG/GA	FGYPL▼RKGSS
		SmoPPO11 (gw1.112.43.1)	885-934 (50)	AG/GT-AG/GA	FGYPI▼RKGSS
Class F	3' Linker Region	*CsaPPO1* (cucsa.126490.1)	899-2485 (1587)	AG/GT-AG/GA	DEPSP▼GGGSV
Class G	PPO1_DWL	*GmaPPO7* (Glyma13g25180.1)	1352-1390 (39)	AG/GT-AG/AG	WKTIP▼ESAFF
Class H	PPO1_DWL	*PpePPO4* (ppa003608m)	1183-1221 (39)	TT/GT-AG/CT	EKKNL▼LGYVY

Supplemental 2.3 (continued) Intron/exon gene structures identified in *PPO* genes.

PPO Group and Location		Gene Model Name	Intron Position and Sequence[a]		Sequence Motif[b]
Class I	PPO1_DWL	*PpaPPO8* (gw1.83.3.1)	1089-1181 (93)	AG/GT-AG/AT	NTTLL▼RYEYE
		PpaPPO7 (estExt_Genewise1.C_830181)	939-1171 (233)	AG/GT-AG/GT	DTDLL▼RYKFE
		PpaPPO9 (estExt_gwp_gw1.C_900109)	1098-1236 (139)	AG/GT-AG/GT	DSRKL▼RYKYA
		PpaPPO13 (fgenesh1_pg.scaffold_559000003)	1008-1095 (88)	AG/GT-AG/GT	NTPLL▼RYEYE
		PpaPPO1 (estExt_Genewise1.C_1210016)	1056-1149 (94)	AG/GT-AG/AT	NAGNL▼RYGYQ
		PpaPPO10 (e_gw1.167.6.1)	1077-1214 (138)	AG/GT-AG/AT	NLANL▼RYDYD
		PpaPPO12 (estExt_fgenesh1_pg.C_4910010)	1059-1186 (128)	AG/GT-AG/GT	SLRNL▼RYGYQ
Class J	PPO1_DWL	*SmoPPO2* (estExt_Genewise1Plus.C_120580)	1222-1311 (90)	CC/GT-AG/GC	ARPST▼AVSGK
Class K	PPO1_KFDV	*SmoPPO1* (estExt_Genewise1Plus.C_10273)	1372-1515 (144)	CC/GT-AG/AA	VDSPA▼KLRIG
Class L	PPO1_KFDV	*VviPPO2* (GSVIVT00011776001)	1325-1375 (51)	CC/GT-AG/CG	SRPKP▼PNAAE
Class M	PPO1_KFDV	*MtrPPO4* (Medtr4g042630.1)	1362-1434 (73)	AG/GT-AG/AT	INDEI▼RFGLT
Class N	C-terminus	*MguPPO4* (mgf019188m)	1700-2580 (881)	GG/GT-AG/TT	VPRHE▼ (566) VPRLN

[a] Denotes the dinucleotide sequences (underlined) at the 5' and 3' termini of the intron sequence.
[b] Denotes the encoded PPO sequence motif in which the intron is inserted.
* Denotes missing sequence in the respective gene model.

Supplemental Table 3.1 Primers used in this study for the analysis of the poplar *PPO* gene family.

***PPO* Gene**	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**	**Expected Product Size (bp)**
PtrPPO1	ctactttcagcctcaccg	ggtgcagcaaaaccaaca	866
PtrPPO2	ccccacaccaaaccaaacta	ccaagtcctggaatggga	967
PtrPPO3	ccatggcaaaggtgttcag	tgatcttgattggagtgg	849
PtPPO4/5	acccttgtggtgactttc	gaatactacaatggtttcttag	253
PtrPPO9	cacgatctgtattgtctttc	ccaagccctggaatgggt	998
PtrPO10	gttaaagatactcttgacacc	gattgatcatgatgatgagcagg	417
PtrPPO11	ctgaaaaaacacgctccatttg	cgacgcagcatttcttttctac	1092
PtrPPO12	actacttcgaaagaatcttgg	ttgtcgctgactgatcatgg	1145
PtrPPO13	ctccggggaaggatcaacag	cagtgttgtgaggggctag	780
actin*	gtgcttctaagtccgaacagtgc	gactaccaaagtgtctgaccacca	
PtrPPO11N168-GFP	gcggccgcgatggcctataacctttct	gcggccgcggacatggcttgctgtg	505
PtrPPO13N59-GFP	gcggccgcaatggaagcaaagaaattggg	gcggccgcgctcctgttgatcctt	177
PtrPPO13N40-GFP	gcggccgcaatggaagcaaagaaattggg	gcggccgcgattgaattcccccaaaagcgca	120
PtrPPO13N35-GFP	gcggccgcaatggaagcaaagaaattggg	gcggccgcgaagcgcatgctggacttccgg	105
PtrPPO13NΔ19-GFP	gcggccgcaaatgactcggattcttcatat	gcggccgcgctcctgttgatcctt	120

*Actin primer sequences are as listed in Supplemental Table S2 from **Tsai et al.** (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytologist **172:** 47-62

Underlined sequence gcggccgc represents NotI restriction site.

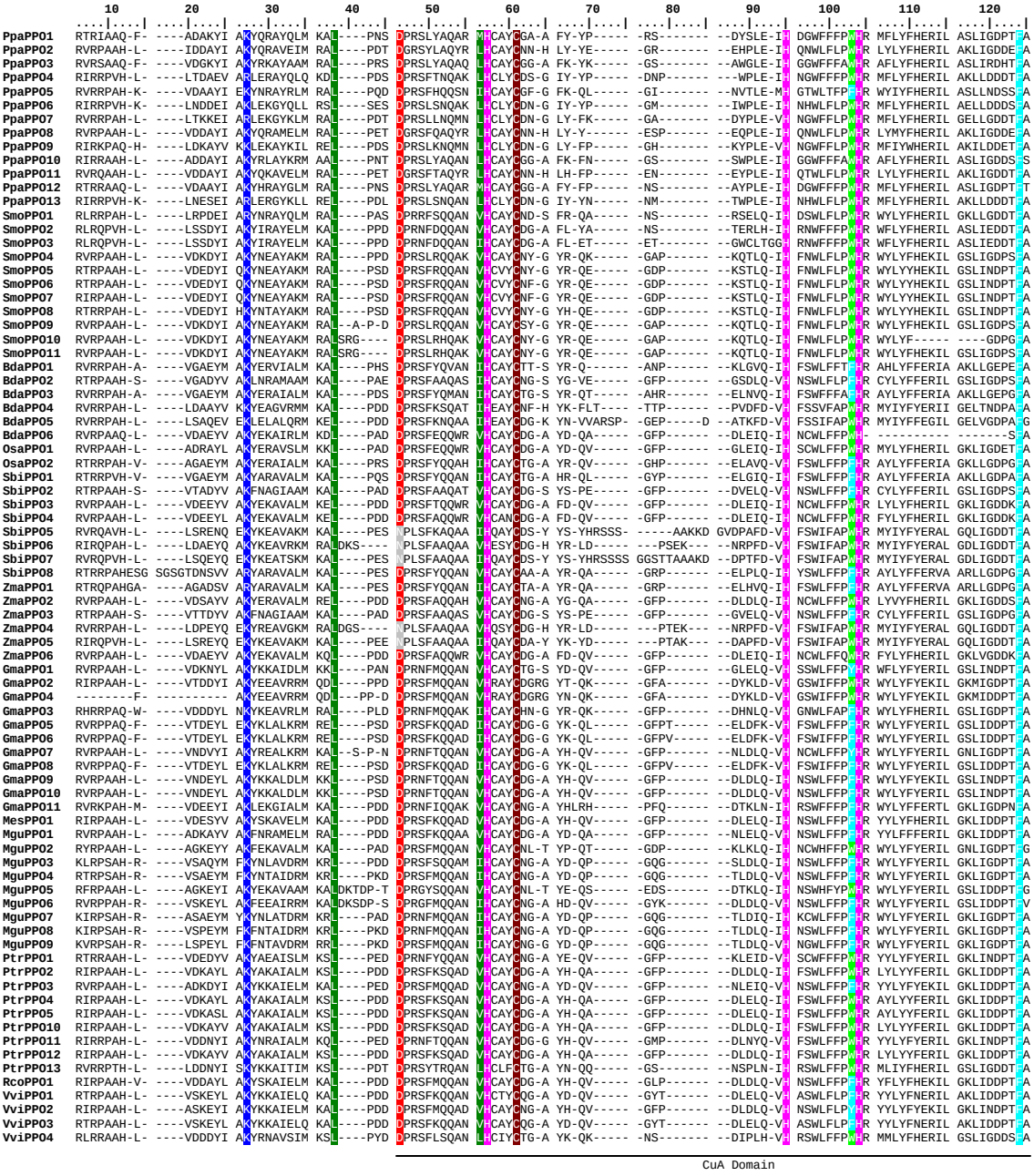
Supplemental Table 3.2 Genome-wide identification of *PPO* Genes from the poplar (*P. trichocarpa*) genome, version 1.1

PPO Gene	Protein ID	Gene Model	Location and Position	Sequence Length (bp)
PtrPPO13	546083*	fgenesh4_pm.C_LG_IV000465	LG_IV:14615253-14616755	1764
	555541	eugene3.00040254	LG_IV:2337488-2338580	745
PtrPPO1	235529	gw1.XI.3509.1	LG_XI:10322630-10324129	1692
PtrPPO8	235527	gw1.XI.3507.1	LG_XI:10308145-10310490	1116
	568258	eugene3.00110272	LG_XI:2765536-2766488	953
PtrPPO3	568257	eugene3.00110271	LG_XI:2759038-2761535	1773
PtrPPO2	275859	gw1.178.49.1	scaffold_178:208709-210609	1764
	275857	gw1.178.47.1	scaffold_178:187389-188467	841
PtrPPO9	275848	gw1.178.38.1	scaffold_178:175580-177482	1701
	585480	eugene3.01780015	scaffold_178:117425-118080	1043[α]
	585479	eugene3.01780014	scaffold_178:116692-116913	222[α]
	786653	fgenesh4_pg.C_scaffold_178000012	scaffold_178:109889-110966	891
PtrPPO4	585475	eugene3.01780010	scaffold_178:77919-80250	1746
PtrPPO7	275813	gw1.178.3.1	scaffold_178:26685-29864	1530[α]
PtrPPO11	276236	gw1.182.27.1	scaffold_182:128196-130052	1722
PtrPPO10	276228	gw1.182.19.1	scaffold_182:115108-116889	1737
PtrPPO5	674097	grail3.0182000101	scaffold_182:69713-72090	1746
	585742	eugene3.01820005	scaffold_182:66413-67304	738[α]
PtrPPO12	594316	eugene3.06370001	scaffold_637:8165-9789	1746
PtrPPO6	790794	fgenesh4_pg.C_scaffold_1529000001	scaffold_1529:15-516	414
	585390	eugene3.17730001	scaffold_1773:606-1145	447
	296394	gw1.8119.4.1	scaffold_8119:311-1899	1056
	794816	fgenesh4_pg.C_scaffold_14069000001	scaffold_14069:642-1328	744

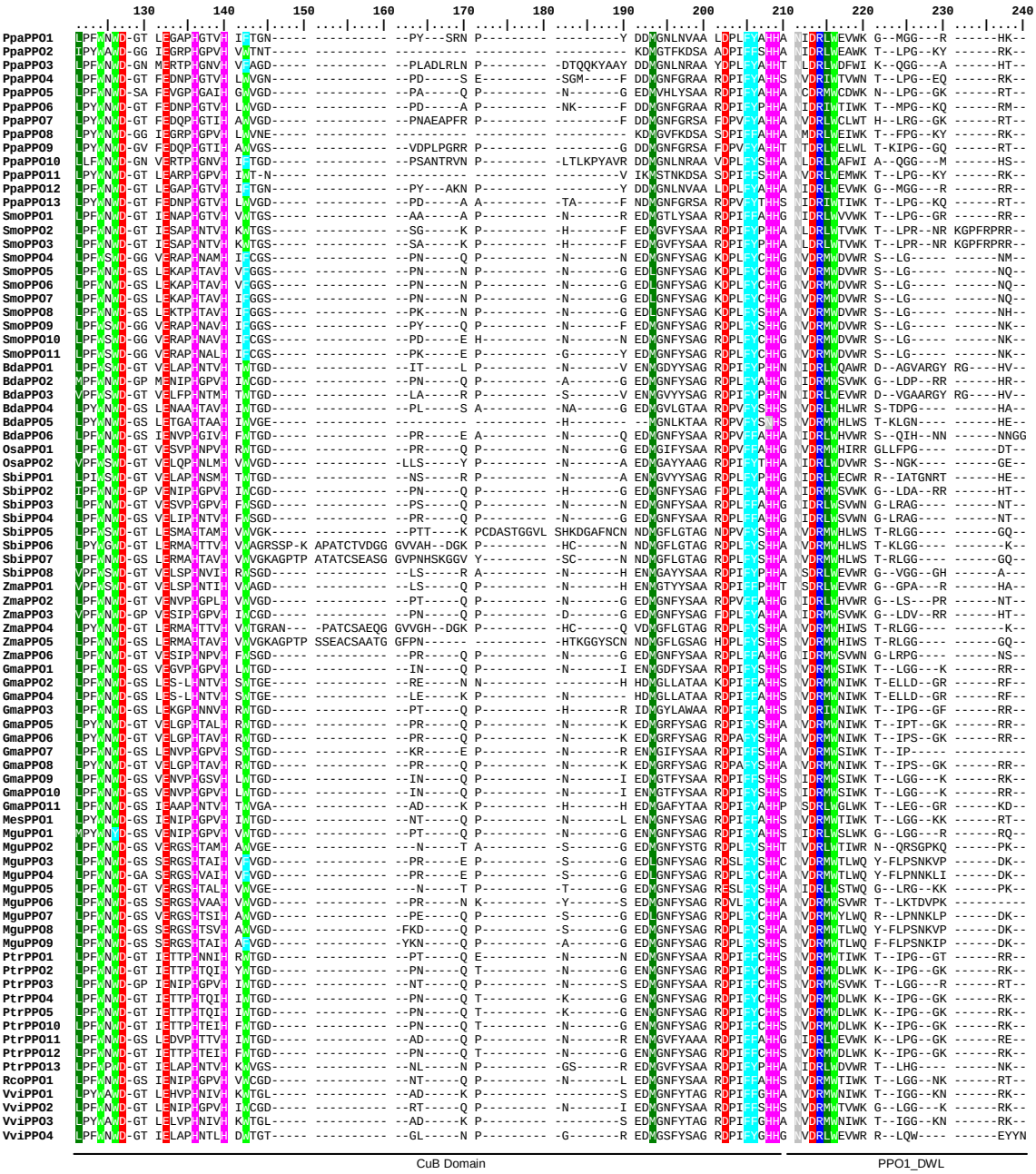
PtrPPO1-PtrPPO8 were previously annotated in the genome. For this analysis, only putative full-length *PPO* genes were assigned names (*PtrPPO9-PtrPPO13*).

[α] Denotes nucleotide sequences that appear to be incomplete as a series of Ns were present in the annotation.

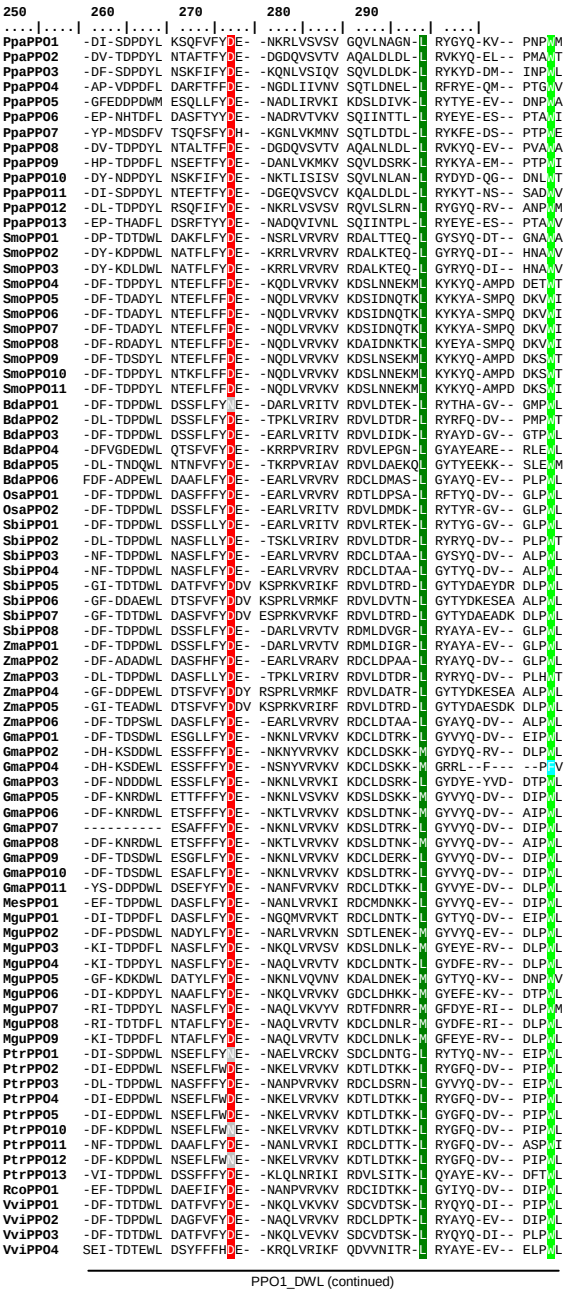
* Protein ID and gene model name corresponds to genome version 1.0. In version 1.1, the Protein ID is 800778.



Supplemental Figure 2.1 Muscle alignment for neighbour-joining phylogenetic reconstruction of PPOs from land plants.



Supplemental Figure 2.1 (continued) Muscle alignment for neighbour-joining phylogenetic reconstruction of PPOs from land plants.



Supplemental Figure 2.1 (continued) Muscle alignment for neighbour-joining phylogenetic reconstruction of PPOs from land plants.