

Supplementary File S2

Comparative genomics, Gene family analysis details

Redwood gene families (unique): A total of 460 gene families were present in coast redwood but absent in all other species compared. Among these, the majority were associated with resistance genes, protein kinases, Cytochrome P540, and genes that relate to pseudogene proliferation, such as TKP5. Overall, aspects of the Glycosylphosphatidylinositol (GPI)-anchor biosynthesis pathway were enriched as indicated by KEGG. Gene Ontology (GO) enrichment analysis highlighted xanthophyll metabolic process, nitrogen compound metabolic process, cellular nitrogen compound metabolic process, cellular aromatic compound metabolic process, and heterocycle metabolic process. In terms of molecular function, mannose-ethanolamine phosphotransferase was enriched and related to the GPI pathway. Heat shock protein (HSP70) ATPase SSB1 (OG0035595) was specifically observed. This is over-expressed in response to cold transitions in yeast and *Arabidopsis* (Heinen *et al.* 2006). In addition, other families with similarities to HSPs were present, and putatively unique to coast redwood (OG0035580; OG0016779; OG0035498; OG0035580). In addition, gene families associated with abiotic and biotic stress were shown to form unique families in coast redwood. Bifunctional abietadiene synthase (OG0022263) is involved in defensive oleoresin formation in conifers in response to insect attack or other injury (Celedon and Bohlmann 2019). This is specifically involved in diterpene olefins biosynthesis. In addition, a terpene synthase (OG0026422), involved in sesquiterpene, diterpene, and monoterpene biosynthesis was present. An Ent-kaurene oxidase-like gene family (OG0035511), was identified and is involved in gibberellin biosynthesis. These diterpenoid hormones are precursors for most terpenoid products involved in defense. A

flavonoid 3'-monooxygenase-like gene family (OG0026401) was also identified. Flavonoid biosynthesis is an important branch of phenylalanine metabolism, and these compounds are regulated by the transcription profiles of several genes and manipulated by MYB transcription factors (Li 2014). This network regulates the accumulation of flavonoid components, such as flavone, flavonol, flavanols, procyanidins, and anthocyanins. When confronted with abiotic stress, the flavonoid composition in plant tissues and organs changes in response to possible cell damage, which can induce changes in the coloring of these organs (Isah 2019). Among the broad category of genes known as lipases, a lipase class 3-like protein family (OG0035575) was identified. These were recently associated with activating salicylic acid-dependent defense responses in *Arabidopsis* (Lai *et al.* 2017). Related to this, GDSL esterases/lipases (OG0035632) were identified and are broadly involved in the regulation of plant development, morphogenesis, synthesis of secondary metabolites, and general defense response (regulation of immune function).

Coast redwood gene families (unique among the Cupressaceae): When examining gene families present in coast redwood but not present in other species of Cupressaceae, we identified 1706 gene families in total. Gene families common in this space included ubiquitins, transcription factors (MYBs), Cytochrome C, PRRs, and ATPases. GO enrichment for the following processes were observed: oxoacid metabolic process, response to heat, response to metal ions, response to cadmium ions, response to bacterium, nitrogen and organonitrogen compound metabolic processes, and heterocyclic metabolic process. HSPs were even more dominant and several were identified (OG0035580; OG0035595; OG0035498; OG0034328; OG0022010; OG0019973; OG0018899; OG0017834; OG0017068; OG0016779; OG0014359;

OG0012598; OG0010685; OG0007862). Studies in yeast noted that modifications of TIM-barrel signal transduction proteins (OG0035536) impacts overall fitness in regards to increased temperature (Chan *et al.* 2017). Drought and abiotic stress response elements, including a Dehydration-responsive element-binding protein (OG0004424), were identified. This, as well as the related ethylene insensitive gene family (OG0010932), ethylene responsive transcription factor (OG0011134; OG0026433; OG0021414), and 2-methylene-furan-3-one reductase-like (OG0025450) were prevalent. The enzyme methylene-furan-3-one reductase-like was discovered in strawberry (*Fragaria x ananassa*), where it produces furaneol, one of the major aroma compounds in the fruits (Zhang *et al.* 2020). In coast redwood, it may have a role here in pathogen defense. Related to plant stress and the proliferation of TEs common in conifers, Chromo (OG0011748; CHRromatin Organization MODifier) as well as components of the FACT complex (OG0013343) and YT521 (OG0031005) were identified. These genes are associated with regulating chromatin structure and with epigenetic repression. The WAX2 gene family was also identified here and this gene is directly associated with the synthesis of the cuticle. Variations on WAX2 (OG0013236) in plants, as studied in *Arabidopsis*, leads to variable structures and drought tolerance (less water loss) (Chen *et al.* 2003). Several xyloglucan endotransglycosylation families were identified (OG0015084; OG0020157; OG0025511; OG0026451; OG0034404; OG0035623; OG0012689). These genes cleave and regulate xyloglucan polymers, an essential constituent of the primary cell wall, and participate in cell wall construction in growing tissues. In the case of mechanical stress, it may contribute to adaptive changes in morphogenesis by being recruited to alter the tissue's tensile strength, or flexibility, enabling adaptation to mechanically stressful environments (Peña *et al.* 2004). Heparanase-like genes (OG0032932) are associated with abiotic stress in several plant species and also with

structural remodeling (Ni *et al.* 2020). In terms of nutrient acquisition, a formamidase-like gene family was identified (OG0033463). It has been shown to induce proteoid roots in nutrient deprived soils (phosphorus and nitrogen) (Rath *et al.* 2010). In addition, a bark storage protein family was identified (OG0032932) and these have been associated with nitrogen storage and differentially induced by nitrogen in *Populus* (Coleman *et al.* 1993). From a biotic stress perspective, a gene family (OG0030431) associated with mediating resistance to sphinganine analog mycotoxins (SAMTs) was identified. Genes in these families are known to regulate programmed cell death (PCD) in plants.

Coast redwood gene families (unique among all conifers). When identifying gene families unique to redwood and not found in any of the other conifers - a total of 670 were noted. We once again noted an abundance of heat-shock proteins, transcription factors, resistance genes, and protein kinases. GO enrichment was noted for single-stranded telomeric DNA binding. Here, we also note a syntaxin (OG0035569) - these large and well conserved gene families in eukaryotes are key membrane associated proteins. In plants, studies have indicated their role in drought response as well as salt tolerance (Chen *et al.* 2019). A gene family associated with pachytene checkpoint 2 was also identified in coast redwood (OG0028372). This gene is very typically diverged or completely absent in most plant species. The pachytene checkpoint in animals and yeast has been found to rely on many of the components of the mitotic DNA damage checkpoint (Wijnker and Schnittger 2013). In plants, it is possible that this is very different since different mechanisms exist to arrest the cell cycle after DNA damage. Both pantothenate kinase (OG0028500) and the associated CoA biosynthesis genes were identified as

up-regulated in root tissue in response to drought in differential expression studies in pine (de María *et al.* 2020).

Expanded in redwood compared to all species: Among the 7,724 gene families expanded specifically in redwood, enriched pathways include: carotenoid biosynthesis, RNA transport, and Porphyrin and chlorophyll metabolism

We note that Gene Ontology enrichment categories included porphyrin-containing compound metabolic process, ncRNA processing, cellular response to DNA damage/DNA repair, heterocycle metabolic process, metalloproteinase activity, pigment metabolic process, and seed development. In terms of genes, heat shock proteins, transcription factors, lipases, disease resistance genes, and metal ion transporters were prevalent. Seen specifically in this set were metal-ion related genes, including MATE transporters (OG0000181; OG0000326; OG0026382) and heavy metal associated proteins (OG0000170; OG0003233; OG0003549; OG0016719; OG0018282). Detoxification proteins and Transparent Testa-12 like proteins were observed in large numbers. Detoxification 27 (OG0012979; OG0000322), Detoxification 29 (OG0013597; OG0019025; OG0000383), Detoxification 42 protein (OG0004540), Detoxification 40 (OG0003366), and Detoxification 56 protein (OG0011608) were observed in large numbers. TESTA genes are key regulators of flavonoid biosynthesis (Zhang and Shrader 2017). Also related to flavonoid biosynthesis, UDP-rhamnose-dependent rhamnosyltransferase (OG0000043) is involved in anthocyanin modification (Hsu *et al.* 2017). Giant sequoia seedlings demonstrate a shade-avoidance response when given end-of-day far-red light, including reduced anthocyanin concentration (Peer *et al.* 1999). The largest gene families in coast redwood were associated with

disease resistance, and in specific disease resistance protein RPM1-like (OG0000038). This protein specifically recognizes the AvrRpm1 type III effector avirulence protein from *Pseudomonas syringae* (Guttman and Greenberg 2001). Resistance proteins guard the plant against pathogens that contain an appropriate avirulence protein via an indirect interaction with this avirulence protein. That triggers the hypersensitive response, which restricts the pathogen growth. Acidic endochitinase (OG0000527) is another very large family uniquely expanded in coast redwood. This protein functions as a defense against chitin containing fungal pathogens. Finally, oxophytodienoate reductase (OG0000500) was identified and is associated with pathogen response.

Absent and reduced in redwood: Among the 1054 gene families thought to be absent in redwood when compared to all other species in our comparison, meristem determination as a general process was pathway enriched in terms of its absence. Among the 719 gene families missing in Cupressaceae compared to the remaining conifers, only genes belonging to the Aminoacyl-tRNA biosynthesis process were identified as missing. When examining the 884 missing in redwood compared to the other three Cupressaceae, the missing genes were enriched for Linoleic acid metabolism. Cupressaceae seed oils differ from Pinaceae by the absence of specific Δ^5 -acids with 18 carbon atoms and higher concentrations of Δ^5 -acids with 20 carbon atoms (Wolff *et al.* 1996). A total of 44 gene families were reduced in Cupressaceae when compared to all conifers. Among these, genes related to Triacylglycerol Biosynthesis were noted as reduced. Among the small set of seven genes reduced in Cupressaceae compared to all species, root phototropism protein 3 (OG0003536) was most notable. This gene family is a signal transducer of the phototropic response and photo-induced movements. Involved in the phot1

pathway under low blue light and in the phot2 pathway under higher blue light (Zhang *et al.* 2013). This relates to root and hypocotyl phototropisms.

LITERATURE CITED

- Celedon, J. M., Bohlmann, J., 2019 Oleoresin defenses in conifers: chemical diversity, terpene synthases and limitations of oleoresin defense under climate change. *New Phytol.* 224: 1444-1463.
- Chan, Y. H., Venev, S. V., Zeldovich, K. B., Matthews, C. R., 2017 Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. *Nat. Commun.* 8: 14614.
- Chen, X., Goodwin, S. M., Boroff, V. L., Liu, X., Jenks, M. A., 2003 Cloning and characterization of the WAX2 gene of *Arabidopsis* involved in cuticle membrane and wax production. *Plant Cell* 15: 1170-1185.
- Chen, X., Wan, L., Almers, W., 2019 Mechanism of syntaxin clustering near secretory granules in live cells. *Microsc. Microanal.* 25: 1236-1237.
- Coleman, G. D., Englert, J. M., Chen, T., Fuchigami, L. H., 1993 Physiological and environmental requirements for poplar (*Populus deltoides*) bark storage protein degradation. *Plant Physiol.* 102: 53-59.
- de María, N., Guevara, M. A., Perdiguero, P., Vélez, M. D., Cabezas, J. A., *et al.*, 2020 Molecular study of drought response in the Mediterranean conifer *Pinus pinaster* Ait.: Differential transcriptomic profiling reveals constitutive water deficit-independent drought tolerance mechanisms. *Ecol. Evol.* 10: 9788-9807.

- Guttman, D. S., Greenberg, J. T., 2001 Functional analysis of the type III effectors AvrRpt2 and AvrRpm1 of *Pseudomonas syringae* with the use of a single-copy genomic integration system. *Mol. Plant. Microbe. Interact.* 14: 145-155.
- Heinen, R. C., Diniz-Mendes, L., Silva, J. T., Paschoalin, V. M. F., 2006 Identification of the divergent calmodulin binding motif in yeast Ssb1/Hsp75 protein and in other HSP70 family members. *Braz. J. Med. Biol. Res.* 39: 1399-1408.
- Hsu, Y.-H., Tagami, T., Matsunaga, K., Okuyama, M., Suzuki, T., Noda, N., Suzuki, M., Shimura, H., 2017 Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in *Lobelia erinus*. *Plant J.* 89: 325-337.
- Isah, S., 2019 Anti microbial and phytochemical studies on selected Nigerian medicinal plants: (1) Anti microbial and phytochemical properties of *Phyllanthus discoideus* (Euphorbiaceae). University of Nigeria, Nsukka, Institutional Repository.
<http://repository.unn.edu.ng/xmlui/handle/123456789/7935>
- Lai, C.-P., Huang, L.-M., Chen, L.-F. O., Chan, M.-T., Shaw, J.-F., 2017 Genome-wide analysis of GDSL-type esterases/lipases in Arabidopsis. *Plant Mol. Biol.* 95: 181-197.
- Li, S., 2014 Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal. Behav.* 9: e27522.
- Ni, M., Guerrini, M., Naggi, A., Petitou, M., 2020 Efficient selective deacetylation of complex oligosaccharides using the neutral organotin catalyst $[t\text{Bu}_2\text{SnOH}(\text{Cl})]_2$. *Carbohydr. Res.* 498: 108172.
- Peer, W. A., Briggs, W. R., Langenheim, J. H., 1999 Shade-avoidance responses in two common coastal redwood forest species, *Sequoia sempervirens* (Taxodiaceae) and *Satureja*

- douglasii* (Lamiaceae), occurring in various light quality environments. *Am. J. Bot.* 86: 640-645.
- Peña, M. J., Ryden, P., Madson, M., Smith, A. C., Carpita, N. C., 2004 The galactose residues of xyloglucan are essential to maintain mechanical strength of the primary cell walls in *Arabidopsis* during growth. *Plant Physiol.* 134: 443-451.
- Rath, M., Salas, J., Parhy, B., Norton, R., Menakuru, H., Sommerhalter, M., Hatlstad, G., Kwon, J., Allan, D. L., Vance, C. P., Uhde-Stone, C., 2010 Identification of genes induced in proteoid roots of white lupin under nitrogen and phosphorus deprivation, with functional characterization of a formamidase. *Plant Soil* 334: 137-150.
- Wijnker, E., Schnittger, A., 2013 Control of the meiotic cell division program in plants. *Plant Reprod.* 26: 143-158.
- Wolff, R. L., Deluc, L. G., Marpeau, A. M., 1996 Conifer seeds: Oil content and fatty acid composition. *J. Am. Oil Chem. Soc.* 73: 765-771.
- Zhang, B., Schrader, A., 2017 TRANSPARENT TESTA GLABRA 1-dependent regulation of flavonoid biosynthesis. *Plants* 6: 65.
- Zhang, H., Zhao L., Fan, C., Wang, P., Cui, M., Liu, L., Yang, H., Wang, J., 2020 Impact of methyl salicylate on storage quality, ethylene action, and protein profiling of ‘Zaosu’ pear (*Pyrus bretschneideri*). *Sci. Hortic.* 264: 109196.
- Zhang, K.-X., Xu, H.-H., Yuan, T.-T., Zhang, L., Lu, Y.-T., 2013 Blue-light-induced PIN3 polarization for root negative phototropic response in *Arabidopsis*. *Plant J.* 76: 308-321.