2018

Large Effect Quantitative Trait Loci for Salicinoid Phenolic Glycosides in Populus: Implications for Gene Discovery

Scott A. Woolbright
Brian J. Rehill
Richard L. Lindroth

Follow this and additional works at: https://digitalcommons.usu.edu/aspen_bib

Part of the Agriculture Commons, Ecology and Evolutionary Biology Commons, Forest Sciences Commons, Genetics and Genomics Commons, and the Plant Sciences Commons

Recommended Citation

This Article is brought to you for free and open access by the Aspen Research at DigitalCommons@USU. It has been accepted for inclusion in Aspen Bibliography by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
Large effect quantitative trait loci for salicinoid phenolic glycosides in *Populus*: Implications for gene discovery

Scott A. Woolbright\(^1\) | Brian J. Rehill\(^2\) | Richard L. Lindroth\(^3\) | Stephen P. DiFazio\(^4\) | Gregory D. Martinsen\(^5\) | Matthew S. Zinkgraf\(^6\) | Gerard J. Allan\(^5\) | Paul Keim\(^7\) | Thomas G. Whitham\(^5\)

\(^1\)Department of Biology, University of Arkansas at Little Rock, Little Rock, AR, USA
\(^2\)Department of Chemistry, US Naval Academy, Annapolis, MD, USA
\(^3\)Department of Entomology, University of Wisconsin-Madison, Madison, WI, USA
\(^4\)Department of Biology, West Virginia University, Morgantown, WV, USA
\(^5\)Environmental Genetics and Genomics Laboratory (EnGGen), Department of Biological Sciences, Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ, USA
\(^6\)Department of Biology, Western Washington University, Bellingham, WA, USA
\(^7\)Department of Biological Sciences, Pathogen and Microbe Institute, Northern Arizona University, Flagstaff, AZ, USA

**Abstract**

Genomic studies have been used to identify genes underlying many important plant secondary metabolic pathways. However, genes for salicinoid phenolic glycosides (SPGs)—ecologically important compounds with significant commercial, cultural, and medicinal applications—remain largely undescribed. We used a linkage map derived from a full-sib population of hybrid cottonwoods (*Populus* spp.) to search for quantitative trait loci (QTL) for the SPGs salicortin and HCH-salicortin. SSR markers and primer sequences were used to anchor the map to the V3.0 *P. trichocarpa* genome. We discovered 21 QTL for the two traits, including a major QTL for HCH-salicortin (\(R^2 = .52\)) that collocated with a QTL for salicortin on chr12. Using the V3.0 *Populus* genome sequence, we identified 2,983 annotated genes and 1,480 genes of unknown function within our QTL intervals. We note ten candidate genes of interest, including a BAHD-type acyltransferase that has been potentially linked to *Populus* SPGs. Our results complement other recent studies in *Populus* with implications for gene discovery and the evolution of defensive chemistry in a model genus. To our knowledge, this is the first study to use a full-sib mapping population to identify QTL intervals and gene lists associated with SPGs.

**KEYWORDS**

community and ecosystem genetics, defensive chemistry genes, foundation species, *Populus*, quantitative trait loci mapping, salicinoid phenolic glycoside

1 | INTRODUCTION

Unlike primary metabolites that govern "fundamental life processes," secondary metabolites often mediate interactions between plants and their biotic and abiotic environments (Neuman, Kumar, & Imani, 2009). Many plant secondary metabolites have important ecological functions (Lindroth & St. Clair, 2013; Steppuhn et al., 2004) as well as economical (e.g., Pickett et al., 2014), medicinal (Devore, Kang, Breteler, & Grodstein, 2013), and ethnobotanical applications. To fully exploit the potential benefits of plant secondary metabolites, we need to know the genes that underlie the metabolic pathways governing their synthesis.
and regulation (Redding-Johanson et al., 2011). Likewise, a genetic perspective of secondary metabolism is important to our understanding of how plant chemistry evolves (and coevolves) to affect the ecology and evolution of diverse organisms whose survival is directly or indirectly dependent upon the availability of suitable plant hosts (Lamit et al., 2015).

The genetic bases of metabolic pathways controlling the synthesis of many important secondary metabolites, including nicotine (Steppuhn, Gase, Krock, Halitschke, & Baldwin, 2004), various flavonoids (Saito et al., 2013), and terpenes (Redding-Johanson et al., 2011), are well characterized. However, the pathways leading to the salicinoid phenolic glycosides (SPGs) remain almost entirely undescribed (Tsai, Harding, Tschaplinski, Lindroth, & Yuan, 2006; Tsai, Kayal, & Harding, 2006; Boeckler, Gershezon, & Unsicker, 2011; but see Chedgy, Kollner, & Constabel, 2015). The lack of genetic data for SPGs is conspicuous given their potential agricultural, medicinal, and ecological applications. For example, salicortin, one of the SPGs in our study, has recently been investigated for its potential application toward the treatment and prevention of obesity (Lee et al., 2013), insulin resistance (Harbilas et al., 2013), and inflammation (Kwon et al., 2014). Ecologically, salicortin and other SPGs serve a variety of functions (reviewed in Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006) that include protection from UV damage (Turtola et al., 2005; Warren, Bassman, Fellman, Mattinson, & Eigenbrode, 2003), responses to drought (Turtola et al., 2005), and defense against herbivores (e.g., Lindroth & St. Clair, 2013; Holeski et al., 2016; Tahvanainen et al., 1991). In the model tree genus *Populus*, plant chemistry has been shown to be the key bridge between host plant genetic variation and insect community organization (Bangert et al., 2006; Bernhardson et al., 2013; Martinsen et al., 1998). For example, variation in leaf salicortin and other SPGs can alter herbivory by important arthropod pests, such as gypsy moths (Couture et al., 2016; Donaldson & Lindroth, 2007), forest tent caterpillars (Hwang & Lindroth, 1997), and other herbivores (e.g., Wooley, Walker, Vernon, & Lindroth, 2008). In contrast, SPGs are sequestered by keystone, leaf-feeding beetles of the genus *Chrysomela* and used in their own defenses (Soetens, Pasteels, Daloze, & Kaisin, 1998; see also Kearsley & Whitham, 1992; Martinsen, Dribee, & Whitham, 1998; Waltz & Whitham, 1997). Foliar SPGs have also been shown to affect symbiotic microbes. For example, studies of aspen—gypsy moth interactions at the "foliar-gut interface" have shown that SPGs can influence midgut community composition (Mason, Rubert-Nason, Lindroth, & Raffa, 2015). Some midgut bacteria are able to metabolize SPGs (Mason, Lowe-Power, Rubert-Nason, Lindroth, & Raff, 2016) and could provide new models for investigating the evolution of symbioses as driven by plant defensive chemistry.

The examples above suggest that the ability to manipulate SPG gene expression could contribute to a better understanding of the evolution of plant chemical diversity and its consequences for other aspects of *Populus* biology, such as trade-offs between growth and defense (Osier & Lindroth, 2006). Such an understanding is also likely to play a role in the eventual exploitation of plant genes for increased disease resistance, wood quality, herbal flavor and fragrance products, nutraceuticals, and pharmaceuticals (Tsai, Harding, et al., 2006).

The discovery of SPG genes has been hampered, in part, by their apparent absence in model herbaceous species (Tsai, Harding, et al., 2006). However, forest trees from the model genus *Populus* express a diverse array of SPGs and other secondary metabolites of interest (Boeckler et al., 2011; Chen, Liu, Tschaplinski, & Zhao, 2009; Constabel & Lindroth, 2010; Keefover-Ring et al., 2014). *Populus* species have been studied extensively in genomic (Tuskan et al., 2006), metabolomic (Morrel et al., 2006; Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006), ecological (Boeckler et al., 2011; Casey, Stritt, Glauser, Blanchard, & Lexer, 2015; Lindroth & St. Clair, 2013), and commercial improvement (Jansson & Douglas, 2007; Taylor, 2002; Wullschleger, Jansson, & Taylor, 2002) research. The North American black cottonwood (*P. trichocarpa*) was the first forest tree to have its genome sequenced (Tuskan et al., 2006), and it and many of its congeners are valuable resources for wood fiber, carbon sequestration, and biofuels development (Taylor, 2002). Across the northern hemisphere, *Populus* species frequently act as foundation species—organisms that modulate and stabilize fundamental community and ecosystem processes (Ellison et al., 2005; Whitham et al., 2006) in riparian and other forest ecosystems. Ease of sexual and vegetative propagation, fast growth rate, short time to sexual maturity, and widespread interspecific hybridization predispose *Populus* to experimental manipulation, and numerous genetic mapping pedigrees have been created to study the genetic basis of many traits of commercial or ecological significance (e.g., DeWoody et al., 2013; Rae et al., 2008; Robinson et al., 2012; Woolbright et al., 2008).

Here, we used a previously published (Woolbright et al., 2008) backcrossed mapping population of naturally hybridizing cottonwoods (i.e., *P. fremontii* and *P. angustifolia*) to identify QTL associated with the SPGs salicortin and HCH-salicortin. While the ecological effects of HCH-salicortin have yet to be established in empirical studies using the purified compound, it differs from salicortin only in the addition of a second hydroxycyclohexane-on-oyl (HCH) functional group (Rehill, Clauss, Wieczorek, Whitham, & Lindroth, 2005). As this functional group confers toxicity to SPGs (Lindroth, Scriber, & Hsia, 1988), HCH-salicortin is likely as biologically active as, or more so than salicortin and related compounds that contain the chemical moiety.

Our study was designed with three long-term objectives in mind: First, we sought to contribute to the discovery of genes controlling an important, but largely uncharacterized metabolic pathway. Second, we aimed to complement other recent studies investigating the evolution of plant secondary chemistry in *Populus* (Casey et al., 2015; Chedgy et al., 2015; see also Bernhardson et al., 2013). Third, we looked to establish a basis for genomic approaches to community ecology that will link changes at the DNA sequence level with variation in ecologically important chemical traits that influence other species, shape community structure, and drive ecosystem processes (Schuman, Allmann, & Baldwin, 2015). Given patterns of inheritance observed in previous studies (Bailey, Woolley, Lindroth, & Whitham, 2006; Rehill et al., 2005, 2006; Stevens & Lindroth, 2005), we made the following predictions:
1. Expression of salicortin and HCH-salicortin should be correlated given potential precursor–product relationships or competition for shared substrates.

2. Only those progeny that received donor parent (P. fremontii) alleles at certain loci would express the trait because HCH-salicortin is expressed at very low levels in some P. angustifolia genotypes (Holeski, Hillstrom, Whitham, & Lindroth, 2012) and is not detectable in others, including the recurrent parent genotype in our study.

2 | MATERIALS AND METHODS

2.1 | Pedigree and linkage map

QTL analyses were performed using a previously described mapping pedigree of hybridizing cottonwoods (Woolbright et al., 2008). Briefly, we crossed two individuals (F1 genotype WSU-6 and P. angustifolia genotype #996) from a natural population from the Weber River of northern Utah (Keim et al. 1989, Martinsen et al. 1998) to create a backcross mapping population of 246 backcross progeny. Both parental genotypes were of unknown parentage and were identified using RFLP markers. Progeny were raised under greenhouse conditions at Northern Arizona University’s research greenhouse facility. Leaf samples for both genetic and chemical analyses were collected on dry ice from cuttings that were at least 2 years old and propagated in five-gallon pots. Samples were lyophilized, ground in a Wiley mill, and stored at −20°C until use.

The mapping pedigree was used to construct a framework linkage map composed of 326 AFLP markers distributed among nineteen linkage groups (Woolbright et al., 2008). Chromosome assignments and marker order were established using MapMaker 3.0 (Lander et al., 1987) with “default linkage criteria” of LOD = 8.0 and recombination fraction (rf) = 0.37. One hundred eleven microsatellite (SSR) markers conserved across Populus species were also included in the complete map allowing us to anchor our linkage groups to the nineteen Populus chromosomes (see Tuskan et al., 2006). Thirty-eight AFLP markers were dropped from the QTL analysis due to segregation distortion (Woolbright et al., 2008). In addition, four others were dropped due to unacceptable levels of missing data. Finally, the order of four markers was changed following reanalysis of marker data using the software package R/QTL (Arends, Prins, Jansen, & Broman, 2010; Broman & Sen, 2009, two on chromosome chr2 and two on chromosome chr18). New maps were generated using the switch.order function in R/QTL and the entire linkage map for all 19 chromosomes was reestimated using est.map function with an error prob = 0.03 (the estimated scoring error rate for AFLP markers).

2.2 | Chemistry data

Phenolic glycoside concentrations (salicortin and HCH-salicortin) were determined for 172 progeny as per Lindroth, Sibber, & Hisa, 1986. Briefly, 25 mg of freeze-dried leaf powder was placed in 1 ml of methanol at 0°C and sonicated for 30 min. The extract was analyzed by high-performance thin-layer chromatography using salicortin and HCH-salicortin as standards. These compounds are the major phenolic glycosides in this system (Rehill et al., 2005); standards were purified from cottonwood leaves by liquid–liquid extraction (Lindroth et al., 1996), followed by medium-pressure, “flash” chromatography (Still, Kahn, & Mitra, 1978). Correlation analysis (Spearman’s rank) and nonlinear regression (LOESS) of the two traits were performed in R (R Core Team, 2016) and graphed using ggplot2 (Wickham, 2009). Prior to QTL analyses, salicortin data were transformed using the Box–Cox method via online software from Wessa (2016) to establish a normal distribution. All leaf salicortin and HCH-salicortin phenotypic measurements are reported as percent dry weight (%dw).

2.3 | QTL analyses

Salicortin QTL were identified using the MQM (multiple qtl mapping) function of R/QTL. For missing data, the most likely marker genotype was estimated via multiple imputation using the mapaugment feature with the maximum number of augmented genotypes set at 1,024; a minimum probability of 0.30; and strategy = “impute.” For QTL mapping, we used the function set.cofactors to generate potential cofactors at 5 cm spacing across the entire linkage map. QTL were identified using the mapscan function with cofactor significance of 0.002. Experiment-wise significant thresholds (α = 0.05 and 0.10) were estimated via 1,000 permutations using the mappermute function and used to identify significant and suggestive QTL, respectively (see Lander & Kruglyak, 1995). Chromosome-wise significance thresholds (α = 0.10; van Oojen, 1999) were also used to identify suggestive QTL. We further investigated significant QTL identified in the MQM analysis by fitting them to multiple QTL models (α = 0.10) using the MIM (multiple interval mapping) function of the WinQTLCartographer software (Wang, Basten, & Zeng, 2012). MIM allowed us to validate QTL peaks from the MQM analysis, refine QTL position and effect, and provided estimates of trait variance explained by each QTL (R²).

HCH-salicortin is usually absent or at low concentration in most P. angustifolia genotypes, including our recurrent parent (genet #996), but is present in nearly all P. fremontii and F1 hybrids studied to date (Holeski et al., 2012; Rehill et al., 2005, 2006). As such, it is possible that differences among Populus species largely stem from simple (i.e., Mendelian) modes of inheritance. For example, a backcross toward P. angustifolia is predicted to result in a bimodal, “spiked” distribution where approximately one-half of the progeny (those receiving the recurrent parental allele) will fail to express the trait. Our data followed such a distribution (Figure 1) and were not suitable for traditional mapping analyses that require normal distributions. Thus, we used R/qtl’s scanone function with the “zpart” nonparametric model for bimodal traits and “upper=FALSE” (indicating “a spike” at the zero value). The experiment-wise LOD threshold for HCH-salicortin was determined by 1,000 permutations. Although HCH-salicortin data were indeed bimodal, we obtained rough estimates of QTL effect and percent trait variance explained by the major HCH-salicortin QTL using the composite interval mapping (CIM) function of the WinQTLCartographer software.
Figure 1: HCH-salicortin phenotypes in the mapping population. Histogram of HCH-salicortin measurements showing the "spike" at zero. We searched for QTL using R/qtl's "2part" model and scanone mapping function with "upper=FALSE" for bimodal data with a "spike" at the zero value.

2.4 | Candidate Genes

Using linear interpolation to align our map with the *Populus trichocarpa* genome sequence (Tuskan et al., 2006), we created lists of candidate genes near each QTL peak. We aligned QTL intervals with the genome sequence (V.3.0) using BLASTN (Altschul, Gish, Miller, Myers, & Lipman, 1990) to determine the physical locations of syntenic SSR markers. Specific criteria for accepting SSR positions were hits with a minimum length of seventeen base pairs, seventy percent (0.70) minimum identity, and agreement with expected SSR product size. Using that information, we aligned the physical positions of SSR markers with their positions in the framework map (Woolbright et al., 2008), and used regression to estimate the physical locations of each QTL interval. When QTL intervals included chromosome ends, we used either zero or total chromosome lengths (cm vs. bp) for the relevant flanking positions.

Gene lists were created using the *Populus trichocarpa* V3.0 sequence and the Biomart function on the Phytozome v12.1 database (Goodstein et al., 2012). We compiled lists from the 2.6-Mb intervals centered on each QTL peak which corresponded to the average ±1-LOD "drop" for the significant QTL (approximately 15 cM). By definition, LOD scores for suggestive QTL are relatively low and can result in LOD drops that are prohibitively large. Using the *P. trichocarpa* genome to create gene lists for a map based on *P. fremontii* (our donor parent) presented challenges that are discussed further below.

3 | RESULTS

3.1 | Salicortin

Leaf salicortin levels in the experimental backcrosses differed by 14-fold, ranging from 1.9 to 26.8 percent dry weight (% dw), with an average of 14.1% dw (σ = 4.94). After Box-Cox transformation to establish a normal distribution, multiple QTL mapping (MQM) revealed ten significant (LOD ≥ 2.67, α = 0.05) peaks for salicortin (Figure 2), and one (LOD ≥ 3.34) for HCH-salicortin (Figure 3). However, we chose to treat these as seven QTL because the +/1.3 Mb flanking sequences used to create our 2.6-Mb intervals overlapped for several peaks (see Table 1). In addition, we identified two suggestive QTL for salicortin (LOD ≥ 2.40) using MQM mapping (α = 0.10) and six using van Ooijen’s (1999) chromosome-wise method (LOD ≥ 1.8, α = 0.10). An additional six QTL for HCH-salicortin were also identified using the chromosome-wise threshold (LOD ≥ 1.8). QTL were distributed across 14 chromosomes. Data for significant peaks are summarized in Table 1. Five of the eleven significant QTL peaks were retained by the MIM model and are summarized in Table 2. For those five, R² values ranged from .03 to .13, and QTL effect varied where substitution of a *P. fremontii* allele resulted in a decrease in salicortin for three QTL and an increase for the remaining two. When both significant and suggestive QTL were included, substitution of *P. fremontii* alleles varied in effect, with nine QTL showing an increase in the trait, and nine showing a decrease. QTL effects ranged from 1.9% dw to ~2.34% dw.

Suggestive QTL are summarized in Table 3 (only the sign of the QTL effect is included). Estimates of effect and R² were not calculated for suggestive QTL given inherent difficulty with fitting large numbers of QTL to MIM models and the fact that estimates based on suggestive QTL are less likely to be accurate (see Beavis, 1994). Tables 1 and 3 also include SSR markers that are shared with the association study by Caseys et al. (2015).

3.2 | HCH-salicortin

HCH-salicortin ranged from 0.0 to 13.3% dw in the experimental backcross, with an average of 3.4% dw and a standard deviation of 3.5. Supporting our second prediction, we identified a major QTL on chromosome chr12 (LOD = 23.4) that colocated with a salicortin QTL and that explained at least 52% of the variance. In addition, we identified six suggestive QTL (LOD > 1.8, Table 3). Substitution of *P. fremontii* alleles varied in the direction of effect, with an average increase in HCH-salicortin for the major QTL on chr12. For the six suggestive QTL, three resulted in an increase in the trait, and three resulted in a decrease.

There was no simple linear relationship between salicortin and HCH-salicortin as illustrated in Figure 4. While a Spearman’s rank test showed that salicortin and HCH-salicortin were negatively correlated (rₛ = −0.2903, p < .0001), a LOESS line (nonlinear regression) fitted to the data illustrates complexity in the relationship where the two traits are positively correlated from low- to midrange values, but are negatively correlated when salicortin levels are highest. However, note that several individuals expressed both traits at high values.

3.3 | Gene lists

The total number of genes and transcripts in our 2.6 Mb QTL intervals was 4,463, representing approximately 10.8% of the total genes (41, 335) for the *Populus* V3.0 sequence (see Sundell et al., 2015).
Our lists include genes models that lack general identifiers or descriptions. When relevant, the positions of SSR markers used in the study by Caseys et al. (2015) were also included. Table S1 provides total chromosome-wise gene lists and includes locations of our QTL peaks and SSRs from Caseys et al. (2015) as red text.

We identified eight genes that are associated with secondary metabolites (primarily hydroxycinnamate precursors and derivatives) and thought to play roles in predicted steps in the SPG pathway (see Tsai et al., 2006a,b) or similar metabolic pathways (see Bonpart et al., 2015). Candidate genes are summarized on Table S2 and include two BAHD-type acyltransferases, eight cinnamoyl-coA reductases, and Shikimate dehydrogenase. Supplemental tables are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.8h67t.

### TABLE 1 Results of MQM analyses. Peaks were identified using R/qtl's MQM (salicortin) and "2-part" (HCH-salicortin) models ($\alpha = 0.05$). LOD thresholds were $\geq 2.67$ for salicortin and $\geq 3.34$ for HCH-salicortin. QTLs are named for the chromosome on which they occur and the order in which they were discovered. For example, QTL Sal_4.2 represents the second QTL described for chr04. Pos (Mb) = position in the genome. LOD is the maximum LOD score at the QTL peak. Column labeled “a” indicates the phenotypic effect (increase or decrease of the trait value) when the donor parent (P. fremontii) allele is substituted for the recurrent parent (P. angustifolia) allele at each QTL. QTL from Caseys et al. (2015) and the SPGs they correspond to were included when they occurred on the same chromosomes. Associated SSR is the marker nearest the QTL described by Caseys et al. (2015) and pos(Mb) is that marker’s position in the v3.0 genome. Periods (".") denote data on associated marker or QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>pos (cM)</th>
<th>pos (Mb)</th>
<th>LOD</th>
<th>a</th>
<th>P. tremula and P. alba QTL (Caseys et al., 2015)</th>
<th>Associated SSR</th>
<th>Pos (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sal_2.1</td>
<td>62.50</td>
<td>8.77</td>
<td>3.14</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Sal_5.1</td>
<td>22.63</td>
<td>3.77</td>
<td>4.82</td>
<td>−</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Sal_5.2</td>
<td>79.79</td>
<td>15.81</td>
<td>3.33</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Sal_5.3</td>
<td>100.60</td>
<td>20.58</td>
<td>5.41</td>
<td>+</td>
<td>Tremuloidin, tremulacin isomer</td>
<td>W15</td>
<td>25.78</td>
</tr>
<tr>
<td>Sal_7.1</td>
<td>0</td>
<td>0</td>
<td>3.04</td>
<td>−</td>
<td>Salicortin Isomer 2</td>
<td>O312</td>
<td>3.35</td>
</tr>
<tr>
<td>Sal_12.1</td>
<td>74.00</td>
<td>13.08</td>
<td>18.86</td>
<td>−</td>
<td>Acetyl-salicortin isomer 1, Isomer 2</td>
<td>G1186</td>
<td>12.95</td>
</tr>
<tr>
<td>Sal_12.2</td>
<td>77</td>
<td>14.00</td>
<td>19.89</td>
<td>−</td>
<td>Acetyl-salicortin isomer 1, Isomer 2</td>
<td>G1186</td>
<td>12.95</td>
</tr>
<tr>
<td>HCH_12.1</td>
<td>82.66</td>
<td>14.94</td>
<td>25.53</td>
<td>+</td>
<td>Acetyl-salicortin isomer 1, Isomer 2</td>
<td>G1186</td>
<td>12.95</td>
</tr>
<tr>
<td>Sal_15.1</td>
<td>46.80</td>
<td>9.32</td>
<td>6.17</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Sal_15.2</td>
<td>52.57</td>
<td>10.43</td>
<td>5.99</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Sal_15.3</td>
<td>58.49</td>
<td>11.58</td>
<td>5.33</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

### FIGURE 2 LOD scores for salicortin at all nineteen chromosomes. Horizontal lines represent the experiment-wise (green, $\alpha = 0.05$) and chromosome-wise suggestive (red, $\alpha = 0.10$) QTL thresholds.

4 | DISCUSSION

In this study, we identify several QTL associated with the SPGs salicortin and HCH-salicortin. The most noteworthy QTL occur at the distal end of chr12 and overlap for the two traits. The major QTL for HCH-salicortin on chr12 represent a likely Mendelian locus, and, as such, it could prove to be an important starting point for identifying undescribed SPG genes. Our results also complement other recent studies focusing on association mapping of metabolomic data (Caseys et al., 2015) including gene expression and function of candidate genes associated with SPG synthesis (Chedgy et al., 2015; see also Tsai et al., 2016a,b). Data from our study create foundations for verifying the results of future investigations of the genetic basis of SPG synthesis and the ultimate discovery of SPG genes.
TABLE 2  QTL retained in the MIM Model. MQM QTL peaks that were also supported by MIM model-fitting ($\alpha = .10$). Columns are as in Table 1, with the addition of a column for the $R^2$ value for each QTL as estimated by MIM mapping.

<table>
<thead>
<tr>
<th>QTL</th>
<th>pos (cM)</th>
<th>LOD</th>
<th>$a$</th>
<th>$R^2$</th>
<th>$P. tremula$ and $P. alba$ QTL (Caseys et al., 2015)</th>
<th>Associated SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sal_2.1</td>
<td>53.0</td>
<td>3.23</td>
<td>1.96</td>
<td>.09</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Sal_5.3</td>
<td>96.4</td>
<td>2.68</td>
<td>1.90</td>
<td>.06</td>
<td>Tremuloidin, tremulacin isomer</td>
<td>W15</td>
</tr>
<tr>
<td>Sal_7.1</td>
<td>0.0</td>
<td>2.19</td>
<td>1.81</td>
<td>.06</td>
<td>Salicortin Isomer 2</td>
<td>O312</td>
</tr>
<tr>
<td>Sal_12.1</td>
<td>58.9</td>
<td>6.07</td>
<td>2.35</td>
<td>.13</td>
<td>Acetyl-salicortin isomer 1, Isomer 2</td>
<td>G1186</td>
</tr>
<tr>
<td>Sal_15.1</td>
<td>46.8</td>
<td>1.32</td>
<td>1.57</td>
<td>.043</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3  Suggestive QTL from the MQM analyses and chromosome-wise thresholds. Data are arranged as per Table 1, but for suggestive rather than significant peaks.

<table>
<thead>
<tr>
<th>QTL</th>
<th>pos (cM)</th>
<th>Pos (Mb)</th>
<th>LOD</th>
<th>$a$</th>
<th>$P. tremula$ and $P. alba$ QTL (Caseys et al., 2015)</th>
<th>Associated SSR</th>
<th>Pos (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sal_1.1</td>
<td>0.00</td>
<td>0.00</td>
<td>1.90</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCH_1.1</td>
<td>10.40</td>
<td>2</td>
<td>1.87</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sal_3.1</td>
<td>69.50</td>
<td>16.14</td>
<td>2.39</td>
<td>-</td>
<td>Acetyl-salicortin, isomer 1, isomer 2; tremuloidin; tremuloidin; tremulacin isomer</td>
<td>G1869</td>
<td>15.28</td>
</tr>
<tr>
<td>Sal_4.1</td>
<td>6.84</td>
<td>1.62</td>
<td>2.45</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sal_4.2</td>
<td>44.91</td>
<td>6.54</td>
<td>2.06</td>
<td>-</td>
<td>Salicortin isomer 3</td>
<td>O127</td>
<td>6.64</td>
</tr>
<tr>
<td>Sal_5.4</td>
<td>12.02</td>
<td>2.24</td>
<td>2.40</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCH_5.1</td>
<td>16.20</td>
<td>2.84</td>
<td>2.31</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCH_5.2</td>
<td>26.56</td>
<td>4.34</td>
<td>2.65</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sal_6.1</td>
<td>112.62</td>
<td>23.10</td>
<td>1.93</td>
<td>-</td>
<td>HCH-tremulacin</td>
<td>O369</td>
<td>23.66</td>
</tr>
<tr>
<td>Sal_8.1</td>
<td>31.92</td>
<td>3.70</td>
<td>2.34</td>
<td>+</td>
<td>HCH-Salicortin</td>
<td>O374</td>
<td>6.69</td>
</tr>
<tr>
<td>HCH_11.1</td>
<td>56.00</td>
<td>10.83</td>
<td>2.44</td>
<td>+</td>
<td>Tremulacin</td>
<td>G1037</td>
<td>6.99</td>
</tr>
<tr>
<td>Sal_17.1</td>
<td>20.52</td>
<td>3.11</td>
<td>1.97</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCH_18.1</td>
<td>76.62</td>
<td>15.03</td>
<td>2.91</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCH_19.1</td>
<td>45.10</td>
<td>10.4</td>
<td>2.22</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 3  LOD scores for HCH-salicortin at all nineteen chromosomes. Horizontal lines represent the experiment-wise (green, $\alpha = 0.05$) and chromosome-wise suggestive (red, $\alpha = 0.10$) QTL thresholds.
4.1 | Identification of suggestive versus significant QTL

Choice of mapping strategy can influence the identification of QTL, especially where small populations are involved. We chose to search for suggestive QTL using relaxed LOD thresholds ($\alpha = 0.10$) for both the MQM mapping (Arends et al., 2010; Broman & Sen, 2009) and chromosome-wise threshold (van Ooijen, 1999) methods in order to increase the number of QTL identified. That choice, as well as the failure to retain all loci in the MIM model, resulted in 14 of 21 QTL being categorized as “suggestive” rather than “significant” (see Lander & Kruglyak, 1995). In the past, concern that erroneous QTL (i.e., type I errors) would lead to “bloated literature” has led to relatively conservative suggestions for setting QTL thresholds (see Lander & Kruglyak, 1994). However, given the exploratory nature of most QTL studies, Beavis (1998) noted that, in the long run, type II errors (missing QTL) might be costlier than type I errors (false QTL; see also Holland, 2005), despite his own work demonstrating the increased risk of type I errors when dealing with small populations (i.e., the so-called Beavis Effect). Likewise, Lander & Kruglyak (1994) did not completely dismiss suggestive QTL, but called for caution when reporting them. Today, when whole genomes and massively parallel sequencing make it possible to assess entire transcriptomes (e.g., RNA-seq), virtually any QTL might prove useful if it helps to validate gene expression data. In keeping with the standards set by Lander and Kruglyak (1995), we acknowledge the “tantalizing but unproven” nature of our suggestive QTL, with an eye toward ongoing and future genomic and transcriptomic research in Populus. Thus, we chose to include all intervals and their corresponding gene lists in this article.

4.2 | Trait variation and correlation between salicortin and HCH-salicortin

For overlapping QTL intervals, _P. fremontii_ alleles resulted in decreased values for both traits for the QTL on chr05. In contrast, substitution of _P. fremontii_ alleles at our largest QTL (sal 12 and HCHsal 12) increased HCH-salicortin but decreased salicortin. While this could be explained by precursor-product or competitive relationships, the lack of a linear correlation between the two traits (Figure 4) prevents us from accepting our first prediction that the two traits would be correlated in a manner explicable by our QTL data (e.g., by the conversion of salicortin to HCH-salicortin as governed by a QTL such as occurs on chr12). Instead, it appears that the relationship between the two traits is complex and understanding their relative fluxes requires a broader metabolomic perspective that was outside the scope of our study (but see Tsai et al., 2006a; Casey’s et al., 2015). To date, the SPG pathway remains poorly described, underscoring the need for a genomic basis of SPG synthesis and flux.

Variation in the sign of the QTL effects for each trait could result from species-level differences (e.g., preferred pathways, substrate specificity, allelic, or isozyme efficiency) or from similar allelic variation within the wild (outcrossed) parental genotypes. On average, salicortin levels are higher in _P. angustifolia_ than _P. fremontii_; however, they vary considerably within and among species and hybrids (Rehill et al., 2005, 2006). Despite the overlap in salicortin levels, we were able to identify QTL potentially reflecting allelic or isozyme differences between hybridizing _Populus_ species. In contrast, the fact that _P. angustifolia_ does not produce HCH-salicortin allowed us to identify six potential QTL for HCH-salicortin, including a major locus on chr12 where substitution of a _P. fremontii_ allele appears to be required for expression of the trait (thus supporting Prediction 2, above). The concentration of leaf HCH-salicortin differs significantly among _Populus_ generic sections, and our QTL data complement the only other published study that identifies QTL intervals associated with SPGs, including salicortin and HCH-salicortin (Caseys et al., 2015). Likewise, our HCH-salicortin QTL data support one other study that investigated specific candidate genes that could be directly involved with SPG synthesis (Chedgy et al., 2015; discussed below).

4.3 | Candidate genes

Our study was intended to generate preliminary data on regions of the _Populus_ genome associated with SPG synthesis. Given the length of our QTL intervals, the candidate genes discussed here are informed speculation at best, and so conclusions should be treated with caution. However, our choice to further discuss specific genes is supported by previous studies. For example, starting with the strongest QTL on chr12, we searched for candidate genes that were orthologous with other plant genes controlling similar metabolic pathways, especially those that have been tentatively linked to proposed steps within the _Populus_ SPG pathway (Chedgy et al., 2015; Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006). Our search revealed two hydroxycinnamoyl/benzoyltransferases that belong to the BAHD-type acyltransferase family of genes Potri.001G042900 and Potri.012G144500. The first (on chr01) encodes HCT6 which is
associated with lignin biosynthesis in vascular tissues (Touminen, Johnson, & Tsai, 2011; Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006; Wegryzn et al., 2010) and has not been shown to affect SPGs. In contrast, the function of Potri12G144500, which occurs within the interval for our strongest QTL (on chr12), remains undescribed. However, others have noted similar N-hydroxycinnamoyl/benzoyletransferases that have substrates or products that are also thought to play roles in SPG expression via control over the synthesis of benzyl benzoate and salicyl benzoate—possible intermediates in the SPG pathway (Chedgy et al., 2015; see also Bapst, Harding, & Tsai, 2010). Interestingly, Potri.012G144500 is one of the final fourteen transcripts identified for that end of chromosome chr12 under the current (V3.0) annotated genome sequence. While MQM mapping resulted in slight map expansion due to the introduction of pseudomarkers, CIM of the original, nonimputed maps placed QTL peaks for salicortin and HCH-salicortin at terminus of the linkage map for chromosome chr12—which includes intriguing, albeit circumstantial support for candidate gene status. These findings are important because Chedgy et al. (2015) showed that enzyme activity of similar genes (e.g., PtACT49 on chromosome chr19) is consistent with hypothesized steps in the SPG pathway (see Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006).

Several other candidate genes identified in our study are linked to pathways that might be coregulated or otherwise linked with SPG synthesis. Examples include shikimate dehydrogenase (chr05) which catalyzes the conversion of 3-dehydroshikimate to shikimate—a key entry point to the shikimic acid pathway (aka phenylpropanoid pathway). That pathway eventually leads to production of phenylalanine, which is deaminated to produce cinnamic acid—very likely an initial precursor for PGs (Tsai, Kayal, et al., 2006). Eight cinnamoyl CoA-reductase (CCR) genes were found on chr01). Cinnamic acid is thought to be one of the likely starting points of the SPG pathway (Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006). Cinnamic acid leads to production of benzoyl-CoA, which can be catalyzed by CCR to produce benzaaldehyde. Benzaldehyde can then be metabolized to salicylaldehyde which is also one of the several precursor compounds (another being salicylic acid) that could lead to production of SPGs. While these compounds and genes create potential starting points for understanding SPG metabolism, it is important to emphasize that the main pathway(s) leading to remains unresolved.

4.4 | Alignment with the P. trichocarpa genome

Creating candidate gene lists was challenging given the lack of a genome sequence for our parent species. This was particularly true for P. fremontii, our donor parent and source of the marker alleles used to create our map. While P. angustifolia is from the same generic section as P. trichocarpa (sect. Tacamahaca), P. fremontii is within a different section (sect. Aigeiros). However, previous studies have shown high levels of synteny among Populus species from different generic sections, as well as a high transferability of SSR markers (Berlin et al., 2010). Still, small-scale differences between our map and P. trichocarpa have been noted (see Woolbright et al., 2008) and could be the product of real differences among species/sections or the result of typical mapping errors—especially those associated with relatively small sample sizes. Thus, in order to choose the size of the interval used to search for candidate genes, we had to balance the need for concise gene lists with the possibility that our QTL intervals did not align precisely with the reference genome. We created gene lists from physical intervals of 2.6 Mb, which corresponded to the average ±1-LOD drop around our significant QTL peaks. In addition, two of our intervals contained SSR markers shared with the recent SPG association mapping study by Caseys et al. (2015). This supports the notion that these loci could be useful starting points for future genetic and genomic studies of SPG synthesis, especially those employing expression-based approaches such as eQTL mapping. RNA-seq and knockdown/out/up studies.

5 | CONCLUSIONS

While our results are preliminary steps toward identifying SPG genes, our study has five important conclusions. First, it is the first study of salicinoid phenolic glycosides to use a full-sib mapping population and a linkage map anchored to the Populus genome sequence to identify QTL interval for salicinoid phenolic glycosides. Second, it complements a recent association mapping study by Caseys et al. (2015) that used European Populus species to investigate the
evolution of SPG diversity, and constraints therein, for a more distantly related section of the genus. Third, we identify a major likely Mendelian locus for HCH-salicortin that provides a focal point for future research on SPG gene discovery. Fourth, our QTL intervals contain candidate genes known to be associated with compounds that are thought to play potential roles in the SPG pathway (i.e., hydroxycinnamoyl and its precursors or derivatives; Tsai, Harding, et al., 2006; Tsai, Kayal, et al., 2006) and are similar to those discussed by Chedgy et al. (2015). Finally, our results are an important first step toward linking genomic data with community-level patterns—a major goal in the developing field of community genetics and evolution (e.g., Allen et al., 2012; Whitham et al., 2006, 2008). These conclusions emphasize the role of *Populus* as a model organism for studying the chemical ecology and evolution of foundation forest trees and for the advancement of the developing field of community genetics and genomics.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the following individuals: Phil Service, Tom Sisk, Jen Schweitzer, Joe Bailey, Celine Caseys, Christian Lexer, Justin McGrath, Karl Broman, Danny Arends, Stephanie Woolbright, and Tricia Covington. Special thanks to Brad Blake and Phil Patterson (NAU Greenhouse); the students, faculty, and staff of the NAU Department of Biological Sciences; the NAU Cottonwood and Pinyon Ecology Groups; and the Postdoctoral Fellows Program and the Genomic Ecology of Global Change (GEGC) theme at the University of Illinois Carl R. Woese Institute for Genomic Biology (IGB). Funding and other support were provided by the UA Little Rock Department of Biology and College of Arts, Letters, and Sciences; the Phoenix chapter of the ARCS Foundation; NSF FIBR grant DEB-0425908, NSF MacroSystems grant DEB-1340852, and NSF DBI-1126840 for establishing the Southwest Experimental Garden Array; and the students, Christian Lexer, Justin McGrath, Karl Broman, Danny Arends, Stephanie Woolbright, and Tricia Covington. Special thanks to Brad Blake and Phil Patterson (NAU Greenhouse); the students, faculty, and staff of the NAU Department of Biological Sciences; the NAU Cottonwood and Pinyon Ecology Groups; and the Postdoctoral Fellows Program and the Genomic Ecology of Global Change (GEGC) theme at the University of Illinois Carl R. Woese Institute for Genomic Biology (IGB). Funding and other support were provided by the UA Little Rock Department of Biology and College of Arts, Letters, and Sciences; the Phoenix chapter of the ARCS Foundation; NSF FIBR grant DEB-0425908, NSF MacroSystems grant DEB-1340852, and NSF DBI-1126840 for establishing the Southwest Experimental Garden Array.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

TGW, PK, GDM, and SAW planned and designed the experiments; RLL and BJR conducted the chemical analyses and contributed to writing and editing; GDM and SAW conducted the experimental crosses and performed the QTL experiments; SPD and MSZ conducted genomic data analyses, and participated in writing and editing; GJA oversaw laboratory work and contributed to data analysis, writing, and editing; SAW wrote the manuscript.

ORCID

Scott A. Woolbright [http://orcid.org/0000-0002-7886-1009](http://orcid.org/0000-0002-7886-1009)

REFERENCES


SUPPORTING INFORMATION
Additional Supporting Information may be found online in the supporting information tab for this article.