Association Analysis Identifies *Melampsora × columbiae* Poplar Leaf Rust Resistance SNPs

Jonathan La Mantia¹, Jaroslav Klášťér J, El-Kassaby¹, Shofiul Azam¹, Robert D. Guy¹, Carl J. Douglas², Shawn D. Mansfield³, Richard Hamelin¹,5

¹Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, British Columbia, Canada, ²Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada, ³Department of Wood Science, University of British Columbia, Vancouver, British Columbia, Canada, ⁴Natural Resources Canada, Laurentian Forestry Center, Québec, Canada, ⁵Department of Dendrology and Forest Tree Breeding, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká, Czech Republic

---

Abstract

*Populus* species are currently being domesticated through intensive time- and resource-dependent programs for utilization in phytoengineering, wood and paper products, and conversion to biofuels. Poplar leaf rust disease can greatly reduce wood volume. Genetic resistance is effective in reducing economic losses but major resistance loci have been race-specific and can be readily defeated by the pathogen. Developing durable disease resistance requires the identification of non-race-specific loci. In the presented study, area under the disease progress curve was calculated from natural infection of *Melampsora × columbiae* in three consecutive years. Association analysis was performed using 412 *P. trichocarpa* clones genotyped with 29,355 SNPs covering 3,543 genes. We found 40 SNPs within 26 unique genes significantly associated (permutated \( P < 0.05 \)) with poplar rust severity. Moreover, two SNPs were repeated in all three years suggesting non-race-specificity and three additional SNPs were differentially expressed in other poplar rust interactions. These five SNPs were found in genes that have orthologs in Arabidopsis with functionality in pathogen induced transcriptome reprogramming, Ca\(^{2+}\)/calmodulin and salicylic acid signaling, and tolerance to reactive oxygen species. The additive effect of non-R gene functional variants may constitute high levels of durable poplar leaf rust resistance. Therefore, these findings are of significance for speeding the genetic improvement of this long-lived, economically important organism.

---

Introduction

Rust fungi cause some of the most important crop and tree diseases worldwide. In *Populus* species (poplar trees), leaf rust disease is caused by several species of *Melampsora*. Severe poplar leaf rust infections decrease photosynthetic capacity, reduce biomass, and increase susceptibility to additional pathogens [1]. Reductions in dry weight and fiber volume of *P. deltoides* × *P. balsamifera* ‘Northwest’, have been estimated at 57% and 65%, respectively [2]. In North America, hybridization of rust species, *M. occidentalis* and *M. medusae*, has produced a new rust pathogen, *M. × columbiae*. This hybrid rust has demonstrated high pathogenic diversity. In an initial collection 13 pathotypes (race with unique virulence to specific hosts) were identified. Furthermore, host resistance loci to pathotypes *Mxe3* and *Mxe4* map to unique chromosomes in poplar and are race-specific [3,4].

Rust virulence and poplar resistance interact in a classical gene-for-gene model, where host R gene recognition of a pathogen avirulence gene is necessary for resistance [5]. The pathogenic diversity of *Melampsora* is facilitated by obligate sexual reproduction on alternative hosts (*Larix* species) during winter and migration of wind-dispersed spores. Previous studies of *M. medusae* populations suggest that inoculum source and genetic composition vary from year to year outside the range of alternative host sympaty [6]. This is also evident in *M. scabiana* populations in the coastal Pacific Northwest, where *Larix* spp. are not native and rust pathotypes varied across years at single locations [4].

Poplar leaf rust resistance has been extensively studied using the European counterpart, *M. larici-populina* [7–11]. Bi-parental linkage mapping has identified major resistance loci that map to nucleotide binding site – leucine rich repeat (NBS-LRR) R genes on chromosome 19 [7–9,11]. During the mid-1990s, *P. trichocarpa* × *P. deltoides* ‘Beaupré’ was bred for rust resistance and then exclusively planted in European plantations. Yet, a rust epidemic occurred after the R gene was defeated by the evolution of virulence factor 7 [1,12]. In wheat, NBS-LRRs have also conferred race-specific resistance to stem rust; however, association mapping has recently been applied to validate non-R gene non-race-specific resistance loci [13].

Association mapping in plants has the capability to precisely identify a greater number of functional variants which explain smaller proportions of the phenotypic variance than traditional linkage analysis [14]. The development of genomic resources necessary for association analysis in *Populus* species have been facilitated by its value to wood and paper industries and its potential as a biofuels feedstock. With large unstructured

---

Citation: La Mantia J, Klášťér J, El-Kassaby Y, Azam S, Guy RD, et al. (2013) Association Analysis Identifies *Melampsora × columbiae* Poplar Leaf Rust Resistance SNPs. PLoS ONE 8(11): e78423. doi:10.1371/journal.pone.0078423

Editor: Ivan Baxter, United States Department of Agriculture, Agricultural Research Service, United States of America

Received May 9, 2013; Accepted September 19, 2013; Published November 13, 2013

Copyright: © 2013 La Mantia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Genome British Columbia (103BIO) Applied Genomics Innovation Program and Genome Canada (168BIO) Large Scale Applied Research Project funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declare that no competing interests exist.

* E-mail: richard.hamelin@ubc.ca
populations and wide phenotypic diversity, wood traits in *P. trichocarpa* have begun to be studied via association analysis [15]. Here, we report the first multi-year association analysis of poplar leaf rust resistance SNPs in *P. trichocarpa*.

**Methods**

**Plant Material and Phenotyping**

The ramets of 456 genotypes of native black cottonwood were collected from the common garden of British Columbia Ministry of Forest, Lands and Natural Resource Operations (MOFLNRO) at Surrey, BC in March of 2008. MOFLNRO collected these native poplar genotypes from 136 provenances from 44.00 degrees north latitude (Oregon USA) to 59.34 degrees north latitude of (Alaska USA) under the authority of Dr. Alvin Yanchuk, Technical Advisor for the Tree Improvement Branch (Victoria, BC Canada). In June of 2008, four replicates of each genotype were planted in a common garden situated at the University of British Columbia in Vancouver, Canada (49.27 degree north latitude). Replicates were planted in a completely randomized design with 1.5×1.5 meter spacing. In 2008 and 2009, fields were watered daily by rainfall or drip irrigation. No fertility or soil amendments were applied at any time. In 2009, 2010, and 2011 amendments were applied at any time. In 2009, 2010, and 2011, fields were watered daily by rainfall or drip irrigation. No fertility or soil amendments were applied at any time.

**SNP Genotyping**

We genotyped a total of 456 clones of the *P. trichocarpa* population using an Illumina Infinium® genotyping array with a set of 34,131 SNPs in 3,543 candidate genes. The 34K SNP genotyping array we employed was designed to take linkage disequilibrium (LD) into account. SNPs in any given candidate set of 34,131 SNPs in 3,543 candidate genes. The 34K SNP genotyping array we employed was designed to take linkage disequilibrium (LD) into account. SNPs in any given candidate gene represented on the array were chosen to “tag” as many other

| Table 1. Analysis of variance (ANOVA) testing the effect of clone, year, and clone × year interaction of AUDPC. |
|---------------------------------------------------------------|
| **Sources of Variance** | **Estimate** | **St. Error** | **Z value** | **Pr>z** |
| Clone | 44.6077 | 3.0644 | 14.56 | <0.0001 |
| Year | 27.9939 | 28.0179 | 1 | 0.1589 |
| Clone × Year | 5.1717 | 0.5507 | 9.39 | <0.0001 |
| Residual | 27.2434 | 0.5616 | 48.5 | <0.0001 |

Data normality was tested using Lilliefors (Kolomorov-Smirnov) test in R package “nortest”.

**SNP Genotyping**

We genotyped a total of 456 clones of the *P. trichocarpa* population using an Illumina Infinium® genotyping array with a set of 34,131 SNPs in 3,543 candidate genes. The 34K SNP genotyping array we employed was designed to take linkage disequilibrium (LD) into account. SNPs in any given candidate gene represented on the array were chosen to “tag” as many other

on the basis of pustules present on the leaves. Ratings were taken on a 0–4 scale where (0) = no pustules, (1) = less than five pustules per leaf on less than five leaves, (2) = less than five pustules per leaf on more than five leaves, (3) = more than five pustules per leaf on more than five leaves, (4) = more than five pustules on all leaves. Ratings were taken once a week for 11 consecutive weeks (Julian Day 200–279). Ratings were used to calculated area under the disease curve (AUDPC) using the following equation [16]:

$$\sum_{i=1}^{n} \left( Y_{i} + Y_{i+1} \right) / 2 \left( X_{i+1} - X_{i} \right)$$

where, $Y_{i}$ is the disease rating at the $i^{th}$ observation, $X_{i}$ is the time at the $i^{th}$ observation, and $n$ is the total number of observations. Genotypes with missing scores from all four replicates were removed from the study and reduced the population to 412 genotypes. Date of bud set was taken concurrently with disease ratings and varied widely (data not shown). Host age can interact with disease resistance [17] thus, AUDPC scores were adjusted for bud set using ANCOVA in Minitab v16 (Minitab® Statistical Software). Finally, all adjusted AUDPC scores were transformed for normalization using the following equation:

$$X_{i} = \sqrt{(AUDPC_{i} + 1)}$$

Data normality was tested using Lilliefors (Kolomorov-Smirnov) test in R package “nortest”.  

| Table 2. Correlation coefficients of the population structure, latitude, and AUDPC in each year. |
|---------------------------------------------------------------|
| **Latitude** | **Q1** | **Q2** | **Q3** | **PC1** | **AUDPC09** | **AUDPC10** |
| Q1 | −0.37016 | <0.0001* |
| Q2 | −0.08427 | −0.74541 | 0.0876 | <0.0001 |
| Q3 | 0.65247 | −0.49093 | −0.2148 | <0.0001 | <0.0001 |
| PC1 | −0.62464 | 0.65285 | −0.17031 | −0.73393 | <0.0001 | −0.0001 |
| AUDPC09 | 0.73339 | −0.42216 | 0.09914 | 0.06058 | −0.67059 | <0.0001 | <0.0001 |
| AUDPC10 | 0.67009 | −0.53882 | 0.05547 | 0.71694 | −0.73558 | 0.77123 | <0.0001 | <0.0001 |
| AUDPC11 | 0.73398 | −0.50088 | 0.09273 | 0.61266 | −0.69105 | 0.79764 | 0.8213 | <0.0001 | <0.0001 |

*p* value of Pearson’s correlation coefficient (n = 412).  
Q1, Q2, Q3 represent the three sub-populations revealed by GENELAND analysis where pairwise FST were calculated. Q1 × Q2 = 0.0118, Q1 × Q3 = 0.0226, and Q2 × Q3 = 0.02. PC1 represents the first principal component used to correct for population stratification.

[10.1371/journal.pone.0078423.t0001](http://doi.org/10.1371/journal.pone.0078423.t0001)
Figure 1. Quantile-quantile plots of expected and observed \( P \) values evaluating the type-1 error in a simple model (simple), the top ten principal components (PCA-TOP10), BIC selected PCs (PCA-BIC), the K model (KINSHIP), and the \( Q \) model (\( Q \) MATRIX) with goodness of fit test using Bayesian Information Criterion for 2009, 2010, and 2011 (top to bottom).

doi:10.1371/journal.pone.0078423.g001
target SNPs as possible (based on LD calculations), with a SNP density of approximately 1–2 SNPs per candidate gene kb [18]. We eliminated SNPs with: i) minor allele frequency below 0.05, ii) more than 10% missing values, and iii) an Illumina GeneTrain score below 0.5. These three selection criteria reduced the number of SNPs to 29,355. These remaining SNPs were used in all subsequent analyses.

Population Structure

To fit population structure effect, we used a subset of 899 randomly selected SNPs distributed across all 19 chromosomes with complete information (i.e., no missing data) and meeting HWE expectation (tested using “HWChisq” function implemented in “HardyWeinberg” R package [19]. Population fit was done by performing principal component analysis (PCA) in TASSEL [20] and 263 principal components accounting for 90% of the total variance in the SNP data were retained for further determination of their impact. Principal components affecting AUDPC in each year were selected through regression in a stepwise manner using the function “stepwise” implemented in R package “Rcmdr” with “backward” direction and Bayesian information criterion “BIC” as the selection criterion.

Kinship matrix was calculated using the above mentioned 899 SNPs in SPAGeDi [21]. All negative values were set as zero and diagonal elements were set to one [22]. $Q$ matrix and $F_{ST}$ were calculated using GENELAND software with a subset of 200 SNPs randomly selected from the 899 used in PCA. Pearson’s product moment correlations for latitude, AUDPC, $Q$ matrix, and PC1 were calculated in R package “Rcmdr”.

Association Analysis

We applied a two-step approach to analyze SNP-AUDPC association [23]. First, a simple linear regression with AUDPC in each year and every SNP was used to pre-select SNPs with the following equation:

$$ Y = \mathbf{X}\beta + e $$

(3)

where, $Y$ is the observations vector, $\beta$ is the fixed effects of population mean and SNP effect vector, $\mathbf{X}$ is the incidence matrix assigning fixed effects to observations, and $e$ is the residual effect. SNP genotypes were coded as 0, 1, and 2 for common allele homozygote, heterozygote, and rare allele homozygote, respectively. SNPs with significant effect ($P<0.001$) were included in the second analysis.

Finally, the selected principal components were included in a regression model along the screened SNP individually as follows:

$$ Y = \mu + S\mathbf{x} + \sum_{j=1}^{K} P_j\beta_j + e $$

(4)

where, $Y$ is vector of measurements, is the population mean, is the SNP effect, the term represents the effect of selected principal components resulting from backward stepwise selection procedure (above), and $e$ is the residual effect. Association analysis was performed in TASSEL [20] employing the GLM procedure. Permutated $P$ value was calculated in TASSEL with 1,000 permutations. The correction for multiple testing was applied at $\alpha <0.05$. Cumulative effect of SNPs within each year was calculated using method described by Ingvarsson et al. [24]. Pairwise LD plots was calculated using the “LDheatmap” function implemented in the LDheatmap R package [25].

Variance Components and Spatial Analysis

A REMLmixed linear model was used to estimate variance components in SAS and test the effect of clone, year, and clone x year interaction as follows:

$$ Z_{ijk} = \mu + Y_i + C_j + Y_i \times C_j + e_{ijk} $$

(5)

where, $Z$ is vector of measurements (AUDPC) in the $ith$ year, of the $jth$ clone, in the $kth$ ramet, $\mu$ is the population mean, $Y_i$ = effect of the $ith$ year, $C_j$ = effect of the $jth$ clone, $Y_i \times C_j$ = effect of the $ith$ year x $jth$ clone interaction, $e$ is the residual effect. Broad-sense heritability in each year was calculated using methods described in Lynch & Walsh [26].

In each year the mixed linear model implemented in ASReml [27] was used to plot the residuals to their location in the field as follows:

$$ Y = X\beta + Z\mu + e $$

(6)

where $Y$ is vector of measurements, $\beta$ and $\mu$ are vectors of fixed (intercept and population) and random (genotypic values) effects assuming $U \sim N(0, I)$ and $\text{Var}(\mu) = I$. $e$ is vector of residual effects assuming $E \sim N(0, \text{Var}(\epsilon)) = I$ where $I$ is identity matrix containing 1’s on diagonal and 0’s at diagonal-off elements, and $X$ and $Z$ are index matrices assigning both fixed and random effects to measurements. $Q$ matrix from GENELAND analysis was used to fit population structure effect.

Results

Disease Analysis

To identify SNPs that confer non-race-specific resistance to $M. \times$columbiana, we performed association analysis on 412 unrelated $P. \times$incisa genotype from a North American provenance trial ranging from Alaska to Oregon. Poplar leaf rust severity was scored (0–4 worst) on natural infection in a replicated (ramets = 4) common garden experiment where ratings were taken over 11 continuous weeks (Julian days 200 – 279) in each of three consecutive years. Rust severity ranged from zero rust pustules after 11 weeks (complete resistance) to 100% of the leaves covered after four weeks (Julian day 229).

Area under the disease progress curve (AUDPC) was calculated from the disease ratings over time. Previously, AUDPC had the highest broad-sense heritability ($HF = 0.69$) among four other measures of Melampsora resistance in growth chamber assays with artificial inoculations [28]. We estimated broad-sense heritability for AUDPC at $HF = 0.72$, $0.65$, and $0.58$ for each of the three years, respectively. Analysis of variance indicated that clone and clone x year interaction were significant, while year was not significant (Table 1). Spatial analysis of the experimental plot also demonstrated a change in the pattern of infection across years (Fig. S1).

Population Structure and Association Analysis

Analysis of population structure was tested using GENELAND software [29]. An uncorrelated allele frequency model did not detect any population structure while a correlated allele frequency model revealed three sub-populations with weak systemic structure.
| Year | Scaffold SNP | Gene Model | Gene | Annotated Gene | Marker R² | Permutated P value | Marker P value | Marker R² |
|------|--------------|------------|------|----------------|-----------|-------------------|---------------|-----------|
| 2009 | 10 1921751   | POPTR_0010s22230 | exon | At5g02810.1   | PRR7 (PSEUDO-RESPONSE REGULATOR 7) | 7.343–E-07 | 0.003             | 0.0289     |
|      | 14 3245282   | POPTR_0011s26790 | intron | non arabidopsis blast hit | unknown protein | 1.45–E-06 | 0.003             | 0.0271     |
|      | 12 1814218   | POPTR_0012s02170 | intron | At4g08920.1 | pectinesterase family protein | 4.489–E-06 | 0.012             | 0.0280     |
|      | 15 5749284   | POPTR_0005s25750 | intergenic | At1g19870.1 | IQD32 (IQ-domain 32) | 7.869–E-06 | 0.025             | 0.0238     |
|      | 2 13131622   | POPTR_0002s17360 | intergenic | no arabidopsis blast hit | unknown protein | 7.735–E-06 | 0.022             | 0.0259     |
| 2010 | 10 1921751   | POPTR_0010s22230 | exon | At5g02810.1   | PRR7 (PSEUDO-RESPONSE REGULATOR 7) | 4.493–E-11 | 0.001             | 0.0476     |
|      | 5 23949327   | POPTR_0005s25750 | intergenic | At1g19870.1 | IQD32 (IQ-domain 32) | 2.713–E-07 | 0.003             | 0.0302     |
|      | 5 10782555   | POPTR_0005s13780 | intergenic | At2g23760.1 | BLH4 (BEL1-LIKE HOMEODOMAIN 4) | 4.678–E-07 | 0.004             | 0.0292     |
|      | 5 23952538   | POPTR_0005s25750 | exon | At1g19870.1 | IQD32 (IQ-domain 32) | 7.994–E-07 | 0.004             | 0.0281     |
|      | 6 1402770    | POPTR_0006s02140 | intergenic | At4g15090.1 | FAR1 (FAR-RED IMPAIRED RESPONSE 1) | 3.648–E-06 | 0.016             | 0.0253     |
|      | 6 1397889    | POPTR_0006s02140 | 3' UTR | At4g15090.1 | FAR1 (FAR-RED IMPAIRED RESPONSE 1) | 4.079–E-06 | 0.017             | 0.0244     |
|      | 9 10970414   | POPTR_0009s13880 | intergenic | At4g02390.1 | APP (ARABIDOPSIS POLY(ADP-RIBOSE) POLYMERASE) | 3.522–E-07 | 0.003             | 0.0342     |
|      | 6 1402469    | POPTR_0006s02140 | intergenic | At4g15090.1 | FAR1 (FAR-RED IMPAIRED RESPONSE 1) | 8.057–E-06 | 0.031             | 0.0242     |
|      | 10 21451968  | POPTR_0010s26100 | 5' UTR | At3g54540.1 | AtGCN4; transporter | 8.236–E-06 | 0.007             | 0.0318     |
|      | 17 12392905  | POPTR_0017s12210 | 3' UTR | At5g61430.1 | ANAC100 (ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 100) | 1.030–E-06 | 0.017             | 0.0286     |
|      | 9 1676227    | POPTR_0009s01420 | intergenic | At1g08090.1 | PtNRT2.1 (NITRATE TRANSPORTER 2.1) | 1.924–E-06 | 0.007             | 0.0259     |
|      | 9 1678826    | POPTR_0009s01420 | intergenic | At1g08090.1 | PtNRT2.1 (NITRATE TRANSPORTER 2.1) | 1.924–E-06 | 0.007             | 0.0259     |
|      | 9 1606213    | POPTR_0009s01330 | exon | At3g45040.1 | phosphatidate cytidylyltransferase family protein | 2.986–E-06 | 0.014             | 0.0270     |
|      | 9 1857142    | POPTR_0009s01490 | intergenic | At5g60720.1 | unknown protein | 4.830–E-06 | 0.017             | 0.0308     |
|      | 8 4165833    | POPTR_0008s06920 | intergenic | At5g05610.1 | AL1 (ALFIN-LIKE 1) | 5.717–E-06 | 0.017             | 0.0266     |
|      | 9 1679212    | POPTR_0009s01420 | 3' UTR | At1g08090.1 | PtNRT2.1 (NITRATE TRANSPORTER 2.1) | 5.744–E-06 | 0.017             | 0.0266     |

Poplar Leaf Rust Resistance SNPs

PLOS ONE | www.plosone.org 5 November 2013 | Volume 8 | Issue 11 | e78423
were repeated in all years (PSEUDO RESPONSE REGULATOR 7 (NRT2.4)) were also significant (Table 3).

Bonferroni correction for multiple testing was applied at \( \alpha = 0.05/34,136 \) where \( P = 1.70 \times 10^{-6} \). Single SNPs located in PSEUDO RESPONSE REGULATOR 7 (PRR7) and IQD32 were repeated in three and two years respectively. In addition, two SNPs in high linkage disequilibrium (LD) \( (R^2 > 0.8) \) located in the intergenic region between NITRATE TRANSPORTER 2.1 and NITRATE TRANSPORTER 2.4 were also significant (Table 3).

In the final analysis, we used a simple linear regression with AUDPC in each year to preselect SNPs and reduce the constraint of multiple testing correction. Significant SNPs at \( P < 0.001 \) were selected and re-run in a linear regression with the PCA-BIC model to correct for population structure. In 2009, 2010, and 2011 a total of 9, 11, and 20 SNPs achieved experiment-wide significance at \( P < 1.45 \times 10^{-5} \). Bonferroni correction for multiple testing was applied at \( \alpha = 0.05/29,355 \) where \( P = 1.70 \times 10^{-6} \). Single SNPs located in PSEUDO RESPONSE REGULATOR 7 (PRR7) and IQD32 were repeated in three and two years respectively. In addition, two SNPs in high linkage disequilibrium (LD) \( (R^2 > 0.8) \) located in the intergenic region between NITRATE TRANSPORTER 2.1 and NITRATE TRANSPORTER 2.4 were also significant (Table 3).

In the final analysis, we used a simple linear regression with AUDPC in each year to preselect SNPs and reduce the constraint of multiple testing correction. Significant SNPs at \( P < 0.001 \) were selected and re-run in a linear regression with the PCA-BIC model to correct for population structure. In 2009, 2010, and 2011 a total of 9, 11, and 20 SNPs achieved experiment-wide significance at \( P < 1.45 \times 10^{-5} \). Bonferroni correction for multiple testing was applied at \( \alpha = 0.05/29,355 \) where \( P = 1.70 \times 10^{-6} \). Single SNPs located in PSEUDO RESPONSE REGULATOR 7 (PRR7) and IQD32 were repeated in three and two years respectively. In addition, two SNPs in high linkage disequilibrium (LD) \( (R^2 > 0.8) \) located in the intergenic region between NITRATE TRANSPORTER 2.1 and NITRATE TRANSPORTER 2.4 were also significant (Table 3).

Single SNPs located in PRR7 and IQD32 were repeated in all three years (Table 1). In addition, SNPs in FAR RED IMPAIRED RESPONSE 1 (FAR1), a phosphatidylinositol 4-phosphate 5-kinase (POPR_0008126780; PIPE5), and NITRATE TRANSPORTER 2.4 (NRT2.4) were associated in a single year. In host infection transcriptome analysis, these genes were differentially expressed in incompatible interactions with additional Melampsora species (unpublished data). Arabidopsis orthologs corresponding to genes housing these five SNPs indicate functions in host defense through transcriptome reprogramming, calcium and salicylic acid signaling, and tolerance to reactive oxygen species.

Discussion

In prior descriptions of this population, growth traits and population stratification differentiated in a north to south pattern [15,18,30,33,34]. This differentiation may be driven by intense selection pressure for adaptation to day-length and physical barriers impeding gene flow [33]. Rust severity was also correlated to latitude. Rust aggressiveness can be reduced in below optimal temperatures (20°C) in both poplar leaf rust and wheat rust (FST <0.0227) consistent with our previous results [18,30]. Components of the population structure also displayed strong correlation to AUDPC and latitude (Table 2; Fig. S2). Due to this correlation of phenotype and population structure, the trait-SNP simple model (simple linear regression) produced 941, 1220, and 1093 significant associations at \( P < 1.72 \times 10^{-6} \) in each of the three years, respectively (data not shown) and a prodigious inflation of type-I error (Fig. 1).

During association analysis, we tested the effects of \( Q \) matrix, PCA-based model that uses the first 10 PC’s (PCA-TOP10) [31] PCA-based model with PCs that affect AUDPC selected through a backwards step-wise regression (PCA-BIC), and kinship matrix [32]. In a goodness of fit test, kinship matrix had the lowest Bayesian Information Criterion (BIC) value; however in the 2010 and 2011 quantile-quantile plots (q-q plots), kinship matrix eliminated all of the expected associations. The PCA-BIC displayed the next best model fit without overcorrecting for structure (Fig. 1).

Association analysis was performed with 29,355 SNPs covering 3,543 genes in a linear regression with AUDPC using the PCA-BIC model in each year. In 2009, 2010, and 2011 a total of three, four, and three SNPs were significantly associated at \( P < 1.46 \times 10^{-6} \). Bonferroni correction for multiple testing was applied at \( \alpha = 0.05/29,355 \) where \( P = 1.70 \times 10^{-6} \). Single SNPs located in PSEUDO RESPONSE REGULATOR 7 (PRR7) and IQD32 were repeated in three and two years respectively. In addition, two SNPs in high linkage disequilibrium (LD) \( (R^2 > 0.8) \) located in the intergenic region between NITRATE TRANSPORTER 2.1 and NITRATE TRANSPORTER 2.4 were also significant (Table 3).
Figure 2. Manhattan plot of the results from association analysis for AUDPC in 2009, 2010, and 2011 (from top to bottom). The red line represents the \( P \) value (\( P < 1.45 \times 10^{-2} \), \( P < 1.15 \times 10^{-3} \), and \( P < 1.35 \times 10^{-3} \) in 2009, 2010, and 2011, respectively) corresponding to permuted \( P \) of \( \alpha = 0.05 \) as the threshold for multiple testing corrections. SNPs repeated in time are highlighted in green and identified by gene name in black. SNPs within genes showing expression profile changes in response to \textit{M. larici-populina} are highlighted in green and identified by gene name in red. doi:10.1371/journal.pone.0078423.g002

Figure 3. Pairwise linkage disequilibrium plot of \textit{FAR-RED IMPAIRED RESPONSE1} and \textit{FAR-RED ELONGATED HYPOCHOTYL3} with gene structures. SNPs significant in 2010 are indicated with an asterisk; SNPs significant in 2011 are indicated with a cross. Scaffold_6_1402770 in the intergenic region had the highest significance (\( P = 3.64 \times 10^{-6} \)) and explained 2.5\% (\( R^2 = 0.025 \)) of the phenotypic variance in 2010. doi:10.1371/journal.pone.0078423.g003
et al and IQD32 genome v3. SNPs, FAR1 3.52 were in low PIK5 PRR7 2 is a gene within a small family of circadian = 0.034). Scaffold_143_2955 was annotated to Potri.009G008500 in JGI Populus trichocarpa PIK5 (Fig. S3). In matrix P prr7 IQD PRR7 IQD1 PRR9 has been shown to function in defense PRR7 prr5 loss of function single mutant phenotypes: PRR7 prr9 has been shown to function in defense PRR7 (Fig. S4). This gene is orthologous to a gene NITRATE TRANSPORTER2.1 and was down-regulated NITRATE TRANSPORTER2.4 on AUDPC in 2010 (R our study, GENELAND analysis indicated very weak stratification considered to fit both confounding factors efficiently [31,40,41]. In complex polygenic traits. Moreover, the kinship matrix itself is improved the estimate of the genetic relatedness [37–39] especially in the model, which is usually fit solely by kinship matrix, has also considered to fit both confounding factors efficiently [31,40,41]. In our study, GENELAND analysis indicated very weak stratification among three sub-populations (FST <0.0227). The use of Q matrix in the association model resulted in decreased fit and inflated type-1 error. Kinship matrix had the lowest BIC value in a goodness of fit test, suggesting that it is the best model to correct for the confounding structure; however the q-q plots indicated that kinship matrix eliminated the expected associations and overcorrected the model (Fig. 1).

Alternatively, Price et al. [42,43] employed principal component analysis to improve the correction for population stratification and the confounding effects of phenotype – population structure correlation. They proposed using a fixed number of principal components (first 10) or ones selected on the basis of Tracy-Widom statistics [44] when admixture occurred in population regardless of their relationship to phenotype. Methods using a stepwise regression to select a set of SNPs [23,46] or principal components [23,48] have been suggested to fit the confounding structure and used as regressors in the final association analysis model. Novembre & Stephens [47] also indicated that inclusion of principal components not correlated with the trait may reduce power. In our study, the inflation of significant associations and the goodness of fit in the PCA-BIC model further supports this hypothesis (Fig. 1).

We also used SNP pre-selection to reduce the constraint of multiple testing corrections on inflation of false negative associations [23,48]. We reason that elevating the pre-selection threshold from P<0.05 to P<0.001 would remove erroneous SNPs that would have been selected via the correlation of AUDPC and population stratification and thus increase type-2 error. In 2009, pre-selection at P<0.05 would have selected 10,828 SNPs for AUDPC; where P<0.001 reduced the SNP selection to 3,905 (data not shown). Moreover, SNP associations in FAR1 and PIP5 were only achieved via SNP pre-selection but correlated to rust resistance through transcriptome analysis during incompatible poplar leaf rust interactions.

Associated SNPs within PRR7, IQD32, and PIP5 were in low LD with the adjacent SNPs. Thus, these SNPs may be causative variants or in high LD with the unrepresented causative SNP. Conversely, several SNPs within the neighboring gene pairs; FAR1 and FH13, and NRT2.4 and NRT2.1, respectively, were in high LD which condules the elucidation of the true causative SNP. Scaffold_10_ 19,215,715 is a non-synonymous polymorphism in the fifth exon of a sequence orthologous to PRR7 (Fig. S3). In Arabidopsis, PRR7 is a gene within a small family of circadian clock gene transcription factors [49]; however, it was not associated with phenological traits in this population (personal communications, Athena McKown). The prr7 loss of function mutants has an ambiguous phenotype, but double and triple mutants accentuate the prr5 and prr9 single mutant phenotypes: arrhythmia with increased hypocotyl elongation, leaf number, and days to flowering [49]. More recently, PRR7 was down-regulated in response to chitoctoaase (chitin oligomer; chitin is a component of fungal cell walls). These results would suggest a role of PRR7 transcriptional regulation during host infection [50].

Scaffold_3_23949327 is located 1,826 bp downstream from the 3’UTR of IQD32 (Fig. S4). This gene is orthologous to a calmodulin binding protein that serves as an integral component of Ca2+/calmodulin signaling. In Arabidopsis and rice, IQD gene family members share as many as three calmodulin binding motifs IQ, 1-5-10, and 1-8-14. While IQD gene function has not been well characterized, IQD1 has been shown to function in defense response to herbivory [51].
These single SNPs within PRR7 and IQD32 were associated in all three years. Bi-parental linkage mapping has identified major resistance loci that confer race-specific resistance to M. ×columbiana [4]. In our study, ANOVA and spatial analysis suggest the pathogenicity of the rust population varied across the three years. This is consistent with the reproductive biology of the rust where the non-overlap of poplar-alternative host ranges would affect the genetic composition of the rust population in time. Therefore, we propose that SNP associations replicated in time and in the diversity of M. ×columbiana across the three years confer non-race-specific resistance.

Numerous signals within EAR1 were also significant in 2010, but not repeated in time (Fig. 3). A homolog of EAR1, EAR-RED ELONGATED HYPOCOTYL3 (FHY3), is a clock gene that indirectly mediates the phytochrome A response, but has additional functions. In Arabidopsis, fhy3 mutants regulate plant architecture and abiotic stress tolerance through suppression of axillary bud outgrowth and repressed leaf growth with decreased tolerance to oxidative stress. Loss of function mutants in far1 and resoluta (REV), a leucine-zipper transcription factor, enhance the fhy3 phenotype [52]. In 2011, scaffold_6_1405713 within FHY3 and scaffold_9_5,263,210 within REV were also significant (Table 3). These signals were not repeated in time; however, EAR1 expression was increased 2-fold 96 hours after inoculation with M. larici-populina (unpublished data).

Likewise, scaffold_8_8261867 in the 12 exon of POPTR_0006s12780; encoding a phosphatidylinositol 4-phosphate 5-kinase (PIP5K), was significant in 2011 (Fig. S5) and has been previously implicated in resistance to Melampsora. In Arabidopsis, lower expression of PIP5Ks leads to accumulation of the raffinose family oligosaccharides that act as osmoprotectants and antioxidantrants and protect mitochondria and chloroplasts from stress-induced production of reactive oxygen species [53]. In resistant P. trichocarpa × deltoides, this PIP5K gene was shown to have more than a 2-fold decrease in its expression 48 hours after inoculation with M. larici-populina [54].

The SNP at scaffold_143_2955 encodes a non-synonymous mutation in the third exon of NRT2.4 and is in high LD (R2 > 0.8) with three other associated SNPs in the neighboring NRT2.1 (Fig. 4). These genes are orthologous to the nitrate transporter AtNRT2.1. Nitrate transporters are transmembrane proteins that primarily function in nitrate transport; however, they also function as environmental signal receptors and regulators of biotic and abiotic stress pathways. Recently, it was reported that the mutant nrt2 that lacks the function of both AtNRT2.1 and the adjacent AtNRT2.2 shows decreased susceptibility to Pseudomonas syringae. The decrease in susceptibility is coordinated through an earlier and more robust induction of salicylic acid and up-regulation of defense genes PRI1 and PR5 [55]. In poplar, NTR2.4 was down-regulated in incompatible interactions of P. deltoides with M. occidentalis and M. larici-populina (unpublished data). The change in expression levels of genes housing associated SNPs during incompatible poplar leaf rust interactions further implicates a functional role in host defense.

In summary, we identified two independent loci that were strongly associated with host defense to M. ×columbiana and through repetition in time confer non-race-specific resistance. Furthermore, three other associated loci have been correlated to poplar leaf rust resistance through transcriptome analysis and may form a functional network with additional genes involved in tolerance to reactive oxygen species. In this long-lived ecologically and economically important tree species, these associations lay the foundation to more efficient breeding of durable disease resistance.

Supporting Information

Figure S1 Spatial distribution of residuals for AUDPC in each year. The scale of residuals ranges from −20 (low disease) to 20 (high disease). (TIFF)

Figure S2 Population structure estimates and geographical distribution of each sampled tree (n = 412). Colors designate the three sub-populations detected using GENELAND analysis (Q matrix). (TIFF)

Figure S3 Pairwise linkage disequilibrium plot of PSEUDO-RESPONSE REGULATOR7 and gene structure. Scaffold_10_19215715 is indicated with an asterisk. In 2009, 2010, and 2011 (from left to right) each box plot shows the lower quartile, the median, and the upper quartile values, and the whiskers show the range of the phenotypic variation in the population. (TIF)

Figure S4 Pairwise linkage disequilibrium plot of IQDOMAIN32 and gene structure. Scaffold_5_23949327 is indicated with an asterisk. In 2009, 2010, and 2011 (from left to right) each box plot shows the lower quartile, the median, and the upper quartile values, and the whiskers show the range of the phenotypic variation in the population. (TIF)

Figure S5 Pairwise linkage disequilibrium plot of PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE and gene structure. Scaffold_8_8261867 is indicated with an asterisk. In 2011, the box plot shows the lower quartile, the median, and the upper quartile values, and the whiskers show the range of the phenotypic variation in the population. (TIF)

Acknowledgments

The authors acknowledge Nicolas Feau’s work on RNA-Seq transcriptome analysis of Populus × Melampsora interactions and his review of the manuscript.

Author Contributions

Conceived and designed the experiments: YEK RG CD SM RH. Performed the experiments: JL SA. Analyzed the data: JL JK. Wrote the paper: JL JK.

References

1. Steenackers J, Steenackers M, Steenackers V, Stevens M (1996) Poplar diseases, consequences on growth and wood quality. Biomass Bioenerg 10: 267–274.
2. Widin KD, Schipper AL (1981) Effect of Melampsora medusae inoculation with hybrid poplars. Eur J Forest Pathol 11: 438–448.
3. Newcombe G, Stirling B, Mcdonald S, Bradshaw HD (2000) Melampsora ×columbiana, a natural hybrid of M. medusae and M. occidentalis. Mycol Res 104: 261–274.
4. Newcombe G, Stirling B, Bradshaw HD (2001) Abundant pathogenic variation in the new hybrid rust Melampsora ×columbiana on hybrid poplar. Phytopathology 91: 981–989.
5. Flor HH (1955) Host-parasite interaction in flax rust - its genetics and other implications. Phytopathology 45: 680–683.
6. Bourassa M, Bernier L, Hamelin RC (2007) Genetic diversity in poplar leaf rust (Melampsora larici-populina f. sp. deltoidae) in the zones of host sympatry and allopatry. Phytopathology 97: 603–610.

7. Zhang J, Steenackers M, Storme V, Neyrinck S, Van Montagou M, et al. (2008) Fine mapping and identification of nucleotide binding site-leucine-rich repeat sequences at the MRR locus in Populus deltoides ‘S9-2’. Phytopathology 91: 1069–1073.

8. Lecont M, Rembaux S, Zhang J, Ashbourn S, Mathe C, et al. (2004) Annotation of a 524-kb Populus deltoides genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements. Theor Appl Genet 109: 10–22.

9. Yin TM, DiFazio SP, Gunter LE, Jawdy SS, Borjian W, et al. (2004) Genetic and physical mapping of Melampsora rust resistance genes in Populus and characterization of linkage disequilibrium and flanking genomic sequence. New Phytol 164: 95–105.

10. Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of genetic and virulence diversity of Melampsora larici-populina populations on wild and cultivated poplar and influence of the alternate host. Phytopathology 96: 1027–1036.

11. Bresson A, Jorge V, Dowkiw A, Guerin V, Bourguin I, et al. (2011) Qualitative and quantitative resistance to leaf rust: finely mapped within two nucleotide-binding site-leucine-rich repeat (NBS-LRR)-rich genomic regions of chromosome 19 in poplar. New Phytol 192: 151–163.

12. Dowkiw A, Voisin E, Bastien C (2010) Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. Plant Pathol 59: 523–534.

13. Crossa J, Burgueño J, Dresseacker S, Vargas M, Herrera-Foessel SA, et al. (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177: 1899–1913.

14. Zhu G, Geore M, Buckler E, Yu J (2008) Status and prospects of association mapping in plants. The Plant Genome 1: 5–20.

15. Wegrzyn JL, Eckert AJ, Choi M, Lee JN, Stanton BJ, et al. (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem. New Phytol 188: 515–532.

16. Shander G, Finney RE (1977) The effect of nitrogen fertilization on the shoot branching and stress tolerance in Arabidopsis in an AXR1-dependent manner. Plant Physiol 71: 907–920.

17. Graffelman J, Morales-Camarena J (2006) Graphical tests for Hardy-Weinberg assumption. Bioinformatics 22: 2633–2635.

18. Geraldes A, Difazio SP, Slavov GT, Ranjan P, Muchero W, et al. (2013) A 34K SNP genotyping array for Populus trichocarpa: Design, application to the study of natural populations and transferability to other Populus species. Mol Ecol Resour, in press.

19. Graffelman J, Morales-Camarena J (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. Human Heredity 65: 77–84.

20. Bradley PJ, Zhang Z, Kroune DE, Castevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

21. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to compute and test different measures of genetic similarity. Bioinformatics 18: 1944–1946.

22. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to compute and test different measures of genetic similarity. Bioinformatics 18: 1944–1946.

23. Pant SD, Schnickel FS, Verschoor CP, You Q, Kelton DF, et al. (2010) A circularized genome assembly of the black cottonwood (Populus trichocarpa) genome reveals a disease resistance gene cluster. PLoS ONE 5: e13031.

24. Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, et al. (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. New Phytol 196: 713–725.

25. Zhao K, Arananza MJ, Kim S, Lister C, Shindo G, et al. (2007) An arabidopsis example of association mapping in structured samples. PLoS Genet 3: e4.

26. Loeuille BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). Am J Bot 82: 1420–1425.

27. Xie C-Y, Ying CC, Yanchuk AD, Holmeschuk DL (2009) Ecotypic mode of regional differentiation caused by restricted gene migration: a case in black cottonwood (Populus trichocarpa) along the Pacific Northwest coast. Can J Forest Res 39: 519–526.

28. Xie C-Y, Carlson MR, Ying CC (2012) Ecotypic mode of regional differentiation of black cottonwood (Populus trichocarpa) due to restricted gene migration: further evidence from a field test on the northern coast of British Columbia. Can J Forest Res 42: 400–405.

29. Levine MN (1928) Biometric studies of the variation of physiologic forms of Puccinia graminis tritici and the effects of ecological factors on the susceptibility of wheat varieties. Phytopathology 18: 7–123.

30. Chandrashekar M, Heather WA (1989) Temperature sensitivity of reactions of Populus spp. to races of Melampsora larici-populina. Phytopathology 71: 421–424.

31. Habier D, Fernando RL, Dekkers J (2003) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389–2397.

32. Karkkainen H, Stallkamp M (2012) Robustness of Bayesian multilocus association models to cryptic relatedness. Annals of Human Genetics 76: 510–523.

33. Wang M, Jiang N, Jia T, Leach L, Cockran J, et al. (2012) Genome-wide association mapping of agronomic traits in highly structured populations of barley cultivars. Theor Appl Genet 124: 233–246.

34. Auše W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. Stat Sci 24: 451–471.

35. Stich B, Mohring J, Piepho HP, Heckenberger M, Buckler ES, et al. (2008) Comparison of mixed-model approaches for association mapping. Genetics 174: 1743–1754.

36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.

37. Price AL, Zaiden NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11: 459–463.

38. Patterson N, Price A, Reich D (2006) Population Structure and Eigenanalysis. PLoS Genet 2: e190.

39. Wietzki D, Stierstadt H, Balding DJ (2006) Logistic regression protects against population stratification in genetic association studies. Genome Res 16: 290–296.

40. Li Q, Wacholder S, Hunter D, Hoover R, Chanock S, et al. (2009) Genetic background comparison using distance-based regression, with applications in population stratification evaluation and adjustment. Genet Epidemiol 33: 432–441.

41. Nordheim J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. Nat Genet 40: 616–619.

42. Quresada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, et al. (2010) Association mapping of quantitative disease resistance in a natural population of fohbodly pine (Pinus taeda L.). Genetics 196: 677–686.

43. Namakuch N, Kitr M, Ikri S, Yamashino T, Mizuno T (2005) PSEUDO-RESPONSE REGULATORS (PPR9, PPR7 and PPR5) together play essential roles close to the circadian clock of Arabidopsis thaliana. Plant Cell Physiol 46: 698–698.

44. Lilibault M, Wan J, Czeckowski T, Ucdarh M, Stacey G (2007) Identification of 110 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. Mol Plant-Microbe In 20: 900–911.

45. Levy M, Wang Q, Kaspi R, Parrella MP, Abiel S (2005) Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. Plant J 45: 79–96.

46. Stirmberg P, Zhao S, Williamsson L, Ward S, Leyser O (2012) FYH1 promotes shoot branching and stress tolerance in Arabidopsis in an AXX1-dependent manner. Plant J 71: 987–997.

47. Valturina V, Van den Ende W (2013) Myo-inositol and beyond – Emerging roles in plant defense. Plant J 725.

48. Kinald C, Kohler A, Frey P, Duschassy F, Ninge R, et al. (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust Melampsora larici-populina. Plant Physiol 144: 347–366.

49. Camates G, Pastor V, Cerezo M, Garcia-Andrade J, Vicedo B, et al. (2012) A deletion in NRT2.1 attenuates Pseudomonas syringae induced hormonal perturbation, resulting in primed plant defenses. Plant Physiol 158: 1054–1066.