**The genetic linkages of nearly everything: Tree genotypes and their diverse correlated communities**

L.J. Lamit, M.K. Lau, Z. Compson, Todd XXX69, P. Busby, A.R. Keith, J. Schweitzer, C.A. Gehring, T.G. Whitham

Abstract

* Studies from the field of community genetics have demonstrated the important role genetics can have in structuring communities. However, these studies have focused on subsets of the entire community
* Here, using a diverse set of existing datasets collected from a single common garden experiment we explore patterns of genetic covariation among communities of organisms associated with *Populus angustifolia*.
* Three main results emerged: 1) correlation among *P. angustifolia* communities were detected, 2) communities using similar resources tended exhibited higher levels of correlation; however, communities from more similar taxonomic lineages were not significantly more correlated than more distantly related communities and 3) neither the level of genotype replication nor the amount of time between samples had a significant effect on the level of correlation among communities.
* These results support the general framework of phenotypic effects above the level of the individual and extends the breadth of the community phenotype to a more complex, multi-guild community of organisms. As plant and other foundation species are predicted to respond both ecologically and evolutionarily to environmental change (e.g., climate change), it is important that we understand the far reaching impacts of genetic variation on communities and ecosystems.

Key words: **Introduction**

Recent advances in community genetics show that individual genotypes and the genetic structure of plant species play an important role in defining community structure and ecosystem processes. A series of reviews have established these relationships in diverse systems around the world (Whitham et al. 2003, 2006, 2012, Bailey et al. 2006, 2009, Wade et al. 2007, Haolin and Strauss 2008, Hughes et al. 2008, Genung et al. 2011, Wymore et al. 2011, Rountree et al. 2011, Schweitzer et al. 2012, Fischer et al. 2014, Woolbright et al. 2014). Numerous studies cited in the above reviews demonstrate that related individuals support similar communities and rates of ecosystem processes, such that a plant genotype can be viewed as having its own community and ecosystem phenotype (*sensu* Whitham et al. 2006).

Community and ecosystem phenotypes are the product of interactions between the multivariate phenotype of a plant genotype and its interacting community of associated organisms (Bangert et al. 2006, Barbour et al…. Holeski et al…). Plant associated organisms are sensitive to quantitative variation in plant traits (e.g., secondary chemistry, phenology, morphology; Bangert et al. 2006, Barbour et al. XX, Lamit et al. 2014) that influence their abundances and how they interact with a plant, and many organisms are likely influence by multiple traits simultaneously. The influence of genetic variation in plant traits can cascade to alter complex interactions between community members (Johnson 2008, Smith et al. 2010, Mooney papers, Lamit et al. 2014, Busby et al. in press) and shape the phenotypes of associated organisms (e.g., aphid gall size, Smith et al. 2010). The community phenotype is therefore the extended consequences of: 1) genetically based plant traits affecting abundances of community members, 2) genotype-specific modulation of interactions among community members, and 3) interspecific indirect genetic effects (IIGEs), where the phenotype of an individual in one species is, in part, due to the expression of genes in one or more different species (Shuster et al. 2006, Wade 2007, Allan et al. 2012). These effects may produce important selection pressures that lead to local adaptation of community members to individual plant genotypes or populations (Shuster et al. 2006, Evans et al. 2008, Smith et al. 2010, CITATIONS), and may also feed back to act as selective forces on the plant (Schweitzer et al. 2008, Smith et al. 2010, Lankau and Straus 200X).

Although most community genetics studies focus on a single community associated with individual plant genotypes, it is important to expand these studies to understand the relationships among the diverse and potentially interacting communities of organisms that can be found on, within and near a single plant. For example, Zytynska et al. (2011) examine how different genotypes of the tropical forest tree affected both their associated epiphytic plants and their invertebrate communities (Zytynska et al. 2011). These diverse communities may be genetically correlated among plant genotypes, in which shifts in the composition of one community among genotypes are mirrored by similar shifts in the composition of another community. A rich literature has already demonstrated the frequent existence of genetic correlations among individual species of plant enemies, primarily herbivorous arthropods (e.g., Maddox & Root 1990, Roche & Fritz 1997, Leimu & Korechieva 2006, Johnson et al. 2007). However, there is evidence that plant genetics link diverse groups of plant-associated species (e.g., Dickson & Whitham 1996, Ahlholm *et al*. 2003, Tagu *et al*. 2005, Sthultz *et al*. 2009); thus, the ecological and evolutionary dynamics of one community is likely to constrain or modulate a genetic connection to another community.

linked plant genes (e.g., Dickson & Whitham 1996, Ahlholm *et al*. 2003, Tagu *et al*. 2005, Sthultz *et al*. 2009).

In this paper, we utilize a diverse set of existing datasets collected from the same common garden to understand patterns of genetic covariation among communities of organisms associated with *P. angustifolia*. Our first goal is to examine whether or not different communities are predictably associated with one another on individual plant genotypes by testing for genetic correlations among communities. Having demonstrated that communities are correlated, we then evaluate three hypotheses that could account for these correlations. Because these data are not suitable for addressing other likely hypotheses, we develop additional hypotheses in the discussion. First, we hypothesize that communities associated with similar plant tissues are more strongly correlated than communities from disparate tissue types (resource similarity hypothesis). Species living in close proximity are sharing habitat and may be connected through the same local food web. Living near each other therefore makes them more likely to interact through direct encounters or indirectly through plant trait modulation, and be sensitive to genetic variation in the same plant traits. Second, as an alternative hypothesis (although not mutually exclusive), organisms more closely related to each other will be more strongly correlated than they are with distantly related organisms (taxonomic similarity hypothesis). This hypothesis is based on the assumption that a gene will be more likely to influence related organisms than unrelated organisms. Third, we hypothesize that the presence and strength of genetic correlations among communities will decline as the time between sampling of each community increases (time of sampling attenuation hypothesis). A greater time interval between sampling the communities will be more likely to decouple the direct or indirect effects of communities on each other, and variation in environmental conditions may alter gene expression between sampling dates. Testing these and other hypotheses will help clarify our understanding of how plant genotype structures biodiversity in an ecological and evolutionary context.

**Methods**

***Study System and Common Garden***

*Populus angustifolia* is a mid to upper elevation foundation species of interior western North American riparian habitats. Genotypic differences within this tree species influence the structures of a variety of communities, including foliar arthropods (Keith, Bailey & Whitham 2010), bark lichens (Lamit *et al*. in review), twig endophytes (Lamit *et al*. 2014), foliar pathogens (Busby *et al*. 2014) and soil bacteria and fungi (Schweitzer *et al*. 2008, Lamit 2013), as well as ecosystem processes such as leaf litter decomposition (Schweitzer *et al*. 2005), and there is evidence for genotype-dependent positive plant-soil feedbacks (Smith *et al*. 2011). Genetic influences on communities and ecosystem process are likely due to heritable variation in traits such as foliar chemistry (Bailey *et al*. 2006), productivity (Lojewski *et al*. 2009, Lamit *et al*. 2014) and source-sink relationships within a tree (Compson *et al*. XXXX), although the mechanistic links of *P. angustifolia* genes and traits to community and ecosystems processes remain unresolved in some cases. Because *P. angustifolia* is a model species for community genetics research (Whitham *et al*. 2008), it represents an ideal taxon with which to examine genetic correlations between plant-associated communities.

All community datasets incorporated into this paper were collected from replicated *P. angustifolia* genotypes growing in the North Ogden Nature Center common garden in Ogden, Utah, USA. The garden is located in the core of the *P. angustifolia*’s latitudinal range, and at the lower end of the specie’s elevational distribution along the nearby Weber and Ogden Rivers (latitude = 41.248146, longitude = -111.999830, elevation = 1302 m). Garden soils are coarse-loams, which are common in riparian habitats of the region (Schweitzer *et al.* 2008). The garden was planted in1991 with cuttings that were clonally propagated from naturally established *P. angustifolia*, as well as smaller numbers of *P. fremontii* (Fremont cottonwood) and their natural hybrids, growing along an ~105 km stretch of the Weber River. The garden is contained within an area of ~1.2 hectares, with trees spaced 4-7 m, and replicates of genotypes were planted in random order. By the time community data were collected, trees in the garden were sexually mature. Re-analysis of the thirty-five codominant RFLP marker dataset from Martinsen *et al*. (2001) indicated that all genotypes used were *P. angustifolia*, with little to no introgression from *P. fremontii* (Lamit *et al*. 2011, Zinkgraf *et al*. unpublished). See Lojewski *et al*. (2009) and Busby *et al*. (2014) for more information on the garden.

***Community Datasets***

Datasets representing a range of communities, both above and belowground, were included in analyses (Table 1). As criteria for inclusion, datasets contained observations from at least six replicated *P. angustifolia* genotypes sampled in the North Ogden Nature Center common garden. Datasets were considered community data when they contained information representative of two or more taxonomic groups, which for most datasets approximated species-level operational taxonomic units (OTUs). We focused on community composition, which incorporates all individual abundances from OTUs in a multivariate matrix. Compared to reduced metrics, such as species richness, composition contains the most information about community structure, and was shown to vary among genotypes for all of the communities we are examining (see citations in Table 1). One exception was made, where web abundances of Agelenid spiders (made by a guild of at least two or more species), was included as a univariate variable because of their important role as predators in this system (Todd’s Oecologia paper, Todd’s dissertation). Although the quantification of OTUs in each community were performed with different methods (e.g., % cover of lichens, counts of individual arthropods, pathogen leaf area damage) we considered each of these metrics as indicators of abundance. See Table 1 for details on each dataset and citations were specific details on sampling can be found.

Prior to analyses between any two community datasets, observations from trees that were not present in both datasets were removed. Matched datasets for pair-wise comparisons were composed of 6 to 26 genotype means, made from XX to XX total trees, representing 2 to 9 trees per genotype (Appendix XXX). Communities in each of these datasets contained between XX and XX species (Appendix XXX). Genotype means for each OTU’s abundance were then calculated from replicates of genotypes that had at least two trees remaining in the matched datasets (range = 2-9 trees per genotype; Table 2). OTUs that occurred in only one or two trees in the matched datasets were removed prior to calculating mean values because rare species provide little information about their affinity to a genotype (McCune and Grace 2002). Furthermore, rare taxa are more likely to occur by chance on genotypes with larger number of replicate trees, thus distorting a genotype’s mean community composition after averaging its abundances. Comparisons between two datasets that retained five or fewer genotypes after matching trees, were not conducted.

***Statistical Analyses***

*Pair-wise Mantel tests and network analysis*

The first set of analyses consisted of a series of bivariate Mantel tests to establish the existence of genetic correlations among communities. This approach tests for genetic correlations by identifying concomitant shifts in the composition of two community among a set of genotypes, where increasing dissimilarity among genotypes in one community is associated with a similar increase in dissimilarity among genotypes in another community. These Mantel tests are therefore one-tailed and are focused on testing for a positive slope. Mantel tests using Spearman’s Rho were performed with each pair of matrices of mean OTU abundances. Bray-Curtis distance was used for all multivariate matrices (i.e., datasets with 2 or more OTUs), and Euclidean distance was used for the only univariate variable, Agelenid spinder webs. Mantel tests were then repeated for each pair of communities using fourth-root transformed mean species abundance matrices, to down-weight the effect of highly abundant species in the matrices (Anderson XXXX). Given the relatively small and uneven number of genotypes in the comparisons, we place emphasis on the strength of the relationships as indicated by Rho values because some comparisons have low power. However, we considered any test where *P* < 0.1 as potentially important, and present raw *P*-values and *P*-values after adjustment for false discovery rate (CITATION) for each of the two sets of pair-wise analyses. Analyses were conducted in R 3.0.2 using the package *ecodist* (CITATION). Rho values from these pair-wise Mantel tests between each community were the formed the central dataset used in all subsequent analyses.

Two approaches were used to visualize and interpret the structure of correlations among the different communities. First, a cluster analyses (see McCune and Grace 2002), was performed with pair-wise Rho values from Mantel tests described in the previous paragraph. This analysis used Ward’s method and XXXXX………... The cluster analysis was performed twice, using Rho values from Mantel tests with relativized and non-relativized datasets. The foliar arthropod community dataset from 2008 was excluded from cluster and network analyses (see next paragraph) because it did not have enough observations to conduct pair-wise tests with all other communities.

Second, a network approach was used to construct plots representing the structure of genetic correlations among communities. This analysis used Rho values from the pair-wise community Mantel tests. Nodes (individual points) in the networks represent each of the communities. Each node in the network is positioned using a physical force algorithm (Butts 2010), and scaled in proportion to its degree centrality in the network (Freeman 1979). Edges (lines connecting points) represent correlations between communities and are scaled by the magnitude of their associated Mantel Rho value.

*Testing hypotheses with partial-Mantel tests*

Our final step was to test our predictions about the factors that influenced the strength of the Rho values from pair-wise Mantel tests. For this analysis, we used the matrix of Rho values obtained from communities with which we were able to make all pair-wise comparisons with the other communities (i.e., comparisons with Arthropods from 2008 were excluded) as the response variable in a partial-Mantel test that contained four addition pair-wise matrices as explanatory variables. The four additional matrices were constructed using pair-wise values for each comparison between communities (see Table 2), these included: 1) a matrix indicating whether or not two communities lived in the same or different locations on a tree, 2) a matrix of values indicating whether or not two communities were within the same general taxonomic group (e.g., fungi, bacteria, arthropods), 3) a pair-wise matrix indicating the approximate number of months between the sampling of each pair of communities, and 4) as a control on variation in sample size among pair-wise tests between communities, a matrix corresponded to the number of genotypes that were used for each pair-wise community Mantel test between each community. The partial-Mantel tests was run using partial Pearson’s correlation coefficients, and tested the linear relationship between each explanatory matrix and the Rho values from pair-wise community comparisons while holding the effect of the three other matrices constant. The partial-Mantel analysis was conducted twice, once with Rho values obtained from pair-wise Mantel tests between communities using unrelativized matrices, and a second time using Rho values from pair-wise Mantel tests between communities conducted with matrices relativized by species maximum values.

**Results**

*Pair-wise Mantel tests and network analysis*

Pair-wise Mantel tests revealed that many diverse communities were correlated among genotypes of *P. angustifolia* (Table 2), indicating that communities of different organisms are linked by plant genotype. Using unrelativized matrices, XXX out of XXX comparisons (%) were significant or showed trends (*P*<0.1), with the majority (%) being significant (*P*≤0.05). Significant correlations included communities of organisms living in close proximity (e.g., leaf modifying arthropods and foliar pathogens), as well as organisms in distinctly different locations on a tree (e.g., ectomycorrhizal fungi and XXXX). Using matrices relativized by OTU maximum, XXX out of XXX comparisons (%) were significant or showed trends (*P*<0.1), with the majority (%) being significant (*P*≤0.05). This suggests that some, but not all, of the significant Mantel tests prior to relativization were due to the effect of the most abundant species driving the relationship. For relativized and unrelativized analyses, adjustment of *P*-values using false discovery rate yielded very similar results to non-adjusted analyses, with only a small number of tests becoming insignificant or marginal (Table 2).

Distinct cluster’s of correlated communities were revealed by cluster analysis and network analysis….

*Testing hypotheses with partial-Mantel tests*

**Discussion**

General patterns observed…who correlated with who?

Important factors from partial mantel

Reason’s for community correlations

Potential subheadings for discussion –

1. Additional hypotheses for the existence of correlated communities

Community consequences of the multivariate phenotype - Genetic similarity rule, Community & ecosystem phenotypes, Community heritability, Community stability, Correlated communities, Interacting foundation species

1. Evolution in a community context

Unraveling the details of the complex eco-evolutionary processes that occur within a community context represents a major frontier in community genetics.

?). Unraveling the details of the complex eco-evolutionary processes that occur within a community context represents a major frontier in community genetics.

**Following is a full list of the current hypotheses that I am aware of. I added text above to clearly state that our current data can only address 3 of these but we will develop the others in the discussion. I think we must develop the full set somewhere as reviewers won’t buy that the three we chose to address represent the full set.**

**Hypotheses for SIGNIFICANT correlations among community types on the same genotypes in which trees with one community type are likely to support another community type** would support the following hypotheses:

**1.** **Resource similarity hypothesis (tissue similarity hyp) –** Communities requiring similar resources (e.g., fungal mycorrhizae and endophytes requiring plant photsynthate) tend to be predictably associated on the same plant genotypes.

**2.** **Resource interaction hypothesis –** One community predictably alters the host environment (e.g., induction or resource depletion) in a way that favors or inhibits another community type at another level (e.g., herbivores reduce leaf area, which in turn reduces photosynthate to fugal symbionts or gall aphid induction of tannins in leaves, which predictably affects the soil decomposer community).

**3.** **Productivity hypothesis –** Plant genotypes that are highly productive in a given environment are more like to support similar communities than plant genotypes that are less productive. Thus, the productivity/diversity hypothesis that is supported for different species extends to the genotype level. Because productivity is, in part, genetically based and may also be involved with hypotheses 1 & 2, it is not mutually exclusive from these hypotheses. For example, in one community such as mycorrhizae increase plant productivity, then the increase in productivity with one mycorrhizal community may favor another community type in the canopy such as arthropods.

**4. Taxonomic similarity hypothesis** - organisms more closely related to each other will be more strongly correlated than they are with distantly related organisms (taxonomic similarity hypothesis). This hypothesis is based on the assumption that a gene will be more likely to influence related organisms than unrelated organisms.

Key tests of these hypotheses might involve resource addition and subtraction experiments to affect productivity, quantifying the effects of induction or resource depletion, experimentally triggering induction, network analyses

**Hypotheses for the LACK of correlations among community types on the same genotypes in which one community type shows no correlation with other community types.**

**1.** **Neutral assembly hypothesis –** Different communities (e.g., soil mycorrhizae and leaf herbivorous arthropods) do not interact directly or indirectly, so they assemble independently of one another and no correlations should be expected.

**2.** **Community sampling hypothesis –** The sampling of whole communities does not reflect the actual interactions because most species are so rare that their inclusion in analyses does not gain resolution, but could detract in its detection. Based on network analyses or methods that ascertain actual interactions, species used in community analyses should focus on the strong interactors and/or the relatively few that account for most of the biomass. E.g., based on Bridgeland et al. (2010) 10% of the 700 insect species on cottonwoods account for 90% of the biomass (the actual numbers need to be confirmed from the paper). Most species are so rare (e.g., singletons) that they don’t play a functional role in the community and should not be included in the analyses.

**3.** **Lack of genetic variation hypothesis –** Lack of correlations may be due to a lack of genetic variation in the population such that the trait variation that different communities respond to is lacking and the same communities are found throughout the plant populations. This would be expected in highly clonal species such as aspen where Meneses et al. (2012) found no genetic variation and no variation in the associated community.

**4. Time of sampling attenuation hypothesis** in that data from different years might not exhibit a significant correlation because of changes in the communities from year to year and our data were collected over a 6 year period. This should makes our findings conservative and one might predict that the farther apart two data sets were collected, the detection of a significant relationship would decline. However, based on Art’s **Community Stability Hypothesis**, we should still expect some relationships to be significant at least over a 3 year period of his study.

**Key underlying concepts/hypotheses** - The above hypotheses are all predicated on the **Multivariate Plant Resource/Defense Phenotype** in which genetic variation in n-dimensional traits (defense, ontogeny, induction, phenology, sink-source relationships, productivity, etc.; derived in part from Holeski et al. 2013 Oecologia 2012) produce great variation among individual genotypes, which in turn results in distinct **community and ecosystem phenotypes** (community phenotype concept from Whitham et al. 2003, Whitham et al. 2006). The greater the variation in the multivariate plant resource/defense phenotype, the greater the variation in the associated communities a species will support. This may be, in part, what makes a foundation species a foundation species. Because these traits are, in part, genetically based, they are **heritable and subject to natural selection** (Shuster et al. 2006). Genotypes that are more similar in their genetics and the traits they express, support more similar communities than those that differ in their genetics and the traits they express (**Genetic Similarity Rule** of Bangert et al. 2006a,b, Bangert et al. 2008). If correlated communities are detected, this would enhance and perhaps even support the concept of **Evolution in a Community Context** and would suggest that we should consider how whole **community can co-evolve**.

Importantly, the more dissimilar two plant genotypes are in their genetic composition, the more dissimilar they are in the communities they support (e.g., Bangert *et al*. 2006, Barbour *et al*. 2009, Zytynska *et al*. 2011, Cordier *et al*. 2012, Bernhardsson *et al*. 2013); this genetic similarity rule (Bangert et al. 2006) provides a predictive framework with which understand the relationship of plant genotype with associated communities and the ecosystem processes they influence.

It seems likely that the interactions of foundation species (i.e., species that drive their respective ecosystems; Dayton 1972, Ellison et al. 2005) and other species of large effect (e.g., keystone, ecosystem engineers, dominant, umbrella species, etc.) are especially important to understand due to their recognized influences in structuring whole habitats. For example, Busby et al. (2014 or in press?) show that individual genotypes of the foundation species *Populus angustifolia* (narrowleaf cottonwood) differ greatly in their resistance to the leaf pathogen, *Drepanopeziza populi*, and pathogen resistant genotypes support different foliar arthropod communities than pathogen susceptible genotypes. When susceptible tree genotypes are inoculated with the pathogen, associated arthropod communities are very different than on the same genotypes that were not inoculated with pathogens. Similarly, with tall fescue (*Lolium arundinaceum*) and its systemic endophyte (*Neotyphodium coenophialum*), an important pair of introduced foundation species in eastern North America, the genotype of both species can interact to influence herbivory and the structure of the surrounding plant community (Rudgers et al. 2010). In these cases, the genetically based interactions of a plant with a strongly interacting symbiotic fungus define a much larger community. Such experiments clearly highlight how plant genotypes define interactions with other foundational species to structure communities. Although there are other interactions with non-foundation species that affect associated community members (refs), the potential to affect very larger communities involving thousands of associated species in multiple trophic levels is likely limited to foundation species that are clearly recognized as ecosystem drivers.

**Table 1.** Information on each community…Taxa, kingdoms, location on/in tree, richness, number of genotypes and trees in the full datasets, reference to original publication.

**Table 2.** Genotype Mantels and Spearman’s correlations for composition and abundance, respectively, and sample sizes and p-values for each comparison.

**Table 3.** Four matrices used as predictors in the partial Mantel test examining factors that affect the strength of pair-wise community Mantel and correlation analyses (See Table 2).

|  |  |  |  |
| --- | --- | --- | --- |
| Predictor matrix | Variable notes | Data in matrix | Question addressed |
| **Occurring on the same or different location on the tree** | Three locations defined: phylosphere = leaves and young twigs, trunk = lower bole and associated dead branches, below ground = litter layer and soil | Two categories; 0 = same location, 1 = different locations | Do communities more strongly covary when they occur in the same location on a tree? |
| **In the same taxonomic group, or not** | Defined as course taxonomic groups: fungi, bacteria, and arthropods | Two categories; 0 = same taxonomic group, 1 = different locations | Is the strength of a pair-wise relationship between two communities dependent on the relatedness of the organisms? |
| **Number of months between sampling** | Differences in months were calculated between the sampling dates of each community | Numerical; range = 0-? | Does the strength of the covariation between two communities depend on the time between samplings? |
| **Number of genotypes** | The number of genotypes used in the Mantel or Spearman correlation. | Numerical; range = 6-26 genotypes | Does sample size influence the strength of the pair-wise relationships? |

Notes: The pair wise matrices for each variable are presented in Appendix XXX.

**Figure legends**

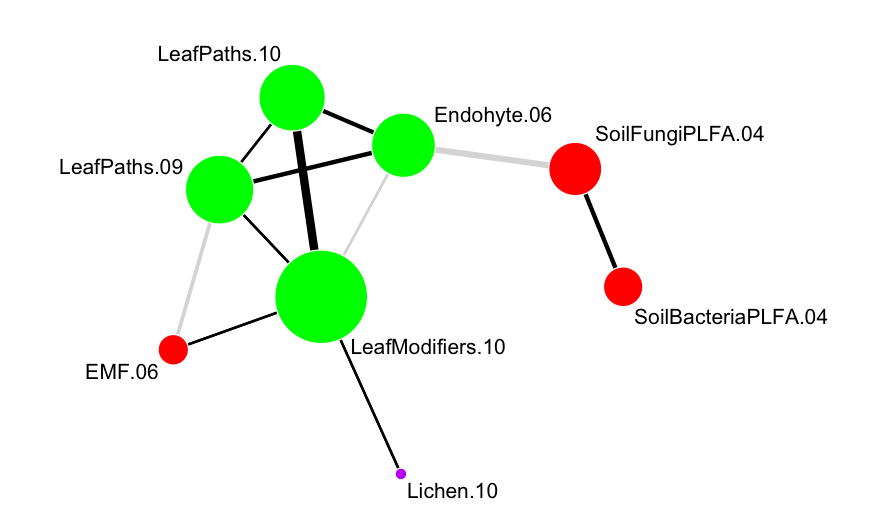
**Figure 1.**

**Figure 2.** Cluster dendrograms (a-c) and network diagrams (d-f) representing the genetic correlation structure among communities on narrowleaf cottonwood. Each network node (a point representing a specific community) is scaled in proportion to its degree centrality in the network and colored green, red or purple for communities of the phylosphere (leaves and young twigs), trunk (lower bole and associated dead branches) or below ground (litter layer and soil), respectively. Edges (i.e., lines connecting communities) in the networks are scaled by the magnitude of their associated Rho value. Signs associated with edges indicate the direction of the correlation.

**Figure 3.**

**Figure 1.** Schematic showing a tree and arrows pointing to the different organisms on it, and their places. Need an artist! Plus a photo of several of the taxa, and of the garden. Reference this in the garden description paragraph.

Figure 2.

Figure 3. Partial mantel model results…