Multi-element fingerprinting and high throughput sequencing identify multiple elements that affect fungal communities in *Quercus macrocarpa* foliage

Ari Jumpponen,[1,2,]* Karen Keating,[3] Gary Gadbury,[3] Kenneth L. Jones[2,4] and J. David Mattox[5]

[1]Division of Biology; Kansas State University; [2]Ecological Genomics Institute; [3]Department of Statistics; Kansas State University; Manhattan, KS USA; [4]Environmental Health Sciences; University of Georgia; Athens, GA USA; [5]City of Manhattan; Manhattan, KS USA

Diverse fungal mutualists, pathogens and saprobes colonize plant leaves. These fungi face a complex environment, in which stochastic dispersal interplays with abiotic and biotic filters. However, identification of the specific factors that drive the community assembly seems unattainable. We mined two broad data sets and identified chemical elements, to which dominant molecular operational taxonomic units (OTUs) in the foliage of a native tree respond most extremely. While many associations could be identified, potential complicating issues emerged. Those were related to unevenly distributed OTU frequency data, a large number of potentially explanatory variables and the disproportionate effects of outlier observations.

Hyperdiverse fungal communities inhabit the foliage of most plants[1,2] and these fungal communities have been reported for virtually every plant that has been examined.[3] Baas-Becking hypothesis states that environment selects microbial communities from the abundant and possibly globally distributed propagule pools.[4] Although the foliage-associated communities—like other microbial communities—are suspected to be sensitive to environmental drivers, determination of the mechanisms that control the assembly of these foliar communities has remained difficult and elusive. Some of the proposed mechanisms include distance limitations to propagule dispersal,[5-7] volume limitations to propagule loads,[7] or limitations set by the environmental conditions either on the scale of the site of fungal colonization[8] or more broadly on a landscape level.[6,9] The forces that may control the fungal community assembly are overlaid by additional biotic controls that include compatibilities between the fungi and host species[10,11] or genotypes[6,12] and the competitive or facilitative interactions among the component fungal genotypes.[6,10-13] Although a variety of potential controls for the foliage-associated fungal communities have been speculated, very little consensus exists on the relative importance of the different drivers. For example, while macronutrient and heavy metal enrichment may have an influence on the composition fungal communities[14] and populations,[15] relative importance of various chemical elements in the foliage remains yet to be investigated.

To evaluate the use of multi-element fingerprinting data produced by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) in combination with high throughput 454-pyrosequencing for determining influential chemical elements in structuring of the leaf-associated fungal communities, we mined a recent dataset[16] that explored the effects of urbanization on the diversity and composition of the fungal communities associated with a native tree *Quercus macrocarpa*. From a total list of more than 700 non-singleton fungal OTUs, we selected fifty with highest overall frequency to provide an observation-rich dataset for elemental effect assessment; these OTUs accounted for 84.5% of all sequences. Even so, many of these OTUs had a number of zero frequencies (**Fig. 1**),
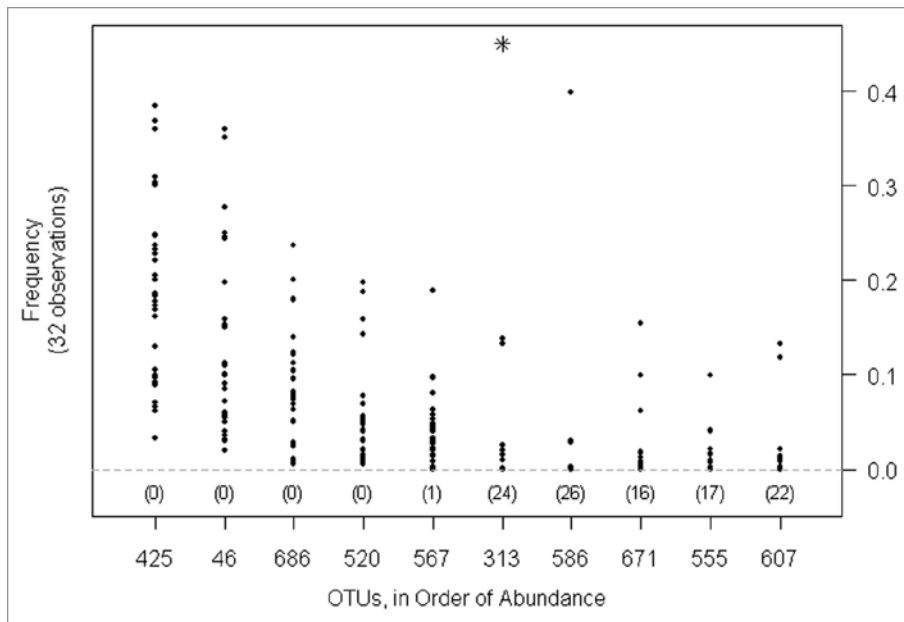
Figure 1. Rank-ordered distribution of observed frequencies for those OTUs whose frequency had an extreme slope when associated with the concentrations of one or more chemical elements in the mixed effects model. The asterisk denotes one extreme frequency for OTU 313 with a value 0.8636. Numbers in parentheses indicate the number of observations with a frequency equal to zero. The OTUs were assigned to approximate taxa using BLAST:[20] 425: *Alternaria alternata* (Pleosporaceae); 46: *Phoma glomerata* (Pleosporaceae); 686: *Aureobasidium pullulans* (Dothioraceae); 520: *Davidiella tassiana* (Davidiellaceae); 567: *Cladosporioum tenuissimum* (Davidiellaceae); 313 *Oidium heveae* (mitosporic Erysiphaceae); 586: *Erysiphe hypogena* (Erysiphaceae); 671: *Mycosphaerella microsora* (Mycosphaerellaceae); 555: Pestalotiopsis sp. (Amphisphaeriaceae); 607: *Pleiochaeta setosa* (incertae sedis).

highlighting one of the difficulties in the use of environmental sequencing data. We omitted one OTU (OTU630 with a likely affinity to *Trimmatostroma cordae* [Mycosphaerellaceae]) that was strongly affected by the original land use design (urbanization; Wilcoxon rank sum test with a Bonferroni adjustment) and therefore unlikely to be representative for the present analyses of elemental drivers. This OTU was replaced with one with the next highest frequency. The frequencies of these 50 OTUs were investigated in the context of concentrations for 29 elements after the omission of five (Ag, Au, C, δ^{13}C, δ^{15}N) in the final analyses because of their strong association with the land use or the difficulty of finding a biological relevance. Of the remaining elements three (Fe, Cr and Ni) had pairwise correlations exceeding 0.98 between the three pairings; others showed no similar high correlations. To allow comparable evaluation across the broad array of elements, all concentrations were standardized to have a mean equal to zero and a standard deviation equal to one.

To rank elements according to their magnitude of association with the abundance of each OTU, a total of 1,450 models (50 OTUs times 29 elements) relating element concentration to OTU abundance were fit to the data. For each model, OTU frequency was the dependent variable, element concentration and time (a factor with three levels) were fixed effects, and—to account for the spatial arrangement of the experimental units—random effects associated with tree nested within site were included in the error structure. Time by element interactions were also investigated and tested using a likelihood ratio test. These mixed effect models were fit using R and the package lme4 (www.r-project.org).

Statistical "tests of significance" that produce p-values can be sensitive to assumptions or outliers. Because of this and the fact that our analyses evaluated a total of 1,450 models, p-values themselves were not considered a reliable measure of importance when associating elements with OTU frequency. Instead,

we emphasized metrics that highlight extraordinary findings rather than rely on tests of statistical significance. This approach facilitates finding few elements that have the strongest effect on OTU frequency. Note that the use of standardized element concentrations (above) provided slope coefficients that are comparable across all models. "Extreme slopes", i.e., models where the OTU response to element concentration was strongest, were identified as those with estimated slope coefficients in the lower or upper 2.5 percentile, i.e., those farther than 1.8 standard deviations from the mean across all estimated slopes (**Fig. 2**). Using this approach, we identified a total of 69 models with extreme slopes (**Table 1**) that we considered the strongest drivers of the OTU frequencies in our data.

Unfortunately, the models with extreme slopes were often affected by high leverage observations (outliers in the explanatory variables) that may have exerted substantial influence on the magnitudes of the slopes. We accounted for this by computing leverage values based on the fixed effect model matrix (element concentration and time) for each model. High leverage was defined as those observations with leverage approximately twice the mean leverage over all samples for a particular model as is considered conventional by some authors.[17] This value was approximately 0.24 for our models. The models with high leverage and extreme slopes were reevaluated by refitting the model to the data after omission of the influential observations. Of the 69 models with extreme slopes only 22 were void of influential observations by our metric (**Table 1**). The model re-evaluation after the omission of observations with the highest leverage identified only seven that were no longer included among those that were in the upper or lower 2.5 percentile of the observed slope estimates (**Table 1**).

Because of the large variability of the frequencies and element concentrations across the sampling events, we felt it necessary to account for the interaction term between the time of sampling and element concentration. Of the total of 69 models, seventeen included a statistically significant interaction term that complicated the interpretation of the chemical element
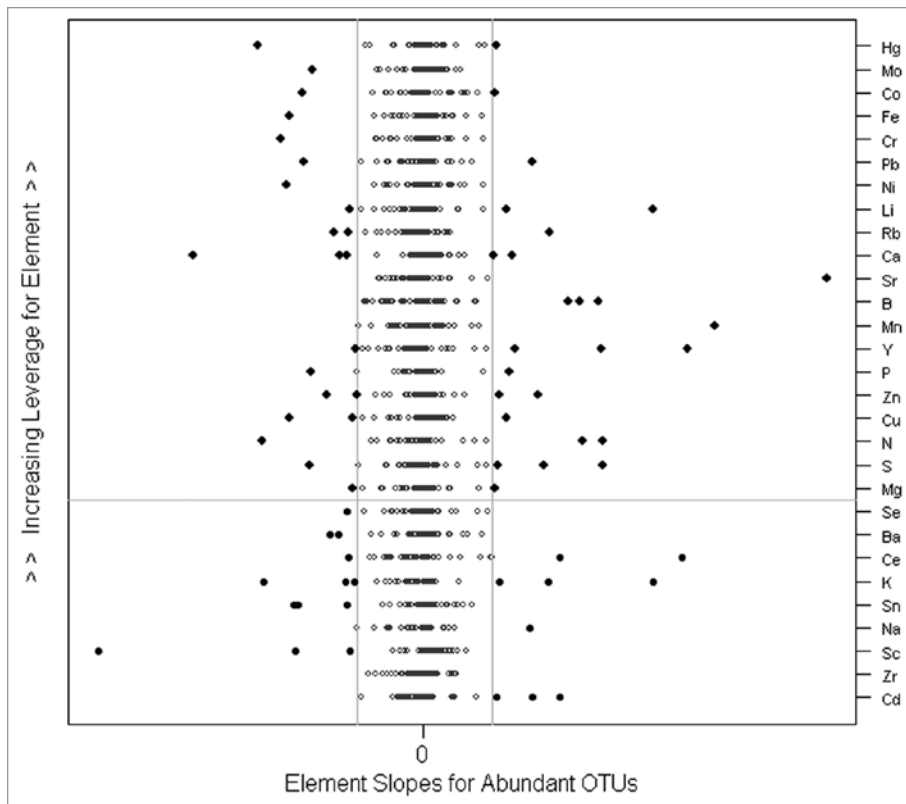
Figure 2. Distribution of estimated slopes (i.e., the slope for element concentration) for a model relating OTU frequency to element concentration, time and a concentration by time interaction, including a tree-nested-within-site random effect. The mean across all 1,450 OTUs is approximately zero; the two vertical lines identify upper and lower 2.5 percentiles, beyond which the slopes were considered extreme (large black symbols). The horizontal line identifies the cut off maximum leverage (0.24), above which the slopes were considered to have observations with high leverage. Models with observations with a high leverage were tested for extreme slopes by refitting without those observations. Models are ranked from bottom to top in order of increasing leverage and the element for which the high-leverage observations and extreme slopes were recorded are identified on the right y-axis.

concentration effects on the OTU frequency (**Table 1**). After omission of the models with interaction terms or those driven by one or more influential observations, 45 models remained. These 45 were the strongest explanatory main effect models for the association of eight of the top 50 most frequent OTUs and explained the association between the OTU frequencies and a total of 27 elements (**Table 1**). These eight OTUs were among the twelve most abundant with six ranked in the top six most abundant and two others ranked 11th and 12th (OTUs 586 and 671).

Our method explored a large number of potential element-frequency relationships by focusing on extreme slopes as a metric to identify those that might merit further investigation. Though no parametric statistical assumptions are needed,

we focused only on linear associations, after accounting for time and random effects as appropriate for the design. Admittedly, complex associations may be missed using this approach. Often, associations between OTU frequency and element concentrations were noisy at lower concentrations and began to take on a pattern at higher concentrations that may approach biologically effective concentrations (e.g., concentrations of heavy metals that may act as selective environmental filters). Unfortunately, many elements had very few observations at higher concentrations complicating clear assessment of any underlying patterns. This is highlighted by the asymptotic likelihood ratio test p-values that test the significance of the element in each of the models with extreme slopes and range from 0.00003

(OTU 46 and element Sr) to 0.75 (OTU 425 and element Cu). The relatively low sample sizes combined with high variance in observations and complex relationships mentioned earlier contribute to the difficulty in confirming any measure of statistical significance. We argue that, for the data analyzed here, the extreme slopes technique is a useful method to explore a large number of models and to highlight patterns warranting further investigation.

Evaluation of the models with extreme slopes brings four interesting issues to light. First, selection of the influential elements is difficult. Most of the top 50 OTUs did not have any elements that very strongly affected their frequency in our analyses. This may indicate that factors other than those considered here may control their distribution and occurrence. It is also of note that these analyses can be influenced strongly by the distribution of the observed frequencies: some of the top 50 OTUs had zero values in frequency for several samples, whereas other OTUs were more evenly distributed in their frequencies across samples (**Fig. 1**). Our analyses included the possibility of identifying those models that were affected by numerous low frequencies and a few high frequency observations. We argue that the few higher frequencies are most likely indicative of those elements that also have extreme concentrations in the same samples; we did not want to miss such findings. Second, no one element controls the occurrence of all or even majority, of the OTUs, but the OTUs appear to respond positively or negatively to different drivers. This is strongly visible even among the eight that remained through our rigorous evaluation of a vast number of models. This can be interpreted in the context of a niche. Foliage represents a complex abiotic physicochemical habitat within which organisms are sorted based by stochastic arrival parameters, but also by either environmental tolerances or nutritional preferences. Those fungi best able to colonize and invade the available substrate under any given combination of the complex physical and chemical environmental matrix will persist and be detected most frequently. Thirdly, even for one OTU, many elements may have strong and occasionally opposing effects. For example, for

Table 1. Slopes identified as extreme in our analyses

Element	OTU 425	OTU 46	OTU 686	OTU 520	OTU 567	OTU 313	OTU 586	OTU 671	OTU 555	OTU 607
B	+*	+*					+*			
Ba						–	–			
Ca			–*			(–)*	–*	(+)*	+**	
Cd	+					+	(+)			
Ce	+	(+)						–		
Co	+**						–*			
Cr							–*			
Cu	+*	–**					–*			
Fe							–*			
Hg	+**					–*				
K	(–)		+	–		+		(–)		(+)
Li	(+)*						(+)*	–*		
Mn		+*								
Mo							–*			
N	–*					+*	(+)*			
Na	+									
Ni							–*			
P	–*									(+)*
Pb		+**					–*			
Rb		+**	+*	–*			–*			
S	(–)*	+*				+*	+*			
Sc	–					(–)	–			
Se		–								
Sn	–					–	(–)			
Sr		+*								
Y		+*			–*	+*	(+)*			
Zn	(–)*	+*				–**				(+)*

Positive slopes are indicated by +, negative by –. Parentheses indicate where a statistically significant (α = 0.05) interaction was observed (likelihood ratio test). Extreme slopes with observations with high leverage are identified by an asterisk (*) and those where omission of high-leverage observations lead to a non-extreme slopes are identified by two asterisks (**). Note that eight of the ten OTUs in the table had an extreme slope with at least one element concentration after accounting for high leverage and interactions in the model.

OTU425, B, Cd, Ce, Cu, Na, had positive effects, whereas N, P, Sc had negative effects (**Table 1**). Interpreting observations of this kind is complicated and often unattainable. However, it is good to bear in mind that foliage represents a complex abiotic environment within which organismal interactions take place. It is unlikely that the physicochemical environment alone determines community composition, but the combination of the abiotic environment and biological interactions defines the observed communities. Furthermore, in absence of strong correlations among the elements, each of the elements varies independently depending on the edaphic conditions and plant physiological activities leading likely to a

somewhat stochastic multidimensional abiotic data matrix. Finally, not all OTUs seem responsive to the same elements. While our analyses were able to pinpoint some potential candidates for the most influential elements for the occurrence of some OTUs, the mechanistic explanations require rigorous experimentation.

More important than the identification of the potentially influential elements for some of the analyzed OTUs is their biological relevance for organismal occurrence. We highlight two specific groups here—important macronutrients N and P as well as heavy metals Cd, Hg, Pb and Zn. OTUs 313 and 425—both likely leaf-inhabiting pathogens or parasites— were strongly responsive to N. While the

former showed a strong positive response to increasing N concentrations, the latter declined in response to both N and P. Because many fungal communities have been shown sensitive to nutrient enrichment,[18,19] it is tempting to speculate on species replacement or on tolerance to nutrient enrichment as a result of changes in the abiotic chemical environment. However, one must exercise caution: as we point out above, a number of other alternative factors come to play when a correlative relationship like this is considered across two discrete and complex datasets. Several heavy metal concentrations also showed either positive or negative associations with the fungal OTU frequencies. To exemplify, the frequencies of OTUs 313 and 425 were

positively associated with the concentrations of Cd and OTU 46 was positively associated with Zn, whereas OTUs 313 and 586 were negatively associated Hg and Pb concentrations, respectively. Does this mean that these species differ in their sensitivities to these particular heavy metals? Not necessarily, but these observational data provide a starting point for more explicit hypothesis-driven experiments that allow for specific elucidation of the fungal responses to these elements and may guide future experimentation.

We conducted a high-dimensional exploratory analysis to evaluate potential effects of element concentration on OTU frequencies. Using a repeated measures mixed effects model, we were able to compile a brief list of chemical elements with the most likely (based on these data) strongest effects on the abundances of the dominant components of the phyllosphere-associated fungal communities. Complicating the use of usual methods of statistical inference (i.e., use of p-values) was the sparseness in the occurrence of many OTUs across samples and outlying observations in the concentration of some elements. We chose the extreme slopes approach that allowed ranking associations between OTU frequency and element concentration with no assumptions regarding normality or equivariance that may be violated using traditional tools of inference (e.g., Analysis of Variance). Still,

some of the observed associations may have been affected by extreme leverage points (outliers in the explanatory variables) and these were accounted for in the present analyses by model re-evaluation after omission of the high-leverage observations. While our analyses identified a number of biologically meaningful associations between chemical elements and molecular OTUs, rigorous experimentation is mandatory to establish causative relationships.

References

1. Arnold A, Lutzoni F. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? Ecology 2007; 88:541-9.
2. Arnold A, Maynard Z, Gilbert G, Coley P, Kursar T. Are tropical fungal endophytes hyperdiverse? Ecol Lett 2000; 3:267-74.
3. Saikkonen K, Faeth S, Helander M, Sullivan T. Fungal endophytes: A continuum of interactions with host plants. Annu Rev Ecol Syst 1998; 29:319-43.
4. Martiny J, Bohannan B, Brown J, Colwell R, Fuhrman J, Green J. Microbial biogeography: Putting microorganisms on the map. Nat Rev Microbiol 2006; 4:102-12.
5. Helander M, Ahlholm J, Sieber T, Hinneri S, Saikkonen K. Fragmented environment affects birch leaf endophytes. New Phytol 2007; 175:547-53.
6. Elamo P, Helander M, Saloniemi I, Neuvonen S. Birch family and environmental conditions affect endophytic fungi in leaves. Oecologia 1999; 118:151-6.
7. Arnold A, Herre E. Canopy cover and leaf age affect colonization by tropical fungal endophytes: Ecological pattern and process in *Theobroma cacao* (Malvaceae). Mycologia 2003; 95:388-98.
8. Valkama E, Koricheva J, Salminen J, Helander M, Saloniemi I, Saikkonen K. Leaf surface traits: Overlooked determinants of birch resistance to herbivores and foliar micro-fungi? Trees 2005; 19:191-7.
9. Helander M. Responses of pine needle endophytes to air pollution. New Phytol 1995; 131:223-9.
10. Cannon P, Simmons C. Diversity and host preference of leaf endophytic fungi in the Iwokrama forest reserve, Guyana. Mycologia 2002; 94:210-20.
11. Arnold A, Mejia L, Kyllo D, Rojas E, Maynard Z, Robbins N. Fungal endophytes limit pathogen damage in a tropical tree. Proc Natl Acad Sci USA 2003; 100:15649-54.
12. Pan J, May G. Fungal-fungal associations affect the assembly of endophyte communities in maize (*Zea mays*). Microb Ecol 2009; 58:668-78.
13. Rakotoniriana E, Munaut F, Decock C, Randriamampionona D, Andriambololoniaina M, Rakotomalala T. Endophytic fungi from leaves of *Centella asiatica*: Occurrence and potential interactions within leaves. Antonie Van Leeuwenhoek 2008; 93:27-36.
14. Seghers D, Wittebolle L, Top E, Verstraete W, Siciliano S. Impact of agricultural practices on the *Zea mays* L. endophytic community. Appl Environ Microbiol 2004; 70:1475-82.
15. Colpaert J, Muller L, Lambaerts M, Adriaensen K, Vangronsveld J. Evolutionary adaptation to zn toxicity in populations of suilloid fungi. New Phytol 2004; 162:549-59.
16. Jumpponen A, Jones K. Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. New Phytol 2010; 186:496-513.
17. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied Linear Statistical Models. Fifth ed. New York: McGraw-Hill/Irwin 2005.
18. Wallenda T, Kottke I. Nitrogen deposition and ecto-mycorrhizas. New Phytologist 1998; 139:169-87.
19. Treseder K. A meta-analysis of mycorrhizal responses to nitrogen, phosphorus and atmospheric CO_2 in field studies. New Phytol 2004; 164:347-55.
20. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000; 7:203-14.