Examining the roles of the Douglas-fir (*Pseudotsuga menziesii)* microbiome in host stress response

A. Introduction

The plant microbiome may act as a fundamental extension of the host plant genome[1]. In the current trajectory of climate change, manipulation of plant microbiomes may therefore represent one of few viable methods for increasing plant drought-stress and pathogen-resistance in plants necessary for human use[2–5]. Despite the urgency, development of techniques to quantify metabolic activity of the plant-associated microbiome has advanced slowly. This is in part due to the diffuse nature and diverse origins of microbiome RNA - any search for signals of RNA expression from the transcriptome of the microbiome that also involves collection of host tissue is presumed to be swamped by host RNA using existing sequencing technologies.

Here we propose an exploration of the Douglas-fir (*Pseudotsuga menziesii*) hologenome, by monitoring changes of expression in both the Douglas-fir transcriptome and its microbiome transcriptome, in response to an (1) environmental gradient and to (2) an environmental change. To monitor changes in enrichment of molecular pathways of the *microbiome* of Douglas-fir, we propose to augment standard RNAseq analysis pipelines with an “ecological” bioinformatics approach.

B. Objectives, methods, and significance

Hypothesis: The metabolisms of plant host and its microbiome are responsive to environmental change and interact between themselves. These interactions in the context of environmental change can be antagonistic (disease and immune response), but also may be mutualistic/complementary[1].

To test these studies of the Douglas-fir microbiome transcriptome are proposed, one survey and one experiment.

Methods:

Study 1: *Changes in host and microbiome gene expression with height and exposure in the canopy of old growth Douglas-fir forest.*

6 trees will be climbed, from several locations in the HJ Andrews forest (andrewsforest.oregonstate.edu). Needles will be sampled every ten meters, as high into the canopy as can be safely climbed. This interval of sampling may vary with the results of our power analysis, see below. The entire leaf microbiome will be sampled including both endophytic and epiphytic microbes. As such, whole needles will be placed into 1.5 mL centrifuge tubes, then into a 1 liter dewars with liquid nitrogen attached to climbers. The smaller dewars will be lowered with drop line occasionally, to

transfer needle samples into a 50L dewars at base of operations on the ground, and to refill liquid nitrogen reservoirs. RNA will be isolated, ribosomal RNA will be removed, cDNA libraries constructed and sequenced via illumina Hi-seq system. Bioinformatic pipelines are discussed below.

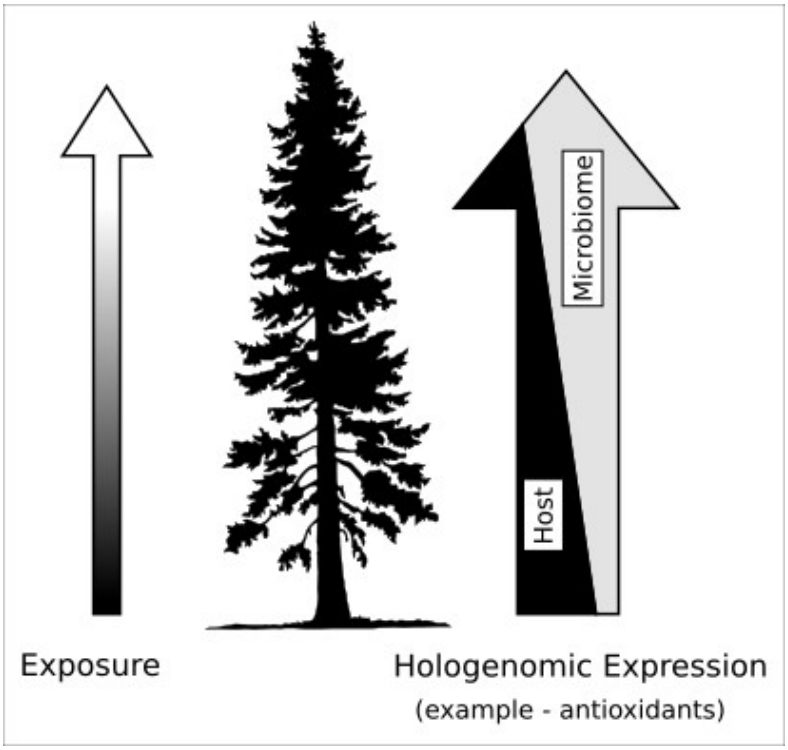


Figure 1 – we predict that in some metabolic pathways, microbiomes may share some mutually useful products such as antioxidants with host, and that the microbiome and host metabolisms can be complementary. Environmental change may influence what products are shared. Here antioxidant and UV-absorbent products such as pigments may be produced by the microbiome with height in the canopy, reducing the need for the host tree to produce non-chlorophyll pigments.

Predictions, study 1:

The combined transcriptome of Douglas-fir and its microbiome will show increased expression in antioxidant and pigmentation pathways, such as melanin[6] and anthocyanin production[7], with increasing height/exposure in the canopy and with increased exposure resulting from effects of deforestation.

We predict a mutualistic, inverse relationship of microbiome vs. host antioxidant and pigementation pathways – i. e. if the microbiome seems to be producing a mutually useful metabolite, host expression for that product or similar products in the host tree may be allowed to remain at relatively lower rates. Thus, for transcriptomes that contain microbial UV-, and drought/oxidative-stress-tolerance related reads such as high melanin pigmentation, we predict will show less upregulation of host tree production of analogous antioxidants or pigments (anthocyanins and carotenoids).

Study 2) *Changes in host and microbiome transcriptome of newly exposed remnant or edge trees.*

Presence of forest and stage of forest succession has been shown to affect fungal microbiome community assembly[8,9]. We will investigate changes in host and microbiome RNA response to the

increased exposure and isolation resulting from timber harvest. In cooperation with Bauman Tree Farm (http://www.foreststodayandforever.org/pages/bauman.html) and associates, we will sample microbiome leave trees from cutblocks (https://oregonforests.org/Harvest_Regulations). Trees marked to be left will be sampled before and following harvest events for two years. Both RNA and DNA will be isolated from leaf samples, to examine both changes in expression and shifts in microbial community composition. DNA will be purified from leaves, selectively PCR-amplified using fungal and bacterial-specific primers , and prepped for sequencing with illumina MiSeq sequencer system. RNA related laboratory preparation of samples will be as above, and bioinformatic pipelines are explained below.

Predictions, study 2:
Using a classical insular ('island') biographical theoretical framework[10] and following observations from Arnold et al.[8] and Chaverri and Vilchez[9], we predict a shift in species composition of the microbiome following isolation of trees. This shift will be temporal – in the first year following harvest, forest associated microbes will be lost, microbial load diversity will be generally reduced (fig 2d). In the second season following harvest, microbial species richness may rebound, but community composition will be fundamentally changed (2e).

We also predict upregulation of antioxidant pathways, and other pathways related to drought and UV stress tolerance, in both microbiome and host transcriptome, with the increased exposure from deforestation.

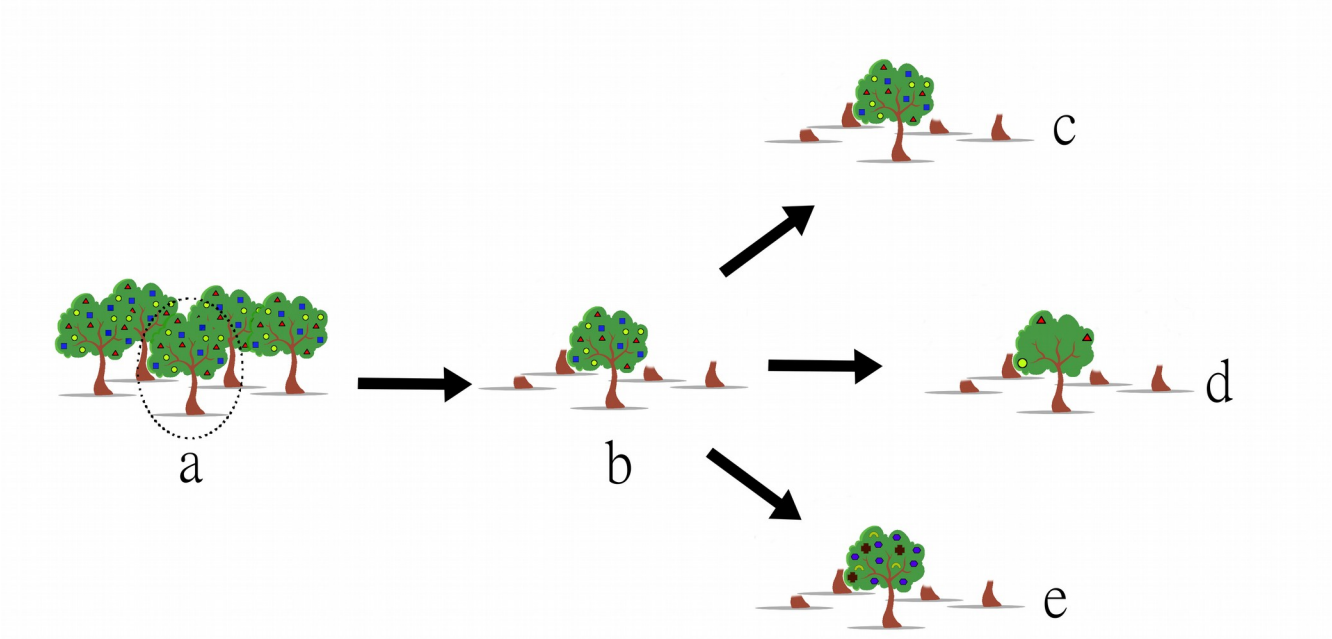


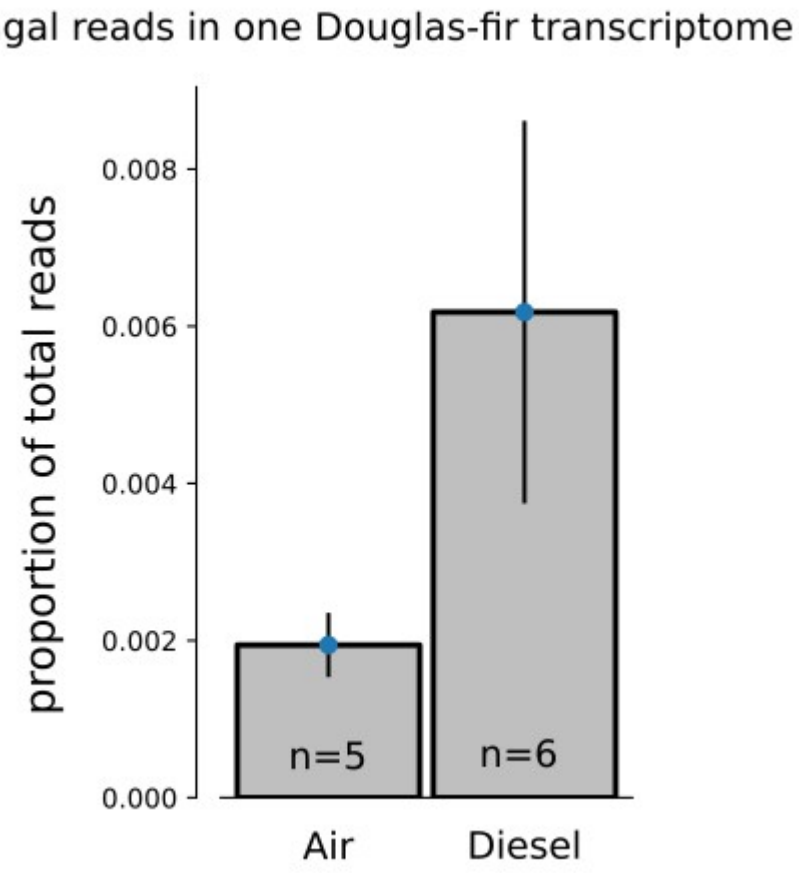
Figure 2 – We propose to sample the microbiome DNA and the transcriptome of Douglas-fir leave trees and their microbes. Sampling will occur prior (a) to the harvest (b), and for two years following. Possible outcomes include no change in microbial community composition (c), reduced diversity of microbes (d), or unchanged/increased species richness, but with a significant shift in microbiome community composition (e).

Bioinformatics:

Fold changes in Douglas-fir gene expression and enrichment of pathways will be examined for response to environmental gradients (increasing height) and change (isolation following removal of surrounding forest) using standard RNAseq bioinformatic pipelines[11–15]

RNAseq studies of large organisms very often leave 15-20% of total reads unaligned to the genome of the targeted organism. These reads are typically discarded, as RNA expression in the large organism is usually the focus of RNAseq studies. These discarded reads can originate from sequencer error and laboratory contamination, but many also originate from the transcriptome of the microbiome. Through careful curation of these "trash" read libraries, without minimum abundance cutoffs, we can draw an outline of metabolic activity in the microbiome.

In a first exploration of existing data, we aligned 11 transcriptome samples from a publicly available dataset[16] to host *P. menziesii* genome[17]. For each of these samples, approximately 15% of reads are unalignable to the hostX We blasted the reads not aligned to host genome against a database of ~10,000 fungal genes randomly selected from NCBI genbank. The average proportion of total reads (including those aligned to host) with high confidence matches to fungal genes in this database was 0.4 %. This amounted to average of ~50,000 fungal reads per sample, with each sample containing an average of ~11,700,000 reads before alignment to host . If this rate applies to other sequencing events, in a single lane of illumina HiSeq 3000/4000 with 50bp per read, we may expect to see as many as 2,700,000 fungal reads. Some amount of these reads must be due to laboratory contamination, but presumably the majority represent coocurring fungi, as handling of samples for any other reason than RNA isolation is necessarily minimal, due to rapid degradation of RNA. Prokaryotic reads were not examined but are of equal interest, and could be estimated similarly, provided RNAseq libraries were not created via poly-A selection methods. It is worth noting that this amount of read coverage is comparable to results from Roche 454 Pyrosequencing platforms, which have been used successfully used for RNAseq studies[18,19].



Proportion of total reads in samples from a publicly available dataset[16]. Douglas-fir seedlings were subjected to various environmental pollutants. Here we compared a subset of pure-air treated samples with a subset of diesel treated samples. Average number of raw reads per sample was ~11,000,000.

As an initial step, power analysis will be conducted. This will involve examining additional existing plant transcriptomes from public databases in the same manner as above to more confidently estimate expected ratio of microbiome reads. Following this, RNAseq power analysis will be conducted[20,21]. However, the sparse gene-expression matrices generated from the low-coverage of the microbiome in RNAseq libraries may be underpowered for detecting fold changes in individual genes and thus also for standard methods of detecting enriched molecular pathways.

Thus as a second stage of analysis, we propose attempting both traditional RNAseq bioinformatics software pipelines (differential expression tests, annotation, and pathways enrichment tests), but also some non-standard approaches to characterizing changes in expression, borrowing from numerical ecological methods. While the shallow sampling of the microbiome in host tissue presents obstacles, it also presents opportunity: the smaller number of microbiome reads allows more computational resources and time to be allocated to each read.

Since the diverse origins of microbiome reads may cause difficulties in assigning gene ontologies, additional care must be placed on identifying the taxonomic origin of reads. Following de novo assembly of reads, a naive bayesian classifier[22] will be trained on SNP calls and other small polymorphisms from aligned homologous sequences from genomic databases (JGI mycocosm, NCBI Microbial Genomes. EnsemblBacteria, etc) to call most-likely taxonomic origins. Once uncertainty

from taxonomic origin is reduced, gene annotation may be more accurately targeted, using standard methods for non-model annotation[15].

After removal of host reads and categorization of reads into functional groups, we are presented with a situation often faced by ecologists – comparison of rows of a sparse matrix. Where data are too sparse for standard pathways enrichment tests, we will attempt to use distance-based ordinations compare visualize (dis)similarity of samples, from both a taxonomic and functional standpoint. PERMANOVA[23] or other dissimilarity pseudo-linear models will be used to statistically test dissimilarity of samples by height or pre/post isolation treatment. Distance metrics will be further researched to find best practices for this somewhat novel application. Hypothesis testing is also possible, by testing groupings of genes into pathways of interest (antioxidant or pigmentation synthesis pathways) as explanatory factors in a PERMANOVA model).

Intellectual significance:

This study will enhance understanding of response by both host-plants and their microbiomes in response to environmental changes that are increasingly common – increased exposure, drought, and variability in climatic conditions, and isolation by habitat fragmentation. This study will also result in increased understanding of appropriate statistical methods and statistical power available for observing microbiome metabolic activity directly from host tissue with current sequencing technologies. Most fundamentally, this study fundamentally questions the role of the microbiome – does the microbiome complement the host genome in times of stress? Are metabolic pathways shared in some meaningful way between plant hosts and microbiome?

C. Training objectives and plan for achieving them (these may include scientific as well as other career preparation activities)

Bayesian methods – Applicant has an interest in understanding Bayesian statistical methods more completely, and several aspects of the proposed project would benefit greatly from their application. In general, the move to bayesian methods is one of the current paradigms shifts in ecology, and the applicant will benefit from mastering bayesian statistical analysis. Several classes are offered at OSU.

Cython and C++: The volume of reads and computational cost of above methods will require optimization for speed. However, most of necessary bioinformatics tools are implemented in the python-based SciPy[24] software ecosystem. The applicant will seek therefore training in Cython and (and C++ where necessary) to find lower level solutions to computational problems where ever possible.

Tree climbing: The applicant has a continued interest in canopy research, and will engage PTCI to learn more about single- and double-rope techniques for accessing the canopy for sampling.

D) An explanation of how the fellowship activities will enhance your career development and future research directions as well as describing how this research differs from your dissertation research, thus providing you an opportunity to broaden your scientific horizon.

The applicant's dissertation work focused on spatial patterns in forest fungal endophytes, a subset of the plant microbiome. Ecological studies of endophytes were conducted using traditional culture-and-barcode techniques, and ITS-marker[25] metabarcoding using Illumina MiSeq sequencer as sampling methods for changes or differences in fungal community. The current proposal differs in its experimental methods (RNAseq), and basic ecological questions (functional complementary nature of the microbiome, and response by the hologenome to environmental gradients and change), and in bioinformatic methods.

This research will enhance my career development through acquisition of new in skillsets (see section D, above). Also enhancing my career are collaborations with the canopy research community of the HJ Andrews experimental forest, and the small-scale tree w industry of western Oregon, institutions with whom I look forward to working in the future.

E) A justification of the choice of sponsoring scientist(s) and host institution(s);

Dr. Posy Busby is a logical choice as adviser for the applicant: her interest in fungal endophytes[26,27], microbial ecology and biogeography[27] in forests ecosystems. Her work contains both theoretical and practical aspects[5,26,28,29], with applications to current challenges facing humans and their plant partners. This research with an eye to the future is one of her strongest assets as a postdoctoral advisor – she has open mind for new biological theory and for attempting new techniques in novel field settings such as proposed above. This is tempered with by her extensive hands-on ecological and mycological experience.

Dr. Busby, her doctoral student Kyle Gervers and I have an existing collaboration examining changes in microbial community composition with height in the old growth Douglas-fir canopy, using ITS and 16s meta-barcoding. Samples have been submitting for sequencing. This collaboration is so far unpaid and research costs were not funded by a major grant. Some of the RNA-based studies proposed here are an expansion of this existing collaboration.

F) A timetable with yearly goals with benchmarks for major anticipated outcomes

Year 1

Summer/fall:

- climb trees at HJA, process RNA, sequence.
- collect needles from cut blocks leave trees, pre-harvest

Winter/spring

- isolate RNA and DNA from samples, prepare for sequencing.
- Recieve sequence data, begin characterization of microbial community from ITS-barcoded (DNA) libraries.
- Characterize differential expression in height study.
- Begin development of bioinformatic pipeline for microbiome pathway analysis

Year 2

Summer/fall:

- continued analysis and writing for canopy height RNAseq study
- continue/finish development of bioinformatic pipeline for microbiome pathway analysis
- collect needles from cut block leave trees, post-harvest.

Winter/spring

- isolate RNA and DNA from samples, prepare for sequencing.
- Recieve sequence data, begin analysis

Year 3

Summer/fall

- second collection of needles from cut block leave trees, post-harvest.
- isolate RNA and DNA from samples, prepare for sequencing.

Winter/spring

- Recieve sequence data, begin analysis on these
- writing: 2-3 manuscripts anticipated: 2 ecological papers, one from each study, and 1 possible methods paper describing microbiome RNAseq methods.

G) A separate section within the narrative that describes in detail the broader impacts of the proposed activities

Manipulation of the microbiome as a tool for mitigating climate change to protect human food systems is mentioned above. This project will also increase understanding of plant response to environmental change, and habitat fragmentation, both relevant in the current regime of climate change and land cover changes throughout the world. The targeted host organism, Douglas-fir, is a tree species of great economic importance to the north American Pacific Northwest[30], which is under increasing pathogen load, in part due to climate change[31,32]

The importance of microbiomes to health extends to all multicellular life. Despite the oft-lauded importance of the microbiome, every study of gene-expression in a multicellular host via RNAseq methods discard millions of reads that may originate from the microbiome. This amounts to a regular, systemic loss of information occurring throughout all the life sciences currently. We hope to be part of the effort to recover this signal of the microbiome that may be available in most high-throughput PCR studies.

Students, both graduate and undergraduate, will be mentored in all aspects of the project – field, lab, and bioinformatic. This last portion of biological projects, bioinformatics and statistical analysis, is very often overlooked as a mentoring opportunity. The bridge from classroom-to-lab in statistics and computational biology is often a pain point for students of all levels. Though this is unfortunate, this

can also be an equalizing opportunity for students with a strong desire to do research, even if they are not succeeding entirely in a classroom setting, in the applicant's experience. The applicant will actively seek to train interested students, including students of underprivileged backgrounds, in bioinformatic and statistical methods that are sometimes not avaialable or are poorly explained in statistical classes.

1. Rosenberg, E. & Zilber-Rosenberg, I. Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. *mBio* **7**, e01395-15 (2016).
2. Naylor, D. & Coleman-Derr, D. Drought Stress and Root-Associated Bacterial Communities. *Front. Plant Sci.* **8**, (2018).
3. Vurukonda, S. S. K. P., Vardharajula, S., Shrivastava, M. & SkZ, A. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. *Microbiol. Res.* **184**, 13–24 (2016).
4. Woodward, C., Hansen, L., Beckwith, F., Redman, R. S. & Rodriguez, R. J. Symbiogenics: An Epigenetic Approach to Mitigating Impacts of Climate Change on Plants. *HortScience* **47**, 699–703 (2012).
5. Busby, P. E. *et al.* Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLOS Biol.* **15**, e2001793 (2017).
6. Shcherba, V. V., Babitskaya, V. G., Kurchenko, V. P., Ikonnikova, N. V. & Kukulyanskaya, T. A. Antioxidant properties of fungal melanin pigments. *Appl. Biochem. Microbiol.* **36**, 491–495 (2000).
7. Steyn, W. J., Wand, S. J. E., Holcroft, D. M. & Jacobs, G. Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytol.* **155**, 349–361 (2002).
8. Arnold, A. E. & Herre, E. A. Canopy cover and leaf age affect colonization by tropical fungal endophytes: Ecological pattern and process in Theobroma cacao (Malvaceae). *Mycologia* **95**, 388–398 (2003).
9. Chaverri, P. & Vilchez, B. Hypocrealean (Hypocreales, Ascomycota) Fungal Diversity in Different Stages of Tropical Forest Succession in Costa Rica 1. *Biotropica* **38**, 531–543 (2006).
10. MacArthur, R. H. & Wilson, E. O. *The theory of island biogeography*. **1**, (Princeton university press, 2001).
11. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).
12. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).

13. Klaus, B. Differential expression analysis of RNA–Seq data using DESeq2. 24

14. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).

15. Conesa, A. & Götz, S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics* (2008). doi:10.1155/2008/619832

16. Reichman, J. R. *et al.* Douglas-Fir (Pseudotsuga menziesii (Mirb.) Franco) Transcriptome Profile Changes Induced by Diesel Emissions Generated with CeO2 Nanoparticle Fuel Borne Catalyst. *Environ. Sci. Technol.* (2018). doi:10.1021/acs.est.8b02169

17. Neale, D. B. *et al.* The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. *G3 Genes Genomes Genet.* g3.300078.2017 (2017). doi:10.1534/g3.117.300078

18. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

19. Hook, S. E. *et al.* 454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: Transcriptome assembly and toxicant induced changes. *Aquat. Toxicol.* **153**, 73–88 (2014).

20. Zhao, S., Li, C.-I., Guo, Y., Sheng, Q. & Shyr, Y. RnaSeqSampleSize: real data based sample size estimation for RNA sequencing. *BMC Bioinformatics* **19**, 191 (2018).

21. Wu, H., Wang, C. & Wu, Z. PROPER: comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics* **31**, 233–241 (2015).

22. Rosen, G., Garbarine, E., Caseiro, D., Polikar, R. & Sokhansanj, B. Metagenome Fragment Classification Using N-Mer Frequency Profiles. *Advances in Bioinformatics* (2008). doi:10.1155/2008/205969

23. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). in *Wiley StatsRef: Statistics Reference Online* 1–15 (American Cancer Society, 2017). doi:10.1002/9781118445112.stat07841

24. Millman, K. J. & Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **13**, 9–12 (2011).

25. Bellemain, E. *et al.* ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189 (2010).

26. Busby, P. E., Ridout, M. & Newcombe, G. Fungal endophytes: modifiers of plant disease. *Plant Mol. Biol.* **90**, 645–655 (2016).

27. Busby, P. E., Peay, K. G. & Newcombe, G. Common foliar fungi of Populus trichocarpa modify Melampsora rust disease severity. *New Phytol.* **209**, 1681–1692 (2016).

28. Newcombe, G., Harding, A., Ridout, M. & Busby, P. E. A Hypothetical Bottleneck in the Plant Microbiome. *Front. Microbiol.* **9**, (2018).

29. Busby, P. E., Newcombe, G., Dirzo, R. & Whitham, T. G. Differentiating genetic and environmental drivers of plant–pathogen community interactions. *J. Ecol.* **102**, 1300–1309 (2014).

30. Talbert, C. & Marshall, D. Plantation Productivity in the Douglas-Fir Region Under Intensive Silvicultural Practices: Results from Research and Operations. *J. For.* **103**, 65–70 (2005).

31. Agne, M. C. *et al.* Interactions of predominant insects and diseases with climate change in Douglas-fir forests of western Oregon and Washington, U.S.A. *For. Ecol. Manag.* **409**, 317–332 (2018).

32. Stone, J. K., Coop, L. B. & Manter, D. K. Predicting effects of climate change on Swiss needle cast disease severity in Pacific Northwest forests. *Can. J. Plant Pathol.* **30**, 169–176 (2008).