Supplementary Appendix for Adams *et al*., "Mountain Pine Beetles Colonizing Historical and Naïve Host Trees are Associated with a Bacterial Community Highly Enriched in Genes Contributing to Terpene Metabolism"

Table S1. Summary and phylogenetic assignment of the twelve DGGE bands sequenced.

Phyla/Class	OTU. no.	Accession no./bp	Closest match	% Identity to match
Actinobacteria	27	JF810915 / 342	*Micrococcus endophyticus*	100
β-Proteobacteria	15	JF810916 / 345	*Massilia aurea*	98
γ-Proteobacteria	2	JF810917 / 213	*Pseudomonas viridiflava*	99
	6	JF810918 / 331	*Serratia fonticola*	100
	7	JF810919 / 298	*Acinetobacter johnsonii*	97
	16	JF810920 / 266	*Serratia fonticola*	97
	17	JF810921 / 327	*Pseudomonas viridiflava*	99
	18	JF810922 / 331	*Pantoea agglomerans*	99
	23	JF810923 / 331	*Pseudoxanthomonas spadix*	100
	24	JF810924 / 333	*Serratia nematodiphila*	99
	25	JF810925 / 333	*Serratia nematodiphila*	99
	29	JF810926 / 327	*Serratia fonticola*	100

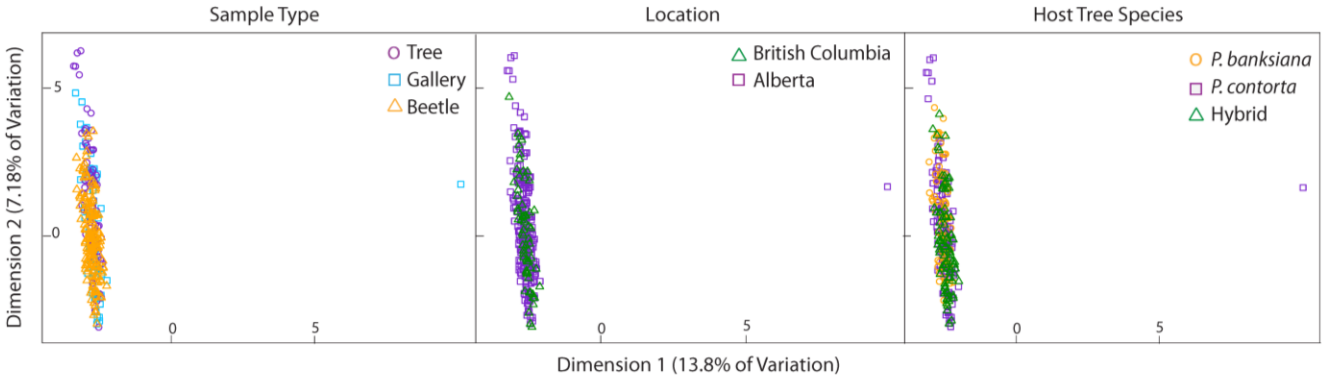

Figure S1. Principle Component Analysis of DGGE data (A) Tree vs Beetle; (B) Hybrid pine vs Lodgepole pine (C) Alberta (AB) vs British Colombia (BC).

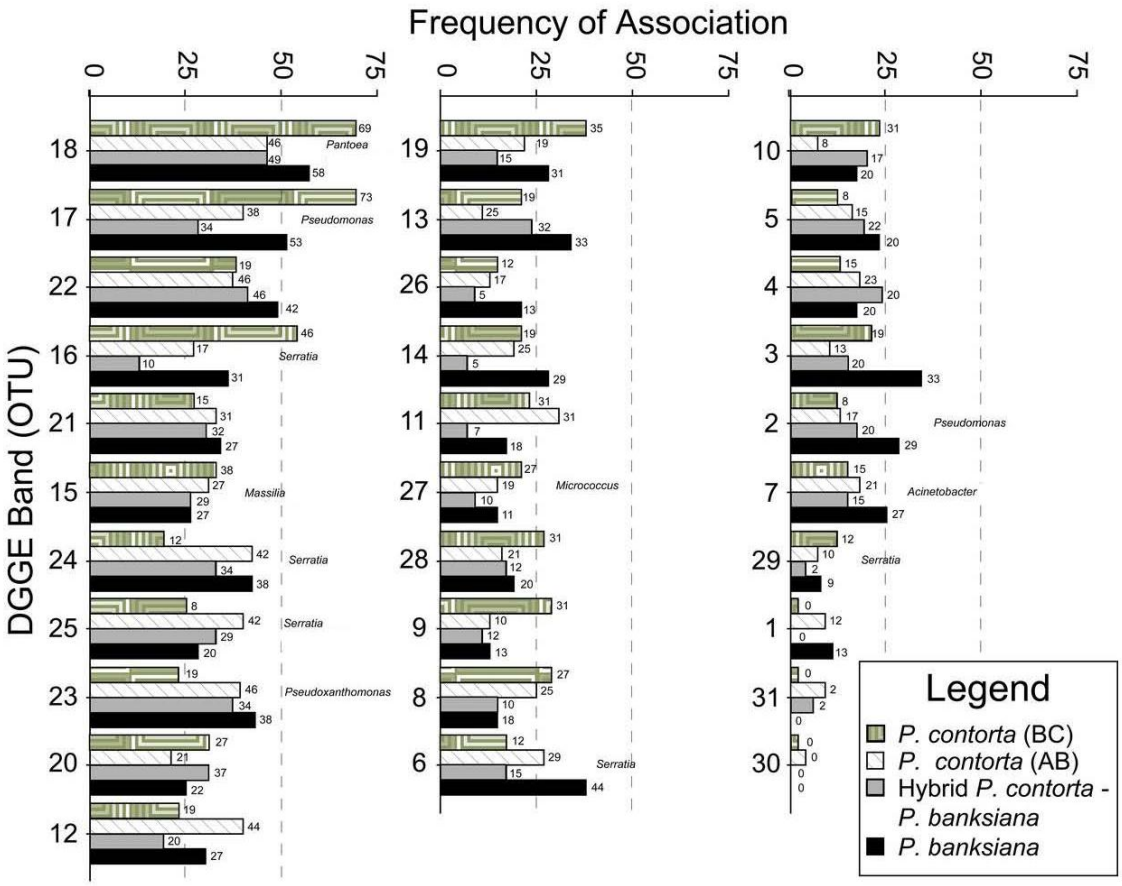

Figure S2. Frequency of association for DGGE bands across all samples from Alberta (AB) and British Colombia (BC). An OTU analysis was performed to identify the identity of specific sequences as shown at the genus level.

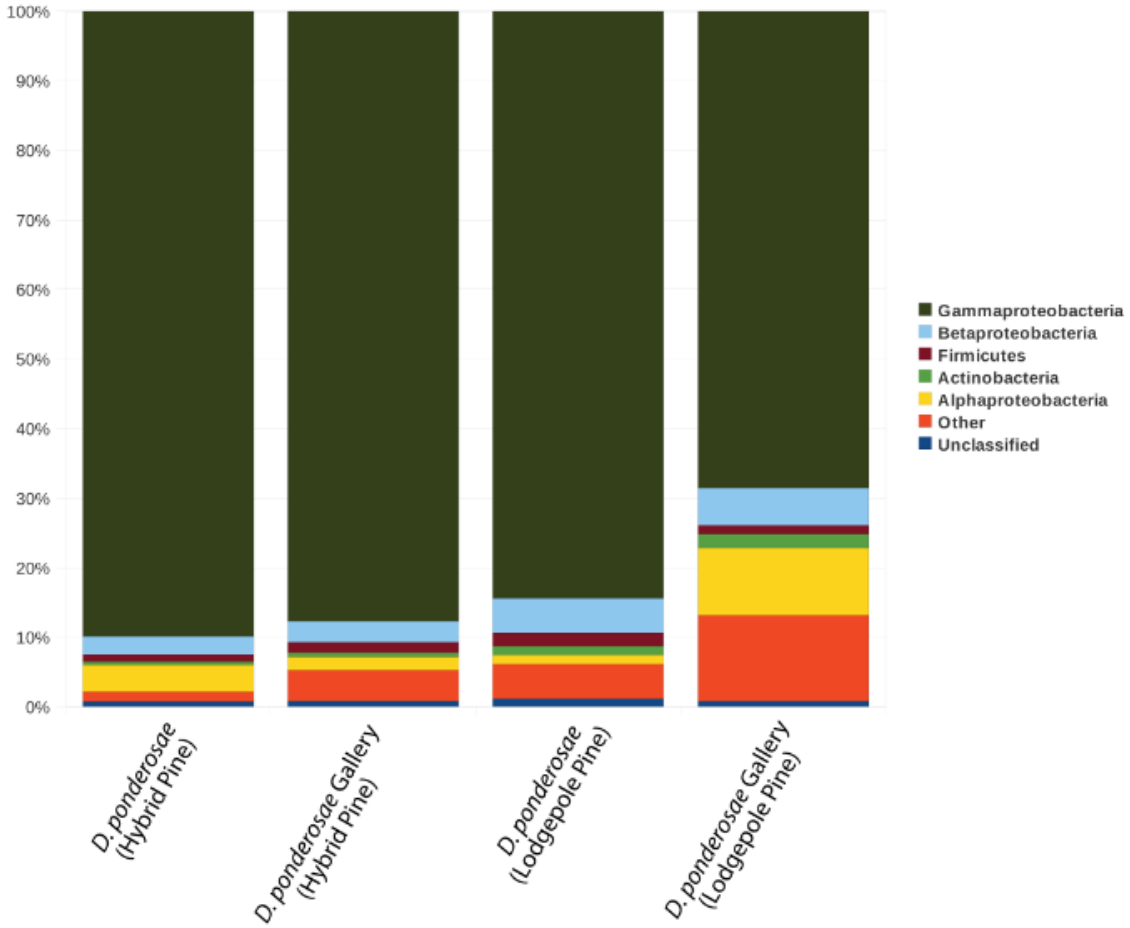

Figure S3. Phylogenetic binning analysis of four beetle-associated community metagenomes. Only those phyla/groups with greater than 1% representation are shown.

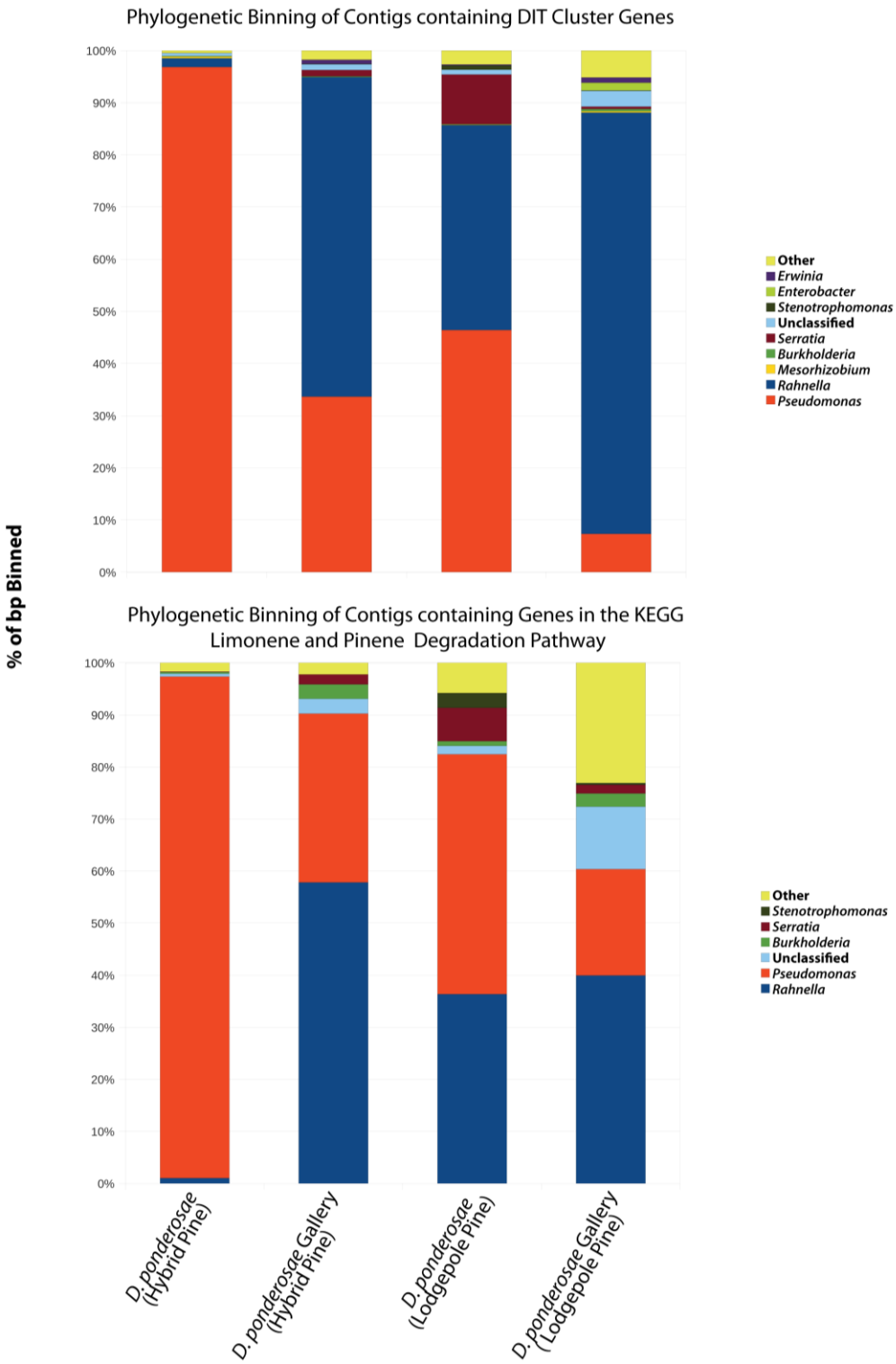

Figure S4. Phylogenetic binning analysis of monoterpene and diterpene genes identified in four beetle-associated metagenomes. Only those genera with greater than 1% representation are shown.

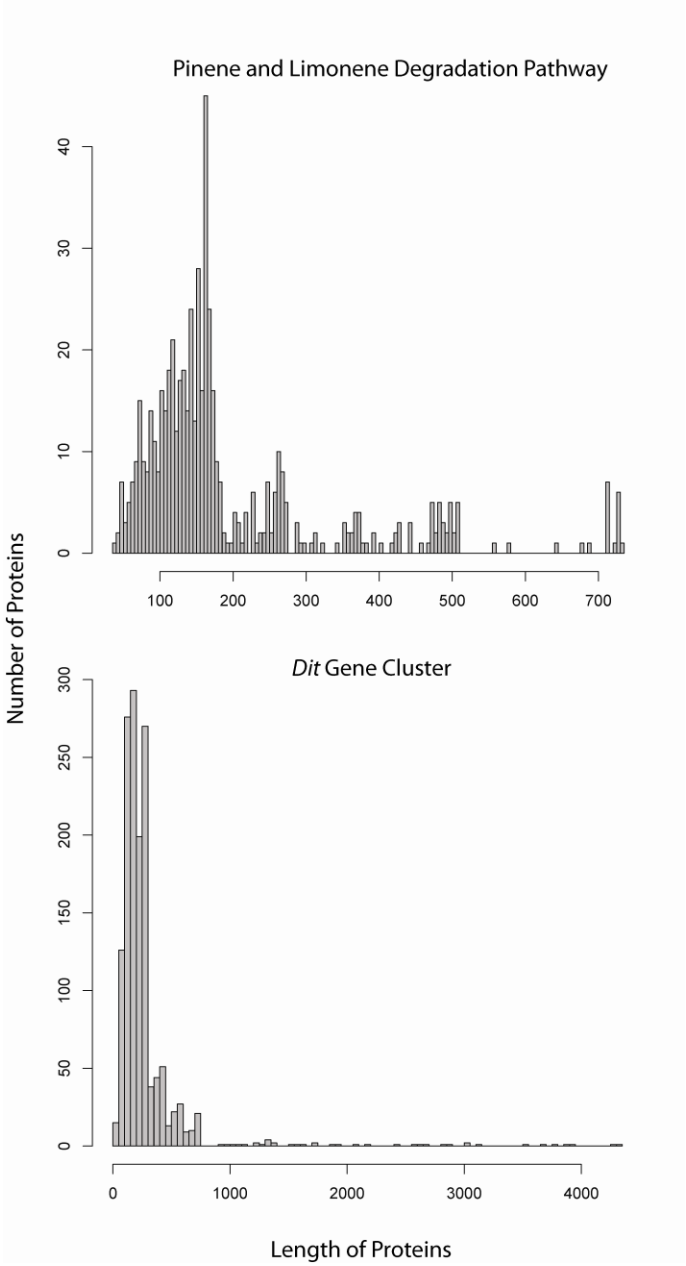

Figure S5. Length in amino acids of the proteins in the four mountain pine beetle-associated community metagenomes that are annotated as belonging to the KEGG Limonene and Pinene Degradation Pathway or have homology to proteins in the *dit* gene cluster.

SI Text

Materials and Methods

DGGE PCR Processing and Analysis. Individual adult beetles were surface washed in the laboratory in 70% ethanol for 1 minute, and ground separately in 1 ml PBS. Phloem from gallery and unattacked tree samples were rinsed in PBS, and ground separately in 1 ml PBS. Total DNA was extracted as described in the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA). PCR amplification and DGGE conditions were performed as described previously (1, 2). Briefly, nested PCR was performed on all samples under sterile conditions using the conserved 16S rRNA gene primers. Initial PCR was performed using 27f (5'-AGA GTT TGA TCM TGG CTC AG-3') and 1492r (5'-GGT TAC CTT GTT ACG ACT T-3') primers (2). Reaction conditions were as follows: 1 ul of template DNA was amplified with Promega Taq (Promega, Madison, WI, USA) with thermal cycler conditions: 20 cycles of 94 ºC for 30 sec, 55 ºC for 30 sec and 72 ºC for 30 sec. Targeted PCR was performed with the clamped primer 536fC (5'-CGC CCG CCG CGC CCC GCG CCC GGC CCG CCG CCC CCG CCCC CWT AAT GGC GCC GMC GAC-3') and 907r (5'-CCC CGT CAA TTC CTT TGA GTT T-3'), that span variable region V4 of bacterial 16S rDNA. PCR amplicons were separated via DGGE using a D-Code system (Bio-Rad Laboratories, Hercules, CA). Each gel contained two sets of internal standards to allow cross-gel comparisons: 100 bp ladder (Sigma-Aldrich, St. Louis, MO, USA) and a separate lane with approximately 700 ng of PCR product amplified from two *D. ponderosae* samples. A linear gradient of denaturant ranging from 40 to 60 % (7 M urea : 40 % (wt / vol) formamide) in a 6 % 2 acrylamide gel matrix was used. Gels were run at 130 V for 5 h. Bands were visualized by staining with SYBR Safe Stain (Invitrogen Corp., Carlsbad, CA, USA) and photographically captured (Fotodyne, Hartland, WI, USA). DGGE bands that were common across multiple

samples were excised from gels (Gene Catcher, The Gel Company, San Francisco, CA, USA) and reamplified with 535f and 907r non-clamped primers and sequenced (University of Wisconsin Biotechnology Center, Madison, WI, USA). Sequences of bands were compared to those deposited in GenBank.

Band locations were analyzed using the band match function in Bionumerics (Applied Maths, Inc., Austin, TX, USA) with a 1.5% tolerance level. Each unique band was considered an operational taxonomic unit (OTU). DGGE bands found common across multiple samples were excised and sequenced at the University of Wisconsin Biotechnology Center, Madison, WI. Taxonomic assignments were made by comparing these sequences to the NCBI microbial 16S dataset available on RefSeq (3) using BLASTN. Principle Component Analysis was performed on a binary matrix of OTU presence and absence for all samples using the R module FactoMineR (4). A nested analysis of similarities (ANOSIM, Primer E software v. 6) was used to test the null hypothesis that association of OTUs was independent of the sampling environment, i.e. *D. ponderosae,* their galleries, and phloem from unattacked trees.

Results

DGGE Microbial Community Analysis. Thirty-one unique OTUs were detected by DGGE from *D. ponderosae* adults, galleries, and phloem from unattacked trees from *P. contorta, P. banksiana,* and their hybrids (Figure S1), of which bands from the 12 most commonly identified OTUs were sequenced for identity (Table S1). OTU #18, most closely related to *Pantoea agglomerans,* was more frequently detected that any other OTU across all host tree species, with frequencies ranging from 50% to 71%. OTU #17, most closely related to *Pseudomonas viridiflava,* was next most frequent across all host tree species, and it was the most frequently detected OTU from *D. ponderosae* adults from any host species at 73% in *D. ponderosae* from

P. contorta. OTU #22 (not sequenced) was detected in between 40% and 50% of samples from each host species, however it was detected in only 19% of *D. ponderosae* from *P. contorta*. OTU #16, most closely related to *Serratia fonticola*, was detected in over 50% of samples from *P. contorta* while it was relatively infrequent in other host species. The opposite trend was observed with OTU #24, also most closely related to the genus *Serratia*. Other sequenced OTUs that were detected in relatively high frequencies were most closely related to the Betaproteobacteria *Massilia aurea* (#15) and the Gammaproteobacteria *Serratia nematodiphila* (#24 & #25) and *Pseudoxanthomonas spadix* (#23). Less frequently detected OTUs include those most closely related to *Micrococcus endophyticus* (#27), *Serratia fonticola* (#6 & #29), *Pseudomonas viridiflava* (#2), and *Acinetobacter johnsonii* (#7).

Supplementary References

1. **Feris, K. P., P. W. Ramsey, C. Frazar, M. C. Rillig, J. E. Gannon, and W. E. Holben.** 2003. Structure and seasonal dynamics of hyporheic zone microbial communities in free-stone rivers of the western United States. Microb. Ecol. **46:**200-15.
2. **Newberry, C. J., G. Webster, B. A. Cragg, R. J. Parkes, A. J. Weightman, and J. C. Fry.** 2004. Diversity of prokaryotes and methanogenesis in deep subsurface sediments from the Nankai Trough, Ocean Drilling Program Leg 190. Environ. Microbiol. **6:**274-87.
3. **Pruitt, K. D., T. Tatusova, G. R. Brown, and D. R. Maglott.** 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. **40:**D130-5.
4. **Lê, S., J. Josse, and F. Husson.** 2008. FactoMineR: An R Package for Multivariate Analysis. J. Stat. Softw. **25:**1-18.