Extended Data

Plant genes influence microbial hubs that shape beneficial leaf communities

Benjamin Brachi, Daniele Filiault, Paul Darme, Marine Le Mentec, Envel Kerdaffrec, Fernando Rabanal, Alison Anastasio, Matthew Box, Susan Duncan, Timothy Morton, Polina Novikova, Matthew Perisin, Takashi Tsuchimatsu, Roderick Woolley, Man Yu, Caroline Dean, Magnus Nordborg, Svante Holm, Joy Bergelson

correspondence to: jbergels@uchicago.edu

Extended Data Fig. 1 to 7
Extended Table 1 to 3
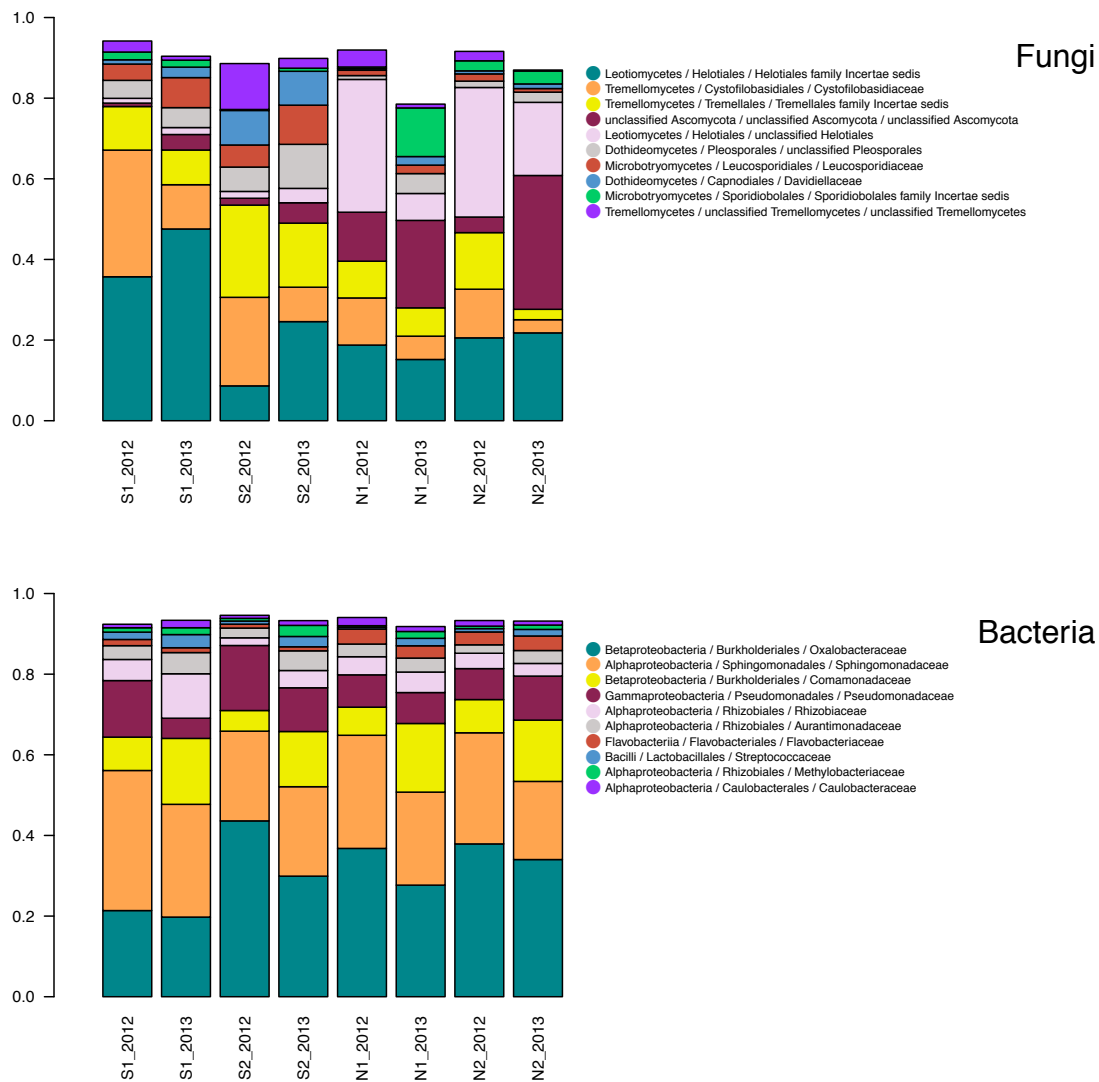
Extended Data Fig. 1: Relative frequency of the first 10 most frequent OTUs. Each stacked bar (x-axis) corresponds to a location/year combination. The y-axis gives the proportion of the 10 most frequent OTUs. The colors correspond to the taxonomical assignments of OTUs given in the legend (class / order / family).
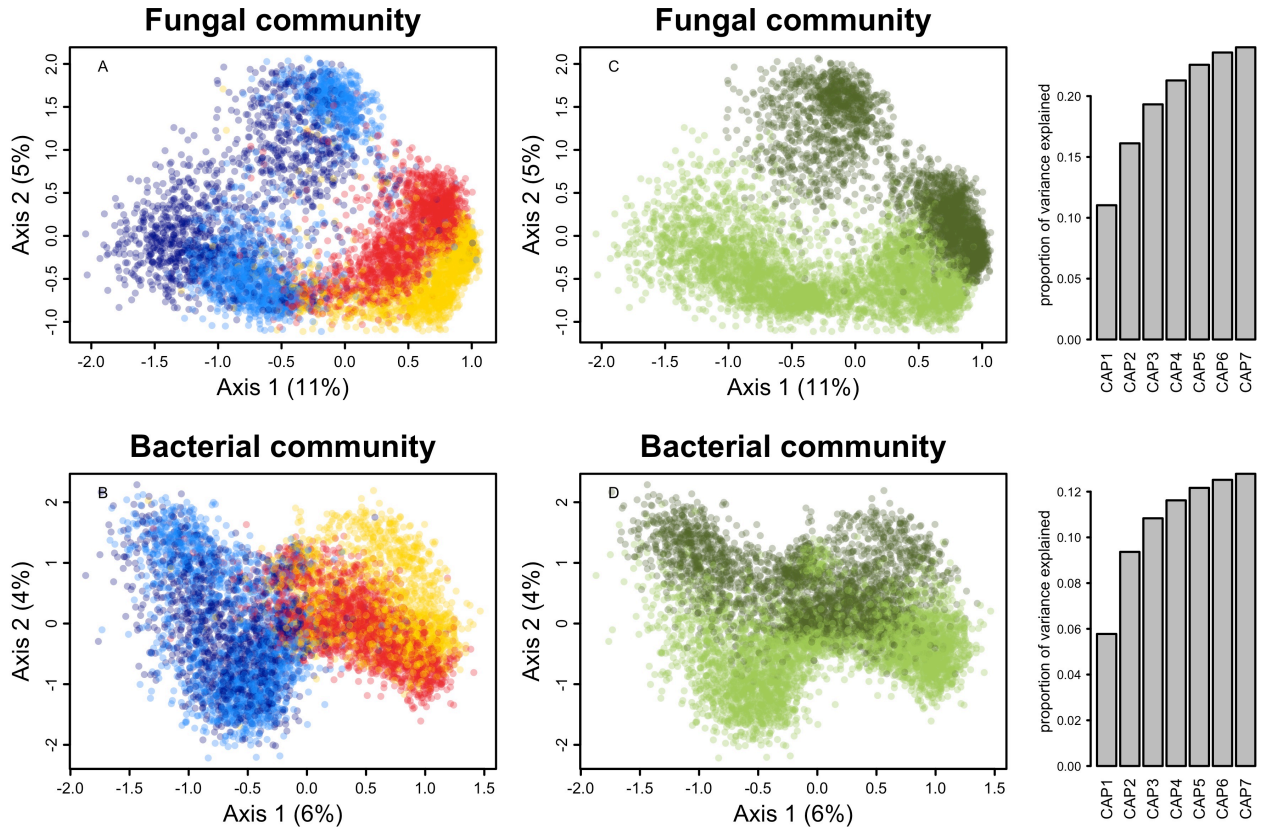
Extended Data Fig. 2: Plants grown in different environments have different microbial communities. The plots represent the projection of each sample on the plane defined by the first two constrained components of the fungal (A and C) and bacterial (B and D) communities, describing variation among sites and year. The percentages in parentheses are the proportion of the total inertia (square root of the Bray-Curtis distance) explained by each component. In all plots each dot represents a sample. In A and B, colors indicate the site from which samples where collected. Experiments from the South are represented in red (S1) and yellow (S2), and experiments from the North in blue (N1) and dark blue (N2). In C and D, dark and lighter green identifies samples collected in 2012 and 2013, respectively. The barplots on the right represent the cumulative proportion of variance explained by the first 7 constrained components (CAP1 to 7) for the fungal and bacterial communities, respectively.
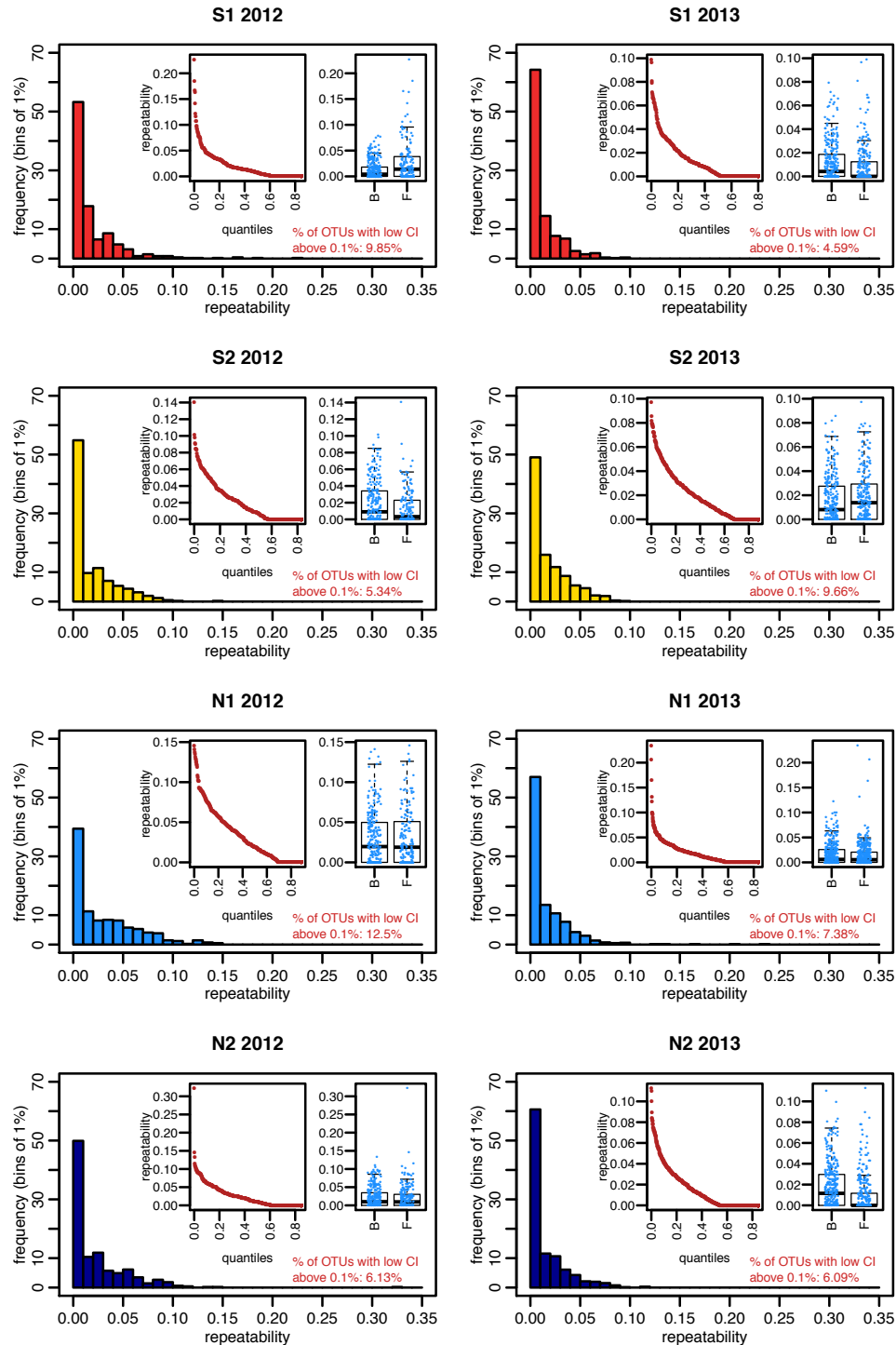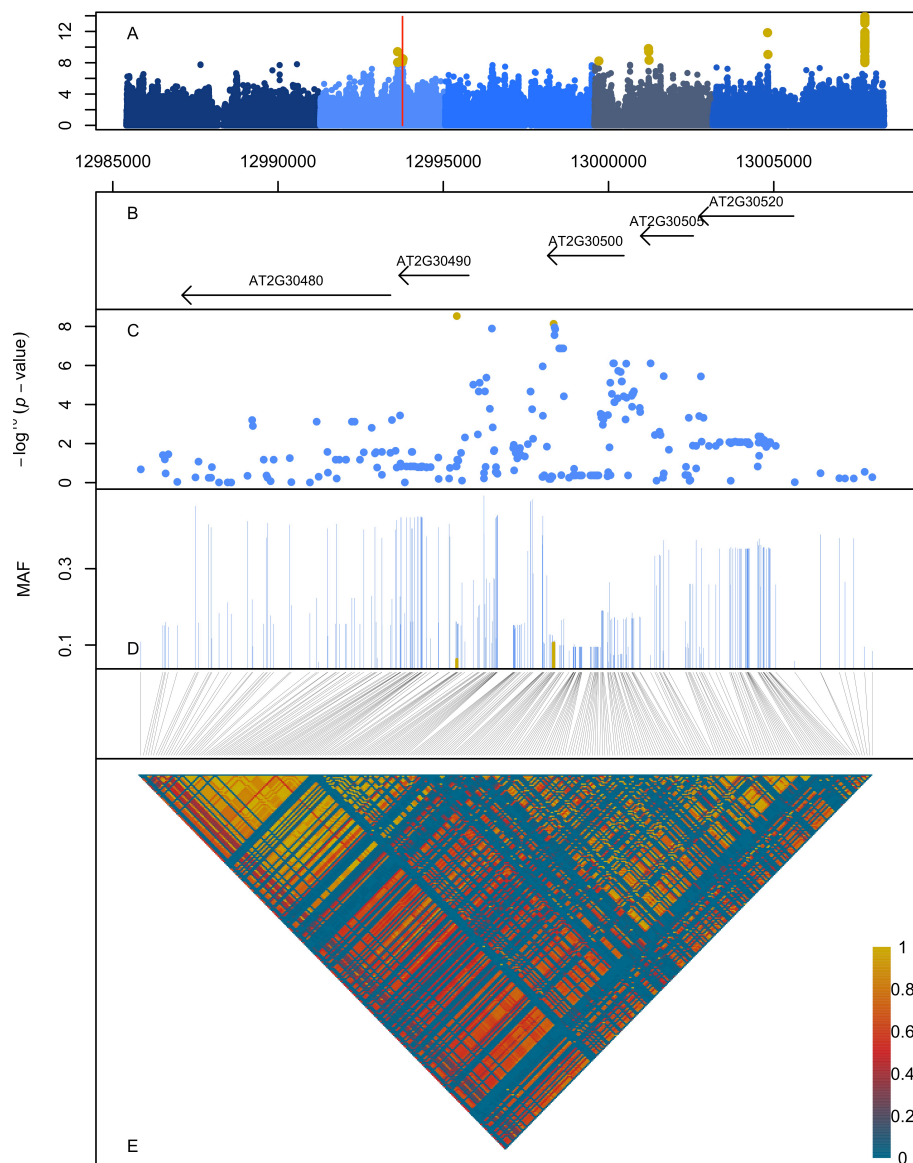
Extended Data Fig. 3: Distribution of heritability values estimated for individual OTUs. Each frame presents the distribution of heritability in one year/location combination, and the two smaller graphics within each frame represent the empirical cumulative frequency distribution of the heritability distribution and the boxplot contrasting bacterial (B) and fungal (F) OTUs.
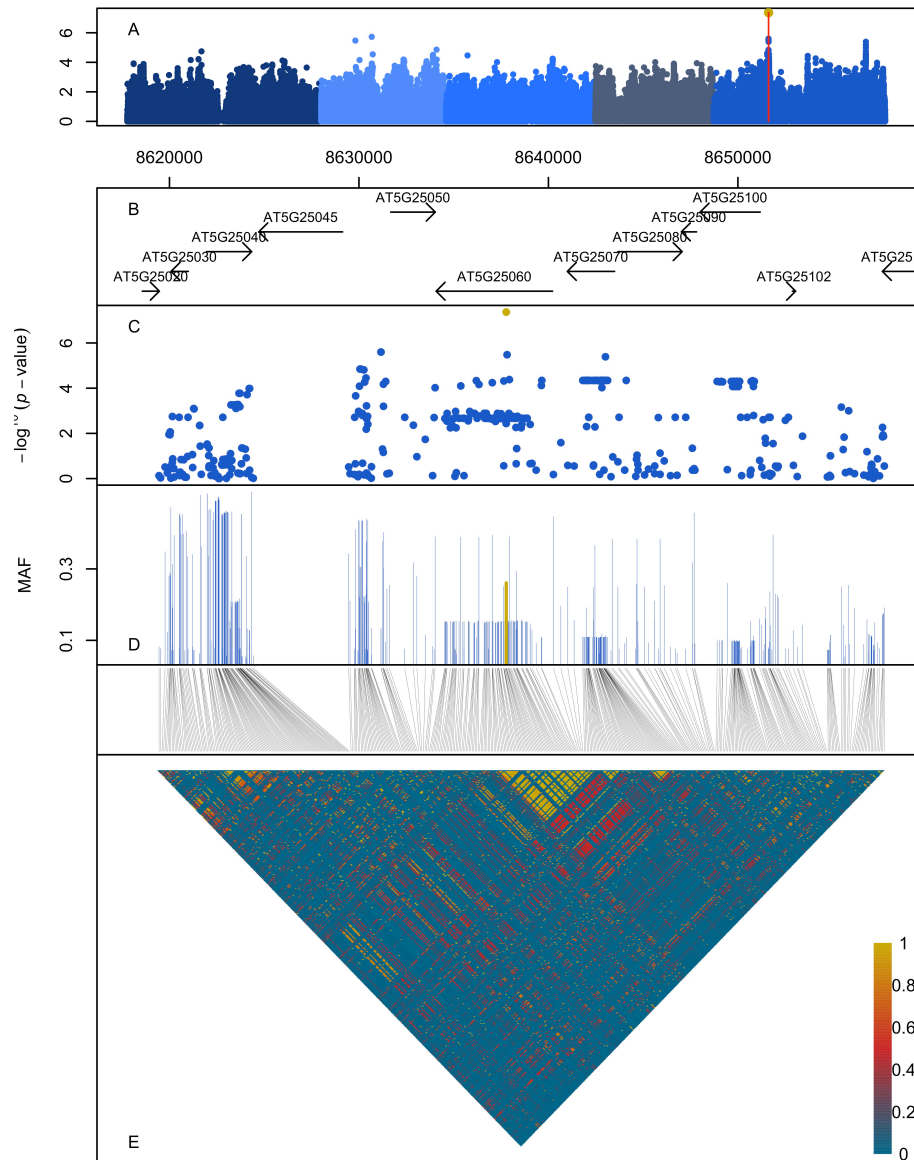
Extended Data Fig. 4: Genome wide association results for F60 in S2 2013. A. Genome-wide Manhattan plot. The y-axis represents the $-\log_{10}$ of the *p*-value. The x-axis is the position along the 5 chromosomes of *A. thaliana*, represented in different shades of blue. The points in yellow correspond to SNPs with $-\log(\text{combined } p\text{-value}) \geq 8$ and the red vertical line indicates the center of the focal region for the following frames. B. Candidate genes in a ~40kb focal region indicated by the red vertical line in A. The arrows indicate the 5'-3' orientation of genes (vertical position is random). C. Associations in the ~20kB region around associated SNP. Axes are the same as in A. D. The x-axis is the same as in C. (coordinate in bp). The y-axis gives the minor allele frequency of all SNPs in the focal region. The segments in yellow correspond to the associated SNPs. E. Pattern of pairwise linkage disequilibrium in the focal region. The grey lines in the frame between D. and E. translate the coordinates from position in bp (in E. and above) to the SNP order.

Extended Data Fig. 5: Association peak 2, chromosome 2, between 12.99 and 13.01 Mbp. A. Genome-wide Manhattan plot. The y-axis represents the $-\log_{10}$ of the combined *p*-value for the 10 genetically correlated heritable hubs. The x-axis is the position along the 5 chromosomes of *A. thaliana*, represented in different shades of blue. The points in yellow correspond to SNPs with $-\log(\text{combined } p\text{-value}) \geq 8$ and the red vertical line indicates the center of the focal region for the following frames. **B.** Candidate genes in a ~40kb focal region indicated by the red vertical line in A. The arrows indicate the 5'-3' orientation of genes (vertical position is random). **C.** Associations in the ~20kB region around associated SNP. Axes are the same as in A. **D.** The x-axis is the same as in C. (coordinate in bp). The y-axis gives the minor allele frequency of all SNPs in the focal region. The segments in yellow correspond to the associated SNPs. **E.** Pattern of pairwise linkage disequilibrium in the focal region. The grey lines in the frame between D. and E. translate the coordinates from position in bp (in E. and above) to the SNP order.

Extended Data Fig. 6: Association peak 6, chromosome 4, between 8.53 and 8.65 Mbp. See Extended Data Fig. 5.
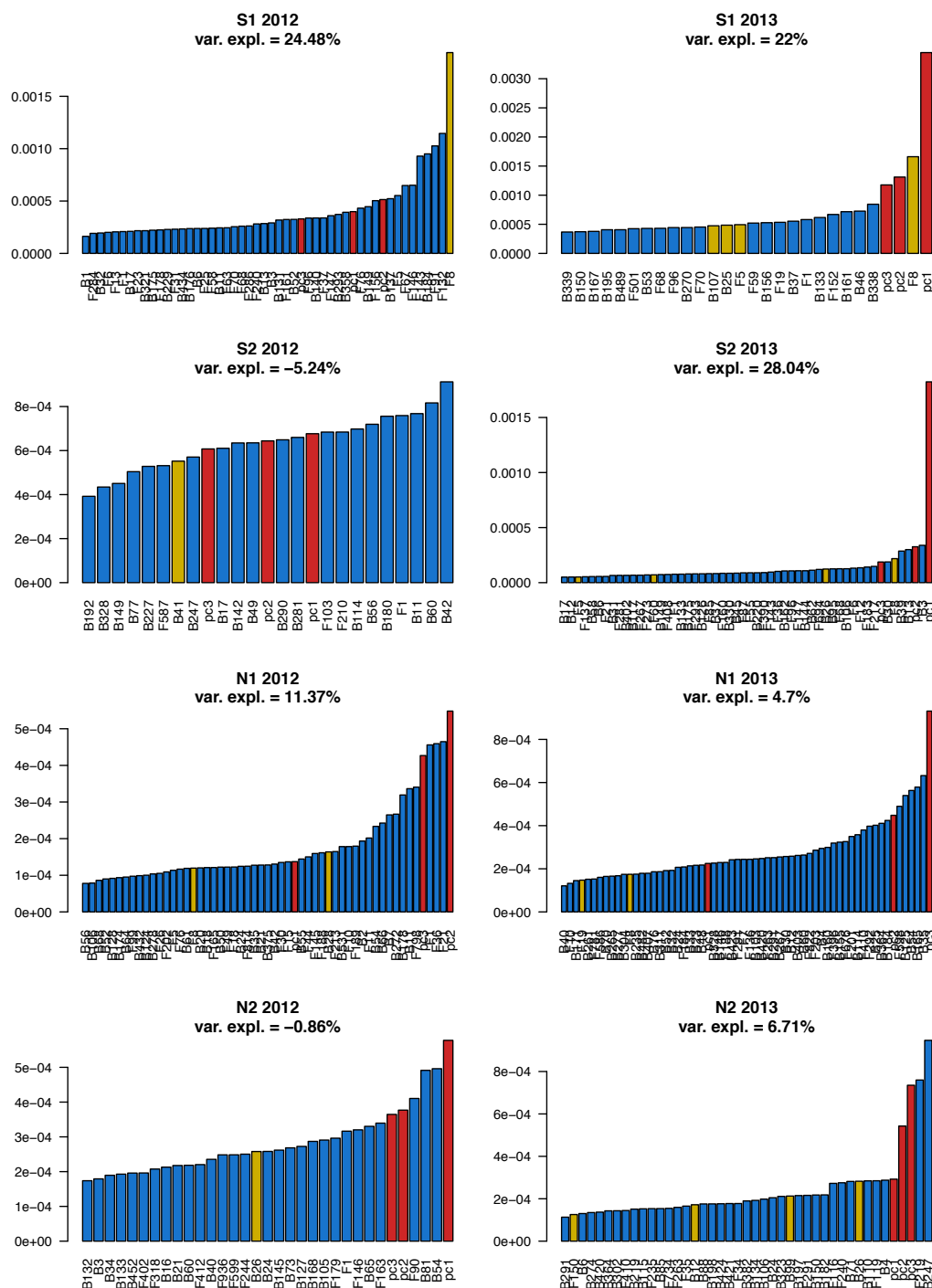
Extended Data Fig. 7: Heritable hubs contribute to fecundity variation. For each location/year, we fitted random forests regression models using breeding values per accession for fecundity as dependent variables. The explanatory variables were breeding values for heritable OTUs (including heritable hubs) and the first 3 components of the pairwise genetic distance (1-Kinship) matrix. In each frame, the y-axis gives the variable importance. Variables include heritable OTUs (blue), heritable hubs (yellow), and the genetic distance components (red).

Extended Data Table 1: Host variation has subtle impact on overall community variation. The first 3 columns indicate the community, location and year for which the analysis was performed. Nh stands for the number of principal coordinate components with significant broad sense heritability estimates (95% confidence intervals not overlapping 0). A total of 10 components were computed for each community/location/year combination. "VE" indicates the total amount of microbial community variation captured by the first 10 components and "he" provides an estimated proportion of total variation explained by the identity of host accessions. The overall host effects reported in the main text reflect the distribution of VE*he in this table. N samples and N OTUs indicate the number of samples and OTUs analyzed in each community/location/year combination.

gene	location	year	Nh	VE	he	N samples	N OTUs
16S	ada	2012	7	55.185	1.244	570	279
16S	ada	2013	5	51.78	1.593	912	338
16S	ram	2012	7	57.683	3.826	589	256
16S	ram	2013	7	45.852	2.115	1021	375
16S	rat	2012	4	57.188	1.204	695	250
16S	rat	2013	8	45.904	1.958	973	295
16S	ull	2012	7	49.242	2.529	1051	321
16S	ull	2013	6	41.472	1.002	982	303
Total number of bacterial OTU/location/year combinations							2417
ITS	ada	2012	6	61.578	1.656	530	210
ITS	ada	2013	5	54.646	0.607	889	320
ITS	ram	2012	9	60.073	2.743	578	163
ITS	ram	2013	7	42.543	1.379	996	425
ITS	rat	2012	4	47.986	1.990	696	163
ITS	rat	2013	9	49.498	2.650	981	233
ITS	ull	2012	9	58.717	5.248	996	156
ITS	ull	2013	3	56.253	0.724	990	244
Total number of fungal OTU/location/year combinations							1914

Extended Data Table 2: Hubs are enriched for interkingdom connections (edges). For each site (first column) and year (second column), the table presents the results from a χ^2 testing for enrichment in interkingdom edges when considering all edges, or edges involving at least one hub. B_B, B_F, F_F give the number of edges between 2 bacterial OTUs, a bacterial and a fungal OTU, and 2 fungal OTUs, respectively.

site	year	edges	B_B	B_F	F_F	χ^2	p-value	FDR adj. p-value
S1	2012	all	1148	134	556	107.325	2.00E-05	0.0002
		with hubs	369	61	34			
S1	2013	all	976	140	844	26.922	4.00E-05	0.0003
		with hubs	132	44	190			
S2	2012	all	981	137	484	41.968	2.00E-05	0.0002
		with hubs	277	41	50			
S2	2013	all	1039	157	873	25.536	2.00E-05	0.0002
		with hubs	132	49	162			
N1	2012	all	1212	157	637	71.122	2.00E-05	0.0002
		with hubs	240	29	23			
N1	2013	all	1427	288	1185	25.593	2.00E-05	0.0002
		with hubs	271	104	295			
N2	2012	all	951	71	430	15.604	6.20E-04	0.0050
		with hubs	183	16	42			
N2	2013	all	1720	229	1473	62.111	2.00E-05	0.0002
		with hubs	254	94	341			

Extended Data Table 3: Breeding values of heritable hub F8 have a significant positive effects on breeding-values for fecundity in linear mixed model including a random term accounting for IBS kinship (K) among accessions. For each location and year, models were defined as follow $Y_i \sim hh_{1i} + \ldots + hh_{ni} + K + error_i$, where Y_i is the fecundity breeding value for the ith accession, hh_{ni} are fixed effects for the breeding values for the ith accessions for the nth heritable hub identified in the location/year combination, K is a random term capturing genome wide relatedness among accessions (kinship) and $error_i$ captures residuals using a normal distribution.

| experiment | sample size | Fixed coefficients | | | | | Random effects | | |
		coefficients	Value	Std Error	z	*p*-value	Group	Std Dev	Variance
S1 2012	196	Intercept	0.00014	0.00049	0.29	7.70E-01	K	8.61E-03	7.41E-05
		F8	**0.00701**	**0.00157**	**4.48**	**7.60E-06**	Residual error	0.006816532	
S1 2013	198	Intercept	0.00022	0.00052	0.42	6.70E-01	K	8.79E-03	7.73E-05
		B107	-0.00476	0.00596	-0.80	4.20E-01	Residual error	0.007283165	
		B25	0.00284	0.00744	0.38	7.00E-01			
		F5	-0.00499	0.00705	-0.71	4.80E-01			
		F8	**0.01529**	**0.00428**	**3.57**	**3.60E-04**			
S2 2012	177	Int	0.00064	0.00055	1.16	2.50E-01	K	9.22E-03	8.49E-05
		B41	-0.00657	0.00523	-1.26	2.10E-01	Residual error	0.006883981	
S2 2013	197	Intercept	0.00007	0.00032	0.21	8.40E-01	K	4.49E-03	2.02E-05
		B26	0.00789	0.00387	2.04	4.10E-02	Residual error	0.004501831	
		F5	-0.01309	0.00351	-3.73	1.90E-04		4.08E-03	1.66E-05
		F60	0.01130	0.00369	3.06	2.20E-03		0.007389621	
		F8	**0.00883**	**0.00394**	**2.24**	**2.50E-02**			
N1 2012	160	Intercept	0.00080	0.00060	1.33	1.80E-01	K	4.08E-03	1.66E-05
		B28	0.00816	0.00692	1.18	2.40E-01	Residual error	0.007389621	
		B38	-0.00707	0.00585	-1.21	2.30E-01			
		F8	0.00387	0.00289	1.34	1.80E-01			
N1 2013	198	Intercept	0.00031	0.00051	0.61	5.40E-01	K	8.75E-03	7.65E-05
		F19	0.00590	0.00713	0.83	4.10E-01	Residual error	0.007104794	
		F60	-0.02173	0.00817	-2.66	7.80E-03			
		F8	0.00276	0.00532	0.52	6.00E-01			
N2 2012	179	Intercept	0.00064	0.00056	1.15	2.50E-01	K	1.49E-04	2.22E-08
		B26	0.00262	0.00457	0.57	5.70E-01	Residual error	0.007457504	
N2 2013	197	Intercept	0.00005	0.00043	0.12	9.10E-01	K	6.44E-03	4.15E-05
		B12	-0.00403	0.00534	-0.75	4.50E-01	Residual error	0.005948529	
		B26	0.01793	0.00914	1.96	5.00E-02			
		B99	0.00164	0.00358	0.46	6.50E-01			
		F150	-0.00073	0.00460	-0.16	8.70E-01			