Electronic Supplementary Material (ESM)

ESM.1) Colony resistance profile

We set up laboratory colonies with field caught *Bombus terrestris* queens from Neunforn, Switzerland in spring 2013. The resistance profile of thirty parasite-free colonies was determined by administering a "cocktail" infection to five workers per colony. The cocktail contained 2'000 parasite cells each of five genetically distinct *Crithida bomi* strains (project tags: 08.068, 08.075, 08.091, 08.161, 08.192) in 10 μl of 50% sugar water and was presented to the worker after a starvation period of 2h. Seven-days post infection [1,2] bees were frozen and DNA was extracted from dissected guts with Qiagen DNeasy Blood & Tissue Kit following the manufacturer's protocol.

Infection intensity was assessed by quantitative real-time PCR (see [3] for primers and cycling protocol). The total number of parasite cells in a sample was determined by absolute quantification using the standard curve method. For this, we extracted DNA from a known number of *C. bombi* cells to generate the standard curves. Each biological sample was run in triplicates in a total reaction volume of 10μl containing: 0.2μl [10μM] of each forward and reverse primer, 2μl of 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (ROX) (Solis BioDyne) and 1μl DNA template.

Infection diversity was determined by genotyping the samples for three *C. bombi* microsatellite markers Cri4G9, Cri4 and Cri2F10 [4]. Microsatellite were amplified in one multiplexed reaction in a total reaction volume of 10μl containing: 2μl of 5x Colorless GoTaq® Reaction Buffer (Promega), 0.5μl of dNTPs [2.5mM each], 0.1μl (0.15μl Cri4G9) [10μM] of each forward and reverse primer, 0.05μl GoTaq DNA polymerase (Promega, 5U / μl) and 2μl DNA template. A total of 40 PCR cycles were performed with the following steps: denaturation (94°C, 30s), annealing (48°C, 30s) and extension (72°C, 30s). PCR products were run on a 3730xl DNA Analyzer (Applied Biosystems).

ESM.2) Microbiota faecal transplant experiment

To transplant the microbiota, we followed the faecal transplant protocol by Koch and Schmid-Hempel [5] with a few modifications. Briefly, all materials used in the context of the microbiota transplant experiment were either autoclaved or washed in 80% alcohol before use. In order to raise germ-free workers (referred to as recipients), we removed and surface sterilized brood from each of the selected colonies by submerging the brood in a 3% bleach solution for 90s. Subsequently, the brood was placed, for each colony individually, in sterilized housing boxes.

Daily, newly emerged recipients were transferred into individual housing boxes and kept for 1-3 days before a faecal transplant was offered to them. Upon emergence, recipients were provided with ad libitum pollen (gamma-irradiated, LEONI Studer Hard AG, Switzerland, dose: 35.5kGy [6], kept frozen at -20° until use) and 50% sugar water (autoclaved).

Faeces used for the microbiota transplants were collected from workers (referred to as donors), which emerged and remained in their source colonies (i.e. naturally acquired their microbiota). A single transplant consisted of faeces collected from at least three source colony workers; faeces collected from different colonies were never mixed. A microbiota transplant consisted of 5μl faeces mixed with 10μl 50% sugar water and was offered to a recipient after a 30min starvation period. Once recipients were observed to have fed on the inoculum, they were put back in their housing boxed and kept for 18h before being snap frozen in liquid nitrogen and stored at -80°C. Any recipient bee that did not feed or only partially fed on the inoculum was excluded from the experiment.

ESM.3) RNA and DNA extraction

RNA and DNA were simultaneously extracted from whole abdomen. For this, we disrupted the abdomen with 0.5 g 1.4mm Zirconium Oxide beads in 1ml peqGold TriFast™ reagent (peQlab) at room temperature (Omni Bead Ruptor 24 Homogenizer). After a 10min centrifuging step (12'000g, at 4°C) 450μl of the supernatant was transferred to a new tube containing 50μl BCP (1-Bromo-3-chloropropane), mixed well, incubated for 10min at room temperature and then centrifuged for 10min (12'000 g, at 4°C). The result was a phase separation into three phases, an aqueous phase, an interphase and a phenol phase. For RNA extraction, 180μl of the upper aqueous phase was transferred into a RNAse-free tube containing 225μl Isopropanol, vortexed for 5-10s, incubated for 10min at room temperature and centrifuged for 10min (17'785g, 4°C). The supernatant was carefully discarded and 300μl 75% EtOH (prepared with DEPC treated H_2O) was added, vortexed and centrifuged for 5min (17'785g). To improve RNA quality, this washing step was repeated. Finally, the supernatant was removed and the pellet was allowed to air dry for 3-5min before RNA was resolved in 50μl DEPC treated H_2O and stored at -80 °C. Samples were extracted in four randomly selected extraction batches.

For DNA extraction 100μl of the phenol phase was removed and discarded before 300μl 100% EtOH was added to the sample, vortexed, incubated at room temperature for 5min and centrifuged for 15min (17'785g, 4°C) to precipitate DNA. The supernatant was discarded, and the DNA pellet was washed in 200μl 100% EtOH and centrifuged for 5min (17'785g, 4°C) before the supernatant was carefully discarded. The pellet was resuspended in 50μl H_2O and incubated for 1h at 56°C. Following the incubation period, 150μl PBS were added before DNA was extracted with Qiagen DNeasy Blood & Tissue Kit following manufacturer's protocol, however the final elution volume was reduced to 70μl.

ESM.4) Gene expression protocol

DNA quality and purity was checked for 12 randomly selected samples on a 2100 Bioanalyzer (Agilent Technologies) with the RNA 600 Nano Kit and all samples were quantified and normalized based on Nanodrop 8000 (ThermoScientific) measurements. Even though no genomic DNA contamination was indicated, we used the Turbo DNA-free kit (Ambion) to purify 1.52μg RNA before reverse transcribing 0.5μg RNA using the QuatiTec® Reverse Transcription Kit (Qiagen). Manufactures protocol (Advanced Development Protocol 14, PN 100-1208B) was followed to measure gene expression with a Fluidigm 96.96 dynamic array IFCs (BioMark) using EvaGreen DNA Binding Dye (Biotium) as a reporter.

Each biological sample was run in triplicates for 31 genes (ESM Table S1). Samples with a Ct standard deviation >0.2 were checked for quantification outliers among the technical replicates. The technical replicate with the largest deviation to the mean was excluded given three replicates (i.e. no failed quantification) otherwise all measurements were retained.

The most stable combination of housekeeping genes was identified by the geNorm algorithm within qbase+ (Biogazelle). Thus, we excluded the housekeeping gene PLA2 (M = 0.909, CV = 0.284) and used the arithmetic mean of the measured Ct values for the genes ef1α (M = 0.556, CV = 0.184) and RPL13 (M = 0.556, CV = 0.204) as the normalization factor within samples [7]. Measurements from transplants of the same donor-recipient colony combinations were averaged (ESM Fig. S1B). Every recipient colony received transplants from both donor phenotypes; as well as every donor colony was transplanted into both recipient phenotypes (ESM Fig. S1A).

ESM Table S1: Genes and primers used for the quantification of the expression response.

Functional gene Class	Gene	Putative gene function (pathway)	NCBI accession	Forward primer	Reverse primer	Product length	Primer ref.[1]
Recognition	BGRP1	Recognition receptor (Toll)	XM_003397996	AACGTGGAAGTCAAAGATGG	GCGAACGATGACTTGGTATT	206	[1]
	BGRP2	Recognition receptor (Toll)	XM_003394713	TAACTCCCTTTGGAAACACG	GGCGGTAAAATACTGAACGA	249	[1]
	Dome	Recognition receptor (JAk/STAT)	XM_012310386	AAAGCCGTTCACTCTAAGCA	GACTTGCGAAAGAAGAAACG	116	
	Hemomucin	Surface glycoprotein, potential recognition receptor	XR_131963	AGCATTCCCAGATTTAGCACT	TAACAGTTGATTTCGGAGGTA	173	[2]
	PGRP-LC	Recognition receptor (Imd)	XM_003396463	CAGCCACCTACGACAGATTT	GTACATTCCGCTTGTGTCCT	101	[1]
	PGRP-S3	Recognition receptor (Toll)	XM_003401893	CGTGAAGGAGCTCATACCAT	CCAGGACTCATAGTGGCTGT	200	[1]
	Toll-1	Recognintion receptor (Toll)	XM_012307988	CGAATGGAGTTTAGAGCAGC	ATTTATCCCAGAACCAAGGG	172	
Signaling	Basket	Signal molecule (JNK)	XM_003402794	GGAACAAGATAATCGAGCAACTG	CTGGCTTTCAATCGGTTGTG	177	[1]
	Hopscotch	Signal molecule (JAK/STAT)	XM_003401903	CACAGACTGAAGCAGGTTGA	CATATGGGTAATTTGGTGCC	353	[1]
	MyD88	Signal molecule (Toll)	XM_003394153	GCATTAGGCATTGACAAACGAC	CAGAAGTCATACAAACCCACTCTG	125	
	Relish	Signal molecule (Imd)	XM_003399472	CAGCAGTAAAAATCCCCGAC	CAGCACGAATAAGTGAACATA	156	[2]
Effectors	Abaecin	Antimicrobial peptide	XM_003394653	GCCACAATATGTGGAATCCT	ATGACCAGGGTTTGGTAATG	141	[1]
	Apidaecin	Antimicrobial peptide	XM_003402966	CCCGACTAATGTACCTGCCA	GAAGGTGCGAATGTGTTGGA	131	[1]
	Defensin	Antimicrobial peptide	XM_003395924	GTCTGCCTTTGTCGCAAGAC	GACATTAGTCGCGTCTTCTTCG	139	[1]
	Ferritin	Iron transportation	XM_003393332	AAAGAATTGGACGCAAATGG	CAGCGAACTGATGTCCAAGA	259	[1]
	Hymenoptaecin	Antimicrobial peptide	XR_132450	TTCATCGTACTGGCTCTCTTCTG	AGCCGTAGTATTCTTCCACAGC	85	[1]
	Lysozyme3	Bacteriolytic effector	XM_003394052	TATGGGCAAGAAGATTCGAC	GTGTACATCGTTCACGCATC	219	[1]
	TEPA	Effector molecule (JAK/STAT)	XM_003399699	GCGTTCTATGACCACCTGTT	TACAGGTTACTCCACAGCCC	212	[1]
	Transferrin	Iron-binding, antibacterial	XM_003401163	CAATTTCTTCACCGCATCCT	CCTCGTTATTTGGCTTGCAT	131	[1]
Metabolism	Apolipophorin.III	Lipid transport	XM_003402572	ATCAGGCTCAAACGAACATC	TTCGTTCACTTGTTGCTGAG	269	[3]
	CYP4G11	Stress response	XM_003399563	GAATGCGCAAAGAAGGTAGC	CGCTTTCCGCTCTTGTAATC	313	[3]
	Vitellogenin	Metabolic, endorinological	XM_003402655/ XM_003402656	GTGACAAGCGAAGAGACTATTATG	CCGTGTTATCTGGCGTGAC	154	[4]
ROS	Jafrac	Peroxiredoxin, ROS regulation	XM_003401245	CTCACTTCAGTCACTTGGCA	GCCAGCAGGACATACTTCTC	290	[1]
	Peroxiredoxin5	Peroxiredoxin, ROS regulation	XM_003394777	TCACACCAGGATGTTCCAAGAC	TTCTGCTCCGTGTTCTTTACCC	146	[1]
Melanisation	Catsup	Enzyme, melanin synthesis	XM_003398173	TTACCATGACGAGTCACCAA	ATGAGGAACCAAAGCATGAG	355	[1]
	PPO	Prophenoloxidase, melanin synthesis	HM142999	AGCGGCATAATACGTTGTGT	CCGAGGGATAGAAAGTCTCC	329	[3]
	Punch	Enzyme, melanin synthesis	XM_012320347	ATTGCCAGGACACTTTCAAC	TACAAGCTGGAAACGGAAAC	212	[1]
	Serpin27a	Serine protease inhibition (PPO)	XM_003392985	CCGATCATCCATTCGTATTC	ACCTGCACTTGATATCCCTG	164	[1]
Housekeeping	ef1α	Elongation factor 1 α	XM_003401944	GCTGGTGACTCGAAGAACAATC	GGGTGGTTCAACACAATAACCTG	74	[3]
	PLA2	Phospholipase A2	FN391388	TATCTTTCAATGCCCAGGAG	GTCGTAACAAATGTCATGCG	129	[5]
	RPL13	Ribosomal protein L13	FN391387	GGTTTAACCAGCCAGCTAGAAA	CTTCACAGGTCTTGGTGCAA	83	[5]

[1]Primers with no reference were designed with either Primer3 [6] or Quantprime [7] with the following specifics: primer length 20±2bp; melting temperature 60±1°C with a maximum temperature difference between forward and reverse primer of 0.5°C. Primers were evaluated with an annealing temperature of 60°C and found to be specific with an amplification efficiency of 1.9 - 2.1.

1. Brunner, F. S., Schmid-Hempel, P. & Barribeau, S. M. 2013 Immune Gene Expression in Bombus terrestris: Signatures of Infection Despite Strong Variation among Populations, Colonies, and Sister Workers. *PLoS One* **8**, e68181. (doi:10.1371/journal.pone.0068181)
2. Schlüns, H., Ben M Sadd, Schmid-Hempel, P. & Crozier, R. H. 2010 Infection with the trypanosome Crithidia bombi and expression of immune-related genes in the bumblebee Bombus terrestris. *Dev Comp Immunol* **34**, 705–709. (doi:10.1016/j.dci.2010.02.002)
3. Barribeau, S. M. & Schmid-Hempel, P. 2013 Qualitatively different immune response of the bumblebee host, Bombus terrestris, to infection by different genotypes of the trypanosome gut parasite, Crithidia bombi. *Infect Genet Evol* **20**, 249–256. (doi:10.1016/j.meegid.2013.09.014)
4. Li, J., Huang, J., Cai, W., Zhao, Z., Peng, W. & Wu, J. 2009 The vitellogenin of the bumblebee, Bombus hypocrita: studies on structural analysis of the cDNA and expression of the mRNA. *J Comp Physiol B* **180**, 161–170. (doi:10.1007/s00360-009-0434-5)
5. Horňáková, D., Matoušková, P., Kindl, J., Valterová, I. & Pichová, I. 2010 Selection of reference genes for real-time polymerase chain reaction analysis in tissues from Bombus terrestris and Bombus lucorum of different ages. *Anal Biochem* **397**, 118–120. (doi:10.1016/j.ab.2009.09.019)
6. Rozen, S. & Skaletsky, H. 1999 Primer3 on the WWW for General Users and for Biologist Programmers. In *Bioinformatics Methods and Protocols*, pp. 365–386. New Jersey: Humana Press. (doi:10.1385/1-59259-192-2:365)
7. Arvidsson, S., Kwasniewski, M., Riaño-Pachón, D. M. & Mueller-Roeber, B. 2007 QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* **9**, 465–465. (doi:10.1186/1471-2105-9-465)

ESM Fig. S1: Schematic of the experimental design. Panel (A) visualizes all possible faecal transplant crosses of donor and recipient resistance phenotypes (arrows). Panel (B) shows all possible donor-recipient colony combinations as grey squares; each bar within a square represents a successful independent faecal microbiota transplant (n = 58) of a particular donor-recipient colony combination (n = 31). Axes are colony ID numbers.

ESM.5) Statistical analyses of gene expression data

All statistical analyses for gene expression data were performed on dCt-values. Only for illustration purposes we show fold changes values of gene expression relative to the resistant phenotype for both recipient and donor effects (ddCt). We applied MANOVAs to test for effects of recipients or donor resistance phenotype (ESM Fig. S1A) on the gene expression response of the functional gene groups. The dCt-values within each gene class were subjected to Yeo-Johnson power transformation to achieve multivariate normality. Backwards model selection was performed for each gene class on the initial full model: donor phenotype (resistant, susceptible) * recipient phenotype (resistant, susceptible) retaining only effects with a p-value < 0.1. We checked for outliers using Mahalanobis distances. However, results were robust to outlier exclusion and outliers were therefore not excluded for the reported results (ESM Table S5 for results with outliers excluded).

To further evaluate MANOVA results, we used linear discriminant analysis. We identified genes that contributed most to the separation of resistance phenotypes in the multivariate setting based on partial correlation coefficients of the linear discriminant function. We used leave-one-out (jack-knifed) cross-validation of the linear discriminant analysis to assess the accuracy of the phenotype discrimination by the linear function within a gene class for either donor or recipient effects (ESM Table S3/S4). If MANOVA results indicated a significant effect, we checked the univariate output of the MANOVA to investigate differential expression between phenotypes for each gene individually and we used the Benjamin and Hochberg method [8] to control the false discovery rate.

To test for a difference in gene expression variation among the resistant phenotypes of either recipient or donor effect we calculated Euclidean distances between samples, and used a permutation test on distances to the centroid to compare to the null hypothesis of equal variance among the resistant phenotypes [9].

ESM Table S3: Percentage of correctly classified cases of the cross-validation of the linear discriminant function for each functional gene class.

		Recipient effect		Donor effect	
		Predicted phenotype		Predicted phenotype	
		resistant	susceptible	resistant	susceptible
Melanisation					
True phenotype	Resistant	54.55	45.45	81.82	18.18
	Susceptible	37.50	62.50	22.22	77.78
Proportions[1]		40.74	59.26	37.93	62.07
Effectors					
True phenotype	Resistant	81.82	18.18	58.33	41.67
	Susceptible	22.22	77.78	38.89	61.11
Proportions[1]		37.93	62.07	40.00	60.00
Recognition					
True phenotype	Resistant	63.64	36.36	66.67	33.33
	Susceptible	33.33	66.67	33.33	66.67
Proportions[1]		37.93	62.07	40.00	60.00
Metabolism					
True phenotype	Resistant	0	100	38.46	61.54
	Susceptible	72.22	27.78	22.22	77.78
Proportions[1]		41.94	58.06	41.94	58.06
ROS					
True phenotype	Resistant	27.27	72.73	83.33	16.67
	Susceptible	50.00	50.00	27.78	72.22
Proportions[1]		37.93	62.07	40.00	60.00
Signaling					
True phenotype	Resistant	36.36	63.64	63.64	36.36
	Susceptible	56.25	43.75	27.78	72.22
Proportions[1]		40.74	59.26	37.93	62.07

[1]Proportions gives percentage of phenotypes in the data set. Proportions vary for some gene classes due to gene expression failure for some samples and genes. However note, the prior classification probability for the cross-validation of the linear discriminant function was set to 0.5.

ESM Table S4: Partial correlation coefficient of linear discriminant functions for either recipient or donor phenotypes.

	Recipient phenotype	Donor phenotype
Melanisation		
Catsup	4.130598	-5.824617
PPO	7.484474	1.427122
Punch	-2.967775	-1.082558
Serpin27a	-28.734568	2.721916
ROS		
Jafrac	-1.293329	-2.512490
Peroxiredoxin5	8.873714	5.542687
Metabolism		
Apolipophorin.III	-0.4304080	-0.3563731
CYP4G11	0.0173596	-0.1947528
Vitellogenin	0.3259856	0.7014135
Effectors		
Abaecin	-0.1818116	0.1281196
Ferritin	-0.3394093	-1.7669270
Apidaecin	-1.3002121	1.2680009
Defensin	0.2550820	-0.5152214
Hymenoptaecin	-0.5381095	-0.2040488
TEPA	99.8821058	32.3809248
Lysozyme3	-1.8080832	-0.8738062
Transferrin	0.8915005	1.4742006
Signalling		
Basket	-50.42431	-14.78089
MyD88	50.35166	-16.51957
Hopscotch	73.90705	-73.37020
Relisch	-196.73343	471.90052
Recognition		
PGRP-S3	-11.928997	20.5579004
BGRP1	-3.574761	-1.4058511
BGRP2	-16.837104	-12.6516326
Dome	-1.135818	6.5397240
PGRP-S3	6.236538	-1.6064176
Toll-1	-11.435283	1.4474364
Hemomucin	9.321592	0.7595999

ESM Table S5: MANOVA results for all gene classes and summary of linear discriminant analysis (LDA) with outliers excluded.

Functional gene class	Factor[1]	Df, Residuals	Pillai	Approx F	Num Df, Den Df	P-value	LDA accuracy[2]	Two genes with highest coefficient
Recognition (-10 outliers)	Recipient phenotype	1,18	0.641	3.062	7,12	0.043	60.00%	PGRP-LC, dome
Signalling (-6 outliers)	Donor phenotype	1,22	0.435	3.653	4,19	0.023	73.91%	Relish, hopscotch
Effectors (-10 outliers)	Recipient phenotype	1,16	0.786	4.584	8,10	0.014	90.91%	TEPA, defensin
Metabolism (-5 outliers)	Donor phenotype	1,24	0.261	2.583	3,22	0.079	69.23%	Vitellogenin, apolipophorin.III
ROS (-3 outliers)	Donor phenotype	1,25	0.334	6.019	2,24	0.008	74.07%	Peroxiredoxin5, jafrac
Melanisation (-6 outliers)	Recipient phenotype	1,22	0.471	4.007	4,18	0.017	95.65%	Serpin27a, PPO
	Donor phenotype	1,22	0.528	5.036	4,18	0.007	69.57%	Catusp, punch

[1]Statistics of the minimal model (i.e. retaining only independent variables with a p-value ≤ 0.1) for the MANOVA results are reported.
[2]Summary values for LDA classification function are given (i.e. accuracy), as well as the two genes contributing most to the discriminant function.

ESM.6) 16S amplicon library preparation and sequencing

We generated a multiplexed 16S amplicon paired-end library for sequencing on the MiSeq® Illumina platform in order to investigate microbial gut communities of recipients that received a microbiota transplant. We amplified the variable region V3-V4 of the 16S rRNA gene with universal primers [10,11] (ESM Table S2). Primers were designed after principles described in [12]. We thus had three primer pairs containing frameshifting nucleotides between the region-specific part and the Illumina overhang adapter in order to increase general sequence diversity in the generated 16S amplicon libraries (ESM Table S2). We followed the manufacturer's suggested two-step amplification workflow.

Thus an initial PCR was set up for each sample and each primer pair as followed: 1.5µl [10µM] forward and reverse primer mix, 12.5µl 2 x KAPA HiFi Hot Start Ready Mix and 2µl DNA template in a total reaction volume of 25µl. The cycling protocol used a 5min initial degradation step at 95°C, followed by 22 cycles of 98°C for 20s, 58°C for 15s and 72°C for 15s and finished with a final elongation step for 5min at 72°C. All four independent PCR reaction products were pooled and purified using Agencourt AMPure XP beads (Beckman Colter) with a ratio of 0.8:1 (beads to PCR product) and eluted in 30µl [10mM] Tris-Buffer.

A second limited-cycle PCR was performed in order to attach indices and sequencing adapters to the template libraries. A total reaction volume of 50µl contained 10µl of the purified PCR product, 25µl 2x KAPA HiFi Hot Start Ready Mix and 5µl of each forward and reverse primers of the Nextera® XT Index Kit v2 Set A (Illumina). The indexing cycle protocol used a 3min initial denaturation step at 95°C followed by 10 cycles of 30s at 95°C, 30s at 55°C and 30s at 72°C and 5min final extension step at 72°C. The indexed amplicon libraries were purified using Agencourt AMPure XP beads (Beckman Colter) with a ration of 1:1 (beads to amplicon product) and eluted in 30µl [10nM] TrisBuffer.

Amplicon library quality and library fragment size were checked for 10 randomly selected samples on a Bioanalyzer (Aligent Technologies) DNA HS 1000 chip. All libraries were 1:10'000 diluted and quantified in duplicates by quantitative PCR using the 2 x KAPA SYBER qPCR Ready Mix on an ABI 7500 Real-time PCR System (Applied Biosystems). We used the Library Quant Illumina Kit (KAPA Biosystems) with quantification standards ranging from 0.002pM to 20pM for absolute quantification. Following quantification, equimolar amounts of all libraries were pooled. The final, multiplexed pool was quantified by qPCR as previously described. We followed the manufacturer's manual (Preparing Libraries for Sequencing on the MiSeq® #15039740 Rev. D) to prepare the library for paired-end sequencing. We loaded a final library concentration of 17.5pM and a 15% PhiX spike-in with MiSeq® Reagent Kit v3.

ESM.7) Amplicon processing and OTU clustering

A total of 18,122,556 paired-end raw reads were generated in a single 600-cycle MiSeq run. Qualities of reads were checked with FastGQ (v0.11.2). 96% (17,432,405) of the paired reads successfully merged using FLASH (v1.2.9) with the following specifications: minimum overlap of 15bp, maximum overlap of 250bp and maximum mismatch density of 0.25. Primer sequences were trimmed from the merged reads using cutadapt (v1.5) requiring a full-length error-free overlap but allowing wildcards.

Merged and primer trimmed reads were quality filtered with PRINSEQ-lite (v0.20.4) and 13,878,054 reads (76.6%) passed filtering, given a fragment range of 350-550bp, GC range of 20-80 and a minimum quality mean of 30, and allowing no ambiguous nucleotides.

OTU clustering was performed with the pooled merged and trimmed reads using UPARSE-OTU algorithm (usearch v7.0.1090_i86linux64) [13,14]. For this the dataset was sorted (sortbylength), dereplicated (derep_fulllength), abundance sorted (sortbysize; size = 2), and clustered (cluster_otus, minimum identity = 97%.)

ESM Table S2: Amplicon Primer to generate V3-V4 16S template library.

Primer Name	Sequence 5′- 3′
U341F_nex0	tcgtcggcagcgtcagatgtgtataagagacagCCTACGGGDGGCWGCA
U341F_nex1	tcgtcggcagcgtcagatgtgtataagagacag**N**CCTACGGGDGGCWGCA
U341F_nex2	tcgtcggcagcgtcagatgtgtataagagacag**NN**CCTACGGGDGGCWGCA
U341F_nex3	tcgtcggcagcgtcagatgtgtataagagacag**NNN**CCTACGGGDGGCWGCA
U806R_nex0	gtctcgtgggctcggagatgtgtataagagacagGGACTACHVGGGTMTCTATTC
U806R_nex1	gtctcgtgggctcggagatgtgtataagagacag**N**GGACTACHVGGGTMTCTAATC
U806R_nex2	gtctcgtgggctcggagatgtgtataagagacag**NN**GGACTACHVGGGTMTCTAATC
U806R_nex3	gtctcgtgggctcggagatgtgtataagagacag**NNN**GGACTACHVGGGTMTCTAATC

Lower case letters represent Illumina-specific overhang adapters, to which in a subsequent limited-cycle amplification step the sample specific indices and sequencing adapter attach. Frameshifting nucleotides are indicated in cursive, bold capital letters. The gene-specific sequences targeting 16S V3 and V4 region are shown in capital letters.

The clustering step detects and removes chimera sequences. In addition we also applied a reference based chimera removal step using the Green Genes database (version May 2013, http://greengenes.secondgenome.com).

The UPARSE-OTU workflow resulted in 352 OTUs. 98.7% of the reads could be successfully mapped back to OTU reference centroid sequences (usearch_global; id = 97%). The OTUs were blasted (blastn [15]) against the GreenGenes database (v13_5) [16] to assign taxonomic information with a bit-score cut-off of equal or bigger 100. Additionally, we checked taxonomic assignment of the best hit using MEGABLAST search against the NCBI nucleotide collection database The phylogenetic tree for the OTUs was built using PyNAST [17] and FastTree [18] as implemented in Qiime (v1.7) [19].

In total 178 OTUs were excluded because assigned taxonomy was either of non-bacterial origin, classified as mitochondria or chloroplast, unique to negative probes run in parallel through whole extraction and sequencing process, or unique to samples that do not belong to this data set, but were included in the previous data processing. Read counts for the same donor-recipient colony combinations were averaged (ESM Fig. S1B) and rounded, thus the final data set contained a total of 3’961’051 read counts mapping to 159 OTUs. The mean read count for the unique donor-recipient colony combinations was 127’776 (n = 31, SD = 58’551).

To show that achieved sequencing depth was adequate to reflect sample complexity we created rarefaction curves by randomly resampling the sequencing pool of each sample without replacement twenty times at different sampling depths. ESM Fig. S2A shows that the alpha diversity index (Shannon Index) of all samples reached the plateau well below the smallest library size (50’792 reads). ESM Fig. S2B plots sample richness (i.e. number of different OTUs discovered) at given subsampling depth and indicates that for most samples the majority of different OTUs were sampled.

ESM.8) Statistical analyses of microbiota community composition

We performed statistical analyses on both the non-rarefied data set, as suggested by McMurdie & Holmes [20] and on data sets rarefied to the smallest library size for comparison. Rarefaction to the smallest library size was repeated a 100 times in order to evaluate chance effects of the sampling process. For all datasets, we defined “ecologically” common and rare OTUs.

ESM Table S6: Taxonomic information for ecologically common OTUs

OTU	Kingdom	Phylum	Class	Order	Family	Genus	Species
OTU_15	Bacteria	OD1	ZB2	NA	NA	NA	NA
OTU_14	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	NA
OTU_19	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	NA	NA
OTU_1	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	NA	NA
OTU_318	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	NA	NA	NA
OTU_2	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	NA	NA	NA
OTU_103	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	NA	NA	NA
OTU_6	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales	NA	NA	NA
OTU_9	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartonellaceae	NA	NA
OTU_16	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	NA	NA
OTU_20	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	NA	NA
OTU_23	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium	NA
OTU_18	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	NA	NA
OTU_5	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	NA
OTU_11	Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bombiscardovia	NA
OTU_12	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
OTU_3	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
OTU_302	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
OTU_44	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA

ESM Table S7: Taxonomic classification of the best hit of BLAST search against NCBI database for ecologically common OTUs.

OTU	Kingdom	Phylum	Class	Order	Family	Genus	Species	Accession[1]	Bit score[1]	Identity %[1]	Source[1,2]
OTU_15	**uncultured bacterium**							JX222661.1	699	98	NA
OTU_14	**Bacteria**	**Proteobacteria**	**Gammaproteobacteria**	**Pseudomonadales**	**Pseudomonadaceae**	**Pseudomonas**	*Pseudomonas libanesis*	KT382242	793	100	compost
OTU_19	**Bacteria**	**Proteobacteria**	**Betaproteobacteria**	**Burkholderiales**	**Oxalobacteraceae**	Massilia	*Massilia suwonensis*	LN774642	793	100	air sample
OTU_1	**Bacteria**	**Proteobacteria**	**Betaproteobacteria**	**Neisseriales**	**Neisseriaceae**	Snodgrassella	*Snodgrasella alvi*	JQ746646	793	100	*Bombus bimaculatus* / bee gut
OTU_318	**Bacteria**	**Proteobacteria**	**Gammaproteobacteria**	**Pasteurellales**	Pasteurellaceae			JQ363618	734	100	*Pyrobombus hypnorum* / **digestive tract**
OTU_2	**Bacteria**	**Proteobacteria**	**Gammaproteobacteria**	Orbales	Orbaceae	Gilliamella	*Gilliamella apicola*	JQ936676	793	100	*Bombus vagans* / bee gut
OTU_103	**Bacteria**	**Proteobacteria**	**Gammaproteobacteria**	**Pasteurellales**	Pasteurellaceae			JQ363618	760	99	*Pyrobombus hypnorum* / digestive tract
OTU_6	**Bacteria**	**Proteobacteria**	**Gammaproteobacteria**					HM215025	793	100	Bumblebee / gut
OTU_9	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Rhizobiales**	**Bartonellaceae**	Bartonella	*Bartonella apis* sp. nov.	KP987884	747	100	*Bartonella apis*
OTU_16	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Rhizobiales**	**Bradyrhizobiaceae**	Bradyrhizobium	*Bradyrhizobium erythrophlei*	NR_135877	747	100	*Erythrophloeum fordii*
OTU_20	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Rhizobiales**	**Methylobacteriaceae**	Methylobacterium	*Methylobacterium aquaticum*	AP014704	747	100	*Methylobacterium aquaticum*
OTU_23	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Rhizobiales**	**Methylobacteriaceae**	**Methylobacterium**	*Methylobacterium oryzae* CBMB20	CP003811	747	100	Rice
OTU_18	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Caulobacterales**	**Caulobacteraceae**	Brevundimonas	*Brevundimonas* sp. ADMK76	KU851032	747	100	soil
OTU_5	**Bacteria**	**Proteobacteria**	**Alphaproteobacteria**	**Sphingomonadales**	**Sphingomonadaceae**	**Sphingomonas**	*Sphingomonas* sp. Sph10	KP866800	747	100	membrane biofilm
OTU_11	**Bacteria**	**Actinobacteria**	**Actinobacteria**	**Bifidobacteriales**	**Bifidobacteriaceae**	**Bifidobacterium**		KC477410	756	100	*Bombus terrestris* / gut
OTU_12	**Bacteria**	**Firmicutes**	**Bacilli**	**Lactobacillales**	**Lactobacillaceae**	**Lactobacillus**	*Lactobacillus sp.* G5_12_5MO2	KF600199	793	100	*Apis mellifera* / hindgut
OTU_3	**Bacteria**	**Firmicutes**	**Bacilli**	**Lactobacillales**	**Lactobacillaceae**	**Lactobacillus**	*Lactobacillus bombicola*	LK054485	793	100	Bumblebee / gut
OTU_302	**Bacteria**	**Firmicutes**	**Bacilli**	**Lactobacillales**	**Lactobacillaceae**	**Lactobacillus**	*Lactobacillus melliventris*	KM068135	726	97	*Apis mellifera* / digestive tract
OTU_44	**Bacteria**	**Firmicutes**	**Bacilli**	**Lactobacillales**	**Lactobacillaceae**	**Lactobacillus**	*Lactobacillus apis*	NR_125702	787	99	*Apis mellifera* / digestive tract

Taxonomic classification in bold corroborates bastn results against the currated 16S GreenGenes database

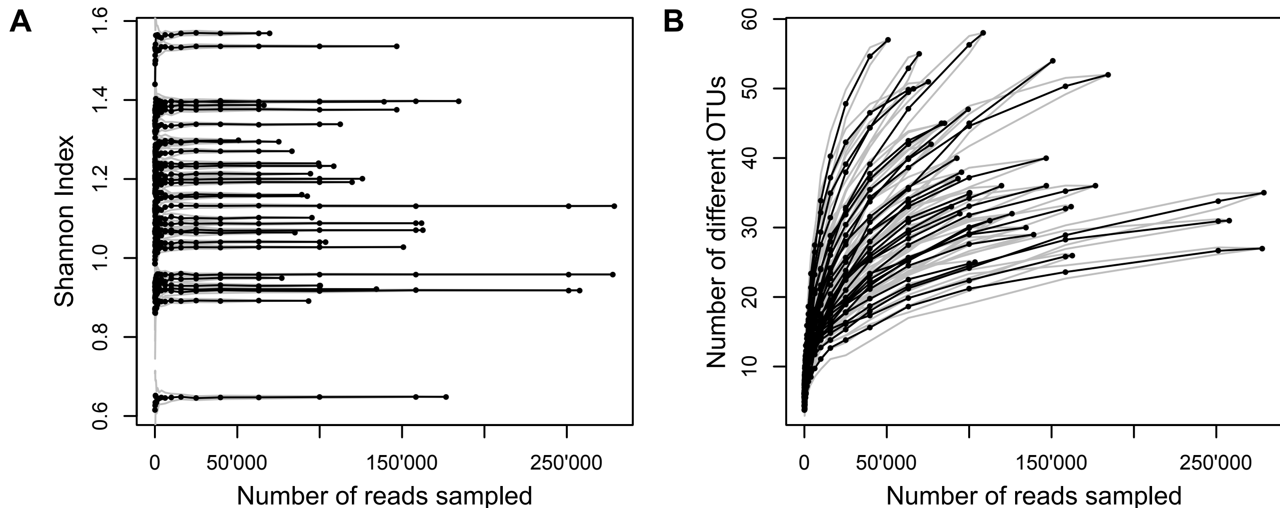
[1]Top hit of the first one hundred BLASTN (2.3.1+) hits against NCBI nucleotide collection database (nr/nt; accessed 5 April 2016); given the same hit quality, preference was given to full lenght 16S sequences

[2]Indicates host and/or isolation source

ESM Table S8: Representative sequence for each ecologically common OTU

OTU	Sequence[1]
OTU_15	GTCGAGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCGACGCCGCGTGGTGGATGAAGTCCTTAGGGACGTAAACACCTTTTATGAGGGAGAAAGTATATTGATGTTACCTCATGAATAAGGGGCTCCCAACTCTGTGCCAGCAGGAGCGGTAATACA GAGGCCCCAAGCATTATCCGGATTTACTGGGCGTAAAGGGTGCGTAGGCGGTTATATTAGTCGGGTGTTAAATCCTGAGGCTCAACCTCAGGCTCGCATTCGAAACGGTATAACTAGAAGGAGTCAGAGGCAAGCAGAACTCATGGTGTAGGGGTGAAAT CCGTTGATATCATGGGGAATACCAAATGCGAAGGCAGCTTGCTGGGACTTTCTTGACGCTGAGGCACGAAAGCGTGGGTAGCGAATCG
OTU_14	GTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAATTAATACTTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGT GCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCAAAACTGACGAGCTAGAGTATGGTAGAGGGTGGTGGAA TTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACA
OTU_19	GTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGAGTGAAGAAGGCCTTCGGGTTGTAAAGCTCTTTTGTCAGGGAAGAAACGGTAGAGGCTAATATCCTTTGCTAATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTTTTGTAAGTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATTGCGATGGAGACTGCAAGGCTAGAATCTGGCAGAGGGGGGTAGAA TTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGTCAAGATTGACGCTCATGCACGAAAGCGTGGGGAGCAAACAG
OTU_1	GTGGGGAATTTTGGACAATGGGGGGAACCCTGATCCAGCCATGCCGCGTGTCTGAAGAAGGCCTTCGGGTTGTAAAGGACTTTTGTTAGGGAAGAAAAGCTGGGTGCTAATACCATCTAGTGCTGACGGTACCTAAAGAATAAGCACCGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGAGCGCAGACGGTTAAATAAGTCAGATGTGAAATCCCCGAGCTCAACTTGGGACGTGCATTTGAAACTGTGTAACTAGAGTGTGTCAGAGGGAGGTAGAA TTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAATACCGATGGCGAAGGCAGCCTCCTGGGATAACACTGACGTTCATGCTCGAAAGCGTGGGTAGCAAACAG
OTU_318	GTGGGGAATATTGCATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCGGTGATGAGGAAGGCGGTGTATCTAATAGGTGCATCGATTGACGTTAATCACAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAAT ACGGAGGGTGCGAGCGTTAATCGGAATGACTGGGCGTAAAGGGCATGTAGGCGGATAATTAAGTTAGGTGTGAAAGCCCTGGGCTCAACCTAGGAATTGCACTTAAAACTGGTTAACTAGAGTATTGTAGAGGAAGGTAGAATTCCACGTGTAGCGGTGA AATGCGTAGAGATGTGGAGGAATACCGGTGGCGAAGGCGGCCTTCTGGACAGATACTGACGCTGAGATGCGAAAGCGTGGGGAGCAAACAG
OTU_2	GTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCGGTGATGAGGAAGGTGATGAATCTAATAGGTTTATTAATTGACGTTAATCACAGAAGAAGCACCGGCTAACTCCGT GCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATGACTGGGCGTAAAGGGCATGTAGGCGGATAATTAAGTTAGGTGTGAAAGCCCTGGGCTCAACCTAGGAATTGCACTTAAAACTGGTTAACTAGAGTATTGTAGAGGAAGGTAGAA TTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGGTGGCGAAGGCGGCCTTCTGGACAGATACTGACGCTGAGATGCGAAAGCGTGGGGAGCAAACAG
OTU_103	GTGGGGAATATTGCACAATGGGGGGAACCCTGATTCCTCCCTTCCTCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCGGTGATGAGGAAGGCGGTGTATCTAATAGGTGCATCGATTGACGTTAATCACAGAAGAAGCACCGGCTAACTCCGT GCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATGACTGGGCGTAAAGGGCATGTAGGCGGATAATTAAGTTAGGTGTGAAAGCCCTGGGCTCAACCTAGGAATTGCACTTAAAACTGGTTAACTAGAGTATTGTAGAGGAAGGTAGAA TTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGGTGGCGAAGGCGGCCTTCTGGACAGATACTGACGCTGAGATGCGAAAGCGTGGGGAGCAAACAG
OTU_6	GTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCGGTGATGAGGAAGGTAGTGTATTTAATAGATGCACTAATTGACGTTAATTACAGAAGAAGCACCGGCTAACTCCGT GCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATGACTGGGCGTAAAGGGCATGTAGGCGGATGATTAAGTTAGGTGTGAAAGCCCCGGGCTCAACCTGGGAATTGCATTTAAAACTGGTCGTCTGGAGTATTGTAGAGGAAGGTAGAA TTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGGTGGCGAAGGCGGCCTTCTGGACAGATACTGACGCTGAGATGCGAAAGCGTGGGGAGCAAACAG
OTU_9	GTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG GGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGATATCTTGAGTATGGAAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGAAATTCGT AGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG
OTU_16	GTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG GGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGTGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTTTGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAATTCGT AGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAACAG
OTU_20	GTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGGTTGTAAAGCTCTTTTCTCCGGGACGATAATGACGGTACCGGAGGAATAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG GGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGCGCGTAGGCGGCTGATTTAGTCGAGGGTGAAAGCCCGTGGCTCAACCACGGAATGGCCTTCGATACTGATTGGCTTGAGACCGGAAGAGGACAGCGGAACTGCGAGTGTAGAGGTGAAATTCGT AGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTGTCTGGTCCGGTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAG
OTU_23	GTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTTATCCGGGACGATAATGACGGTACCGGAGGAATAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG GGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGCGCGTAGGCGGCGTTTTAAGTCGGGGGTGAAAGCCTGTGGCTCAACCACAGAATGGCCTTCGATACTGGGACGCTTGAGTATGGTAGAGGTTGGTGGAACTGCGAGTGTAGAGGTGAAATTCGT AGATATTCGCAAGAACACCGGTGGCGAAGGCGGCCAACTGGACCATTACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAG
OTU_18	GTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAATTCTTTCACCGGGGACGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG GGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGAGCGTAGGCGGACATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCTCGGAATTGCCTTTGATACTGGGTGTCTTGAGTATGAGAGAGGTGTGTGGAACTCCGAGTGTAGAGGTGAAATTCGT AGATATTCGGAAGAACACCAGTGGCGAAGGCGACACACTGGCTCATTACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAACAG
OTU_5	GTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAAGCTCCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGG GAGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTGTAAGTTAGAGGTGAAAGCCTGGAGCTCAACTCCAGAACTGCCTTTAAGACTGCATCGCTTGAATCCAGGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGT AGATATTCGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGACTGGTATTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG
OTU_11	GTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTTAGCAAGGAGCAAGCGAGAGTGAGTGTACTTGTTGAATAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATAC GTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGGTGTGAAAGTCCACTGCTTAACGGTGGATTGGCGCCGGGTACGGGCAGGCTAGAGTGCAGTAGGGGAGACTGGAATTCCCGGTGTAACGGTGGAA TGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGTCTCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAG
OTU_12	GTAGGGAATCTTCCACAATGGACGCAAGTCTGATGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAGGATAGGGATAGTAACTGATTCTTATTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAACGCAGGCGGGAGAACAAGTCAGCTGTGAAAGCCCTCGGCTTAACCGGGGAAGTGCAGCTGAAACTGTTTTTCTTGAGTGCAGAAGAGGAGAGTGGAA CTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCTGTAACTGACGCTGAGGTTCGAAAGCATGGGTAGCGAACAG
OTU_3	GTAGGGAATCTTCCACAATGGACGAAAGTCTGATGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAAGACAAGGATAGTAACTGATTCATGTTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAACGCAGGCGGGAAGATAAGTCAGCTGTGAAAGCCCTCGGCTCAACCGGGGAACGGCAACTGAAACTATTTTTCTTGAGTGCAGAAGAGGAGAGTGGAA CTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCTGTAACTGACGCTGAGGTTCGAAAGCATGGGTAGCGAACAG
OTU_302	GTAGGGAATCTTCCACAATGGACGCAAGTCTGATGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAGGACAAGGATAGTAACTGATTTTTGTTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAACGCAGGCGGGAAGATAAGTCAGCTGTGAAAGCCCTCGGCTCAACCGGGGAAGTGCAGCTGAAACTATATTTCTTGAGTGCAGAAGAGGAGAGTGGAA CTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCTGTAACTGACGCTGAGATGCGAAAGCGTGGGGAGCAAACAG
OTU_44	GTCGGGAATCTTCCACAATGGACGCAAGTCTGATGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAGGACATGGGTAGTAACTGATCTATGTTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGT GCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAACGCAGGCGGGAGAACAAGTCAGCTGTGAAAGCCCTCGGCTTAACCGAGGAACGGCAACTGAAACTGTTTTTCTTGAGTGCAGAAGAGGAGAGTGGAA CTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCTGTAACTGACGCTGAGGTTCGAAAGCATGGGTAGCGAACAG

[1]centroid sequences



ESM Fig. S2: Rarefaction curves each sample. (A) shows the Shannon-Wiener diversity index, and (B) number of different OTUs (richness) at given sampling depth (x-axis). Lines connect means and SD (grey lines) of 20 independent draws at a given sampling depth for a given library.

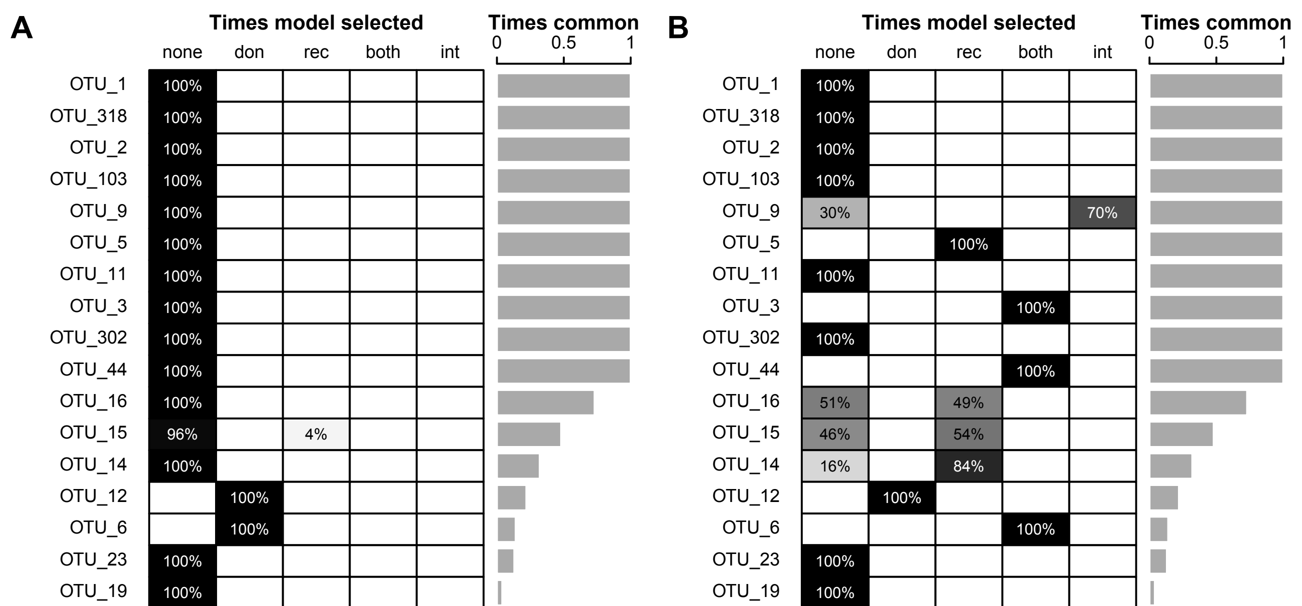
Common OTUs were observed in at least 85% of the samples and the representative sequences are reported in ESM Table 8. We tested, similar to gene expression analysis, for effects of recipient and donor resistance phenotypes on the microbiota community structure.

To investigate potential effects of donor and/or recipient resistant phenotypes on the total proportion of reads defined as common OTUs, we fitted a generalized linear model (glm) with a quasibinomial error distribution using a logit link function to the count data. The model weights the proportions according to sample size, thus controlling for differences in sequencing depth. We performed backwards model selection (i.e. retaining only independent variables with a p-value ≤ 0.1) on the full model: recipient phenotype (resistant, susceptible) * donor phenotype (resistant, susceptible), using F-tests to find the minimal adequate model.

To test for differential abundance of common OTUs, we fitted a negative binomial generalized linear model of the form: recipient phenotype (resistant, susceptible) * donor phenotype (resistant, susceptible), as described in Love *et al.* [21] to the non-rarefied dataset. Briefly, the modelling process statistically accounts for differences in sequencing depth, applies independent filtering to increase detection power, and uses Wald-tests to test for significant coefficients (i.e. log2-fold changes) in the negative binomial generalized linear model. We accounted for false discovery rate [8] at a cut-off of $\alpha = 0.05$.

We use linear discriminant analysis on regularized logarithm transformed count data to find the linear combination of common OTUs that best discriminates between phenotypes. Influential OTUs were identified based on partial correlation coefficients of the linear discriminant function and leave-one-out cross-validation (jack-knifed) assessed accuracy of phenotype discrimination by the discriminant function.

Complementary to the above analysis, we applied to each rarefied data set a negative binomial generalized linear model and performed backwards model selection from the initial full model: recipient phenotype (resistant, susceptible) * donor phenotype (resistant, susceptible), using chi-square tests to identify the minimal adequate model for each common OTU within a rarefied data set. We accounted for false discovery rate [8] at a cut-off of $\alpha = 0.05$ (ESM Fig. S3).



ESM Fig. S3: Differential abundance in microbiota – analysis of rarefied data sets. Panel (A) shows a result summary with correction for false discovery rate, panel (B) shows a result summary without correction. Shown are the percentage of times (grey-scaled) a particular model was selected when it explained significantly more variation with the addition of an explanatory variable (None = model with intercept only; don = model wit donor main effect; rec = model with recipient main effect; both = donor and recipient main effects; int = full model with main effects and interaction). Grey bars to the right illustrate how often a OUT was defined as common OTU after the rarefaction process (1 = always common, 0 = never common, n=100).

References:

1. Schmid-Hempel, P. & Schmid-Hempel, R. 1993 Transmission of a pathogen in Bombus terrestris, with a note on division of labour in social insects. *Behav Ecol Sociobiol* **33**, 319–327. (doi:10.1007/BF00172930)

2. Korner, P. & Schmid-Hempel, P. 2004 In vivo dynamics of an immune response in the bumble bee Bombus terrestris. *J Invertebr Pathol* **87**, 59–66. (doi:10.1016/j.jip.2004.07.004)

3. Ulrich, Y., Sadd, B. M. & Schmid-Hempel, P. 2011 Strain filtering and transmission of a mixed infection in a social insect. *J Evol Biol* **24**, 354–362. (doi:10.1111/j.1420-9101.2010.02172.x)

4. Schmid-Hempel, P. & Reber Funk, C. 2004 The distribution of genotypes of the trypanosome parasite, Crithidia bombi, in populations of its host, Bombus terrestris. *Parasitology* **129**, 147–158. (doi:10.1017/S0031182004005542)

5. Koch, H. & Schmid-Hempel, P. 2011 Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci USA* **108**, 19288–19292. (doi:10.1073/pnas.1110474108)

6. 1999 High-dose irradiation: wholesomeness of food irradiated with doses above 10 kGy. Report of a Joint FAO/IAEA/WHO Study Group. *World Health Organ Tech Rep Ser.* **890**, i–vi– 1–197.

7. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19. (doi:10.1186/gb-2007-8-2-r19)

8. Benjamini, Y. & Hochberg, Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Stat Methodol)* (doi:10.2307/2346101)

9. Oksanen, J. et al. In press. vegan: Community Ecology Package. R package version 2.2-1.

10. Liu, C. M., Aziz, M., Kachur, S., Hsueh, P.-R., Huang, Y.-T., Keim, P. & Price, L. B. 2011 BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. *BMC Microbiol* **12**, 56–56. (doi:10.1186/1471-2180-12-56)

11. Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F. O. 2013 Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation se-

quencing-based diversity studies. *Nucleic Acids Res* **41**, e1–e1. (doi:10.1093/nar/gks808)

12. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. 2013 Practical innovations for high-throughput amplicon sequencing. *Nat Methods* **10**, 999–1002. (doi:10.1038/nmeth.2634)

13. Edgar, R. C. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. (doi:10.1093/bioinformatics/btq461)

14. Edgar, R. C. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**, 996–998. (doi:10.1038/nmeth.2604)

15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J Mol Biol* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)

16. McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R. & Hugenholtz, P. 2011 An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**, 610–618. (doi:10.1038/ismej.2011.139)

17. Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. & Knight, R. 2010 PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267. (doi:10.1093/bioinformatics/btp636)

18. Price, M. N., Dehal, P. S. & Arkin, A. P. 2010 FastTree 2-Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**. (doi:10.1371/journal.pone.0009490)

19. Caporaso, J. G. J. et al. 2010 QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336. (doi:10.1038/nmeth.f.303)

20. McMurdie, P. J. & Holmes, S. 2014 Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* **10**, e1003531. (doi:10.1371/journal.pcbi.1003531)

21. Love, M. I., Huber, W. & Anders, S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550. (doi:10.1186/s13059-014-0550-8)