doi:10.1038/nature10571

Supplementary Methods

Quartet mapping

To test whether phylogenetic reconstruction supports our inference of HGT, we performed quartet mapping, in which all possible four member trees are generated and analyzed to simulate analysis of the larger and more computationally challenging parent tree. We followed a similar approach to the quartet mapping described by Daubin and Ochman [30]. Briefly, we searched all 2,235 genomes in our analysis for homologs to each HGT event (defined as best reciprocal BLAST hits with > 60% nucleotide identity over > 60% of the length of the transferred gene; see note on homology below). For HGT events with at least two homologs, we used MUSCLE (with default settings) to construct an alignment of the HGT sequences and all other non-HGT sequences. Events with fewer than two non-HGT homologs - 23% of the total - cannot be used to generate a quartet and so could not be analyzed by quartet mapping. For the quartets that remained, we used Tree Puzzle to analyze all possible quartet topologies among the aligned HGT and non-HGT sequences. With Tree Puzzle we used exact parameter estimates and gamma distributed rates with four rate categories. To provide phylogenetic confirmation of our putative HGT events, we computed the likelihood of obtaining a quartet grouping the HGT events together, versus the alternative, vertical model that would group sequences by the topology of the species phylogeny. A previously published likelihood ratio [30] was then used to place phylogenetic confidence in each HGT event. We used the most stringent confidence threshold possible, requiring a likelihood ratio of 1.0 to support HGT inference. With this conservative approach, more than 99% of the HGTs we analyzed were supported.

Gene loss analysis

We explored whether vertical inheritance is a plausible alternative explanation for each inferred HGT by determining the minimum number of independent loss events that would be needed to support a model of vertical inheritance. We mapped all inferred transfers and their homologs to the IMG species tree and calculated the number of independent loss events that would be required to explain the sparse phylogenetic distribution of these events. Here, we define homologs as best BLAST hits with > 90% identity and > 80% length (see note below). These parameters allow for considerable variation in evolutionary rates within the gene family.

As shown in Supplementary Fig. 1, for the majority of HGT events, over 100 independent loss events would be required to accept a model of vertical descent. To contextualize this remarkable observation, most parsimony based HGT detection tools use an empirically derived estimate of approximately 3:1 as the parsimony cost of losses relative to HGT [31]. Using this 3:1 parsimony metric, over 99% of our events can be explained by HGT.

Note on the detection of homologs

We varied the parameters that define homology for the two approaches above in order to maximize our ability to detect vertical transmission. We used an especially permissive definition of homology for quartet mapping to allow a maximal number of potentially homologous genes to disrupt the pairing of the putatively transferred sequences, thereby

increasing the opportunity to return a quartet that does not support HGT. We employed a more moderate definition of homology for the loss analysis to avoid spuriously inserting unrelated proteins that may have appeared as false loss events.

Controlling for contamination

To control for the potential effect of contamination derived from genomes processed at the same sequencing facility, we repeated our principal analysis, but only compared genomes sequenced at different facilities. This restricted analysis confirmed that our main findings are not caused by contamination between projects at the same sequencing center. In Supplementary Fig. 4, we show that there is more HGT among human-associated bacteria than among non-human associated bacteria, across all phylogenetic distances. The enrichment in HGT among bacteria occupying the same body site relative to bacteria occupying different body sites is similarly replicated in this restricted analysis (as found in Figure 1 of the main text).

In Supplementary Fig. 5, we also show that the most distantly related comparisons with shared ecology continue to exchange more DNA than the most closely related comparisons with different ecology when only HGT between sequencing centers is allowed (as found in Figure 2 of the main text).

Controlling for cosmopolitan genomes

To control for the potential effect of cosmopolitan genomes that inhabit multiple environments, we repeated our principal analysis, excluding all genome clusters containing at least two representatives from different body sites, hosts, or other environmental categories. This removed cosmopolitan groups of organisms like *Escherichia coli*, which is found in the gut, skin, blood, and non-human environments for example. This restricted analysis robustly yields the pattern of ecological enrichment found in the main text (Supplementary Fig. 6).

Limitation of HGT detection

Our method is only able to detect horizontal gene transfer between distantly related lineages. Another limitation is that our method can only detect recent events that share 99% nucleotide identity. Consequently the dynamics discussed in our analysis may not apply to more ancient HGT or to HGT between less divergent strains. However, because a stringent phylogenetic distance cut-off is used to inform each HGT classification our method avoids many of the limitations of previous BLAST-based approaches to HGT detection [32].

Limitations of geographic inference

There are a few important caveats to consider when reviewing our geographic findings. First, due to limited sample size, we only explored the effects of geography at continental scales. It is possible that strong effects may persist at finer spatial scales, although these may be primarily driven by ecological overlap, which is difficult to distinguish from local geography. Second, the location of isolation is only a proxy for the overall geographic range of a sequenced strain. When a strain is isolated from a particular site, it may have a

range that extends across a much larger geographic range, obscuring the validity of geographic inference from a single sample.

Annotation of mobile genetic elements

For this analysis we were interested in exploring the approximate magnitude of mobile elements relative to other functional groups. In the interests of defining the minimum number of mobile elements in our analysis, we chose a rapid and highly specific method at the expense of sensitivity. We aligned all transferred sequences to the NCBI nr database using BLASTX. We extracted the annotations for the best BLASTX hit in nr (with an e-value of $e < 1E\text{-}50$). Next we used keyword search text mining coupled with manual curation to count the frequency of each functional category. Our keywords are designed to reduce false positives - we understand that valid mobile elements may not be detected with this simple approach.

The keywords used to identify each functional group are listed below (case sensitive):

Transposons: transpos*, TN, insertion element, is element, IS element
Phage: phage, tail protein, tegument, capsid
Plasmid: relaxase, conjugal transfer, Trb, relaxosome, Type IV secretion, conjugation, Tra[A-Z], Mob[A-Z], Vir[A-Z][0-9], t4ss, T4SS, resolvase
Other MGE: recombinase, integrase

The percent of total proteins (27%) is calculated by counting each of the functional classes as a fraction of all transferred sequences. In order to account for redundancy in the set, we extract the NCBI gene identifier for the best BLASTX hit for each transferred sequence. We then remove all redundancies from this list of gene identifiers and count the fraction of unique gene identifiers that fall into each of the functional classes described. This analysis suggests that a relatively small group of promiscuous mobile elements accounts for a large fraction of total transferred sequences.

Definition of environments

Farm samples are taken directly from animals used in agriculture (horse, cow, sheep, goat, pig). As with human subjects, samples from animals vary (blood, stool, rumen etc). Metadata to define environments, such as isolation site, oxygen tolerance, and pathogenicity were downloaded from IMG[27].

Treatment of ambiguous metadata annotations

We only consider genome comparisons for which we have appropriate metadata. For genomes with partial metadata (ie oxygen tolerance is annotated, but continent and disease are missing), we include the genome when possible (for oxygen tolerance) and ignore it in other analyses (continent and disease).

When comparing the frequency of HGT in the same environment with the frequency of HGT between different environments it is necessary to handle ambiguous genome annotations with multiple annotated environments (e.g. gut and skin). In these cases, we consider this strain once for each metadata label. Thus when a strain from the gut is

compared to a strain annotated as gut and skin, this comparison will contribute to both comparisons of gut-gut transfer and gut-skin transfer.

Computation of error bars

Error bars reflect our estimated uncertainty in the sampling of a binomial random variable (the observation of HGT). We compute error bars as the standard deviation in %HGT by modeling the total number of transfers as a binomial random variable with parameters p and n. We take n to be the number of independent species cluster comparisons and we estimate p as the total %HGT observed at each phylogenetic distance. From these considerations, it follows that the variance is given by $\mathrm{Var}[\%\mathrm{HGT}] = p(1-p)/n$ which is used to calculate the standard deviation at each distance bin.

Counting HGT

When measuring the frequency of HGT between environments we only consider the fraction of genomes that share at least one HGT. We do not consider the length of a transfer because high variance in event length would add significant noise to our results and overweight rare, large transfer events that do not reflect evolutionary independence. We do not consider the number of distinct regions of HGT shared between two genomes because transposition or poor assembly might falsely inflate this metric by splitting a single large event into many smaller apparent events.

In Fig. 3 of the main text, HGT is computed as the average across all distance bins in contrast to Fig. 1, where HGT is computed in separate distance bins. As a result, the frequencies of HGT cannot be directly compared between the two figures.

Clustering similar genomes

In order to avoid over-counting transfers, we use average linkage clustering to group similar genomes (with < 2% average 16S divergence). This ensures that transfers between clusters reflect evolutionary independence and avoids the problem of counting a single transfer in a densely sampled lineage many times. All comparisons discussed in the text reflect transfers across clusters constructed in this manner.

Because the sequenced flexible genome is larger when more isolates from a single cluster are considered, the probability of observing at least one transfer between two clusters with many sequenced isolates is greater than between two clusters with fewer sequenced isolates. To account for this effect, for each cluster comparison we consider the fraction of genomes that share an HGT. We equally weight all genome comparisons between two clusters. If 50% of a genome cluster has a hit with at least one member of another genome cluster, we consider this cluster comparison as 50% of an HGT.

Statistical methods

To test for overall enrichment in HGT between two metadata labels (e.g. human vs. non-human) we perform separate statistical tests for enrichment within each phylogenetic distance bin, then combine these test results into a single p-value using Fisher's method. Within each phylogenetic distance bin, we determine if there is a significant difference in HGT frequencies between all pairs of genome clusters belonging to the two different

metadata labels. With our counting and clustering protocols (described above), we create two vectors (each corresponding to a metadata label) of HGT frequencies (with continuous values) that we compare with a Mann-Whitney U-test. This approach is applied to assess differences in observed frequencies of HGT and to assess the statistical significance of the data underlying Fig. 1, Fig. 2, and Fig. 3 in the main text. This approach controls for the effect of phylogeny by restricting comparisons of HGT frequency to isolates of similar phylogenetic divergences (distance bins of 1% 16S distance).

After establishing the strong effect of body-site on HGT frequency in the human microbiome, further analyses (such as oxygen tolerance and pathogenicity as in Fig. 4, main text) must control for both the effects of phylogeny and body-site. We achieve this by calculating the frequency of HGT for all possible combinations of body-sites and phylogenetic divergences. For example, the expected value for skin-gut transfer at 3-4% 16S divergence is the average of all observations that meet these metadata criteria. Our null model assumes that further constraining our analysis with additional metadata labels will not lead to values that deviate from these expected values. To test this model, we compare the expected value to the observed frequency of HGT when the analysis is further conditioned on a new metadata label (e.g. anaerobes in skin and gut at 3-4% 16S divergence). We determine whether this further metadata constraint is associated with elevated HGT by using a Chi-Square test to compare the expected values with the observed values.

Public access to identified HGT sequences and Metadata

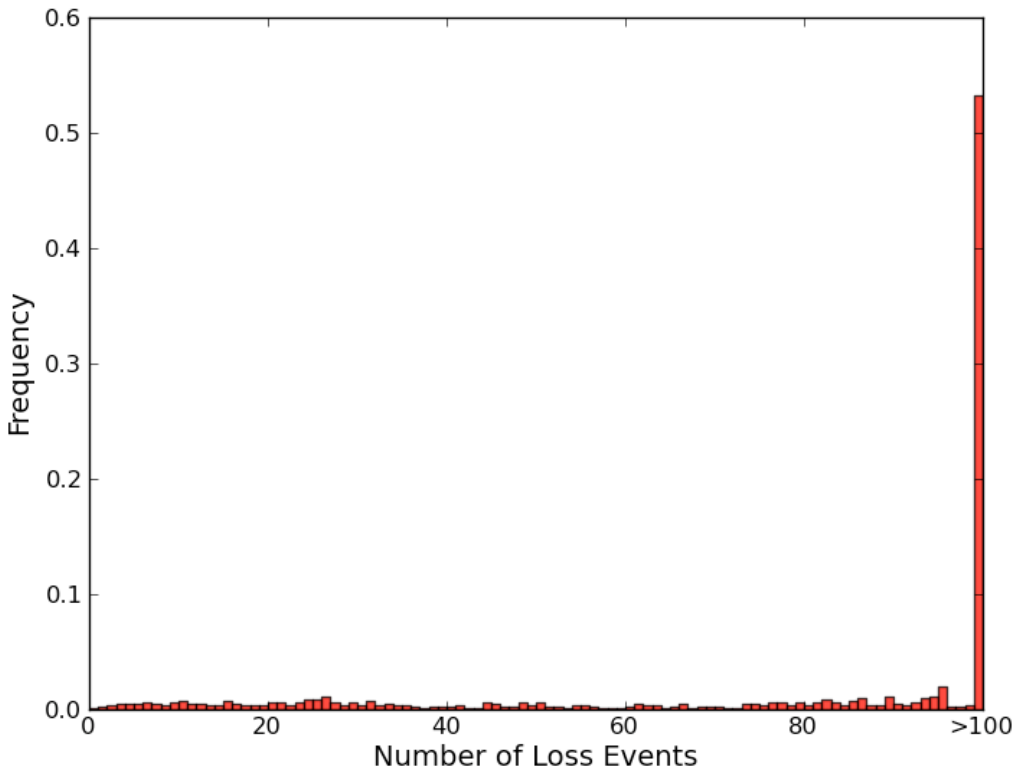
A FASTA file with all horizontal gene transfers identified in this study is publicly available at the following location:
http://almlab.mit.edu/sites/default/files/hgt_seqs.zip

A text file with a complete list of genomes used in our analysis and the corresponding metadata labels (originally from IMG) is available here:
http://almlab.mit.edu/sites/default/files/Metadata.txt

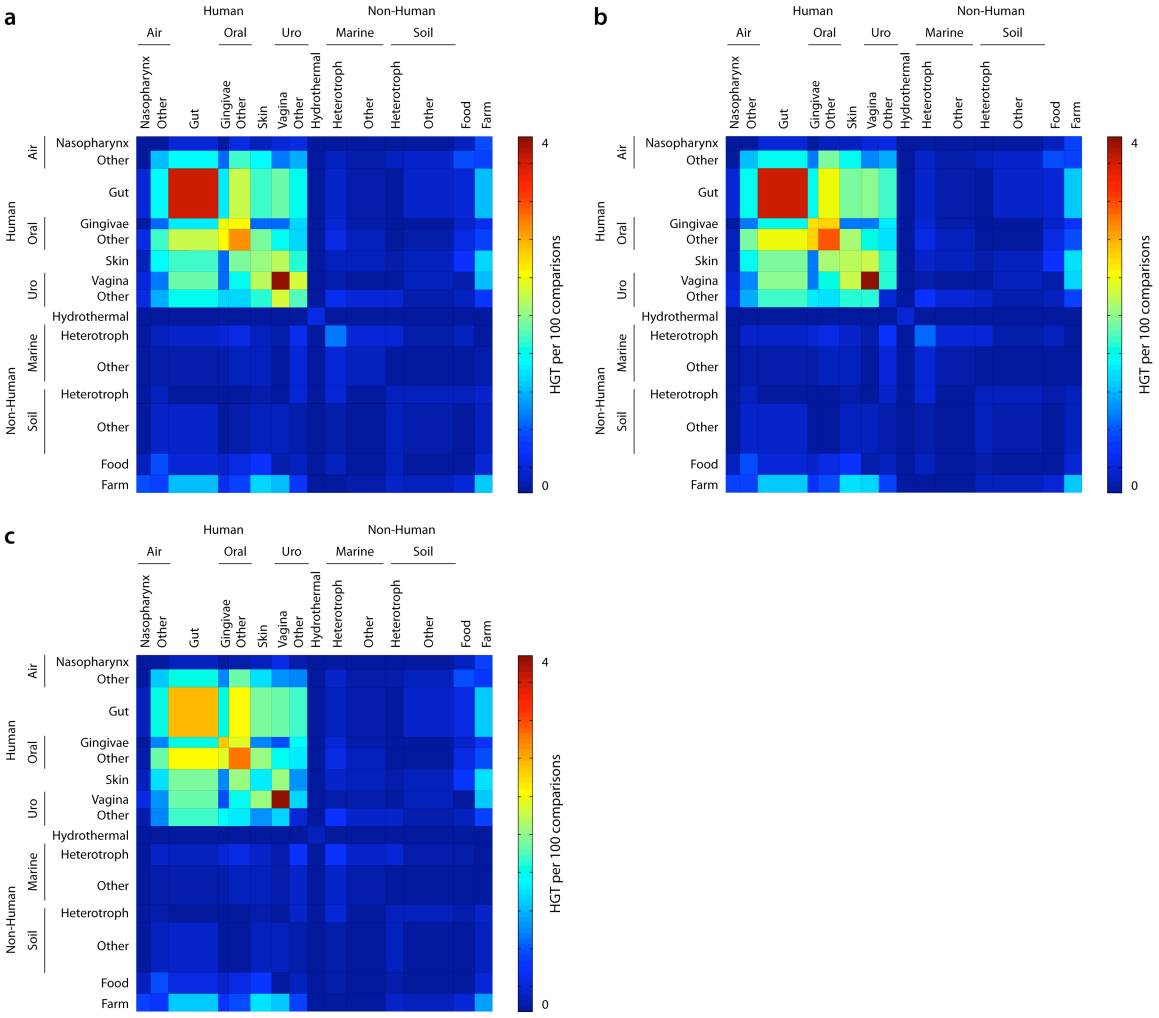
Supplemental References

30. Daubin, V. & Ochman, H. Quartet Mapping and the Extent of Lateral Transfer in Bacterial Genomes. *Molecular Biology and Evolution* **21**, 86 -89 (2004).
31. David, L.A. & Alm, E.J. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93-96 (2011).
32. Stanhope, M.J. *et al.* Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* **411**, 940-944 (2001).

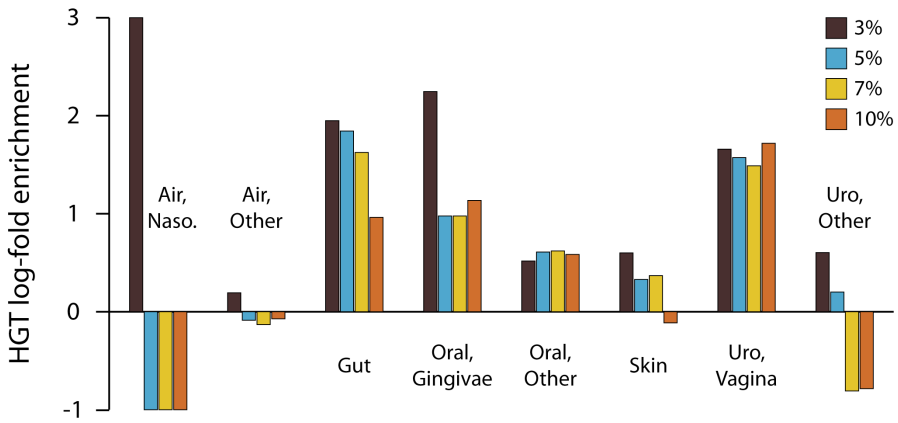
Supplemental Figures and Legends 1-6



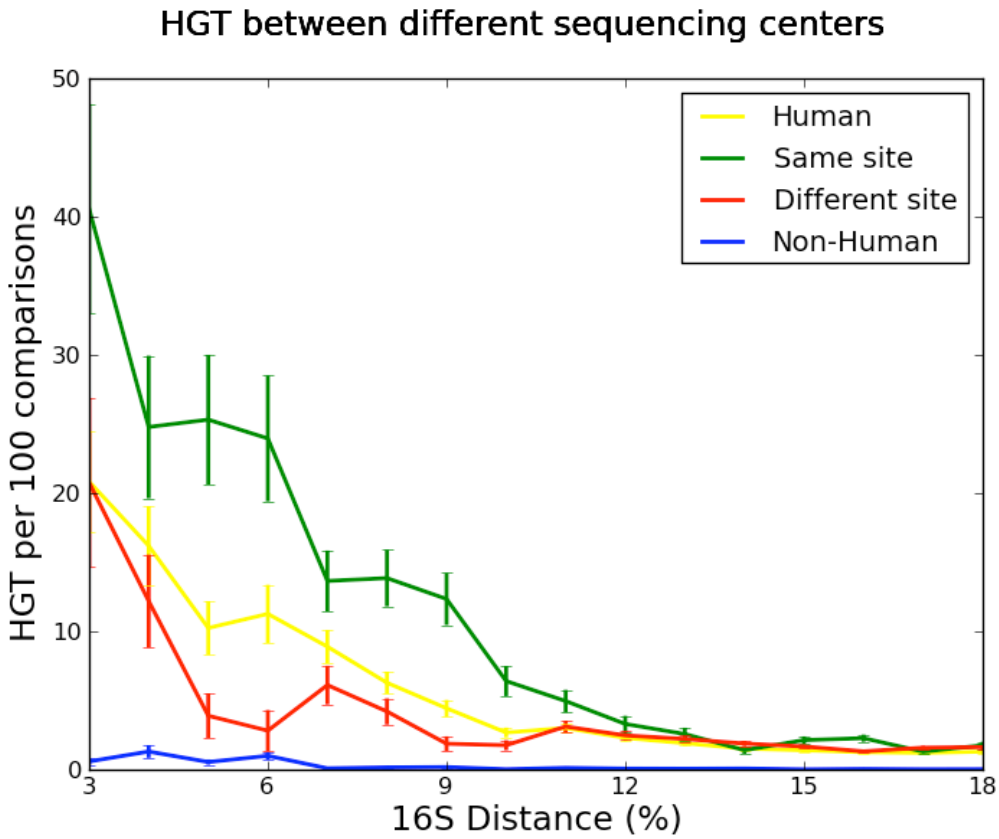
Supplemental Figure 1: The majority of inferred HGT events require over 100 independent loss events in order to accept a model of vertical descent. For each inferred transfer we map homologs onto the species tree and infer the minimum number of independent loss events needed to support a model of vertical inheritance. This figure depicts the frequency with which loss events are inferred – most inferred transfers would require extensive loss events in order to accept the alternative model of vertical transmission, supporting our approach to HGT detection.



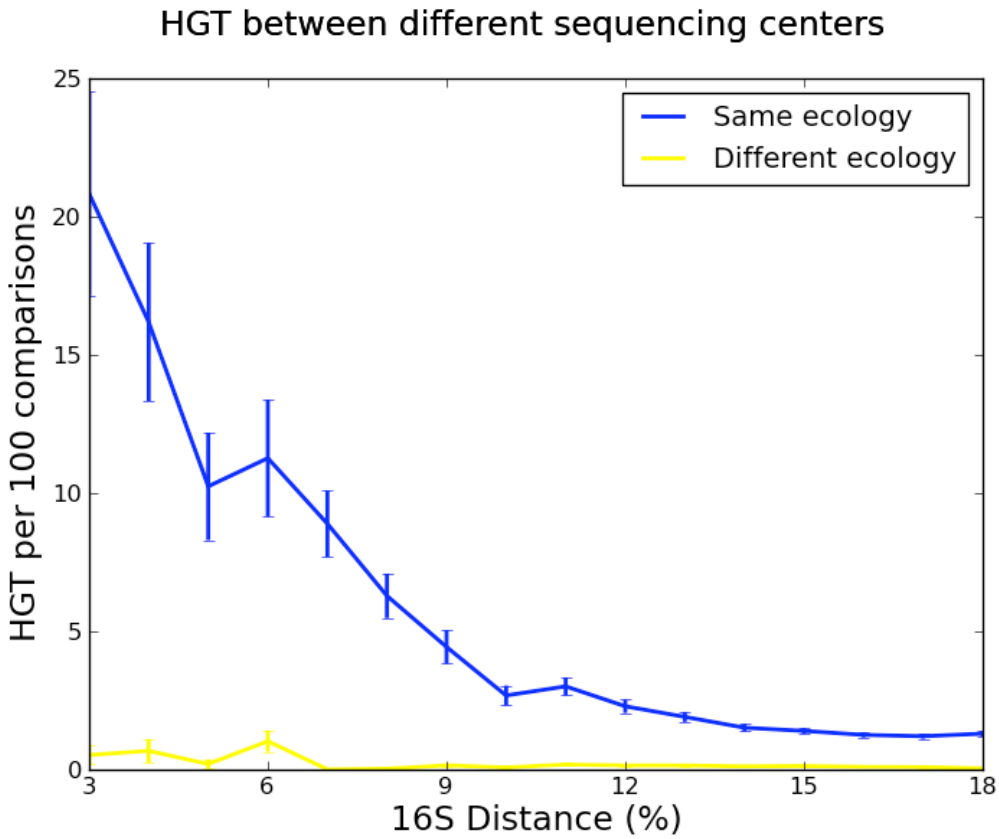
Supplemental Figure 2: Heatmap of HGT among isolates in different environments at 5%, 7% and 10% 16S divergences. This figure shows the frequency of HGT between each of the environments included in this study across three different distance cutoffs, in addition to the overall plot shown in Figure 3a of the main text. Each distance cutoff includes all comparisons satisfying the given separation criteria (e.g. 5% includes comparisons of all clusters of bacteria separated by at least 5% 16S rRNA divergence). Although the specific values of enrichment vary across different distance cutoffs, the overall pattern of human, body site and body sub-site enrichment persists across all distance groupings. We show only the heatmap for all gene classes (excluding the inset heatplot for antibiotics that appears in the main text Figure 3c) because there are insufficient counts to yield reliable estimates for rates of long distance transfer when only antibiotic resistance genes are considered.



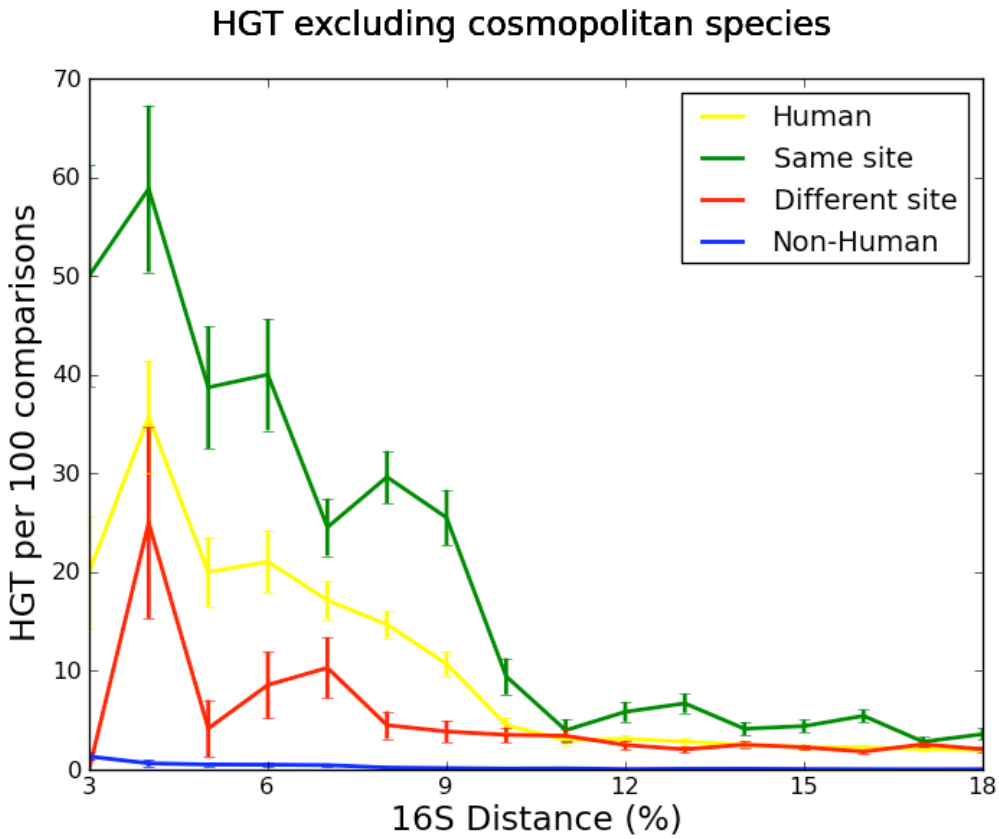
Supplemental Figure 3: Barplot of HGT for each body site at 3%, 5%, 7% and 10% distance cutoffs. This figure summarizes the persistence of body-site and sub-site enrichment across four distance cutoffs. As in Supplemental Figure 2, distance cutoffs reflect all comparisons with at least the given 16S distance. The log-fold enrichment indicated on the vertical axis describes the ratio of observed transfers within the given body site at each distance relative to HGT among all human isolates at the same phylogenetic distance cutoff. The poorly sampled nasopharynx (n = 25) and non-vaginal urogenital sites (n = 46) are the only categories for which the enrichment in transfer does not persist across phylogenetic distances (likely due to uncertainty arising from small sample sizes). Otherwise, the majority (n = 480) of isolates belong to body sites for which enrichment persists across all observed distances.



Supplemental Figure 4: Ecological structure persists when only genome comparisons from different sequencing centers are allowed. We compute the frequency of transfer within human associated isolates (yellow), non-human isolates (blue) human isolates from the same body site (green) and human isolates from different body sites (red), while only allowing genome comparisons between different genome sequencing centers. This controls for contamination that might arise in the sequencing and assembly process.



Supplemental Figure 5: Ecology is the dominant force shaping recent HGT in the human microbiome, even when HGT is only allowed between different sequencing centers. This figure compares the effects of ecology relative to phylogeny on HGT, when HGT is only allowed between different sequencing centers. The frequency of HGT is shown among human-associated isolates (same ecology, blue) and between human-associated and non human-associated isolates (different ecology, yellow). Even the most distantly related bacteria with shared ecology engage in more HGT than the most closely related bacteria with different ecology when we control for contamination caused by sequencing projects from the same sequencing center.



Supplemental Figure 6: Ecological structure persists when cosmopolitan species are excluded. We compute the frequency of transfer within human associated isolates (yellow), non-human isolates (blue) human isolates from the same body site (green) and human isolates from different body sites (red), while excluding species that are present in multiple environments (cosmopolitan species). This controls for the potential confounding effect of cosmopolitan species.

Supplementary Tables and Legends 1-7

Supplementary Table 1: Recently transferred genes shared by meningitis isolates. Forty-four transposon proteins have been filtered from the list of genes listed below to reduce functional redundancy and create a more concise table.

E-value	Accession Number	Annotation
0	AAB60031.1	ORF3
0	AAC36982.1	ORF2
0	ADL64465.1	conserved hypothetical protein
0	CBJ57220.1	hypothetical protein
0	CBJ57221.1	hypothetical protein
0	EFU87609.1	putative ATP/GTP-binding protein
0	NP_269362.1	putative phosphoglucomutase
0	NP_687937.1	tetracycline resistance protein
0	NP_688256.1	mercuric reductase
0	NP_689009.1	mercuric reductase
0	NP_735546.1	hypothetical protein gbs1100
0	NP_802984.1	hypothetical protein SPs1722
0	NP_838069.1	putative integrase
0	YP_003028246.1	tetracycline resistance protein TetM 1
0	ZP_00787447.1	tetracycline resistance protein (tetM)
0	ZP_01786915.1	mercuric reductase
0	ZP_03079517.1	NLP/P60 family protein
0	ZP_07467524.1	tetracycline resistance protein TetP
0	ZP_07473555.1	phage-related integrase
0	ZP_07568822.1	putative ATP/GTP-binding protein
1.00E-107	NP_838057.1	resolvase
1.00E-113	ZP_05682278.1	plasmid recombination enzyme
1.00E-135	CBA06498.1	putative phage protein
1.00E-141	CAE53425.1	dihydropteroate synthase
1.00E-153	ZP_06864307.1	aminopeptidase
1.00E-171	NP_269968.1	putative laminin adhesion
1.00E-176	CBY90211.1	haemagglutinin/hemolysin-related protein TpsA
1.00E-61	YP_002342940.1	hypothetical protein NMA1636
1.00E-61	YP_975359.1	hypothetical protein NMC1358
1.00E-62	YP_002995720.1	hypothetical protein pRAx_0065
1.00E-75	NP_687933.1	Tn916 hypothetical protein
1.00E-79	CAX50817.1	putative NADPH-quinone dehydrogenase (modulator of drug activity B)
1.00E-83	NP_274107.1	acetylglutamate kinase
2.00E-102	YP_001155450.1	aminopeptidase N

2.00E-134	ZP_07370145.1	phenylalanyl-tRNA synthetase alpha subunit
2.00E-144	YP_001083717.1	TrbL/VirB6 plasmid conjugal transfer protein
2.00E-166	NP_838066.1	putative replication protein C
2.00E-168	ZP_04464048.1	type III restriction/modification system enzyme
2.00E-51	CBY91030.1	A/G-specific adenine glycosylase
2.00E-51	NP_687948.1	Tn916 hypothetical protein
2.00E-71	EFV63654.1	mafB1
2.00E-72	YP_004137897.1	NADPH quinone reductase
3.00E-108	NP_838057.1	resolvase
3.00E-129	ZP_03714830.1	hypothetical protein EIKCOROL_02540
3.00E-137	BAG80626.1	hypothetic protein
3.00E-168	CBY90211.1	haemagglutinin/hemolysin-related protein TpsA
3.00E-53	YP_975344.1	adenine glycosylase
3.00E-63	YP_002995720.1	hypothetical protein pRAx_0065
3.00E-81	NP_687943.1	Tn916 hypothetical protein
4.00E-119	NP_043522.1	ORF beta
4.00E-141	CAE53425.1	dihydropteroate synthase
4.00E-141	ZP_07678265.1	P-type conjugative transfer protein TrbL
4.00E-151	YP_133682.1	hypothetical protein Tn916_12
4.00E-164	NP_269968.1	putative laminin adhesion
4.00E-57	NP_687935.1	Tn916, transcriptional regulator, putative
4.00E-63	YP_975868.1	hypothetical protein NMC1940
4.00E-87	NP_838064.1	conjugal transfer protein TrbJ
4.00E-89	CBY90989.1	conserved hypothetical protein
5.00E-63	YP_003082440.1	hypothetical protein NMO_0203
5.00E-73	ZP_07851952.1	conserved domain protein
5.00E-76	NP_687933.1	Tn916 hypothetical protein
5.00E-79	CBA04893.1	arginine biosynthesis bifunctional protein
6.00E-172	NP_838061.1	putative sodium bile acid symporter family protein
6.00E-57	YP_002002524.1	hypothetical protein NGK_1899
6.00E-96	YP_975756.1	hypothetical protein NMC1815
7.00E-129	ZP_03714830.1	hypothetical protein EIKCOROL_02540
7.00E-140	CBA06498.1	putative phage protein
7.00E-60	CBA04486.1	conserved hypothetical protein
7.00E-60	CBA07330.1	conserved hypothetical protein
7.00E-60	ZP_07369440.1	fic family protein
7.00E-64	ZP_00604430.1	protein of unknown function DUF961
7.00E-80	CBA04893.1	arginine biosynthesis bifunctional protein
8.00E-134	CAX50488.1	phenylalanyl-tRNA synthetase alpha chain (phenylalanine--tRNA ligase alpha chain; PheRS)
8.00E-66	CBA04945.1	conserved hypothetical protein

8.00E-70	ZP_03079541.1	sigma-70, region 4 family
8.00E-72	YP_004137897.1	NADPH quinone reductase
8.00E-84	EFV64600.1	arginine biosynthesis bifunctional protein ArgJ
9.00E-165	ZP_07699709.1	DNA-binding helix-turn-helix protein
9.00E-167	CAA09003.1	restriction modification system-R protein
9.00E-72	CAX50553.1	hypothetical protein
9.00E-80	CBA04893.1	arginine biosynthesis bifunctional protein
9.00E-138	2FZV	Chain A, Crystal Structure Of An Apo Form Of A Flavin-Binding Protein From Shigella Flexneri

Supplementary Table 2: Recently transferred genes shared by pneumonia isolates. Thirty-three transposon proteins have been filtered from the list of genes listed below to reduce functional redundancy and create a more concise table.

E-value	Accession Number	Annotation
0	AAB60031.1	ORF3
0	AAC36982.1	ORF2
0	ADH21996.1	Tn916, tetracycline resistance protein
0	ADL64465.1	conserved hypothetical protein
0	CBJ57220.1	hypothetical protein
0	CBJ57221.1	hypothetical protein
0	EFU87609.1	putative ATP/GTP-binding protein
0	NP_044257.1	hypothetical protein R751p48
0	NP_269362.1	putative phosphoglucomutase
0	NP_370922.1	tetracycline resistance protein
0	NP_687937.1	tetracycline resistance protein
0	NP_687946.1	Tn916, FtsK/SpoIIIE family protein
0	NP_735546.1	hypothetical protein gbs1100
0	NP_775042.1	class I integron integrase
0	NP_802984.1	hypothetical protein SPs1722
0	YP_001338812.1	chloramphenicol resistance protein
0	YP_001694718.1	tetracycline resistance protein
0	YP_001694730.1	Cro/CI family transcriptional regulator
0	ZP_00787447.1	tetracycline resistance protein (tetM)
0	ZP_03079540.1	transcriptional regulator, Cro/CI family
0	ZP_07363085.1	tetracycline resistance protein TetP
0	ZP_07568822.1	putative ATP/GTP-binding protein
0	ZP_07699709.1	DNA-binding helix-turn-helix protein
1.00E-122	CBQ72270.1	related to mfs-multidrug-resistance transporter
1.00E-130	CBQ72270.1	related to mfs-multidrug-resistance transporter
1.00E-141	CAE53425.1	dihydropteroate synthase
1.00E-145	AAG45721.1	aminoglycoside 3'-(9)-O-adenylyltransferase
1.00E-145	ACJ46476.1	aminoglycoside 3"-adenylyltransferase/c-myc 9E10 epitope fusion
1.00E-151	AAF42994.1	geneticin-resistant protein
1.00E-159	NP_478145.1	aminoglycoside 3'-phosphotransferase
1.00E-161	BAA78093.1	aminoglycoside adenyltransferase
1.00E-170	YP_001816598.1	dihydropteroate synthase
1.00E-171	NP_269968.1	putative laminin adhesion
1.00E-60	CAP45531.1	adenyltransferase
1.00E-90	ZP_00603166.1	conserved hypothetical protein
1.00E-91	CAP17166.1	hypothetical protein

1.00E-97	YP_001102120.1	resolvase
2.00E-120	CAV31159.1	hypothetical protein
2.00E-125	EFU06856.1	phosphotransferase enzyme family protein
2.00E-140	YP_209330.1	hypothetical protein SC025
2.00E-145	ZP_07123872.1	nucleotidyltransferase domain protein
2.00E-149	AAG41774.1	Cat
2.00E-151	AAO21433.1	extended-spectrum beta-lactamase/aminoglycoside modifying enzyme fusion protein
2.00E-151	CAJ32372.1	beta-lactamase
2.00E-170	YP_001816598.1	dihydropteroate synthase
2.00E-51	NP_687948.1	Tn916 hypothetical protein
2.00E-63	YP_002995720.1	hypothetical protein pRAx_0065
2.00E-64	NP_687947.1	Tn916 hypothetical protein
2.00E-67	YP_001715334.1	putative GCN5-related N-acetyltransferase
2.00E-73	YP_001338843.1	hypothetical protein KPN_pKPN5p08233
2.00E-73	ZP_02226649.1	acetyltransferase, gnat family
2.00E-94	YP_001569063.1	strepthothricin acetyltransferase
2.00E-97	ZP_00603166.1	conserved hypothetical protein
3.00E-101	YP_001715361.1	2"-aminoglycoside nucleotidyltransferase aadB
3.00E-108	YP_003937697.1	aminoglycoside N(6')-acetyltransferase
3.00E-127	CBQ72270.1	related to mfs-multidrug-resistance transporter
3.00E-137	BAG80626.1	hypothetic protein
3.00E-140	YP_209330.1	hypothetical protein SC025
3.00E-151	AAO21433.1	extended-spectrum beta-lactamase/aminoglycoside modifying enzyme fusion protein
3.00E-169	YP_001816598.1	dihydropteroate synthase
3.00E-51	NP_687948.1	Tn916 hypothetical protein
3.00E-63	YP_002995720.1	hypothetical protein pRAx_0065
3.00E-81	NP_687943.1	Tn916 hypothetical protein
3.00E-81	XP_001234258.1	PREDICTED: similar to bacterial IS-element
4.00E-140	YP_209330.1	hypothetical protein SC025
4.00E-147	ZP_07123872.1	nucleotidyltransferase domain protein
4.00E-151	CAA48214.1	beta-lactamase,human IgG3 hinge fusion
4.00E-151	CAJ32372.1	beta-lactamase
4.00E-164	NP_269968.1	putative laminin adhesion
4.00E-57	NP_687935.1	Tn916, transcriptional regulator, putative
4.00E-94	YP_001569063.1	strepthothricin acetyltransferase
5.00E-105	YP_209365.1	resolvase
5.00E-133	YP_001598130.1	TnpA_rve
5.00E-136	ZP_02433069.1	hypothetical protein CLOSCI_03331

5.00E-159	NP_862254.1	AphA1-IAB
5.00E-163	CAJ77053.1	sul1delta fusion protein
5.00E-168	ZP_01974621.1	DNA integrase IntI1
5.00E-68	ZP_03754880.1	hypothetical protein ROSEINA2194_03310
5.00E-76	NP_687933.1	Tn916 hypothetical protein
6.00E-106	NP_838057.1	resolvase
6.00E-137	ZP_02433069.1	hypothetical protein CLOSCI_03331
6.00E-165	YP_002317674.1	dihydropteroate synthase
6.00E-66	YP_003446809.1	hypothetical protein smi_1707
7.00E-152	AAF42994.1	geneticin-resistant protein
7.00E-64	ZP_00604430.1	protein of unknown function DUF961
8.00E-142	CAE53425.1	dihydropteroate synthase
8.00E-59	ZP_06061528.1	conserved hypothetical protein
8.00E-64	ZP_00604430.1	protein of unknown function DUF961
8.00E-70	ZP_03079541.1	sigma-70, region 4 family
8.00E-97	YP_187540.1	streptothricin acetyltransferase
9.00E-112	CAJ77031.1	sul1delta fusion protein
9.00E-141	YP_209330.1	hypothetical protein SC025
9.00E-165	ZP_07699709.1	DNA-binding helix-turn-helix protein
9.00E-65	YP_003446809.1	hypothetical protein smi_1707
4.00E-145	1L8T	Chain A, Crystal Structure Of 3',5"-Aminoglycoside Phosphotransferase Type Iiia Adp Kanamycin A Complex

Supplementary Table 3: Recently transferred genes shared by endocarditis isolates. Forty-two transposon proteins have been filtered from the list of genes listed below to reduce functional redundancy and create a more concise table.

E-value	Accession Number	Annotation
0	AAB60031.1	ORF3
0	ADL64465.1	conserved hypothetical protein
0	CBI48350.1	conserved hypothetical protein
0	CBI50745.1	mercuric reductase
0	NP_040422.1	tetracycline resistance protein
0	NP_040434.1	replication protein
0	NP_043526.1	RepU
0	NP_370922.1	tetracycline resistance protein
0	NP_370926.1	putative ATP/GTP-binding protein
0	NP_370932.1	putative transcriptional regulator
0	NP_370933.1	FtsK/SpoIIIE family protein
0	NP_687937.1	tetracycline resistance protein
0	NP_763635.1	regulatory protein
0	NP_763640.1	dihydrolipoamide dehydrogenase
0	YP_001198287.1	translation elongation factor (GTPases)
0	YP_001198319.1	DNA recombinase, putative
0	YP_002559372.1	bifunctional aminoglycoside N-acetyltransferase and aminoglycoside phosphotransferase
0	YP_003028246.1	tetracycline resistance protein TetM 1
0	YP_003028699.1	tetracycline resistance protein TetO
0	YP_003028728.1	tetracycline resistance protein TetL
0	YP_003431099.1	plasmid replication protein
0	YP_003446426.1	hypothetical protein smi_1318
0	ZP_00603114.1	Small GTP-binding protein domain
0	ZP_00787447.1	tetracycline resistance protein (tetM)
0	ZP_02206446.1	hypothetical protein COPEUT_01215
0	ZP_03079517.1	NLP/P60 family protein
0	ZP_03624201.1	small GTP-binding protein
0	ZP_03948336.1	DNA recombinase
0	ZP_03984378.1	tetracycline resistance protein Tet (M)
0	ZP_04435342.1	tetracycline resistance protein Tet (M)
0	ZP_05426923.1	6'-aminoglycoside N-acetyltransferase
0	ZP_05557914.1	tetracycline resistance protein
0	ZP_06334345.1	FtsK/SpoIIIE family protein
0	ZP_06334349.1	LOW QUALITY PROTEIN: conserved hypothetical protein

0	ZP_06628531.1	6'-aminoglycoside N-acetyltransferase/2-aminoglycoside phosphotransferase
0	ZP_06701789.1	6'-aminoglycoside N-acetyltransferase
0	ZP_07363085.1	tetracycline resistance protein TetP
0	ZP_07363100.1	FtsK/SpoIIIE family protein
0	ZP_07467523.1	tetracycline resistance protein
0	ZP_07568822.1	putative ATP/GTP-binding protein
1.00E-104	ADM29139.1	Transporter
1.00E-108	CAD35323.1	streptomycin resistance protein
1.00E-109	NP_370928.1	hypothetical protein SAV0404
1.00E-113	ZP_05682278.1	plasmid recombination enzyme
1.00E-118	NP_043522.1	ORF beta
1.00E-118	ZP_03565861.1	integrase-like protein
1.00E-121	ZP_04839156.1	alkylmercury lyase
1.00E-122	EFU06283.1	integrase core domain protein
1.00E-125	EFU06856.1	phosphotransferase enzyme family protein
1.00E-129	NP_863160.1	hypothetical protein pRUM_p17
1.00E-138	ZP_06334341.1	conserved hypothetical protein
1.00E-160	ZP_04564063.1	conserved hypothetical protein
1.00E-160	ZP_06926619.1	DNA recombinase, putative
1.00E-171	ZP_07851954.1	replication protein
1.00E-51	NP_370936.1	hypothetical protein SAV0412
1.00E-51	NP_687948.1	Tn916 hypothetical protein
1.00E-54	YP_002454911.1	hypothetical protein BCAH820_B0024
1.00E-57	NP_370921.1	putative transcriptional regulator
1.00E-67	ZP_03754880.1	hypothetical protein ROSEINA2194_03310
1.00E-68	NP_878007.1	acetyltransferase GNAT family protein
1.00E-75	NP_687933.1	Tn916 hypothetical protein
1.00E-76	YP_003750567.1	p-loop containing nucleoside triphosphate hydrolase domain
1.00E-91	CAP17166.1	hypothetical protein
1.00E-91	EFT48340.1	antirestriction protein
1.00E-94	YP_232761.1	N-methyltransferase
1.00E-98	ZP_05585337.1	conserved hypothetical protein
2.00E-100	ZP_07847049.1	integrase core domain protein
2.00E-108	CAD35323.1	streptomycin resistance protein
2.00E-119	NP_043522.1	ORF beta
2.00E-120	NP_763641.1	alkylmercury lyase
2.00E-125	EFU06856.1	phosphotransferase enzyme family protein
2.00E-130	NP_878023.1	beta-lactamase
2.00E-133	ZP_06820952.1	beta-lactamase

2.00E-142	NP_370930.1	hypothetical protein SAV0406
2.00E-149	NP_370934.1	hypothetical protein SAV0410
2.00E-150	ZP_07363101.1	conserved hypothetical protein
2.00E-51	NP_370936.1	hypothetical protein SAV0412
2.00E-54	ZP_05563488.1	conserved hypothetical protein
2.00E-57	NP_370935.1	hypothetical protein SAV0411
2.00E-57	NP_687935.1	Tn916, transcriptional regulator, putative
2.00E-61	ZP_07549672.1	replication protein
2.00E-65	ZP_07563512.1	acetyltransferase, GNAT family
2.00E-70	NP_763636.1	hypothetical protein SE0081
2.00E-70	ZP_05585325.1	mercuric resistance operon regulatory protein merR
2.00E-72	ZP_07851952.1	conserved domain protein
2.00E-76	YP_003750567.1	p-loop containing nucleoside triphosphate hydrolase domain
2.00E-90	YP_003864117.1	erythromycin ribosome methylase
2.00E-91	CAP17166.1	hypothetical protein
2.00E-91	EFT48340.1	antirestriction protein
2.00E-92	YP_001198324.1	putative adenine phosphoribosyltransferase
2.00E-96	ZP_00603166.1	conserved hypothetical protein
2.00E-96	ZP_05582794.1	predicted protein
3.00E-101	ADM29139.1	Transporter
3.00E-109	CAD35323.1	streptomycin resistance protein
3.00E-118	NP_043522.1	ORF beta
3.00E-125	EFU06856.1	phosphotransferase enzyme family protein
3.00E-151	ZP_00790462.1	conserved hypothetical protein
3.00E-169	CBJ57221.1	hypothetical protein
3.00E-51	ZP_05605038.1	mercuric transporter
3.00E-57	NP_687935.1	Tn916, transcriptional regulator, putative
3.00E-58	ZP_07771428.1	putative phage head-tail adaptor
3.00E-59	YP_001220640.1	hypothetical protein SaurJH9_2762
3.00E-72	ZP_07851952.1	conserved domain protein
3.00E-82	ZP_07860106.1	conserved domain protein
3.00E-85	ZP_05658352.1	conserved hypothetical protein
3.00E-92	NP_370929.1	hypothetical protein SAV0405
4.00E-106	ZP_03565873.1	hypothetical protein SauraJ_07038
4.00E-130	NP_863160.1	hypothetical protein pRUM_p17
4.00E-151	YP_003028726.1	membrane protein
4.00E-151	YP_133682.1	hypothetical protein Tn916_12
4.00E-57	NP_370935.1	hypothetical protein SAV0411
4.00E-57	NP_687935.1	Tn916, transcriptional regulator, putative
4.00E-64	ZP_00604430.1	protein of unknown function DUF961

4.00E-70	ZP_03079541.1	sigma-70, region 4 family
4.00E-73	ZP_07851952.1	conserved domain protein
4.00E-83	EFR90504.1	resolvase domain-containing protein
4.00E-86	NP_370916.1	integrase-like protein
4.00E-87	ZP_05658352.1	conserved hypothetical protein
5.00E-102	ZP_00603166.1	conserved hypothetical protein
5.00E-130	NP_863160.1	hypothetical protein pRUM_p17
5.00E-152	YP_133682.1	hypothetical protein Tn916_12
5.00E-153	ZP_07549672.1	replication protein
5.00E-164	YP_001198321.1	hypothetical protein SSU05_0955
5.00E-54	ZP_05563488.1	conserved hypothetical protein
5.00E-58	EFT95203.1	site-specific recombinase, phage integrase family
5.00E-76	NP_687933.1	Tn916 hypothetical protein
5.00E-83	YP_001198317.1	DNA recombinase, putative
6.00E-101	ZP_07363082.1	sigma-70 protein
6.00E-120	YP_003028701.1	methyltransferase
6.00E-133	ZP_06820952.1	beta-lactamase
6.00E-164	YP_001200520.1	hypothetical protein SSU98_0962
6.00E-68	ZP_03754880.1	hypothetical protein ROSEINA2194_03310
6.00E-76	NP_687933.1	Tn916 hypothetical protein
7.00E-116	ZP_05592053.1	endonuclease/Exonuclease/phosphatase
7.00E-122	ZP_05581297.1	streptomycin aminoglycoside 6-adenyltransferase
7.00E-161	ZP_04564063.1	conserved hypothetical protein
7.00E-70	ZP_03079541.1	sigma-70, region 4 family
8.00E-113	ZP_05682278.1	plasmid recombination enzyme
8.00E-117	ZP_04848819.1	DNA polymerase
8.00E-121	CAV31159.1	hypothetical protein
8.00E-51	ZP_05605038.1	mercuric transporter
8.00E-70	EFT38859.1	streptomycin adenylyltransferase
8.00E-70	ZP_03079541.1	sigma-70, region 4 family
8.00E-76	NP_687933.1	Tn916 hypothetical protein
9.00E-116	ZP_07516606.1	hypothetical protein EcolT_04009
9.00E-121	ZP_04839156.1	alkylmercury lyase
9.00E-139	ZP_06334341.1	conserved hypothetical protein
9.00E-72	ZP_07851952.1	conserved domain protein
3.00E-145	1L8T	Chain A, Crystal Structure Of 3',5"-Aminoglycoside Phosphotransferase Type Iiia Adp Kanamycin A Complex

Supplementary Table 4: Recently transferred genes shared by hot spring isolates.
Transposons are not filtered from this list. All proteins are included.

E-value	Accession Number	Annotation
0	NP_622395.1	IS30 family transposase
0	NP_624269.1	ABC-type multidrug/protein/lipid transport system, ATPase component
0	YP_001179121.1	hypothetical protein Csac_0283
0	YP_001179122.1	regulator of chromosome condensation, RCC1
0	YP_001430345.1	hypothetical protein Rcas_0194
0	YP_001431060.1	hypothetical protein Rcas_0928
0	YP_001664625.1	extracellular solute-binding protein
0	YP_001664937.1	ABC transporter related
0	YP_001665911.1	glycoside hydrolase family protein
0	YP_001666031.1	PTS system transcriptional activator
0	YP_002574406.1	Hedgehog/intein hint domain-containing protein
0	YP_002574410.1	transglutaminase domain-containing protein
0	YP_003676016.1	PTS system transcriptional activator
0	YP_003852496.1	extracellular solute-binding protein family 1
0	YP_475558.1	sensor histindine kinase/response regulator
1.00E-68	YP_001179696.1	hypothetical protein Csac_0889
1.00E-85	YP_002572798.1	hypothetical protein Athe_0920
1.00E-89	YP_001664938.1	ABC-type multidrug transport system ATPase and permease components-like protein
2.00E-103	YP_003677146.1	anthranilate synthase component I
2.00E-110	YP_003476262.1	major facilitator superfamily MFS_1
2.00E-128	YP_001662990.1	radical SAM domain-containing protein
2.00E-132	YP_001039596.1	integrase catalytic subunit
2.00E-164	YP_001666026.1	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding
2.00E-178	YP_001179122.1	regulator of chromosome condensation, RCC1
2.00E-60	YP_003477502.1	hypothetical protein Thit_1694
2.00E-64	YP_003951479.1	regulator of chromosome condensation-like protein
2.00E-72	YP_001179123.1	hemerythrin-like metal-binding protein
2.00E-89	YP_001308094.1	sigma-54 factor interaction domain-containing protein
3.00E-107	YP_001665992.1	hypothetical protein Teth39_2031
3.00E-107	YP_003676052.1	protein of unknown function DUF820
3.00E-156	ZP_04788155.1	binding-protein-dependent transport systems inner membrane component
3.00E-169	YP_001179128.1	regulator of chromosome condensation, RCC1
3.00E-68	YP_002572618.1	hypothetical protein Athe_0730
3.00E-91	YP_478942.1	hypothetical protein CYB_2753

4.00E-125	ZP_06974940.1	conserved hypothetical protein
4.00E-140	ZP_05493182.1	PTS system mannose/fructose/sorbose family IID component
4.00E-56	XP_001033296.2	ABC transporter family protein
4.00E-74	ZP_04787687.1	PTS system sorbose subfamily IIB component
4.00E-74	ZP_05491481.1	CRISPR-associated protein Cas4
4.00E-87	YP_002572798.1	hypothetical protein Athe_0920
5.00E-124	ZP_06974940.1	conserved hypothetical protein
5.00E-67	YP_001666030.1	PTS system fructose subfamily IIA component
6.00E-136	YP_002574410.1	transglutaminase domain-containing protein
6.00E-173	YP_001179341.1	ABC-2 type transporter
7.00E-123	YP_516917.1	hypothetical protein DSY0684
7.00E-158	YP_001664626.1	binding-protein-dependent transport systems inner membrane component
7.00E-173	YP_002574462.1	ABC-2 type transporter
7.00E-73	YP_475556.1	hypothetical protein CYA_2154
8.00E-108	YP_003853012.1	PTS system mannose/fructose/sorbose family IID component
8.00E-131	YP_001666028.1	PTS system mannose/fructose/sorbose family IID component
8.00E-91	YP_002574406.1	Hedgehog/intein hint domain-containing protein
8.00E-98	YP_001664532.1	CRISPR-associated Cas4 family protein
9.00E-126	YP_001179121.1	hypothetical protein Csac_0283
9.00E-77	YP_001666027.1	PTS system sorbose subfamily IIB component

Supplementary Table 5: Recently transferred genes shared by soil isolates. Seventy-four transposon proteins have been filtered from the list of genes listed below to reduce functional redundancy and create a more concise table.

E-value	Accession Number	Annotation
0	ACN58394.1	MerA
0	ACT97186.1	MmfS
0	CAA86212.1	FyuA precursor
0	CAJ77060.1	Mercuric ion reductase
0	EEE73071.1	predicted protein
0	EEE77804.1	predicted protein
0	EEF08500.1	predicted protein
0	EFU55740.1	salicylate synthase
0	NP_669701.1	salicylate synthase Irp9
0	NP_669706.1	HMWP2 nonribosomal peptide synthetase
0	NP_754310.1	prophage P4 integrase
0	NP_858035.1	putative mercuric reductase
0	NP_993007.1	yersiniabactin biosynthetic protein
0	YP_001096379.1	hypothetical protein pLEW517_p54
0	YP_001100248.1	DNA topoisomerase III
0.00E+00	YP_001173822.1	phage integrase family site specific recombinase
0.00E+00	YP_001452491.1	hypothetical protein CKO_00905
0.00E+00	YP_001452495.1	hypothetical protein CKO_00909
0.00E+00	YP_001452497.1	hypothetical protein CKO_00911
0.00E+00	YP_001452502.1	salicylate synthase Irp9
0.00E+00	YP_001596803.1	hypothetical protein COXBURSA331_A1044
0.00E+00	YP_001899332.1	heavy metal efflux pump, CzcA family
0.00E+00	YP_001899334.1	outer membrane efflux protein
0.00E+00	YP_002255023.1	integrase/recombinase protein
0.00E+00	YP_002398185.1	permease and ATP-binding protein of yersiniabactin-iron ABC transporter YbtQ
0.00E+00	YP_002412997.1	hypothetical protein ECUMN_2279
0.00E+00	YP_002440259.1	COG1196: Chromosome segregation ATPases
0.00E+00	YP_002798173.1	ABC transporter, transmembrane region, ATP binding component
0.00E+00	YP_002798176.1	Non-ribosomal peptide synthetase, with condensation, AMP binding and thioesterase modules
0.00E+00	YP_002801292.1	MFS family sugar transporter
0.00E+00	YP_002920242.1	integrase
0.00E+00	YP_002920245.1	putative ABC transporter protein

0.00E+00	YP_002920249.1	yersiniabactin biosynthetic protein
0.00E+00	YP_002920252.1	yersiniabactin siderophore biosynthetic protein
0.00E+00	YP_003324256.1	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase
0.00E+00	YP_025335.1	hypothetical protein pRA2_13
0.00E+00	YP_025381.1	putative Tn5504 resolvase
0.00E+00	YP_070123.1	yersiniabactin biosynthetic protein
0.00E+00	YP_145595.1	cation efflux system transmembrane protein
0.00E+00	YP_145601.1	putative porin signal peptide protein
0.00E+00	YP_145603.1	cadmium translocating p-type atpase
0.00E+00	YP_145605.1	putative transmembrane protein
0.00E+00	YP_556423.1	putative mercuric reductase
0.00E+00	YP_581791.1	cation proton antiporter efflux protein CzcC
0.00E+00	YP_581983.1	putative mercuric reductase
0.00E+00	YP_584440.1	hypothetical protein Rmet_2292
0.00E+00	YP_584446.1	DNA topoisomerase III
0.00E+00	YP_584451.1	heavy metal translocating P-type ATPase
0.00E+00	YP_584473.1	hypothetical protein Rmet_2327
0.00E+00	YP_584476.1	helicase-like protein
0.00E+00	YP_584482.1	hypothetical protein Rmet_2336
0.00E+00	YP_584484.1	hypothetical protein Rmet_2338
0.00E+00	YP_585175.1	ParB family protein
0.00E+00	YP_984373.1	integrase catalytic subunit
0	YP_985596.1	mercuric reductase
0	YP_986892.1	multicopper oxidase, type 3
0	YP_986893.1	putative copper resistance-related lipoprotein
0	YP_986896.1	acetate kinase
0	YP_986898.1	cyclopropane-fatty-acyl-phospholipid synthase
0	YP_986900.1	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase
0	YP_986901.1	hypothetical protein Ajs_2679
0	YP_986903.1	2-octaprenylphenol hydroxylase
0	YP_986904.1	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding
0	YP_986911.1	thymidine phosphorylase
0	YP_986912.1	beta-lactamase domain-containing protein
0	ZP_04762039.1	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding
0	ZP_04762040.1	ABC-1 domain protein
0	ZP_04762045.1	Cyclopropane-fatty-acyl-phospholipid synthase
0	ZP_04762047.1	acetate kinase

0	ZP_04762051.1	putative copper resistance-related lipoprotein
0	ZP_04762052.1	multicopper oxidase type 3
0	ZP_04763698.1	histidine kinase
0	ZP_04934712.1	hypothetical protein PA2G_02085
0	ZP_05132972.1	hypothetical protein SSKA14_39
0	ZP_05231481.1	ATP-dependent Clp protease
0	ZP_07193592.1	TonB-dependent siderophore receptor
0	ZP_07793687.1	Predicted transcriptional regulator
1.00E-101	NP_669714.1	hypothetical protein y2407
1.00E-104	YP_986360.1	polar amino acid ABC transporter, inner membrane subunit
1.00E-105	EEE70611.1	predicted protein
1.00E-109	YP_002800768.1	Integrase, catalytic domain-containing protein
1.00E-110	YP_243131.1	hypothetical protein XC_2050
1.00E-119	ZP_06501249.1	glyceraldehyde-3-phosphate dehydrogenase, type II
1.00E-122	EFU06283.1	integrase core domain protein
1.00E-126	YP_584438.1	cobyrinic acid a,c-diamide synthase
1.00E-128	YP_788484.1	putative binding protein component of ABC transporter
1.00E-139	ZP_05290280.1	ATP-dependent Clp protease, ATP-binding subunit ClpL
1.00E-141	YP_002302647.1	hypothetical DNA-binding protein
1.00E-145	NP_820760.2	hypothetical protein CBU_1780
1.00E-149	YP_585136.1	phosphonate-binding periplasmic protein
1.00E-152	YP_002317674.1	dihydropteroate synthase
1.00E-155	ZP_04761600.1	conserved hypothetical protein
1.00E-157	CAD60669.1	hypothetical protein
1.00E-178	NP_754310.1	prophage P4 integrase
1.00E-50	YP_584475.1	hypothetical protein Rmet_2329
1.00E-53	YP_145590.1	mgtC protein
1.00E-53	ZP_08004433.1	hypothetical protein HMPREF1013_01038
1.00E-56	YP_584469.1	hypothetical protein Rmet_2323
1.00E-57	YP_293612.1	hypothetical protein Reut_D6455
1.00E-57	ZP_02537516.1	ribosome recycling factor
1.00E-57	ZP_07148706.1	acetyltransferase (GNAT) family protein
1.00E-61	YP_584470.1	hypothetical protein Rmet_2324
1.00E-64	YP_002980631.1	hypothetical protein Rpic12D_0658
1.00E-65	YP_585144.1	putative signal peptide protein
1.00E-65	ZP_07796774.1	hypothetical protein PA39016_002870009
1.00E-70	NP_085423.1	Tn501 orf, hypotheical
1.00E-71	YP_335188.1	IrlS

1.00E-73	ZP_05135172.1	conserved hypothetical protein
1.00E-74	ZP_04762053.1	blue (type 1) copper domain protein
1.00E-78	YP_002984397.1	Cupin 2 conserved barrel domain protein
1.00E-81	YP_746363.1	hypothetical protein Neut_0110
1.00E-83	YP_788482.1	hypothetical protein PA14_04070
1.00E-83	YP_986908.1	hypothetical protein Ajs_2687
1.00E-85	YP_584444.1	hypothetical protein Rmet_2296
1.00E-85	ZP_03824995.1	pyridine nucleotide transhydrogenase
1.00E-87	AAC53669.1	Flp recombinase
1.00E-87	NP_668948.1	NADH dehydrogenase subunit A
1.00E-87	YP_001899048.1	Resolvase domain
1.00E-89	YP_985618.1	hypothetical protein Ajs_1322
1.00E-98	YP_585149.1	resolvase-like protein
1.00E-98	ZP_04763700.1	Resolvase domain protein
2.00E-100	EFV84648.1	methylenetetrahydrofolate reductase
2.00E-105	EEE70611.1	predicted protein
2.00E-105	YP_584468.1	hypothetical protein Rmet_2322
2.00E-105	ZP_07847049.1	integrase core domain protein
2.00E-108	ZP_06845918.1	integral membrane protein
2.00E-110	YP_584477.1	hypothetical protein Rmet_2331
2.00E-119	ADA74134.1	IS1 ORF2
2.00E-125	YP_002798178.1	hypothetical protein Avin_09710
2.00E-134	ZP_04934717.1	hypothetical protein PA2G_02090
2.00E-144	ZP_01771019.1	probable short-chain alcohol dehydrogenase protein
2.00E-150	YP_002135620.1	hypothetical protein AnaeK_3271
2.00E-151	AAD31025.1	aminoglycoside 3'-phosphotransferase
2.00E-155	AAL91683.1	putative transporter
2.00E-158	ZP_04934726.1	Transposase Tn3
2.00E-159	NP_478145.1	aminoglycoside 3'-phosphotransferase
2.00E-162	YP_002255093.1	2,3-dihydroxybenzoate-amp ligase protein
2.00E-164	YP_556417.1	putative chromate resistance protein
2.00E-174	ZP_05231481.1	ATP-dependent Clp protease
2.00E-53	YP_001899338.1	MgtC/SapB transporter
2.00E-60	NP_780783.1	hypothetical protein CTC00065
2.00E-61	YP_004129006.1	hypothetical protein Alide_4423
2.00E-62	YP_025338.1	MerR
2.00E-66	YP_584450.1	MerR family transcriptional regulator
2.00E-73	YP_585145.1	hypothetical protein Rmet_3003
2.00E-73	ZP_05111911.1	cobalt-zinc-cadmium resistance protein
2.00E-74	NP_249006.1	hypothetical protein PA0315

2.00E-75	NP_903722.1	amino acid ABC transporter permease
2.00E-76	YP_556414.1	hypothetical protein Bxe_C1204
2.00E-79	YP_584435.1	hypothetical protein Rmet_2287
2.00E-81	YP_986362.1	hypothetical protein Ajs_2116
2.00E-82	YP_584452.1	signal peptidase II
2.00E-83	EEE72779.1	predicted protein
2.00E-85	YP_001899048.1	Resolvase domain
2.00E-85	YP_070116.1	hypothetical protein YPTB1588
2.00E-86	ZP_01887574.1	NADH dehydrogenase I chain A
2.00E-88	NP_858031.1	putative resolvase
2.00E-90	NP_085422.3	Tn501 repressor
3.00E-102	YP_001899051.1	conserved hypothetical protein
3.00E-106	YP_002798179.1	ABC transporter, transmembrane region, ATP binding component
3.00E-128	YP_002798174.1	thioredoxin-like fold protein
3.00E-151	ZP_04762034.1	ribose-phosphate pyrophosphokinase
3.00E-161	YP_002798843.1	Integrase, catalytic domain-containing protein
3.00E-164	YP_556417.1	putative chromate resistance protein
3.00E-168	YP_001096379.1	hypothetical protein pLEW517_p54
3.00E-171	YP_981981.1	integrase catalytic subunit
3.00E-172	ABW39611.1	StrA
3.00E-173	YP_003064448.1	ATP-dependent Clp protease, ATP-binding subunit ClpL
3.00E-174	CBY26680.1	integrase
3.00E-176	YP_986902.1	glyceraldehyde-3-phosphate dehydrogenase
3.00E-53	YP_003982631.1	lipoprotein signal peptidase 2
3.00E-57	ZP_04762044.1	conserved hypothetical protein
3.00E-58	YP_587453.1	major facilitator protein family permease
3.00E-59	ZP_05137048.1	cation efflux protein
3.00E-66	YP_004111325.1	GCN5-like N-acetyltransferase
3.00E-66	ZP_06839949.1	hypothetical protein BCh11DRAFT_1213
3.00E-70	YP_986897.1	MerR family transcriptional regulator
3.00E-72	YP_145589.1	hypothetical protein pMOL30_059
3.00E-75	AAC53669.1	Flp recombinase
3.00E-75	CAA23425.1	unnamed protein product
3.00E-75	ZP_05134612.1	putative secreted protein
3.00E-76	YP_556414.1	hypothetical protein Bxe_C1204
3.00E-76	YP_788486.1	D-3-phosphoglycerate dehydrogenase
3.00E-76	ZP_03829064.1	helix-turn-helix domain-containing protein
3.00E-79	YP_320423.1	hypothetical protein Ava_C0146
3.00E-83	YP_002354579.1	HtrA2 peptidase

3.00E-86	NP_858031.1	putative resolvase
3.00E-89	YP_001350673.1	integrase/recombinase
3.00E-96	NP_820372.1	ribosome recycling factor
4.00E-104	AAL10408.1	AacA4 aminoglycoside (6') acetyltransferase
4.00E-125	YP_584443.1	hypothetical protein Rmet_2295
4.00E-141	YP_584466.1	hypothetical protein Rmet_2320
4.00E-151	AAZ73126.1	Sul1
4.00E-151	ZP_03543711.1	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding
4.00E-152	AAF42994.1	geneticin-resistant protein
4.00E-153	ZP_01771016.1	transcriptional regulator, LysR family
4.00E-157	CAD60669.1	hypothetical protein
4.00E-161	YP_002798182.1	Integrase, catalytic domain-containing protein
4.00E-166	YP_002798168.1	hypothetical protein Avin_09590
4.00E-166	ZP_07793689.1	hypothetical protein PA39016_001010043
4.00E-170	YP_113802.1	putative mercuric reductase
4.00E-173	NP_669705.1	AraC family transcriptional regulator
4.00E-173	YP_003064448.1	ATP-dependent Clp protease, ATP-binding subunit ClpL
4.00E-176	YP_002038873.1	aminoglycoside resistance protein A
4.00E-176	YP_584442.1	hypothetical protein Rmet_2294
4.00E-178	ZP_05231481.1	ATP-dependent Clp protease
4.00E-53	YP_001899052.1	Cupin 2 conserved barrel domain protein
4.00E-56	ZP_03700455.1	conserved hypothetical protein
4.00E-61	CAA23425.1	unnamed protein product
4.00E-69	YP_584460.1	MerR family transcriptional regulator
4.00E-72	YP_585139.1	hypothetical protein Rmet_2997
5.00E-102	EFU17052.1	ATPase family protein
5.00E-104	YP_274080.1	helicase/SNF2 domain-containing protein
5.00E-104	ZP_00603144.1	Integrase, catalytic region
5.00E-119	YP_584472.1	hypothetical protein Rmet_2326
5.00E-133	YP_002553260.1	chromate transporter, chromate ion transporter (chr) family
5.00E-139	ZP_05290280.1	ATP-dependent Clp protease, ATP-binding subunit ClpL
5.00E-146	YP_584459.1	hypothetical protein Rmet_2311
5.00E-150	YP_986359.1	extracellular solute-binding protein
5.00E-151	NP_820760.2	hypothetical protein CBU_1780
5.00E-152	ACH99098.1	Aph
5.00E-154	AAF42994.1	geneticin-resistant protein
5.00E-155	AAL91683.1	putative transporter

5.00E-161	YP_863796.1	hypothetical protein Shewana3_4281
5.00E-175	YP_001452490.1	hypothetical protein CKO_00904
5.00E-175	YP_070118.1	hypothetical protein YPTB1590
5.00E-51	YP_003982636.1	sterol desaturase-like protein
5.00E-52	ZP_01771027.1	dsba oxidoreductase, putative
5.00E-61	EFS80503.1	hypothetical protein HMPREF9597_00208
5.00E-63	YP_002440292.1	transcriptional regulator, MerR family
5.00E-66	YP_986899.1	hypothetical protein Ajs_2677
5.00E-69	ZP_01363215.1	hypothetical protein PaerPA_01000309
5.00E-77	YP_584447.1	cation efflux protein
5.00E-78	ZP_04763627.1	conserved hypothetical protein
5.00E-81	YP_548042.1	phosphoribosylpyrophosphate synthetase
6.00E-102	NP_820905.2	acyl-CoA synthetase
6.00E-102	ZP_04763628.1	aminoglycoside phosphotransferase
6.00E-112	YP_274080.1	helicase/SNF2 domain-containing protein
6.00E-115	NP_820373.1	uridylate kinase
6.00E-141	YP_002302647.1	hypothetical DNA-binding protein
6.00E-146	YP_584467.1	hypothetical protein Rmet_2321
6.00E-147	YP_001563700.1	LysR family transcriptional regulator
6.00E-166	YP_002798168.1	hypothetical protein Avin_09590
6.00E-54	YP_986888.1	hypothetical protein Ajs_2662
6.00E-56	YP_585148.1	hypothetical protein Rmet_3006
6.00E-71	YP_001899339.1	hypothetical protein Rpic_1769
6.00E-71	YP_335188.1	IrlS
6.00E-75	ZP_03829064.1	helix-turn-helix domain-containing protein
6.00E-88	AAC53669.1	Flp recombinase
6.00E-89	NP_052900.1	resolvase for Tn21
6.00E-94	YP_584479.1	hypothetical protein Rmet_2333
7.00E-111	ZP_01365962.1	hypothetical protein PaerPA_01003092
7.00E-119	YP_985617.1	hypothetical protein Ajs_1321
7.00E-137	YP_584457.1	hypothetical protein Rmet_2309
7.00E-150	AAZ73126.1	Sul1
7.00E-154	AAF42994.1	geneticin-resistant protein
7.00E-170	ZP_04762035.1	2-alkenal reductase
7.00E-62	ZP_05826564.1	mercury resistance inner membrane protein
7.00E-68	YP_001173906.1	hypothetical protein PST_3436
7.00E-72	CAA23425.1	unnamed protein product
7.00E-74	ZP_01165046.1	uridylate kinase
7.00E-79	NP_569360.1	putative transcriptional regulator MerR
7.00E-80	NP_941195.1	putative transcriptional regulator MerR
7.00E-99	YP_001338852.1	hypothetical protein KPN_pKPN5p08242

8.00E-107	NP_608307.1	aminoglycoside 6'-N-acetyltransferase type Ib
8.00E-112	ZP_07673475.1	sensor protein CzcS
8.00E-131	YP_585131.1	hypothetical protein Rmet_2989
8.00E-133	YP_002553260.1	chromate transporter, chromate ion transporter (chr) family
8.00E-51	YP_145599.1	ORF131 protein
8.00E-62	ZP_05826564.1	mercury resistance inner membrane protein
8.00E-65	ZP_04762037.1	conserved hypothetical protein
8.00E-68	YP_002798175.1	AMP-dependent synthetase and ligase family protein
8.00E-81	ZP_01363211.1	hypothetical protein PaerPA_01000305
8.00E-84	ZP_06686109.1	conserved hypothetical protein
8.00E-88	EEE76668.1	predicted protein
9.00E-107	YP_293616.1	hypothetical protein Reut_D6459
9.00E-112	ZP_07673475.1	sensor protein CzcS
9.00E-139	ZP_06880786.1	hypothetical protein PaerPAb_24286
9.00E-140	ZP_05290280.1	ATP-dependent Clp protease, ATP-binding subunit ClpL
9.00E-148	YP_124671.1	hypothetical protein lpp2360
9.00E-151	AAZ73126.1	Sul1
9.00E-156	AAL91683.1	putative transporter
9.00E-173	ABW39611.1	StrA
9.00E-51	YP_145599.1	ORF131 protein
9.00E-53	YP_001747182.1	hypothetical protein PputW619_0307
9.00E-56	ADR71220.1	TEM-183 beta-lactamase
9.00E-57	ZP_07148706.1	acetyltransferase (GNAT) family protein
9.00E-61	EFS80503.1	hypothetical protein HMPREF9597_00208
9.00E-63	YP_986940.1	blue (type1) copper domain-containing protein
9.00E-65	ADD63334.1	hypothetical protein
9.00E-78	ZP_01767210.1	CBS domain protein
9.00E-82	YP_584445.1	single-stranded DNA-binding protein
9.00E-88	EEE76668.1	predicted protein
9.00E-89	YP_001173820.1	double-stranded beta-helix domain-containing protein

Supplementary Table 6: Antibiotic resistance genes transferred between human and farm animals. Each of the nine gene families that includes HGT between human and farm animals is shown along with the specific drugs targeted. Drugs with known uses in agriculture are marked with an asterisk. All nine gene families include at least one drug used in agriculture[33]. The genera engaged in transfer are listed for each gene family and broken into groups based on whether they are Gram-positive or Gram-negative and whether they are human or farm animal associated. The phylogenetic breadth of transfer highlights the mobility of these resistance genes.

Gene family	Targeted Antibiotics	Human		Agricultural Mammals	
		Gram +	Gram -	Gram +	Gram -
aac	amikacin* dibekacin isepamicin netilmicin sisomicin tobramycin	Enterococcus Staphylococcus Streptococcus		Enterococcus	
aph	amikacin* butirosin gentamincin_b* isepamicin kanamycin* lividomycin* neomycin* paromomycin* ribostamycin streptomycin*	Bacillus Corynebacterium Enterococcus Staphylococcus Streptococcus	Acinetobacter Edwardsiella Escherichia Klebsiella Pseudomonas Salmonella Vibrio Yersinia	Enterococcus Ruminococcus	Actinobacillus Escherichia
cata	bicyclomycin* chloramphenicol* florfenicol*	Enterococcus Streptococcus Bifidobacterium Ruminococcus	Acinetobacter Edwardsiella Escherichia Parabacteroides Klebsiella Providencia	Enterococcus	Actinobacillus
erm	lincosamide* macrolide*	Enterococcus Clostridium		Enterococcus	

	streptogramin_b*	Streptococcus			
sul	sulfonamides*		Acinetobacter Enterobacter Escherichia Klebsiella Salmonella Vibrio Yersinia		Actinobacillus Escherichia
tet	tetracycline*	Enterococcus Mycobacterium Staphylococcus Streptococcus Ureaplasma Faecalibacterium Anaerostipes Clostridium Dorea Anaerostipes Ruminococcus Methanobrevibacter Lactobacillus Coprococcus	Actinobacillus Brucella Campylobacter Edwardsiella Escherichia Desulfovibrio Subdoligranulum Fusobacterium Proteus Providencia Salmonella Vibrio	Ruminococcus Staphylococcus	Actinobacillus Campylobacter Escherichia
erm	lincosamide* macrolide* streptogramin_b*	Enterococcus Clostridium Streptococcus		Enterococcus	
sul	sulfonamides*		Acinetobacter Enterobacter Escherichia Klebsiella Salmonella Vibrio Yersinia		Actinobacillus Escherichia
tet	tetracycline*	Enterococcus	Actinobacillus	Ruminococcus	Actinobacillus

		Mycobacterium	Brucella	Staphylococcus	Campylobacter
		Staphylococcus	Campylobacter		Escherichia
		Streptococcus	Edwardsiella		
		Ureaplasma	Escherichia		
		Anaerostipes	Fusobacterium		
		Clostridium	Desulfovibrio		
		Coprococcus	Proteus		
		Dorea	Providencia		
		Faecalibacterium	Salmonella		
		Lactobacillus	Subdoligranulum		
		Methanobrevibacter	Vibrio		
		Ruminococcus			

Supplementary Table 7: Sample sizes used in statistical comparisons. This table shows the sample sizes used in the Mann-Whitney U-tests in Figs. 1, 2, and 4.

	16S Distance Bins								
Environment	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Human	166	232	378	383	827	1327	1638	2672	3544
Human within	88	120	184	193	416	546	562	786	973
Human between	62	132	200	198	372	576	857	1689	2300
Non-Human	1658	1169	2859	2657	3810	6526	6891	10380	13841
Same ecology	166	232	378	383	827	1327	1638	2672	3544
Different ecology	552	425	948	911	1876	3831	3979	6990	8263
Same continent	84			253			652		
Different continent	108			372			1186		
Same oxygen tolerance	106	190	268	247	500	708	683	988	1297
Same pathogenicity	30	70	122	84	174	233	266	355	434

	16S Distance Bins						
Environment	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Human	5001	7262	10802	15319	18587	18125	14944
Human within	1417	1896	2807	3935	4556	4275	3451
Human between	3101	4704	6840	9586	11739	10982	8946
Non-Human	22295	29688	45310	64169	76375	74374	58819
Same ecology	5001	7262	10802	15319	18587	18125	14944
Different ecology	13433	18909	27956	39740	50099	50868	40880
Same continent	2120			4709			1974
Different continent	3375			7831			3510

Supplemental Reference

33. Aarestrup, M, F. *Antimicrobial resistance in bacteria of animal origin*. (ASM Press: Washington, DC, 2006).