The ISME Journal (2015) 9, 207–216
© 2015 International Society for Microbial Ecology All rights reserved 1751-7362/15
www.nature.com/ismej

ORIGINAL ARTICLE

Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology

Molly K Gibson[1], Kevin J Forsberg[1] and Gautam Dantas[1,2,3]
[1]*Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO, USA;* [2]*Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA and* [3]*Department of Biomedical Engineering, Washington University, St Louis, MO, USA*

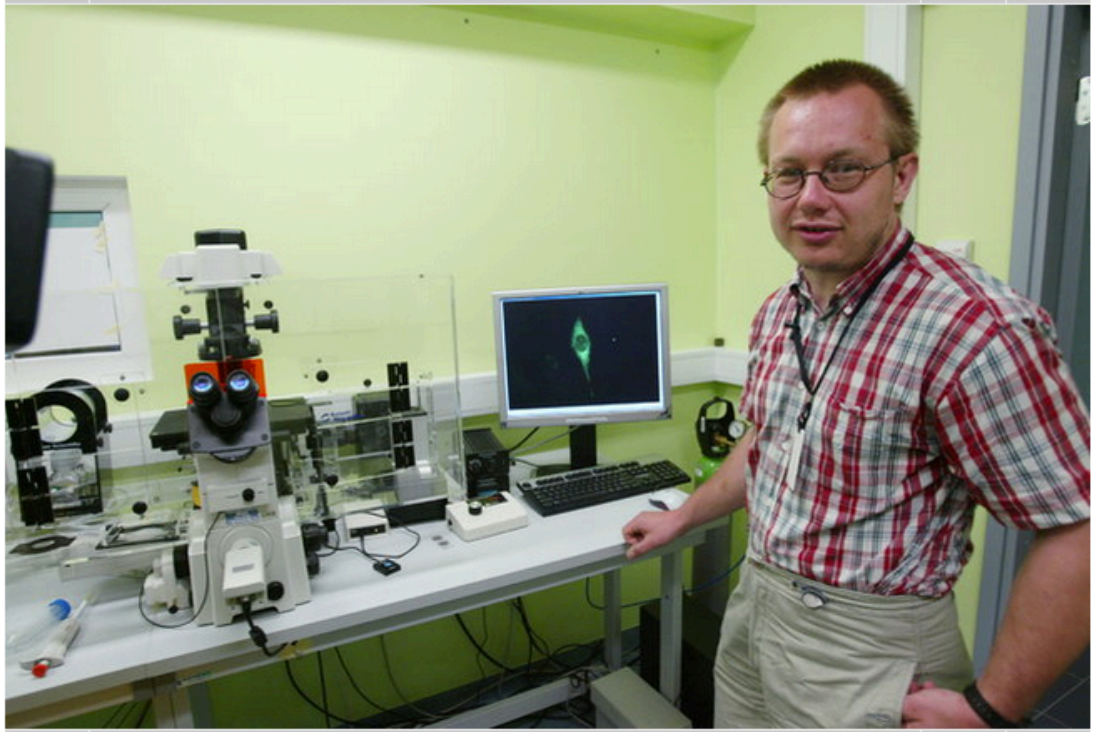
Age Brauer

Journal Club
18th of March 2015

Present situation
Resistance in ba
WHO's 2014 re
resistance reveal
prediction for the
world, and is putti
in the community
action, the world is
common infecti
treatable for dec

Tervis

2/2



2/2

Professor Tanel Tensoni juhtimisel uuritakse, kuidas resistentsusgeenid kanduvad ühelt bakterilt teisele – see teadmine aitaks võidelda ravile allumatute pisikutega. Foto: Margus Ansu

Ravile mittealluvad haigustekitajad võtavad võimust

Postimees 24. veebruar 2015 **23:59** Lisa järjehoidja

Antibiootikumidele resistentsed ehk ravile mittealluvad haigustekitajad võtavad kogu maailmas üha enam võimust. Üheks nende leviku põhjuseks on antibiootikumide sattumine keskkonda.

stance is
atest
an health.

mic Forum
ts

idrug-
pean

tibiotic-
result.

ooo
ant

een
DR-TB
hat
ctive
TB.

Antibiotic resistance in bacteria

- Intrinsic resistance
  - The absence of a susceptible target
  - Outer membrane barrier (Gram-negative)

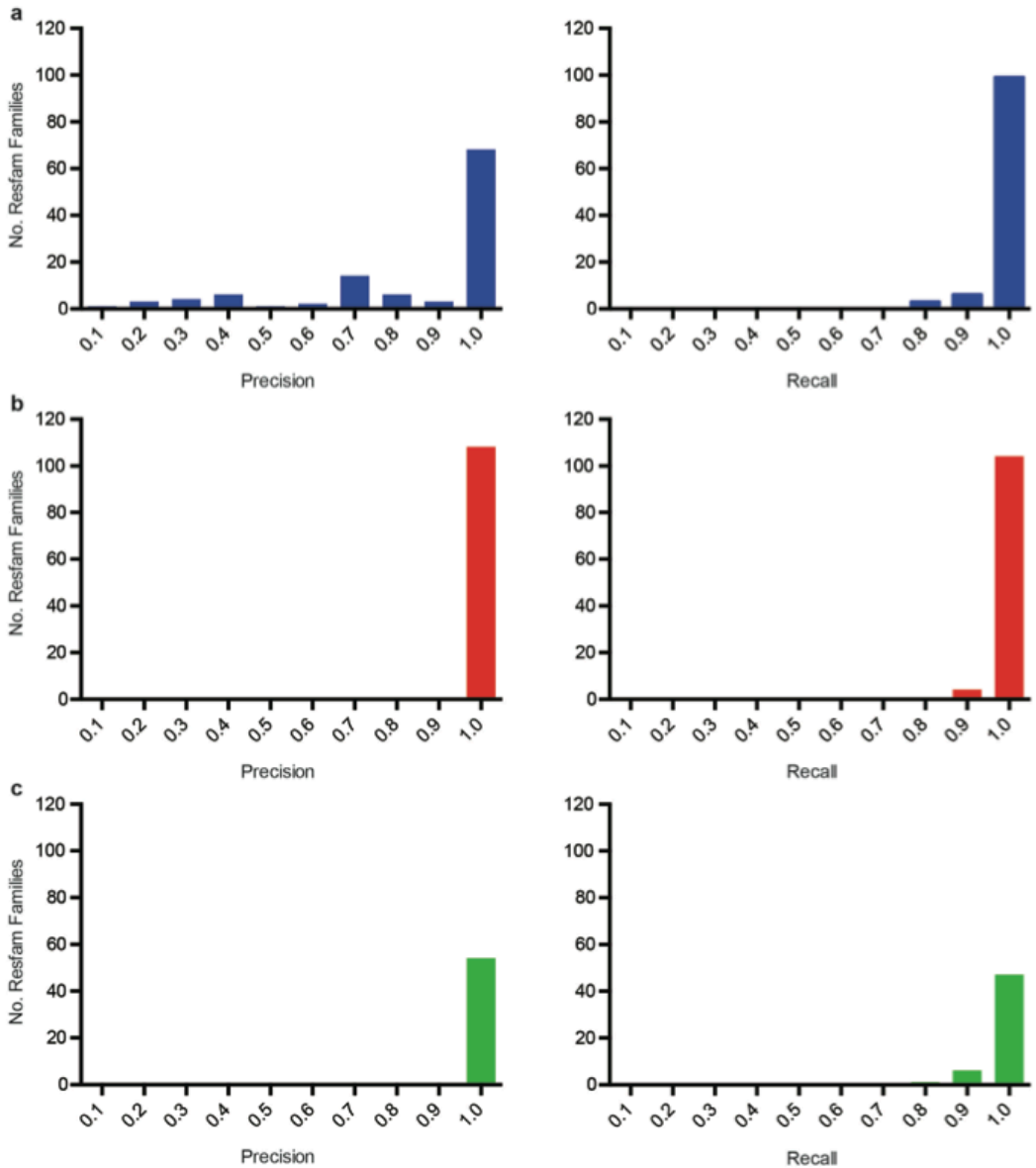
- Acquired resistance
  - Minimizing the intracellular concentrations of the antibiotic (poor penetration into the bacterium, antibiotic efflux)
  - Modification of the antibiotic target (genetic mutations, post-translational modification)
  - Inactivation of the antibiotic by hydrolysis or modification

- Resfams – a curated database of protein families and associated profile hidden Markov models (HMMs),organized by ontology and confirmed for antibiotic resistance (AR) function.
- To quantitatively understand the relationship between human-associated and environmental resistomes.

Training and benchmarking of Resfams profile HMMs

- Proteins from Comprehensive Antibiotic Resistance Database (**CARD**), The Antibiotic Resistance Database (**ARDB**), the Lactamase Engineering Database (**LacED**) and collection of **curated beta-lactamase proteins**.
- All proteins were hand-curated using phylogenetic analysis and literature searches.
- 119 Resfams profile HMMs were trained using CARD, LacED and curated collection of beta-lactamases.
- ARDB was an independent test set.
- Gathering thresholds were set to achieve the precision and recall metrics as close to 1.0 as possible.

- Precision = (the total number of correct annotations)/(total number of annotations)
- Recall = (the total number of correct annotations)/(the total number of proteins in ARDB with annotations matching the annotation of the profile HMM)



Supplementary Figure S1: Resfams family precision and recall distributions. The x-axis represents the precision or recall values while the y-axis represents the number of Resfams families of a given value. (A) Precision and recall distributions of Resfams families on the proteins used to train the original profile using universal thresholds. (B) Precision and recall distributions of Resfams families on the proteins used to train the original profile HMMs using defined gathering thresholds. (C) Precision and recall distributions of Resfams families on an independent validation set of antibiotic resistance proteins not used to train the Resfams profile HMMs using defined gathering thresholds.

Functional metagenomic selections

1) Multidrug resistant cultured soil isolates
2) Pediatric human fecal samples
3) Cedar Creek and Kellogg Biological Station Soils

- Metagenomic plasmid libraries in *E.coli*
- Kanamycin + the antibiotic of interest
- Resistant inserts were amplified and sequenced using Illumina Hi-Seq PE 76 or 101 bp sequencing protocol
- PARFuMS (Parallel Annotation and Re-assembly of Functional Metagenomic Selections) pipeline
- Open reading frame predictions – MetaGeneMark
- Protein annotation using Resfams database of profile HMMs
- Resistome comparisons

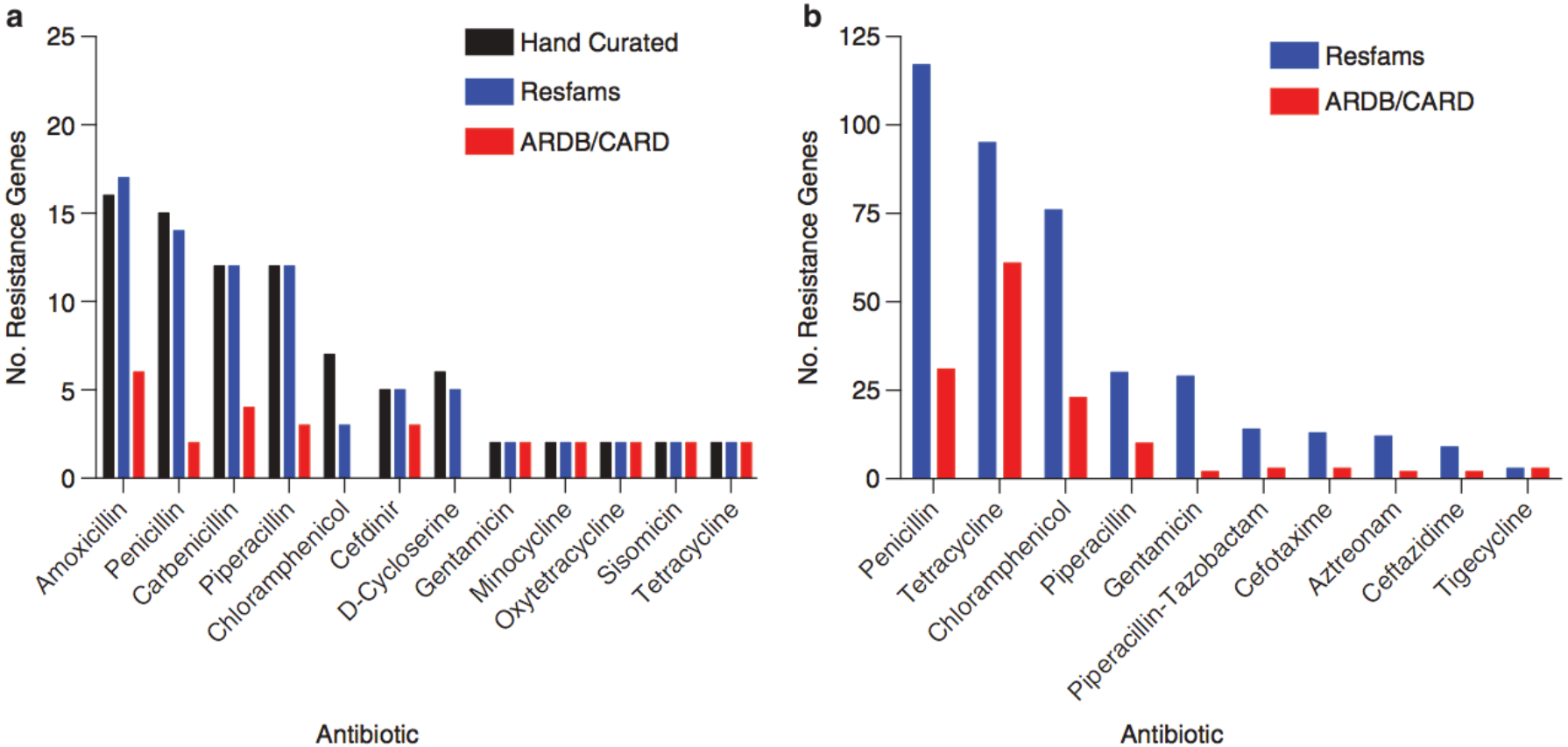


Figure 1 Annotation of functional metagenomic selections using Resfams. Unique full-length open reading frames (ORFs; >90 amino acids (a.a.)) annotated as AR proteins from functional metagenomic selections using Resfams HMM database (blue) compared with hand curation (black) and BLAST to AR databases (red) from (**a**) MDR cultured soil bacteria and (**b**) human gut microbiota. The Resfams HMM database identified no false positive annotations as measured by the hand-curated gold standard for cultured soil bacteria. For the amoxicillin selection, MetaGeneMark predicted one ORF in the hand-curated set as two independent ORFs and both were correctly identified by Resfams.

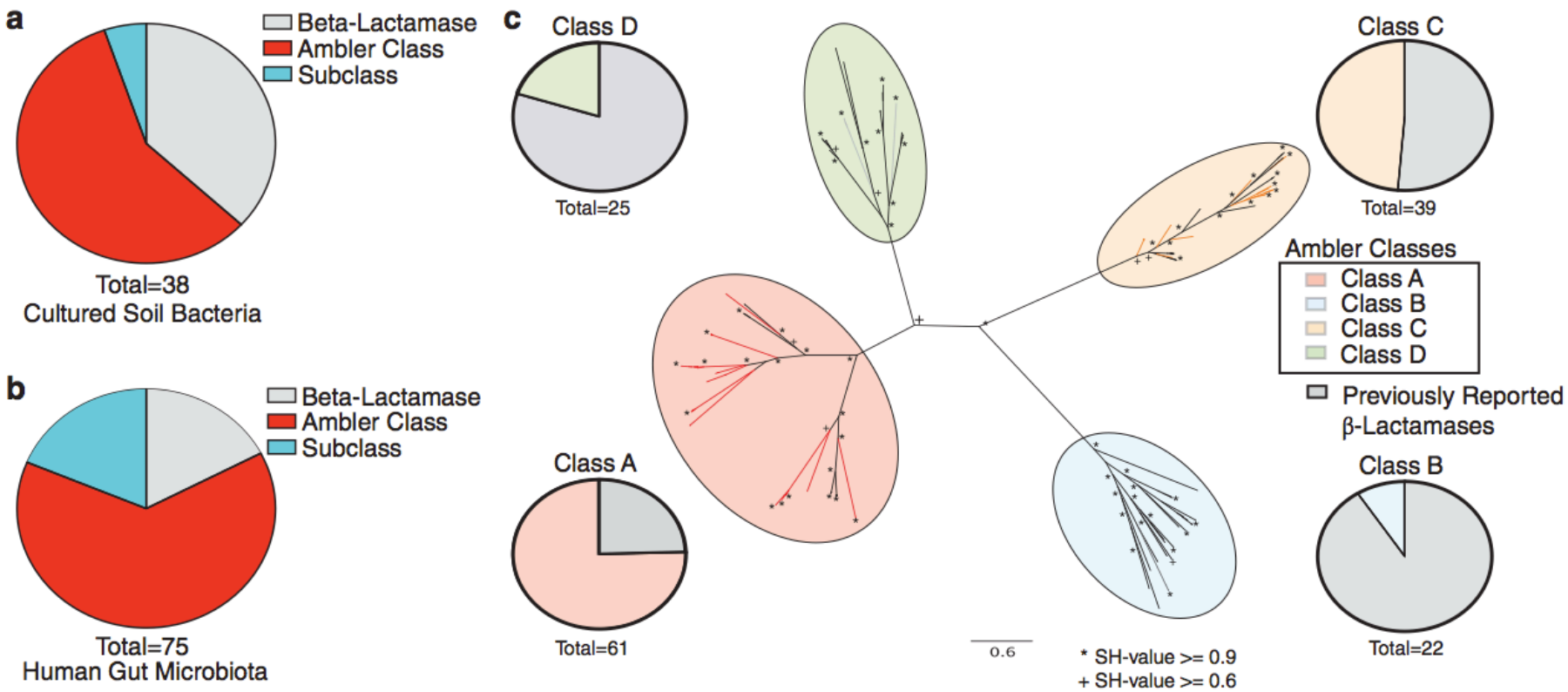
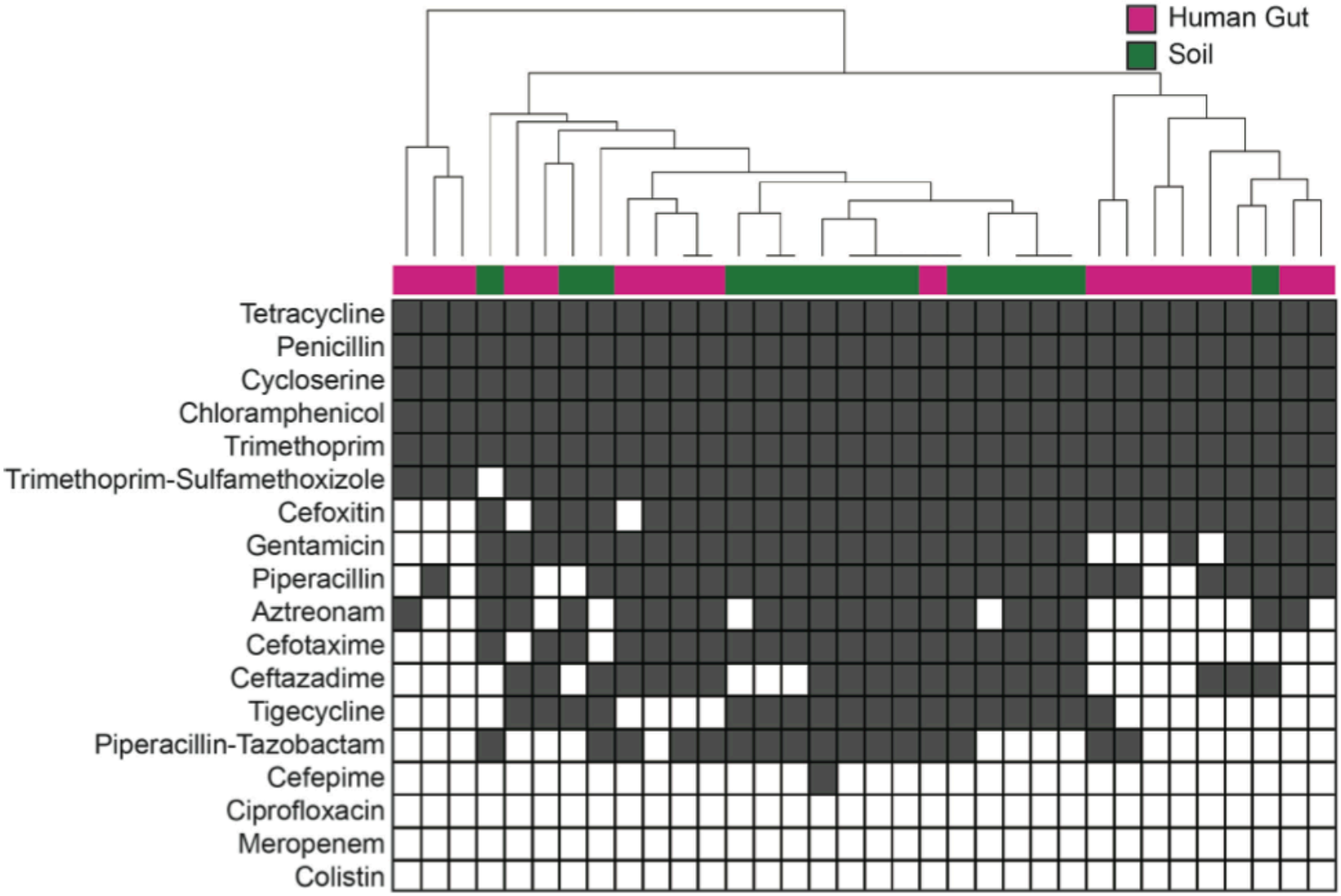


Figure 2 Resfams improves the resolution of annotation of antibiotic resistance proteins. The majority of β-lactamases identified from (**a**) MDR cultured soil bacteria and (**b**) human gut microbiota in functional metagenomic selections that are highly sequence divergent from any protein in the ARDB or CARD databases (<80% amino acid (a.a.) identity over 85% of target sequence) are annotated according to Ambler class (red) or subclass (cyan) level by Resfams. (**c**) All highly divergent β-lactamases from (**a**) and (**b**) that are annotated at the Ambler class and subclass level (colored branches) by Resfams accurately cluster on a phylogenetic tree with previously verified β-lactamases from all four Ambler classes (black branches). Pie charts in (**c**) depict the fraction of Resfams identified β-lactamases and previously verified β-lactamases represented on the phylogenetic tree in each Ambler class clade.

Antibiotic resistomes and ecology

- Environmental, nonpathogenic and commensal organisms have long been shown to harbor functional AR genes.
- It was only recently established that exchange of multiple classes of AR genes has occurred between nonpathogenic environmental bacteria and human pathogens.
- Environmental bacteria serve as potential reservoirs of AR genes primed for exchange with pathogenic bacteria.



Supplementary Figure S2: Phenotypic resistance profiles across ecologies. Binary heatmap (gray, resistance observed; white, no resistance) of phenotypic profiles of 16 soil microbiota (green) and 18 human gut microbiota (magenta) to 18 antibiotics. Samples are clustered using the Jaccard Index and resistance profiles are herarcically clustered.

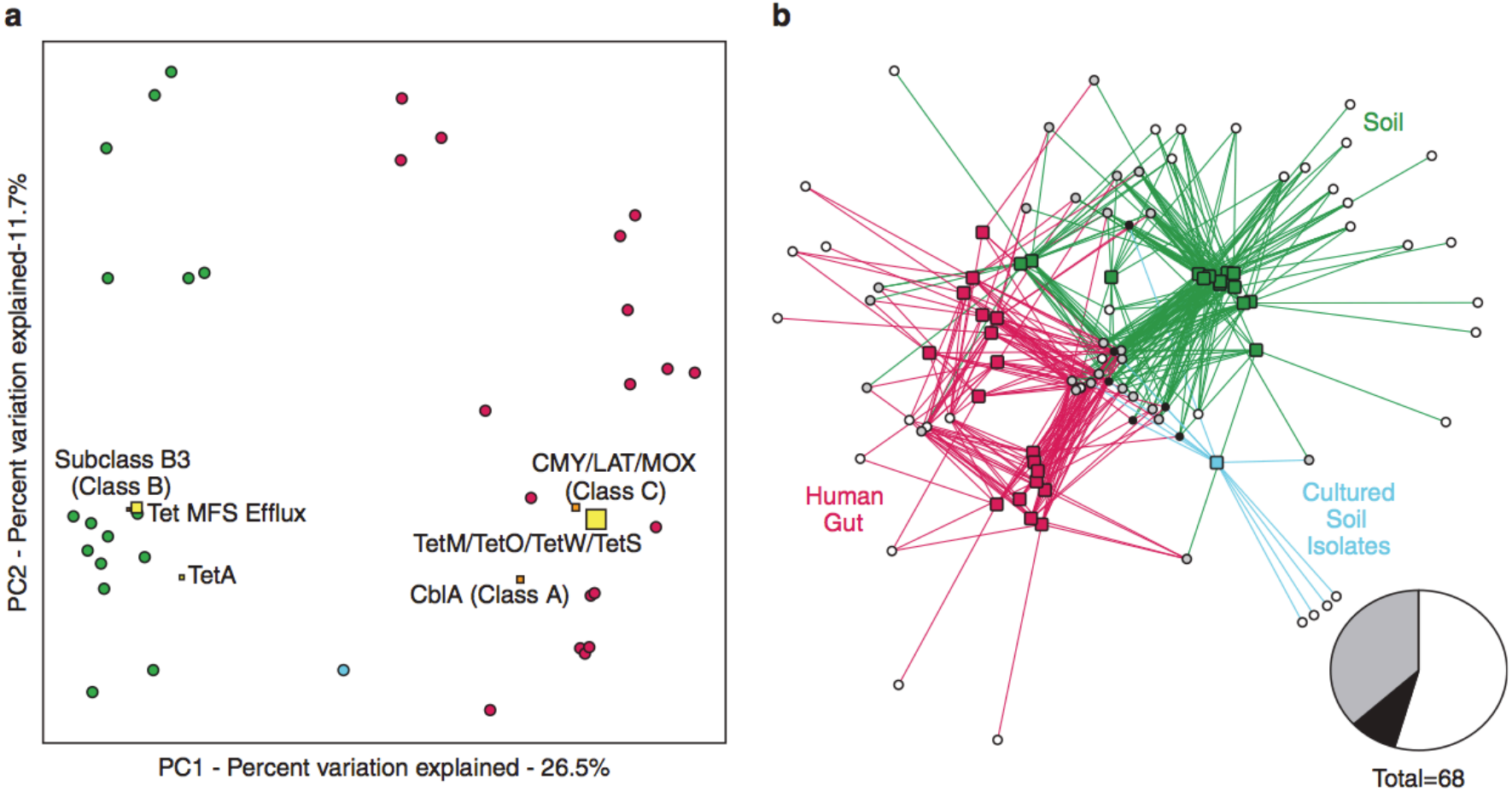


Figure 3 Resistomes differ by ecology. (**a**) Principal coordinate analysis (PCoA) plot depicting Bray-Curtis distances between resistomes of the soil (green), human gut (magenta) and MDR soil isolates (cyan), calculated using unique AR protein counts. Resistomes of different ecologies cluster separately ($P<0.001$, ANOSIM). Function biplot coordinates (squares) represent the weighted average of the top six most discriminating AR functions between ecologies across all samples. The size of the biplot squares represent the aggregate abundance of the unique AR family members. Separation of resistomes is heavily influenced by β-lactamase (orange squares) and tetracycline resistance functions (yellow squares). (**b**) Bipartite network diagram of normalized AR protein counts across all resistomes. Edges connect sample nodes (squares) to AR function (circles). Edges and sample nodes are colored by sample ecology (green, soil; magenta, human gut; cyan, MDR soil isolates) and AR functions are colored by extent of sharing across ecologies (white, unique to ecology; gray, shared between two ecologies; black, shared across all three ecologies). Inset pie chart represents the percentage of AR Resfams families that belong to each group.

- The separation of soil and gut resistomes is driven mainly by beta-lactamase and tetracycline resistance functions.
- Subclass B3 beta-lactamases are mainly associated with soil resistomes.
- CblA (class A) and CMY/LAT/MOX (class C) beta-lactamases are found mainly associated with human gut.
- Class A beta-lactamases in general are prevalent across all environments and samples (94% of soil resistomes and 89% of human gut resistomes).
- class C/class B = 0.42 in soil and 4.3 in gut

- There are **no significant differences in the number of distinct AR genes** between the human gut and soil microbiota when screened for functional AR.
- Resistomes cluster by ecology, suggesting that **the soil and human gut microbiota consist of functionally distinct AR gene architectures**.
- There is **no "core" resistome** – no single Resfams family that was shared across all samples.
- Only two families were shared across >50% of metagenomic samples (class A beta-lactamases and MFS Antibiotic efflux).
- Only six Resfams families were found in at least one sample from every habitat investigated.

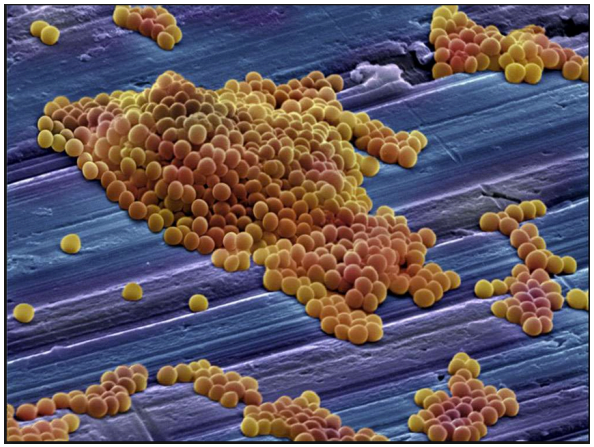
Antibiotic resistance potential encoded in microbial genomes

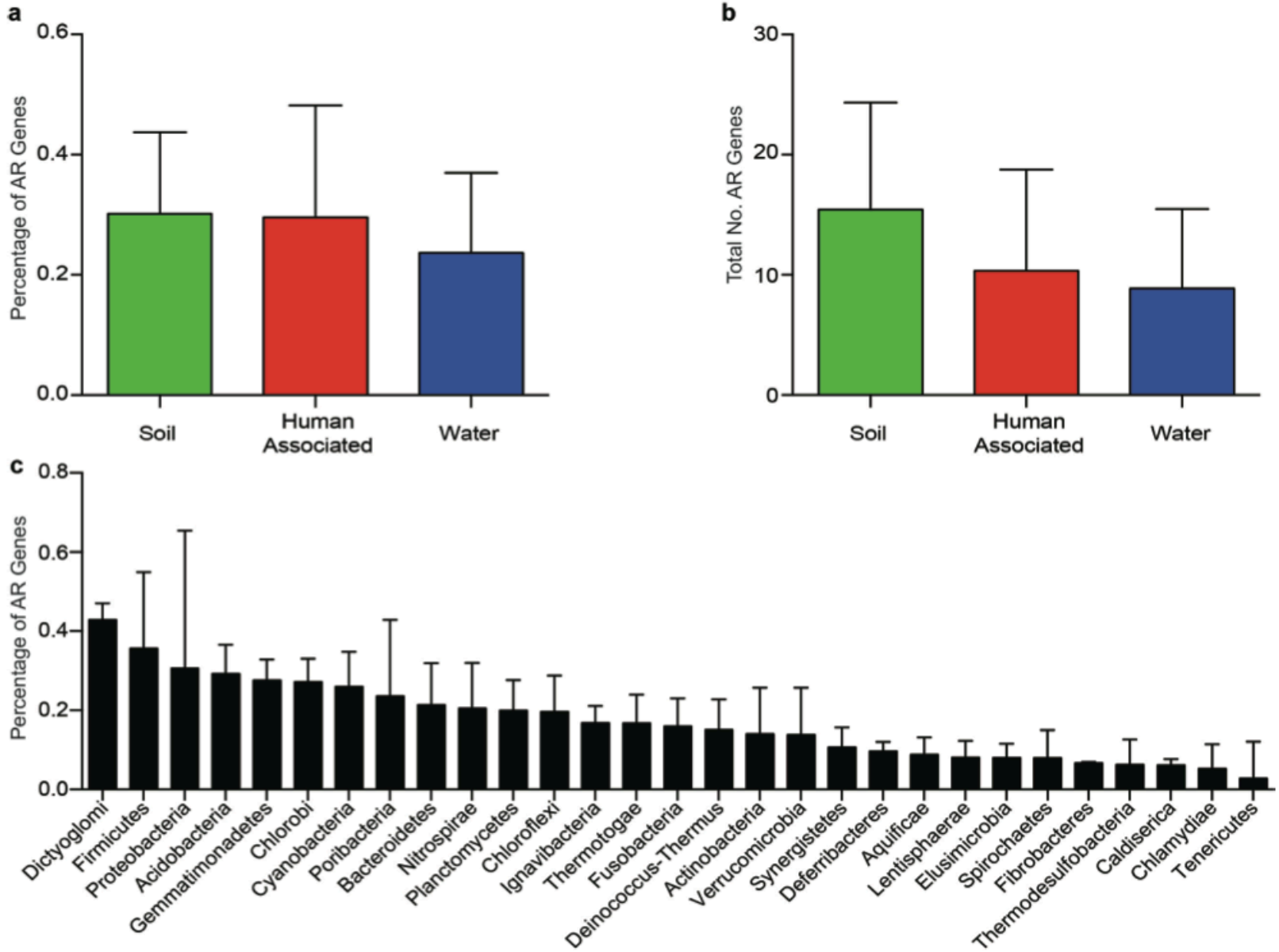
JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

10TH ANNIVERSARY

img INTEGRATED MICROBIAL GENOMES AND METAGENOMES

- 84% of microbial genomes investigated encode at least one AR gene.
- There is significant AR mechanism enrichment by bacterial phyla and habitat.





Supplementary Figure S4: Distribution of total antibiotic resistance (AR) genome composition across habitat and bacteria phyla. The total AR genome composition was calculated by taking the total number of genes annotated by Resfams profile HMMs as AR genes in the genome divided by the total number of genes in the genome. The total AR genome composition was averaged across (A) habitats and (C) bacterial phyla. Error bars represent one standard deviation. The raw number of total AR genes per genome was averaged across all genomes by (B) habitat. Error bars represent one standard deviation.

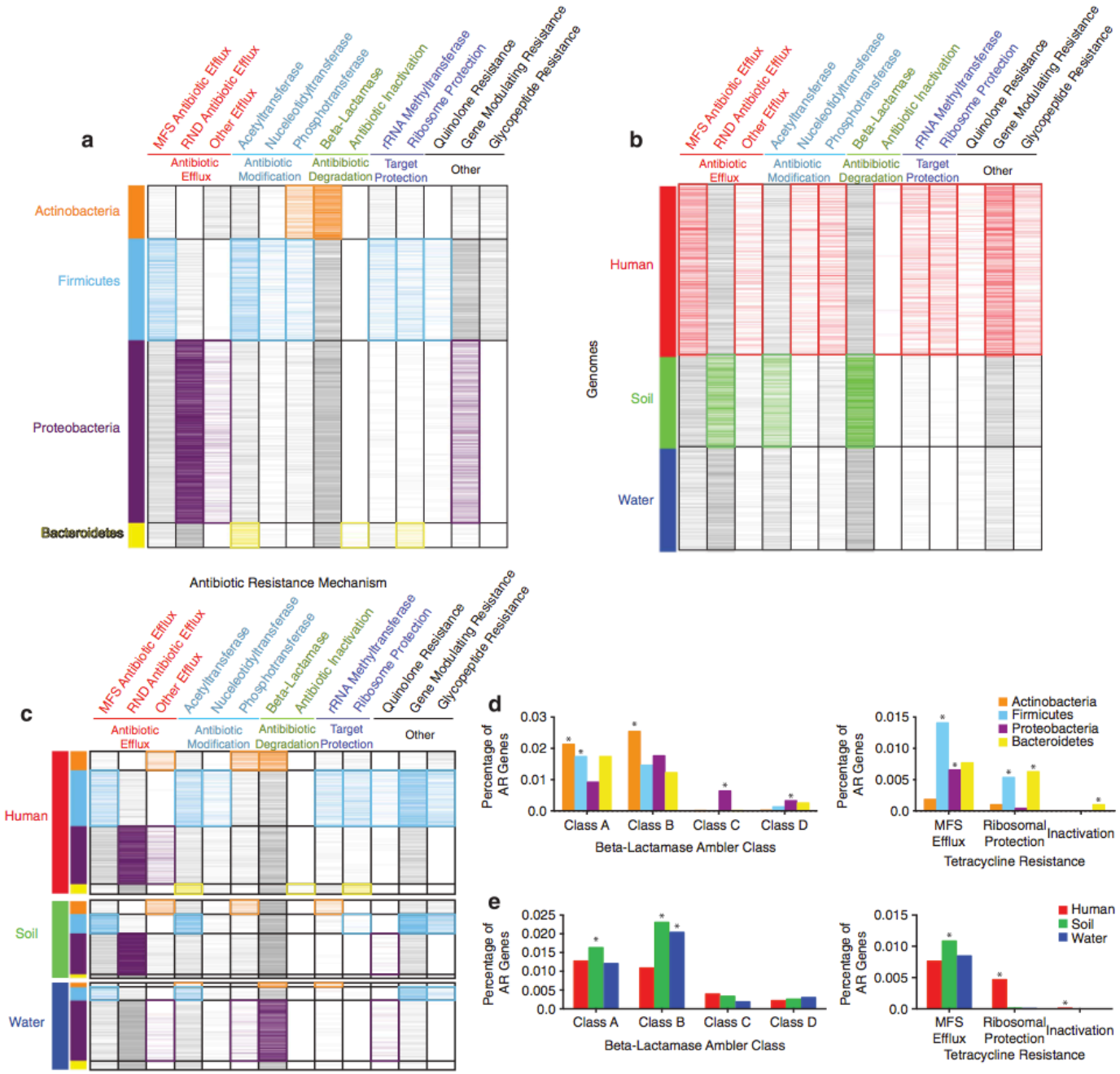
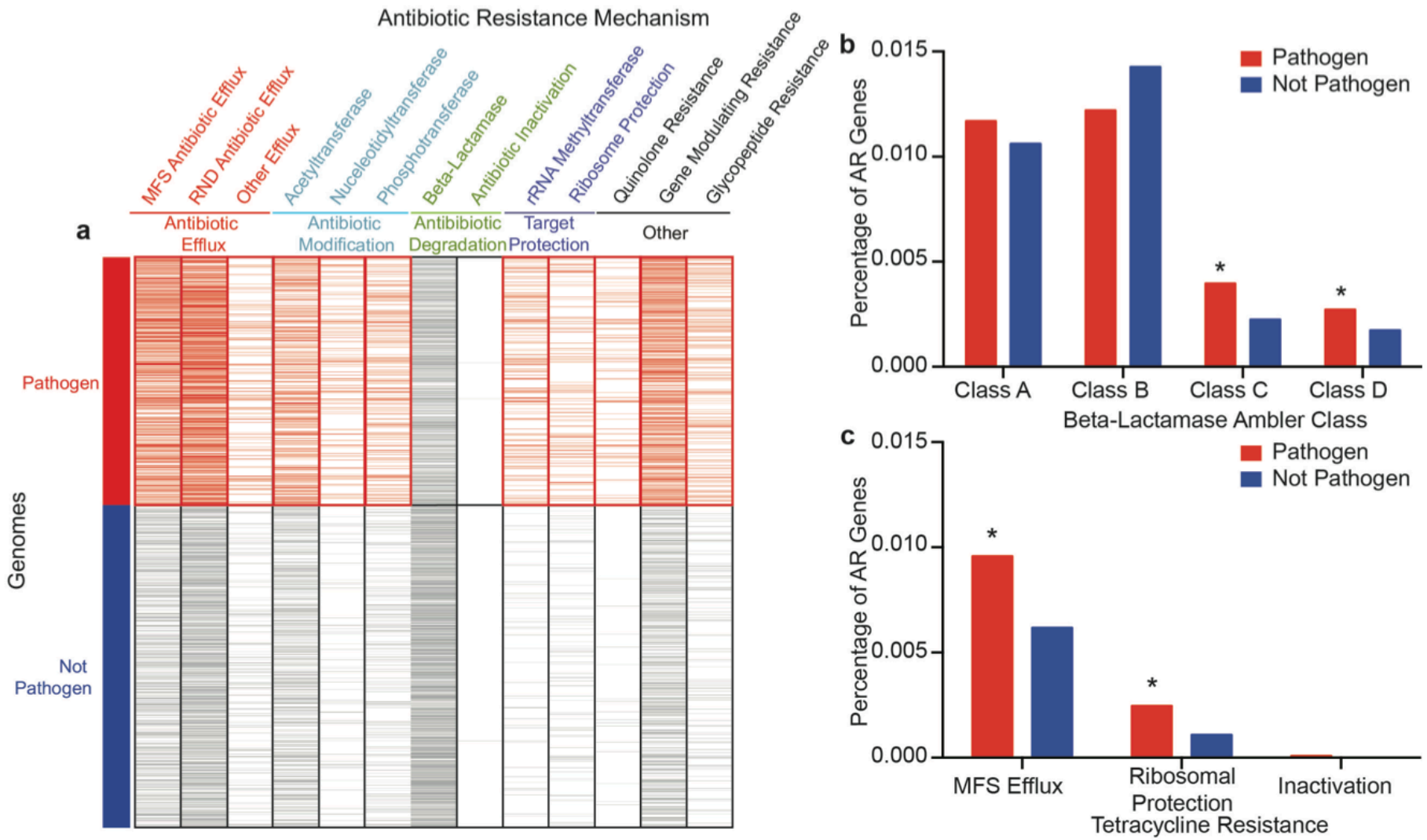


Figure 4 Resistomes annotated by Resfams across phylogeny and habitats of 6179 sequenced bacterial isolate genomes. Binary heatmaps of genomes organized by (**a**) phylogeny, (**b**) habitat and (**c**) phylogeny within habitat. Sections of the heatmaps are colored if a particular AR mechanism is significantly enriched within a particular phyla or habitat ($P<0.01$, Fisher's exact). Enrichment of β-lactamase Ambler class and tetracycline resistance functions is depicted across (**d**) phyla and (**e**) habitat (*$P<0.01$, Fisher's exact).



Supplementary Figure S6: Enrichment of antibiotic resistance functions in pathogens. (A) Binary heatmap of resistomes organized by pathogen status of 2,966 genome sequenced bacterial isolates. The heatmap is colored by enrichment of a particular AR mechanism within pathogenic or non-pathogenic organisms ($P < 0.01$, Fisher's exact). Enrichment of (B) β-lactamase ambler class and (C) tetracycline resistance functions within pathogenic or non-pathogenic organisms (*$P < 0.01$, Fisher's exact).

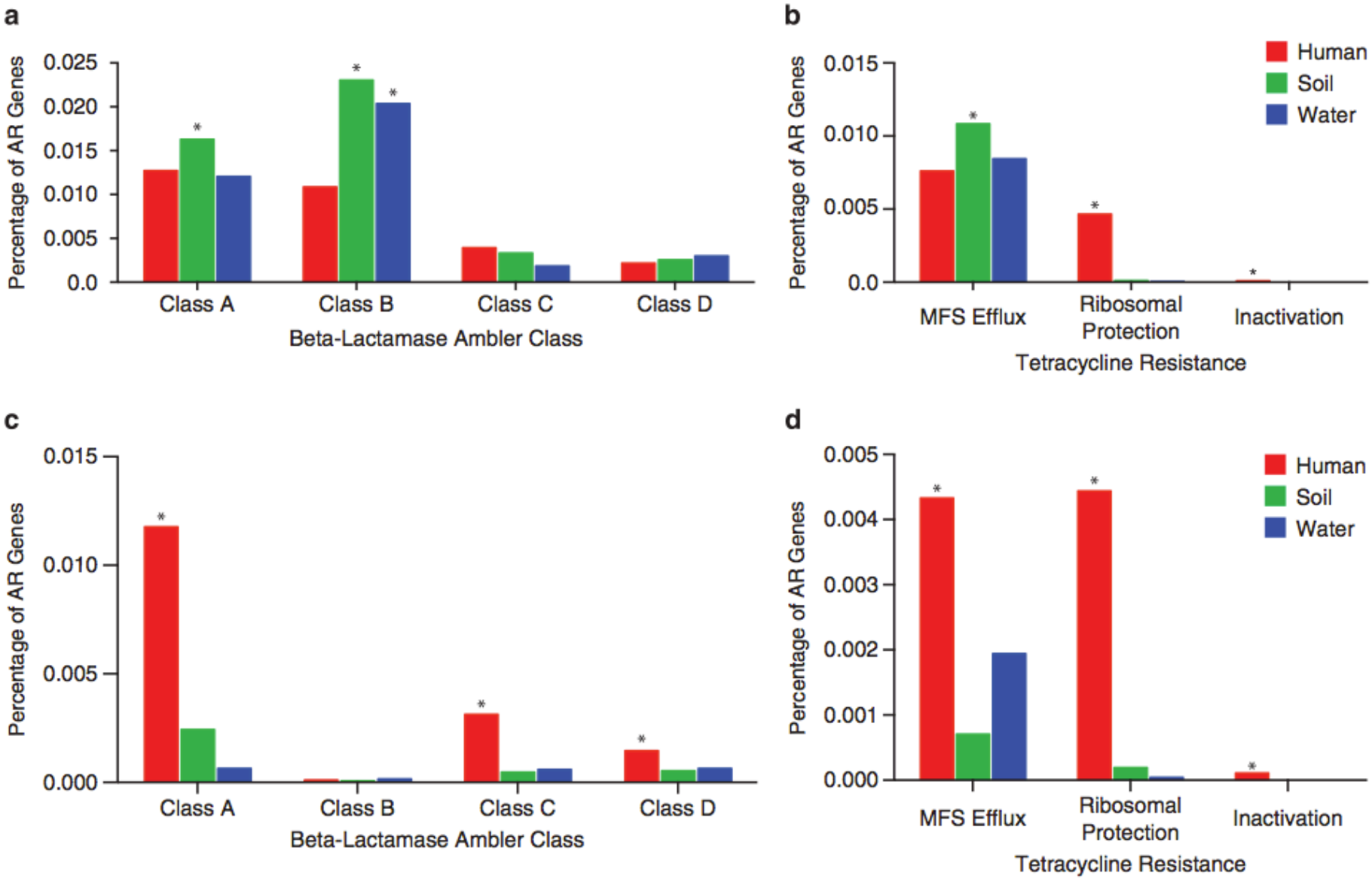


Figure 5 Annotation of antibiotic resistance across habitats using Resfams compared with pairwise sequence alignment. Predicted enrichment of β-lactamase Ambler classes and tetracycline resistance mechanisms (*$P<0.01$, Fisher's exact) in sequenced genomes across habitats using (**a**, **b**) Resfams family profile HMMs compared with (**c**, **d**) BLAST to the ARDB and CARD databases.

Conclusions

- Antibiotic resistomes cluster by ecology, with no core resistome shared between all samples.
- Pairwise sequence alignment method to characterize resistome composition is biased toward human-associated organisms, vastly underestimating the potential impact of environmental resistance reservoirs on AR in pathogens.
- By using a consensus model approach, we are able to significantly increase our ability to characterize highly diverse and understudied reservoirs of resistance while minimizing bias.