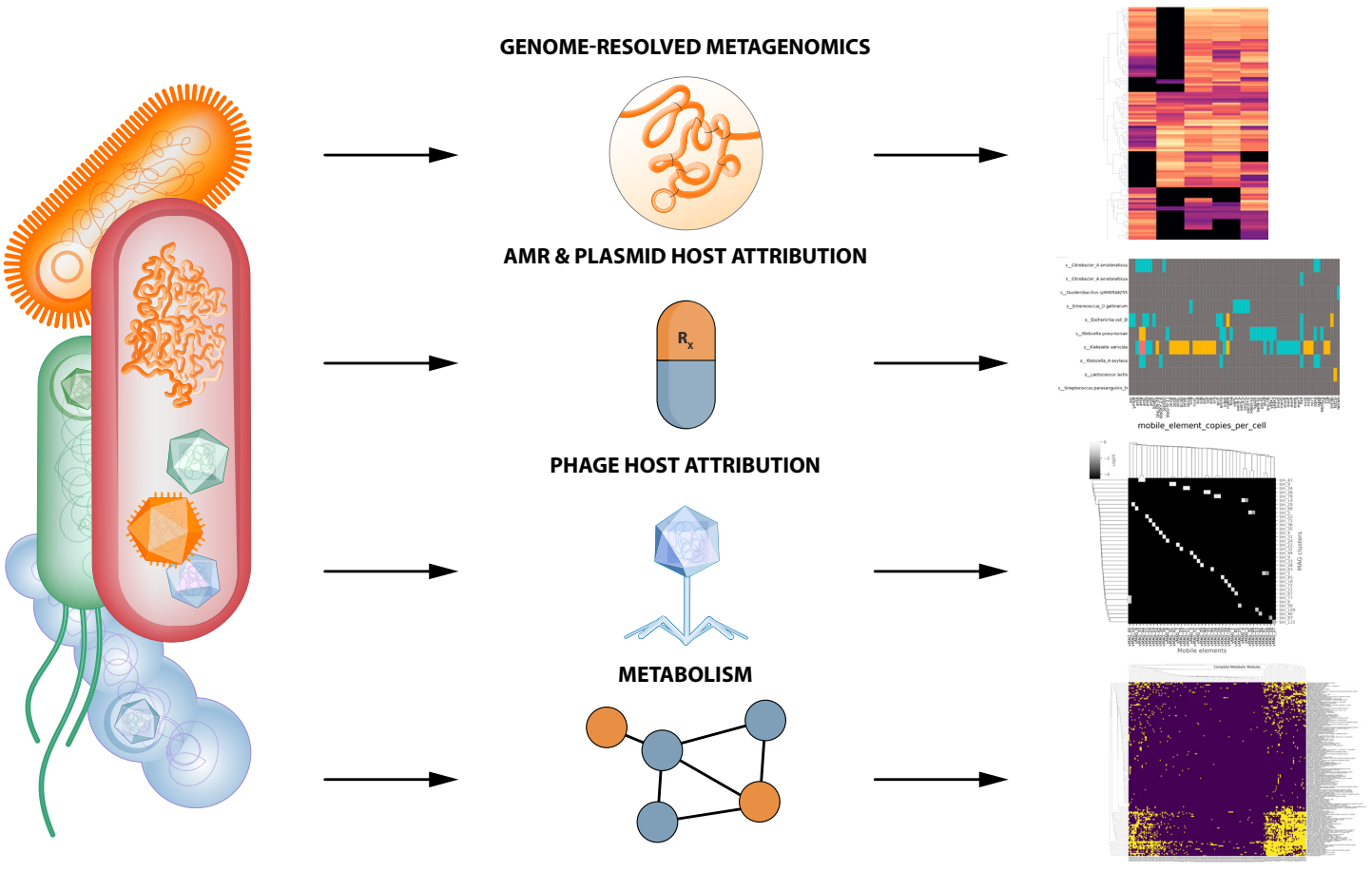
# Capture a complete picture of complex microbial communities, including the moving parts

Unlike conventional metagenomics approaches, the ProxiMeta™ Metagenomics Platform enables genome-resolved metagenomics, accurate host attribution of mobile genetic elements, and metabolic analysis and functional characterization—directly from complex communities, without culturing.

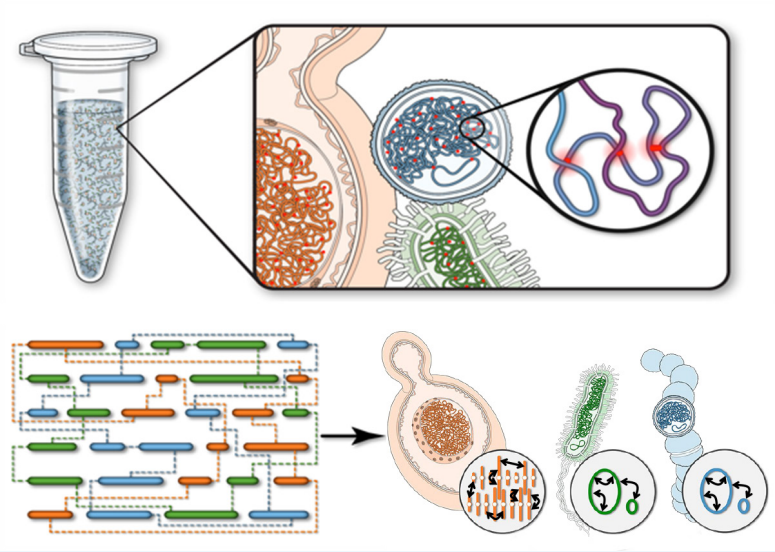



*Decoding metagenomic samples requires innovative technologies and computational tools. Phase Genomics' ProxiMeta Platform uses short-read sequencing technology to generate high-quality, genome-resolved metagenomes. This unique solution enables the identification of novel bacterial and viral species and strains (irrespective of culturability), accurate host attribution of mobile genetic elements (such as plasmids harboring antibiotic resistance genes and phages), and significantly improved annotation to discover new metabolic features and support functional genomics.*

# Ultra-long-range Sequencing Technology

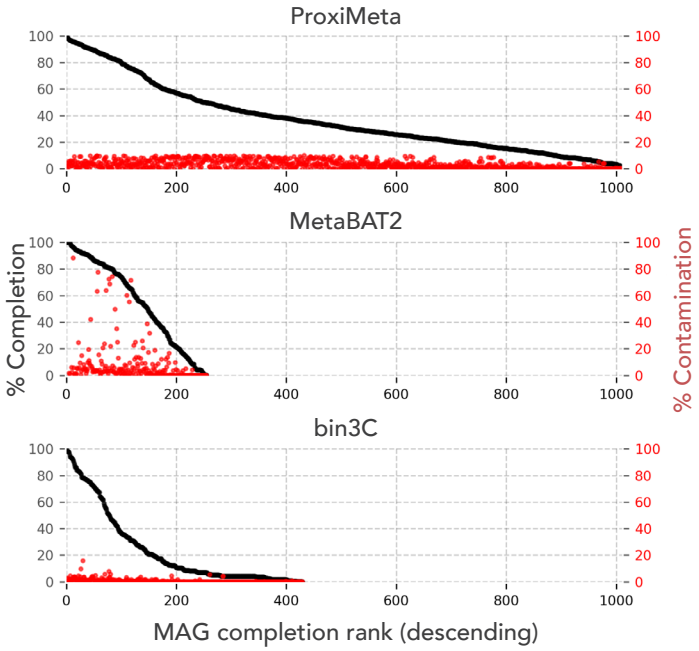
Established metagenomic sequencing strategies such as 16S/ITS rRNA sequencing and shotgun metagenomics offer taxonomic and some functional insights, but lack long-range genomic contiguity information. These methods simply can't tell with certainty which sequences originated from which organism in a complex microbial sample. As a result, metagenome assembly pipelines have to rely on *a priori* knowledge, statistical assumptions, and the accuracy of binning algorithms; leaving you with an incomplete and/or incorrect picture of the members of the community, their basic biology, and the interactions between them.

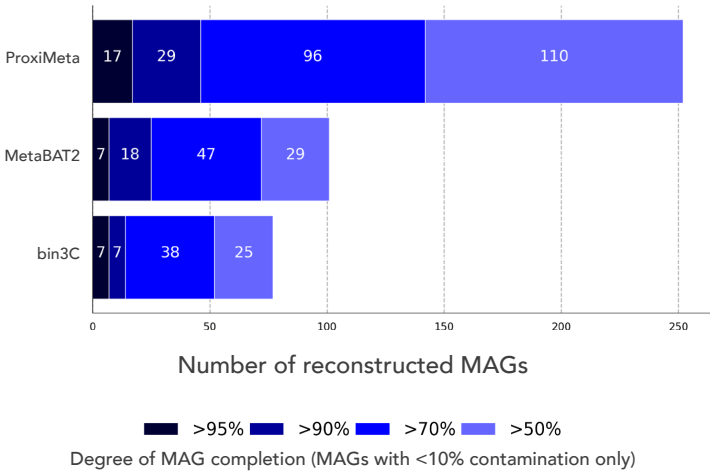
The ProxiMeta Platform combines proximity ligation (Hi-C) and novel computational approaches to achieve ultra-long-range sequencing. Proximity ligation is utilized in of a suite of chromosome conformation capture techniques originally devised to study the spatial organization of chromatin.[1,2] It employs cost-effective, high-throughput, short-read sequencing to identify the genomic loci that are co-located in three-dimensional space, but may be separated by significant distances along the genome. These direct and quantitative measurements of DNA sequences that interact *in vivo* greatly improve the quality and reliability of assembled genomes from metagenomic samples.



**Proximity ligation-based metagenome deconvolution.** Inter- and intrachromosomal DNA contacts (red foci) within intact microbial cells in metagenomic samples are trapped by *in vivo* crosslinking. Crosslinked DNA is fragmented and proximity ligated, creating chimeric junctions between sequences originating from the same cell. These chimeric junctions are recovered, converted into a sequencing library, and subjected to paired-end, short-read sequencing. The resulting genomic contiguity information is used to deconvolute contigs into their original cellular groupings, including chromosomes (orange) and mobile genetic elements (green and blue). Metagenome deconvolution with the ProxiMeta Platform is accomplished directly from crude samples, without the need for culturing or the extraction of high-molecular weight DNA.

The ProxiMeta computational tool combines proximity ligation data with short- or long-read metagenome shotgun sequencing data (reads or assemblies) to cluster contigs originating from the same cell. Binning is performed on the metagenomic assembly, using industry-leading binning tools. The binning process is augmented by applying long-range contiguity information to the assembly. A connectivity network is generated from multiple binning solutions—with and without proximity ligation information. A supervised algorithm is subsequently employed to deconvolve this connectivity network and export the highest confidence set of genomes or bins. This produces more, higher-quality genomes than metagenomic binning or other proximity-based approaches alone. Proximity-augmented metagenome-assembled genomes (MAGs) are more complete and less contaminated than metagenome assemblies derived from shotgun sequencing approaches that lack linkage information.




**The ProxiMeta computational tool improves metagenome deconvolution and assembly.** Metagenome-assembled genomes (MAGs) were recovered from a short-read shotgun assembly of a complex wastewater sample using three different binning methods. The ProxiMeta algorithm combines proximity ligation linkage information with traditional binning to produce more, higher-quality MAGs than using either a conventional binning tool (MetaBAT2)[3] or a Hi-C based method (bin3C)[4] alone. In plots on the left, each point along the x-axis represents a MAG, ranked in order of decreasing completeness (using CheckM).[5] The level of contamination of each genome is indicated with its corresponding red dot. Only those MAGs with <10% contamination are shown on the right. These are grouped by degree of completeness (blue bars).

## Resources

### Publications

- Burton JN, et al. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3* (Bethesda, Md.) 2014; 4(7): 1339–1346. doi: 10.1534/g3.114.011825.

### ProxiMeta Kits and Services

- phasegenomics.com/products/proximeta/

### ProxiMeta Analysis and Sample Reports

- proximeta.phasegenomics.com

# Applications

The use of contact probability maps for the deconvolution of metagenomic samples was first demonstrated in 2014 by scientists from Jay Shendure's lab at the University of Washington. They successfully clustered the genome content of fungal, bacterial, and archaeal species with >99% agreement to published reference genomes in two synthetic metagenome samples.[6] Phase Genomics was founded in 2015 to develop robust ultra-long-range sequencing platforms and innovative computational tools to enable deeper insights into the complexity, architecture, variation, function, and complexity of metagenomes.

Ultra-long-range sequencing takes the guesswork out of metagenomic deconvolution and yields genome-resolved metagenomes. Whether you are working with biological or environmental samples, the ProxiMeta™ Metagenomics Platform uniquely enables:

- Genome-resolved metagenomics
- Host attribution of mobile genetic elements, such as antimicrobial resistance genes
- Accurate viral genome reconstruction and host attribution of phages
- Improved metabolic annotation

This unlocks the power of metagenomics for a variety of applications, including surveillance and control of antibiotic resistance, viral therapy, fecal microbiota transplantation (FMT), bioprospecting, food processing, bioremediation, and other biotechnological applications.

Learn more about ProxiMeta applications at
phasegenomics.com/applications/metagenomics-microbiology/

# Host attribution of mobile genetic elements, including antimicrobial resistance genes

*Enabling effective surveillance and control of antibiotic resistance*

Horizontal gene transfer of mobile genetic elements (MGEs) is a driving force in the genetic composition and diversity of environments with a high microbial density. In metagenomic samples, up to 25% of the genetic material can be composed of MGEs,[7] such as bacterial plasmids, phages associated with bacteria or viruses, integrons, and transposons—including those carrying antibiotic resistance genes (ARGs). Yet very little is known about the diversity and structure of MGEs, and the mechanisms by which they are transmitted.

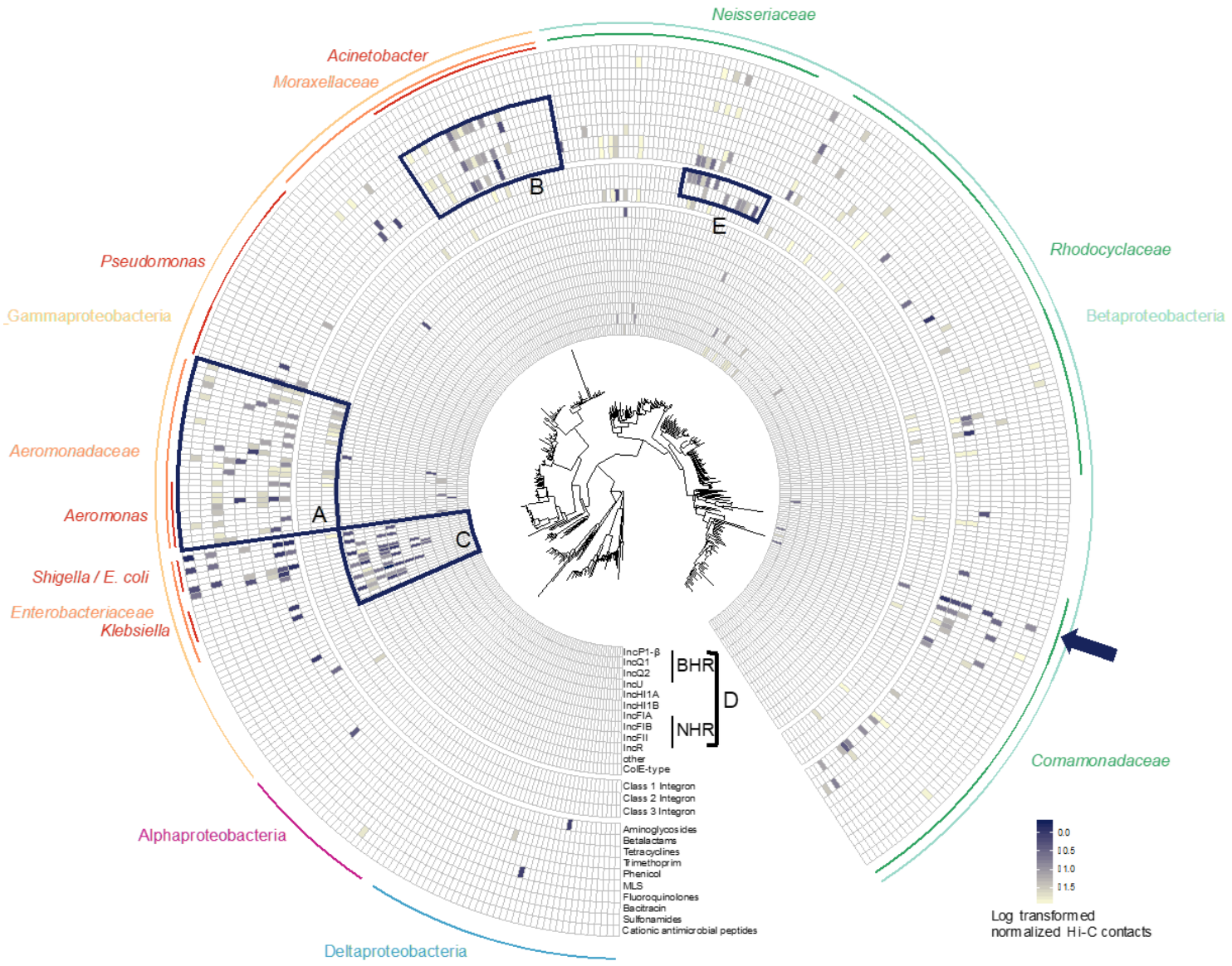
In metagenomic shotgun sequencing, the DNA from a complex, uncharacterized population of micro-organisms is fragmented and mixed during library preparation—leaving it to assumption-driven computational methods to untangle. Even if MGEs are identified, it is virtually impossible to attribute them to the individual organisms and strains with which they were associated at the time the sample was taken.

Attribution of MGEs to bacterial hosts is straightforward with the ProxiMeta Platform—since it is based on direct, physical and quantitative measurements of the interactions between the genomes assembled from a metagenomic sample and the MGEs identified using shotgun sequencing data. This linkage information is indispensable in the study of complex microbial communities, particularly those containing previously unsequenced pathogens which may be reservoirs of disease, toxicity, or antibiotic resistance. A thorough understanding of the diversity, abundance, distribution, and host range of ARGs, and the MGEs that mediate their spread, is essential to design effective surveillance, prevention, and control strategies.

## Resources

### Publications


- Stadler T, et al. Linking the resistome and plasmidome to the microbiome. *ISME J* 2019; 13: 2437–2446. doi: 10.1038/s41396-019-0446-4.
- Bickhart DM, et al. Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biol* 2019; 20: 153. doi: 10.1186/s13059-019-1760-x.




**Linking the plasmidome and resistome to the microbiome.** Physical intracellular linkage information obtained from ProxiMeta analysis was used to identify the *in situ* host range of MGEs among Proteobacteria in a wastewater sample. Each tip of the circular phylogenetic tree in the center represents a bacterial genome. Each concentric circle corresponds to an ARG, plasmid marker or integron. The presence or absence of a link is shown in the heat map, with color shading representing the intensity of the Hi-C linkage signal (normalized according to the abundance of clusters/elements). IncQ plasmids and class 1 integrons had the broadest host range in this sample. The analysis identified bacteria belonging to *Moraxellaceae*, *Bacteroides*, and *Prevotella*, and especially *Aeromonadaceae* as the most likely reservoirs of ARGs in this community.

# Accurate viral genome reconstruction and host attribution of phages

*Supporting emerging applications such as viral therapy and FMT*

Viruses are ubiquitous and are the most abundant biological entities in the biosphere. Yet, they represent the largest unexplored genetic information space on earth. Viruses infect bacteria, archaea, and eukaryotes. As important vectors of horizontal gene transfer, they shape the evolution and population dynamics of their microbial hosts, as well as the natural and man-made ecosystems in which they occur.

The ProxiMeta Platform enables significant improvements in the quantity, completeness, and quality of reconstructed viral genomes. In addition, the pipeline employs a sophisticated statistical approach to dynamically assign bacterial hosts to viruses with high sensitivity and specificity, making it the first platform for reliable viral host attribution in metagenomic samples.

As the "perfect predators" of microbial communities,[8] phages and DNA viruses impact the genomic plasticity of their hosts and ecosystems, with functional impacts all the way up the proverbial food chain. As such, the ProxiMeta Platform provides a powerful tool for emerging applications, such as viral therapy and fecal microbiota transplantation (FMT).

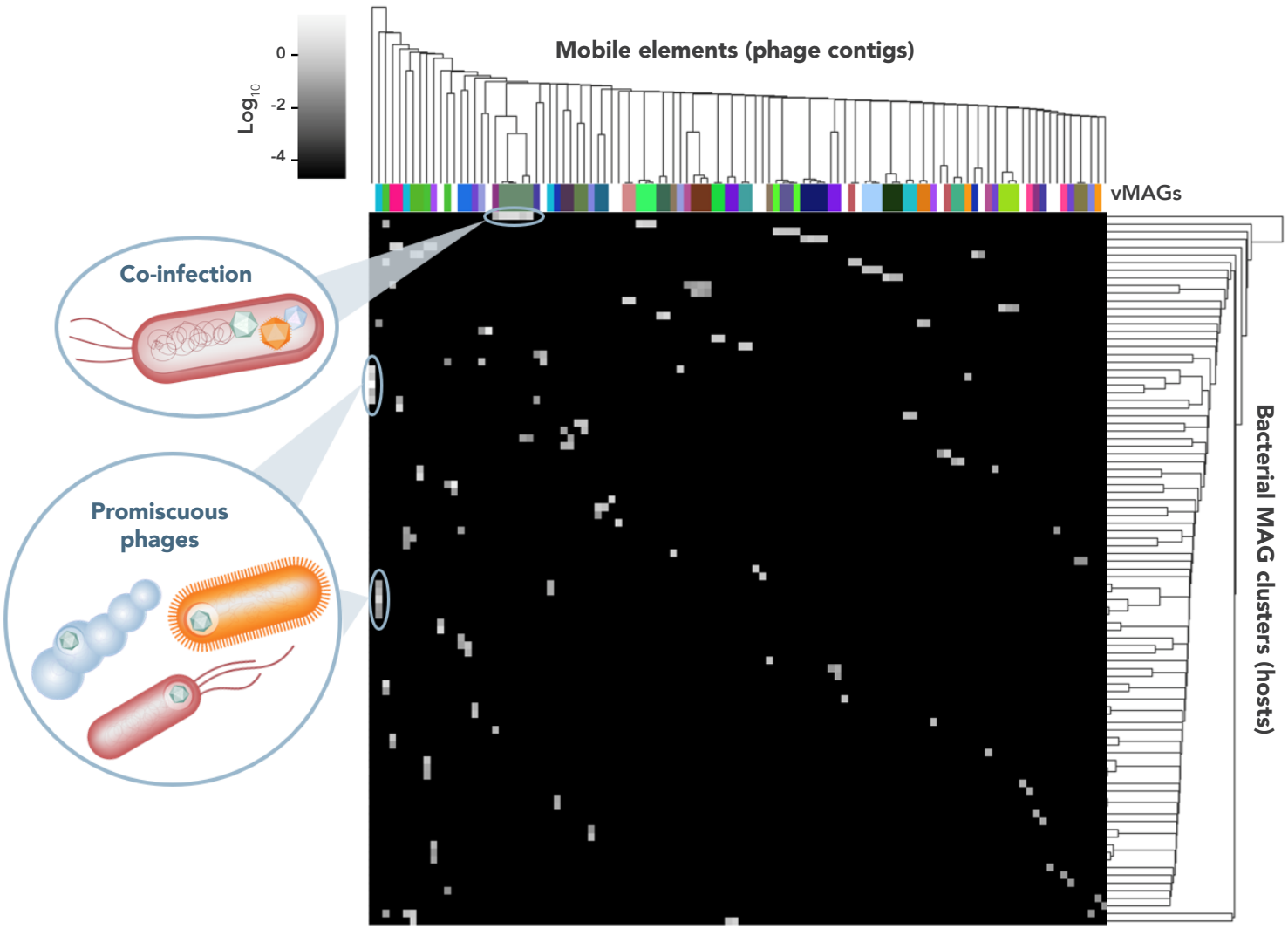
## Resources

### Application Note

- Accurate viral genome reconstruction and host attribution with proximity-guided metagenomics. https://phasegenomics.com/wp-content/uploads/2021/06/ProxiMeta_Phage-Analysis-App-Note_June-2021.pdf

### Publications

- Uritskiy G, et al. Accurate viral genome reconstruction and host assignment with proximity-ligation sequencing. *bioRxiv* 2021.06.14.448389. doi: 10.1101/2021.06.14.448389.




**Bacterial hosts identified for viral contigs reconstructed from an animal fecal sample with the ProxiMeta Platform.** The color map represents the log of the estimated average copy count of each phage genome in its host. Colors above phage contigs designate the vMAG that they belong to. Only viral contigs from near-complete vMAGs are shown. Call-outs emphasize examples of putative co-infection events (horizontal lines) and promiscuous phages infecting multiple hosts (vertical lines).

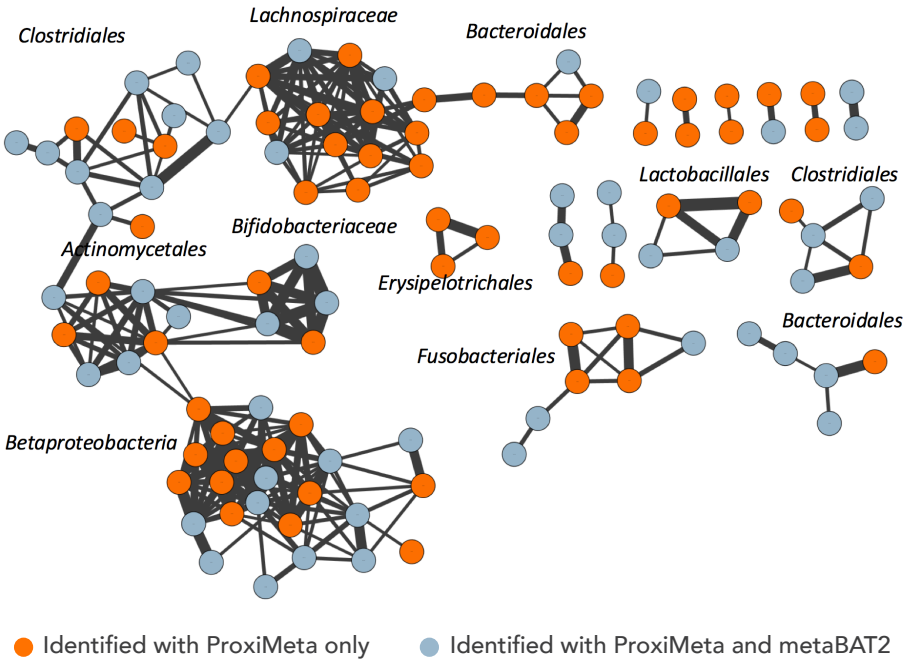
# Improved metabolic discovery and annotation

*Allowing for the discovery of new metabolic features and supports accurate functional analysis*

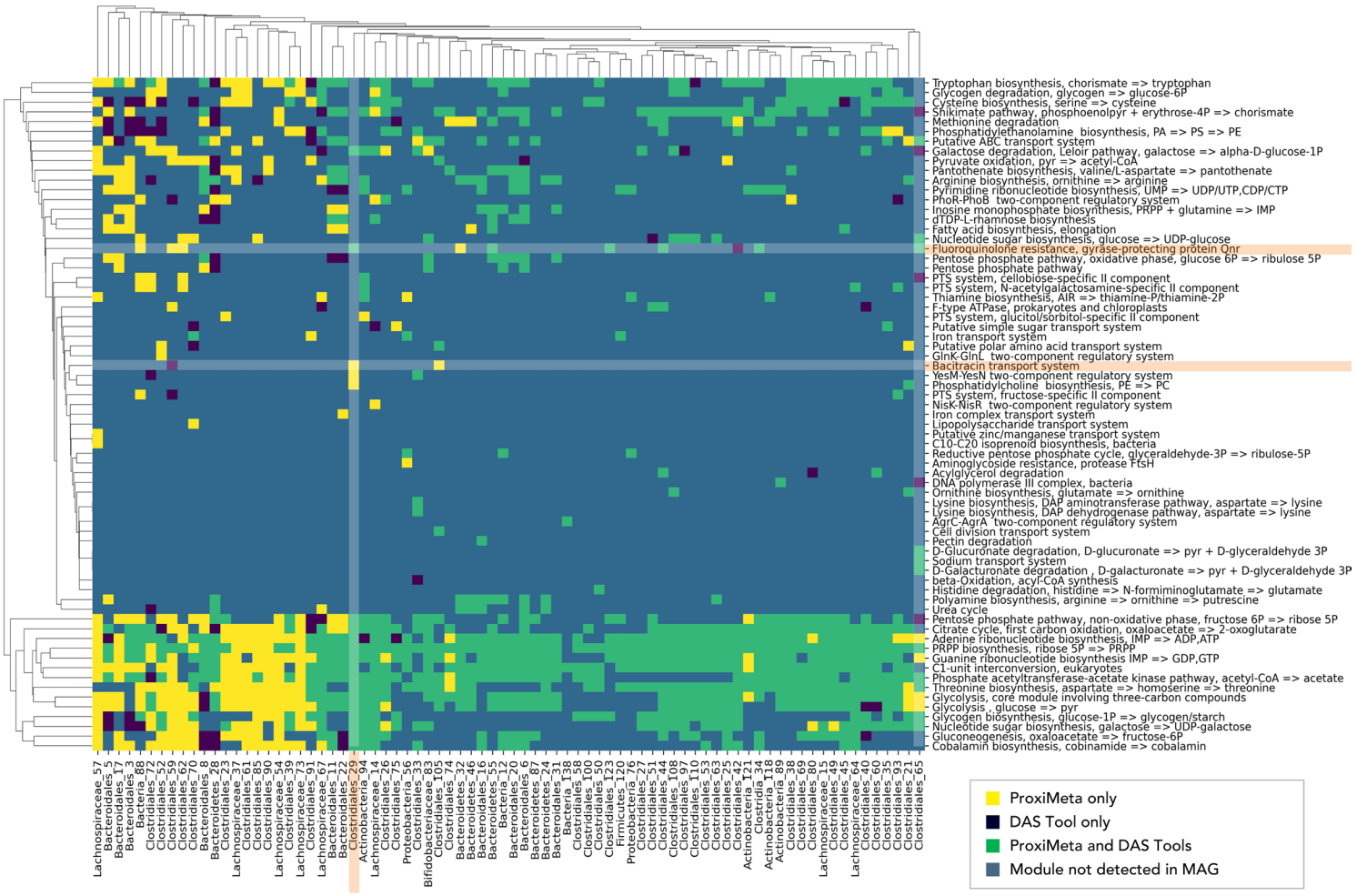
Earth may be home to as many as one trillion ($10^{12}$) microbial species, of which 99.999% remain to be discovered.[9] The identification and characterization of new micro-organisms is not only important for the prediction of biodiversity and ecological studies, but also for the search of valuable products from nature—known as "bioprospecting". Discovery and functional metagenomics from complex environmental samples have traditionally been hampered by the fact that many microbes are unculturable or present in very low abundance, and the reliance of data analysis methods on annotated reference genomes. In addition, functional analysis of metagenomic samples is hampered by the loss of intracellular contiguity information during sample preparation, caused by the bulk extraction of input DNA.

The ProxiMeta Platform uniquely enables discovery and bioprospecting, as it does not require culturing, or rely on any *a priori* data. The ability to reconstruct high-quality genomes for microbes with as little as 0.05% cellular abundance in a sample,[10] combined with the preservation of genome contiguity information enable the recovery of more genomes and significantly improved annotation and superior metabolic pathway analysis.

The benefits of more accurate and complete genome recovery for metabolic pathway analysis are illustrated on the next page. Even though the fecal sample from a healthy donor analyzed in this study was expected to be relatively "unremarkable" in terms of metabolic annotation, this study produced evidence of multi-drug resistance. A module associated with fluoroquinolone resistance was identified in six MAGs with both analytical approaches, in three additional MAGs with the ProxiMeta Platform only, and in one MAG as a potential false positive with binning/DAS Tool. In addition, a bacitracin transport system was identified with the ProxiMeta Platform only in two MAGs (and was incorrectly attributed to a third MAG using binning/DAS Tool). Only the ProxiMeta Platform identified both antibiotic resistance mechanisms in one of the MAGs (Clostridiales_29). These results suggest that the limitations of conventional shotgun sequencing and binning could have considerable consequences in applications such as microbiome analysis or pathogen surveillance.




**Discover more.** Data from a complex wastewater sample was analyzed using ProxiMeta, or by conventional binning with MetaBAT2. Resulting genomes reported to be >80% complete were mapped against one another to measure their similarity. Orange nodes (circles) represent genomes that were found only with ProxiMeta, whereas blue nodes correspond to genomes found in both the ProxiMeta and MetaBAT2 outputs. The thickness of the edges (lines) connecting the nodes represents average sequence identity. The thickest edges connect genomes that are likely sub-strains of the same species. These results demonstrate the power of the ProxiMeta Platform to deconvolve both distant and closely related genomes from a single, complex sample.




**Comparison of metabolic modules identified in a short read assembly of a human fecal sample.** Of the modules discovered with the ProxiMeta Platform only (yellow), 89.5% were confirmed to be accurate. Of the modules identified using binning/DAS Tool only (black), 84.8% were confirmed to be false positives introduced by conventional binning. Highlighted modules provide evidence of antibiotic resistance in several MAGs. Only high-quality MAGs (completion >50%, contamination <10%) were included in the analysis. The MAGs listed on the x-axis exclude ProxiMeta MAGs for which a corresponding MAG with a sequence overlap ≥50% could not be found in the DAS Tool bin set.

## Resources

### Application Note

- Discover more metabolic features with proximity-guided metagenome deconvolution.
  https://phasegenomics.com/wp-content/uploads/2021/02/ProxiMeta_Metabolism-App-Note_Feb-2021.pdf

## References


1. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; 326 (5950): 289–293. doi: 10.1126/science.1181369.
2. Van Berkum NL, et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* 2010; 39: e1869. doi: 10.3791/1869.
3. Kang DD, et al. MetaBAT2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; 7:e7359. doi: 10.7717/peerj.7359.
4. DeMaere M and Darling A. bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes. *Genome Biol* 2019; 20: 46. doi: 10.1186/s13059-019-1643-1.
5. Parks DH, et al. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2014; 25: 1043–1055. doi: 10.1101/gr.186072.114.
6. Burton JN, et al. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3* (Bethesda, Md.) 2014; 4(7): 1339–1346. doi: 10.1534/g3.114.011825.
7. Qu A, et al. Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLOS ONE* 2008; 3(8):e2945. doi: 10.1371/journal.pone.0002945.
8. Strathdee S and Patterson T. The perfect predator: a scientist's race to save her husband from a deadly superbug. New York: Hachette Books, 2019. ISBN: 9780316418089.
9. Locey KJ and Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA* 2016; 113 (21): 5970–5975. doi: 10.1073/pnas.1521291113.
10. Stadler T, et al. Linking the resistome and plasmidome to the microbiome. *ISME J* 2019; 13: 2437–2446. doi: 10.1038/s41396-019-0446-4.


PHASE GENOMICS

info@phasegenomics.com
www.phasegenomics.com

Phone: 1-833-742-7436
Twitter: @PhaseGenomics