

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

Unlike conventional metagenome sequencing approaches, the ProxiMeta™ Platform provides the long-range information needed to assemble high-quality metagenomes and associate mobile genetic elements with their hosts.

Introduction

Metagenomic samples are treasure troves of information. Decoding that information to understand evolution, ecology, health, and disease requires innovative technologies and computational tools. Established metagenomic sequencing strategies offer different levels of insight (Table 1), 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 deconvolution and assembly pipelines have to rely on *a priori* knowledge, statistical assumptions and binning algorithms, leaving you with an incomplete and/or incorrect picture of the members of the community, and the genetic interactions between them.

Phase Genomics' ProxiMeta Metagenome Deconvolution Platform employs proximity ligation technology (also known as Hi-C) to capture physical interactions between sequences within the same cell. The ProxiMeta computational tool augments metagenomic binning with this additional layer of linkage information to generate more and higher-accuracy genomes from highly complex microbial communities than traditional metagenomic approaches.

This unique solution enables the identification of novel species and strains, irrespective of culturability, and provides for accurate annotation of mobile genetic elements (MGEs)—such as plasmids harboring antibiotic resistance genes (ARGs)—within host genomes.

Table 1. Comparison of commonly used metagenomic analysis methods and the ProxiMeta Platform.

	16S/ITS rRNA Sequencing	Shotgun metagenomics	ProxiMeta Platform
Information interrogated	Bacterial 16S rRNA or fungal ITS sequences only	All genomic material in a metagenomic sample	Physical interactions between genomic sequences in the same cell
Library construction	Amplify, barcode, and pool 16S/ITS amplicons	Fragment genomic DNA, ligate barcoded adapters, amplify library	Crosslink DNA <i>in vivo</i> , fragment, ligate proximity junctions, ligate barcoded adapters, amplify library
Computational strategy	Alignment to reference 16S/ITS database	Alignment using reference genome and marker databases, binning	Combines Hi-C data with short or long reads (no reference needed) to augment binning approaches
Key advantages	High sensitivity and coverage, low cost per sample	High resolution (species/strain), functional information	More complete, higher-quality assemblies, high-confidence host attribution of mobile elements
Disadvantages	Low resolution (genus/species only), no functional information	Low coverage, incomplete and contaminated assemblies, higher cost per sample	Not a standalone method (shotgun assembly or reads also required)

Experimental Strategy: Proximity Ligation

Proximity ligation or Hi-C is one of a suite of "chromosome conformation capture" (3C) techniques originally devised to study the spatial organization of chromatin.^{1,2} Hi-C 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 in the linear genome. This powerful methodology has enabled significant improvements in genome assembly of humans and other species³, as well as structural variant and epigenetic analysis.⁴ In addition, it has unlocked many applications in metagenomics and microbiology.

The use of Hi-C contact probability maps to deconvolute 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.⁵ Their experimental approach is summarized in Figure 1.

Proximity ligation library preparation has subsequently been streamlined and optimized by Phase Genomics, and is available in the form of easy-to-use ProxiMeta Kits (which include bioinformatic analysis), or as a service.

Computational Strategy: ProxiMeta

The ProxiMeta computational tool was developed by Phase Genomics bioinformaticians to harness the power of proximity ligation data for metagenomic applications. The tool combines Hi-C data with short- or long-read metagenome shotgun data (reads or assemblies) to cluster contigs originating from the same cell. Binning is performed on the metagenomic assembly, using industry-leading binning tools (e.g. MetaBAT2).⁶ The binning process is augmented by applying Hi-C connectivity information to the assembly. A connectivity network is generated from multiple binning solutions—with and without proximity information. A supervised algorithm is subsequently employed to deconvolve this connectivity network and export the highest confidence set of genomes or bins.

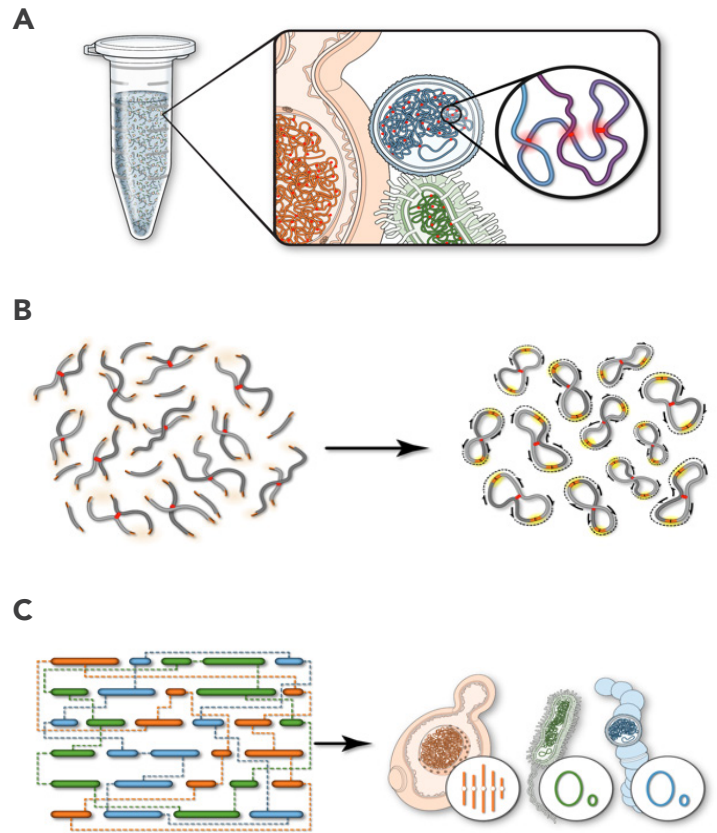


Figure 1. Proximity ligation-based metagenome deconvolution workflow. **A.** *In vivo* crosslinking traps both interchromosomal and intrachromosomal DNA contacts (red foci) within intact cells. **B.** Crosslinked DNA is fragmented and proximity-ligated, creating chimeric junctions between sequences originating from the same cell. These chimeric junctions are recovered, converted into a sequencing library, and subjected to paired-end sequencing on the Illumina® platform. **C.** Hi-C links are used to deconvolute contigs into their original cellular groupings, including chromosomes (orange) and mobile genetic elements (green and blue).

As shown in Figure 2, ProxiMeta produces more higher-quality genomes than either binning or other proximity-based approaches alone, particularly for short-read assemblies.

Proximity-augmented metagenome-assembled genomes (MAGs) are typically more complete and less contaminated than metagenome assemblies derived from shotgun sequencing approaches that lack linkage information. In addition, MGEs and ARGs can be annotated accurately within the assembly and attributed to their host genomes (see Applications section on p. 4).

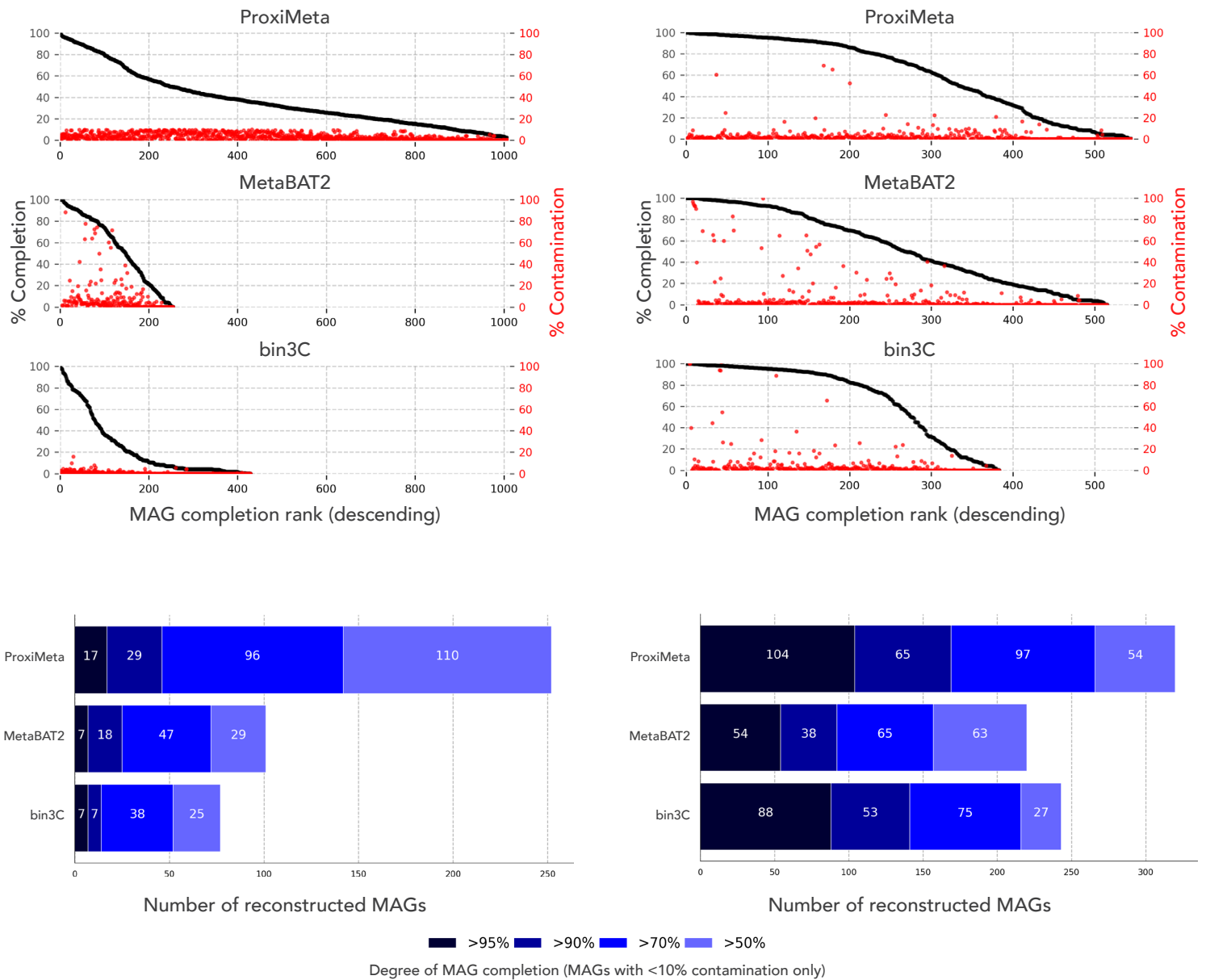


Figure 2. The ProxiMeta computational tool improves metagenome deconvolution and assembly from both short- and long-read sequencing data. Metagenome-assembled genomes (MAGs) were recovered from a short-read shotgun assembly of a complex wastewater sample (left), and a long-read assembly of an animal fecal sample (right), using three different binning methods (ProxiMeta, MetaBAT2 or bin3C⁷). The ProxiMeta tool was used to combine the linkage information provided by Hi-C with traditional binning to produce more higher-quality MAGs than using either a conventional binning tool (MetaBAT2) or a Hi-C based method (bin3C) alone. In the top half of the figure, each point along the x-axis represents a MAG, ranked in order of decreasing completeness (using CheckM⁸). The level of contamination of each genome is indicated with its corresponding red dot. In the bottom half of the figure, only those MAGs with <10% contamination are shown. These are grouped by degree of completeness (indicated by blue bars). ProxiMeta significantly improves results, particularly for short-read assemblies.

Visit proximeta.phasegenomics.com to view example reports or perform your analysis

Applications

The long-range information obtained with Hi-C takes the guesswork out of metagenomic deconvolution and provides a much more complete picture of complex microbial communities. Whether you are working with environmentally or biologically sourced samples, the ProxiMeta Platform uniquely enables the discovery and characterization of more genomes, more mobile genetic elements, and more gene functions—as well as the relationships between them.

Host attribution of mobile elements enables effective surveillance and control of antibiotic resistance

Horizontal gene transfer of 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 comprised of MGEs,⁹ 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 can be 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.

As shown in Figure 3, host attribution of MGEs is straightforward with the ProxiMeta Platform—as 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.

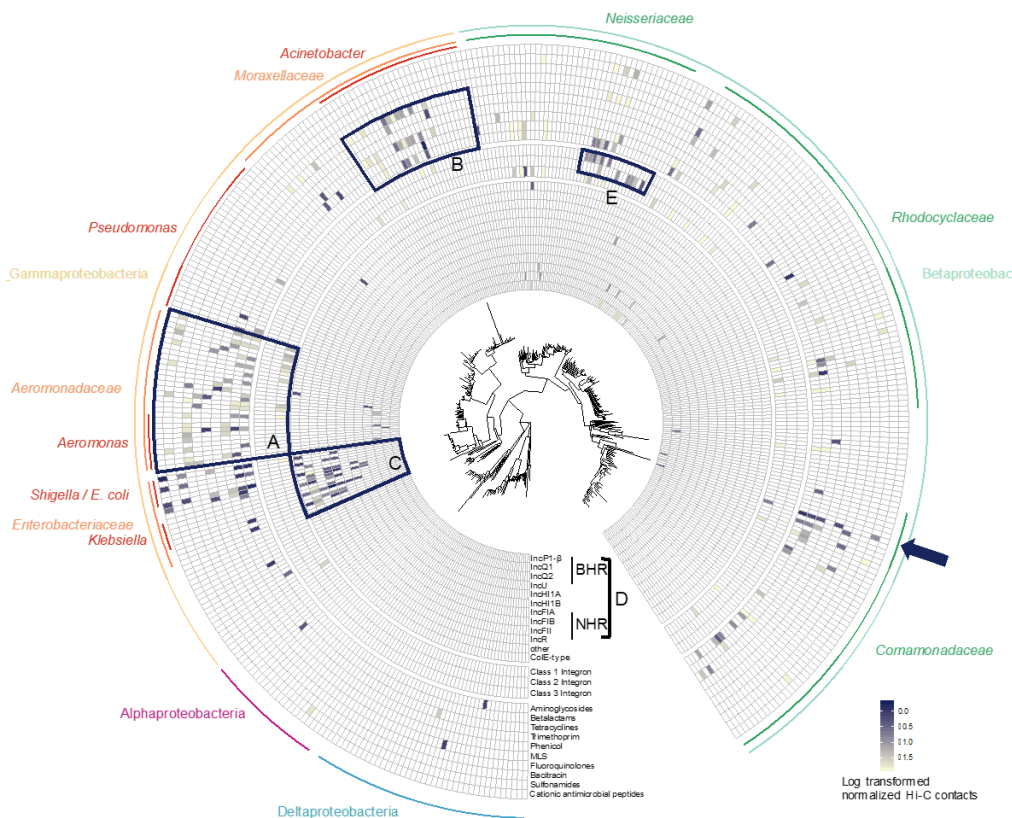


Figure 3. Linking the plasmidome and resistome to the microbiome. Physical intracellular linkage information obtained from ProxiMeta analysis was used to identify the *in situ* host range of MGEs among Proteobacteria in a wastewater sample. Each tip of the circular phylogenetic tree in the center represents 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.

Reproduced without change from Stadler T, et al. *ISME J* 2019; 13: 2437–2446. doi: 10.1038/s41396-019-0446-4, under the provisions outlined at <http://creativecommons.org/licenses/by/4.0/>.

Discover more species, strains, genes, and gene functions

Earth may be home to as many as one trillion (10^{12}) microbial species, of which 99.999% remain to be discovered.¹⁰ 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.

The ProxiMeta Platform uniquely enables discovery and bioprospecting, as it does not require culturing, nor relies on any *a priori* data. With the ability to assemble high-quality genomes for microbes with as little as 0.05% cellular abundance¹¹ in a sample, it is easy to discover more novel organisms, strains, and genes; and to predict metabolic functions more accurately (Figures 4 and 5).

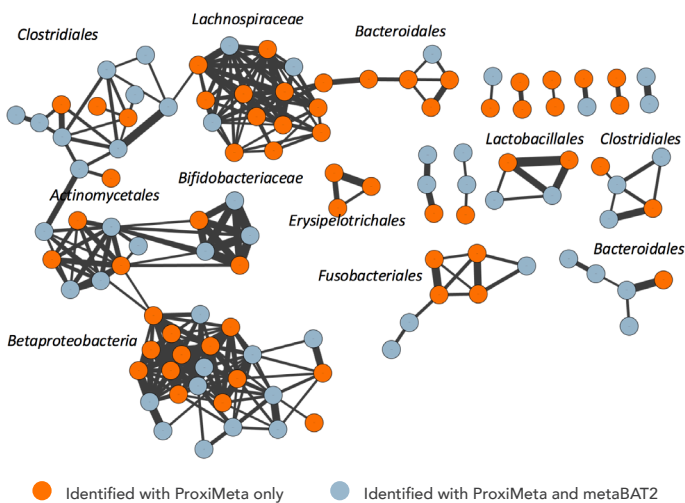


Figure 4. Discover more. Data from a complex wastewater sample was analyzed using ProxiMeta, or by conventional binning with MetaBAT2. Resulting genomes reported to be >80% complete were mapped against one another to measure their similarity. Orange nodes (circles) represent genomes that were found only 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 ProxiMeta to deconvolve both distant and closely related genomes from a single, complex sample.

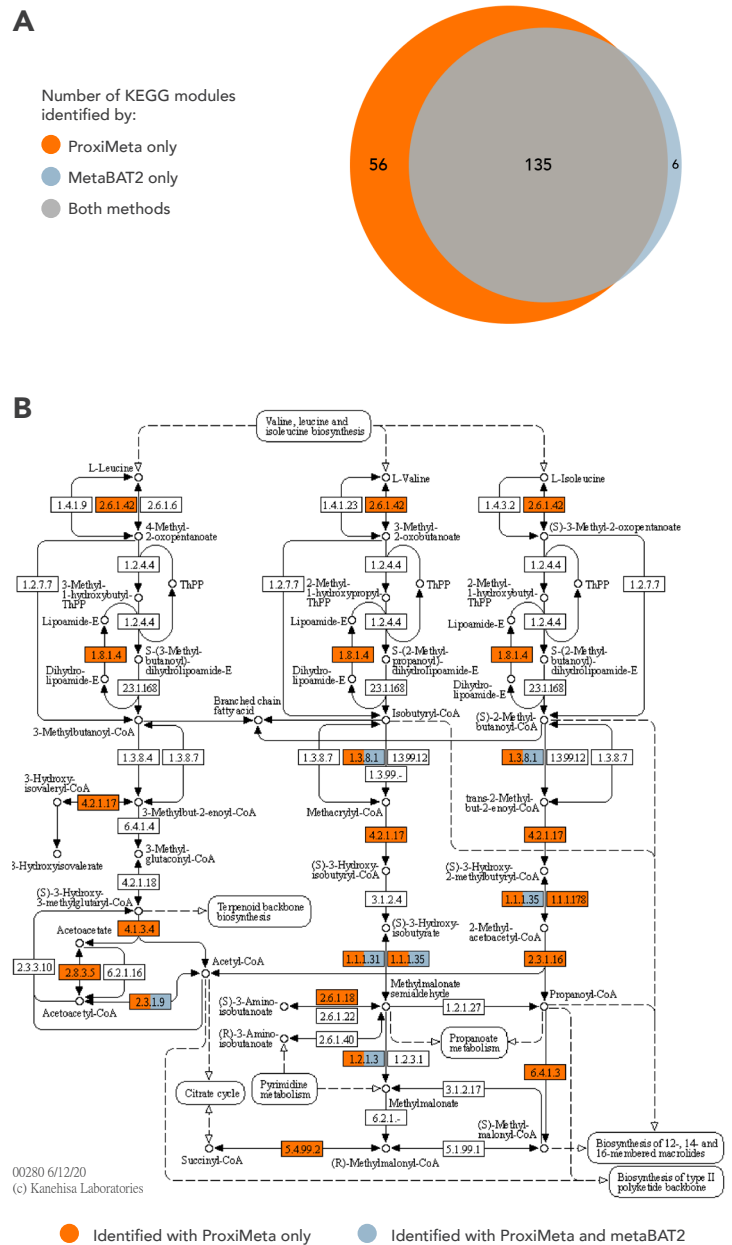


Figure 5. More high-quality MAGs offer more metabolic insights. **A.** MAGs assembled from the wastewater sample referenced in Figure 4 were used to perform a KEGG pathway analysis.¹² Within the set of MAGs assembled with ProxiMeta, 191 KEGG modules were identified, as compared to 141 for the set of MAGs assembled with MetaBAT2. Of these modules, 135 overlapped between the two analyses. ProxiMeta identified almost 10-fold more gene modules that can be linked to specific metabolic capacities and other phenotypic features.

B. The KEGG pathway for the degradation of the amino acids valine, leucine and isoleucine in the Burkholderiales (an order of Proteobacteria). Nodes highlighted in orange designate modules identified exclusively in ProxiMeta MAGs, whereas blue nodes represent modules identified in both the ProxiMeta and MetaBAT2 MAGs. In this pathway, 70% of the nodes were not identified using the conventional binning approach.

Summary

The ProxiMeta Metagenome Deconvolution Platform combines proximity ligation data with shotgun sequencing data, to enable the assembly of high-quality metagenomes and accurate host attribution of mobile genetic elements. This allows for more complete insights and new discoveries from metagenomic and microbial samples.

Key features and benefits of the platform are summarized below:

- ProxiMeta Library Prep Kits provide a streamlined Hi-C protocol that does not require culturing of microbes, nor extraction of high-molecular weight (HMW) DNA.
- Cost-effective, short-read Hi-C sequencing data may be combined with shotgun assemblies obtained from short- or long-read sequencing.
- Cloud-based ProxiMeta analysis recovers more known and novel metagenomes that are more complete and less contaminated—even from low-abundance and unculturable species strains.
- Direct and quantitative measurements between sequences that originated from the same cell allow for accurate attribution of MGEs to their hosts.
- Available as a complete, easy-to-use sample-to-analysis solution, or full service.

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. <http://phasegenomics.com/applications/plant-animal-genomics/>
4. <http://phasegenomics.com/applications/human-genomics-epigenomics/>
5. 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.
6. 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.
7. 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.
8. Parks DH, et al. 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.
9. 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.
10. 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.
11. 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.
12. Kanehisa M and Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucl Acids Res* 2000; 28(1): 27–30. doi:10.1093/nar/28.1.27.



Learn more about the ProxiMeta Platform at phasegenomics.com/products/proximeta/



info@phasegenomics.com
www.phasegenomics.com

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

Unless otherwise stated, data on file.

For Research Use Only. Not for use in diagnostic procedures.

PROXIMETA is a trademark of Phase Genomics, Inc. All other product names and trademarks are the property of their respective owners.

© 2020 Phase Genomics, Inc. All rights reserved.

AN001-081320