Targeted Genomics for Pathogen and Microbial Communities

## COMPREHENSIVE GENE OR GENOME SEQUENCING VIA HYBRIDIZATION CAPTURE

Next-generation sequencing (NGS) is a powerful method for rapid whole-genome sequencing of microbial genomes or overall microbial species content in complex samples. However, most biological samples with bacteria or viral species of interest are dominated by non-microbial DNA. This necessitates extremely deep sequencing in order to accurately resolve the genomes or genes of interest, or to fully characterize the variation within microbial communities. Targeted sequencing, wherein the abundance of genetic regions of interest are increased in the sample prior to sequencing, drastically reduces the overall costs of sequencing and data analysis per sample. Hybridization capture is currently the most versatile technique currently for comprehensive, cost-effective genome or gene family sequencing of both viruses and bacteria in complex samples. With included custom probe design and high-sensitivity, high-performance reagents and protocols, myBaits® Custom hybridization capture kits from Daicel Arbor Biosciences provide the best value on the market for any microbial DNA and/or RNA sequencing project.

WHAT IS HYBRIDIZATION CAPTURE?

If a microbial species of interest can be grown into pure culture, or otherwise fully physically isolated from host or environmental background cells, then its genomic DNA can typically be directly sequenced without any background DNA removal beforehand. However, in many experimental contexts, this approach is either impossible due to biological growth constraints, or infeasible because the research questions require metagenomic or microbiomic data contexts. In such cases, traditional NGS approaches have been to either perform total sample shotgun sequencing – wherein only a small percentage of reads may come from the microbial genome(s) of interest – or perform targeted techniques to increase the percentage of microbial reads prior to NGS. In both scenarios, hybridization capture offers significant cost and bioinformatic advantages compared to both of these approaches, with minimal expected loss of target sequence diversity.

Hybridization capture works by leveraging the flexible power of complex long oligo pools to target molecules of interest within an NGS library (see Fig. 1). In brief, NGS libraries (for DNA- or RNA-seq) are denatured, and allowed to hybridize to target-specific synthetic biotinylated RNA probes. Then the probe:library complexes are bound to streptavidin-coated magnetic beads, and washed to remove non-specifically bound library molecules. These "enriched" libraries are then amplified, and are ready for sequencing on the appropriate NGS platform.
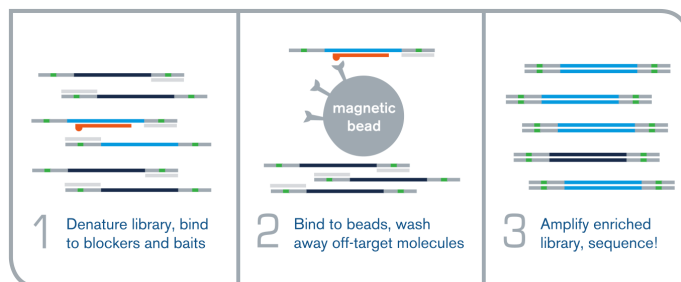
Figure 1. In-solution hybridization capture, or "hyb cap". *Biotinylated oligo probes hybridize to NGS library molecules built from all DNA in a sample, and then the hybrids are sequestered on streptavidin-coated magnetic beads prior to NGS.*

CASE STUDIES

The numerous benefits of hyb capture are now routinely exploited for microbial genome sequencing and strain identification.

Viral whole genome sequencing from total tissue DNA

Infectious virus strain identification often benefits from full genome sequencing, especially when that strain is novel or exhibits previously unobserved characteristics. However, the vast majority of DNA in host or cell culture samples are from non-viral sources. In Forth et al. 2019, genomes of the large dsDNA virus African swine fever virus (ASFV), up to almost 191 Kbp, were effectively retrieved using hyb capture from both cultured cells and porcine spleen tissue. This increased the percentage of viral DNA by over 100 fold (Fig. 2) compared to shotgun sequencing, allowing for the generation of high coverage genomes and various analyses including investigations into viral variants with a fraction of the sequencing effort.
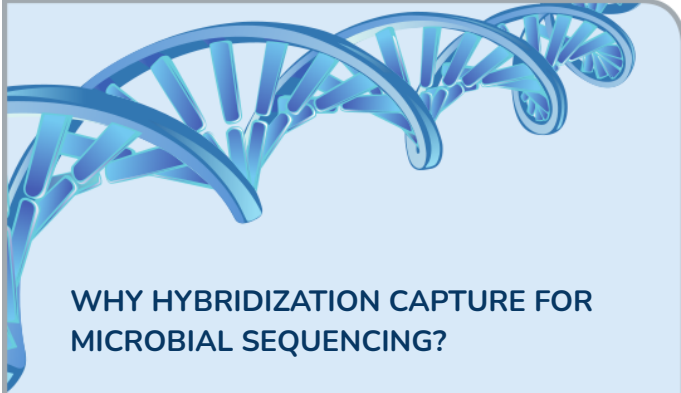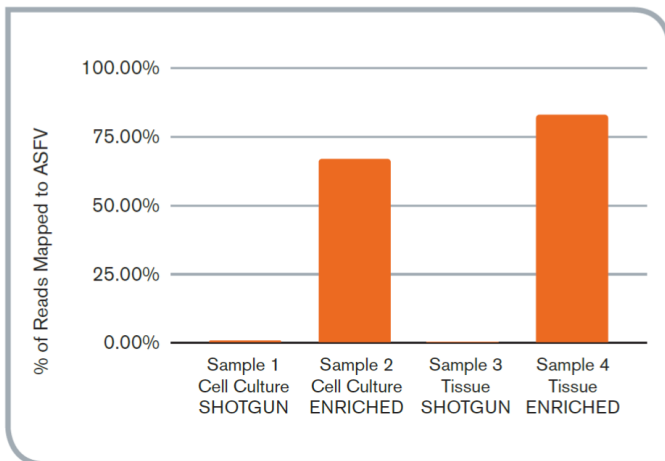
Figure 2. High percentage of ASFV observed in enriched samples (cell culture and tissue). *Based on data from Forth et al. (2019).*

Bacterial genome sequencing from environmental DNA

Pathogenic bacteria typically exist at low abundance in natural environments, but represent a reservoir from which epidemics can arise. Detecting and characterizing ultra-low-abundance strains can serve as a critical tool in outbreak prevention and treatment. Vezzulli et al. (2017) demonstrate that by using hyb capture, extremely low-frequency *Vibrio cholerae* genomes, including specific virulence factors, could be retrieved from raw river water despite being embedded in an extremely diverse background of other DNA sources (Fig. 3). Since these bacteria were so low abundance, they would have been otherwise impossible to study with tissue culture or more traditional molecular techniques.

WHY HYBRIDIZATION CAPTURE FOR MICROBIAL SEQUENCING?

- Able to efficiently recover even highly divergent molecules

- Accommodates any NGS library, whether built from DNA or RNA (cDNA)

- Reconstruct any type of mutation: SNP, indel, or rearrangement

- Platform agnostic: use same kit for both short- and long-read sequencing
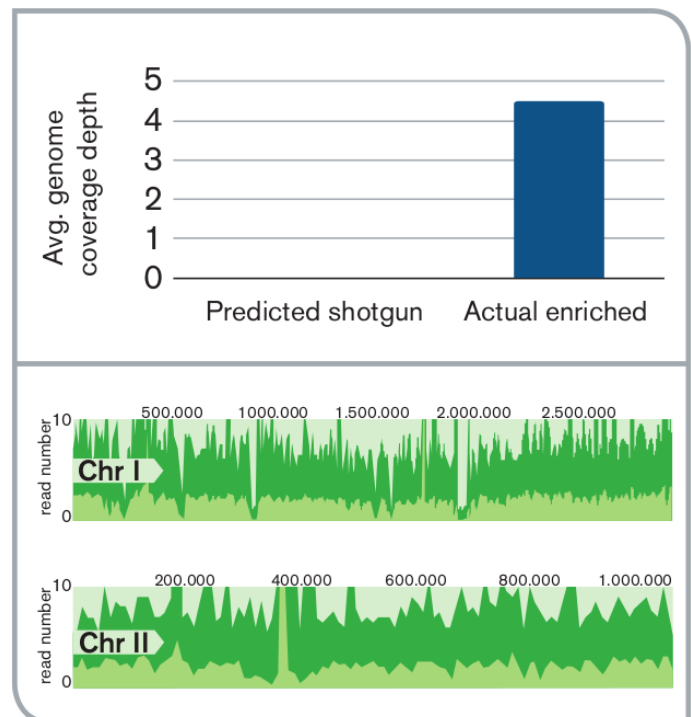
Figure 3. Effective genome-wide enrichment of *V. cholerae* DNA from river water. *Vezzulli et al. (2017) estimate that their enrichment approach was ~2500 times more effective than shotgun sequencing alone (0.0018X avg genome coverage with shotgun vs 4.5X enriched). Bottom figure redrawn from Vezzulli et al. (2017), Fig. 2, pg. 736.*

Comprehensive 16S rRNA metagenomic profiling

The 16S rRNA locus is a well-established target used for bacterial community profiling research. Sequencing its multiple variable regions can provide broad species-level resolution in a metagenomic context. However 'universal' 16S amplicon methods have known biases, and 16S or total shotgun metagenomic deep sequencing can be costly and require longer turnarounds for sequencing as well as analysis. Beaudry et al. (2021) demonstrate that a 16S hybridization capture panel can offer a key "middle ground" between these two different methods (Fig. 4), by enabling similar profiling power as gold-standard shotgun with significantly less sequencing expenditure and time commitment.
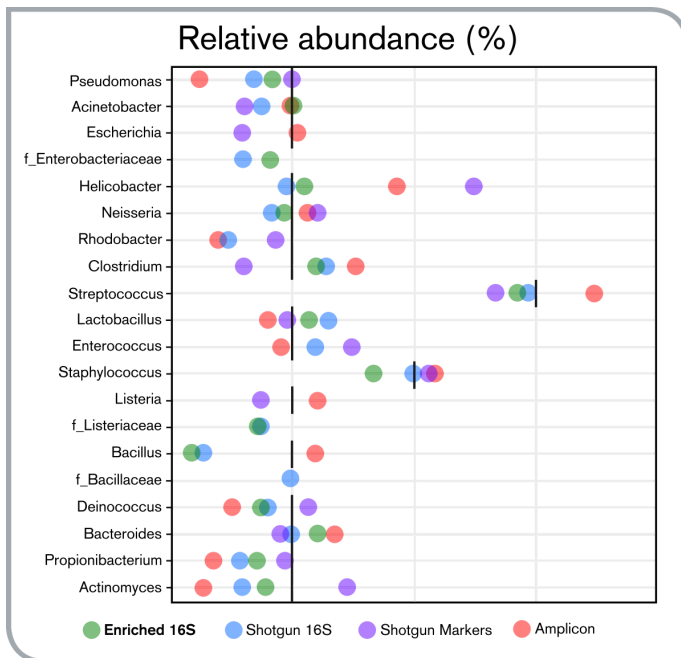
Figure 4. 16S hybridization capture provides similar data compared to shotgun approaches, at a fraction of the sequencing depth required. *Vertical lines show intended genera percentages in mock community samples. Based on Beaudry et al. (2021), Fig 3.*

Antimicrobial resistance gene sequencing

The development of antimicrobial resistance by pathogenic bacteria is considered one of the world's most pressing global public health threats. Guitor et al. (2019) designed a hyb capture panel targeting the antimicrobial resistance gene (ARGs) "resistome" and evaluated its performance on mock community samples and biological samples. They demonstrated that myBaits was an effective tool not only enabling comprehensive ARG profiling but also detecting novel content at considerably lower read depths compared to total metagenomic shotgun sequencing (Fig. 5).
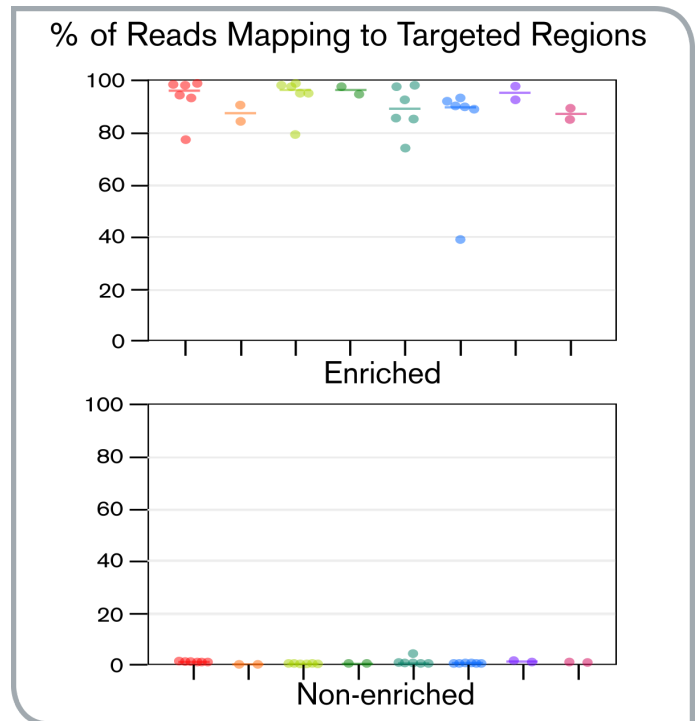
Figure 5. Hybridization capture provides orders of magnitude higher target specificity vs. shotgun. *For control experiments with mock bacterial community samples of 4 or 8 different genomic DNAs, results are shown for percentage of sequencing reads mapping to the 8 different intended target regions both with (left) and without (right) myBaits enrichment with a comprehensive "resistome" ARG baitset. Based on Guitor et al. (2019) Fig 3A.*

myBaits® Hybridization Capture Kits

- **Adaptable** - We design custom probes for your project needs
- **Seamless** - Fits in between NGS library prep and sequencing
- **Flexible** – Perfect for re-sequencing as well as variant discovery
- **Complete** - myBaits kits contain all reagents for hyb capture protocol
- **Easy to use** - No special equipment or training necessary
- **Platform agnostic** - Can use with any NGS platform

Ancient viral genomic DNA sequencing

In addition to the ability of hyb capture to enrich extremely low-abundance microbial DNA from complex samples, another key feature is its capacity for capturing nucleic acids that are damaged and degraded, increasing their relative divergence from the probe sequences. For example, in archaeological or paleontological samples, viral DNA is not only swamped by environmental and host DNA, but it is also heavily damaged by taphonomic processes. Duggan et al. (2016) were able to retrieve even the most trace, most fragmented remaining genomic fragments from a nearly 400 year-old mummy, and fully reconstruct the genome sequence of smallpox – variola virus (VARV) – an endeavor that would have otherwise cost orders of magnitude more in sequencing alone (Fig. 6).
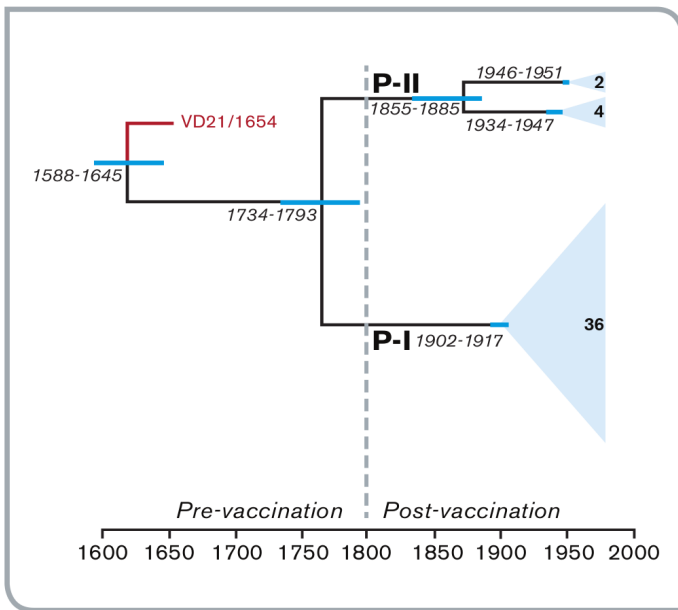
Figure 6. VARV evolutionary timescale tree. *16th century strain in red. Redrawn from Duggan et al. 2016, Fig 3, pg. 3410.*

Simultaneous host and pathogen sequencing

Novel disease reservoirs and transmission routes are often unknown. While ectoparasites like biting arthropods can be easily identified as a disease's potential vector, identifying the spectrum of hosts of that parasite, including the one that might serve as an interim pathogen reservoir, can be much more difficult. As demonstrated by Campana et al. (2016), hyb capture can serve as a quick and inexpensive means of identifying both a parasite's most recent bloodmeal, as well as spectrum of pathogens in that parasite (Fig. 7). This highlights hyb capture's versatility in multi-species profiling in complex DNA samples that may be dominated by host DNA.
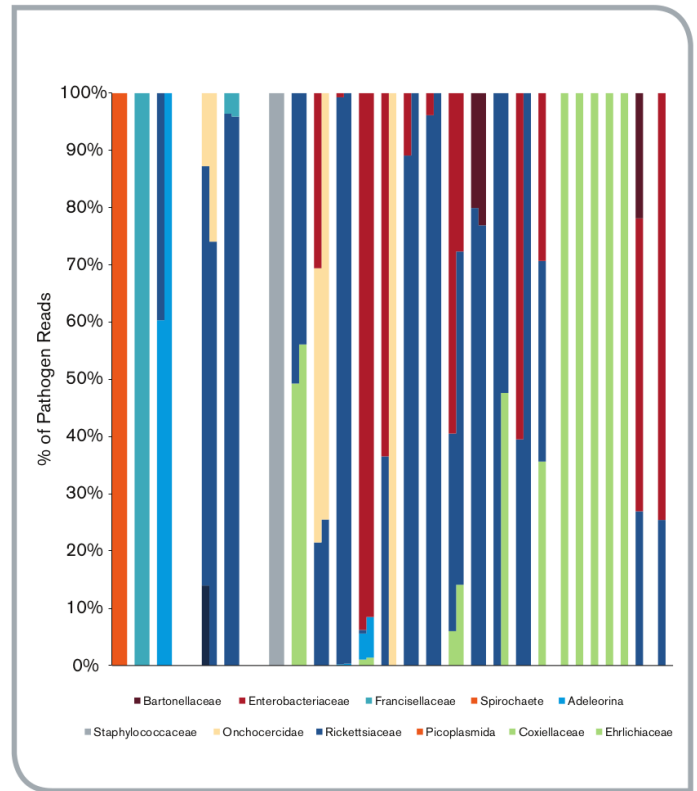
Figure 7. **Abundance of co-enriched pathogen and macroparasite taxa identified in tick and flea samples.** *Redrawn from Campana et al. 2016, Fig. 3, pg 11.*

WHY HYBRIDIZATION CAPTURE

The hybridization capture system has many advantages compared to alternative targeted sequencing approaches such as multiplexed amplicons (Fig. 8). Hyb capture probes function even when there is significant probe-target sequence mismatch, up to 30+% divergence. Since the hyb capture probes do not themselves become directly integrated into the sequenced library molecules, this offers significant experimental flexibility. Since only one probe minimally must hybridize to enrich a molecule, and no probe-probe interaction is required, this minimizes the need to optimize most experiments. Library molecules of any length, even up to several Kb, can be enriched and sequenced with long-read platforms such as PacBio® and Oxford Nanopore®.

These advantages mean that any type of novel sequence feature in or adjacent to the targeted regions – e.g. SNVs, short indels, or rearrangements – can typically be reconstructed in the downstream bioinformatic data analysis.
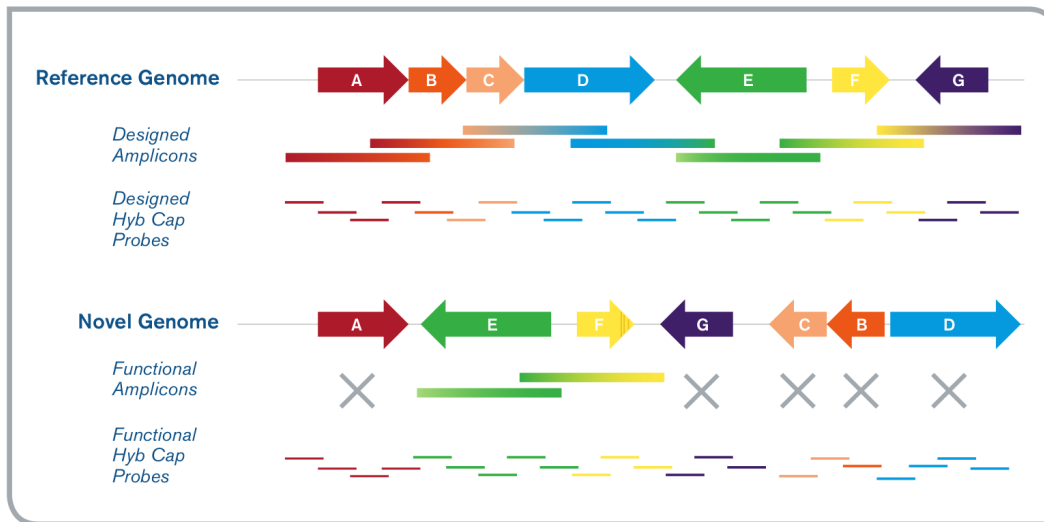
Figure 8. Hybridization capture tolerates both rearrangements and sequence variation. *Both hyb cap and amplicon sequencing can effectively enrich for known sequence regions. However, only hyb cap can retrieve target sequences that have significant rearrangements and/or mutations relative to the reference used for probe design, such as when capturing genomic content from novel viral strain genomes.*

CONCLUSIONS

When embedded in complex host, environmental, or metagenomic DNA backgrounds, microbial genomes can be prohibitively expensive or even impossible to characterize with direct high-throughput sequencing or traditional molecular techniques. Inexpensive, versatile, and platform-agnostic hybridization capture methods, like those afforded by the myBaits system from Daicel Arbor Biosciences, reduce microbial genome sequencing costs by orders of magnitude in most circumstances, and can reliably illuminate novel variations in strain sequence, gene content, and genome structure. The small size of microbial genomes means that their capture probe sets are typically even smaller than a standard mammalian exome or SNP genotyping panel, a feature which confers enhanced specificity and sensitivity in capture experiments. Whether for strain identification, virulence detection, transmission history tracking or illuminating ancient evolutionary origins, hybridization capture offers logistical and budgetary benefits to microbial genome sequencing unmatched by other technologies.

REFERENCES

Beaudry, M.S. et al. (2021). **Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment**. Frontiers in Microbiology.

Campana, M.G. et al. (2016) **Simultaneous identification of host, ectoparasite and pathogen DNA via in-solution capture.** Molecular Ecology Resources.

Duggan, A.T. et al. (2016) **17th Century Variola Virus Reveals the Recent History of Smallpox**. Current Biology.

Forth, J.H. et al. (2019) **A Deep-Sequencing Workflow for the Fast and Efficient Generation of High-Quality African Swine Fever Virus Whole-Genome Sequences**. Viruses.

Guitor, A.K. et al. (2019) **Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes**. Antimicrobial Agents and Chemotherapy.

Vezzulli, L. et al. (2017) **Whole-Genome Enrichment Provides Deep Insights into Vibrio cholerae Metagenome from an African River**. Microbial Ecology.