VirION2: a short- and long-read sequencing and informatics workflow to study the genomic diversity of viruses in nature

BACKGROUND

Microbes play fundamental roles in shaping natural ecosystem properties and functions, but do so under constraints imposed by their viral predators.

In the world's oceans, viruses modulate microbial gene flow and are integral to global oceanic nutrient cycles.

However, current sequencing methods to study them can miss important viruses and genome features.

CURRENT SEQUENCING BARRIERS

short-read
- Microdiverse populations and fine-scale genomic traits can be missed

long-read
- High DNA requirements
- Sequencing errors limit accurate gene prediction

We introduce VirION2, an update to our initial method that uses long-read sequencing to better analyze viromes.

METHODS

Using a viral mock community made up of three *Pseudoalteromonas* phage isolates, we first **optimized laboratory protocols to enable 76% longer reads** from 100-fold less input DNA (now 1 ng).
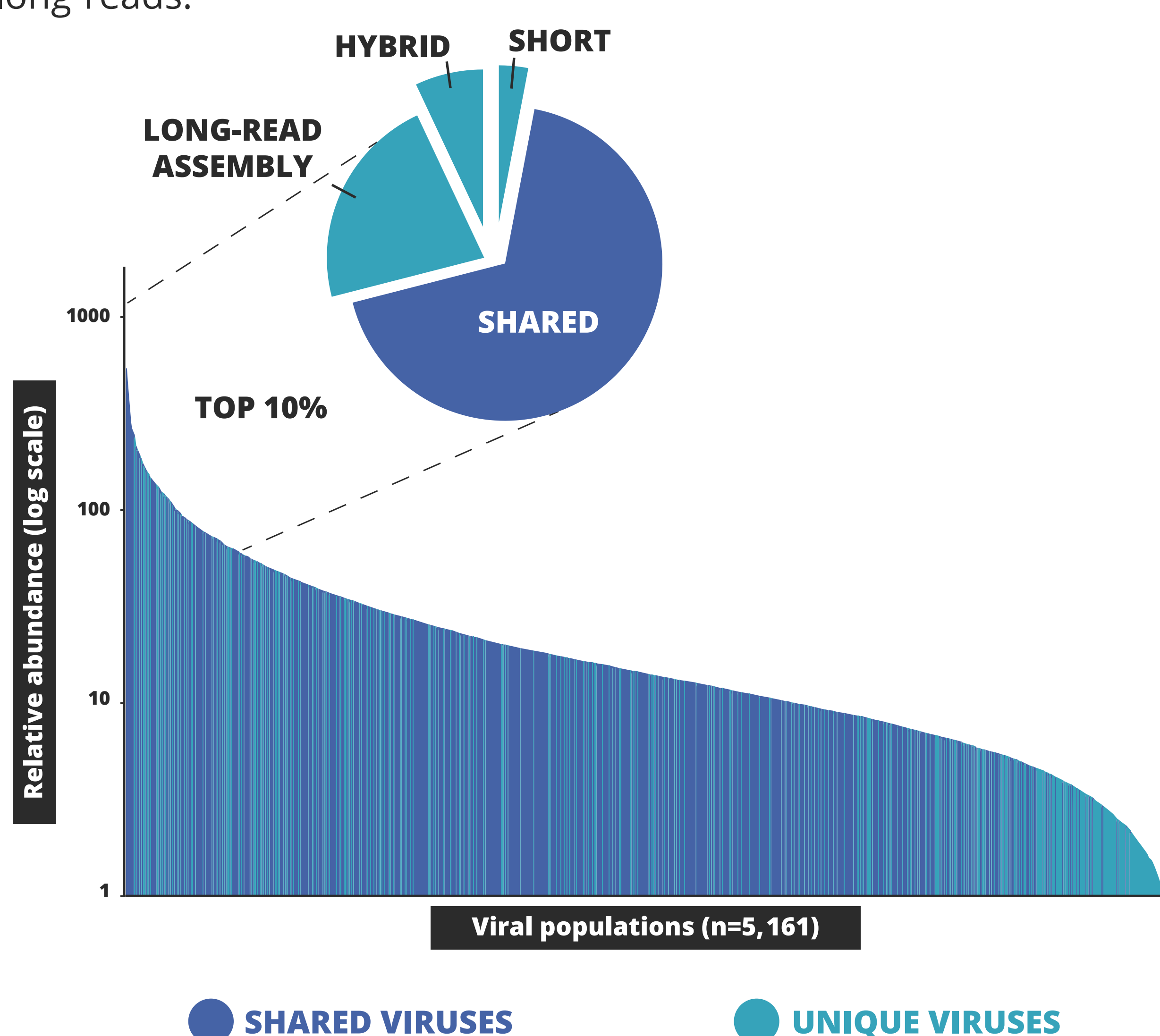
From a natural seawater sample, **we compared viromes generated with VirION2 against three other library preparation options.** Downstream informatics were optimized for improved long-read error correction and assembly.

1 UNAMPLIFIED LONG-READ

2 ORIGINAL VirION

3 SHORT-READ

RESULTS

VirION2 assemblies combined with short-read data provided significant improvements over VirION libraries in the recovery of longer and more complete viral genomes. Our optimized error-correction strategy using long- and short-read data **achieved 99.97% accuracy.**

VirION2 assemblies captured 5,161 viral populations, 30% of which were uniquely assembled through inclusion of long-reads; and 22% of the top 10% most abundant virus populations were derived from the assembly of long-reads.

HYBRID
SHORT
LONG-READ ASSEMBLY
SHARED
TOP 10%

Relative abundance (log scale)
1000
100
10
1

Viral populations (n=5,161)

● SHARED VIRUSES   ● UNIQUE VIRUSES

On average, viral populations unique to VirION2 assemblies had **significantly higher microdiversity,** which may explain why short-read virome approaches failed to capture them.

CONCLUSION

Long-read sequencing, combined with short-reads, can further increase estimates of viral diversity and capture ecologically important taxa.

Though studying viruses in nature can be challenging due to low biomass and the lack of universal gene markers, **VirION2's sample prep and workflow can help researchers better investigate the virosphere.**