**illumına®**

# Using a PhiX Control for HiSeq® Sequencing Runs

A low-concentration spike-in of Illumina PhiX Control v3 provides quality and calibration controls.

---

**Highlights**

- For most libraries, Illumina recommends using PhiX Control v3 (catalog # FC-110-3001) in a low-concentration spike-in (1%) to monitor sequencing quality control.

- For unbalanced samples such as Amplicon libraries, use a higher concentration PhiX spike-in to improve cross-talk and phasing calculations.

- For samples with low diversity, use a high-concentration spike-in (5% or higher) of PhiX to create a more diverse set of clusters.

---

## What are PhiX Control Libraries?

Control libraries generated from the PhiX virus serve as an effective control in sequencing runs. Characteristics of the PhiX genome provide several benefits:

- **Small**—PhiX is a small genome, which enables quick alignment and estimation of error rates.
- **Diverse**—The PhiX genome contains approximately 45% GC and 55% AT.
- **Well-Defined**—PhiX has a well-defined genome sequence.

Illumina cluster detection algorithms are optimized around a balanced representation of A, T, G, and C nucleotides. Illumina PhiX Control v3 is a balanced and diverse library that can help mitigate sequencing challenges in unbalanced and low diversity libraries.

The mean insert size of the PhiX v3 library is approximately 375 bp, corresponding to approximately a 500 bp library size if visualized on a bioanalyzer.

## Benefits of Using a PhiX Control

PhiX libraries provide a quality control for cluster generation, sequencing, and alignment, and a calibration control for cross-talk matrix generation, phasing, and prephasing.

### Cluster Generation

Illumina PhiX Control v3 is shipped as a ready-to-use 10 nM library that can be used as a positive control in the clustering process. If a problem occurs in sample preparation, PhiX still generates clusters. These clusters help to discern whether a lack of clusters is due to sample preparation failure or a failure in the cluster generation process.

### Cross-Talk Matrix Generation

During an Illumina sequencing run, the cross-talk due to spectral overlap between the 4 fluorescently labeled nucleotides is calculated during template generation (cluster identification) in cycles 1–5.

For proper cross-talk calculation, HiSeq Control Software v1.3.8 and later requires approximately equal numbers and at least 50,000 clusters/mm² in each of the 4 bases. Therefore, it is imperative to have a balanced representation of bases at the beginning of each read, excluding the Index Read.

### Phasing and Prephasing

During sequencing by synthesis, each DNA strand in a cluster extends by 1 base per cycle. A small proportion of strands may become out of phase with the current cycle, either falling a base behind (phasing) or jumping a base ahead (prephasing). The phasing and prephasing rates define the fraction of molecules that become phased or prephased per cycle.

Calculation of these rates requires a balanced and random base composition in cycles 2–12. For best results, use a PhiX spike-in as a control with any library that does not comprise a balanced base composition (eg, initial cycle indexing, restriction enzyme libraries).

High GC samples typically show higher phasing rates. However, if the sample has good diversity (for example, whole-genome sequencing libraries with 40–60% GC), a PhiX control is not required.

### Alignment

Because PhiX has a small, well-defined genome sequence, it is an excellent alignment control. If a lane containing PhiX is designated for your run, Real-Time Analysis software aligns complete sequences to the PhiX reference beginning after cycle 25, which provides an indication of sequencing success during the run.

## Using a PhiX Control for Unbalanced Samples

Most mammalian genomic or whole-transcriptome RNA samples have a balanced genomic composition (approximately equal proportions of A, T, G, and C). These samples do not require a PhiX spike-in to generate accurate matrix and phasing estimations. Unbalanced samples contain genomes with high AT or GC content (less than 40% or greater than 60%). Some examples are Arabidopsis, Plasmodium, some bacteria, and bisulfite conversion studies.

Use a PhiX spike-in with unbalanced samples to improve cross-talk and phasing calculations.

## Using a PhiX Control for Low Diversity Libraries

Low diversity libraries are libraries where a significant number of the reads have the same sequence. The base composition shifts because the reads are no longer random.

Low diversity can occur with some expression studies with greater than 25% of 1 type of transcript, amplicon pools, adapter dimer, and initial cycle indexing, for example. Low diversity libraries might require a high concentration spike-in to help balance the overall lack of sequence diversity. For more information, see the *Low-Diversity Sequencing on the Illumina HiSeq Platform* technical note.

### Strategies for Low Diversity Libraries

Use the following strategies when sequencing low diversity libraries on the HiSeq:

- **Reduce Cluster Density**—The optimal cluster density specified for sequencing runs assumes that the lane contains a diverse library. Template generation (cluster identification) algorithms detect clusters across multiple sequencing cycles, which resolves overlapping clusters when their sequence differs between cycles. This advantage is lost when clusters have the same sequence throughout template generation cycles. Reducing the cluster density by 50% when sequencing low diversity libraries reduces the number of overlapping clusters.

- **High-Concentration PhiX Spike-In**—Spike-in a PhiX control for HiSeq X, HiSeq 4000, and HiSeq 3000 at a minimum of 5% of the total library. Spike-in a PhiX control for HiSeq 2500, HiSeq 2000, and HiSeq 1500 at a minimum of 10% of the total library. The percentage must be empirically determined. Additionally, reduce the amount of library used so that the total cluster density does not exceed Illumina recommendations. This method provides the appropriate clusters/mm² of each fluor required for matrix generation. Adding less PhiX (eg 1% spike-in) can determine an alignment error rate, but cannot optimize diversity for matrix generation, phasing, and prephasing calculations.

- **Concatamerize and Fragment Amplicons**—When sequencing PCR amplicons or other samples where diversity is low for the first 12 cycles, concatamerize the amplicons and fragment to create sequencing diversity. Alternatively, increase the probability of an equal representation of A, T, G, and C nucleotides by adding custom primers with a random sequence onto the amplicons on both the 5' and 3' ends. Other options include indexing libraries and then sequencing with other diverse libraries, such as PhiX, or to control the pooling of different amplicons such that the first few bases fit the balancing requirements.

## How to Use a PhiX Control

You can apply a PhiX library to your HiSeq sequencing run in the following 2 ways:

- **Low-concentration spike-in of 1%**—For most libraries, Illumina recommends a low-concentration spike-in as an in-lane positive control for alignment calculations and quantification efficiency.

- **High-concentration spike-in of 5% or higher**—Reduce the cluster density of your library by 40–50%, and then use a high-concentration spike-in to create a more diverse set of clusters for matrix, phasing, and prephasing calculations.

### Specifying PhiX Control Options

You can specify PhiX for lane alignment and error rate calculations in the run setup steps. From the Advanced screen in the HiSeq Control Software, clear the checkbox for lanes that do not contain PhiX.

For more information about setting up a run using the control software, see the user guide for your sequencing instrument:

- *HiSeq X System Guide,* Document # 15050091
- *HiSeq 4000 System Guide,* Document # 15066496
- *HiSeq 3000 System Guide,* Document # 15066493
- *HiSeq 2500 System Guide*, Document # 15035786
- *HiSeq 2000 System Guide*, Document # 15011190
- *HiSeq 1500 System Guide*, Document # 15035788

## Downstream Analysis of PhiX v3 Library Reads

PhiX v3 is not indexed, but it can be used with other indexed libraries. The non-indexed reads are parsed into a folder named Unknown or Undetermined_Indices (depending on the CASAVA version) along with any indexed reads where the index could not be identified. The PhiX library can be further processed using downstream alignment. However, PhiX cannot be used to balance out index read low diversity.

**FOR RESEARCH USE ONLY**