NinthBio

EMPOWERING DNA ASSEMBLY

NinthBio Homology Path – DNA Script SYNTAX Experiment Report

FRIDAY, DEC 29, 2023

INTRODUCTION

The goal of this experiment was to assess the efficacy of the NinthBio *Homology Path* library synthesis design executed on the DNA Script SYNTAX synthesis system. In order to do so, a library of 36 sequences with a single variable region between positions 111 to 121 was created (Figure 1). The goal being to design and synthesize these sequences with minimal cost and maximum accuracy.

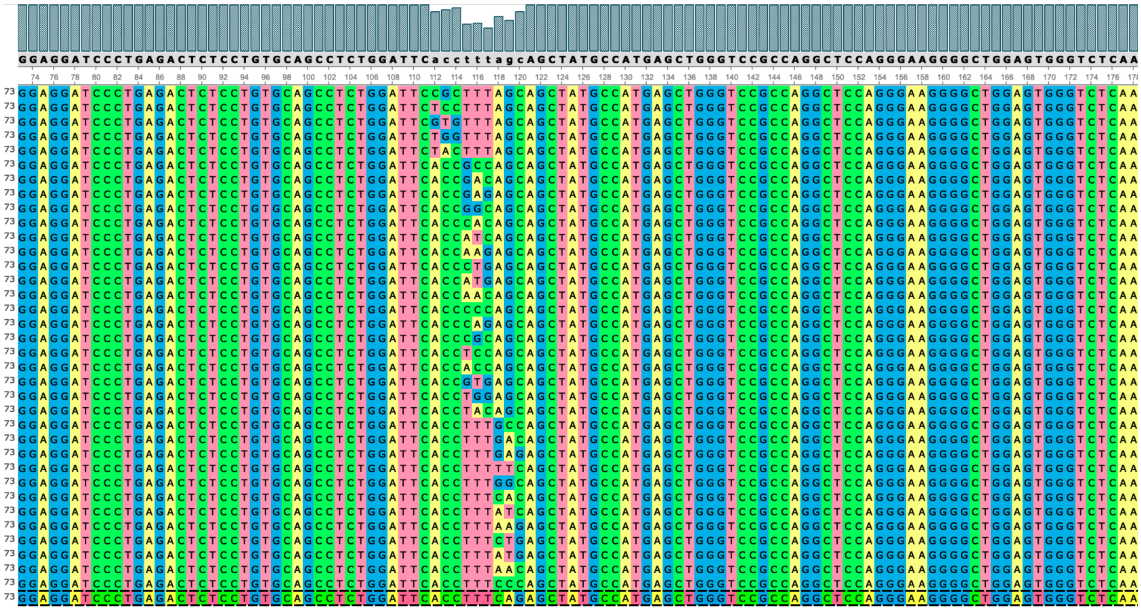

FIGURE 1.

Library of 36 sequences to be synthesized.

HOMOLOGY PATH DESIGN

The first step is to input these sequences into Homology Path Oligo Design (https://homology-path.com/OligoDesign/), with Maximize Recycling chosen as the design type. All other parameters were left as default. The resultant design had the following statistics:

Total sequences	36
Total pre-recycle oligos	792
Total post-recycle oligos	92
Total pre-recycle bases	27216
Total post-recycle bases	3171
Total bases recycled	24045
Recycle efficiency	88.35%

Notice how these 36 sequences can be built with just 92 oligos, when compared to the naïve design of 792 oligos, **an ~88% savings in oligo costs.**

DNA SCRIPT SYNTAX SYNTHESIS

Neochromosome was enlisted to run this design on their SYNTAX system. After the run completed, all sequences were run on a gel to assess basic synthesis success rate (Figure 2). Every sequence shows a band at the expected size.

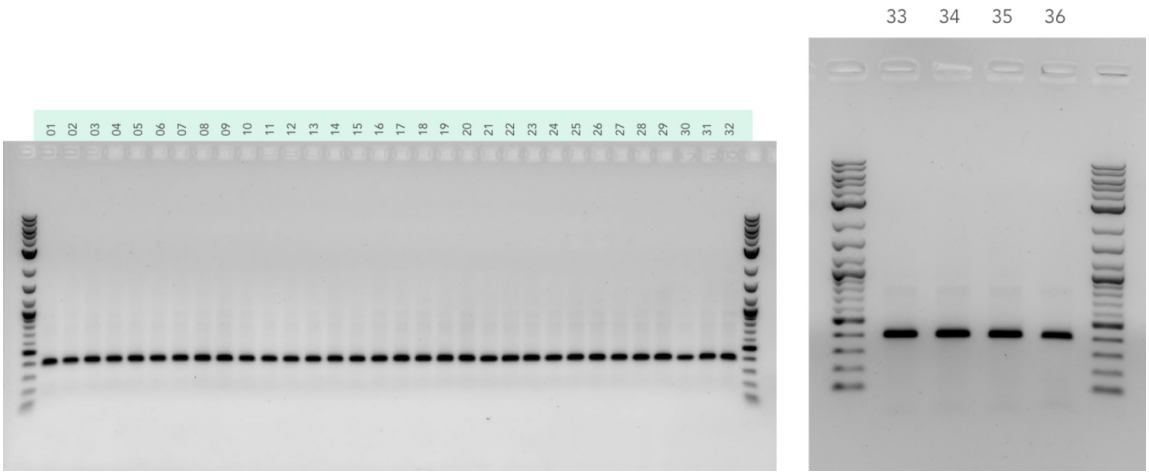

FIGURE 2.

Gel bands for the 36 sequences showing target size achieved.

Next, Neochromosome sequenced each of the 36 samples on an NGS platform resulting in 34 out of 36 (sequences 24 and 33 did not have valid NGS data) successful sequencing FASTQ data files.

NINTHBIO VARIANT ANALYSIS REPORT

The given NGS data from these 34 samples, was run through NinthBio's variant analysis pipeline. Sequencing coverage (Figure 3) for all samples looks consistent and valid, with plenty of coverage to ensure variants are called correctly. Since these sequences are linear products, the coverage of this NGS data falls off dramatically at the tails of the sequence due to the nature of this sequencing mechanism. This is expected and should be considered in the analysis going forward. The reasonable conclusion from this information is that there is valid NGS sequencing data covering all bases for all sequences.

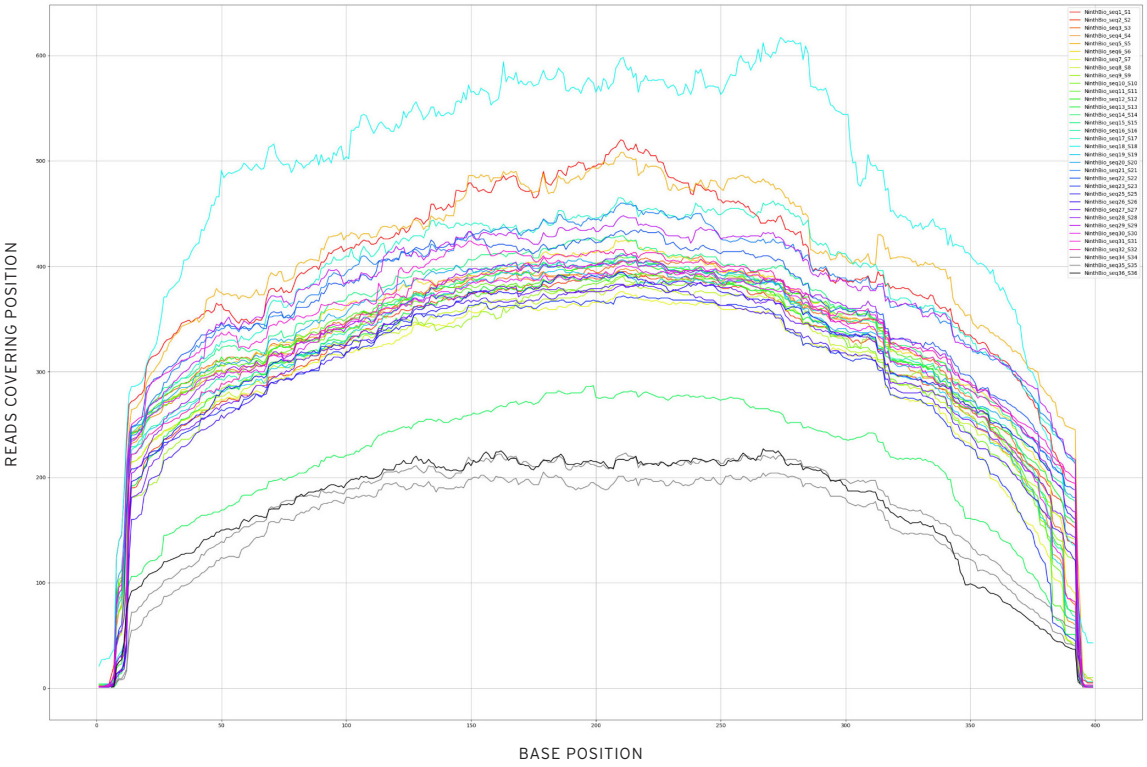

FIGURE 3.

Coverage plots for all 34 samples.

The raw paired end reads were merged into single reads, deduped, and filtered on Phred score >= 30. The variant ratios (Figure 4) were then calculated to assess how accurate these sequences were after synthesis. Ignoring the tail ends due to the coverage issues mentioned earlier, there were only two loci of interest. First, Ninthbio_seq31 (pink outlier in Figure 4) shows about an 18% alternate variable region. Meaning ~18% of the population showed this variant while the rest was matching the reference. Next, several sequences show a ~10% variant at position 360, with the remaining > 90% matching reference. None of these are significant products, and would not affect downstream processing or delivery of target products.

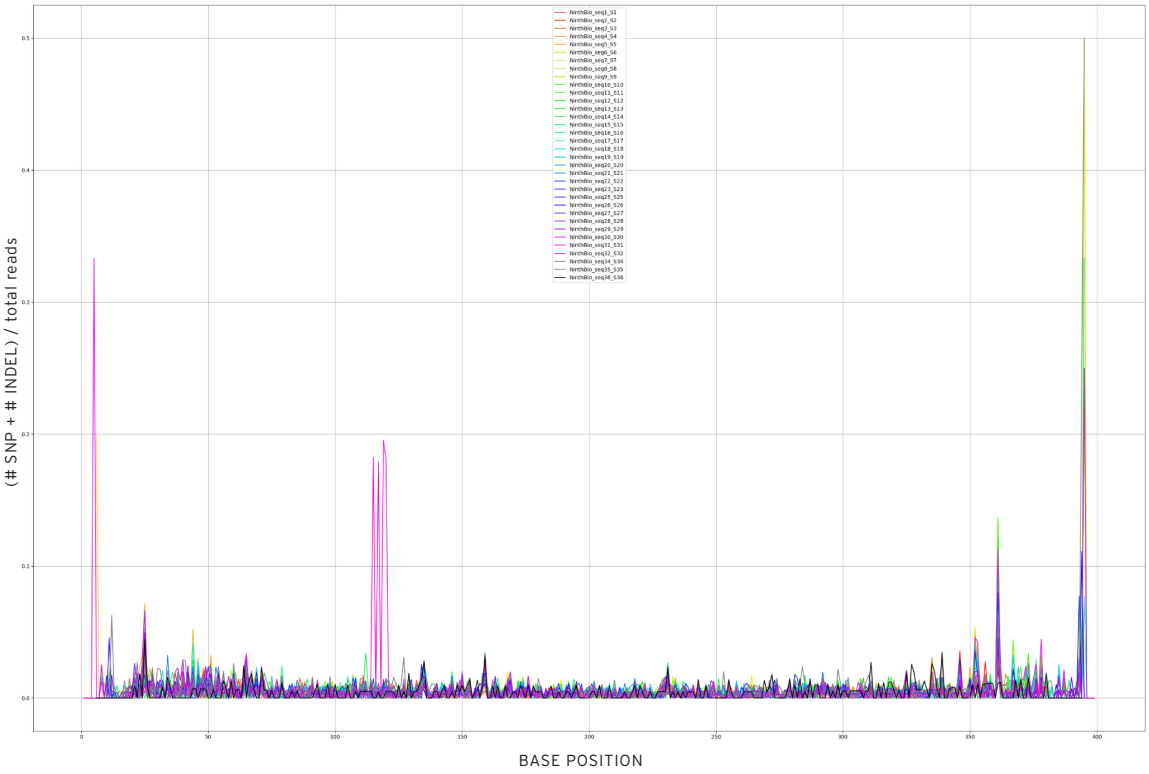

FIGURE 4.

Variant ratios plot showing (# SNP + # INDEL) / Total Reads at each base position.

CONCLUSION

The Homology Path Oligo Design enabled DNA Script SYNTAX to create a high fidelity variant library at approximately 88% cost savings when compared to a naive design. In the worst case example (NinthBio_seq31), if one were to randomly select a sequence from the pool, there would still be an ~82% likelihood of picking the target sequence.

APPENDIX

See the spreadsheet below for the full variant report:
https://docs.google.com/spreadsheets/d/1tHQ-V1NY0tvwQFr9i8bf8NLoQC34hhhuidLGDp03-rs/edit?usp=sharing