Evaluating the impact of error rate on productivity, and cost savings in synthetic DNA fragments

GeneArt High-Q Strings DNA Fragments vs. other suppliers' products

Abstract

Surveys and interviews with researchers revealed that less hands-on time, an easy-to-use method, and a high cloning success rate are important factors to consider to evaluate technologies and vendors for synthetic DNA. To address these needs, we developed Invitrogen™ GeneArt™ High-Q Strings™ DNA Fragments. These DNA fragments are made using our cutting-edge microchip DNA synthesis platform, resulting in superior accuracy for uncloned DNA fragments. To demonstrate the benefits of this technology, we compared these GeneArt fragments to DNA fragments from two different suppliers, and found that the GeneArt technology brings about a 60% reduction in cloning effort with up to 30% cost savings.

Introduction

With innovations in synthesis technology, gene synthesis has become a valuable and affordable way to obtain genetic material. The process of *de novo* gene synthesis involves assembly and PCR amplification of chemically synthesized oligonucleotides. Since this chemistry is not entirely perfect, random sequence errors occur in a small fraction of molecules during oligonucleotide synthesis, which remain through the downstream assembly process. Cloning and sequencing, therefore, is a reasonable way to screen for error-free molecules and discard those with mutations.

After screening, synthetic genes are typically delivered in a cloned and sequenced plasmid to ensure 100% sequence accuracy. However, many scientists prefer to directly obtain linear synthetic DNA fragments to perform cloning and sequencing on their own or for a variety of other applications, including direct assembly and cloning for protein expression, CRISPR-based genome editing, and *in vitro* transcription and translation.

GeneArt High-Q Strings DNA Fragments and alternative DNA fragment synthesis

In a previous white paper from 2015 [1], we had evaluated the performance of Invitrogen™ GeneArt™ Strings DNA Fragments from Thermo Fisher Scientific compared to gBlocks™ Gene Fragments from Integrated DNA Technologies, which is another well-established synthetic DNA product. Briefly, gBlocks Gene Fragments are also uncloned, double-stranded, sequence-verified linear DNA fragments. We demonstrated that the GeneArt Strings DNA Fragments have better sequence accuracy than the gBlocks Gene Fragments, resulting in reduced screening effort to obtain a correct clone, leading to time and cost savings.

Here we introduce GeneArt High-Q Strings DNA Fragments, which have been developed to combine the convenience of a PCR product with the potency of custom-made DNA—available with complementary expression optimization at an affordable price. Use of our cutting-edge and innovative microchip technology platform results in an accuracy of <1 random error in 10,000 bp.

In this paper, we compare the performance of DNA fragments from the GeneArt High-Q Strings portfolio and those from two other suppliers. In doing so, we look at the impact of sequence accuracy on productivity and cost.

Materials and methods

DNA fragments

We obtained 15 gene fragments from supplier 1, and 15 gene fragments from supplier 2; both sets covered lengths from 204 bp to 1,806 bp. A total of 18 GeneArt Strings DNA Fragments with lengths from 204 bp to 1,806 bp, and 16 GeneArt High-Q Strings DNA Fragments with lengths from 204 bp to 1,206 bp, were produced in our production facility. The DNA fragments were resuspended in water to a final concentration of 20 ng/µL and directly used for cloning.

Molecular cloning

The resuspended DNA fragments were cloned either into a pUC-based vector using the Invitrogen™ GeneArt™ Seamless Cloning and Assembly Kit, or into an Invitrogen™ pCR™ Blunt II-TOPO™ vector using the Zero Blunt™ TOPO™ PCR Cloning Kit (Cat. No. K280020), following the respective kits' manuals. In both cases, Invitrogen™ One Shot™ TOP10 Chemically Competent *E. coli* cells were transformed using a heat shock method following the supplier's instructions, and plated and grown at 37°C on Luria-Bertani (LB) agar plates with the antibiotic required for selection of transformants.

Sequence analysis

In the case of gene fragments from suppliers 1 and 2, we picked 16 colonies for each cloning reaction and checked for clones with an insert of the expected size by colony PCR using primers that flank the insert. The cloned GeneArt Strings DNA Fragments and GeneArt High-Q Strings DNA Fragments were processed the same way, but 96 colonies were screened for each cloning reaction.

Colony PCR products of expected sizes were prepared for sequencing using the Applied Biosystems™ ExoSAP-IT™ PCR Product Cleanup Reagent, and their sequences were determined by Sanger sequencing in-house. Sanger sequencing reads were then aligned with the reference sequence for the cloned DNA fragment, poor-quality reads discarded, and sequence errors counted in the resulting contigs. Accuracy was determined by dividing the total number of errors found in clones by the total number of sequenced base pairs.

Results

Determining error rate and impact on screening effort

To interrogate the accuracy of the GeneArt High-Q Strings DNA Fragments, we looked at the data derived from cloning between 16 and 18 linear DNA fragments from different sources (GeneArt High-Q Strings DNA Fragments, GeneArt Strings DNA Fragments, supplier 1, and supplier 2) and Sanger sequencing of individual clones. For the GeneArt High-Q Strings DNA Fragments we found 59 sequence errors in 929,877 sequenced base pairs. For the GeneArt Strings DNA Fragments we found 178 errors in 1,030,603 sequenced base pairs. For supplier 1's DNA fragments we found 74 errors in 137,530 sequenced base pairs, and for supplier 2's DNA fragments we found 69 errors in 125,405 sequenced base pairs. For further estimation of required screening effort and cost, we used the approximated values shown in Table 1.

Table 1. Comparison of accuracy. Error frequencies were calculated by dividing the total number of found sequenced errors by the total number of sequenced base pairs.

Product	Accuracy
GeneArt High-Q Strings DNA Fragments	Less than 1 error in 10,000 bp
GeneArt Strings DNA Fragments	Less than 2 errors in 10,000 bp
Suppliers 1 and 2	Less than 5 errors in 10,000 bp

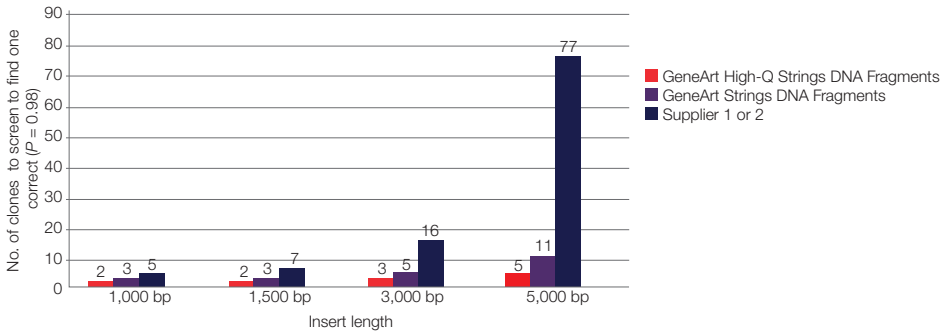

Figure 1. Number of clones that need to be sequenced in order to identify a correct clone with a probability of 98%. At an insert length of 1,000 bp we determined the need for 2 clones with the GeneArt High-Q Strings DNA Fragments vs. 5 clones with both other suppliers' products—a 60% reduction in screening effort.

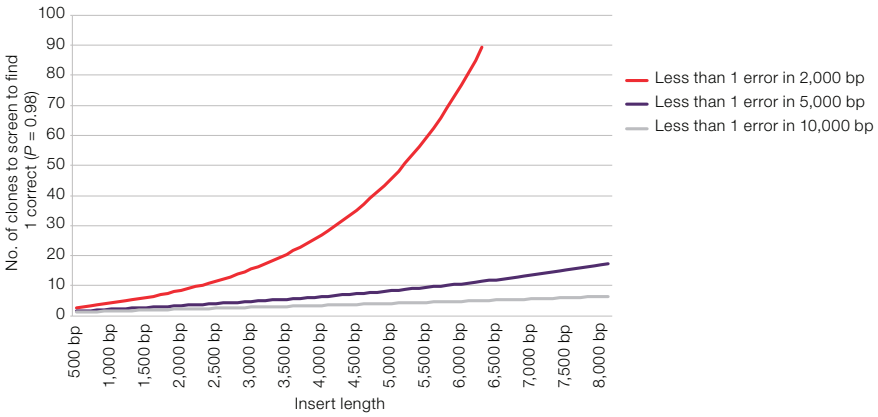

Figure 2. Number of clones required to choose the correct one, using 3 different error rates. The impact of a lower error rate is shown as the insert length increases.

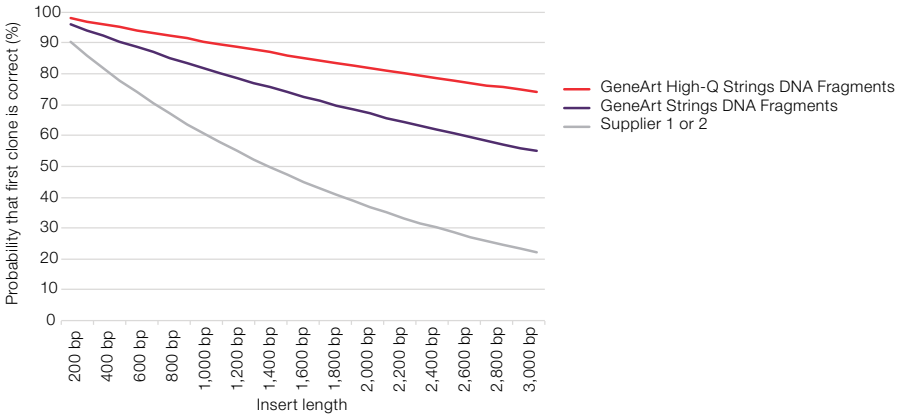

Figure 3. Probability curves showing that GeneArt High-Q Strings DNA Fragments allow screening of only 1 clone with a high probability of success, for inserts up to 1,000 bp.

Table 2. Cost analysis illustrating the savings of GeneArt High-Q Strings DNA Fragments compared to products from suppliers 1 and 2.

	Other suppliers' DNA fragments	GeneArt High-Q Strings DNA Fragments
1,000 bp fragment		
Screening and sequencing cost	$39.00	$16.50
DNA fragment cost	$149.00	$114.00
Total cost	$188.00	$130.50
Cost savings		**30%**
Additional total cost over GeneArt High-Q Strings DNA Fragments	$57.50	–

To evaluate the impact on screening effort, we took the error frequencies determined for the different products, and then we calculated the number of clones that need to be sequenced in order to identify a correct clone with a probability of 98% (Figure 1). Since this number increases exponentially with the size of the DNA fragment, differences in sequence fidelity become particularly substantial as fragments get longer (Figure 2). Furthermore, the probability that the first sequenced clone is correct can be calculated for inserts of different lengths (Figure 3).

Calculation of cost savings

In order to calculate the cost savings, we took the screening and sequencing costs plus the cost of a 1,000 bp fragment. Screening and sequencing costs were calculated from the number of clones required to identify a correct clone with the following assumptions: 90% of clones contain an insert of expected size; associated costs of $1.50 per colony PCR reaction or mini-prep, and $3.00 per sequencing reaction; 1,000 bp inserts require 2 sequencing reactions, 1,500 bp inserts require 4 sequencing reactions, and 3,000 bp inserts require 6 sequencing reactions. Our calculations showed that the GeneArt High-Q Strings DNA Fragments resulted in at least 30% lower cost when compared to fragments from suppliers 1 and 2, and 8% lower cost than the GeneArt Strings DNA Fragments, due to the improved error rate (Table 2).

invitrogen

Conclusion

Today's researchers are increasingly under pressure to obtain reproducible and reliable data in the shortest time possible. At Thermo Fisher Scientific we strive to meet the needs of the scientific community, and the cost-effective, productive technology of the GeneArt High-Q Strings DNA Fragments enables us to do so. The sequence accuracy of <1 error in 10,000 base pairs has resulted in 60% less effort in identifying a correct clone with >95% probability, compared to DNA fragments from suppliers 1 and 2. Furthermore, this optimization process translates into 30% cost savings when compared to these offerings suppliers 1 and 2—making the GeneArt High-Q Strings DNA Fragments a superior choice when evaluating linear gene synthesis products up to 1,200 bp.

Reference

1. Evaluation of linear synthetic DNA fragments from separate suppliers. (2015) White paper, Thermo Fisher Scientific.

Find out more at **thermofisher.com/highqstrings**

ThermoFisher SCIENTIFIC