

Access More Sequences with Ansa Biotechnologies

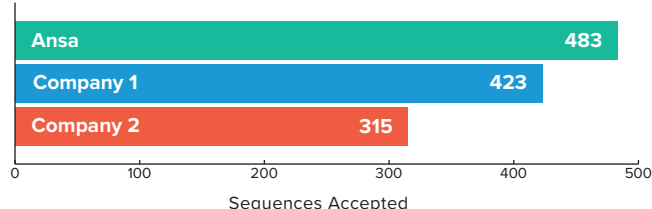
Modern biological experiments require the ability to interrogate large and diverse sets of DNA. The collection of naturally occurring DNA sequences is immense, and even larger is the set of potential synthetic DNA sequences. But DNA synthesis vendors struggle to deliver a vast portion of possible DNA designs. Their processes impose constraints on DNA accessibility that have prevented or delayed countless experiments.

To date, no DNA synthesis vendor has devised a process that can access all DNA sequences, but Ansa's enzymatic synthesis technology greatly increases the number and diversity of sequences that can now be produced. One of our core strengths is our ability to contiguously synthesize highly complex oligos up to 900 nucleotides long with industry-leading accuracy and with almost no sequence limitations. Using these oligos as input for our unique, sequence-agnostic assembly method, we are now building and shipping Ansa Clonal DNA up to 50 kb.

Examples of Accessible Sequences

Human Genome

From a random sample of 500 base-pair regions of the hg18 human genome reference, 96% of these sequences were given a standard complexity rating for clonal gene synthesis by Ansa, while other DNA synthesis companies accepted just 63% to 85% of these same sequences. The higher rate of rejection seen by other vendors is largely associated with complex regions inherent throughout the genome. As the length of the sequence increases, so does the probability of the sequence containing complex features, making it more likely that other vendors will reject the sequence.



Extreme GC Content

In a separate analysis, the human genome was randomly sampled to identify 600 base-pair regions containing a 50

base-pair window of high or low GC content. These regions were then filtered to 144 sequences representing an even distribution of local GC content from 10% to 20% and 80% to 90% across the set of sequences. Ansa successfully synthesized 96% of these sequences as gene fragments. Other DNA synthesis companies reject 25% to 58% of these sequences.

Promoters and 5' UTRs

Promoters and untranslated regions (UTRs) often contain features that can complicate manufacturing, such as low GC regions, repeat regions for protein binding, and inverted repeats that can form secondary structures. A random sample of promoters with 5' UTRs (the 600 bp directly 5' of the start codon) from the human, maize, and *Arabidopsis* genomes were synthesized as fragments by Ansa with a 90% success rate. Other DNA synthesis companies reject 20% to 65% of these same sequences.

AAV ITRs

Adeno-associated virus (AAV) inverted terminal repeats (ITRs) have a characteristic T-shaped secondary structure that consists of around 150 base pairs. Clonal gene synthesis of the ITRs from AAV serotypes 1-13 were manufactured by Ansa and all but one were synthesized successfully. Other DNA synthesis companies reject all of these sequences at the outset or fail to deliver them.

GC Content

Ansa can manufacture sequences containing regions of high and low GC content, including within a sequence.

Accepted local GC % within any 50 base-pair region	
Clonal DNA	10%-94%
DNA Fragments	6%-90%

Homopolymers

Long homopolymers present a unique challenge as length can be difficult to measure using sequencing. Ansa can synthesize long homopolymers and we employ homopolymer-specific QC measures to assess them.

Homopolymer Acceptance Criteria

Sizes of homopolymers accepted		
	DNA Fragments	Clonal DNA
A or T homopolymers	≤ 140 bp	≤ 140 bp
C or G homopolymers	≤ 20 bp	≤ 16 bp

Homopolymer QC

Ansa uses long-read sequencing to assess the quality of the DNA we produce, including measuring the length of all homopolymers. Ansa can accurately measure short homopolymer lengths; a construct passes only if at least 70% of the reads match the expected length.

Range of short homopolymer lengths subject to exact length QC		
	DNA Fragments	Clonal DNA
A or T homopolymers	≤ 19 bp	
C or G homopolymers	≤ 9 bp	

All sequencing technologies have increased error rates on homopolymers, making it difficult to have high confidence when measuring the length of a long homopolymer.

Range of homopolymer lengths subject to variable length QC		
	DNA Fragments	Clonal DNA
A or T homopolymers	≥ 20 bp	≥ 20 bp
C or G homopolymers	≥ 10 bp	≥ 10 bp

Long homopolymers pose the additional challenge of amplifying and propagating stably in *E. coli*. For this reason, a long homopolymer passes if at least 70% of reads fall within the length variation specified below.

Allowed homopolymer length variation		
	DNA Fragments	Clonal DNA
A or T homopolymers	± 10%	± 20%
C or G homopolymers	± 10%	± 10%

Repeats

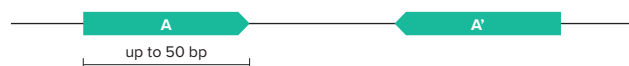
Tandem Direct Repeats

Repeated sequences that are adjacent to each other and in the same orientation, especially those with a repeat unit 2-6 base pairs long, can be difficult to manufacture. Ansa can manufacture these short repeat sequences.

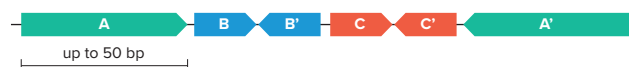
Length of repeat sequence accepted	
Dinucleotide repeat	≤ 20 bp
Trinucleotide repeat	≤ 30 bp
Tetranucleotide repeat	≤ 40 bp
Pentanucleotide repeat	≤ 50 bp
Hexanucleotide repeat	≤ 60 bp

Inverted Repeats

A repeated sequence where one copy is the reverse-complement of the other is an inverted repeat. There are often additional bases between the two regions of repeated sequence. Sequences where the length of one side of the inverted repeat is up to 50 base pairs are classified as standard complexity.



Multiple inverted repeats often occur within a region of DNA. For example, in adeno-associated virus inverted terminal repeats, a "T" shaped secondary structure is the result of three sets of inverted repeats. Multiple inverted repeats are considered standard complexity if one side of every individual repeat is 50 base pairs or less.



Conclusion

Accessing complex DNA sequences is imperative to enable continued advancements across healthcare, life science research, and other industries. Ansa's DNA synthesis service, powered by its proprietary enzymatic synthesis technology, marks a significant evolution in the ability to access new levels of complexity for synthetic gene fragments and clonal DNA.