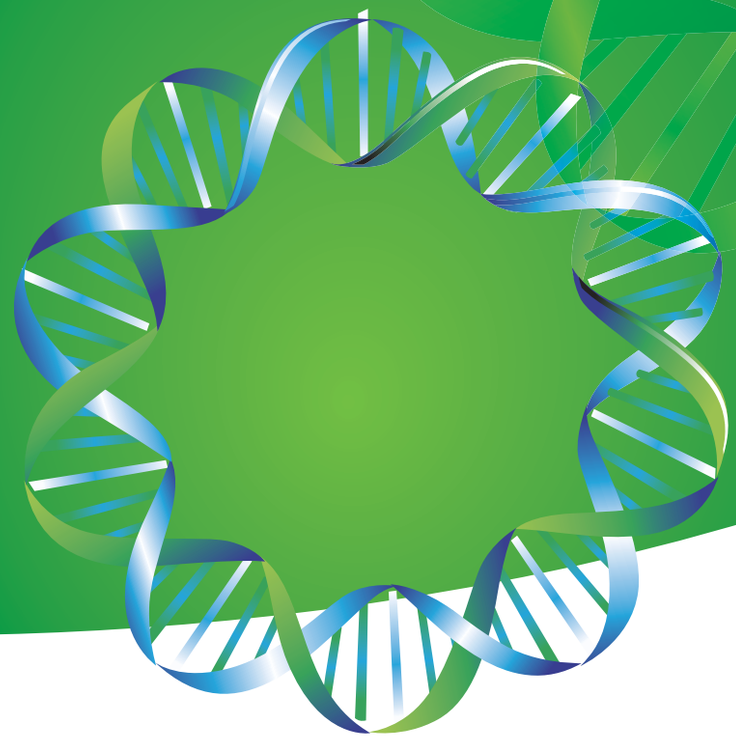
www.GenScript.com
GenScript USA Inc.
860 Centennial Ave.
Piscataway, NJ 08854 USA
Phone: 1-732-885-9188
Toll-Free: 1-877-436-7274
Fax: 1-732-885-5878

Molecular Cloning Handbook

Past, Present, and Future Techniques of Molecular Cloning



with ultimate molecular cloning troubleshooting guide

GenScript
Make Research Easy

Table of Contents

Traditional Molecular Cloning Overview

The History of Molecular Cloning

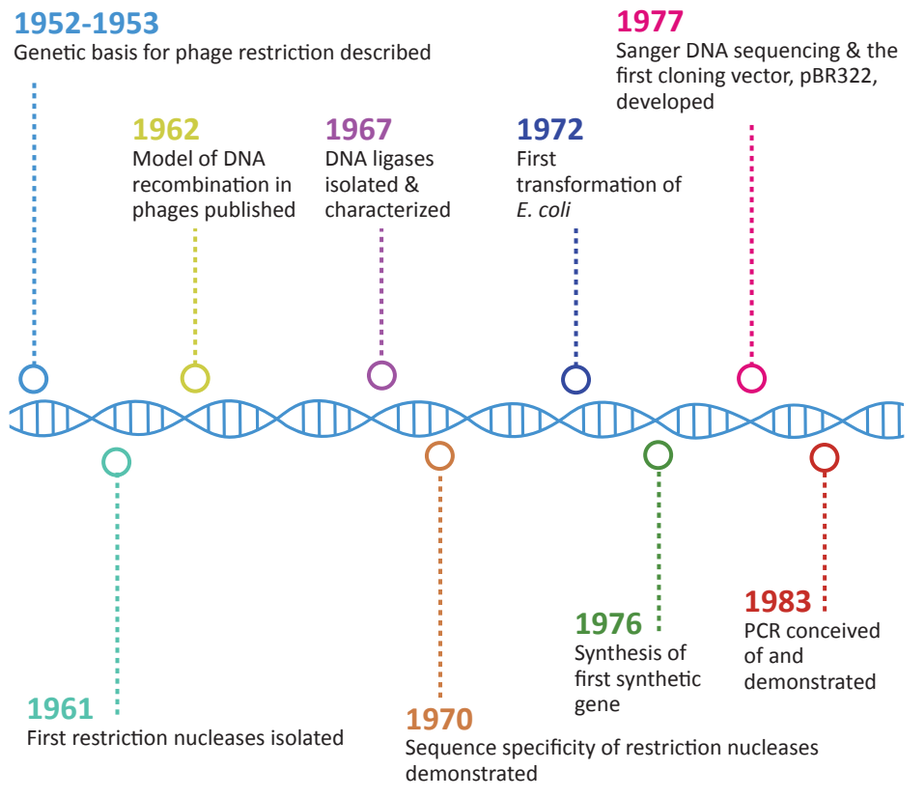
The key principles of molecular cloning were discovered a little over 50 years ago. Since then, molecular cloning has become one of the most powerful tools of the molecular biology laboratory enabling the expression of the smallest genes, as well as the engineering of whole genomes.

The advent of molecular cloning was spawned from a number of observations that centered around DNA recombination, namely the exchange of DNA bewteen bacterial and bacteriophage genomes [1, 2]. Key to these pioneering observations was the discovery that bacteriophage λ formed a circle when it entered the host bacterial cell. In particular, phage DNA was observed to have single stranded DNA flanking each end. The ends were complementary to each other and were called cohesive sites or "cos sites". The cos sites could reanneal to make the phage DNA circular. Then, excision and insertion events led to the incorporation of bacteriophage DNA into the host genome [3]. Campbell's model of λ prophage excision and insertion led to the classic principle of DNA recombination. In particular, today's "sticky ends", formed by DNA digestion using restriction enzymes mimic the functionality of cos sites allowing for the insertion of a target sequence into its destination vector.

Next came the discovery of another key tool in molecular cloning, the restriction enzyme. It was discovered that methylation of phage DNA by host methyltransferases prevented phage DNA from being destroyed by host enzymes called restriction nucleases. Foreign DNA molecules not having the methylation patterns in accordance with their host (or that were unmethlyated) were recognized as foreign and destroyed by host restriction nucleases [4, 5]. The first restriction nucleases were characterized by Meselson and Yuan [6]. Then, Kelly and Smith showed that restriction enzymes recognize and cleave specific nucleotide sequences [7]. Today, restriction enzymes are used as biological scissors to cut out the precise DNA desired to create recombinant sequences. The *Escherichia coli* strains of today used for transformation of cloned recombinant DNA lack these enzymes.

Finally, came the discovery of enzymes called ligases, which could join two DNA molecules together [8-12]. The discovery of ligases solidified the groundwork for the basic principles of molecular cloning – the creation of recombinant DNA via the cutting and gluing of DNA molecules together. Later innovations such as transformation, PCR, Sanger sequencing, and more recently, gene synthesis would make molecular cloning one of the most prolific tools of the molecular biology laboratory.

Figure 1. Timeline of molecular cloning history



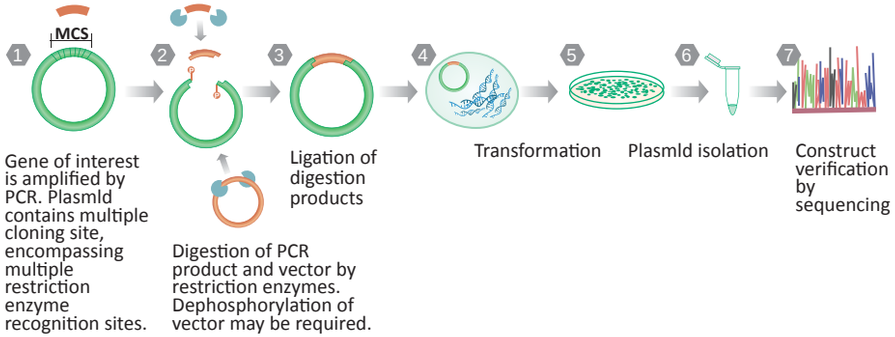
Molecular Cloning Strategies

Since the seminal discoveries of the basic principles underlying molecular cloning, a number of cloning strategies have been developed to improve on the ease and speed at which DNA fragments can be recombined.

Traditional cloning

Traditional cloning, also called PCR cloning, requires the use of the polymerase chain reaction (PCR) to amplify the template sequence of interest (usually the gene of interest) and add restriction sites to the ends of the sequence. Restriction enzymes are used to cut both the template of interest and the target vector, and DNA ligase is used to join the sticky ends of the template and vector together. Traditional cloning allows for flexible DNA sequence manipulation, which facilitates the building of nearly any desired construct. However, the checkpoints and optimization procedures required for traditional cloning can be cumbersome, and the reagents required can be expensive.

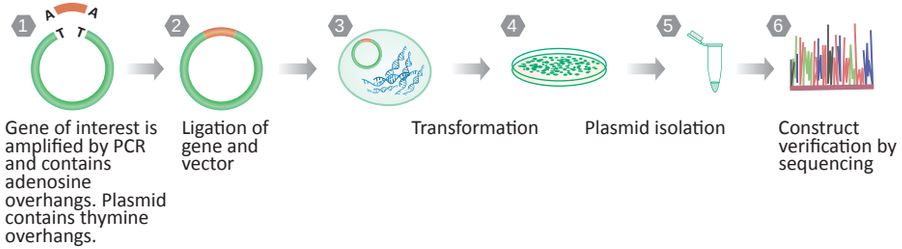
Figure 2: Schematic of PCR cloning



TA cloning

TA cloning is one of the simplest forms of cloning. In this method, vectors containing 5' thymine overhangs are used to accept PCR products in which additional 3' adenosine overhangs have been added on by the nature of TAQ polymerase amplification. TA cloning has the advantage of ease and speed, since no restriction digestion step is required. In addition, TA cloning kits contain reaction buffers that contain the pre-mixed vector, ligase, and buffer, cutting ligation reaction time to as few as 5 minutes. The disadvantage of TA cloning technology is that the cloning is not directional, meaning the gene of interest may be inserted into the target vector in either the sense or antisense orientation. Normally, half of the subsequent transformants will contain the gene in the sense direction and half will contain the gene in the antisense direction. However, cells transformed with toxic genes may all display the genes in the antisense direction, since cells containing the sense directed genes will not survive. In addition, survivor cells containing toxic genes oriented in the sense direction may be mutated to encode a less toxic protein.

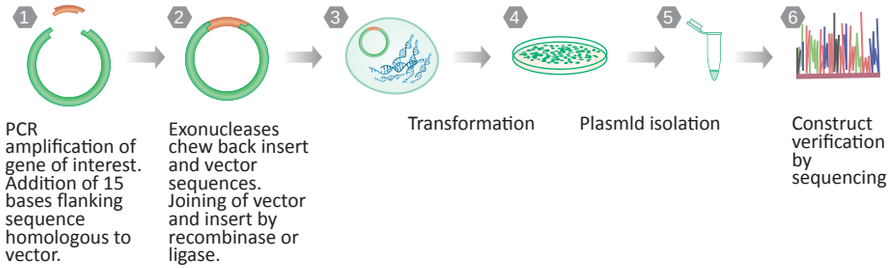
Figure 3: Schematic of TA cloning



Seamless cloning

Seamless cloning technologies eliminate the requirement for restriction enzymes. This can be advantageous when an insert contains a number of restriction sites within its sequence, making it difficult to identify restriction enzymes that will not cut the gene of interest during the cloning procedure. Seamless cloning takes advantage of homologous recombination and there are numerous variations on the technique. In general, the procedure consists of adding flanking sequences approximately 15 bp in length to both the insert and vector via PCR. Exonucleases are used to chew back the insert and vector sequences and the DNA is joined using recombinase enzymes or DNA ligase. Seamless cloning has been simplified by the development of kits that already contain the target vector and a proprietary mix of enzymes required for the recombination reaction. For instance, GenScript's CloneEZ™ Kit can clone inserts up to 10 kb in 30 minutes, and can also be used for high-throughput cloning projects.

Figure 4: Schematic of seamless cloning



Traditional Molecular Cloning Bottlenecks

Technical bottlenecks of traditional molecular cloning can cause the process to be time consuming and expensive. Smoothly run cloning experiments can take about 2-3 weeks to complete. However, many cloning experiments require troubleshooting in at least one phase of the process, increasing both cost and time consumption. Below is a list of common bottlenecks that plague the traditional cloning process.

Unstable, low yield RNA extraction – Molecular cloning often begins with total RNA extraction from the host containing the gene of interest. For the cloning of protein coding sequences, mRNA extraction is required. mRNA copy numbers for some genes can be low, making it difficult to amplify the sequence via RT-PCR. In addition, RNA extraction from some host tissues may prove difficult due to the harsh conditions under which some cell types must be lysed. Finally RNA is inherently less stable than DNA, and great precautions must be taken to prevent the action of RNases and the degradation of RNA by repetitive freezing and thawing.

Difficult PCR program optimization – Following the synthesis of cDNA from mRNA molecule templates, a PCR program must be designed to amplify the gene of interest, as well as add additional elements such as restriction sites or detection/purification tags. Intrinsic properties of gene sequences such as high GC content, long stretches of the same polynucleotide, and sequences encoding hairpin loop structures can all hinder PCR efficiency.

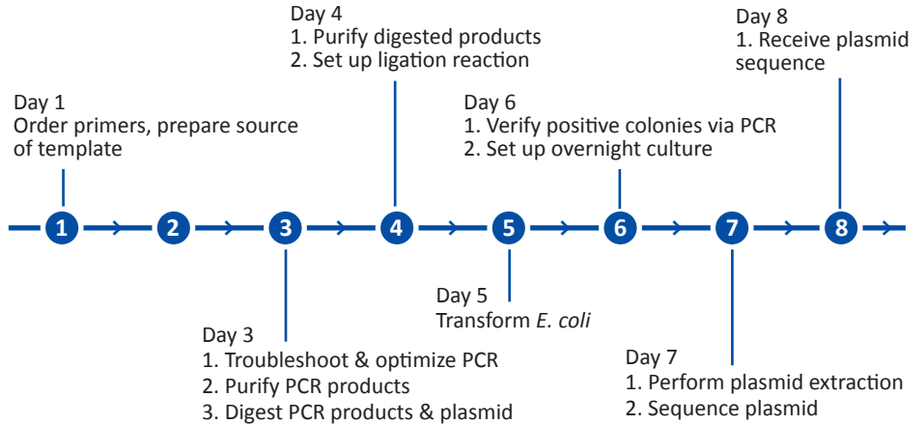
DNA losses from purification steps – DNA must be purified at numerous steps in the traditional cloning process. PCR products must be purified from reaction components; digested genes and vectors must be purified from restriction enzymes, buffers, and digestion products; and plasmid DNA must be extracted and purified from cellular material. At each step, DNA recovery % can vary from as high as 95% to as low as 60% depending on the purification method and quality of the DNA. Low recovery percentages or contamination during the purification procedure can result in DNA losses so great that the next step of cloning cannot be carried out, requiring steps such as PCR, digestion, or transformation to be repeated.

Low ligation efficiencies – Following the digestion of the gene insert and the target vector, a ligation reaction is performed to join the two molecules. The efficiency of ligation reactions are dependent on a number of variables including vector to insert ratio and salt concentration. If the ratio is too low, or plasmid digestion is incomplete, excess vector may re-anneal without the insert. In some cases the vector may require an additional de-phosphorylation reaction step or the insert may require an extra phosphorylation step, to prevent vector re-annealing and improve insert-vector ligation efficiency, respectively. In addition, solvent contaminants from DNA purification procedures such as agarose gel extraction can lower ligation efficiency.

Mutations – Mutations can be introduced into the molecular cloning process during PCR amplification or plasmid propagation. Typical Taq DNA polymerases do not contain proofreading subunits. Though high-fidelity Taq DNA polymerases that contain proofreading subunits, such as Pfu, have been developed, reaction conditions such as sustained high temperatures can still cause DNA damage that results in mutations.

During plasmid propagation, toxic genes may be expressed at low levels due to RNA polymerase read through of antibiotic resistance genes, or leaky promoters. Proteins or enzymes encoded by toxic genes may cause the transformed cells to die, resulting in only the survival of transformants containing mutated genes encoding less potent versions of the target protein or enzyme.

Figure 5: Traditional molecular cloning steps and timeline
At best, traditional cloning may take only 2 weeks, but troubleshooting after each phase may extend the cloning process to multiple weeks or months.



Molecular Cloning Troubleshooting Guide

Problem	Potential Causes	Solutions
Low or no PCR product yield	Number of PCR cycles is insufficient	• Increase number of PCR cycles by 5.
	Template is degraded	• Use electrophoresis to check DNA quality.
	Template is contaminated with PC inhibitors	• Check DNA ratio of absorbance at 260 and 280 nm. Pure DNA should have a 260/280 ratio of ≥ 1.8. • Use less volume of the template in the reaction. • Use DNA clean-up kit or ethanol precipitation to remove contaminants.
	Thermocycler program annealing and extension temperatures are not optimal	• Follow general rules of PCR design: Annealing temperature = lowest primer T_m - 5 °C, Extension temperature = 72 °C. • Decrease annealing temperature by 6 to 10°C in stepwise fashion.
	Reaction is missing Taq polymerase or other reaction component	• Make sure each component was added to PCR reaction.
	Primer concentration too low	• Check primer concentration; increase concentration if necessary.
	Target sequence is not in DNA template	• Re-extract DNA from source. • Test another region of template.
	Not enough template	• Increase concentration of DNA template.
	Reaction component concentrations not optimal	• Check recommended primer concentrations (normally from 0.05–1 mM) and Mg^{++} concentrations (normally from 0.2–1 mM).
	Reaction mix components are compromised	• Check expiration date of components. • Aliquot biological components of reaction mixture and avoid multiple freeze-thaw cycles.
	Primer design not optimal (causing non-specific annealing, or primer dimer formation)	• Follow general rules of primer design: Length from 18-30 nucleotides, GC content from 40-60%, T_m of primers within 5°C of each other. • Avoid stretches of 4 or more of the same nucleotide or dinucleotide repeats. • Avoid self-complementary sequences within primers.
Multiple/non-specific products from PCR reaction	Primer design not optimal (causing non-specific annealing, or primer dimer formation)	• Follow general rules of primer design: Length from 18-30 nucleotides, GC content from 40-60%, T_m of primers within 5°C of each other. • Avoid stretches of 4 or more of the same nucleotide or dinucleotide repeats. • Avoid self-complementary sequences within primers.
	Template or reaction mixture components are contaminated	• Re-extract template. • Try new reaction mixture. • Use filter pipette tips and wear gloves during reaction set-up.
	Annealing temperature too low	• Incrementally increase annealing temperature.
	Primer concentration too high	• Use less primer.
	Template concentration is not optimal	• For plasmids use 1 pg–10 ng of DNA / 50 µl reaction. • For genomic DNA use 1 ng–1 µg of DNA / 50 µl reaction.
	Enzyme quality is compromised	• Check enzyme expiration date. • Avoid multiple freeze thaw cycles of enzyme.
	Methylation of template sequence prevents enzyme from cutting	DNA from a bacterial source or eukaryotic source may be blocked by Dam and Dcm methylation or CpG methylation, respectively. • Use a Dam- or Dcm- strain.

Incomplete or no template/vector digestion	Not enough incubation time	• Add 1 to 2 hours to incubation time.
	Template/plasmid contamination with inhibitors	Inhibitor contamination can originate from miniprep kits, or leftover PCR reaction components: • Use dialysis or spin column to remove contamination. • Dilute or use less volume of DNA - DNA solution should compose no more than 25% of digestion reaction. • Use PCR clean up kit to remove PCR reaction components.
	Enzyme concentration too low	• Use 3–5 units of enzyme / µg of DNA.
	Incompatible buffer was used	• Use recommended buffer with enzyme.
Digestion product is smeared on gel	Template is contaminated with nucleases	• Use commercial kit to clean up template. • Use running buffer and agarose gel made with fresh nuclease-free water.
	Enzyme is bound to DNA substrate	• Decrease units of enzyme. • Remove enzyme from DNA by adding 0.1–0.5% SDS to loading buffer.
Template/vector appears to be over digested	Star activity (enzyme cleaved similar recognition sequence)	• Use minimum incubation time. • Reduce number of enzyme units in reaction. • Use compatible buffer with enzyme. • Use a high-fidelity enzyme.
	Wrong antibiotic was used or antibiotic concentration was too high	• Ensure the correct antibiotic was applied to plates. • Use only concentration recommended by competent cell or antibiotic manufacturer.
	Competent cell viability is low	• Thaw competent cells on ice and use immediately. • Check expiration date of cells. • Do not re-freeze cells. • Do not vortex cells - gently tap to mix.
Few or no colony transformants	DNA insert encodes protein that is toxic to cells	• Use a lower incubation temperature (25 – 30°C). • Use a cell strain and vector designed for tightly controlled transcription.
	Heat-shock incubation too long	• Reduce incubation time from 45 to 25 seconds.
	Construct is too big	• Use electroporation for vectors over 10 kb.
	Too much ligation mixture was used for the transformation	Ligation reaction components can inhibit transformation. • Dilute ligation reaction with TE buffer (up to 5 times).
	Too much DNA in reaction	• Use no more than 1-10 ng of DNA in 5 µl for a 100 µl reaction or in 1-3 µl for a 50 µl reaction.
	Low ligation efficiency	• Vector insert ratio not optimal. Use a vector:insert molar ratio from 1:1 to 1:10. • Use a DNA concentration of 1-10 µg/ml.
	Construct recombined with genomic DNA	• Switch to a Rec A- cell strain.
No plasmid in colony transformants	Antibiotic concentration too low	• Use antibiotic concentration recommended by manufacturer.
	Antibiotic is degraded	• Aliquot working volumes of antibiotic and avoid freeze-thaw cycles. • Add antibiotic to liquid plate media after sufficient cooling.
No insert in colony transformant plasmids	Vector re-ligation	• Vector insert ratio not optimal. Use a vector:insert molar ratio from 1:1 to 1:10. Use a DNA concentration of 1-10 µg/ml. • Dephosphorylate DNA with phosphatase to prevent re-ligation.
Sequencing of transformant plasmid reveals wrong plasmid sequence	DNA insert encodes protein that is toxic to cells	• Use a lower incubation temperature (25 – 30°C). • Use a cell strain and vector designed for tightly controlled transcription.
	Mutations introduced by initial PCR	• Use a high-fidelity polymerase.
	Inconclusive sequencing artifacts	• Repeat sequencing reaction.

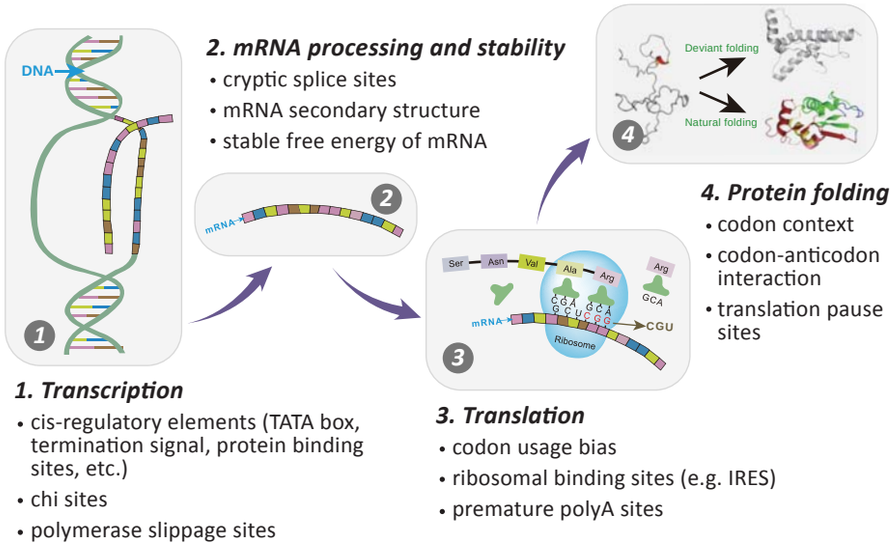
Other Limitations of Traditional Molecular Cloning

In addition to the bottlenecks encountered in traditional molecular cloning, other limitations avail. Traditional cloning is most useful for stitching together naturally occurring sequences. Engineering optimized or non-natural sequences can be much more difficult. In particular, gene sequences may be engineered to contain mutations for performing functional assays of the resulting protein or for characterization of the disease state of a protein. In other cases, a researcher may desire to optimize the codons of a natural gene sequence to enhance protein expression. In both cases, additional manipulation of the cloned gene is required, which may result in additional DNA losses and unwanted mutations.

Codon Optimization

In order to characterize a gene's function or purify a protein for structural study, researchers frequently need to express a gene in a host cell or model organism different from the one in which it appears in nature. Unfortunately, heterologously expressed genes often suffer from low rates of protein expression. A common cause of low expression levels is the variation in codon usage frequency between different species. This can cause the translation of heterologously expressed genes to stall due to tRNA scarcity, leading to lower protein levels and increased rates of improper protein folding. The degeneracy of the genetic code makes it possible to change the DNA sequence in a way that does not alter the final amino acid sequence, but does have significant effects on the efficiency of transcription, mRNA stability, translation, and protein folding. Aside from codon usage bias, many other features of a DNA sequence affect the efficiency of transcription, the proper splicing and processing of mRNA, and mRNA stability, all of which in turn reduce the ultimate protein yield (Figure 6).

Figure 6: DNA sequence features that influence protein expression levels



Modifying a DNA sequence to optimize it for efficient heterologous protein expression while preserving the amino acid sequence is called codon optimization. In a study by Burgess-Brown *et al.* [13] that investigated the expression of 30 human genes in *E. coli*, codon optimization increased both the total protein yield and the protein solubility, compared to heterologous expression of native gene sequences. Codon optimization algorithms have improved since this 2008 study, and codon optimization is now widely used by researchers to facilitate the purification of proteins for structural studies, enzyme kinetics, and other biochemical investigations.

Mutagenesis

The mutation of gene sequences is useful for identifying amino acids crucial to protein function, studying disease states of native proteins, facilitating drug discovery or developing new protein variants (mutants) with properties desired for medical, industrial, agricultural, or other applications. Traditional methods for inducing random mutations include subjecting live cells or animals to UV radiation or chemical mutagens in order to induce DNA damage. DNA repair enzymes can, in some cases, introduce or preserve mutations resulting from damage such as single- or double strand breaks, but this method leads to high rates of lethality and low rates of useful mutant phenotypes.

Depending on the type of mutagen used, mutations may be biased, limiting the range of possible mutants to only a fraction of those that are theoretically possible or that can be obtained through other strategies.
Site-directed mutagenesis may be designed based on experimental data such as X-ray crystallography data that identifies residues of interest. Site-directed mutagenesis can be performed using PCR-based methods with mutated primer sequences to create and selectively amplify mutated sequences. This type of mutagenesis must be followed by sequence verification, and can be tedious, error-prone, and costly.

Gene Synthesis – The Solution to the Limitations of Traditional Cloning

Luckily, a method that circumvents the limitations of traditional molecular cloning strategies has been developed – gene synthesis. Gene synthesis is the *de novo* synthesis of a DNA strand. No template is required for gene synthesis and nearly any sequence that can be designed can be synthesized. In the next section the applications and techniques that demonstrate how the versatility of gene synthesis has both accelerated and revolutionized molecular cloning are discussed.

How Gene Synthesis Has Revolutionized Molecular Cloning

What is Gene Synthesis?

Gene synthesis technologies have revolutionized biology research and provide a more straightforward, faster, and less costly alternative to traditional molecular cloning. Gene synthesis involves the *de novo* chemical synthesis of DNA, differing from traditional molecular cloning in that no template is required. Gene synthesis allows researchers to specify a desired sequence (native or engineered) and custom-build it directly.

Gene synthesis refers to chemically synthesizing a strand of DNA base-by-base. Unlike DNA replication that occurs in cells or by PCR, gene synthesis does not require a template strand. Rather, gene synthesis involves the step-wise addition of nucleotides to a single-stranded molecule, which then serves as a template for creation of a complementary strand.

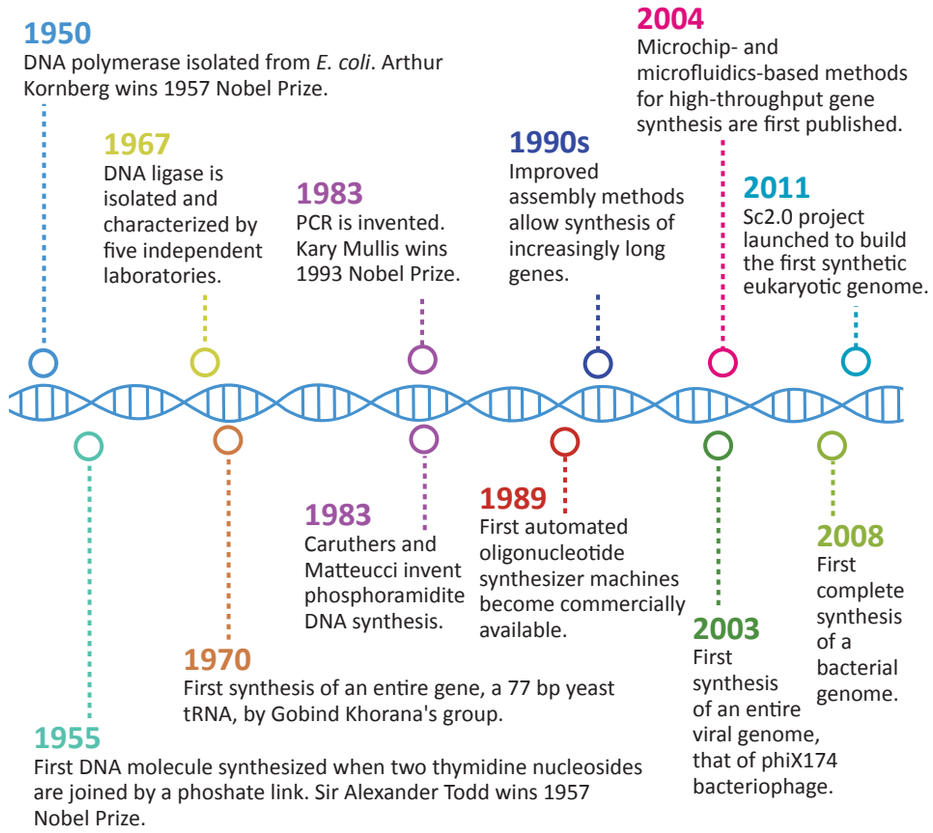
Any DNA sequence can be synthesized, including sequences that do not exist in nature, or variants on naturally-occurring sequences that would be tedious to produce through site-directed mutagenesis, such as codon-optimized sequences for increased heterologous protein expression. Synthetic DNA can be cloned into expression vectors and used in any protocol that requires natural or recombinant DNA. Synthetic genes are used to study all the diverse biological roles that nucleic acids play, from encoding proteins and regulating gene expression in the nucleus, to mediating cell-cell communication and building biofilms from extracellular DNA.

Due to gene synthesis, scientists are no longer limited to classical methods of manipulating a single gene at a time; they now have the power to design or reprogram entire genomes and cells which can advance the fields of healthcare, agriculture, energy, and other fields of human endeavor. For instance, newly identified viral genomes can be quickly synthesized to accelerate vaccine development. Novel enzymes can be engineered to fight cancer and produce sustainable biofuels. Genes can be engineered to enhance crop yields and reduce vulnerability to the common pests and plant diseases that endanger food supplies and contribute to global hunger. Designer metabolic circuits using interchangeable synthetic parts can be built to create synthetic genomes and artificial cells for studying the basic requirements of life.

Methods for *de novo* chemical synthesis of DNA have been refined over the past 60 years. Synthetic short oligonucleotides (oligos) serve as custom primers and probes for a wide variety of applications. Longer sequences that serve as genes or even whole genomes can be synthesized as well; these sequences are typically produced by synthesizing 40-200 bp oligos and then assembling them in the proper order. Many methods for oligo assembly have been developed that rely upon a DNA polymerase enzyme for PCR-based amplification, a DNA ligase enzyme for ligation of oligos, or enzymes that mediate homologous recombination *in vitro* or *in vivo*. Most sequences up to 1000 base pairs (1 kb) can be assembled in a standard molecular biology lab, and commercial gene synthesis providers routinely synthesize sequences over 10 kb.

The history of gene synthesis began in 1955, when Sir Alexander Todd published a chemical method for creating a phosphate link between two thymidine nucleosides, effectively describing the first artificial synthesis of a DNA molecule [14] The first successful synthesis of an entire gene was reported by Gobind Khorana's group in 1970; the 77 bp DNA fragment took 5 years to synthesize [15] Subsequent improvements in DNA synthesis, sequencing, amplification, and automation have made it possible now to synthesize genes over 1 kb in just a few days, and to synthesize much longer sequences including entire genomes. Gene synthesis can now be easily and cost-effectively outsourced to commercial providers. **GenScript, a pioneer in gene synthesis, was founded in 2002 and is the largest gene synthesis supplier in the world.**

Figure 7: Timeline of gene synthesis technology development



What Can Be Synthesized?

Gene synthesis can generate recombinant, mutated, or completely novel DNA sequences without a template, simplifying the creation of DNA tools that would be laborious to produce through traditional molecular cloning techniques. A wide variety of types of sequences can be produced to aid in diverse research applications (See Table 1). In addition to DNA sequences, RNA and oligos containing modified bases or chimeric DNA-RNA backbones can also be synthesized. However, the most widely used synthetic sequences are customized DNA of the following types:

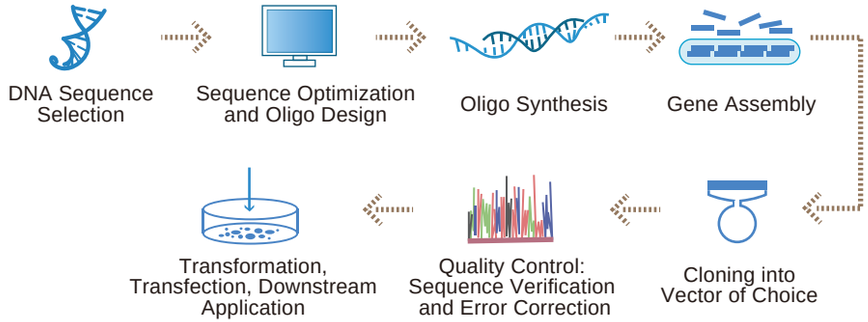
Table 1: Types of synthetic DNA and their research applications

Types of Synthetic DNA	Research Applications
cDNA / ORFs	Over-expression or heterologous expression
Customized coding sequences	Expressing fusion proteins; high-level protein expression from codon-optimized sequences for purification for enzymatic or structural studies
Promoter-reporter constructs	Monitoring gene expression downstream of manipulations to transcription factors, signaling cascades, etc.
Genomic DNA	Creating synthetic genes or genomes; studying gene structure, regulation, and evolution.
Mutant sequences	Confirming function (promoter-bashing; amino acid substitutions); Protein engineering (rational design or screening/directed evolution)
RNAi constructs (shRNA, miRNA, siRNA)	Regulating (suppressing) gene expression; intercellular communication
Extracellular DNA	Biofilm formation, intercellular signaling as in cancer metastasis

How is Gene Synthesis Performed?

The basic steps of gene synthesis are:
1. sequence optimization and oligo design
2. oligo synthesis
3. gene assembly
4. sequence verification and error correction
5. preparation of synthetic DNA for downstream applications

Figure 8: Steps in the *de novo* gene synthesis process



Step 1: Sequence Optimization and Oligo Design

Sequence Optimization

Once a gene of interest has been selected, the sequence to be synthesized must be designed with the end application in mind. For example, codon optimization is appropriate if the goal is to maximize heterologous protein expression levels, however, it may not be appropriate for the study of endogenous regulation of gene expression. For constructs containing multiple segments, the intended reading frame should be maintained throughout the entire coding region. The addition of short flanking sequences can facilitate later excision or recombination, via restriction enzymes or similar tools. Sequences should be designed to minimize restriction enzyme recognition sites or other sequences that might interfere with downstream workflow. Careful attention should be paid to functional domains such as *cis*-regulatory elements or RNase splice sites, which may be unintentionally introduced during codon optimization. The presence of biologically functional sequences such as *cis*-regulatory elements or RNase splice sites on oligos may hinder *in vivo* assembly, maintenance, or expression.

Some sequence features make synthesis more challenging, including: extremely high or low GC content; highly repetitive sequences; complex secondary structures such as hairpins; unstable structural elements; polyA stretches; and longer sequences, especially over 1 kb. These are all important features to consider when selecting and optimizing sequences to synthesize, and they will increase the cost and turnaround time of synthesis, whether it is performed in-house or outsourced to a commercial service provider.

Oligo Design

After finalizing the sequence that will be synthesized, sequence analysis is required to determine the best way to divide the whole gene into fragments that will be synthesized and then assembled. Very large synthetic gene sequences, should be divided into chunks of 500-1000 kb to be synthesized separately and assembled later.

While synthesis platforms utilizing phosphoramidite chemistry have very low error rates, errors do accumulate as strand length increases; therefore, gene synthesis typically uses oligos with lengths of 40-200 bp. The optimal oligo length depends upon the assembly method that will be used, the complexity of the sequence, and the researcher's preferences. As a general rule, shorter oligos may have a lower error rate, but will be more expensive to synthesize because more overlaps will be required. Longer overlaps increase the likelihood of correct assembly by decreasing the rate of nonspecific annealing. Oligos should be adjusted for even lengths and equal melting temperatures. Each oligo design tool has its own default parameters; for example, GeneDesign uses default settings of 60 bp oligos with 20 bp overlaps-values which typically work well with yeast and mammalian sequences which are ~40% GC.

In addition to oligo length, some factors to consider in selecting oligo sequences include GC content, sequence repeats, and the tendency for hairpin formation. GC content determines the stability of DNA strands and thus the melting temperature. All *in vitro* assembly methods, both ligase and polymerase based, rely upon the melting and annealing of oligos and thus require that oligos have equal melting temperatures.

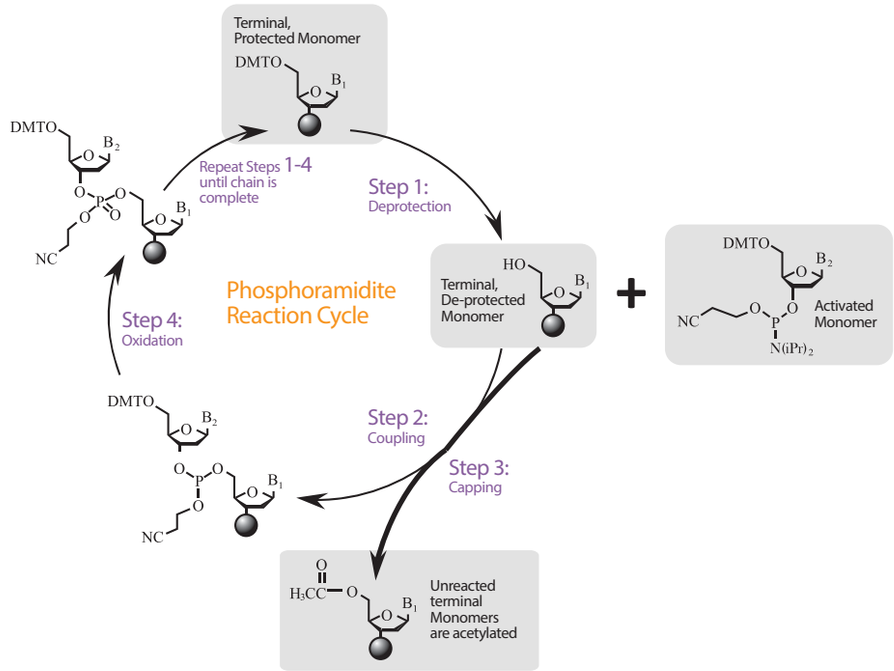
GenScript's proprietary DSOpt™ Sequence Analysis tool, developed through years of experience as the volume leader in gene synthesis, takes the guesswork out of selecting the best assembly method and oligo design strategy for each project.

Repeated sequences may pose challenges for correct annealing or ligation of oligos during assembly. Repeats may occur in many forms, including direct, inverted, palindromic, or tandem repeats. Improper hybridization or intramolecular binding such as hairpin structure formation can be avoided through careful design of oligo sequences, including the overhangs or primers used to assemble them.

Step 2: Oligo Synthesis

All DNA fabrication today begins with the step-wise addition of nucleotide monomers via phosphoramidite chemistry to form short oligonucleotides. Oligo synthesis via phosphoramidite chemistry uses modified nucleotides, called phosphoramidites, to ensure that nucleotides assemble in the correct way and to prevent the growing strand from engaging in undesired reactions during synthesis. The phosphoramidite group is attached to the 3' O and contains both a methylated phosphite and a protective di-isopropylamine to prevent unwanted branching. Phosphite is used because it reacts faster than phosphate. Methyl groups are attached to the phosphite, and amino-protecting groups are added to the bases, to protect against unwanted reactions until oligonucleotide synthesis is complete. Although DNA synthesis in living cells always occurs in the 5' to 3' direction, phosphoramidite synthesis proceeds in the 3' to 5' direction. The first monomer is attached via the 3' O to a solid support such as a glass bead, and its 5' O is initially protected from nonspecific reactions by conjugation of a dimethyloxytrityl (DMT) group.

Figure 9: Phosphoramidite reaction cycle



In Step 1: Deprotection, DMT is removed by washing with a mild acid such as trichloroacetic acid, exposing the 5' O for reaction. The second nucleotide is introduced to the reaction with its own 5' O protected with DMT, while its 3' O is activated by the conjugated phosphoramidite group. **Step 2:** Coupling occurs when the 3' O of the second nucleotide forms a phosphate triester bond with the 5' O of the first nucleotide. **Step 3:** Capping involves acetylating the 5' OH of any unreacted nucleotides to prevent later growth of an incorrect sequence. Acetic anhydride and dimethylaminopiridine are typically added to acetylate any unreacted terminal 5' OH groups, but will have no effect on terminal nucleotides that are protected by DMT. **Step 4:** Oxidation occurs upon the addition of iodine to convert the phosphate triester bond into the phosphodiester bond that forms the familiar backbone of DNA.

Once the desired chain is complete, all of the protecting groups must be removed, and the 5' end of the oligo is phosphorylated. Prematurely terminated strands can be removed by purifying the eluted product via gel electrophoresis and cutting out the band with the correct length. Further characterization of synthetic oligonucleotides may include sequencing or simply verifying the predicted molecular mass by mass spectrometry.

Step 3: Gene Assembly

Many methods of assembling oligos into complete genes or larger genome building blocks have been developed and successfully used [16-32]. For relatively short sequences (up to 1 kb), polymerase-based or ligase-based *in vitro* assembly methods are sufficient. For longer sequences, *in vivo* recombination-based methods may be preferred. Correct assembly requires a high-fidelity enzyme (e.g. DNA polymerase or ligase).

Polymerase chain assembly (PCA) is a standard technique for polymerase-based oligo assembly in a thermocycler. This reaction is also called templateless PCR. The principle is to combine all of the single-stranded oligos into a single tube, perform thermocycling to facilitate repeated rounds of annealing, extension, and denaturation. Then use the outermost primers to amplify the full-length sequence. The success of this method depends upon the accurate synthesis of oligos designed to possess sufficient regions of overlap, have sufficiently similar melting and annealing temperatures, and have minimal opportunities for mishybridization. Protocols for PCA to produce final sequences up to ~750 kb have been published[13].

Sequence- and Ligation-Independent Cloning (SLIC) is a method of *in vitro* homologous recombination employing a T4 DNA polymerase that allows the assembly of up to five gene fragments via simultaneous incorporation into a plasmid vector [14]. Synthetic oligos are prepared for SLIC by PCR extension to introduce flanking regions of sequence homology. This facilitates recombination of fragments without any sequence restrictions or the introduction of restriction enzyme sites that will produce permanent seams. The exonuclease activity of T4 DNA polymerase generates single-stranded DNA overhangs in the insert and vector sequences. Homologous regions are annealed *in vitro* and undergo gap repair after transformation into *E. coli*.

Step 4: Sequence Verification and Error Correction

Due to the inherent potential for error in each step of gene synthesis, all synthetic sequences should be verified before use. Sequences harboring mutations must be identified and removed from the pool or corrected. Internal insertions and deletions, as well as premature termination, are common in synthetic DNA sequences. The accumulation of errors from phosphoramidite chemical synthesis alone can lead to only about 30% of any synthesized 100-mer being the desired sequence [33]. Improper annealing during oligo assembly can also introduce heterogeneity in the final pool of synthetic gene products.

Cloning newly synthesized sequences into a plasmid vector can simplify the process of sequence verification. Sequencing primers that bind to vector regions flanking the gene insert ensure correct sequencing of the ends of the synthesized gene insert. Further, plasmid DNA can be clonally amplified to create a homogeneous pool of DNA with the correct sequence.

In the event that the correct sequence cannot be obtained and amplified from the pool of synthesized DNA, numerous methods for error detection and correction have been developed and successfully used. These include stringent hybridization using carefully designed oligos; exhaustive purification using electrophoresis, mass spectrometry and other biochemical methods; mismatch-binding or mismatch-cleavage using prokaryotic endonucleases; selection of correct coding sequences via functional assays; and site-directed mutagenesis after sequencing [34]. PAGE or agarose gel purification is the technique in longest use, but it is costly and labor intensive and does not identify or correct for substitutions or for small insertions/deletions. Enzyme-based strategies are limited by the enzyme's capabilities; for example, the widely used *E. coli* MutHLS is not very effective for substitutions other than G-T, A-C, G-G, A-A [34]. Site-directed mutagenesis after sequencing can introduce point mutations using mutant primers and high-fidelity DNA polymerase followed by selection for unmethylated molecules, but can be an unwieldy technique. Because error correction can be so time-consuming and costly, especially for long or complex sequences, efforts continue to improve the accuracy of oligo synthesis and assembly.

GenEZ™ Custom Molecular Clones and ORF Clones Powered by Gene Synthesis

What are Custom Clones?

While it is possible to perform both traditional molecular cloning and gene synthesis in the laboratory, both of these procedures can be tedious, time-consuming, and laborious. Fortunately a number of companies offer gene synthesis and custom cloning as a service, which delivers the desired gene in the desired vector in less time than it would take to build the construct in the lab.

Not only can cloning and gene synthesis services accelerate the generation of gene constructs in the molecular biology lab, but they can allow for an influx of contributions from scientists who may previously have been excluded from certain interdisciplinary research due to their inexperience in molecular biology techniques. For instance, a bioprocess engineer who conceives of a novel genetic circuit that can increase lipid production for a biofuel application may not necessarily need to become well-versed in molecular biology, nor purchase molecular biology research equipment to implement his ideas. He could simply order the custom genetic circuits that will function as he has designed, and then immediately begin his experimentation.

Below we review two of the most powerful services in the custom cloning industry – GenEZ™ Custom Molecular Clone and GenEZ™ ORF Clone services. Powered by GenScript's proprietary gene synthesis technologies, these services deliver mutation-free gene sequences in a time and cost-effective manner.

GenEZ™ Custom Molecular Clones – The next-generation of molecular cloning

GenScript's GenEZ™ Custom Molecular Clone service combines GenScript's proprietary gene synthesis technologies with custom cloning to deliver sequence-verified, custom genes in your vector of choice in as few as 2 weeks. GenEZ™ Molecular Clones are ideal if you desire non-natural engineered sequences, mutant sequences, codon optimized sequences, or other customized recombinant DNA constructs.

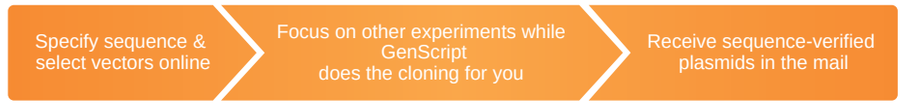


Table 2. Top advantages of GenEZ™ Custom Molecular Clones

No sequence restrictions	None. GenEZ™ sequences are not restricted by sequence complexity. If you can design it, we can synthesize it
Expansive vector selection	100 popular in-stock vectors or any vector of your choice
Sequence-verification	Each clone is guaranteed 100% sequence-verified
Downstream services	• Custom cloning • Custom plasmid preparation (research and transfection grade) • Mutagenesis service • Custom protein expression

What are ORF Clones?

Despite the rise in popularity of fields such as genetic engineering and synthetic biology, molecular cloning is still largely used for the expression of native proteins. Structural and functional characterization of these proteins was spurred in part by the completion of the Human Genome Project, and more recently has been further fueled by advances in proteomics and proteome mapping which postulate that the human genome encodes for some 17,000+ proteins [35, 36].

Native proteins and enzymes are encoded by open reading frame (ORF) sequences, which exclude introns, and 5' or 3' untranslated regions such as terminators and promoters. To facilitate protein expression, these protein encoding ORF sequences are cloned into expression vectors that may contain N- or C-terminal tags for protein detection or purification.

GenEZ™ ORF Clones – The easiest way to clone

GenScript's GenEZ™ ORF Clone service is powered by our proprietary gene synthesis technology, which has allowed us to synthesize over 40,000 ORF sequences *de novo*, which are currently in stock. Our gene synthesis technology has also allowed us to expand our ORF clone collection to any protein encoding sequence registered in the NCBI RefSeq database – over 2 million sequences synthesized on demand. GenEZ™ ORF clones can be shipped in as few as 5-days.

Cost-efficient

- Starting at **$99/clone***
- Fully sequence verified

While most ORF clone databases are limited to human, mouse, or rat ORF sequences, the combination of our in-stock clones and on demand sequence synthesis capability affords our customers the opportunity to select their desired ORF clone from the largest ORF sequence database in the world.

GenEZ™ ORF species coverage

The GenEZ™ ORF Clone database includes **2 million** ORFs from **186 species** including but not limited to:

- **Human**
- **Mouse**
- **Rat**

- **Mammals**
  - *Ovis Aries*
  - *Oryctolagus cuniculus*

- **Plants**
  - *Arabidopsis thaliana*
  - *Chlamydomonas reinhardtii*
  - *Nicotiana benthamiana*
  - *Oryza sativa*

- **Vertebrates**
  - *Danio rerio*
  - *Gallus gallus domesticus*
  - *Xenopus laevis*

- **Invertebrates**
  - *Caenorhabditis elegans*
  - *Drosophila melanogaster*

- **Fungi**
  - *Saccharomyces cerevisiae*
  - *Schizosaccharomyces pombe*
  - *Neurospora crassa*

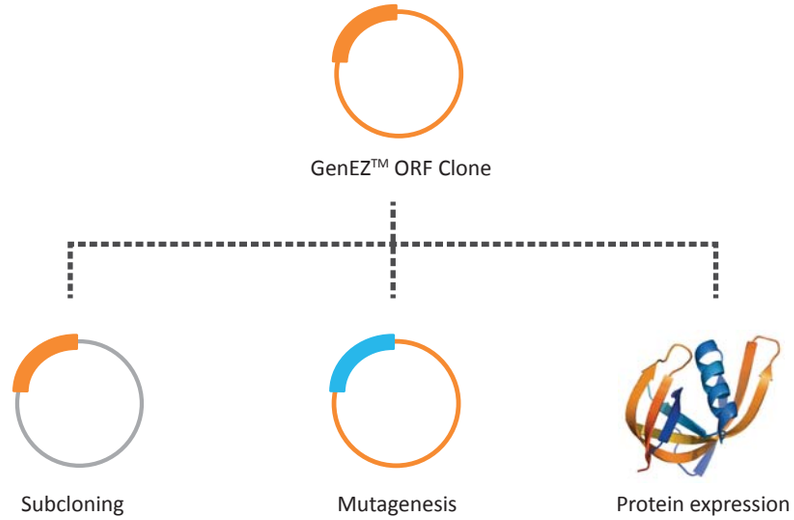
Largest ORF database in the world

- **> 2 million** ORFs from **186** species*
- **> 40,000** ORFs in-stock

**Available for on-demand synthesis*

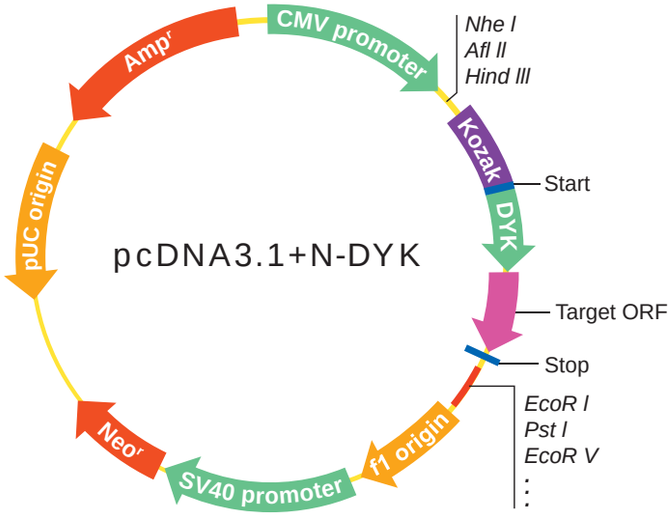
GenEZ™ ORF Custom Cloning and Mutant Clones

GenEZ™ ORF clones are highly customizable and can be custom cloned into the vector of your choice, mutated, or undergo our downstream protein expression service.



Standard expression vector

GenScript's GenEZ™ ORF clones are seamlessly cloned into our standard mammalian expression vector, pcDNA3.1⁺N-DYK using our CloneEZ™ seamless cloning technology. Driven by a CMV promoter, pcDNA3.1⁺N-DYK is equipped with an N-terminal DYKDDDDK tag for easy protein detection and purification. GenEZ™ ORF clones are delivered expression and transfection ready, saving researchers valuable time in the lab.



Vector name	pcDNA3.1$^+$-DYK
Vector length (w/o insert)	5453 bp
Promoter	CMV
N-terminal flag	DYDDDDK
Bacterial selection	Amp^R
Mammalian selection	Neo^R

Custom vectors

For enhanced flexibility, GenEZ™ ORF clones can be directly custom cloned into one of over 100 expression vectors which are compatible with mammalian, bacterial, and yeast expression systems, including **pcDNA, pET, pPIC,** and **pUC** vectors series.

GenEZ™ custom cloning vectors include but are not limited to the following vector series.

Customized vectors

- Only **$50** for subcloning
- Choose from **> 100** expression vectors
- Add **up to 30 bp** of 5' and 3' flanking sequence for free

pCDNA	pUC	pCold	pENTR	pGEX
pET	pBluescript	pEGFP	pFastBac	PQE
pPIC	pYES	pGL	pMAL	

In addition ORFs cloned into custom vectors can be further customized by the addition of up to 30 bp of 5' or 3' flanking sequence – free of charge.

Selected GenEZ™ ORF Clone vectors

Vector name	Promoter	Bacterial Selection	Epitope tag	Length
pcDNA3.1+N-DYK	CMV	Ampicillin	N-DYK	5453 bp
pcDNA3.1+N-HA	CMV	Ampicillin	N-HA	5450 bp
pcDNA3.1+N-6His	CMV	Ampicillin	N-6*HIS	5441 bp
pcDNA3.1+N-Myc	CMV	Ampicillin	N-MYC	5453 bp
pcDNA3.1+N-GST(a)	CMV	Ampicillin	N-GST-(TEV)	6092 bp
pcDNA3.1+N-GST(b)	CMV	Ampicillin	N-GST-(Thrombin)	6086 bp
pcDNA3.1+C-DYK	CMV	Ampicillin	c-DYK	5438 bp
pcDNA3.1+C-HA	CMV	Ampicillin	C-HA	5441 bp
pcDNA3.1+C-6His	CMV	Ampicillin	C-6*HIS	5432 bp
pcDNA3.1+C-Myc	CMV	Ampicillin	C-MYC	5444 bp

Mutant ORF Clones

Unlike ORF clones from other companies, GenEZ™ ORF clones can undergo mutagenesis service to create the mutant variants you need for your functional studies.

Mutagenesis services include:

- Point mutations
- Deletions
- Insertions
- 5' or 3' Truncations
- Protein domains
- 5' or 3' end sequence modifications

Mutant ORFs

- Mutant ORF clones from **$149**
- Just **3 days** additional production time

GenEZ™ ORF Clonesets

Clonesets are large collections of related ORF clones that contain the key genes of biological signaling pathways or even whole genomes (also called ORFeomes). Clonesets are useful for high-throughput screening assays in the arenas of drug discovery and systems biology.

Selected GenEZ™ Clonesets

Pathway	Species	# of genes
Methylation	*Homo sapiens*	10
ERK/MAPK targets	*Homo sapiens*	21
p38 MAPK signaling pathway	*Homo sapiens*	34
PI3K/AKT signalling	*Homo sapiens*	102
mTOR signaling pathway	*Homo sapiens*	61
G1/S DNA damage checkpoints	*Homo sapiens*	62
G2/M DNA damage checkpoint	*Homo sapiens*	9
DNA damage response	*Homo sapiens*	67
Signaling by GPCR	*Homo sapiens*	1068
IL-6 Signaling Pathway	*Homo sapiens*	99
TNF-alpha/NF-kB signaling pathway	*Homo sapiens*	196
TGF-beta signaling pathway	*Homo sapiens*	80
B cell receptor signaling pathway	*Homo sapiens*	72
ErbB signaling pathway	*Homo sapiens*	87
Angiogenesis	*Homo sapiens*	23
Wnt signaling pathway	*Homo sapiens*	139
Notch signaling pathway	*Homo sapiens*	48
Nuclear receptors	*Homo sapiens*	38

References

1. Meselson M. and Weigle J.J. (1961) Chromosome breakage accompanying genetic recombination in bacteriophage. *Proc. Natl. Acad. Sci. USA* **47**, 857-868.

2. Kellenberger G., Zichichi, M.L. and Weigle J.J. (1961) Exchange of DNA in the recombination of bacteriophage λ. *Proc. Natl. Acad. Sci. USA* **47**, 869-878.

3. Campbell A. (1962) Episomes. *Adv. Genet.* **11**, 101-145.

4. Luria S.E. and Human M.L. (1952) A nonhereditary host-induced variation of bacterial viruses. *J. Bacteriol.* **64**, 557-569.

5. Bertani G. and Weigle J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.* **65**, 113-121.

6. Meselson M. and Yuan R. (1968) DNA restriction enzyme from *E. coli.* Nature. 217: 1110−1114.

7. Kelly T.J. and Smith HO. (1970) A restriction enzyme from *Heomophilus influenzae.* II. Base sequence of the recognition site. *Journal of Molecular Biology.* **51**: 393−409.

8. Cozzarelli N.R., Melechen N.E., Jovin T.M., and Kornberg A. (1967) Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. *Biochem. Biophys. Res. Commun.* **28**, 578-586.

9. Gefter M.L., Becker A., and Hurwitz J. (1967) The enzymatic repair of DNA. I. Formation of circular lambda-DNA. Proc. *Natl. Acad. Sci. USA* **58**, 240-247.

10. Gellert M. (1967) Formation of covalent circles of lambda DNA by E. coli extracts. *Proc. Natl. Acad. Sci. USA* **57**, 148-155.

11. Olivera B.M. and Lehman I.R. (1967) Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. *Proc. Natl. Acad. Sci. USA* **57**, 1426-1433.

12. Weiss B. and Richardson C.C. (1967) Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *Proc. Natl. Acad. Sci. USA* **57**, 1021-1028.

13. Burgess-Brown N.A. *et al.* (2008) Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr. Purif.* **59** 94-102.

14. Michelson, A.M. and Todd, A.R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3: 5-internucleotidic linkage. *Journal of the Chemical Society* (Resumed) 2632–2638 (1955).

15. Agarwal, K.L. et al. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**: 27–34 (1970).

16. Marchand J.A. and Peccoud J. (2012) Building block synthesis using the polymerase chain assembly method. *Methods Mol. Biol.* **852**, 3-10.

17. Kelly J.R. *et al.* (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3**, 4.

18. Gibson D.G. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220.

19. Horton R.M. *et al.* (1993) Gene splicing by overlap extension. , Gene splicing by overlap extension. *Meth. Enzymol.* **217**, 270-279.

20. Stemmer W.P., Crameri A., Ha K.D., Brennan T.M. and Heyneker H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49-53.

21. Wooddell C.I. and Burgess R.R. (1996) Use of asymmetric PCR to generate long primers and single-stranded DNA for incorporating cross-linking analogs into specific sites in a DNA probe. *Genome Res.* **6**, 886-892.

22. Gao X., Yo P., Keith A., Ragan T.J., and Harris T.K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**, e143.

23. Tian J. *et al.* (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050-1054.

24. Wu G. *et al.* (2006) Simplified gene synthesis: a one-step approach to PCR-based gene construction. *J. Biotechnol.* **124**, 496-503.

25. Kong D.S., Carr P.A., Chen L., Zhang S. & Jacobson J.M. (2007) Parallel gene synthesis in a microfluidic device. *Nucleic Acids Res.* **35**, e61.

26. Yehezkel T.B., Linshiz G., Buaron H., Kaplan S., Shabi U., and Shapiro E. (2008) De novo DNA synthesis using single molecule PCR. *Nucleic Acids Res.* **36**, e107.

27. Huang M.C., Cheong, W.C., Ye, H. & Li M.-H. (2012) Top down real-time gene synthesis. *Methods Mol. Biol.* **852**, 23-34.

28. Eren M. & Swenson R. P. (1989) Chemicalsynthesis and expression of a synthetic gene for the flavodoxin from Clostridium MP. *J. Biol. Chem.* **264**, 14874-14879.

29. Mehta D.V., DiGate R.J., Banville D.L. & Guiles R.D. (1997) Optimized gene synthesis, high level expression, isotopic enrichment, and refolding of human interleukin-5. *Protein Expr. Purif.* **11**, 86-94.

30. Au L.C., Yang F.Y., Yang W.J., Lo S.H., and Kao C.F. (1998) Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli. Biochem. Biophys. Res. Commun.* **248**, 200-203.

31. Li M.Z. and Elledge S.J. (2012) SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51-59.

32. Ho-Shing O., Lau K.H., Vernon W., Eckdahl T.T. & Campbell A.M. (2012) Assembly of standardized DNA parts using BioBrick ends in *E. coli. Methods Mol. Biol.* **852**, 61-76.

33. Young L. and Dong Q. (2004) Two-step total gene synthesis method. *Nucleic Acids Res.* **32**, e59.

34. Hughes R.A., Miklos A.E., and Ellington A.D. (2011) Gene synthesis: methods and applications. *Meth. Enzymol.* **498**, 277-309.

35. Ma S., Saaem I. and Tian J. (2012) Error correction in gene synthesis technology. *Trends Biotechnol.* **30**, 147-154

36. Wilhelm M. et al. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587.

37. Kim M-S. (2014) A draft map of the human proteome. *Nature* **509**, 575–581.