wjpmr, 2020,6(8), 257-269

SJIF Impact Factor: 5.922

WORLD JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.wjpmr.com

Research Article

ISSN 2455-3301

WJPMR

GENETIC ENGINEERING – THE FOUNDATION OF CUTTING-EDGE EXTRAMURAL RESEARCH

***[1]Kushal Nandi, [1]Subarna Mahanti, [1]Dr. Dhrubo Jyoti Sen, [2]Dipra Dastider and [3]Dr. Sudip Kumar Mandal**

[1]Department of Pharmaceutical Chemistry, School of Pharmacy, Techno India University, Salt Lake City, Sector-V, EM-4, Kolkata-700091, West Bengal, India.

[2]Department of Pharmaceutical Technology, Brainware University, 398-Ramkrishnapur Road, Barasat, Kolkata-700125, West Bengal, India.

[3]Dr. B. C. Roy College of Pharmacy and A.H.S, Dr. Meghnad Saha Sarani, Bidhan Nagar, Durgapur-713206, West Bengal, India.

***Corresponding Author: Kushal Nandi**

Department of Pharmaceutical Chemistry, School of Pharmacy, Techno India University, Salt Lake City, Sector-V, EM-4, Kolkata-700091, West Bengal, India.

Article Received on 03/06/2020 **Article Revised on 04/06/2020** **Article Accepted on 14/07/2020**

ABSTRACT

A molecular genetic technique use for the direct manipulation, alteration or modification of genes or genome of organisms in order to manipulate the phenotypes is called genetic engineering. Genetic engineering, also called Genetic modification or Genetic manipulation, is the direct manipulation of an organism's genes using biotechnology. It is a set of technologies used to change the genetic makeup of cells, including the transfer of genes within and across species boundaries to produce improved or novel organisms. New DNA is obtained by either isolating and copying the genetic material of interest using recombinant DNA methods or by artificially synthesizing the DNA. A construct is usually created and used to insert this DNA into the host organism. The first recombinant DNA molecule was made by Paul Berg in 1972 by combining DNA from the monkey virus SV40 with the lambda virus. As well as inserting genes, the process can be used to remove, or "knock out", genes. The new DNA can be inserted randomly, or targeted to a specific part of the genome. Humans have altered the genomes of species for thousands of years through selective breeding, or artificial selection as contrasted with natural selection. More recently, mutation breeding has used exposure to chemicals or radiation to produce a high frequency of random mutations, for selective breeding purposes. Genetic engineering as the direct manipulation of DNA by humans outside breeding and mutations has only existed since the 1970s. The term "genetic engineering" was first coined by Jack Williamson in his science fiction novel Dragon's Island, published in 1951 – one year before DNA's role in heredity was confirmed by Alfred Hershey and Martha Chase, and two years before James Watson and Francis Crick showed that the DNA molecule has a double-helix structure – though the general concept of direct genetic manipulation was explored in rudimentary form in Stanley G. Weinbaum's 1936 science fiction story Proteus Island. In 1972, Paul Berg created the first recombinant DNA molecules by combining DNA from the monkey virus SV40 with that of the lambda virus. In 1973 Herbert Boyer and Stanley Cohen created the first transgenic organism by inserting antibiotic resistance genes into the plasmid of an Escherichia coli bacterium. A year later Rudolf Jaenisch created a transgenic mouse by introducing foreign DNA into its embryo, making it the world's first transgenic animal These achievements led to concerns in the scientific community about potential risks from genetic engineering, which were first discussed in depth at the Asilomar Conference in 1975. One of the main recommendations from this meeting was that government oversight of recombinant DNA research should be established until the technology was deemed safe.

KEYWORDS: Genetics, Genome, GMO, PCR, Gel electrophoresis, DNA mapping, DNA fragmentation, Gene coding, Genetic coding, Gene sequencing.

INTRODUCTION

An organism that is generated through genetic engineering is considered to be genetically modified (GM) and the resulting entity is a genetically modified organism (GMO). The first GMO was a bacterium generated by Herbert Boyer and Stanley Cohen in 1973. Rudolf Jaenisch created the first GM animal when he inserted foreign DNA into a mouse in 1974. The first company to focus on genetic engineering, Genentech, was founded in 1976 and started the production of human proteins. Genetically engineered human insulin was produced in 1978 and insulin-producing bacteria were commercialised in 1982. Genetic engineering is a process that alters the genetic structure of an organism by either removing or introducing DNA. Unlike traditional animal and plant breeding, which involves doing multiple crosses and then selecting for the organism with the desired phenotype,

genetic engineering takes the gene directly from one organism and delivers it to the other. This is much faster, can be used to insert any genes from any organism (even ones from different domains) and prevents other undesirable genes from also being added.

Genetic engineering could potentially fix severe genetic disorders in humans by replacing the defective gene with a functioning one. It is an important tool in research that allows the function of specific genes to be studied. Drugs, vaccines and other products have been harvested from organisms engineered to produce them. Crops have been developed that aid food security by increasing yield, nutritional value and tolerance to environmental stresses.[1]

Figure-1: Genome.

The DNA can be introduced directly into the host organism or into a cell that is then fused or hybridised with the host. This relies on recombinant nucleic acid techniques to form new combinations of heritable genetic material followed by the incorporation of that material either indirectly through a vector system or directly through micro-injection, macro-injection or micro-encapsulation.

Genetic engineering does not normally include traditional breeding, *in-vitro* fertilization, induction of polyploidy, mutagenesis and cell fusion techniques that do not use recombinant nucleic acids or a genetically modified organism in the process. However, some broad definitions of genetic engineering include selective breeding. Cloning and stem cell research, although not considered genetic engineering, are closely related and genetic engineering can be used within them. Synthetic biology is an emerging discipline that takes genetic engineering a step further by introducing artificially synthesised material into an organism. Such synthetic DNA as Artificially Expanded Genetic Information System and Hachimoji DNA is made in this new field.

Plants, animals or microorganisms that have been changed through genetic engineering are termed genetically modified organisms or GMOs. If genetic material from another species is added to the host, the resulting organism is called transgenic. If genetic material from the same species or a species that can naturally breed with the host is used the resulting organism is called cisgenic. If genetic engineering is used to remove genetic material from the target organism the resulting organism is termed a knockout organism. In Europe genetic modification is synonymous with genetic engineering while within the United States of America and Canada genetic modification can also be used to refer to more conventional breeding methods.[2]

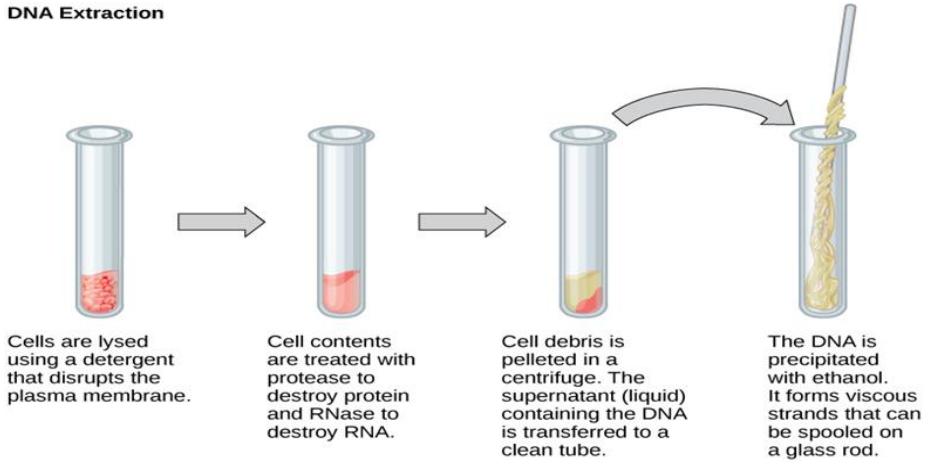

Figure-2: Extraction of DNA.

Isolation of Nucleic Acids

To study or manipulate nucleic acids, the DNA must first be extracted from cells. Various techniques are used to extract different types of DNA (Figure-2). Most nucleic acid extraction techniques involve steps to break open the cell, and then the use of enzymatic reactions to destroy all undesired macromolecules. Cells are broken open using a detergent solution containing buffering compounds. To prevent degradation and contamination, macromolecules such as proteins and RNA are inactivated using enzymes. The DNA is then brought out of solution using alcohol. The resulting DNA, because it is made up of long polymers, forms a gelatinous mass. RNA is studied to understand gene expression patterns in cells. RNA is naturally very unstable because enzymes that break down RNA are commonly present in nature. Some are even secreted by our own skin and are very difficult to inactivate. Similar to DNA extraction, RNA extraction involves the use of various buffers and enzymes to inactivate other macromolecules and preserve only the RNA.

Gel Electrophoresis

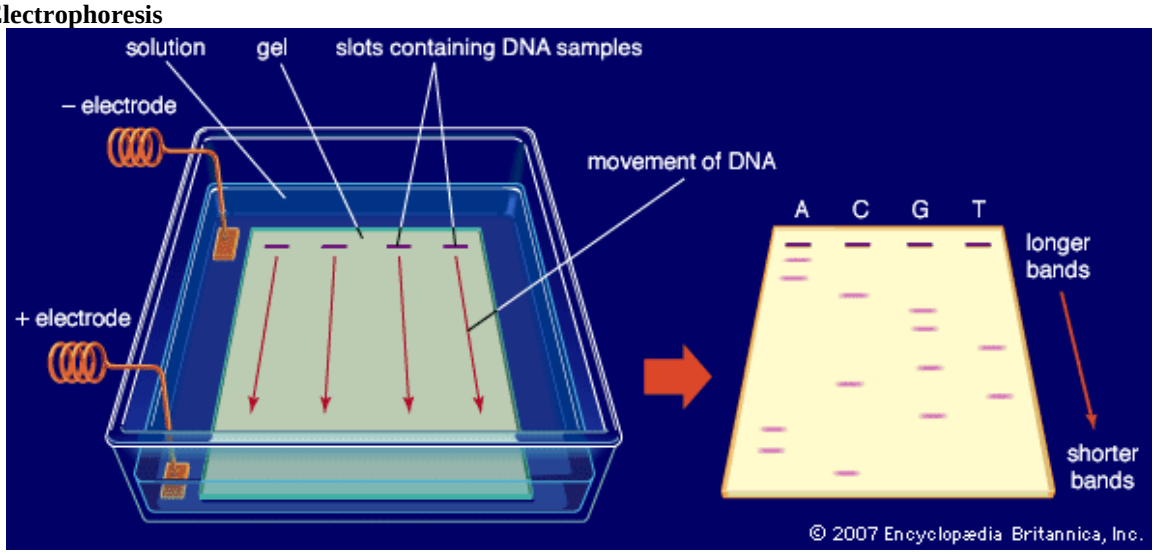

Figure-3: Gel electrophoresis.

Because nucleic acids are negatively charged ions at neutral or alkaline pH in an aqueous environment, they can be moved by an electric field. Gel electrophoresis is a technique used to separate charged molecules on the basis of size and charge. The nucleic acids can be separated as whole chromosomes or as fragments. The nucleic acids are loaded into a slot at one end of a gel matrix, an electric current is applied, and negatively charged molecules are pulled toward the opposite end of the gel (the end with the positive electrode). Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. The nucleic acids in a gel matrix are invisible until they are stained with a compound that allows them to be seen, such as a dye. Distinct fragments of nucleic acids appear as bands at specific distances from the top of the gel (the negative electrode end) that are based on their size (Figure-3). A mixture of many fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.[3]

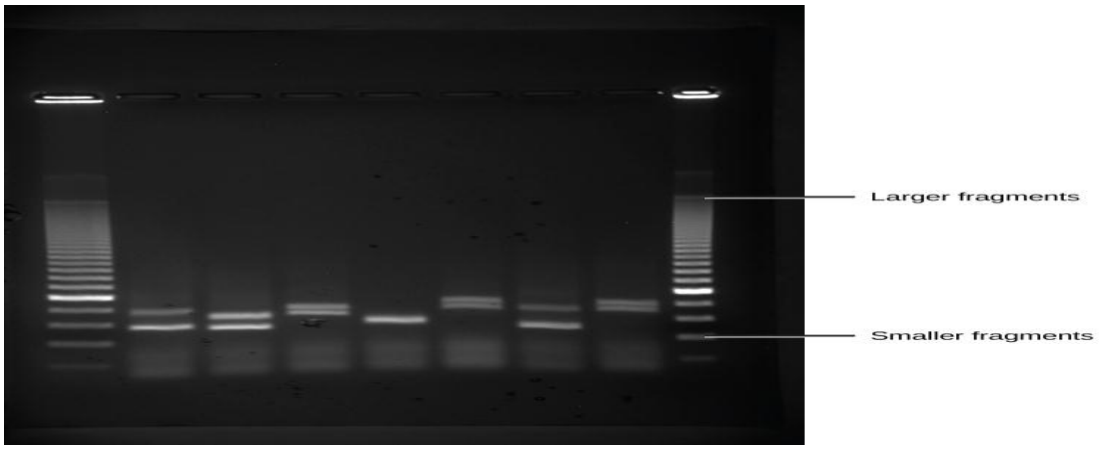

Figure-4: DNA fragments from six samples run on a gel, stained with a fluorescent dye and viewed under UV light.

Polymerase Chain Reaction

DNA analysis often requires focusing on one or more specific regions of the genome. It also frequently involves situations in which only one or a few copies of a DNA molecule are available for further analysis. These amounts are insufficient for most procedures, such as gel electrophoresis. Polymerase chain reaction (PCR) is a technique used to rapidly increase the number of copies of specific regions of DNA for further analyses (Figure-5). PCR uses a special form of DNA polymerase, the enzyme that replicates DNA, and other short nucleotide sequences called primers that base pair to a specific portion of the DNA being replicated. PCR is used for many purposes in laboratories. These include: 1) the identification of the owner of a DNA sample left at a crime scene; 2) paternity analysis; 3) the comparison of small amounts of ancient DNA with modern organisms; and 4) determining the sequence of nucleotides in a specific region.

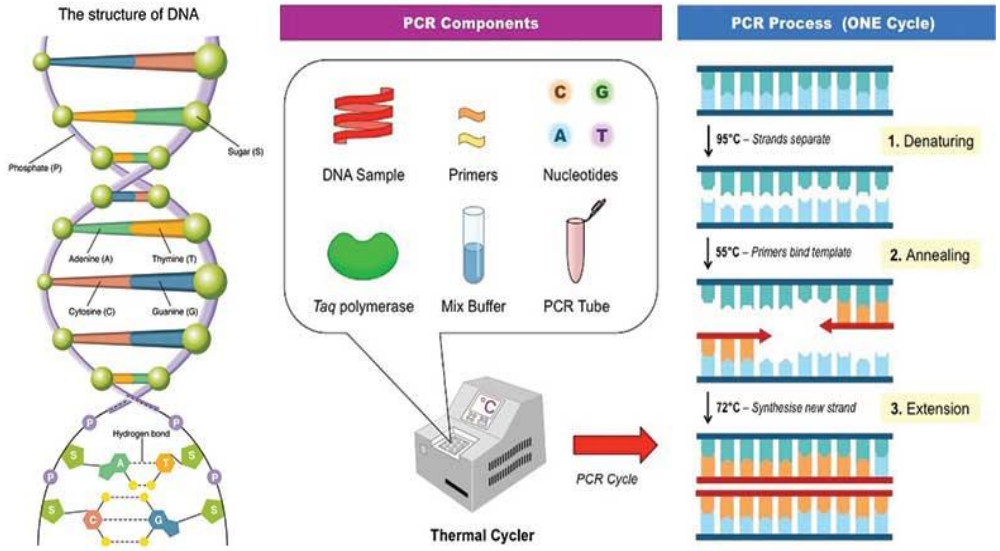

Figure-5: Polymerase chain reaction.

PCR, is used to produce many copies of a specific sequence of DNA using a special form of DNA polymerase. Figure-5 showing PCR in 4 steps. First, the double strand of DNA is denatured at 95 degrees Celsius to separate the strands. The 2 strands are then annealed at approximately 50 degrees Celsius using primers. DNA polymerase then extends the new strands at 72 degrees Celsius. The fourth step shows that this procedure takes place many times, resulting in an increase in copies of the original DNA.[4]

Gene Coding

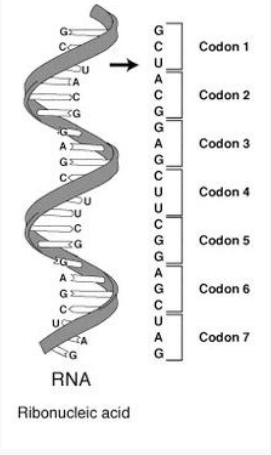

Figure-6: Gene coding.

A series of codons in part of a messenger RNA (mRNA) molecule. Each codon consists of three nucleotides, usually corresponding to a single amino acid. The nucleotides are abbreviated with the letters A, U, G and C. This is mRNA, which uses U (uracil). DNA uses T (thymine) instead. This mRNA molecule will instruct a ribosome to synthesize a protein according to this code (Figure-6).

The genetic code is the set of rules used by living cells to translate information encoded within genetic material (DNA or mRNA sequences of nucleotide triplets, or codons) into proteins. Translation is accomplished by the ribosome, which links amino acids in an order specified by messenger RNA (mRNA), using transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA three nucleotides at a time. The genetic code is highly similar among all organisms and can be expressed in a simple table with 64 entries. The code defines how codons specify which amino acid will be added next during protein synthesis. With some exceptions, a three-nucleotide codon in a nucleic acid sequence specifies a single amino acid. The vast majority of genes are encoded with a single scheme (see the RNA codon table). That scheme is often referred to as the canonical or standard genetic code, or simply the genetic code, though variant codes (such as in human mitochondria) exist.

While the "genetic code" is what determines a protein's amino acid sequence, other genomic regions determine when and where these proteins are produced according to various "gene regulatory codes".

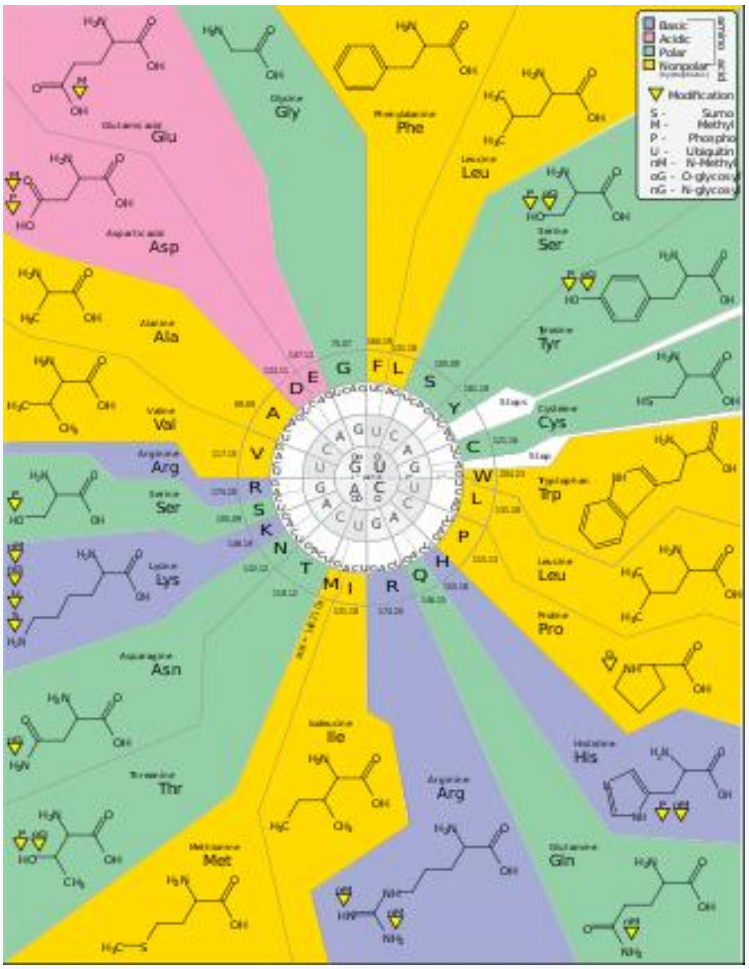

Figure-7: The Genetic Codes.

Efforts to understand how proteins are encoded began after DNA's structure was discovered in 1953. George Gamow postulated that sets of three bases must be employed to encode the 20 standard amino acids used by living cells to build proteins, which would allow a maximum of $4^3 = 64$ amino acids.[5]

Codons

The Crick, Brenner, Barnett and Watts-Tobin experiment first demonstrated that codons consist of three DNA bases. Marshall Nirenberg and Heinrich J. Matthaei were the first to reveal the nature of a codon in 1961. They used a cell-free system to translate a poly-uracil RNA sequence (i.e., UUUUU...) and discovered that the polypeptide that they had synthesized consisted of only the amino acid phenylalanine. They thereby deduced that the codon UUU specified the amino acid phenylalanine. This was followed by experiments in Severo Ochoa's laboratory that demonstrated that the poly-adenine RNA sequence (AAAAA...) coded for the polypeptide poly-lysine and that the poly-cytosine RNA sequence (CCCCC...) coded for the polypeptide poly-proline. Therefore, the codon AAA specified the amino acid lysine, and the codon CCC specified the amino acid proline. Using various copolymers most of the remaining codons were then determined (Figure-7).

Subsequent work by Har Gobind Khorana identified the rest of the genetic code. Shortly thereafter, Robert W. Holley determined the structure of transfer RNA (tRNA), the adapter molecule that facilitates the process of translating RNA into protein. This work was based upon Ochoa's earlier studies, yielding the latter the Nobel Prize in Physiology or Medicine in 1959 for work on the enzymology of RNA synthesis. Extending this work, Nirenberg and Philip Leder revealed the code's triplet nature and deciphered its codons. In these experiments, various combinations of mRNA were passed through a filter that contained ribosomes, the components of cells that translate RNA into protein. Unique triplets promoted the binding of specific tRNAs to the ribosome. Leder and Nirenberg were able to determine the sequences of 54 out of 64 codons in their experiments. Khorana, Holley and Nirenberg received the 1968 Nobel for their work. The three stop codons were named by discoverers Richard Epstein and Charles Steinberg. "Amber" was named after their friend Harris Bernstein, whose last name means "amber" in German. The other two stop

codons were named "ochre" and "opal" in order to keep the "color names" theme.

Expanded genetic codes (synthetic biology)

In a broad academic audience, the concept of the evolution of the genetic code from the original and ambiguous genetic code to a well-defined ("frozen") code with the repertoire of 20 (+2) canonical amino acids is widely accepted. However, there are different opinions, concepts, approaches and ideas, which is the best way to change it experimentally. Even models are proposed that predict "entry points" for synthetic amino acid invasion of the genetic code. Since 2001, 40 non-natural amino acids have been added into protein by creating a unique codon (recoding) and a corresponding transfer-RNA:aminoacyl – tRNA-synthetase pair to encode it with diverse physicochemical and biological properties in order to be used as a tool to exploring protein structure and function or to create novel or enhanced proteins. H. Murakami and M. Sisido extended some codons to have four and five bases. Steven A. Benner constructed a functional 65th (*in-vivo*) codon. In 2015 N. Budisa, D. Söll and co-workers reported the full substitution of all 20,899 tryptophan residues (UGG codons) with unnatural thienopyrrole-alanine in the genetic code of the bacterium *Escherichia coli*. In 2016 the first stable semisynthetic organism was created. It was a (single cell) bacterium with two synthetic bases (called X and Y). The bases survived cell division.[6]

In 2017, researchers in South Korea reported that they had engineered a mouse with an extended genetic code that can produce proteins with unnatural amino acids. In May 2019, researchers, in a milestone effort, reported the creation of a new synthetic (possibly artificial) form of viable life, a variant of the bacteria *Escherichia coli*, by reducing the natural number of 64 codons in the bacterial genome to 59 codons instead, in order to encode 20 amino acids.

Features

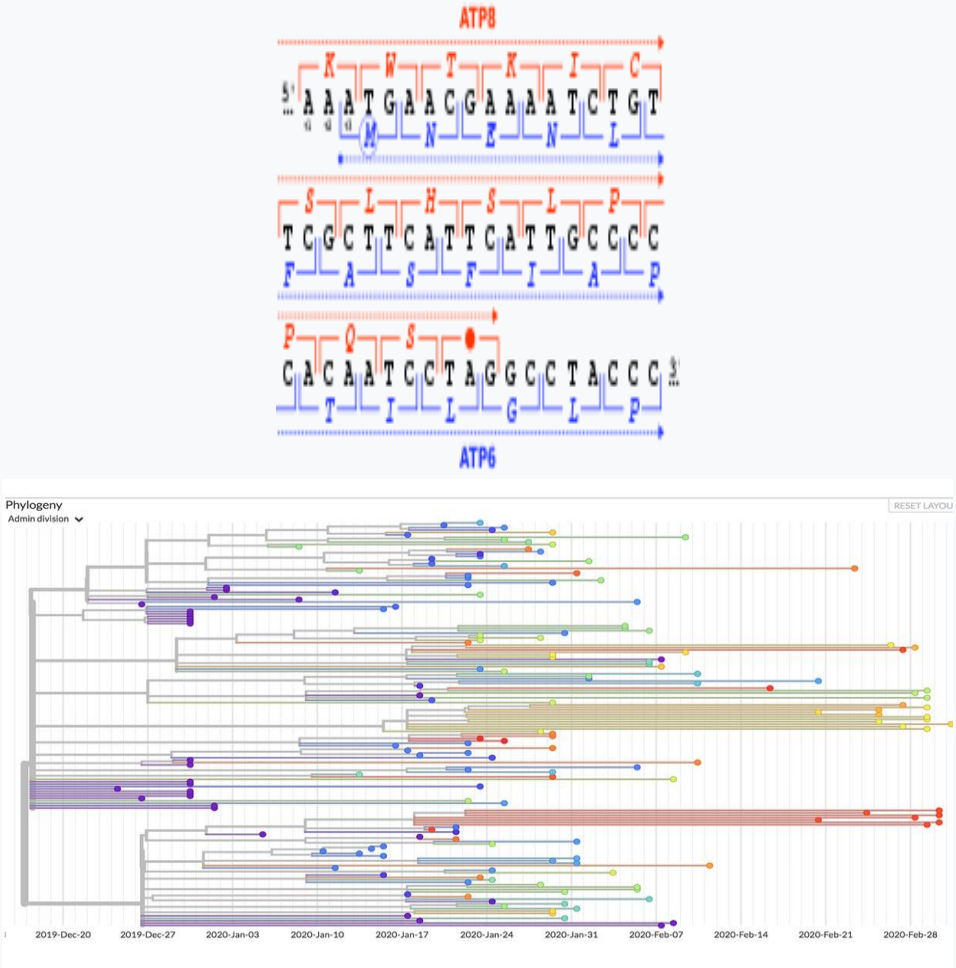

Figure-8: Genetic sequence & Software Genetic Sequencing Track.

Reading frames in the DNA sequence of a region of the human mitochondrial genome coding for the genes MT-ATP8 and MT-ATP6 (in black: positions 8,525 to 8,580 in the sequence accession NC_012920). There are three possible reading frames in the 5' → 3' forward direction, starting on the first (+1), second (+2) and third position (+3). For each codon (square brackets), the amino acid is given by the vertebrate mitochondrial code, either in the

+1 frame for MT-ATP8 (in red) or in the +3 frame for MT-ATP6 (in blue). The MT-ATP8 genes terminates with the TAG stop codon (red dot) in the +1 frame. The MT-ATP6 gene starts with the ATG codon (blue circle for the M amino acid) in the +3 frame (Figure-8).

Reading frame

A reading frame is defined by the initial triplet of nucleotides from which translation starts. It sets the frame for a run of successive, non-overlapping codons, which is known as an "open reading frame" (ORF). For example, the string 5'-AAATGAACG-3', if read from the first position, contains the codons AAA, TGA, and ACG; if read from the second position, it contains the codons AAT and GAA; and if read from the third position, it contains the codons ATG and AAC. Every sequence can, thus, be read in its 5' → 3' direction in three reading frames, each producing a possibly distinct amino acid sequence: in the given example, Lys (K)-Trp (W)-Thr (T), Asn (N)-Glu (E), or Met (M)-Asn (N), respectively (when translating with the vertebrate mitochondrial code). When DNA is double-stranded, six possible reading frames are defined, three in the forward orientation on one strand and three reverse on the opposite strand. Protein-coding frames are defined by a start codon, usually the first AUG (ATG) codon in the RNA (DNA) sequence. In eukaryotes, ORFs in exons are often interrupted by introns.

Start/stop codons

Translation starts with a chain-initiation codon or start codon. The start codon alone is not sufficient to begin the process. Nearby sequences such as the Shine-Dalgarno sequence in E. coli and initiation factors are also required to start translation. The most common start codon is AUG, which is read as methionine or, in bacteria, as formylmethionine. Alternative start codons depending on the organism include "GUG" or "UUG"; these codons normally represent valine and leucine, respectively, but as start codons they are translated as methionine or formylmethionine. The three stop codons have names: UAG is amber, UGA is opal (sometimes also called umber), and UAA is ochre. Stop codons are also called "termination" or "nonsense" codons. They signal release of the nascent polypeptide from the ribosome because no cognate tRNA has anticodons complementary to these stop signals, allowing a release factor to bind to the ribosome instead.

Degeneracy

Grouping of codons by amino acid residue molar volume and hydropathicity. A more detailed version is available.
Degeneracy is the redundancy of the genetic code. This term was given by Bernfield and Nirenberg. The genetic code has redundancy but no ambiguity (see the codon tables below for the full correlation). For example, although codons GAA and GAG both specify glutamic acid (redundancy), neither specifies another amino acid (no ambiguity). The codons encoding one amino acid may differ in any of their three positions. For example, the amino acid leucine is specified by YUR or CUN (UUA, UUG, CUU, CUC, CUA, or CUG) codons (difference in the first or third position indicated using IUPAC notation), while the amino acid serine is specified by UCN or AGY (UCA, UCG, UCC, UCU, AGU, or AGC) codons (difference in the first, second, or third position). A practical consequence of redundancy is that errors in the third position of the triplet codon cause only a silent mutation or an error that would not affect the protein because the hydrophilicity or hydrophobicity is maintained by equivalent substitution of amino acids; for example, a codon of NUN (where N = any nucleotide) tends to code for hydrophobic amino acids. NCN yields amino acid residues that are small in size and moderate in hydropathicity; NAN encodes average size hydrophilic residues. The genetic code is so well-structured for hydropathicity that a mathematical analysis (Singular Value Decomposition) of 12 variables [(4 nucleotides × 3 positions)] yields a remarkable correlation (C = 0.95) for predicting the hydropathicity of the encoded amino acid directly from the triplet nucleotide sequence, without translation. Note in the table, below, eight amino acids are not affected at all by mutations at the third position of the codon, whereas in the figure above, a mutation at the second position is likely to cause a radical change in the physicochemical properties of the encoded amino acid. Nevertheless, changes in the first position of the codons are more important than changes in the second position on a global scale. The reason may be that charge reversal (from a positive to a negative charge or vice versa) can only occur upon mutations in the first position, but never upon changes in the second position of a codon. Such charge reversal may have dramatic consequences for the structure or function of a protein. This aspect may have been largely underestimated by previous studies.[7]

Codon usage bias

The frequency of codons, also known as codon usage bias, can vary from species to species with functional implications for the control of translation. The following codon usage table is for the human genome. The codon AUG both codes for methionine and serves as an initiation site: the first AUG in an mRNA's coding region is where translation into protein begins. The other start codons listed by GenBank are rare in eukaryotes and generally codes for Met/fMet. The historical basis for designating the stop codons as amber, ochre and opal is described in an autobiography by Sydney Brenner and in a historical article by Bob Edgar.

Non-standard amino acids

In some proteins, non-standard amino acids are substituted for standard stop codons, depending on associated signal sequences in the messenger RNA. For example, UGA can code for selenocysteine and UAG can code for pyrrolysine. Selenocysteine became to be seen as the 21st amino acid, and pyrrolysine as the 22nd. Unlike selenocysteine, pyrrolysine-encoded UAG is translated with the participation of a

dedicated aminoacyl-tRNA synthetase. Both selenocysteine and pyrrolysine may be present in the same organism. Although the genetic code is normally fixed in an organism, the achaeal prokaryote Acetohalobium arabaticum can expand its genetic code from 20 to 21 amino acids (by including pyrrolysine) under different conditions of growth.

Variations

Figure-9: Genetic code.

Genetic code logo of the *Globobulimina pseudospinescens* mitochondrial genome. The logo shows the 64 codons from left to right, predicted alternatives in red (relative to the standard genetic code). Red line: stop codons. The height of each amino acid in the stack shows how often it is aligned to the codon in homologous protein domains. The stack height indicates the support for the prediction.[8]

Variations on the standard code were predicted in the 1970s. The first was discovered in 1979, by researchers studying human mitochondrial genes. Many slight variants were discovered thereafter, including various alternative mitochondrial codes. These minor variants for example involve translation of the codon UGA as tryptophan in Mycoplasma species, and translation of CUG as a serine rather than leucine in yeasts of the "CTG clade" (such as *Candida albicans*). Because viruses must use the same genetic code as their hosts, modifications to the standard genetic code could interfere with viral protein synthesis or functioning. However, viruses such as totiviruses have adapted to the host's genetic code modification. In bacteria and archaea, GUG and UUG are common start codons. In rare cases, certain proteins may use alternative start codons. Surprisingly, variations in the interpretation of the genetic code exist also in human nuclear-encoded genes: In 2016, researchers studying the translation of malate dehydrogenase found that in about 4% of the mRNAs encoding this enzyme the stop codon is naturally used to encode the amino acids tryptophan and arginine. This type of recoding is induced by a high-readthrough stop codon context and it is referred to as functional translational readthrough.

Variant genetic codes used by an organism can be inferred by identifying highly conserved genes encoded in that genome, and comparing its codon usage to the amino acids in homologous proteins of other organisms. For example, the program FACIL infers a genetic code by searching which amino acids in homologous protein domains are most often aligned to every codon. The resulting amino acid probabilities for each codon are displayed in a genetic code logo, that also shows the support for a stop codon.

Despite these differences, all known naturally occurring codes are very similar. The coding mechanism is the same for all organisms: three-base codons, tRNA, ribosomes, single direction reading and translating single codons into single amino acids.

Origin

The genetic code is a key part of the history of life, according to one version of which self-replicating RNA molecules preceded life as we know it. This is the RNA world hypothesis. Under this hypothesis, any model for the emergence of the genetic code is intimately related to a model of the transfer from ribozymes (RNA enzymes) to proteins as the principal enzymes in cells. In line with the RNA world hypothesis, transfer RNA molecules appear to have evolved before modern aminoacyl-tRNA synthetases, so the latter cannot be part of the explanation of its patterns. A hypothetical randomly evolved genetic code further motivates a biochemical or evolutionary model for its origin. If amino acids were randomly assigned to triplet codons, there would be 1.5×10^{84} possible genetic codes. This number is found by calculating the number of ways that 21 items (20 amino acids plus one stop) can be placed in 64 bins, wherein each item is used at least once. However, the distribution of codon assignments in the genetic code is nonrandom. In particular, the genetic code clusters certain amino acid assignments. Amino acids that share the same biosynthetic pathway tend to have the same first base in their codons. This could be an evolutionary relic of an early, simpler genetic code with fewer amino acids that later evolved to code a larger set of amino acids. It could also reflect steric and chemical properties that had another effect on the codon during its evolution. Amino acids with similar physical properties also tend to have similar codons, reducing the problems caused by point mutations and mistranslations. Given the non-random genetic triplet coding scheme, a tenable hypothesis for the origin of genetic code could address multiple aspects of the codon table, such as absence of codons for D-amino acids, secondary codon patterns for some amino acids, confinement of synonymous positions to third position, the small set of only 20 amino acids (instead of a number approaching 64), and the relation of stop codon patterns to amino acid coding patterns. Three main hypotheses address the origin of the genetic code. Many models belong to one of them or to a hybrid.[9]

Random freeze: the genetic code was randomly created. For example, early tRNA-like ribozymes may have had different affinities for amino acids, with codons emerging from another part of the ribozyme that

exhibited random variability. Once enough peptides were coded for, any major random change in the genetic code would have been lethal; hence it became "frozen". Stereochemical affinity: the genetic code is a result of a high affinity between each amino acid and its codon or anti-codon; the latter option implies that pre-tRNA molecules matched their corresponding amino acids by this affinity. Later during evolution, this matching was gradually replaced with matching by aminoacyl-tRNA synthetases.

Optimality: the genetic code continued to evolve after its initial creation, so that the current code maximizes some fitness function, usually some kind of error minimization. Hypotheses have addressed a variety of scenarios: Chemical principles govern specific RNA interaction with amino acids. Experiments with aptamers showed that some amino acids have a selective chemical affinity for their codons. Experiments showed that of 8 amino acids tested, 6 show some RNA triplet-amino acid association. Biosynthetic expansion. The genetic code grew from a simpler earlier code through a process of "biosynthetic expansion". Primordial life "discovered" new amino acids (for example, as by-products of metabolism) and later incorporated some of these into the machinery of genetic coding. Although much circumstantial evidence has been found to suggest that fewer amino acid types were used in the past, precise and detailed hypotheses about which amino acids entered the code in what order are controversial. Natural selection has led to codon assignments of the genetic code that minimize the effects of mutations. A recent hypothesis suggests that the triplet code was derived from codes that used longer than triplet codons (such as quadruplet codons). Longer than triplet decoding would increase codon redundancy and would be more error resistant. This feature could allow accurate decoding absent complex translational machinery such as the ribosome, such as before cells began making ribosomes.

Information channels: Information -theoretic approaches model the process of translating the genetic code into corresponding amino acids as an error-prone information channel. The inherent noise (that is, the error) in the channel poses the organism with a fundamental question: how can a genetic code be constructed to withstand noise while accurately and efficiently translating information? These "rate-distortion" models suggest that the genetic code originated as a result of the interplay of the three conflicting evolutionary forces: the needs for diverse amino acids, for error-tolerance and for minimal resource cost. The code emerges at a transition when the mapping of codons to amino acids becomes nonrandom. The code's emergence is governed by the topology defined by the probable errors and is related to the map coloring problem.[10]

Game theory: Models based on signaling games combine elements of game theory, natural selection and information channels. Such models have been used to suggest that the first polypeptides were likely short and had non-enzymatic function. Game theoretic models suggested that the organization of RNA strings into cells may have been necessary to prevent "deceptive" use of the genetic code, i.e. preventing the ancient equivalent of viruses from overwhelming the RNA world.[11]

Stop codons: Codons for translational stops are also an interesting aspect to the problem of the origin of the genetic code. As an example, for addressing stop codon evolution, it has been suggested that the stop codons are such that they are most likely to terminate translation early in the case of a frame shift error. In contrast, some stereochemical molecular models explain the origin of stop codons as "unassignable".

It has been claimed that the genetic code contains patterns and arithmetic coincidences that are very unlikely by chance and that would not arise through evolution. The authors of this claim contend that this is basically a message indicating that life on Earth was seeded by a previous civilization, similar to panspermia.[12]

CONCLUSION

Playing with the embryo or fetus is against the natural law, people strongly believe in it, thus genetically modified food and plant products are always becoming a centre of controversy. However, using genetic engineering tools such as gene therapy and gene transfer technique, inherited disorders and cancer like lethal diseases can be prevented. Positive use of genetic engineering techniques can change the fate of mankind. Genetic engineering is a technique using which the genetic composition of an organism can be altered." The technique is often known as genetic manipulation, genetic modification or genetic alterations, broadly categorized as genetic engineering. By inserting a gene of interest or by deleting the unwanted DNA sequences from the genome, altered gene or DNA is constructed called as recombinant DNA which is transferred into the host genome using the vectors. The first recombinant DNA was constructed by Paul Berg in 1972. Using the genetic engineering technique genetically modified organism can be constructed which are economically very important. It is employed for the production of improved plant species, therapeutic drugs or proteins, prevention of inherited genetic disorders and construction of a genetically modified organism.

In the present article, we will discuss on genetic engineering and its applications. The content of the article is [What is genetic engineering, Definition, History, Types, Process, Application of genetic engineering, Limitations of genetic engineering, Conclusion].

Humans are manipulating the genetic material of many organisms since long. Using selective breeding and cross-hybridization, economically important plant species were created by humans. The purpose of developing the genetic engineering or genetic manipulating technique is to produce organisms or phenotypes which are useful to us. Genetic engineering techniques are used for [Construction of genetically modified plant species, Abiotic and biotic stress resistant plant species, Economically important plant species, Commercially valuable organism, For the production of therapeutic drugs]

Prevention of genetic abnormalities. "In genetic engineering, two different cell's DNA are combined and inserted into the host genome via vector." Important components in any of the gene manipulation experiments are:
Gene of interest: A DNA sequence which we want to insert in our target cells.

Vector: using the plasmid DNA like vectors the gene of interest is inserted into the host genome. Vectors are kind of vehicles which transfer the genetic material.

Target cells: target cells are the population of cells whose genome we wish to manipulate or change. The general process of gene therapy.

History of genetic engineering:
The term genetic engineering was first used by the science-fiction novelist, not by any scientist. In the year, 1951, Jack Williamson used the term "genetic engineering" for the first time in his novel "Dragon's island". Soon after that, the molecular structure of the DNA was discovered by Watson and Crick, although the genetic experiments were popular since the time of Mendel. The first recombinant DNA was constructed by Paul Berg in 1972. In the same year, Herbert Boyer and Stanley Cohen performed gene transfer experiments. In 1974, Rudolf Jaenisch had created genetically modified mice the first time in the history of genetics. After the success of Rudolf, the genetically modified or genetically engineered tobacco plant species was developed in 1976. During this period (between 1960 to 1990) restriction digestion, ligation and PCR like techniques were discovered which gave wings to the genetic engineering technology. Type of genetic engineering techniques:

Recombinant DNA- A recombinant DNA technology is a type of genetic engineering technique in which an artificial DNA molecule is constructed using some of the physical methods. For that, the gene of interest is inserted into the plasmid vector and used for gene transfer experiments.

Gene delivering- Gene delivering technique is employed for the insertion of a gene of interest into the host genome. Electrophoration, solicitation and viral vector-mediated gene transfer, liposome-mediated gene transfer, transposon-mediated gene transfer are some of the methods used for that.

Gene editing- A gene-editing technique is used to edit the genome in which undesired DNA sequence is removed or a new gene can be inserted into the host genome. CRISPR-CAS9, TALEN and ZFN are some known gene-editing tools used in gene therapy experiments.

Process of genetic engineering: The genetic engineering technique is used for many different purposes thus we must have to decide first the purpose of the experiment. The entire process of genetic engineering can be divided into 5 broader steps: Selecting and isolating the candidate gene, Selection and construction of plasmid, Gene transformation.

Insertion of DNA into the host genome, Confirmation of insert, Selecting and isolating the candidate gene:
The gene must contain a sequence of DNA which we want to study. The candidate gene does not have repeated DNA sequences and higher GC content. In addition to this, the gene of interest must not be too long-only a few kb genes can be successfully inserted. Longer the gene higher the chance of failure. The candidate gene must have a start and stop codon in it. Now, the gene of interest can be isolated from the rest of the DNA using either restriction digestion or polymerase chain reaction. The restriction endonucleases are the bacterial enzyme having the power to digest DNA sequence at a specific location. Using a specific type of restriction endonuclease we can cut and isolated our gene of interest. The restriction digestion method is explained in our previous article: What is restriction digestion?

In the polymerase chain reaction, using the information of the sequence of a gene, the gene of interest or the candidate gene is amplified in the thermocycler. The machine, using the polymerase chain reaction amplifies millions of copies of a gene of our interest. Through the process of agarose gel electrophoresis, the amplified gene can be isolated.

If the gene of interest is well studied, previously, then the information of a gene is accessible in the genetic library and we can use it for the artificial synthesis of a gene of our interest. (using the genetic library information, the gene can also be artificially synthesized). In the next step, perform DNA purification, if required, now our DNA is ready for the construction of plasmid. Selection and construction of plasmid: Selecting plasmid for the genetic engineering experiment is one of the crucial steps in the entire experiment. Before selection the plasmid, we must understand why the plasmid is used in the gene transfer experiments. The plasmid DNA is a circular, double-stranded cytoplasmic DNA of the bacteria that replicate independently. Scientists are using it as a vehicle for transferring the gene of interest to the target

location in the genome. It can efficiently transfer the gene of interest at the target location.

Preparation of plasmid: Select the plasmid which suits your experiment. The plasmid must have the origin of replication, promoter region, antibiotic resistance gene and other important sequences. Using the restriction digestion method, an insertion site is introduced in the plasmid at which our gene of interest is ligated. Utilizing the T4 DNA ligase like power sealer, the DNA of our interest in inserted and ligated in the plasmid. Along with the plasmid, a selectable marker is also introduced in the plasmid DNA to identify the recombinant DNA. In addition to this, a promoter region and terminator sequences are also introduced in the plasmid for effective expression of a gene of our interest. A plasmid with our gene of interest and some other important sequences is called a recombinant DNA molecule. Now our recombinant DNA is ready for the expression. If we are performing gene cloning than the plasmid is inserted in the bacterial host, for that generally *E.coli* are commonly used. Once the bacteria start dividing, our recombinant plasmid DNA is also replicated along with it. Now we have the multiple copies of our plasmid DNA which are extracted using the plasmid DNA extraction kit and used for the transformation experiments. The process of Genetic engineering. Transformation into the host genome: Transporting the recombinant DNA into the recipient cell or the host genome is yet another tedious and difficult task. Various methods for recombinant DNA insertion are used for various cell types because a single method is not suitable for all cell types. Various method for transformation: Using stress- bacteria easily uptake the plasmid DNA using some stress factor such as heat or electrical shock. Microinjection- a sharp needle is used for insertion of DNA directly into the nucleus of a cell; however, the method is less effective and required a higher level of expertise for that.

Electroporation- one of the best methods having a great success rate is an electrophoration method in which the recombinant DNA is inserted into the host genome by permeabilizing the cell with electrical current.

Sonication- sonication is yet another method sometimes used in the gene transfer experiment in which the recombinant DNA is inserted into the target cell using ultrasonic waves. The ultrasonic waves also increase the permeability of the cell. Liposome mediated gene transfer- Using an artificial cell-like outer coat known as a liposome- recombinant DNA can be inserted in the host genome. Gene transfer using bacterial infection- This method is one of the popular methods and routinely used in plant genetic engineering experiments. Here, the plant species is infected with the transformed bacteria for inserting a gene of interest. *Agrobacterium tumifecian* is utilized to insert recombinant DNA into the plant cell. A gene of interest is inserted into the T- plasmid of the Agrobacterium. The plant cells are infected by this bacteria cell culture and the transformed cells are regenerated using the plant tissue culture methods. Chemical in gene transfer- Some metal ions, chemicals and solution of different chemicals are also used in the gene transfer experiments, however, the success rate is too low as compared with the other methods.

Confirmation of insert: In the traditional culturing method, presence or absence of selectable marker is used to differentiate transformed cells from the untransformed cells. Although, it is not necessary for the PCR based detection method. The polymerase chain reaction-based detection method is widely accepted more trusted than other methods.

DNA is extracted from the transformed cell and amplified using the primers complementary to our gene of interest or our recombinant DNA. If the recombinant DNA is present it surely amplified otherwise no amplification obtained. For the two-factor conformation, one primer set complementary to recombinant DNA specific and one set of primer complementary to the selectable marker sequence are taken and multiplex PCR is performed. For conforming results, amplification must be obtained in both the reaction. What happened if any mutation occurred during the experiment in our gene of interest? Because the PCR can only amplify the DNA. We must need sequence information to detect the mutation. For that, the DNA sequencing method is used. DNA is extracted from the transformed cells and the gene of interest is amplified using the PCR. Now the PCR amplicons are used for DNA sequencing in which using the fluorescent chemistry the sequence of our gene of interest is orderly determined. Once all the parameter for determining the gene of interest fulfilled, our cells are now ready to inject in the host organism or for tissue culture experiments.

Applications of Genetic engineering:
Genetic engineering has great industrial and agriculture value. It is practised in the medicine, genetic research, agriculture, crop improvement and for production of therapeutic drugs. It is also used in the development of genetically modified organisms. Here we are discussing some of the important applications of genetic engineering.

The recombinant DNA technology is used in the crop improvement and development of new economically important traits. Some of them are:
Herbicide resistance, Virus resistance, Delayed fruit ripening, Altered oil content, Pollen control, Development of cold and drought-tolerant plant species. A classical example of it is the BT cotton- one of the types of genetically modified species give resistance to the plant against bacillus thuringiensis. Process of developing genetically modified plant species:

A gene of interest is isolated from the organism using restriction digestion or amplified using the polymerase chain reaction. Recombinant DNA is constructed by

inserting a gene of interest into the plasmid, here the T-plasmid is used. In the next step, the T- plasmid is inserted into the agrobacterium. In the last step, the plant species is infected with the transformed bacterial cells and cultured. Agrobacterium-mediated gene transfer in plant species. GMF- genetically modified food is another best application of genetic engineering in which economically important food products are constructed using recombinant DNA technology. The classical example of it is Flavr Savr tomato, a genetically modified tomato species made up of the antisense RNA technology. It has great economic values as the GM-tomato can easily be transported from one region to another region of the country.

Another important application of genetic engineering is genetically modified or genetically engineered food. The quality of some of the food products such as cotton, corn and soybeans are improved using the present recombinant DNA technology. The aim of developing genetically modified crop or plant species is to make them economical important, nutritious, protein-rich, disease and stress resistance. Even, using genetic engineering and tissue culture techniques insecticides resistance plant species in tobacco, potato, corn and cotton are developed. In addition to this, some modified plants capable of generating their own fertilizers can also be created using the present genetic modification technique. Transgenic model organisms are developed to test different parameters- the function of certain genes can be determined by designing the transgenic microorganism and animal models. Harmful pathogens and insecticidal pasts can be destroyed using genetically modified microorganism capable of degrading toxics.

Medicinal applications: Low-cost drugs, hormones, enzymes and vaccines are created using the genetic engineering tools. The anti-blood-clotting factor is a great example of it in which plasminogen activating enzyme which is capable of dissolving the blood clot is artificially designed and used in the patients with coronary artery disease or heart attack. Other examples are two other therapeutic proteins somatostatin and lymphokines which are used against several disease condition and synthesized artificially. Insulin is yet a classic example of therapeutic protein designed using genetic engineering technology. A gene for insulin is isolated by restriction digestion or through PCR and inserted int the plasmid. The recombinant plasmid DNA is now inserted into the bacterial or yeast cell in which the plasmid is multiplying. As the microorganism starts dividing it starts making artificial insulin. A large amount of insulin produced using the same technique at an industrial scale. The commercial production of insulin started after the FDA approval in 1982. Recombinant vaccines: Vaccines against smallpox, herpes simplex virus and hepatitis are produced using the genetic engineering technique. The vaccines are the inactivated viral particles used to induce an immune response against that pathogen, however, the chance of contamination is high in it. Using the recombinant DNA technology scientists has created a unique type of vaccines which only contains the DNA for viral coat protein thus the pathogen can never be activated again. The main advantage of it is that it is safer, contamination-free and more reactive. Genetic engineering in gene therapy: Using the gene therapy or gene transfer technique, inherited genetic disorders can be cured. Cystic fibrosis, Duchenne muscular dystrophy and sickle cell anaemia like gene therapies are now under final clinical trial phase and ready to use on patients. In the gene therapy, a faulty, non-function or mutated gene is replaced with the wild type one using the same technique as explained above.

Gene Therapy: Types, Vectors [Viral and Non-Viral], Process, Applications and Limitations. What is Gene Therapy? and How Does it Work?

Naked DNA Mediated Gene Therapy, Sleeping Beauty Transposon System: The Future of Gene Therapy. In addition to this, the genetic engineering technology is likewise used in the production of biofuel, disease, bio alcohol and other essential products. Limitations of genetic engineering: There are ethical issues associated with the use of gene therapy and genetically engineered products. Also, to provide an economic value to the food product or any GM product, the nutritional values are compromised. Because of the adverse effect of it, new resistant pathogenic strains are evolved faster. Also, the side effects of gene therapy and the use of viruses in it are harmful to the target organism.

REFERENCES

1. Shu, Jian-Jun. "A new integrated symmetrical table for genetic codes". BioSystems, 2017; 151: 21–26.
2. Turanov AA, Lobanov AV, Fomenko DE, Morrison HG, Sogin ML, Klobutcher LA, Hatfield DL, Gladyshev VN "Genetic code supports targeted insertion of two amino acids by one codon". Science, 2009; 323(5911): 259–61.
3. Crick, Francis "Chapter 8: The genetic code". What Mad Pursuit: A Personal View of Scientific Discovery. Basic Books, 1990; 89–101.
4. Nirenberg MW, Matthaei JH. "The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides". Proceedings of the National Academy of Sciences of the United States of America, 1961; 47(10): 1588–602.
5. Gardner RS, Wahba AJ, Basilio C, Miller RS, Lengyel P, Speyer JF "Synthetic polynucleotides and the amino acid code. VII". Proceedings of the National Academy of Sciences of the United States of America, 1962; 48(12): 2087–94.
6. Wahba AJ, Gardner RS, Basilio C, Miller RS, Speyer JF, Lengyel P "Synthetic polynucleotides and the amino acid code. VIII". Proceedings of the National Academy of Sciences of the United States of America, 1963; 49(1): 116–22.

7. Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O'Neal C "RNA codewords and protein synthesis, VII. On the general nature of the RNA code". Proceedings of the National Academy of Sciences of the United States of America, 1965; 53(5): 1161–8.
8. Edgar B "The genome of bacteriophage T4: an archeological dig". Genetics, 2004; 168(2): 575–82.
9. Klümper W, Qaim M "A meta-analysis of the impacts of genetically modified crops". PLOS ONE, 2014; 9(11): e111629.
10. Dively GP, Venugopal PD, Finkenbinder C "Field-Evolved Resistance in Corn Earworm to Cry Proteins Expressed by Transgenic Sweet Corn". PLOS ONE, 2016; 11(12): e0169115.
11. Gilbert N "Case studies: A hard look at GM crops". Nature, 2013; 497(7447): 24–6.
12. Ashish M. Parmar, Keyur D. Patel, Nilang S. Doshi, Girish M. Kapadiya, Bhavesh S. Patel and Dr. Dhrubo Jyoti Sen Correlation approach between shotgun sequencing with DNA sequencing in molecular genomics: World Journal of Pharmacy and Pharmaceutical Sciences, 2014; 3(11): 963-995.