https://doi.org/10.56946/jspae.v3i2.492
ISSN: 2957-9082
Journal of Soil, Plant and Environment

REVIEW

Designing of Smart Gene Resources and Computational Approaches For Sustainable Environment; Opportunities and Future Challenges

Nimra Riasat[1,2], Ammara Sana[2], Muhammad Hassaan Khan[2,3], Tahira Kabeer[2,4], Laiba Saher[1,2], Muhammad Tahir Khan[1,2], Muhammad Arshad[1,2]*

[1]Pakistan Institute of Engineering and Applied Sciences (PIEAS), Nilore, Islamabad, Punjab, Pakistan.

[2]Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), P.O. Box 577, Jhang Road, Faisalabad, Punjab, Pakistan.

[3]Centre for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad, Punjab, Pakistan.

[4]College of Agriculture and Biotechnology, Zhejiang University, China.

*Corresponding author:
arshadchbt@yahoo.com
Received: 13 October 2024
Revised: 27 November 2024
Accepted: 02 December 2024

ABSTRACT: Gene revolution is very successful to meet the food challenges resulting from climate change and global warming. Genomic and gene database is ever growing owing to advancement and development of modern biotechnological techniques and machinery. During 1970, 80s, gene isolation was utilized to develop gene resources from original biological systems which was inherently-embraced with challenges of unoptimized protein coding DNA sequences as well as unpredicted expression pattern and levels. Originally, DNA sequences are prone to low and unstable expression in target organisms. So, codon optimization process changed the scenario all the way and most of the problems associated with unmodified sequence has been addressed. At present plethora of softwares are available that fairly process the DNA sequence to make it highly expressible and stable in heterologous systems. Different softwares are being used effectively for synthetic gene design such as, EuGene, COOL, D-Tailor, Costar. Bioinformatic tools have two main functions to data gathering and optimization of gene sequence. DNA sequences is retrieved and processed for Codon adaptation index, GC content, relative synonymous codon usage (RSCU), Protein structure, orthologs, codon pair bias (CPB) and kozak sequences. Codon optimization holds a great potential to develop gene resources which are host friendly and stable for desired traits. In future, gene resources and crop improvement will go side by side for precise and accurate crop improvement and in solving the plant improvement issues previously unaddressed.

KEYWORDS: Gene isolation, codon optimization, gene synthesis, gene designing tools, synthetic biology and biopharming.

This is an open-access review article published by the Journal of Soil, Plant and Environment, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The post genomic era has compelling challenges of handling massive data of biological systems and interpretation for discovery and invention. Conventional methods for data handling and processing that involves computation approaches requires more computing energy, posing time constraints various fundamental and applied research. After throughput experimentations and projects, bulk of information are available that requires proper categorization, management and annotation for deep learning and process driven application in the field of biological sciences. With this, a beacon of hope enlightened in the research areas of global food security with the implementation of computational acquired biological

information and its interpretation. The usage of this information may help in effective utilization of genetic resources for screening and mining the useful genes needed for plant improvement in agricultural crops (Naveen & Sontakke, 2024). Among all, the climate change and global warming are the most highlighted issues that are responsible for developing a hostile and harsh environmental conditions reducing the yield of crops by 50-70% under field conditions, leading to disturbed crop production (Kumar, 2020).

The progressive increase in overall temperature is due to emission of greenhouse gases (global warming) that causes unpredictable drought stress and change in precipitation pattern. Grain yield and quality of crop during growth stages are negatively affected by abiotic stresses as temperature, drought and salinity varies (Kumar, 2020; A. Sharma et al., 2020). To adjust according to different environmental changes, breeding practices have been successfully utilized in many crops for stress resistance, but generally it is not convenient to harness all the stress resilience qualities in a single genotype. Moreover, it is difficult to identify the fundamental molecular and biochemical networks and their features to understand how different stresses impacts are being regulated in the field conditions. Nowadays, the tolerance feature in crops can be improved by using genetically engineered procedure (Hussain et al., 2022).

Crop enhancement has been done by using selection and mutation breeding tools since the domestication and civilization. However, the advancement in biotechnological approaches has revolutionized crop improvement for last three decades. These approaches have been very useful to understand the genetic and molecular base of plant traits and manipulation for crop improvement by genetic transformation (foreign genetic elements addition), backcrossing (crossing of offspring lines with parental lines), self and cross pollination (among the flowers of same and adjacent plants, respectively) and clonal propagation (multiplication of same plant using its small part under *in vitro* conditions) etc. (Begna, 2021). With advancements in modern biotechnology, various methods have been employed to enhance crop varieties, ranging from marker-assisted selection (identification of potential traits linked to a particular phenotype or genetic element at genomic level) to newly developed plant breeding techniques. Genetic modifications have greatly expanded the genetic pool (nucleic acid information) available to plant breeders since the mid-90s. As a result, the development of new plants with a wide range of agricultural traits has gained considerable momentum (Ayan et al., 2022). Stable QTLs (Quantitative trait loci) and molecular markers (biological molecule that serves as measurable indicator) are important for molecular breeding and identifying candidate genes, that can be used in transgenic methods or molecular introgression (transfer of genetic elements from one organism to another within the same species to produce hybrid) (Raj & Nadarajah, 2022).

Crop productivity has significantly enhanced by using conventional breeding over the past century. However, traditional breeding methods have now hit their maximum limits because of the complex, multigenic nature of biotic/abiotic stress

factors. Over the past decades, numerous stress-related genes have been identified and confined to produce stress-tolerant plants (Esmaeili et al., 2022). Research has shown that epigenetic changes, including DNA methylation and histone acetylation can influence the gene expression of stress-related genes and improve plant stress tolerance (Mishra et al., 2023).

In modern biotechnology, chemical synthesis of DNA molecules is very trending approach serving as highly effective and resilient approach to enhance or down regulation of target genes. Moreover, heterologous expression of proteins is often required for unraveling the underlying mechanisms in fundamental and applied research. To maximize the expression of DNA sequence in the host, the only possible way is DNA sequence optimization., This will improve the translation efficiency of target gene because raw DNA sequence of one species is convertible to other species synergistically (Gustafsson et al., 2004).

Most of the success stories of GM crops are based on usage of processed DNA sequences of genes because *in vivo* approaches of gene isolation are inefficient and possess limited gene expression. Initially, target genes are expressed in heterologous systems of model organisms and their structure and function are characterized. All the plant trait genes were manipulated by *in-silico* analysis and optimized for different parameters followed by transformation in crop plants that are introgressed into other cultivars (Jana & Deb, 2005; Lou et al., 2007; Peng et al., 2007). About sixty years ago, chemical synthesis of DNA was a dream. HG Khorana, is the pioneer in devising a set of techniques to synthesize oligo nucleotides in 1970s. Chemical synthesis of tRNA genes proved very effective and successful (Gupta et al., 1976). Other artificially synthesized genes were lac operator and tyrosine repressor, somatostatin and cloned in different vectors successfully. During 1980s, different research groups become interested in the chemical synthesis of a large number of genes and expression in *E. coli* (RIGGS et al., 1980). Limited length of synthesized nucleotides is a limiting factor and over time it has been improved significantly. Nowadays, chemical synthesis is a routine practice and has become a powerful tool in creating transgenic organisms. Most of the genetic traits either of plant or animal origin are being genes sequenced thoroughly using computational techniques and chemically synthesized followed by introduced into heterologous systems.

DNA is highly resilient in nature and can be optimized to a range of quantitative gene and protein expression. Oftenly, it is advised to optimize original codon sequence as per natural distribution of host codons aiming to preserve slow translation regions playing critical role in protein folding. It is after successful application of optimized DNA sequences for industry, agriculture and pharmaceutical purposes that different biotech companies are providing custom gene optimization services. Different empirical indexes for optimizing gene resources have been set as standards for key variables such as codon adaptation index (CAI), the codon bias index, frequency of relative codon usage, optimal codon usage and effective codon number as well as duplication of Kozak's sequence (Fu et al., 2020).

Based on deep learning, a novel optimization tools are available which were devised to introduce the concept of codon boxes. Codon boxes emphasize the recoding of the sequences into codon box sequences irrespective of order of the bases. The codon optimization model for *E. coli* was trained for bidirectional long-short term memory conditional random field. Protein expression experiments for *Plasmodium falciparum* candidate vaccine and polymerase acidic protein in comparison with original sequences were performed which showed that optimize sequences by Thermo Fischer and Genewiz are highly efficient and competitive in enhancing expression (Fu et al., 2020). Modern approaches of crop improvement will hinge on a suit of innovative approaches to run crop improvement programs on a fast track by creating novel diversity and their efficient and rapid incorporation into crop cultivars. Deep learning by use of bioinformatics tools for identification, processing and manipulation is highly useful owing to algorithms-based DNA sequence information and performance with high precision.

Desired traits can be created and transferred to other plants possibly done by genetic engineering techniques (as transgenic). Improvement in crop potential has been achieved using manipulated DNA sequences for developing resilient plants against abiotic and biotic stresses that otherwise is impossible by conventional tools of plant improvement (Passricha et al., 2020).

2. Gene Isolation

Availability of gene resources is prerequisite of crop improvement program and are accessed form breeding parents, gene isolation and DNADNA synthesis. DNA synthesis is an emerging field of genetic resources and being benefitted from advancements in molecular biology, sequencing, bioinformatics tools and engineering machines. Principally, DNA synthesis commences from retrieving DNA sequence information from data bases, planning and designing depending upon objectives followed by synthesis and transformation in model plant systems. After the launching of several sequencing projects, the genomic and metagenomic data bases are ever growing but lacks the tangible DNA which are stressing the development of novel DNA synthesis technologies (Sequeira et al., 2016).

Numerous genes involved in abiotic stress tolerance have been identified and characterized using molecular techniques and during plant transformation experiments these genes have been individually overexpressed. However, the phenotypes related to stress tolerance are polygenic traits, that are being dissected by using current genomic tools. The genetic introgression is accelerated by using these tools through using molecular markers or site-directed mutagenesis techniques like CRISPR-Cas9 (Villalobos-López et al., 2022).

Conventionally, gene isolation is a tedious and error-prone method with the use of restriction enzymes, purified from different sources as gene does not exist as a separate unit in cells, but present as a small part within a DNA molecule. The enough yield of specific gene would only be possible by using vast amount of DNA. It can create even more challenges in selecting unique cutters, performing digestion and ligations, and

ensuring that all assembled portions maintain the correct open reading frame (Alberts et al., 2002).

3. Computational Biology

Computational biology relies on the application of computational algorithms and mathematical models to understand the biological systems from various perspectives (Hattaf, 2024). Moreover, this innovative approach has also facilitated to manipulate genomic data, thereby opening new horizons for innovative research work in different domains of science like agriculture, environmental and health sciences (Gayathiri et al., 2023). Comparative genomics is also one of those areas in which large biological data sets are compared to identify distinct genomic patterns (A. Sharma et al., 2020). While, the specific techniques of this particular domain are concerned with molecular modelling, molecular dynamics simulations (predictive models) and machine learning. The later one has significantly changed the landscape of computational biology by integrating the high-throughput data analysis, leading to the formation of efficient predictive models (Chen et al., 2024).

All of these innovative methods are widely used in drug designing, agricultural and environment sciences as well (V. Sharma, 2020; Zhang et al., 2022). In terms of drug designing, virtual screening and molecular dynamics simulations assisted novel drug designing has enable to be optimistic and precise while selecting a potent drug against a particular disease. For example, in recent times one of a published article has elaborated the similar approaches to screen-out phytocompounds against cancer which would definitely be highly crucial for pharmacologists, molecular biologists and other health sciences researchers (Das & Agarwal, 2024).

Whereas, in agricultural sector, computational biology has provided the tailored solution for improvement of crops, yield enhancement and resistance against various diseases and pests by decoding the crop genomics data (Prajval et al., 2024). Another related and crucial aspect is the sustainable agriculture. One of the reported studies has highlighted that there is a certain cyber-agricultural system (CAS) which focuses on the artificial intelligence and scalable cyber-infrastructure in both of the domains that are crop breeding and production agriculture (Sarkar et al., 2024). Similarly, environmental sciences also employ different computational and *in-silico* approaches which give insights regarding the way different toxins interact with environmental factors and cause pollution. Arsenic (As) is one of the environmental toxins which is highly important in terms of public health. It has been studied by the integration of molecular docking and molecular dynamics simulations that how arsenic interacts with different enzymes. The objective of this work was to unveil the binding patterns of As-enzyme complexes so that it could be assessed, which enzyme could be effective against such a hazardous environmental pollutant (Ahmad et al., 2024).

4. DNA Synthesis

Synthetic biology is an interdisciplinary domain dedicated to designing and creating new biological components and systems, along with modifying and constructing existing biological systems (Sharkey et al.,

2024). The emergence of synthetic biology represents a cultural shift having extensive implications for the biotechnology sector. Today the efforts of synthetic biologists anticipate the emergence of a novel era of cyber-biological system that could pave the way for the 5th industrial revolution. Through addressing the scientific publishing requirements of a board and varied community, synthetic biology aims to contribute to the advancement of this new emerging engineering discipline, initiate the necessary cultural shifts and nurture the growth of a new industry well into twenty-first century (Peccoud, 2016).

Nowadays, DNA synthesis is among the most vital biological technologies, playing a significant role in genome studies, research related to gene expression and studies related to gene network, etc. (Fang & Liang, 2022). The ability to read and write DNA is open secret by computational and engineering tools making DNA synthesis a reliable and economic industry (Hoose et al., 2023). The demand for synthetic DNA is increasing day by day beyond many areas of research and economic, marketing and trading activities. The advancements in engineering biotechnology, storage of data and Nanotechnology are set to provide rapid growth of DNA at reduced cost and large quantity (Figure 1). The achievements in gene editing, next generation sequencing (NGS) and gene isolation, DNA synthesis are already emerging as a thriving industry with enormous scope and applications (Hoose et al., 2023).

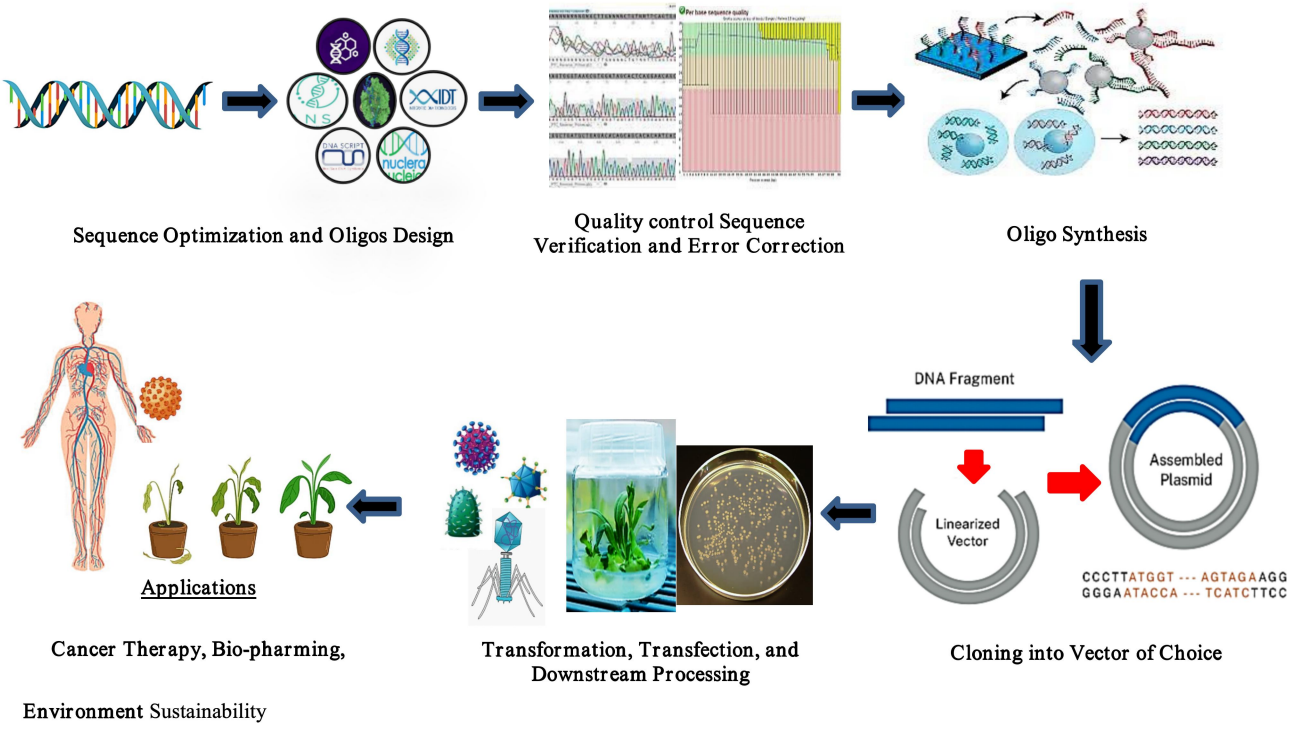

Figure 1. Steps involve in gene synthesis.

DNA synthesis involves the synthesizing a gene in laboratory without relying on an initial template. In contrast to traditional gene cloning methods, which require a template DNA sequence to be amplified (Javed, 2013). Despite the beliefs of many researchers, commercially synthesizing gene service going to progress rapidly to offer a cost-effective alternative to conventional cloning and other methods in molecular biology.

The primary reasons for this include time efficiency, enhanced DNA performance, cost effectiveness and convenience. The multi-step traditional cloning process includes designing a strategy, synthesis of primers, PCR, gel extraction, bacterial transformation and additional intricate procedures. The process demands significant time and human resources, unlike DNA synthesis, which not required these requirements. DNA synthesis does not require a physical template and restriction enzymes to separate gene from source as required in traditional gene cloning; by providing the nucleotide sequence or amino acid sequences, a researcher can obtain a gene of their choosing. In many instances, the cost of ordering a synthetic gene is lower than that ordering oligos, cloning kits and sequencing services. DNA synthesis enables codon optimization, a process demonstrated to enhance DNA performance and protein expression efficiency (Schwartz et al., 2011).

5. Codon Optimization

Codon optimization relies on the concept that replacing codons with synonymous codons present in a gene originating from a donor organism that are more commonly used in the host organism can result in higher levels of recombinant protein expression (Jenkins et al., 2023). Protein synthesis in heterologous hosts is an essential tool in biotechnology. Nevertheless, degeneracy exhibited by genetic code, leading to biased codon biased in a lot of organisms. The level of protein expression significantly affected by changes in synonymous codon, tailored specifically for individuals host organisms. This phenomenon can be quantified using metrics like codon pair bias, relative codon pair bias, codon adaptation index and relative codon bias. The designing of codons that boost multiple objectives is done by codon optimization. Presently, existing software and algorithm solutions either depend on heuristics without optimality assurances or exhibit inflexibility in modelling various objective functions and constraints (Şen et al., 2020).

Codon optimization may be necessary to ensure efficient expression when transferring a natural DNA sequence to a different organism, as the original sequence may not be optimally expressed in the new host (Young & Dong, 2004). Codon optimization, which involves modifying codons within a gene sequence to enhance the expression of recombinant proteins, has become a common practice (Webster et al., 2017). For heterologous gene expression, codon optimization has been a longstanding technique to improve protein expression. Optimizing gene codons leads to significant increases in protein and RNA levels, indicating that codon usage plays a crucial role in governing gene expression (Zhou et al., 2016). A specific amino acid involves in the synthesis of a protein encoded by a sequence of three nucleotides called codon. The genetic code degeneracy refers to that multiple codons specify a single amino acid.

64 distinct codons are present naturally, they only encode for 20 amino acids, arises to the genetic code's degeneracy. As, UGU and UGC two codons that encodes cysteine. Whereas six synonymous codons CUG, UUA, CUC, CUA, UUG and CUU (Figure 2). The degeneracy of codon results in many probable ways of encoding a protein. Codon usage bias suggested that some synonymous codons are more preferred over others when encodes a specific amino acid in a specific organism (Şen et al., 2020).

Codon usage bias refers to the preferred use of specific synonymous codons, naturally occur with different frequencies in different species. The substitutions of synonymous codon regarded as "silent" because they didn't affect the primary structure of a protein but now it is considered that synonymous codons as substitution had critical impact on heterologous protein expression (Diez et al., 2022; Webster et al., 2017).

6. Protein Expression and Codon Optimization

Producing industrial and therapeutic recombinant proteins in plants provides numerous benefits as compared to traditional bacterial or mammalian systems, offering reduced cost, enhanced scalability, simpler growth conditions and improved product safety (Feng et al., 2022). For the enhancement of protein expression recombinant gene technologies utilize codon optimization, which involves replacing codons with synonymous ones. One should have accurate knowledge of codon usage frequencies to understand the chemistry behind its implementation. Assessing codon usage bias accurately across various organisms serves multiple purposes beyond codon optimization. It also aids to explore various studies through codon usage similarities as phylogenetic relations of organisms, host-pathogen relationships, evolutionary and translation research.

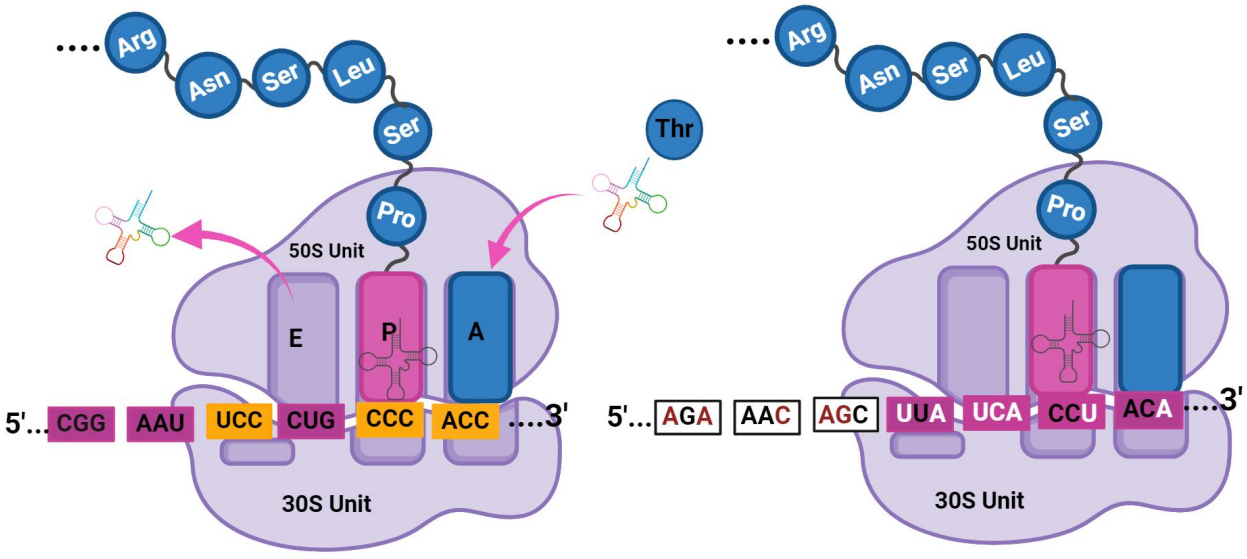

Figure 2. Degeneracy impact, mRNA with different Synonymous codons can code for the same polypeptide.

The co-translational protein folding and translation kinetics also influenced by affecting protein structure and function (Athey et al., 2017). DNA synthesis has become an important tool across many domains, diagnosis of diseases, gene therapy, production of biopharmaceuticals and vaccine development. Synthetic gene are integrated into the genetic material of different host organisms, as yeast and bacteria, facilitating the expression and production of proteins (Şen et al., 2020).

It has been identified that codon biased usage is strongly linked to protein expression and as the most conserved property (Lithwick & Margalit, 2003). The unequal usage of synonymous codons is a widely recognised phenomenon in living organisms. This can significantly contribute to improve the efficiency and accuracy of translation processes (Majeed et al., 2020). The production of functional proteins in different hosts is key to futuristic biotechnology. Unfortunately, achieving successful expression outside their native environment is often challenging. Proteins may include codons infrequently utilized in the desired host, originate from organisms that use non-standard genetic codes, or contain regulatory elements within their coding sequence that restrict expression. Advancements in DNA synthesis, lower cost and high speed have enabled to completely re-design a gene sequences and optimize its protein expression probabilities. There are various re-design strategies, such as adjusting translation initiation regions, modifying mRNA structural elements, and employing alternative codon biases (Gustafsson et al., 2004).

Translation efficiency and expression levels can be enhanced by substituting frequently observed codons instead of rarely observed ones in host organism. Enhancements in expression levels up to 105-fold can be possible when optimal codons are utilized during gene design (Gustafsson et al., 2004). Over time, different computer programs have been created to assist in the process of optimizing gene sequences (Angov, 2011; Gustafsson et al., 2004).

7. Gene Design Tools

Artificial synthesis of DNA fragments is drastically determining the efforts for improvement of biological systems. Use of AI for refining DNA sequences, editing and artificial synthesis are undeniable developments whose applications are penetrating in fundamental and applied research of living organisms (Dixit et al., 2024; Villalobos et al., 2006). Designing tools of DNA are widely used in fundamental and applied biotechnological research. The popularity is reflective of the simplicity, convenience, efficacy and efficiency that it is being utilized in manipulation of biological system and processes. At present, most of the biotech traits either biotic or non-biotic and even biopharming really thanks to availability of user-friendly gene design tools (Table 1).

In comparison to other tools, JCat (Java Codon Adaptation Tool) offers an approach that eliminate the manual designation of highly expressed genes, making it a swift and straightforward method. It offers additional-options for codon adaptation such as Rho-independent transcription terminators and preventing undesired restriction enzymes cleavage sites.

Table 1. Comprehensive review of gene design characteristics involve multi-objective optimization tools.

Synthetic gene design tools	GC/AT Content	Codon usage Optimi-zation*	Codon context bias optimi-zation	Restriction site modification	Repetitive base excision	mRNA secondary structure	Oligo generation	Hidden stop codon	Merits	Demerits	Reference
EuGene	✓	✓	✓	✓	✓	✓		✓	Data gathering and optimization are main functions	Difficult to understand output due to scarce documentation Harder to understand results because uses of percentages	(Gaspar et al., 2012)
JCat	✓	✓		✓					Web- based tool Rapid analysis Remove the rho-independent transcription terminator	bounded to certain expression systems limited optimization parameters	(Gould et al., 2014)
COOL	✓	✓	✓	✓	✓	✓		✓	Efficient algorithms for gene design	Limited scalability Limited output interpretation	(Chin et al., 2014)
D-Tailor	✓	✓		✓		✓			Hydropathy index optimization Multi-objective optimization	Lack of comprehensive documentation	(Webster et al., 2017)
OPTIMIZER		✓		✓			✓		Focus on optimizing rare codons User friendly interface	Limited to specific organisms May not consider all factors	(Gould et al., 2014)
Gene Designer	✓	✓							Allows manual adjustment of codons Flexible customization options	Limited optimization algorithms, not support all expression systems	(Gould et al., 2014)
COStar	✓			✓	✓	✓			Optimizes both the local and global properties of DNA simultaneously	Requires expertise in molecular biology	(Webster et al., 2017)

Note: ✓ represents yes and represents no.*JCat, D-Tailor and Gene Designer optimizes the codon usage of the target gene by increasing its CAI score, codon usage optimization by EuGene based upon RSCU and CAI values, COOL optimizes codon usage by utilizing both the ICU and CAI scores, the adjustment of codon usage in OPTIMIZER can be achieved by three different ways: CAI Optimal, Guided random and Selective optimization.

The results of JCat are provided both graphically and in terms of Codon Adaptation Index (CAI) values for both the adapted sequence and input sequence. Moreover, CAI values calculation can be taken by the list of genes uploaded in FASTA format (Grote et al., 2005).

Gene Designer stands as a standalone software for swift and straightforward design of synthetic DNA segments. Users can effortlessly incorporate, modify, and combine genetic elements like promoters, open reading frames (ORF), and tags using an intuitive drag-and-drop hierarchical DNA/Protein object map coupled with graphic interface. Leveraging advanced optimization algorithms, open reading frames within the DNA construct can be swiftly codon-optimized for enhanced protein expression across diverse host organisms. Gene Designer also encompasses features including an oligonucleotide annealing temperatures calculator as real-time sliding calculator, a sequencing primer generator, tools for selective inversion or insertion of restriction sites, and options for maximizing or minimizing sequence identity to a reference (Villalobos et al., 2006).

Gene expression levels can be enhanced by optimizing codon usage through an online tool known as OPTIMIZER. It provides three optimization methods: a guided random method employing a Monte Carlo algorithm, 'one amino acid-one codon' approach and a new method aimed at maximizing optimization with minimal alterations to the original sequence. A key feature of OPTIMIZER is its capability to optimize sequences of DNA using pre-computed codon usage tables derived from a wide array of prokaryotic species, encircled 150 species characterized by strong translational selection (Puigbo et al., 2007).

There are many codons optimization tools focus on a single design criterion and offer less flexibility, resulting in only one optimal sequence being generated, that may be not the ideal solution. To address this problem, we've introduced COOL (Codon Optimization Online), the initial web tool to providing multi-objectives codon optimization capabilities, thereby enhancing synthetic gene design systematically. COOL is a user-friendly allowing customization of various codon optimization parameters as individual codon usage, codon pairing and codon adaptation index (CAI). Furthermore, COOL enables comparison between user-defined DNA sequence and COOL-optimized sequences, to understand for further improvement in the user's design (Chin et al., 2014).

A software program provides a powerful and convenient tool for custom-made codon optimization: The user-friendly graphical interface allows even untrained scientists to easily create, test, adjust and save intricate codon optimization strategies. It also makes possible the sharing of successful optimization strategies within scientific community. Codon Wizard offers versatile functionalities for customizable modification, optimization of codon usage and sequence analysis for any provide input data sequence (RNA/DNA/Protein) by using a range of algorithms (Rehbein et al., 2019).

D-star Lite is a dynamic search algorithm based on codon optimization. Firstly, the algorithm transforms the codon optimization into acyclic graph, incorporates weights on

its edges using a sliding window technique. Afterwards, by utilizing the D-star Lite algorithm to calculate the shortest path from the starting point to the target site within the generated graph. Therefore, the way of optimizing a gene is transformed into a real time search for the shortest path within an obtained graph (Liu et al., 2014).

COStar is powerful tool and have potential to optimize multiple objectives all together, resulting in improved results such as lower variance of GC content, fewer repetitive sequences and reduced hairpin formation during codon optimization (Webster et al., 2017).

A Multitask software tool, enables the design of sequences of synthetic DNA based on properties. It facilitates the integration of various sequence analysis within a universal Monte Carlo stimulation framework, thereby facilitating the progressive evolution of sequences towards predetermined, delineated properties (Guimaraes et al., 2014).

D-Tailor incorporates codon usage optimization depends on refinement of mRNA secondary structure, optimization of GC content, Codon Adaptation index (CAI) and elimination of repetitive Pattern as restriction sites (Webster et al., 2017).

EuGene Provides two main functionalities: gathering data and gene optimization. In the process of data gathering phase, the information about a gene is collected from various online sources, genome statistical analyses and different offline tools. During codon optimization phase, EuGene modifies the gene considering various factors that affect mRNA decoding, by utilizing the previous data gathered in prior step. The primary function of EuGene is to optimize gene codon sequences based on specific criteria. Six redesign approaches are offered by this tool for customizing genes: hidden stop codon control, codon usage (RSCU and CAI), removal of deleterious sites like Shine-Dalgarno sequences, codon context (CPB), elimination of repetitions and G+C content. These modifications are made while preserving the native amino acid sequence (Gaspar et al., 2012).

8. Synthetic Biology and Biopharming

Over the past decades, progresses in plant molecular farming techniques have transformed plants into as promising manufacturing system, capable of achieving commercial-scale production level quickly (Shanmugaraj et al., 2020). The initial production of first recombinant pharmaceutical protein derived from plants (PDP) was human serum albumin (HAS) that was produced in transgenic tobacco and potato plant in 1990 (Sijmons et al., 1990).

The rising demand for recombinant proteins has garnered global attention, spurring the development of recombinant protein production technologies and engineering of expression systems based on mammalian or bacteria cell cultures. Plants have emerged as valuable hosts for producing economically significant, non-native protein that are biologically active or biopharmaceuticals, known as plant molecular farming (PMF). PMF emerged as a cost-effective technology that experienced substantial growth and advancement over the last two decades. The enhancement and refinement of transient expression systems have led to immense reduction in protein production timelines and improvements in yield within plant-based systems as protein

yield in plants. The key factors making plant-based platforms potential competitors as conventional expression systems includes their flexibility, versatility, cost-effectiveness, robustness and scalability. Plants are now used to produce numerous biopharmaceuticals, as monoclonal antibodies, recombinant vaccine antigen and many other commercially viable proteins, which are progressing through pre-clinical and clinical development stages (Shanmugaraj et al., 2020).

In plant molecular pharming, multiple factors play important role as plant hosts, target gene, vector expression cassette and last one extraction and purification methods. These steps play crucial role in biopharming (Makhzoum et al., 2014). Molecular farming is reaching a point where it could challenge established manufacturing techniques that utilizes yeast, bacteria and cultured mammalian cells (Ma et al., 2005). Molecular manipulation of the plant secretory pathway is now facilitating the production of progressively intricate biomolecules through tailored protein-specific approaches. These methods ensure proper maturation by eliminating unwanted processing events and the incorporation of heterologous biosynthetic machinery to facilitate the production of specific target proteins (Margolin et al., 2020).

9. Future Challenges

Judicial and rational expression would be the aim in efficient management of multigene expression to address simultaneously different traits of plant improvement. Gene optimization will face different challenges depending upon gene type, mode of action and spatial/temporal variations during plant growth and development. Moreover, role of regulatory sequences and transcription factors will require another layer of sequence optimization for sustainable gene level and drive without detrimental consequences on growth and yield. It means different parameters will need to consider while optimizing sequence depending up gene nature, its role and impact on other genes of the genome.

In future, every gene optimization will require rigorous testing in some system that is most convenient to study the gene expression and function in relation with other genes of the host genes without any negative impacts on organism life. As an example, various genes will be modulated for improved and rapid growth of plants and animal for better nutrition management. Various plants can be altered to provide special food requirement in unsubstantial conditions. Moreover, sequencing will also play a crucial role to allow maximum and safe level of expression by optimized sequence for fundamental research and commercial purposes.

10. Conclusion

Crop resilience is priority area of different breeding programs and can be accomplished by rationalizing gene choices, expression levels and fidelity in subsequent generations for the regulation of various genetic elements that targets biotic and abiotic factor. Moreover, conventional crop improvement tools are insufficient to ensure food security and climate challenges threatening sustainability in food crop production. To address major agricultural challenges, gene isolation and DNA synthesis offer powerful tools for crop improvements. As codon

optimized genes expressed at higher rate, they may cope up the relevant stress in various agronomical crops (self-sufficiency for fertilizer utilization, insect and disease resistance, mineral accumulation), as well as animal sector (better inbred populations, more milking potential, edible vaccine development, reduced and improved health risks and statues). Their successful implementation requires careful consideration of the related challenges. The pathway from isolating genes to DNA synthesis offers transformative possibilities for crop enhancement, promising increased yields, improved nutritional content, and greater resilience to environmental variables. However, achieving these benefits necessitates addressing important scientific, regulatory and ethical challenges. The sustainability for future of agriculture, a collaborative effort involving policy-makers, scientists and public is essential to harness these technologies for developing resilient food crops.

Authors contributions: Muhammad Arshad conceived the idea and supervised the whole write up process. Ammara sana and Tahira kabeer contributed by writing the abstract and introduction part as well as computational biology part of the manuscript. Nimra Riasat was directly involved in literature collection, manuscript –write-up and making of graphical diagrams. Muhammad Hassan Khan executed the write-up, draft setting, grammer editing and finalization of the manuscript as per journal requirements. Muhammad Tahir Khan and Laiba Sehar contributed in writing the introduction part of the manuscript.

Acknowledgments: Not Applicable (N/A).

Conflicts of Interest: The authors bear no conflicts of interest

Availability of Data and Materials: Data will be available on a formal request from the corresponding author.

Funding: Not applicable(N/A).

REFERENCES

Ahmad, I., Singh, A. K., Mohd, S., Katari, S. K., Nalamolu, R. M., Ahmad, A., . . . Nadeem, M. S. In silico insights into the arsenic binding mechanism deploying application of computational biology-based toolsets. ACS omega. (2024). 9(7), 7529-7544. https://doi.org/10.1021/acsomega.3c06313

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Isolating, cloning, and sequencing DNA. In Molecular Biology of the Cell. 4th edition: Garland Science.

Angov, E. Codon usage: nature's roadmap to expression and folding of proteins. Biotechnology journal. (2011). 6(6), 650-659. https://doi.org/10.1002/biot.201000332

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L. V., Katneni, U.,Kimchi-Sarfaty, C. A new and updated resource for codon usage tables. BMC bioinformatics. (2017). 18, 1-10. https://doi.org/10.1186/s12859-017-1793-7

Ayan, A., Meriç, S., Gümüş, T., & Atak, Ç. Next generation of transgenic plants: From farming to pharming. Genetically modified plants and beyond. (2022). 49. https://doi.org/10.5772/intechopen.102004

Begna, T. Role of molecular markers in crop improvement. International Journal of Agriculture and Biosciences. (2021). 10(1): 17-26.

Chen, V., Yang, M., Cui, W., Kim, J. S., Talwalkar, A., & Ma, J. Applying interpretable machine learning in computational biology-pitfalls, recommendations and opportunities for new developments. Nature methods. (2024). 21(8), 1454-1461. https://doi.org/10.1038/s41592-024-02359-7

Chin, J. X., Chung, B. K.-S., & Lee, D.-Y. Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. Bioinformatics. (2014). 30(15), 2210-2212. https://doi.org/10.1093/bioinformatics/btu192

Das, A. P., & Agarwal, S. M. Recent advances in the area of plant-based anti-cancer drug discovery using computational approaches. Molecular Diversity. (2024). 28(2), 901-925. https://doi.org/10.1007/s11030-022-10590-7

Diez, M., Medina-Muñoz, S. G., Castellano, L. A., da Silva Pescador, G., Wu, Q., & Bazzini, A. A. iCodon customizes gene expression based on the codon composition. Scientific reports. (2022). 12(1), 12126. https://doi.org/10.1038/s41598-022-15526-7

Dixit, S., Kumar, A., Srinivasan, K., Vincent, P. D. R., & Ramu Krishnan, N. Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions. Frontiers in Bioengineering and Biotechnology. (2024). 11, 1335901. https://doi.org/10.3389/fbioe.2023.1335901

Esmaeili, N., Shen, G., & Zhang, H. Genetic manipulation for abiotic stress resistance traits in crops. Frontiers in Plant Science. (2022). 13, 1011985. https://doi.org/10.3389/fpls.2022.1011985

Fang, G., & Liang, H. An integrated algorithm for designing oligodeoxynucleotides for gene synthesis. Frontiers in genetics. (2022). 13, 836108. https://doi.org/10.3389/fgene.2022.836108

Feng, Z., Li, X., Fan, B., Zhu, C., & Chen, Z. Maximizing the production of recombinant proteins in plants: from transcription to protein stability. International Journal of Molecular Sciences. (2022). 23(21), 13516. https://doi.org/10.3390/ijms232113516

Fu, H., Liang, Y., Zhong, X., Pan, Z., Huang, L., Zhang, H., Liu, Z. Codon optimization with deep learning to enhance protein expression. Scientific reports. (2020). 10(1), 17617. https://doi.org/10.1038/s41598-020-74091-z

Gaspar, P., Oliveira, J. L., Frommlet, J., Santos, M. A., & Moura, G. EuGene: maximizing synthetic gene design for heterologous expression. Bioinformatics. (2012). 28(20), 2683-2684. https://doi.org/10.1093/bioinformatics/bts465

Gayathiri, E., Prakash, P., Kumaravel, P., Jayaprakash, J., Ragunathan, M. G., Sankar, S., Govindasamy, R. Computational approaches for modeling and structural design of biological systems: A comprehensive review. Progress in Biophysics and Molecular Biology. (2023).mhttps://doi.org/10.1016/j.pbiomolbio.2023.08.002

Gould, N., Hendy, O., & Papamichail, D. Computational tools and algorithms for designing customized synthetic genes. Frontiers in Bioengineering and Biotechnology. (2014). 2, 41. https://doi.org/10.3389/fbioe.2014.00041

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research. (2005). 33(suppl_2), W526-W531. https://doi.org/10.1093/nar/gki376

Guimaraes, J. C., Rocha, M., Arkin, A. P., & Cambray, G. D-Tailor: automated analysis

and design of DNA sequences. Bioinformatics. (2014). 30(8), 1087-1094. https://doi.org/10.1093/bioinformatics/btt742

Gupta, R., Randerath, K., & Randerath, E. Sequence analysis of small amounts of nonradioactive oligoribonucleotides containing ribose-methylated nucleosides by a combination of 3H-and 32P-labeling techniques. Analytical Biochemistry. (1976). 76(1), 269-280. https://doi.org/10.1016/0003-2697(76)90285-2

Gustafsson, C., Govindarajan, S., & Minshull, J. Codon bias and heterologous protein expression. Trends in biotechnology. (2004). 22(7), 346-353. https://doi.org/10.1016/j.tibtech.2004.04.006

Hattaf, K. A new mixed fractional derivative with applications in computational biology. Computation. (2024). 12(1), 7. https://doi.org/10.3390/computation12010007

Hoose, A., Vellacott, R., Storch, M., Freemont, P. S., & Ryadnov, M. G. DNA synthesis technologies to close the gene writing gap. Nature Reviews Chemistry. (2023). 7(3), 144-161. https://doi.org/10.1038/s41570-022-00456-9

Hussain, K., Lone, S., Mushtaq, F., Malik, A., Narayan, S., Rashid, M., & Nazir, G. (2022). Abiotic Stresses and Their Management in Vegetable Crop Production. In Advances in Plant Defense Mechanisms: IntechOpen. https://doi.org/10.5772/intechopen.105453

Javed, A. Artificial Gene Synthesis In Vitro. International Journal of Science and Nature(IJSN). (2013). Vol-4, 124-130.

Jenkins, M. C., Parker, C., O'Brien, C., Campos, P., Tucker, M., & Miska, K. Effects of codon optimization on expression in Escherichia coli of protein-coding DNA sequences from the protozoan Eimeria. Journal of Microbiological Methods. (2023). 211, 106750. https://doi.org/10.1016/j.mimet.2023.106750

Kumar, S. Abiotic stresses and their effects on plant growth, yield and nutritional quality of agricultural produce. Int. J. Food Sci. Agric. (2020). 4(4), 367-378. https://doi.org/10.26855/ijfsa.2020.12.002

Lithwick, G., & Margalit, H. Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome research. (2003). 13(12), 2665-2673. https://doi.org/10.1101/gr.1485203

Liu, X., Deng, R., Wang, J., & Wang, X. COStar: a D-star Lite-based dynamic search algorithm for codon optimization. Journal of theoretical biology. (2014). 344, 19-30. https://doi.org/10.1016/j.jtbi.2013.11.022

Lou, X.-M., Yao, Q.-H., Zhang, Z., Peng, R.-H., Xiong, A.-S., & Wang, H.-K. Expression of the human hepatitis B virus large surface antigen gene in transgenic tomato plants. Clinical and Vaccine Immunology. (2007). 14(4), 464-469. https://doi.org/10.1128/CVI.00321-06

Ma, J. K. C., Barros, E., Bock, R., Christou, P., Dale, P. J., Dix, P. J., . . . Pezzotti, M. Molecular farming for new drugs and vaccines: current perspectives on the production of pharmaceuticals in transgenic plants. EMBO reports. (2005). 6(7), 593-599. https://doi.org/10.1038/sj.embor.7400470

Majeed, A., Kaur, H., & Bhardwaj, P. Selection constraints determine preference for A/U-ending codons in Taxus contorta. Genome. (2020). 63(4), 215-224. https://doi.org/10.1139/gen-2019-0165

Makhzoum, A., Benyammi, R., Moustafa, K., & Trémouillaux-Guiller, J. Recent advances on host plants and expression cassettes' structure and function in plant molecular

pharming. BioDrugs. (2014). 28, 145-159. https://doi.org/10.1007/s40259-013-0062-1

Margolin, E. A., Strasser, R., Chapman, R., Williamson, A.-L., Rybicki, E. P., & Meyers, A. E. Engineering the plant secretory pathway for the production of next-generation pharmaceuticals. Trends in biotechnology. (2020). 38(9), 1034-1044. https://doi.org/10.1016/j.tibtech.2020.03.004

Mishra, S., Panda, S., Akcil, A., & Dembele, S. Biotechnological avenues in mineral processing: Fundamentals, applications and advances in bioleaching and bio-beneficiation. Mineral Processing and Extractive Metallurgy Review. (2023). 44(1), 22-51. https://doi.org/10.1080/08827508.2021.1998043

Naveen, A. K., & Sontakke, M. A review on regulatory aspects, challenges and public perception in acceptance of genetically modified foods. Food science and biotechnology. (2024). 33(4), 791-804. https://doi.org/10.1007/s10068-023-01481-0

Passricha, N., Saifi, S. K., Negi, H., Tuteja, R., & Tuteja, N. (2020). Chapter 20 - Transgenic approach in crop improvement. In N. Tuteja, R. Tuteja, N. Passricha, & S. K. Saifi (Eds.), Advancement in Crop Improvement Techniques (pp. 329-350): Woodhead Publishing. https://doi.org/10.1016/B978-0-12-818581-0.00020-6

Peccoud, J. (2016). Synthetic Biology: fostering the cyber-biological revolution. In (Vol. 1, pp. ysw001): Oxford University Press. https://doi.org/10.1093/synbio/ysw001

Peng, Y., Lin, W., Cai, W., & Arora, R. Overexpression of a Panax ginseng tonoplast aquaporin alters salt tolerance, drought tolerance and cold acclimation ability in transgenic Arabidopsis plants. Planta. (2007). 226, 729-740. https://doi.org/10.1007/s00425-007-0520-4

Prajval, V., Krishnamoorthi, A., Thakur, R., Ruchitha, T., Kapoor, M., Singh, A., & Chittibomma, K. From Code to Crop: How Bioinformatics is Transforming Crop Genomics in Modern Agriculture and Bettering Environment. Journal of Advances in Biology & Biotechnology. (2024). 27(4), 27-49. https://doi.org/10.9734/jabb/2024/v27i4737

Puigbo, P., Guzmán, E., Romeu, A., & Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic acids research. (2007). 35(suppl_2), W126-W131. https://doi.org/10.1093/nar/gkm219

Raj, S. R. G., & Nadarajah, K. QTL and candidate genes: Techniques and advancement in abiotic stress resistance breeding of major cereals. International Journal of Molecular Sciences. (2022). 24(1), 6. https://doi.org/10.3390/ijms24010006

Rehbein, P., Berz, J., Kreisel, P., & Schwalbe, H. "CodonWizard"-An intuitive software tool with graphical user interface for customizable codon optimization in protein expression efforts. Protein Expression and Purification. (2019). 160, 84-93. https://doi.org/10.1016/j.pep.2019.03.018

Riggs, A. D., Itakura, K., Crea, R., Hirose, T., Kraszewski, A., Goeddel, D., Heyneker, H. L. (1980). Synthesis, cloning, and expression of hormone genes in Escherichia coli. Paper presented at the Proceedings of the 1979 Laurentian Hormone Conference. https://doi.org/10.1016/B978-0-12-571136-4.50014-0

Sarkar, S., Ganapathysubramanian, B., Singh, A., Fotouhi, F., Kar, S., Nagasubramanian,

K., . . . Krishnamurthy, A. Cyber-agricultural systems for crop breeding and sustainable production. Trends in Plant Science. (2024). 29(2), 130-149. https://doi.org/10.1016/j.tplants.2023.08.001

Schwartz, M., Lo, H., Zhou, J., Yang, P., & Liao, A. Gene Synthesis: A Cost-Effective Alternative to Traditional Molecular Cloning. Journal of Biomolecular Techniques: JBT. (2011). 22(Suppl), S31.

Şen, A., Kargar, K., Akgün, E., & Pınar, M. Ç. Codon optimization: a mathematical programing approach. Bioinformatics. (2020). 36(13), 4012-4020. https://doi.org/10.1093/bioinformatics/btaa248

Sequeira, A. F., Bras, J. L., Guerreiro, C. I., Vincentelli, R., & Fontes, C. M. Development of a gene synthesis platform for the efficient large scale production of small genes encoding animal toxins. BMC biotechnology. (2016). 16, 1-13. https://doi.org/10.1186/s12896-016-0316-3

Shanmugaraj, B., I. Bulaon, C. J., & Phoolcharoen, W. Plant molecular farming: A viable platform for recombinant biopharmaceutical production. Plants. (2020). 9(7), 842. https://doi.org/10.3390/plants9070842

Sharkey, C. M., Lekveishvili, M., de la Rosa, T., & Danskin, K. Enhancing Gene Synthesis Security: An Updated Framework for Synthetic Nucleic Acid Screening and the Responsible Use of Synthetic Biological Materials. Applied Biosafety. (2024). https://doi.org/10.1089/apb.2023.0036

Sharma, A., Kumar, V., Shahzad, B., Ramakrishnan, M., Singh Sidhu, G. P., Bali, A. S., . . . Khanna, K. Photosynthetic response of plants under different abiotic stresses: a review. Journal of Plant Growth Regulation. (2020). 39, 509-531. https://doi.org/10.1007/s00344-019-10018-x

Sharma, V. (2020). Bioinformatics and its applications in environmental science and health and its applications in other disciplines. In: Sambodhi.

Sijmons, P. C., Dekker, B. M., Schrammeijer, B., Verwoerd, T. C., Van Den Elzen, P. J., & Hoekema, A. Production of correctly processed human serum albumin in transgenic plants. Bio/technology. (1990). 8(3), 217-221. https://doi.org/10.1038/nbt0390-217

Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., & Govindarajan, S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC bioinformatics. (2006). 7, 1-8. https://doi.org/10.1186/1471-2105-7-285

Villalobos-López, M. A., Arroyo-Becerra, A., Quintero-Jiménez, A., & Iturriaga, G. Biotechnological advances to improve abiotic stress tolerance in crops. International Journal of Molecular Sciences. (2022). 23(19), 12053. https://doi.org/10.3390/ijms231912053

Webster, G. R., Teh, A. Y. H., & Ma, J. K. C. Synthetic gene design-The rationale for codon optimization and implications for molecular pharming in plants. Biotechnology and bioengineering. (2017). 114(3), 492-502. https://doi.org/10.1002/bit.26183

Young, L., & Dong, Q. Two-step total gene synthesis method. Nucleic acids research. (2004). 32(7), e59-e59.

https://doi.org/10.1093/nar/gnh058

Zhang, Y., Luo, M., Wu, P., Wu, S., Lee, T.-Y., & Bai, C. Application of computational biology and artificial intelligence in drug design. International Journal of Molecular Sciences. (2022). 23(21), 13568.

https://doi.org/10.3390/ijms232113568

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.-h., Fu, J., Liu, Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proceedings of the National Academy of Sciences. (2016). 113(41), E6117-E6125. https://doi.org/10.1073/pnas.1606724113

How to cite this article: Riasat, N., Sana, A., Khan, M.H., Kabeer, T., Sehar, L., Khan, M.T., & Arshad, M. Designing of smart gene resources and computational approaches for sustainable environment: Opportunities and future challenges. *Journal of Soil, Plant and Environment* (2024) 3(2), 86–104.