Dsembler – DNA Assembly Designer: A Design tool for Facilitating Gene Assembly

Gyeongmin Park[1,2,‡], Aporva Gupta[1,2,‡], A-young Park[1,3], Eugene Rha[1], Dae-Hee Lee[1,2], Seung-Goo Lee[1,2,*], Haseong Kim[1,2,*]

[1]Synthetic Biology and Bioengineering Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea, [2]Department of Biosystems and Bioengineering, KRIBB School of Biotechnology, University of Science and Technology, Daejeon, 34113, Republic of Korea. [3]Department of Chemical Engineering and Applied Chemistry, Chungnam National University, Daejeon , 34134, Republic of Korea

*To whom correspondence should be addressed.

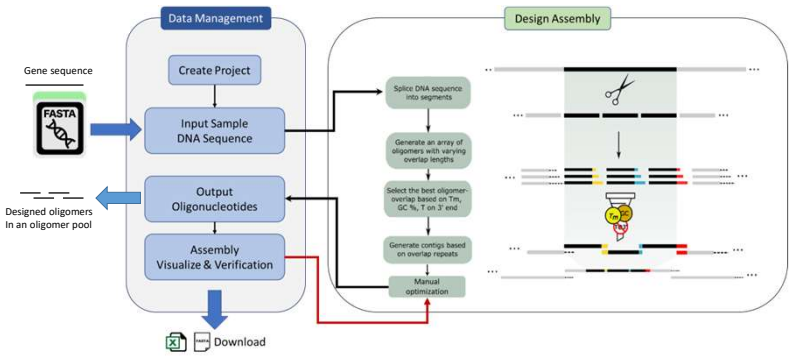
‡These authors contributed equally to this work.

Abstract

The field of synthetic biology has been gaining global interest due to various applications in bio-based production, bio-sensing, living therapeutics, and drug delivery. The increased interest has amplified the demand for long DNA synthesis and ultimately genome assemblies. However, gene assembly of oligomers is challenging since oligomers can be incorrectly assembled to other oligomers, self-anneal, or generate insertions or deletions due to mis-annealed oligomers. Dsembler (DNA Assembly Designer) is a web-based software application that aims to design oligomers for gene assembly. It designs optimized oligomer pool by automatically compute possible combinations of oligomers in the pool. It also provides UI that optimizing Tm between oligomers by manually. The final output sequences for both linear and circular DNA assembly are provided in Excel or FASTA file formats. The efficacy of Dsembler was demonstrated by assembling various length of genes.

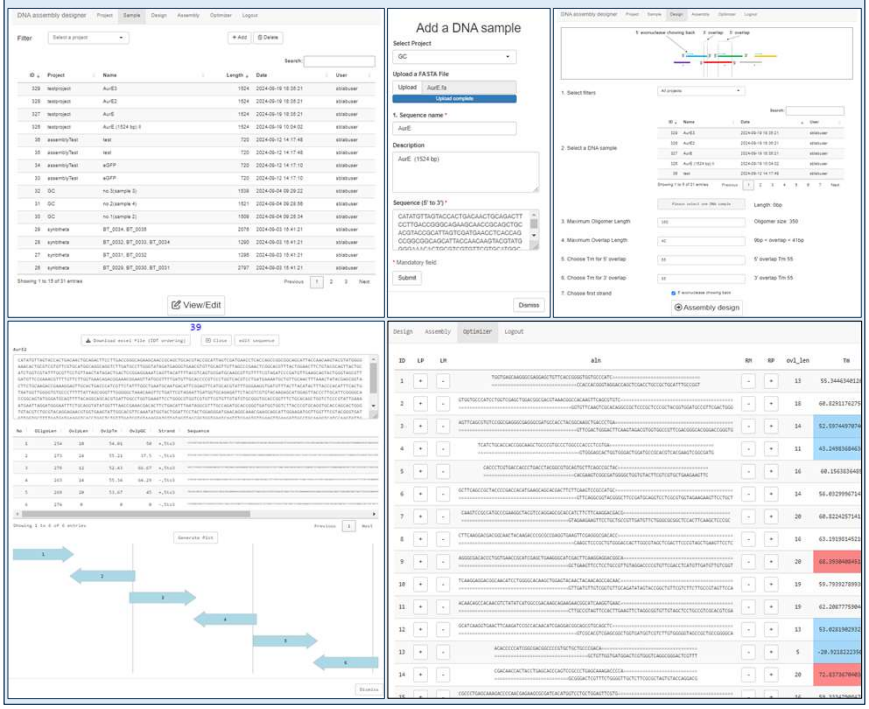
Introduction

Synthetic biology involves designing organisms by assembling short oligonucleotides to create artificial genomes. Despite advancements in gene synthesis and computational power, comprehensive tools that optimize DNA assembly design remain lacking. Here, we introduce **Dsembler (DNA Assembly Designer)**, a novel software tool designed to de novo DNA assembly for the use of PCA and Gibson methods. **Dsembler** is accessible at http://223.130.146.86:8088/.



Results

Dsembler provides an intuitive framework for managing synthetic biology projects by organizing data through a hierarchical structure of projects, samples, and assembly units. Within this system, users can manage their DNA data on a project level, then move down to upload and organize individual DNA samples. At the sample level, users can input synthesis conditions and perform DNA assembly, fine-tuning parameters like Tm, oligomer length, and overlap length. Once the assembly is complete, the results can be visualized and further optimized. This structured approach offers flexibility, allowing users to easily manage and track their experiments while efficiently optimizing DNA synthesis strategies for each project.



Methodology

Dsembler generates optimized assembly oligonucleotides by dividing a target DNA sequence into smaller fragments based on user-defined parameters such as oligonucleotide size, overlap length, GC content, and melting temperature (Tm). It carefully designs overlaps between adjacent oligonucleotides to ensure stability, calculating the GC content and Tm to ensure that both fall within the specified range. The tool also avoids certain problematic features, like thymine at the 3' end, which can reduce binding efficiency. To further enhance accuracy, Dsembler assigns a penalty score to each oligonucleotide, reflecting potential issues such as mismatched Tm values, insufficient GC clamps, and sequence composition. This scoring system enables users to identify and adjust any problematic oligonucleotides, ensuring that the final design is optimized for successful DNA assembly. By allowing users to refine their design based on detailed feedback, Dsembler streamlines the experimental design process and reduces the likelihood of assembly errors.

$$score = |T_t - T_c| + 1_T + GC_{clamp} + R_{within} + R_{between}$$

Evaluation

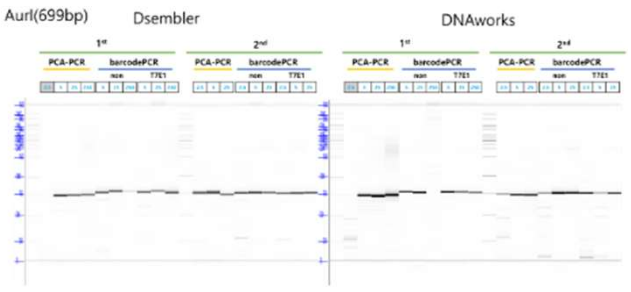
The evaluation of Dsembler was conducted by comparing it to manually designed oligonucleotides and GeneDesign using a 520bp fragment of the M13 bacteriophage genome. The key metrics assessed were the error rates (insertions and deletions per kilobase) and the proportion of error-free colonies after DNA assembly. Dsembler outperformed both GeneDesign and manual design, with no insertions and a significantly lower deletion rate (0.38 ± 0.5 errors/kbp) compared to GeneDesign (2.02 ± 1.0 errors/kbp) and manual design (0.67 ± 0.6 errors/kbp). Additionally, 83% of colonies assembled with Dsembler oligos were error-free, which was higher than both manually designed oligos (71%) and GeneDesign (33%). This indicates that Dsembler offers more accurate oligonucleotide design, reducing the need for error correction during assembly.

Table 1: Oligomer Design Analysis

Design Method	Insertion Rate	Deletion Rate	Total Error Rate	Error Free colonies
Dsembler	0	0.38 ± 0.5	0.38 ± 0.5	83
GeneDesign	1.13 ± 0.7	0.9 ± 0.7	2.02 ± 1.0	33
Manual	0.40 ± 0.6	0.27 ± 0.4	0.67 ± 0.6	71

The error rates (errors/kbp) and proportion of error-free colonies (%) of the 520bp fragment assemblies with oligomers designed manually and by Dsembler, GeneDesign.

Dsembler was also compared to DNAworks in the evaluation. Both tools were used to design fragments of the AurI gene from *Bacteroides thetaiotaomicron* for synthesis and assembly using PCA. The analysis showed that fragments designed by both Dsembler and DNAworks were successfully synthesized and assembled, with no significant differences in the quality or accuracy of the final assembly products. The gene lengthes are ranging from 1~10kbp with maximum oligomer length is 250nt.



Discussion

Dsembler's error-reducing algorithm demonstrated superior performance compared to GeneDesign and manual oligo design by significantly lowering insertion and deletion rates and producing a higher proportion of error-free colonies. This reduction in errors would reduce the need for additional corrective steps during DNA assembly, thus speeding up the assembly process, particularly for long sequences. Additionally, Dsembler's ability to identify potential assembly errors in advance allows users to refine their experimental conditions, contributing to more efficient and reliable DNA synthesis. When compared to DNAworks, Dsembler performed similarly in terms of synthesis and assembly accuracy, which suggests it is a viable tool for large-scale or high-throughput projects. The authors also suggest future improvements for Dsembler, such as incorporating features for codon optimization, secondary structure detection, and insert-backbone-based assembly, which would further enhance its utility in synthetic biology applications.

Acknowledgement

This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ015808022021)" Rural Development Administration, Republic of Korea, the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (2018M3A9H3024746, 2021M3A9I4022731), and the Korea Research Institute of Bioscience and Biotechnology Research (KRIBB) Research Initiative Program (KGM5402113).

References

(1) Hughes,R.A., and Ellington,A.D. (2017) Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. CSH Perspect. Biol. 9(1), a023812.
(2) Hendling,M. and Barišić. (2019) In-silico Design of DNA Oligonucleotides: Challenges and Approaches. Comput. Struct. Biotechnol J.,17:1056-1065.
(3) Richardson,S.M. et al. (2006). GeneDesign: Rapid, automated design of multikilobase synthetic genes, Genome Res, 16): 550-556.
(4) Richardson,S.M. et al. (2010) GeneDesign 3.0 is an updated synthetic biology toolkit. Nucleic Acids Res, 38, 2603-6.
(5) Villalobos,A. et al. (2005) Gene Designer: a synthetic biology tool for constructing artificial DNA egments. BMC Bioinform, 7, 285.
(6) SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs., Annu. Rev. Biophys. Biomol. Struct., 33, 415–440.
(7) Panjkovich,A. and Melo,F. (2005) Comparison of different melting temperature calculation methods for short DNA sequences. Bioinformatics, 21,711-722.