Synthetic Biology, 2020, 5(1): ysaa016

doi: 10.1093/synbio/ysaa016
Advance Access Publication Date: 19 August 2020
Synthetic Biology News

Designing for durability: new tools to build stable, non-repetitive DNA

The survival of genetic information hinges on identifying repetition. Genomes are repaired by mechanisms such as homologous recombination, in which matching DNA sequences are used as a template to replace missing information. This strategy works provided sequences in the genome are mostly unique. While sequence diversity has kept genomes stable enough to replicate for millions of years, it poses a problem for those trying to engineer DNA (1). After all, one of the central tenets of synthetic biology is the reutilization of standard parts.

How, then, can we design stable, non-repetitive genetic systems with a limited toolkit of synthetic parts? Researchers in Howard Salis's lab at Pennsylvania State University set out to address this challenge through the Non-Repetitive Parts Calculator (NRPC), a set of new algorithms described in a recent publication by Hossain *et al.* (2) and available online (https://salislab.net/software/).

As the name implies, NRPC builds collections of biological parts containing minimal repetitive sequences, where the repetitiveness of a collection is defined by $L_{\max}$, the maximum length of the longest shared repeat. Collections can be created using two different modes.

The 'Finder' mode determines the largest subset of non-repetitive elements within any given database of parts, given a maximum $L_{\max}$ set by the user. The sheer number of possible subsets to evaluate can make this computationally impractical for large libraries. The authors solve this problem by representing parts as nodes on a graph and improving on existing algorithms in graph theory to efficiently maximize the number of disconnected components.

The 'Maker' mode creates a new library of non-repetitive parts within the design constraints set by the user, which may include a degenerate DNA sequence or RNA structure template and a set $L_{\max}$ value. In this case, all possible sequences are represented as a decision tree and hash tables are used to store and check for occurrences of sub-sequences within parts.

Hossain *et al.* tested their new 'Maker' algorithm by generating libraries of 4350 synthetic, non-repetitive bacterial promoters and 1722 yeast promoters, designed to have a wide range of transcription rates. The authors validated each library's predicted transcriptional behavior by assembling and characterizing every promoter through next-generation DNA and RNA sequencing in *Escherichia coli* and *Saccharomyces cerevisiae*.

The increased stability of NRPC designs was demonstrated in *E. coli* by comparing versions of a construct with either repetitive or non-repetitive promoters. The former rapidly lost fluorescence and DNA content while the latter remained stable. Finally, the authors applied regression models and neural networks developed elsewhere (3) to explain and predict the strength of the synthetic promoters they created.

This work can have tremendous, immediate impact in two ways. Not only did Hossain *et al.* produce vast libraries of bacterial and yeast promoters with known expression profiles and improved compatibility, but they also published software for researchers to design their own stable libraries for many different applications. This opens the question of what threshold of repetitiveness, whether measured as $L_{\max}$ or with another metric, should be used in a given organismic context. Regardless, NRPC is noteworthy for tackling a pervasive problem in synthetic biology, one seemingly at odds with the principles of the field.

References

1. Jack,B.R., Leonard,S.P., Mishler,D.M., Renda,B.A., Leon,D., Suárez,G.A. and Barrick,J.E. (2015) Predicting the genetic stability of engineered DNA sequences with the EFM calculator. *ACS Synth. Biol.*, 4, 939–943.
2. Hossain,A., Lopez,E., Halper,S.M., Cetnar,D.P., Reis,A.C., Strickland,D., Klavins,E. and Salis,H.M. (2020) Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nat. Biotechnol.* doi:10.1038/s41587-020-0584-2.
3. de Boer,C.G., Vaishnav,E.D., Sadeh,R., Abeyta,E.L., Friedman,N. and Regev,A. (2020) Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.*, 38, 56–65.

Pablo Cárdenas

Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: E-mail: pcarden@mit.edu

Submitted: 8 August 2020; **Accepted:** 11 August 2020