

## ДНК-ТЮНЕР – ПРОГРАММА ДЛЯ КОНСТРУИРОВАНИЯ ПОЛИЛИНКЕРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ МОЛЕКУЛЯРНЫХ ВЕКТОРОВ

В.С. Латынский, Л.Н. Валентович\*

ООО "Белка Технолоджис", Минск, Республика Беларусь

\*ГНУ «Институт микробиологии НАН Беларуси», Минск, Республика Беларусь,

e-mail: valentovich@mbio.bas-net.by

### Введение

Методы конструирования гибридных молекул ДНК *in vitro* основаны на использовании вспомогательных векторных молекул, с помощью которых можно размножить целевой фрагмент ДНК, переместить его в нужную клетку, провести направленный мутагенез и ряд других манипуляций [1]. В качестве векторных молекул в генетической инженерии используют широкий спектр плазмидных и вирусных ДНК, которые в зависимости от функции разделяют на клонирующие, экспрессирующие, интегративные векторы [2].

Наиболее распространенным методом вставки целевой последовательности ДНК в векторную молекулу является рестриктазно-лигазный метод, предложенный Стэнли Коэном с сотрудниками ещё в 1973 году [3], особенно после разработки и улучшения технологии химического синтеза олигонуклеотидных молекул [4], которая позволила целенаправленно вводить в векторные молекулы синтетические участки ДНК, имеющие последовательности, узнаваемые рестриктазами [5; 6]. Появились векторы, содержащие в своём составе полилинкеры (или MCS – от английского «multiple cloning site» – место множественного клонирования) – короткие искусственные нуклеотидные последовательности, содержащие несколько уникальных мест узнавания для эндонуклеаз рестрикции, по которым удобно вставлять чужеродные фрагменты ДНК.

Первое поколение векторов с синтетическими полилинкерами содержало всего несколько, не всегда уникальных, мест узнавания рестриктазами [7]. MCS этих векторов не в полной мере удовлетворяли запросам экспериментаторов и, в связи с этим, шло активное конструирование подобных плазмидных векторов, несущих или другую последовательность мест узнавания тех же рестрикционных эндонуклеаз или дополнительные, ранее не использовавшиеся места узнавания. В 80-х годах 20-го века были сконструированы классические векторы с приемлемыми полилинкерами для общих целей клонирования – семейства pUC [8], pMTL [9], pBlueScript [10]. Однако работа над созданием улучшенных MCS продолжалась далее и велась в направлении конструирования данных участков со всевозможными вариантами мест узнавания для рестриктаз – были созданы так называемые суперполилинкеры, которые содержат все 64 возможных варианта гексануклеотидных палиндромных последовательностей для узнавания рестрикционными эндонуклеазами [11].

Тем не менее, разработка каждого нового полилинкера до последнего времени была очень дорогостоящим и трудозатратным мероприятием, поэтому при конструировании векторов обычно использовались полилинкерные области уже известных готовых плазмид, фазмид, фагов [12]. Только в последнее десятилетие с удешевлением химического синтеза нуклеиновых кислот и развитием методов сборки ДНК [13] для большинства исследователей появилась возможность сконструировать вектор *de novo* и продумать роль каждого нуклеотида. Возникли такие доступные проявления синтетической биологии, как Биоконструктор (BioBrick) [14; 15], заказ синтеза отдельного гена [16], или даже целой плазмиды [17]. Однако до сих пор нет ни одной программы для дизайна полилинкера с оптимизированной последовательностью, несмотря на востребованность таких приложений [18; 19]. Под оптимизацией в данном случае понимается уменьшение длины последовательности MCS при сохранении всех нужных мест узнавания эндонуклеаз рестрикции за счёт удаления одинаковых нуклеотидов или олигонуклеотидов в местах

стыковки структурных элементов полилинкера, что в итоге приводит к уменьшению стоимости работ.

Программа, описанная в данной статье, была создана с целью автоматизации и значительного упрощения конструирования полилинкерных последовательностей векторных молекул.

### **Описание алгоритма работы и его реализация**

Программа «ДНК-тюнер» – это веб-приложение, написанное с использованием языка программирования Java [20] и фреймворка Spring boot [21]. Пользовательский интерфейс реализован с помощью JavaScript [23] и языка разметки HTML [22]. «ДНК-тюнер» работает на виртуальном выделенном сервере (512 Мб ОЗУ) под управлением операционной системы Debian [24]. Программа доступна по ссылке <http://www.bio.bsu.by/molbiol/dna-tuner/> или по служебному адресу <http://95.85.2.49:9000>.

Для своей работы приложение использует базу данных REBASE [25]. На вход программе дается список подстрок – нуклеотидных последовательностей. Поиск минимальной строки, содержащей заданные подстроки, осуществляется следующим алгоритмом: генерируются все перестановки подстрок, и для каждой перестановки выходная строка получается последовательной конкатенацией (с возможным перекрытием) всех подстрок, при этом на каждом шаге конкатенации получается минимально возможная строка. Из строк полученных данным образом в свою очередь выбираются строка (либо строки) минимальной длины.

Вышеописанный алгоритм применяется, если на вход дается 9 либо меньше подстрок. В случае если на вход подается 10 либо более подстрок, то все перестановки не генерируются, а используется только одна. Это обусловлено тем, что сложность описанного алгоритма  $O(M \times N!)$ , где  $M$  – сумма длин подстрок,  $N$  – число подстрок, что уже для 10 подстрок делает время работы программы неприемлемо долгим для пользователя (более 2 минут).

Для удобства пользователя итоговая последовательность транслируется в трех рамках считывания (+1,+2,+3) с использованием стандартной таблицы генетического кода и анализируется на наличие мест узнавания эндонуклеазами рестрикции.

### **Интерфейс программы**

Для запуска и работы программы «ДНК-тюнер» необходим любой компьютер с установленным веб-обозревателем и доступом к сети Интернет.

При загрузке страницы отображается:

один пункт меню «Paste»;

окно для ввода данных с примером записи;

окно-справка с информацией об эндонуклеазах рестрикции и последовательностях ДНК, которую они узнают;

поле с заголовком «Result:» для вывода результата работы программы (рисунок 1).

«ДНК-тюнер» работает в двух режимах: либо данные вводятся в виде названий рестриктаз (пункт меню «Paste Names of RE») или в виде любых последовательностей ДНК (пункт «Paste DNA Sequences»). Каждое название эндонуклеазы рестрикции или последовательность ДНК должны вводиться с новой строки. Буквенные коды для обозначения нуклеотидных оснований должны соответствовать номенклатуре, принятой Международным обществом теоретической и прикладной химии (IUPAC).

Под справочным окном есть кнопка «Add», с помощью которой можно перенести выбранную(ые) рестриктазу(ы) в окно ввода данных. Также присутствует меню сортировки информации внутри справочного окна: можно отобразить только эндонуклеазы рестрикции, образующие после разрезания «тупые» концы ДНК или дающие выступающие 3' или 5' концы; можно выбрать рестриктазы, узнающие определенное количество нуклеотидов (4-9 и более).

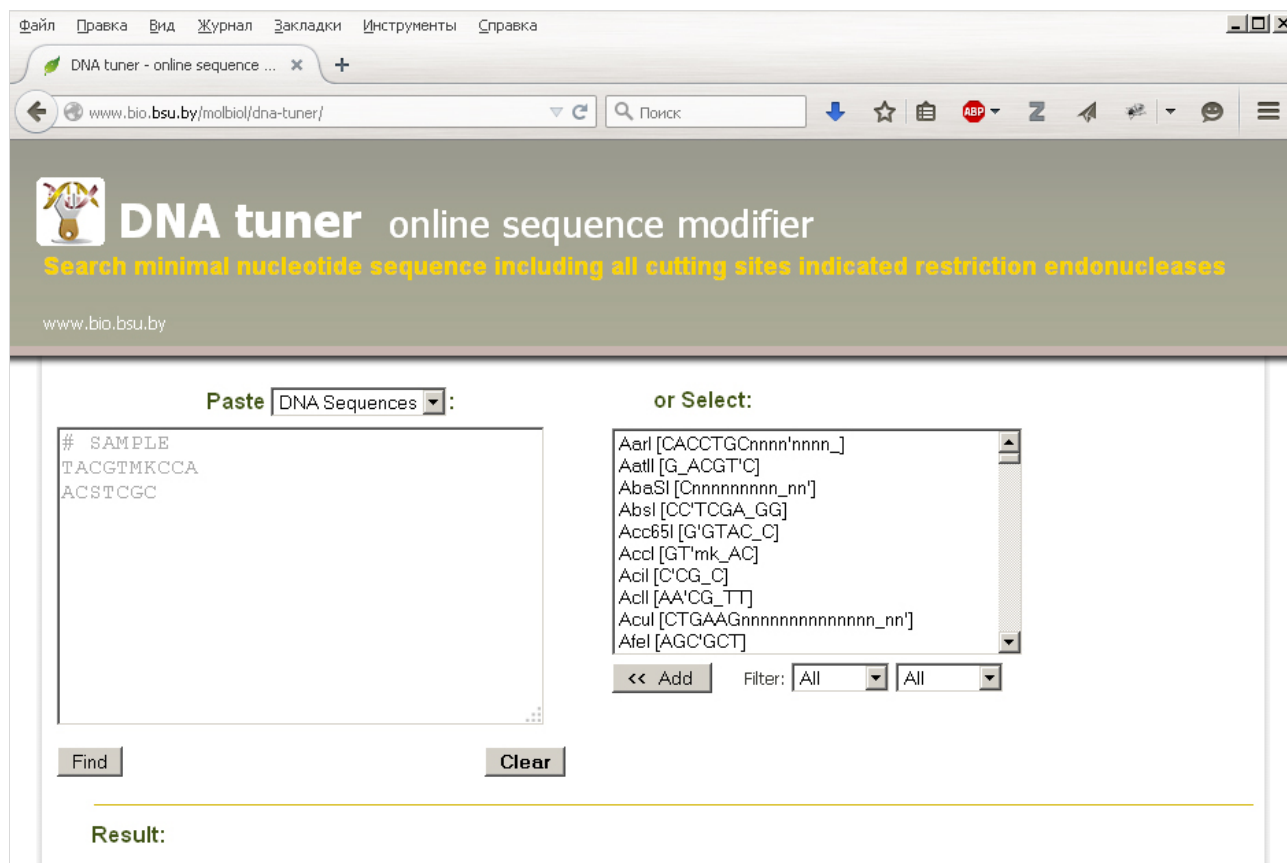


Рисунок 1 – Вид англоязычного интерфейса программы ДНК-тюнер

После ввода данных работа программы инициируется нажатием кнопки «Find».

Чтобы очистить окно для ввода новых данных, можно нажать на кнопку «Clear».

Если были введены некорректные данные, то под кнопкой «Find» отобразится сообщение об ошибке и комментарий. В случае отсутствия ошибки после заглавия «Result:» выведется оптимизированная нуклеотидная последовательность (или несколько последовательностей) и информация о ней. Если кликнуть на ссылку «more» справа от результата, то каждую последовательность ДНК можно отобразить в виде текстовой карты с указанием аминокислотных последовательностей после трансляции и мест узнавания рестриктазами.

#### Основные и вспомогательные функции программы

Предполагается, что пользователь «ДНК-тюнера» заранее определится с набором рестриктаз, места узнавания которых должны присутствовать в полилинкере и решит, будут ли в MCS отсутствовать стоп-кодоны в нужной рамке считывания. Далее он заносит названия рестриктаз или функциональные последовательности ДНК в окно ввода данных и на выходе получает оптимизированную последовательность (или несколько равнозначных). В результате указывается процент оптимизации и количество сокращённых нуклеотидов. Оценить полученный результат можно с помощью текстовой карты (рисунок 2). Например, оптимизированная последовательность для полилинкера с 10 местами узнавания эндонуклеаз рестрикции популярного вектора pUC18 (HindIII, SphI, PstI, SalI, XbaI, BamHI, SmaI, KpnI, SacI, EcoRI) составляет всего 50 нуклеотидов, что на 10 меньше, чем получается при простом объединении последовательностей, узнаваемых указанными рестриктазами. При этом полученная последовательность в рамке считывания +1 и +2 не содержит стоп-кодонов и вполне возможно, что она бы не нарушала способность модифицированного *lacZ'*-гена к  $\alpha$ -комплементации  $\beta$ -галактозидазы в клетках бактерий.

**Result:**

Variants: 1 Optimization: 10 bp [16%]

1. AAGCTTGGTACCCGGGATCCGAGCTCTAGAATTTCGCATGCTGCAGTCGAC [close]

```

      A W Y P G S E L * N S H A A V D
      S L V P G I R A L E F A C C S R
      K L G T R D P S S R I R M L Q S
1 AAGCTTGGTACCCGGGATCCGAGCTCTAGAATTTCGCATGCTGCAGTCGAC 50
HindIII  FspEI  FspEI  ApoI  FatI  HincII
AluI     XmaI  Hpy188I  EcoRI  TseI  TaqI
SetI     BsaJI  AbaSI  MluCI  BlnI
AbaSI    FspEI  AbaSI  AbaSI  SalI
SgeI     SmaI  MspJI  MspJI  AccI
MspJI    HpaII  SgeI   SphI   Hpy166II
BanI     AluI  AluI   NspI
KpnI     BstYI  AbaSI  Cac8I
Acc65I   BstKTI  Hpy188III  BisI
NlaIV    BamHI  XbaI   NlaIII
CviQI    MboI   BfaI   FaiI
RsaI     NlaIV  AbaSI  CviAII
          StyD4I  MspJI  AbaSI
          NciI   Eco53kI  Fnu4HI
          AbaSI   SacI     PstI
          AvaI   Bsp1286I  SfcI
          SgeI   BsiHKAI  HpyCH4V
          ScrFI  BanII
          MspJI   SetI
          StyD4I  SgeI
          NciI   MspJI
          AbaSI
          SgeI
          ScrFI
          MspJI
          LpnPI
          AbaSI
          SgeI
          MspJI
          DpnI

```

Рисунок 2 – Так бы мог выглядеть полилинкер плазмиды pUC18

В качестве дополнительных функций, с которыми справляется «ДНК-тюнер» можно выделить следующие.

1) Возможность сборки нескольких фрагментов ДНК в один протяжённый контиг. Данная операция работает с определёнными ограничениями – неточность даже одного нуклеотида приведёт к отсутствию ожидаемого результата. Также важно, чтобы количество фрагментов не превышало несколько сотен, что вполне пригодно для сборки ДНК после секвенирования с помощью методов первого поколения.

2) При вводе только одной последовательности ДНК программа «ДНК-тюнер» считает её длину, выводит статистическую информацию о ГЦ-составе и процентном содержании каждого нуклеотида, а также транслирует последовательность в трёх рамках считывания.

3) «ДНК-тюнер» можно использовать в качестве простейшей программы для рестрикционного анализа последовательностей ДНК.

4) Информационное окно с указанием рестриктаз и соответствующих им последовательностей ДНК можно использовать как справочник, т.к. названия эндонуклеаз

рестрикции расположены в алфавитном порядке и подвержены сортировке исходя из их свойств.

### Выводы

На данный момент «ДНК-тюнер» является уникальной высокоспециализированной программой, которая может занять свою нишу в некоторых приложениях синтетической биологии. А за счёт выполнения ряда операций по базовому анализу нуклеотидных последовательностей, она может стать простейшим инструментом-справочником исследователей, работающих в области молекулярной биологии, генетики, биохимии.

### Список литературы

1. Current Protocols in Molecular Biology / eds. Ausubel, F.M. [et al.]. – New York: John Wiley & Sons, Inc., 2003. – 1600 p.
2. Щелкунов С.Н. Генетическая инженерия / Щелкунов С.Н. – Новосибирск: Сиб. унив. изд-во, 2004. – 496 с.
3. Construction of Biologically Functional Bacterial Plasmids In Vitro / S.N. Cohen [et al.] // Proceedings of the National Academy of Sciences. – 1973. – Vol. 70. – № 11. – P. 3240–3244.
4. A Modified Triester Method for the Synthesis of Deoxyribopolynucleotides / K. Itakura [et al.] // Canadian Journal of Chemistry. – 1973. – Vol. 51. – № 21. – P. 3649–3651.
5. Bahl C.P. A general method for inserting specific DNA sequences into cloning vehicles / C.P. Bahl, K.J. Mariani, R. Wu // Gene. – 1976. – Vol. 1. – № 1. – P. 81–92.
6. Chemical synthesis of restriction enzyme recognition sites useful for cloning / R.H. Scheller [et al.] // Science. – 1977. – Vol. 196. – № 4286. – P. 177–180.
7. Messing J. A system for shotgun DNA sequencing / J. Messing, R. Crea, P.H. Seeburg // Nucleic Acids Research. – 1981. – Vol. 9. – № 2. – P. 309–321.
8. Yanisch-Perron C. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors / C. Yanisch-Perron, J. Vieira, J. Messing // Gene. – 1985. – Vol. 33. – Improved M13 phage cloning vectors and host strains. – № 1. – P. 103–119.
9. The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing / S.P. Chambers [et al.] // Gene. – 1988. – Vol. 68. – № 1. – P. 139–149.
10. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. / J.M. Short [et al.] // Nucleic Acids Research. – 1988. – Vol. 16. – Lambda ZAP. – № 15. – P. 7583–7600.
11. Brosius J. Superpolylinkers in cloning and expression vectors / J. Brosius // DNA (Mary Ann Liebert, Inc.). – 1989. – Т. 8. – № 10. – С. 759–777.
12. The Cloning Vector Collection (National Institute of Genetics, Japan) [Электронный ресурс]. – Режим доступа: <http://www.shigen.nig.ac.jp/ecoli/strain/cvector/cvectorExplanation.jsp>.
13. DNA Cloning and Assembly Methods : Methods in Molecular Biology. Т. 1116 / ред. S. Valla, R. Lale. – Totowa, NJ: Humana Press, 2014.
14. Main Page - parts.igem.org [Электронный ресурс]. – Режим доступа: [http://parts.igem.org/Main\\_Page](http://parts.igem.org/Main_Page).
15. Shetty R.P. Engineering BioBrick vectors from BioBrick parts / R.P. Shetty, D. Endy, T.F. Knight // Journal of Biological Engineering. – 2008. – Т. 2. – С. 5.
16. Gene Synthesis - DNA Synthesis - from \$0.23/bp - Largest Gene Synthesis Supplier - GenScript- Make Research Easy [Электронный ресурс]. – Режим доступа: [http://www.genscript.com/gene\\_synthesis.html](http://www.genscript.com/gene_synthesis.html).
17. GeneArt® Plasmid Services [Электронный ресурс]. – Режим доступа: <http://www.lifetechnologies.com/by/en/home/life-science/cloning/gene-synthesis/geneart-plasmid-services.html>.
18. Designing a multiple cloning site - Molecular Biology [Электронный ресурс]. – Режим доступа: <http://www.protocol-online.org/biology-forums-2/posts/20775.html>.

19. Does anyone know of an easy way to design an MCS (Multiple Cloning... [Электронный ресурс]. – Режим доступа: [http://www.researchgate.net/post/Does\\_anyone\\_know\\_of\\_an\\_easy\\_way\\_to\\_design\\_an\\_MCS\\_Multiple\\_Cloning\\_Site](http://www.researchgate.net/post/Does_anyone_know_of_an_easy_way_to_design_an_MCS_Multiple_Cloning_Site).
20. Java SE - Downloads | Oracle Technology Network | Oracle [Электронный ресурс]. – Режим доступа: <http://www.oracle.com/technetwork/java/javase/downloads/index.html?ssSourceSiteId=ocomen>.
21. Spring Boot [Электронный ресурс]. – Режим доступа: <http://projects.spring.io/spring-boot/>.
22. HTML 4.01 Specification [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/html401/>.
23. ECMAScript Language Specification - ECMA-262 Edition 5.1 [Электронный ресурс]. – Режим доступа: <http://www.ecma-international.org/ecma-262/5.1/>.
24. Debian – Универсальная Операционная Система [Электронный ресурс]. – Режим доступа: <https://www.debian.org/index.ru.html>.
25. REBASE – a database for DNA restriction and modification: enzymes, genes and genomes / R.J. Roberts [et al.] // Nucleic Acids Research. – 2010. – Vol. 38. – № Database issue. – P. D234–236.