

Моделирование поиска праймеров в цепи ДНК

О.Ю. Кирьянова¹, А.В. Чемерис²

¹Уфимский государственный нефтяной технический университет, Космонавтов 1, Уфа, Россия, 450062

²Институт биохимии и генетики УФИЦ РАН, пр. Октября 71, Уфа, Россия, 450054

Аннотация. Полимеразная цепная реакция (ПЦР) является одним из самых распространенных экспериментальных методов при решении задач анализа ДНК. Олигонуклеотидные праймеры являются важной составляющей любой ПЦР, и поэтому существует ряд требований к их дизайну. От данных структур зависит успешность и возможность проведения успешного эксперимента в целом. В связи с этим появилась необходимость проведения компьютерного анализа подбора праймеров. Разработан алгоритм на основе алгоритма поиска Бойера-Мура для специфичного дизайна праймеров. В данной работе представлены две вариации поиска праймеров. На основе алгоритма разработана программа, позволяющая проводить компьютерный дизайн праймеров перед непосредственным экспериментальным проведением ПЦР. Что значительно упрощает проведение непосредственного эксперимента. На данный момент расчеты проведены для модельных объектов (хромосом арабидопсиса).

1. Введение

Олигонуклеотидные праймеры являются важной составляющей ПЦР, потому что именно от них зависит специфичность и эффективность реакции, а также возможность протекания реакции при наличии всех остальных компонентов. Специфичность ПЦР основана на образовании комплементарных комплексов между матрицей и праймерами, короткими синтетическими олигонуклеотидами длиной от 10 до 30 оснований. Каждый из праймеров комплементарен одной из цепей двуцепочечной матрицы и ограничивает начало и конец амплифицируемого участка. От состава праймеров зависит температура проведения реакции. Поэтому существует целый ряд требований к подбору последовательностей нуклеотидов в праймерах и их протяженности в зависимости от решаемой задачи, объекта эксперимента. Существует довольно большое количество программных продуктов, позволяющих проводить дизайн праймеров. Однако, не всегда в них возможно проводить специфичный поиск, корректировать параметры, начальные условия. Поэтому была разработана программа, позволяющая проводить дизайн праймеров с возможностью введения более жестких условий на искомые праймеры, их местоположение в геноме. Что позволяет неоднократно проводить компьютерный анализ в нескольких вариациях для определения наиболее благоприятных условий проведения ПЦР.

2. Объект исследования

Полимеразная цепная реакция (ПЦР) является одним из самых распространенных экспериментальных методов при решении задач анализа ДНК. Данный метод генной

диагностики широко применяется в различных областях биологии, медицины. Суть данного анализа заключается в многократном увеличении содержания за счет применения особых ферментов, которые многократно копируют фрагменты, характерные для конкретных геномов. ПЦР проводится в три этапа. Первый этап – денатурация, расхождение двух цепей ДНК (температура 94-96 °С). Второй этап – отжиг. Когда цепи разошлись, температуру понижают, чтобы праймеры могли связаться с одноцепочечной матрицей. Далее происходит репликация (синтез дочерней молекулы ДНК), где праймер используется в качестве затравки [1].

Олигонуклеотидные праймеры (искусственно созданные олигонуклеотиды, которые занимаются поиском нужного фрагмента ДНК) являются важной составляющей ПЦР, потому что именно от них зависит специфичность и успешность реакции, а также возможность протекания реакции при наличии всех остальных компонентов [2]. Для непосредственного проведения экспериментальных исследований важно провести предварительный компьютерный анализ. Существует довольно большое количество программных продуктов, которые осуществляют дизайн олигонуклеотидных праймеров [3]. Однако, не всегда возможно в данных программах проводить более специфический дизайн праймеров, решать небольшие рутинные подзадачи.

Одной из задач компьютерного анализа является поиск и определение локализации коротких последовательностей (праймеров) с точностью до нуклеотида. Это необходимо для того, чтобы определить возможные места отжига праймеров, их количество. Фактически задача похожа на поиск слова (длиной порядка 10-20 символов) в некотором тексте (объемом до 1 млрд символов). В данном случае алфавитом являются следующие символы: А, G, С, Т (нуклеотиды аденин, гуанин, цитозин, тимин, из которых состоит молекула). Важно не только определить местонахождение «слова», но и количество его повторов.

Поиск коротких последовательностей имеет два предназначения. Первый – для проведения случайной ПЦР. В данном случае важно знать, сколько раз встречается праймер на всей цепи ДНК и на какой позиции он находится. В этом случае рассматриваемая длина коротких последовательностей колеблется в диапазоне от 8 до 25 нуклеотидов. Чем чаще встречается искомый участок, тем выше вероятность проведения успешного эксперимента. Схема поиска представлена на рисунке 1.

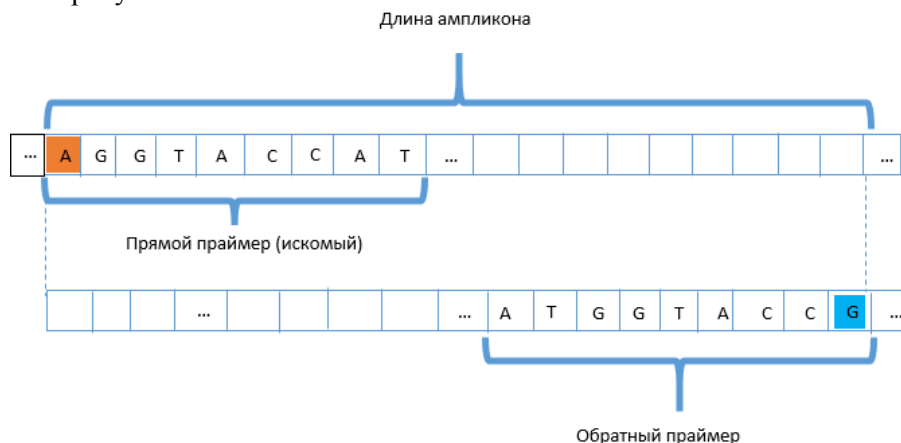


Рисунок 1. Поиск праймеров для проведения случайной ПЦР.

Второе приложение такой поиск будет иметь для обнаружения мест отжига чуть более длинных праймеров, которые должны в среднем встречаться каждые 16 млн. нуклеотидов. В этом случае будет интересовать последовательность нуклеотидов после таких мест, исходя из предположения, что в реакцию амплификации будет браться три, а не четыре нуклеотида, и на отсутствующем будет происходить терминация (завершение синтеза) цепи. В этом случае нас интересует не только положение праймера, но и смежные с ним участки ДНК. Что характерно, для данного метода расчета, это то, что прямой и обратный праймер принимаются равнозначными. И поэтому для обоих проводится поиск места отжига. Схема данного поиска представлена на рисунке 2. Поэтому форма представления таких результатов должна учитывать

те места отжига, где достаточно протяженное расстояние не будет встречаться конкретный нуклеотид (Например, гуанин G, как представлено на рисунке 2). При расчете важно знать длину таких участков, а также суммарную молекулярную массу составляющих их нуклеотидов.

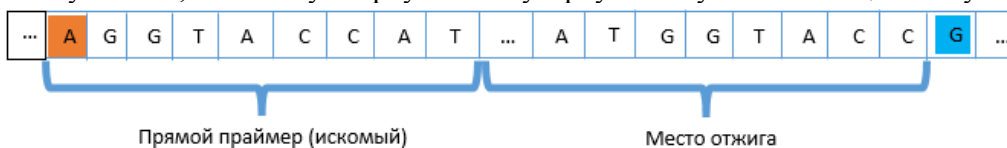


Рисунок 2. Поиск праймеров и мест отжига в последовательности нуклеотидов.

Для решения вышеописанных задач разработан алгоритм поиска коротких праймеров в ДНК-цепи на основе алгоритма Бойера-Мура [4] с добавлением дополнительных условий на выбор праймеров. Данный алгоритм позволяет найти включения конкретных фрагментов последовательностей в цепи ДНК с точностью до нуклеотида, а также определить состав и размер ампликонов. В процессе поиска происходит выбор множества праймеров с учетом заявленных требований. На основе алгоритма создана программа, позволяющая проводить дизайн праймеров.

Программа написана на языке программирования Python 3 с применением библиотеки BioPython [5]. В данной библиотеке собраны инструменты, применяемые для расчетов в области вычислительной биологии и биоинформатики. Кроме того, инструменты библиотеки позволяют работать с файлами в fasta-формате (текстовый формат для нуклеотидных или полипептидных последовательностей, в котором нуклеотиды или аминокислоты обозначаются при помощи однобуквенных кодов) [6].

Результаты работы программы представляются в виде таблицы, представленной на рисунке 3.

	A	B	C	D
1	#GGATCTTT	AAAGATCC	length of amplicon	
2	39883835	39884052	217	
3	55375264	55375548	284	
4	29569657	29569969	312	
5	38393029	38393375	346	
6	49519668	49520023	355	
7	41540764	41541163	399	
8	8231987	8232448	461	
9	37414390	37414862	472	
10	56234084	56234565	481	
11				

а)

	A	B	C	D	E	F	G
#GGATCTTTAC	last primer position	length before G	A	C	T	mass	
GTAAGATCC	1374796	6 3A	2C	1T		1822,2	
GTAAGATCC	45312016	6 2A	1C	3T		1828,2	
GGATCTTTAC	10248026	6 3A	1C	2T		1837,2	
GTAAGATCC	23115254	6 4A	1C	1T		1846,2	
GGATCTTTAC	25932850	6 3A	0C	3T		1852,2	
GGATCTTTAC	38805490	6 3A	0C	3T		1852,2	
GTAAGATCC	23553251	7 0A	2C	5T		2099,4	
GGATCTTTAC	29203996	7 2A	2C	3T		2117,4	
GTAAGATCC	51132579	7 2A	1C	4T		2132,4	
GGATCTTTAC	59725545	7 5A	2C	0T		2144,4	
GTAAGATCC	16126213	7 5A	1C	1T		2159,4	
GGATCTTTAC	61328473	8 1A	3C	4T		2397,6	
GTAAGATCC	22650242	10 1A	3C	6T		3006	
GGATCTTTAC	54848321	10 6A	2C	2T		3066	
GTAAGATCC	1734495	11 3A	2C	6T		3343,2	
GTAAGATCC	2526205	14 7A	4C	3T		4261,8	
GGATCTTTAC	23167659	17 3A	7C	7T		5093,4	

б)

Рисунок 3. Выходные результаты работы программы: а) поиск последовательности GGATCTTT (обратный праймер AAAGATCC) для анализа случайной ПЦР; б) поиск последовательности GGATCTTTAC (обратный праймер GTAAGATCC) для обнаружения мест отжига.

Довольно удобно, что в данной программе можно варьировать размер праймера, варьировать длину ампликона, изменять условия для места отжига (размер, нуклеотидный состав).

Полученные результаты позволяют предсказывать условия экспериментального проведения ПЦР. В ряде случаев на основе компьютерного анализа выявлена нецелесообразность проведения экспериментальных исследований ПЦР-диагностики, так как уже на этапе моделирования в геноме находится небольшое количество участков, содержащих искомые праймеры, либо их вовсе нет. Это в значительной степени упрощает и оптимизирует работу генетиков-экспериментаторов при проведении ПЦР. Так как зная ожидаемый из компьютерного анализа размер ампликонов и их количество, мы должны увидеть их на геле-электрофорезе в виде полос при проведении непосредственного проведения ПЦР. Если их нет, то и проводить амплификацию бессмысленно.

3. Литература

- [1] Kleppe, K. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases / K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, H.G. Khorana. – Mol. Biol. Bd. – 2002. – Vol. 56 – P. 341-361.
- [2] Глик, Б. Молекулярная биотехнология. Принципы и применение / Б. Глик, Дж. Пастернак. – М.: Мир, 2002. – 589 с.
- [3] Чемерис, Д.А. Дизайн праймеров для полимеразной цепной реакции (краткий обзор компьютерных программ и баз данных) / Д.А. Чемерис, О.Ю. Кирьянова, И.М. Губайдуллин, А.В. Чемерис // Биомика. – 2016. – Т. 8, № 3. – С. 215-238.
- [4] Knuth, D.E. Fast pattern matching in strings / D.E. Knuth, J.H. Morris (Jr), V.R. Pratt // SIAM Journal on Computing. – 1977. – Vol. 6(1) – P. 323-350.
- [5] Biopython [Electronic resource]. – Access mode: <https://biopython.org/>.
- [6] Pearson, W.R. Improved tools for biological sequence comparison / W.R. Pearson, D.J. Lipman // Proceedings of the National Academy of Sciences of the United States of America. – 1988. – Vol. 85(8). – P. 2444.

Благодарности

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 17-44-020120.

Simulation of primer search in the DNA chain

O.Y. Kiryanova¹, A.V. Chemeris²

¹Ufa State Petroleum Technological University, Kosmonavtov street 1, Ufa, Russia, 450062

²Institute of Biochemistry and Genetics, Ufa Federal Scientific Centre, prospect Oktyabrya 71, Ufa, Russia, 450054

Abstract. Polymerase chain reaction (PCR) is one of the most common experimental methods for solving DNA analysis problems. Oligonucleotide primers are an important component of any PCR, and therefore there are a number of requirements for their design. The success and the possibility of conducting a successful experiment as a whole depends on these structures. In this regard, there is a need for a computer analysis of primer selection. An algorithm was developed based on the Boyer-Moore search algorithm for a specific primer design. This paper presents two variations of primer search. On the basis of the algorithm, a program has been developed that allows computer-aided design of primers to be performed prior to direct experimental PCR. Which greatly simplifies the conduct of the direct experiment. At the moment, calculations have been carried out for model objects (chromosomes of Arabidopsis).