

УДК 575.22:599.723.2

ПРИНЦИПЫ ИДЕНТИФИКАЦИИ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В СОСТАВЕ СПЕКТРОВ ISSR-МАРКЕРОВ

© 2022 г. Н. В. Бардуков*, @, В. С. Артамонова*

*ФГБУН Институт проблем экологии и эволюции им. А.Н. Северцова РАН,
Ленинский проспект, д. 33, Москва, 119071 Россия

@E-mail: bardukv-nikolajj@mail.ru

Поступила в редакцию 27.03.2020 г.

После доработки 26.02.2021 г.

Принята к публикации 26.02.2021 г.

Разработан алгоритм для теоретического прогнозирования нуклеотидных последовательностей, амплифицирующихся в результате ISSR-PCR, на примере секвенированного генома лошади. Экспериментально показано, что теоретически предсказанные последовательности присутствуют в реальном электрофоретическом спектре ампликонов. Обсуждаются перспективы применения разработанного подхода для расшифровки спектров анонимной амплификации.

Ключевые слова: ISSR-маркеры, анонимная амплификация, *Equus caballus*, секвенированный геном, ПЦР *in silico*

DOI: 10.31857/S102634702201005X

Системы геномного сканирования, т.е. одно-временного анализа множества локусов, очень востребованы. В основном их создают на основе SNP-чипов (Kulbrock *et al.*, 2013; Porto-Neto *et al.*, 2013), хотя более простыми и доступными являются системы, основанные на принципе анонимной амплификации, в частности, ISSR-анализ (Inter Simple Sequence Repeats) (Zietkiewicz *et al.*, 1994). При использовании в ПЦР праймера, комплементарного какому-либо микросателлитному мотиву и имеющего “якорные” нуклеотиды на 3'-конце, на тестируемой ДНК-матрице синтезируется набор ампликонов, который различается у представителей разных видов, пород (сортов), и даже у отдельных животных (растений). Это позволяет сравнивать между собой виды, популяции (породы, сорта), отдельных особей (Basha, Sujatha, 2007; Воронкова и др., 2011; Столповский, 2013; Al-Ameri *et al.*, 2016; Kalia *et al.*, 2017).

Однако многие исследователи скептически относятся к работе с такими системами, поскольку остается неизвестным, из каких участков генома происходят ампликоны, составляющие ISSR-спектр, и чем обусловлен выявляемый полиморфизм (Банникова, 2004).

С другой стороны, несмотря на наличие объективных недостатков, данный тип ДНК-маркеров востребован и в настоящее время. Наибольшее количество исследований, выполняемых с помощью ISSR-анализа, лежит в области изучения генетики растений. Они востребованы при селекции, а также

для оценки разнообразия в генофондах диких видов (Дорогина и др., 2019; Степанов и др., 2019; Kim *et al.*, 2019; Chen *et al.*, 2020). Продолжаются, хотя и с меньшей интенсивностью, исследования генетики животных. В частности, в последнее время ISSR-маркеры применяли при изучении азиатских слонов (Kriangwanich *et al.*, 2018), псовых (Глазко и др., 2020), искусственно выведенных пород и природных популяций рыб (Пелеева и др., 2018). Имеются работы, объектами которых становились лошади. Так, с помощью ISSR-анализа изучали особенности генетики чистокровных лошадей арабской породы (Sheikh *et al.*, 2018), а в работе (Донт и др., 2018) с помощью ISSR-маркеров были найдены локусы, характерные для лошадей тракененской, тяжеловозных и рысистых пород. Авторы последней работы рекомендуют использовать разработанную ими систему для определения породной принадлежности и при создании генетических паспортов животных. Таким образом, можно утверждать, что ISSR-анализ до сих пор имеет свою нишу и активно используется исследователями в разных странах.

Однако лишь в настоящее время, когда стали известны полные последовательности геномов животных тех видов, которых изучают методами анонимной амплификации, стало возможно с помощью *in silico* PCR (моделирование результатов ПЦР по секвенированным последовательностям) рассчитать ISSR-спектры ампликонов и сравнить теоретически ожидаемые спектры с эксперимен-

тальными. Такой способ позволяет понять, какие именно нуклеотидные последовательности формируют ISSR-спектры. Для решения подобных задач были созданы специальные компьютерные программы (Kalendar *et al.*, 2017), но, судя по работам, опубликованным в настоящее время, несмотря на принципиальную возможность их использования для моделирования спектров анонимных ДНК-маркеров, с этой целью они не применяются.

В настоящей работе представлен пример использования подхода *in silico* PCR для расшифровки состава спектров ISSR-маркеров.

МАТЕРИАЛЫ И МЕТОДЫ

В данном исследовании в качестве примера был создан алгоритм для теоретического прогнозирования нуклеотидных последовательностей, получаемых в результате ISSR-PCR при использовании праймеров (GA)_nC или (AG)_nC, записанный на языке Python 2.6. (рис. 1). Расчеты выполняли в программе IDLE (Python GUI) (<http://www.python.org/>). Предварительный вариант алгоритма был опубликован в работе (Бардуков и др., 2014). Преимуществом данного алгоритма является большая гибкость и прозрачность настройки поискового запроса (например, имеется возможность учесть различные варианты неполной комплементарности ДНК-матрицы заданному праймеру) по сравнению с поиском с помощью программ *in silico* PCR. Это условие особенно важно в случае анонимной амплификации, при которой существует высокая вероятность отжига праймера на не полностью комплементарной ему матрице. Важной особенностью алгоритма является то, что первые пять нуклеотидов матрицы должны точно соответствовать 3'-концу праймера, поскольку было экспериментально показано, что именно эта область оказывает определяющее влияние на эффективность синтеза ампликонов (Rejiali, 2018).

При поиске теоретически возможных ампликонов были использованы последовательности секвенированного генома домашней лошади (*Equus caballus*), представленные в GenBank (EquCab2.0.; NC_009144.2 – NC_009175.2).

Поскольку программного обеспечения, позволяющего адекватно оценивать температуру отжига праймера на не полностью комплементарной матрице, в настоящее время не существует, с целью сегрегации найденных последовательностей по степени вероятности их появления в составе спектра, для каждого возможного сайта отжига

праймера оценивали температуру плавления дуплексов, образуемых ими с праймером, по формуле:

$$T_{\text{теор}} = 100.5 + \frac{41(yG + zC)}{(wA + xT + yG + zC)} - \frac{820}{(wA + xT + yG + zC)} + 16.6 \log_{10}([Na^+]),$$

где wA , xT , yG , zC – число нуклеотидов А, Т, G и С в составе праймера, а $[Na^+]$ – концентрация ионов натрия (<http://biotools.nubic.northwestern.edu/OligoCalc.html>). При этом условно принимали, что при наличии в составе дуплекса неспаренных нуклеотидов, праймер укорачивается в соответствии с их количеством.

Теоретически рассчитанный спектр ампликонов сравнивали с экспериментальным, для чего производили генотипирование лошадей алтайской и карачаевской пород, образцы крови которых имелись в коллекции Центра нанобиотехнологий РГАУ-МСХА им. К.А. Тимирязева. Образцы крови лошадей алтайской породы (48 образцов из хозяйства СПК “Энчи”, СПК зарегистрирован в селе Чибиля Улаганского района Республики Алтай, а также 48 образцов из СПК “Чингиз”, СПК зарегистрирован в селе Кара-Кудюр Улаганского района Республики Алтай) были собраны в ходе экспедиции сотрудников Лаборатории сравнительной генетики животных Института общей генетики им. Н.И. Вавилова РАН в Республику Алтай, а 34 образца крови лошадей карачаевской породы были собраны сотрудниками Центра нанобиотехнологий РГАУ-МСХА им. К.А. Тимирязева в частных хозяйствах Карачаево-Черкесской Республики и Ставропольского Края.

ДНК выделяли из образцов периферической крови лошадей с помощью набора реагентов “ДНК-Экстран-1” (ООО “Синтол”, Москва) согласно инструкции фирмы-производителя.

Для выполнения ПЦР использовали амплификатор Терцик (“ДНК-технология”, Россия) и наборы реагентов ПЦР-РВ (ООО “Синтол”). Концентрацию $MgCl_2$ в ПЦР-смеси доводили до 2.5 мМ. Условия ПЦР были следующими: первичная денатурация ДНК ($t = 94^\circ C$, 2 мин); 35 циклов синтеза ПЦР-продукта, включающих три этапа (денатурация ($t = 94^\circ C$, 30 с), отжиг праймера ($t = 55^\circ C$, 30 с), элонгация ($t = 72^\circ C$, 2 мин)); финальная элонгация цепей ДНК ($t = 72^\circ C$, 10 мин). В качестве ISSR-праймеров использовали олигонуклеотиды с последовательностями (GA)_nC и (AG)_nC.

Разделение ПЦР-продуктов проводили методом горизонтального электрофореза в 1.5%-ном агарозном геле. Для уточнения размеров ампликонов при сопоставлении длины теоретически рассчитанного амплифицируемого фрагмента с локусами экспериментального ISSR-спектра выполняли электрофорез в 5%-ном полиакриламидном геле (ПААГ). Во всех случаях для электро-

```

# -*- coding: cp1251 -*-
import re
f=open("chromosome.txt", "r")
for line in f.readlines():
    print
    re.findall(r".....AGAGA[^...]{0,4}AGAGAGAC[^...]{50,2000}GTCT
    CTC[^...]{0,4}TCTCT.....", line)
    print
    re.findall(r".....AGAGA[^...]{0,4}AGAGAGAC[^...]{50,2000}GTCT
    CT[^...]{0,2}CTCT[^...]{0,2}CTC.....", line)
    print
    re.findall(r".....AGAGA[^...]{0,4}AGAGAGAC[^...]{50,2000}GTCT
    C[^...]{0,2}TCT[^...]{0,2}CTC[^...]{0,2}TCT.....", line)
    print
    re.findall(r".....GAG[^...]{0,2}AGAG[^...]{0,2}AGAGAC[^...]{50
    ,2000}GTCTCTC[^...]{0,4}TCTCT.....", line)
    print
    re.findall(r".....GAG[^...]{0,2}AGAG[^...]{0,2}AGAGAC[^...]{50
    ,2000}GTCTCT[^...]{0,2}CTCT[^...]{0,2}CTC.....", line)
    print
    re.findall(r".....GAG[^...]{0,2}AGAG[^...]{0,2}AGAGAC[^...]{50
    ,2000}GTCTC[^...]{0,2}TCT[^...]{0,2}CTC[^...]{0,2}TCT.....",
    line)
    print
    re.findall(r".....AGA[^...]{0,2}GAG[^...]{0,2}AGA[^...]{0,2}GA
    GAC[^...]{50,2000}GTCTCTC[^...]{0,4}TCTCT.....", line)
    print
    re.findall(r".....AGA[^...]{0,2}GAG[^...]{0,2}AGA[^...]{0,2}GA
    GAC[^...]{50,2000}GTCTCT[^...]{0,2}CTCT[^...]{0,2}CTC.....",
    line)
    print
    re.findall(r".....AGA[^...]{0,2}GAG[^...]{0,2}AGA[^...]{0,2}GA
    GAC[^...]{50,2000}GTCTC[^...]{0,2}TCT[^...]{0,2}CTC[^...]{0,2}
    TCT.....", line)
f.close()

```

Рис. 1. Алгоритм поиска теоретически возможных ампликонов, которые могут присутствовать в ISSR-спектре при использовании в ПЦР праймера (GA)₉C. Алгоритм для праймера (AG)₉C аналогичен.

фореза использовали Tris-боратный буфер (1хTBE: 89 mM Tris, 89 mM борная кислота, 2 mM ЭДТА, pH 8.3) (Маниатис и др., 1984).

Окрашивание ПЦР-продуктов производили раствором бромистого этидия (0.5 мкг/мл), визуализацию результатов — на трансиллюминаторе УВТ-1 (“Биоком”, Россия) и системе геле-документирования INGENIUS (“SYNGENE”, Великобритания) при длине волны 312 нм. Для определения длин ампликонов использовали маркеры молекулярных масс “М 27” (“СибЭнзим”, Россия) и “50 bp DNA Step Ladder” (“Promega”, США).

Поиск последовательностей, гомологичных заданным, в геномах животных разных видов, представленных в GenBank, выполняли с использованием встроенной программы BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

С помощью разработанного алгоритма было выявлено 185 теоретически возможных ампликонов длиной не более 2000 п.о. Следует отметить однако, что, несмотря на их большое количество, ни один из этих ампликонов не имел сайтов от-

жига, полностью комплементарных праймеру (GA)₉C.

Теоретическая оценка температуры плавления (T_m) дуплексов показала, что, даже при полной комплементарности праймера и матрицы, она не превышает 57°C, а для теоретически предсказанных сайтов, образуемых праймером и флангами расчетных ампликонов, не превышает 55°C. Между тем, принято считать, что температура отжига праймера $T_{A \text{ теор.}} \approx T_m - 5$, т.е., примерно на пять градусов ниже температуры плавления дуплексов, образуемых праймером и матрицей (ПЦР “в реальном времени”, 2009).

Это означает, в свою очередь, что при температуре отжига праймера, равной 55°C, получение богатого ампликонами ISSR-спектра на матрице, не имеющей участков, полностью комплементарных матрице, теоретически невозможно ($T_{A \text{ теор.}} \approx 50^\circ\text{C}$ для теоретически предсказанных сайтов отжига, наиболее комплементарных праймеру). Однако известно, что при использовании праймера (GA)₉C и температуре отжига 55°C разные исследователи неоднократно получали на ДНК лошадей ISSR-спектры, насыщенные ампликонами (Glazko *et al.*, 2016; Голик и др., 2018).

В связи с этим, мы поставили серию экспериментов и определили, что в реальных условиях амплификация большинства ампликонов для выбранной нами системы матрица-праймер полностью прекращается только при температуре отжига праймера $\geq 65^\circ\text{C}$, т.е. $T_{A \text{ макс. эксп.}} \approx T_{A \text{ теор.}} + 15^\circ\text{C}$.

Далее, мы ранжировали теоретически предсказанные ампликоны по степени вероятности их появления в экспериментальном спектре, исходя из теоретической оценки температуры плавления дуплексов на обоих концах каждого из ампликонов. Для дальнейшего анализа были отобраны только те ампликоны, которые имели расчетную температуру плавления ДНК-дуплексов, образуемых матрицей и праймером, не ниже 40°C (табл. 1). В результате для последующего анализа было отобрано 25 последовательностей.

Теоретически рассчитанный спектр сравнивали с экспериментальным, полученным в ПЦР с использованием праймера (GA)₉C (Бардуков и др., 2014) при температуре отжига праймера равной 55°C.

Сопоставление показало хорошее соответствие между теоретически ожидаемым и экспериментальными спектрами: для всех мажорных и большинства минорных локусов экспериментального ISSR-спектра были найдены расчетные ампликоны, соответствующие им по длине.

Наибольшее число расчетных ампликонов было найдено для широких мажорных полос, состоящих, по всей видимости, из ампликонов сходной длины, но происходящих из разных частей

генома — т.н. гетерогенных кластеров (рис. 2). Теоретически предсказанные ампликоны, совпадающие по длине с мажорными полосами и полосами в составе гетерогенных кластеров, имели более высокие расчетные температуры плавления дуплексов (от 52 до 54°C), по сравнению с другими предполагаемыми ампликонами, найденными в геноме (рис. 2). Это хорошо согласовывалось с особенностями реального электрофоретического спектра, однако являлось лишь косвенным подтверждением того, что были найдены ампликоны, действительно входящие в состав ISSR-спектра.

Дальнейшую проверку возможностей выявления в ISSR-спектрах теоретически предсказанных ампликонов выполняли двумя способами.

На первом этапе был рассмотрен клонированный и секвенированный ранее полиморфный ампликон длиной ~420 п.о. из ISSR-спектра, полученного на геномной ДНК лошадей при использовании в ПЦР праймера (AG)₉C (Глазко и др., 2013). На основе секвенированной последовательности (длина ампликона составляла, как оказалось, 416 п.о.) были спланированы праймеры F-416 (5'-GAGAGAGAGAGAGAGAGCCCATGG-3') и R-416 (5'-GAGAGAGAGAGAGAGCGAAAGAGC-3') для избирательной амплификации данного локуса. Массовое тестирование лошадей показало, что амплификация этого локуса проходит успешно только у жеребцов, что говорит о его локализации в Y-хромосоме, которая в геноме лошади, представленном в GenBank, отсутствует (в GenBank представлен геном кобылы). В связи с этим, как и ожидалось, найти данный фрагмент в геноме домашней лошади с помощью алгоритма, разработанного в настоящей работе, не удалось.

При поиске в GenBank последовательностей, гомологичных клонированному ампликону выяснилось, что участок Y-хромосомы, найденный в ISSR-спектре лошади, секвенирован у лошади Пржевальского (*Equus przewalskii*, NW_007676839.1) и осла (*Equus asinus*, NW_014638651.1). Эти последовательности обозначены как “unplaced genomic scaffold”, т.е. их локализацию в геноме авторам установить не удалось. Однако наше исследование доказывает, что это участки Y-хромосомы.

Применение разработанного алгоритма к участкам хромосом *E. przewalskii* и *E. asinus* показало, что внутри последовательностей NW_007676839.1 и NW_014638651.1 действительно удастся выявить ампликоны, гомологичные последовательности полиморфного ампликона домашней лошади. Таким образом, с помощью разработанного алгоритма поиска в последовательностях Y-хромосом близких видов удалось найти ампликон, который ранее был выделен из ISSR-спектра, полученного

Таблица 1. Соответствие между теоретически рассчитанными и экспериментально полученными спектрами ампликонов

Название теоретически предсказанного ампликона	Последовательности ДНК на сайтах посадки праймера с расчетной температурой плавления дуплексов и длина теоретически предсказанного ампликона	Локализация теоретически предсказанного ампликона на хромосоме	Соответствие спектру
17 ¹ –1326	GGAAGAGAGAGAGAGAGAC ² (52°C); GTCTCTCTCTCTCTTTCTC ³ (54°C); 1326 п.о.	NC_009160.3 (48971749–48973074)	Мажорный кластер около 1250–1350 п.о.
1–1319	ATCCTAGAGAGAGAGAGAC (41°C); GTCTCTCTGGCTCTGTCTT (44°C); 1319 п.о.	NC_009144.3 (163465329–163466647)	
2–1302	GAAAGAGAGAGAGAGAGAC (54°C); GTCTCTCTCCCTCTGTCTG (48°C); 1302 п.о.	NC_009145.3 (28343941–28345242)	
6–1301	TAAAGAGAATGAGAGAGAC (44°C); GTCTCTCTCTCAGTCTTCT (41°C); 1301 п.о.	NC_009149.3 (51069282–51070582)	
25–1281	GAGGGAGAGAGACAGAGAC (52°C); GTCTCTCAGCATCTCTCTG (41°C); 1281 п.о.	NC_009168.3 (29106855–29108135)	
3–1214	GCCTGAGAGAGAAAGAGAC (46°C); GTCTCTCTGTCTCTGTCTG (46°C); 1214 п.о.	NC_009146.3 (30151710–30152923)	Полоса около 1100–1150 п.о.
4–1148	TCTCGAGAGACAGAGAGAC (41°C); GTCTCTCTCTCTCTTTCC (52°C); 1148 п.о.	NC_009147.3 (17092678–17093825)	
25–1281	ATGAGAGAGAGAGAGAGAC (52°C); GTCTCTCAGCATCTCTCTG (41°C); 1140 п.о.	NC_009168.3 (29106996–29108135)	
20–1115	GAGGGGCAGAGAGAGAGAC (51°C); GTCTCTCTCTCTCGCTGGG (46°C); 1115 п.о.	NC_009163.3 (7278613–7279727)	
10–1102	AACTGAGAGACAGAGAGAC (44°C); GTCTCTGTCTCTTCCTCGG (40°C); 1102 п.о.	NC_009153.3 (20333365–20334465)	
21–1060	GAGAGAGGATGAGAGAGAC (51°C); GTCTCTCTCTTGCTCTGAC (46°C); 1060 п.о.	NC_009164.3 (57129665–57130724)	Нет типичных четких полос
15–922	GGAAGAGAGAGAGAGAGAC (52°C); GTCTCTCGCTGTGTCTCTCTCCCTCTCTATG ⁴ (48°C); 910–922 п.о.	NC_009158.3 (88931543–88932464)	Нет типичных четких полос
24–900	TAGAAAGAGAGAGAGAGAC (50°C); GTCTCTCTCTCTAACATGC (44°C); 900 п.о.	NC_009167.3 (24755030–24755929)	
X–876	TTTATAGAGAGAGAGAGAC (44°C); GTCTCTCTCTCTCACTCAC (55°C); 876 п.о.	NC_009175.3 (76760564–76761439)	
18–871	AAGACAGAGAGAGAGAGAGAC ⁴ (50–54°C); GTCTCTCTTTCTCTCTCTC (54°C); 863–871 п.о.	NC_009161.3 (27521579–27522449)	Мажорный кластер около 820–880 п.о.
15–922	GGAAGAGAGAGAGAGAGAC (52°C); GTCTCTCTCTGTCTCTCTCTATTCTTTCT ⁴ (46–54°C); 855–865 п.о.	NC_009158.3 (88931543–88932407)	
18–871	GAGAGAGAGGGAGAGAGAC (55°C); GTCTCTCTTTCTCTCTCTC (54°C); 835 п.о.	NC_009161.3 (27521615–27522449)	
10–833	CTGAATGAGAGAGAGAGAC (46°C); GTCTCTCTGTCTCCATCTCTCTCTCTC ⁴ (48–52°C); 825–833 п.о.	NC_009153.3 (60737537–60738369)	
19–823	TAGAGAGAGAGAGAGAGAC (54°C); GTCTCTCTGTGTCTCTATG (41°C); 823 п.о.	NC_009162.3 (43532121–43532942)	
8–805	GGGAGACAGAGACAGAGAC (48°C); GTCTCTCTTTCTCTCTGTGTC (50°C); 805 п.о.	NC_009151.3 (8521975–8522779)	Нет типичных четких полос

Название теоретически предсказанного ампликона	Последовательности ДНК на сайтах посадки праймера с расчетной температурой плавления дуплексов и длина теоретически предсказанного ампликона	Локализация теоретически предсказанного ампликона на хромосоме	Соответствие спектру
X-637	CAAAACGAGAGAGAGAGAC (44°C); GTCTCTCCATCTCTCTCTC (52°C); 637 п.о.	NC_009175.3 (1358676–1359312)	Минорные полосы 610–690 п.о.
10-572	AAGAAAGAGGGAGAGAGAC (48°C); GTCTCTCTCTCTTTCTCTT (50°C); 572 п.о.	NC_009153.3 (29331545–29332116)	Полосы около 540–580 п.о.
18-871	TAGAGAGAGACAGAGAGAC (50°C); GTCTCTCTTTCTCTCTCTC (54°C); 563 п.о.	NC_009161.3 (27521887–27522449)	
X-637	CAAAACGAGAGAGAGAGAC (44°C); GTCTCTCTATCTCTCTCCA (48°C); 552 п.о.	NC_009175.3 (1358676–1359227)	
10-572	AAGAAAGAGGGAGAGAGAC (48°C); GTCTCTCTCTGTTTCTATC (46°C); 551 п.о.	NC_009153.3 (29331545–29332094)	
10-546	GGGAGGAAGAGAGAGAGAC (51°C); GTCTCTCTGTGTCTCTTTC (46°C); 546 п.о.	NC_009153.3 (23373675–23374221)	Нет типичных четких полос
11-513	GAGGGAGAGAGAGAGAGAC (55°C); GTCTCTCTGGTTCTCTCCT (41°C); 513 п.о.	NC_009154.3 (121598–122110)	
11-513	AGGAGACAGAGAGAGAGAC (48°C); GTCTCTCTGGTTCTCTCCT (41°C); 491 п.о.	NC_009154.3 (121620–122110)	
28-489	GAGAGAAAGAGAGAGAGAC (54°C); GTCTCTCTCTCTCACACAC (53°C); 489 п.о.	NC_009171.3 (34185907–34186395)	Мажорная полоса около 490 п.о.
X-637	CAAAACGAGAGAGAGAGAC (44°C); GTCTCTCCATCTCTCTCCA (46°C); 450 п.о.	NC_009175.3 (1358676–1359125)	Нет типичных четких полос
X-637	CAAAACGAGAGAGAGAGAC (44°C); GTCTCTCCATCTCTCTCCA (46°C); 418 п.о.	NC_009175.3 (1358676–1359093)	
21-377	TGAGGAGAGGGAGAGAGAC (44°C); GTCTCTCTATCTCTCTCTA (50°C); 377 п.о.	NC_009164.3 (41224113–41224489)	Мажорная полоса около 380 п.о.
X-637	CAAAACGAGAGAGAGAGAC (44°C); GTCTCTCTCTCCATCTTTT (44°C); 366 п.о.	NC_009175.3 (1358676–1359041)	
10-394	ACAGCAAGAGAGAGAGACA <u>AGAGAGAAAGAGAGAGAGAGT</u> <u>AGAGT</u> GAGTGAGTGAGAGACA <u>AGAGAGAGAGAC</u> ⁴ (44–52°C); GTCTCTCTCTCTCTCTCCA (52°C); 354–394 п.о.	NC_009153.3 (58797548–58797941)	
7-349	GAGAGATGGAGAGAGAGAC (52°C); GTCTCTTCCTCTCTCTTCC (46°C); 349 п.о.	NC_009150.3 (85378714–85379062)	Нет типичных четких полос

ИЗВЕСТИЯ РАН. СЕРИЯ БИОЛОГИЧЕСКАЯ № 1 2022

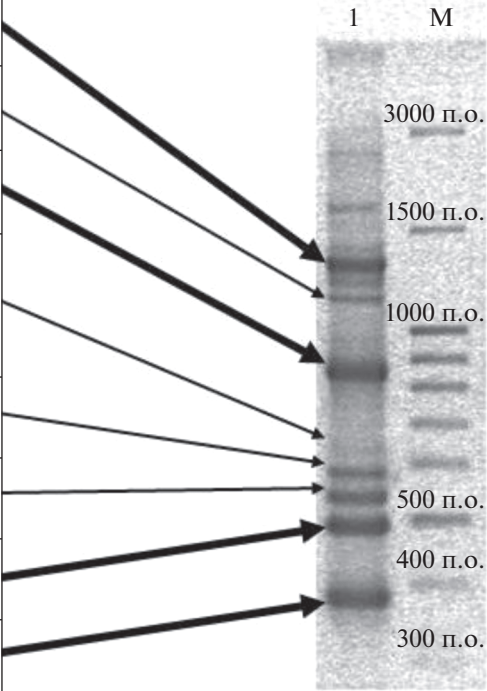
Наиболее характерные зоны в экспериментальном спектре	Количество совпадающих расчетных фрагментов	Совпадение результатов с электрофоретическим спектром
Мажорный кластер 1250–1350 п.о.	5	 М — маркер молекулярных масс; 1 — пример электрофоретического спектра ISSR-маркеров, полученного на ДНК карачаевской лошади с помощью праймера (GA)₉C
Полоса около 1100–1150 п.о.	4	
Мажорный кластер 800–900 п.о.	8	
Минорные полосы со слабым сигналом около 610–690 п.о.	1	
Полоса около 580 п.о.	1	
Полосы около 540–560 п.о.	4	
Мажорная полоса около 490 п.о.	1	
Мажорная полоса около 380 п.о.	1	

Рис. 2. Соответствие между теоретически предсказанными ($T_m \geq 40^\circ\text{C}$) и экспериментально полученными ампликонами. Жирные стрелки указывают на мажорные полосы, для которых теоретически предсказаны ампликоны с высокой температурой плавления (от 52 до 54°C) на расчетных сайтах посадки праймера (GA)₉C.

ампликоном из ISSR-спектра домашней лошади. Это показывает, что единичные точечные мутации могут менять температуру отжига праймера в том или ином его положении, давая преимущества в связывании праймера одному из двух-трех возможных сайтов, с чем, видимо, связана, по крайней мере, часть полиморфизма, который демонстрируют ISSR-маркеры.

На втором этапе проверки возможности выявления в ISSR-спектрах теоретически предсказанных ампликонов была поставлена обратная задача. Предстояло экспериментально подтвердить возможность амплификации теоретически предсказанного ампликона в ISSR-спектре, полученном с применением праймера (GA)₉C при стандартных условиях ПЦР. С этой целью в экспериментальном ISSR-спектре, полученном при исполь-

зовании в качестве матрицы ДНК лошади карачаевской породы, был выбран ампликон, который имел длину порядка 570 п.о., не входил в состав крупных мажорных кластеров и отличался минорным характером проявления (рис. 4). С помощью алгоритма, разработанного в данном исследовании, в 10-й хромосоме лошади был найден схожий по длине ампликон (572 п.о.), фланкированный последовательностями, частично комплементарными праймеру (GA)₉C.

Для экспериментальной проверки соответствия расчетного ампликона ампликону, присутствующему в экспериментальном спектре, были подобраны праймеры F-10-740 (5'-GGCTCTCTTTAAGGATCTAGAAC-3') и R-10-740 (5'-GTAATTCTTTCACTCATCTGCTAC-3'), комплементарные участкам, находящимся от последователь-

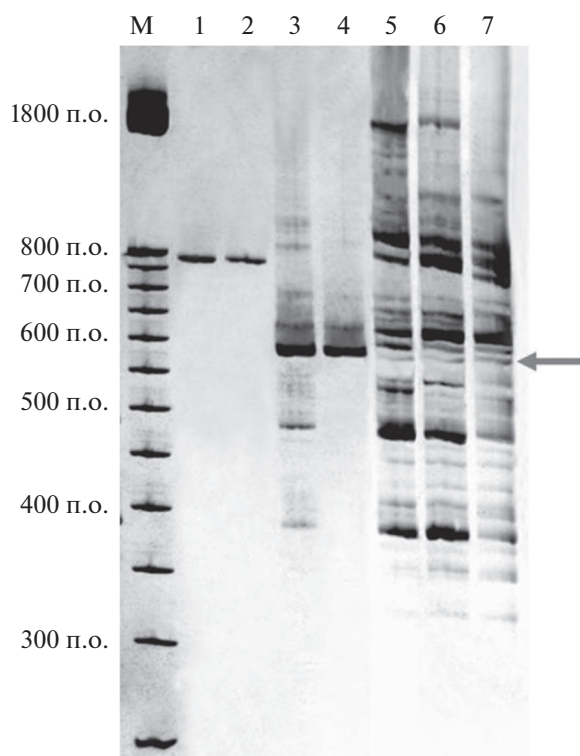


Рис. 4. Электрофорез в 5%-ом ПААГ. М – маркер молекулярных масс; 1, 2 – ПЦР-продукт, полученный с праймеров F-10-740 и R-10-740 на геномной ДНК лошади; 3, 4 – ПЦР-продукт, полученный с использованием праймера (GA)₉C, когда матрицей служил ПЦР-продукт, представленный на дорожках 1 и 2 (синтез дополнительных минорных ампликонов происходил на геномной ДНК, остатки которой присутствовали в пробе); 5–7 – электрофоретический спектр, полученный при использовании в ПЦР праймера (GA)₉C и геномной ДНК лошадей каракаевской породы в качестве матрицы. Стрелка указывает на теоретически предсказанный ампликон длиной 572 п.о.

(GA)₉C и сайтом посадки праймера R-10-740, его аллельные варианты представлены на рис. 5.

Пример ампликона, происходящего из десятой хромосомы, с прилегающим к нему полиморфным микросателлитным локусом (AGAT)_n представляет большой интерес, поскольку внутри

25-ти теоретически рассчитанных ампликонов было найдено множество микросателлитных мотивов (в табл. 1 наряду с фланговыми последовательностями ампликонов даны ссылки на локализацию этих ампликонов в секвенированных хромосомах лошади, что позволяет убедиться в наличии микросателлитных мотивов внутри данных ампликонов). Таким образом, представляется весьма вероятным, что в ряде случаев полиморфизм ISSR-спектров наблюдается из-за различий в длине микросателлитов, расположенных внутри ампликонов.

ЗАКЛЮЧЕНИЕ

Полученные результаты показывают, что моделирование ISSR-спектров с помощью анализа последовательностей секвенированных хромосом позволяет расшифровать значительную часть анонимного спектра, а также определить типичные для конкретного спектра причины полиморфизма его ампликонов. В настоящем исследовании установлено, что основными причинами полиморфизма ISSR-спектров, полученных с использованием праймеров (GA)₉C и (AG)₉C являются, по всей видимости, точечные нуклеотидные замены в т. н. якорных нуклеотидах и полиморфизм микросателлитов, расположенных внутри ампликонов.

Результаты работы показывают, что комплексный подход, включающий методы биоинформатики и экспериментальные приемы работы с последовательностями, способными воспроизводиться при анонимной амплификации, перспективен для расшифровки последовательностей, составляющих наблюдаемый спектр. В будущем он позволит выявлять общие закономерности формирования спектров ампликонов, что значительно упростит интерпретацию получаемых результатов. Не менее важно и то, что данный подход позволяет значительно снизить трудозатраты при создании на основе ISSR-ампликонов узкоспециализированных локус-специфичных маркеров. Упрощение разработки таких ДНК-маркеров позволит исключить целый каскад биотехнологических про-

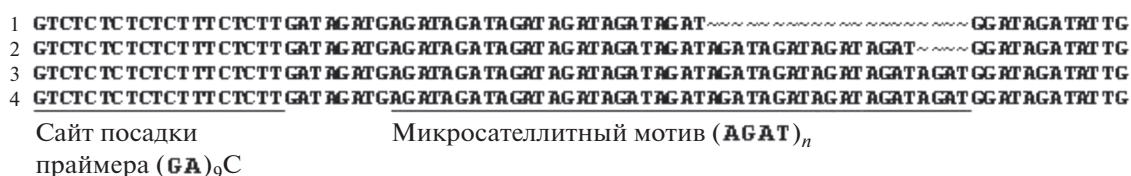


Рис. 5. Аллельные варианты микросателлитного локуса с коровым мотивом (AGAT)_n, находящегося между теоретически рассчитанным сайтом отжига праймера (GA)₉C и сайтом отжига праймера R-10-740. 1 — участок нуклеотидной последовательности из 10-й хромосомы лошади, представленной в GenBank; 2, 3 — аналогичные участки последовательности у лошади карачаевской породы, являющейся гетерозиготной по микросателлитному локусу с коровым мотивом (AGAT)_n; 4 — соответствующий участок последовательности лошади карачаевской породы, гомозиготной по данному микросателлитному локусу.

цедур, а значит, появится возможность быстро создавать их под необходимые задачи — будь то определение принадлежности животного или растения к конкретной популяции, породе, сорту, выявление гибридов, поиск надежных маркеров для выявления филогенетических связей.

Благодарности. Авторы благодарят сотрудников Центра нанобиотехнологий ФГБОУ ВО РГАУ-МСХА им. К.А. Тимирязева за предоставленные пробы ДНК лошадей.

Финансирование. Работа выполнена в рамках госзадания (тема 6: Экология и биоразнообразие водных сообществ 0109-2018-0076 AAAA-A18-118042490059-5).

СПИСОК ЛИТЕРАТУРЫ

- Банникова А.А. Молекулярные маркеры и современная филогенетика млекопитающих // Журн. общей биологии. 2004. Т. 65. № 4. С. 278–305.
- Бардуков Н.В., Феофилов А.В., Глазко Т.Т., Глазко В.И. ISSR-PCR маркеры и мобильные генетические элементы в геноме домашней лошади *Equus caballus* // Сельскохозяйственная биология. 2014. № 4. С. 42–57.
- Воронкова В.Н., Цэндсүрэн Цэдэв, Сулимова Г.Е. Сравнительный анализ информативности ISSR-маркеров для оценки генетического разнообразия пород лошадей // Генетика. 2011. Т. 47. № 8. С. 1131–1134.
- Глазко В.И., Гладырь Е.А., Феофилов А.В., Бардуков Н.В., Глазко Т.Т. ISSR-PCR маркеры и мобильные генетические элементы в геномах сельскохозяйственных видов млекопитающих // Сельскохозяйственная биология. 2013. № 2. С. 71–76.
- Глазко В.И., Косовский Г.Ю., Блохина Т.В., Есин А.Г., Глазко Т.Т. Геномный “портрет” некоторых видов псовых, полученный с помощью ISSR-PCR и IRAP-PCR маркеров // Кролиководство и звероводство. 2020. № 1. С. 28–39.
- Голик Т.В., Эркенов Т.А., Глазко Т.Т., Глазко В.И. Спектры ISSR-PCR маркеров в оценках популяционно-генетической дифференциации карачаевской лошади в хозяйствах Карачаево-Черкесской Республики // Известия ТСХА. 2018. № 3. С. 61–77.
- Донт Ю.У., Тимарова А.В., Комарова Л.В., Боронникова С.В. Генетический полиморфизм трех пород лошади домашней // Вестник Пермского университета. Сер.: Биология. 2018. Вып. 1. С. 50–56.
- Дорогина О.В., Васильева О.Ю., Нуржидина Н.С., Буглова Л.В., Жмудь Е.В., Зуева Г.А., Комина О.В., Кубан И.С., Гусар А.С., Дудкин Р.В. Формирование и изучение коллекционного генофонда ресурсных видов рода *Miscanthus* Anderss. в условиях лесостепи Западной Сибири // Вавиловский журн. генетики и селекции. 2019. Т. 23. № 7. С. 926–932.
- Маниатис Т., Фрич Э., Сэмбрук Дж. Методы генетической инженерии. Молекулярное клонирование. М.: Мир, 1984. 480 с.
- Пелеева А.Р., Комарова Л.В., Васильева Ю.С. Анализ генетического разнообразия естественных популяций и ремонтно-маточных стад стерляди на основании полиморфизма межмикросателлитных маркеров // Бюллетень науки и практики. 2018. Т. 4. № 4. С. 20–29.
- Ребриков Д.В., Саматов Г.А., Трофимова Д.Ю., Семенов П.А., Савилова А.М., Кофиади И.А., Абрамов Д.Д. ПЦР “в реальном времени” / Под ред. Ребрикова Д.В. М.: БИНОМ. Лаборатория знаний, 2009. 215 с.
- Степанов И.В., Супрун И.И., Токмаков С.В., Лободина Е.В. Поиск эффективных IRAP и ISSR маркеров для генетического анализа подвоев яблони // Плодоводство и виноградарство Юга России. 2019. № 60. С. 11–20.
- Столбовский Ю.А. Популяционно-генетические основы сохранения генофондов domesticiрованных видов животных // Вавиловский журн. генетики и селекции. 2013. Т. 17. № 4/2. С. 900–915.
- Al-Ameri A.A., Al-Qurainy F., Gaafar A.-R.Z., Khan S., Nadeem M. Molecular identification of sex in *Phoenix dactylifera* using inter simple sequence repeat markers // BioMed Res. Intern. V. 2016. № 4530846.
- Basha S.D., Sujatha M. Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers // Euphytica. 2007. V. 156. P. 375–386.
- Chen X., Feng H., Du Y., Luo S., Li W., Yu L., Feng Z., Cui T., Zhou L. Genetic polymorphisms in mutagenesis progeny of *Arabidopsis thaliana* irradiated by carbon-ion beams and γ -rays irradiations // Int J Radiat Biol. 2020. V. 96(2). P. 267–275.
- Glazko V.I., Erkenov T.A., Glazko T.T., Dzatoev K.M. Genetic structure of Karachai horses on ISSR-PCR markers // Biogeosystem Technique. 2016. V. 9. № 3. P. 195–204.
- Kalendar R., Khassenov B., Ramankulov Y., Samuilova O., Ivanov K.I. FastPCR: An *in silico* tool for fast primer and probe design and advanced sequence analysis // Genomics. 2017. V. 109(3–4). P. 312–319.
- Kalia P., Saha P., Ray S. Development of RAPD and ISSR derived SCAR markers linked to Xca1Bo gene conferring resistance to black rot disease in cauliflower (*Brassica oleracea* var. *botrytis* L.) // Euphytica. 2017. V. 213. № 232.
- Kim J.Y., Adhikari P.B., Ahn C.H., Kim D.H., Kim Y.C., Han J.Y., Kondeti S., Yong Choi Y.E. High frequency somatic embryogenesis and plant regeneration of interspecific ginseng hybrid between *Panax ginseng* and *Panax quinquefolius* // J. Ginseng Res. 2019 V. 43(1). P. 38–48.
- Kriangwanich W., Nganvongpanit K., Buddhachat K., Brown J.L., Siengdee P., Chomdej S., Bansiddhi P., Thitaram C. Genetic diversity and variation in captive Asian elephants (*Elephas maximus*) in Thailand // Trop. Conser. Science. 2018. V. 11.
- Kulbrock M., Lehner S., Metzger J., Ohnesorge B., Distl O. A genome-wide association study identifies risk loci to equine recurrent uveitis in German warmblood horses // PLoS One. 2013. V. 8(8). e71619.
- Porto-Neto L.R., Sonstegard T.S., Liu G.E., Bickhart D.M., Da Silva M.V., Machado M.A., Utsunomiya Y.T., Garcia J.F., Gondro C., Van Tassell C.P. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping // BMC Genomics. 2013. V. 14. P. 876.

- Rejali N.A., Moric E., Wittwer C.T.* The effect of single mismatches on primer extension // *Clin Chem.* 2018. V. 64. № 5. P. 801–809.
- Sheikh A., Ahmed M.M.M., Mutwakil M.H.Z., Saini K.S., Alsulaimany F.A.S., EL Hanafy A.A., Sabir J.S.M.* Comparative molecular analysis of ISSR markers in Arabian horse breeds // *J. Anim. Plant Sciences.* 2018. V. 28(1). P. 332–336.
- Zietkiewicz E., Rafalski A., Labuda D.* Genome fingerprinting by sequence repeat (SSR) anchored polymerase chain reaction amplification // *Genomics.* 1994. V. 20. P. 176–183.

Principles of Nucleotide Sequences Identification ISSR Markers Spectra

A. V. Bardukov^{1, #} and V. S. Artamonova¹

¹*A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences,
Leninskij prosp. 33, Moscow, 119071 Russia*

[#]*e-mail: bardukv-nikolajj@mail.ru*

An algorithm for theoretical prediction of nucleotide sequences obtained in ISSR PCR was developed. The calculation was performed using one of the sequenced horse genomes as an example. Presence of theoretically calculated sequences in the real electrophoretic spectrum of amplification was experimentally confirmed. The possibility of applying the developed approach for identifying nucleotide sequences in anonymous amplification spectra was discussed.

Keywords: ISSR markers, anonymous amplification, *Equus caballus*, sequenced genome, *in silico* PCR