Биомика, 2020, Том 12, № 2, 194-210.

BIOMICS/БИОМИКА

ISSN 2221-6197 http://biomicsj.ru

МУЛЬТИПЛЕКСНЫЙ *IN SILICO* RAPD-АНАЛИЗ РЯДА РОДСТВЕННЫХ РАСТЕНИЙ С ОТЛИЧАЮЩИМИСЯ РАЗМЕРАМИ ГЕНОМОВ И ПЕРСПЕКТИВЫ ТАКОГО ПОДХОДА ДЛЯ ДНК-ПАСПОРТИЗАЦИИ СОРТОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ РАСТЕНИЙ

[1]Кирьянова О.Ю., [2]Кулуев Б.Р., [2]Кулуев А.Р., [3]Марданшин И.С., [4]Губайдуллин И.М., [2]Чемерис А.В.

[1]Уфимский государственный нефтяной технический университет
Россия, 450062, Уфа, ул. Космонавтов 1, E-mail: olga.kiryanova27@gmail.com
[2]Институт биохимии и генетики – обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, Россия, 450054, Уфа, Проспект Октября, 71,
[3]Башкирский научно-исследовательский институт сельского хозяйства – обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, Россия, 450059, Уфа, ул. Р.Зорге, 19
[4]Институт нефтехимии и катализа – обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, Россия, 450075, Уфа, пр. Октября, 141

Резюме

Для нескольких видов растений проведен мультиплексный *in silico* RAPD-анализ их полных геномов, заметно различающихся своими размерами. Из семейства Крестоцветных исследованы три разновидности резуховидки Таля *Arabidopsis thaliana*, имеющие геном размером около 135 млн. пар нуклеотидов (п.н.). Из семейства Пасленовых проанализированы картофель *Solanum tuberosum* и томат *S.lycopersicum* с геномами 840 и 828 млн.п.н. соответственно. Из семейства Злаковых в анализ взяты два подвида риса *Oryza sativa* indica и *O.sativa* japonica (по 500 млн.п.н.), диплоидные пшеница *Triticum urartu* (5 млрд.п.н.) и эгилопс *Aegilops tauschii* (4,3 млрд.п.н.), а также тетраплоидные пшеницы *T.turgidum* и *T.dicoccoides* (оба вида имеют геномы около 12 млрд.п.н.) и гексаплоидная мягкая пшеница *T.aestivum*, имеющая геном размером около 17 млрд.п.н. Биоинформатический *in silico* анализ на наличие мест отжига в этих геномах комплекта декамерных праймеров проводился с помощью созданной нами ранее программы ABCDNA_GS. Особенностью праймеров, содержащих 40% G и C оснований, является их тринуклеотидный состав с полным отсутствием тиминов, способствующий исключению образования гомо- и гетеродимеров таких праймеров. Показано, что для растений с малыми размерами геномов (на примере семейства Крестоцветных) удовлетворительный уровень полиморфизма ДНК достигается при использовании в мультиплексном анализе комплекта из 12 праймеров, тогда как для крупных геномов вполне достаточно 6 таких праймеров. Предлагается модифицировать метод RAPD-анализа таким образом, что будут учитываться только короткие ампликоны, которые можно разделять капиллярным гель-электрофорезом и измерять их длины с точностью до нуклеотида. Выбранный диапазон разделения ампликонов от 51 до 500 нуклеотидов вмещает 450 воображаемых ДНК-ячеек, которые обозначаются как ДНК[+]-ячейка при наличии в ней фрагмента(ов) ДНК и ДНК[-]-ячейка при их отсутствии, что в двоичном счислении отображается как «1» и «0» соответственно. Приведены подсчеты количества возможных комбинаций при разной заполненности таких воображаемых ДНК-ячеек, превышающих в ряде случаев гугол (10^{100}). На основе полученных *in silico* результатов построены генетические штрих-коды, дающие возможность визуального наблюдения отличий проанализированных видов растений. Выявленные у разновидностей резуховидок и подвидов риса минорные различия позволяют прогнозировать применимость такого подхода к однозначной ДНК-паспортизации сортов растений. Причем данный подход может быть вполне применим и к породам животных, расам грибов и даже штаммам микроорганизмов, требуя лишь корректировки числа мультиплексных праймеров в зависимости от размеров геномов исследуемых видов. Помимо

RAPD-анализа, для целей ДНК-паспортизации могут применяться и иные методы детекции полиморфизма ДНК, обеспечивающие точное установление размеров образующихся ампликонов. Биоинформатический анализ полных геномов исследуемых видов с помощью программы ABCDNA_GS может также служить в этих случаях для прогнозирования результатов и предварительного выбора комплектов мультиплексных праймеров. Отмечается, что подобная ДНК-паспортизация крайне важна для селекционных работ и аграрного производства.

Ключевые слов: полиморфизм ДНК, геном, мультиплексная ПЦР, праймер, RAPD-анализ, биоинформатический анализ, *in silico*, виртуальные ампликоны, генетическое штрих-кодирование, ДНК-паспортизация, сорт, селекция

Цитирование: Кирьянова О.Ю., Кулуев Б.Р., Кулуев А.Р., Губайдуллин И.М., Чемерис А.В. Мультиплексный *in silico* RAPD-анализ ряда родственных растений с отличающимися размерами геномов и перспективы такого подхода для ДНК-паспортизации сортов сельскохозяйственных растений // Биомика. 2020. Т.12. №2. С. 194-210. DOI: 10.31301/2221-6197.bmcs.2020-10

MULTIPLEX *IN SILICO* RAPD-ANALYSIS OF SEVERAL RELATED PLANTS WITH DIFFERENT GENOME SIZES AND PROSPECTS FOR THIS APPROACH FOR DNA-CATALOGUING OF AGRICULTURAL PLANT VARIETIES

[1]Kiryanova O.Yu., [2]Kuluev B.R., [2]Kuluev A.R., [3]Mardanshin I.S., [4]Gubaydullin I.M., [2]Chemeris A.V.

[1]Ufa State Petroleum Technological University,
Russia, Ufa, 450062, 1 Kosmonavtov str., E-mail: olga.kiryanova27@gmail.com
[2]Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences,
Russia, Ufa, 450054, 71 Pr. Oktyabrya
[3]Bashkir Research Institute of Agriculture, Ufa Federal Research Center, Russian Academy of Sciences,
Russia, Ufa, 450059, 19 R. Zorge Str.
[4]Institute of Petrochemistry and Catalysis, Ufa Federal Research Center, Russian Academy of Sciences,
Russia, Ufa, 450075, 141 Pr. Oktyabrya

Resume

A multiplex *in silico* RAPD-analysis of their complete genomes, which differ markedly in size, was performed for several plant species. From Crucifera family three lines of *Arabidopsis thaliana* with a genome size of about 135 million bp were investigated. The potato *Solanum tuberosum* and tomato S. lycopersicum from the Solanaceae family with genomes 840 and 828 million bp respectively were analyzed. Two subspecies of rice *Oryza sativa* indica and *O. sativa* japonica (500 million bp each), diploid wheat *Triticum urartu* (5 billion bp) and *Aegilops tauschii* (4.3 billion bp), as well as tetraploid wheat *T. turgidum* and *T. dicoccoides* (both species have genomes of about 12 billion bp) and hexaploid bread wheat *T. aestivum* (17 billion bp) were analyzed in silico for the presence of annealing sites in these genomes of a set of decamer primers was performed using the ABCDNA_GS program that we created earlier. A feature of primers containing 40% G and C bases is their trinucleotide composition with a complete absence of thymines, which helps to exclude the formation of homo- and heterodimers of such primers. It is shown that for plants with small genome sizes (for example, the Crucifera family), a satisfactory level of DNA polymorphism is achieved when a set of 12 primers is used in multiplex analysis, whereas 6 such primers are quite sufficient for large genomes. It is proposed to modify the RAPD-analysis in such a way that only short amplicons that can be separated by capillary gel electrophoresis and their lengths measured with accuracy to the nucleotide will be taken into account. The selected range of amplicons from 51 to 500 nucleotides accommodates 450 imaginary DNA cells, which are designated as a DNA[+]-cell if it contains a fragment(s) of DNA and DNA[-]-cell in their absence, which in binary is displayed as "1" and " 0 " respectively. The calculations of the number of possible combinations with different filling of such imaginary DNA cells, exceeding in some cases a googol (10^{100}), are given. Based on the results obtained in silico, genetic barcodes are constructed that allow visual observation of differences between the analyzed plant species. Minor differences found in

Arabidopsis lines and rice subspecies allow us to predict the applicability of this approach to unambiguous DNA-cataloguing of plant cultivars. Moreover, this approach can be quite applicable to animal breeds, fungal races, and even microbial strains, requiring only an adjustment of the number of multiplex primers depending on the size of the genomes of the studied species. In addition to RAPD analysis, other methods of detecting DNA polymorphism can be used for the purpose of DNA-cataloguing, which provide accurate determination of the size of the resulting amplicons. Bioinformatic analysis of the complete genomes of the studied species using the ABCDNA_GS program can also serve in these cases to predict the results and pre-select sets of multiplex primers. It is noted that such DNA certification is extremely important for breeding and agricultural production.

Keywords: DNA polymorphism, genome, multiplex PCR, primers, RAPD analysis, bioinformatic analysis, in silico, virtual amplicons, genetic barcoding, DNA-cataloguing, cultivar, selection

Citation: Kiryanova O.Yu., Kuluev B.R., Kuluev A.R., Mardanshin I.S., Gubaydullin I.M., Chemeris A.V. Multiplex *in silico* RAPD-analysis of several related plants with different genome sizes and prospects for this approach for DNA-cataloguing of agricultural plant varieties. *Biomics*. 2020. Vol. 12. No. 2. P. 194-210. DOI: 10.31301/2221-6197.bmcs.2020-10 (In Russian)

Введение

В сельскохозяйственном производстве необходимо использовать высокопродуктивные, устойчивые к комплексу неблагоприятных факторов среды сорта растений. В настоящее время для подтверждения подлинности происхождения семян производитель, совместно с сотрудниками центра по сертификации комиссионно на основе имеющихся документов и осмотра посевов оформляют сертификат соответствия. При этом, помимо габитуса того или иного сорта (видимого лишь при росте и/или плодоношении), для него еще желательно иметь некий генетический паспорт, по которому уже на стадии семян или иного посевного материала можно устанавливать их принадлежность, что весьма важно при приобретении подобного товара, особенно больших партий. Другое применение таких генетических паспортов может заключаться в контроле за возделыванием тех или иных сортов для получения роялти создавшими их селекционерами. Наконец, генетические паспорта сортов и отдельных перспективных линий культурных растений крайне необходимы для селекционной работы при проведении различных скрещиваний и анализе гибридных форм.

Для создания подобных генетических паспортов сельскохозяйственных растений как нельзя лучше подходят молекулы ДНК, характеризующиеся огромным полиморфизмом, который может быть выявлен различными методами. Однако большинство из них обеспечивают получение аналоговых данных, которые даже при их оцифровке являются по сути псевдоцифровыми, поскольку предполагают большие допущения. При этом цифровые технологии все больше проникают в разные сферы жизни общества, но их представленность в сельском хозяйстве пока недостаточна. В этой связи является весьма перспективным использование для ДНК-паспортизации сортов сельскохозяйственных растений метода обнаружения полиморфизма ДНК, позволяющего получать истинно цифровые данные в виде точного (до нуклеотида) определения размеров фрагментов ДНК (в частности, продуктов ПЦР – ампликонов). Одним из таковых является относительно простой метод, получивший название RAPD (Random Amplified Polymorphic DNA) [Williams et al., 1990]. Но его классический вариант предполагает разделение ампликонов в агарозных гелях, исключающих точное определение их размеров и поэтому требуется применение секвенирующего гель-электрофореза, что для RAPD-анализа с флуоресцентно-меченными праймерами было уже показано [Corley-Smith et al., 1997], но об оцифровке первичных данных тогда сообщено не было.

Ранее нами довольно подробно рассмотрен целый спектр различных методов, включая RAPD, используемых для выявления полиморфизма ДНК [Кулуев и др. (Kuluev et al.), 2018; Нигматуллина и др. (Nigmatullina et al.), 2018: Сухарева, Кулуев (Sukhareva, Kuluev), 2018] и поэтому здесь на них останавливаться не будем. Заметим только, что ДНК-паспортизация сорта может быть никак не связана с его кодируемыми теми или иными генами продуктивными признаками. То есть, говоря другими словами, используемые маркеры могут представлять собой некие анонимные участки ДНК, просто присущие конкретным сортам, сортообразцам, линиям. Во избежание недоразумений хотим заметить, что в данной статье идет речь о ДНК-паспортизации / ДНК-каталогизации сортов / сортообразцов / линий растений, а не об их генетической характеристике на основе ДНК, под

которой можно было бы понимать их определенные фенотипические особенности. При этом существует группа методов выявления полиморфизма ДНК, в том числе RAPD-анализ, не требующая от экспериментатора знания нуклеотидных последовательностей ни отдельных генов, ни всего генома, но для тех видов растений, для которых полные геномы известны, несмотря на то, что они полные довольно условно, поскольку по сути квазигаплоидные, можно, тем не менее, *in silico* провести их анализ и заранее прогнозировать ожидаемые результаты выявления полиморфизма ДНК с теми или иными праймерами, чему и посвящена данная статья. Использование биоинформатического подхода на этапе подготовки к широкомасштабным исследованиям полиморфизма ДНК сортов сельскохозяйственных растений с помощью RAPD-анализа и их ДНК-паспортизации позволяет значительно сэкономить расходные материалы и время, поскольку для тех видов, чьи геномы полностью или почти полностью секвенированы, можно *in silico* подобрать наиболее оптимальные комплекты мультиплексных праймеров, отбраковав неподходящие. Причем все большее число полных геномов сельскохозяйственных видов растений становятся известными и поэтому прогнозирование ожидаемого полиморфизма с помощью *in silico* анализа будет востребованным в плане исключения как низкого уровня полиморфизма с теми или иными праймерами, так и заметного превалирования одного или нескольких ампликонов, приходящихся на повторяющуюся ДНК или на геномы органелл, имеющих высокую копийность, что может затруднять считывание первичных данных.

Выбор объектов и инструментов для *in silico* анализа

Известно, что ядерные геномы растений могут сильно отличаться по своим размерам, варьируя от 60 млн.п.н. до 150 млрд.п.н. или более чем в тысячу раз. Однако размеры геномов сельскохозяйственных видов отличаются не столь значительно, хотя в определенных случаях и десятки раз разницы учитывать необходимо. В этой связи в качестве объектов исследования выбраны из семейства Злаковых трибы пшеницевых мягкая гексаплоидная хлебная пшеница *Triticum aestivum*, размер генома которой достигает 17 млрд.п.н., что приблизительно в 20 раз превышает размер генома картофеля из семейства Пасленовых – *Solanum tuberosum* (840 млн.п.н.), также взятого нами в анализ. Для сравнения из этого же семейства взят томат *S.lycopersicum*, геном которого (828 млн.п.н.) близок по размеру с картофельным. Из трибы пшеницевых взяты еще несколько видов – диплоидная пшеница *T.urartu* (5 млрд.п.н.) и диплоидный эгилопс *Aegilops tauschii* (4,3 млрд.п.н.), являющиеся соответственно предполагаемыми донорами субгеномов **A** и **D** гексаплоидной пшеницы *T.aestivum*. При этом *T.urartu* – также предполагаемый донор субгенома **A** ряда тетраплоидных пшениц, среди которых для *T.turgidum* и *T.dicoccoides* полные геномы (около 12 млрд.п.н. каждый) установлены и по этой причине они тоже оказались среди анализируемых в данной работе видов. Еще из семейства Злаковых проанализированы два подвида риса – *Oryza sativa* japonica и *O.sativa* indica, имеющие размер генома около 500 млн.п.н.

Помимо возделываемых культурных растений, в анализ взят сорняк из семейства Крестоцветных - резуховидка Таля *Arabidopsis thaliana*, служащая модельным объектом для Царства растений ввиду малого размера генома, равного 135 млн.п.н., что приблизительно в 125 раз меньше генома мягкой пшеницы и в 6 раз меньше картофельного. Еще одной важной причиной почему был выбран арабидопсис в качестве объекта исследования служит то, что секвенированы полные геномы уже нескольких образцов этого растения, что позволяет лучше прогнозировать возможные внутривидовые (потенциально сортовые) отличия, выявляемые с помощью RAPD-анализа.

Таким образом, *in silico* анализу были подвергнуты несколько близкородственных (имеется ввиду внутри каждого семейства) видов растений из трех семейств (Злаковые, Пасленовые, Крестоцветные), заметно различающиеся размерами своих ядерных геномов. Поскольку в растениях ДНК содержится не только в ядре, но и в органеллах, то определенное внимание уделено и имеющим гораздо меньший размер пластидным (135 - 180 тысяч пар нуклеотидов) и митохондриальным (350 - 500 т.п.н.) геномам вышеупомянутых видов, для которых таковые секвенированы. Забегая вперед, скажем, что ни для одного проанализированного нами плазмона с выбранными (см. ниже) мультиплексными праймерами, виртуальных ампликонов не найдено, что в целом неудивительно, учитывая весьма малые размеры таких геномов, однако, пластомы и хондриомы при прогнозировании возможных ампликонов, безусловно, нужно принимать во внимание, поскольку иная комбинация праймеров может привести к амплификации фрагментов ДНК из них в том или ином выбранном диапазоне. При этом следует иметь ввиду, что гомология геномов органелл у близкородственных видов крайне высока и если ампликоны в них выявятся, то они будут носить скорее родоспецифичный характер или даже будут типичны на уровне семейств.

Нуклеотидные последовательности полных ядерных и плазмонных геномов исследуемых видов растений были взяты из нескольких баз данных –-

Ensemble (http://plants.ensembl.org/species.html), NCBI (https://www.ncbi.nlm.nih.gov), 1001 Genomes Project (https://1001genomes.org).

Для мультиплексного *in silico* RAPD-анализа использована созданная нами ранее программа ABCDNA_GS (Amplified Bar-Coded DNA Genome/Specimen), совмещенная с собственной базой данных [Кирьянова и др. (Kiryanova et al.), 2020]. В качестве виртуальных мультиплексных праймеров использовались шесть подобранных нами декамерных олигонуклеотидов со следующими почти рандомными последовательностями: AACCAGACAA, AAGGGACAAA, AACCGAACAA, AACGCACAAA, AAAACGCCAA, AACGCCAAAA. Ввиду малого размера генома у резуховидки и как следствие относительного небольшого числа ампликонов с этими шестью праймерами для этого вида растений мультиплексный *in silico* RAPD-анализ проводился с еще шестью дополнительными праймерами: AAACGCCAAA, AACCCAGAAA, AACCGAAGAA, AAAGGGACAA, AACGACACAA, ACAGAGACAA. Особенностью всех этих праймеров следует считать их тринуклеотидный состав с полным отсутствием тиминов. Еще одна особенность заключается в наличии на 3'-конце от двух до четырех аденинов, что с одной стороны может ухудшать эффективность амплификации, но с другой будет повышать ее специфичность, поскольку известно, что наличие нескольких G и C оснований на 3'-конце может приводить к неспецифичной ПЦР. Таким образом, подобранные праймеры абсолютно неспособны сформировать как гомо-, так и гетеродимеры. Но все это имеет значение для так называемых «мокрых» экспериментов, как стало принято в последнее время говорить, когда реальную ПЦР в пробирке или иной подходящей емкости необходимо противопоставить виртуальной ПЦР. Что касается GC-состава выбранных праймеров, равного 40%, то этот показатель важен и для *in silico* анализов, поскольку известно, что растительные геномы характеризуются чуть более высоким содержанием AT-пар, а одно из условий однозначной ДНК-паспортизации сортов заключается в выявлении большего (достаточного) полиморфизма ДНК.

Выравнивание нуклеотидных последовательностей отдельных виртуальных ампликонов проводилось с помощью программы MegAlign из пакета Lasergene фирмы DNAStar, Inc.

Теория метода

Для выявления полиморфизма ДНК *in silico* нами выбран подход, использующий в реальности ПЦР амплификацию и не требующий предварительного знания нуклеотидных последовательностей генома того или иного вида растения, а именно уже упоминавшийся выше RAPD-анализ [Williams et al., 1990]. Хотя нужно признать, что RAPD-анализ никогда всерьез не рассматривался в качестве «ДНК-паспортизирующего» метода в силу его не самой хорошей воспроизводимости, однако его некоторое усовершенствование возможно и об этом будет говориться при обсуждении результатов. В настоящее время нарабатываемые в ходе RAPD ампликоны в виде двухцепочечных фрагментов ДНК разделяются обычно в агарозном геле и их анализируемый размерный диапазон, как правило, составляет от 200 до 2000 пар нуклеотидов. В этих условиях длины разделяемых фрагментов соотносятся с маркерными фрагментами ДНК (лестницами), разделяемыми в соседних треках, и определяются с довольно большой погрешностью, где некоторое число пар нуклеотидов может достигать немалых величин ±, хотя даже ±1 уже будет означать, что такие сведения считаться истинно цифровыми данными не могут. Поскольку отличить фрагмент ДНК длиной, например в 1657 пар нуклеотидов от имеющего близкий размер в 1662 пары нуклеотидов, не представляется возможным, то экспериментаторы вынужденно округляют эти значения до 1660 или даже до 1650 пар нуклеотидов, фактически считая их одинаковыми фрагментами, тогда как на деле для разных образцов это могут быть совершенно различные участки генома исследуемого вида. В связи с этим необходимо модифицировать метод RAPD следующим образом. Во-первых, брать в анализ другой диапазон длин разделяемых фрагментов ДНК (ампликонов), позволяющий определять размеры их одноцепочечных вариантов (это – во-вторых) с точностью до нуклеотида (в-третьих), что позволит уже однозначно определять, что у одного сорта, например, имеется RAPD-фрагмент длиной 176 нуклеотидов, а у другого – 179 нуклеотидов. Причем это могут быть и разные участки генома и даже одинаковые, но несущие инделы размером в три нуклеотида, что и обеспечивает выявляемый полиморфизм по длине. Но из какой части генома происходят эти фрагменты уже не суть важно, поскольку гигантское число комбинаций этих признаков в виде фрагментов ДНК с точно определенными размерами должно нивелировать все такие нюансы, не взирая как на аллополиплоидный, так и автополиплоидный статусы исследуемых видов растений. Это уже – в-четвертых. В-пятых, для увеличения числа ампликонов необходимо превратить обычный RAPD в мультиплексный, что в настоящее время применяется не часто, хотя каких-либо серьезных препятствий к этому быть не должно, тем более при применении праймеров с тринуклеотидным составом. Для этого надо использовать не один RAPD-праймер в отдельной реакционной смеси, как это обычно делается сейчас (что приводит к наработке

ампликонов, ограниченных праймерами, обозначенными здесь как A|A), а сразу несколько отличающихся праймеров. Например, если взять в реакцию шесть разных RAPD-праймеров с произвольными последовательностями (A, B, C, D, E, F), то возможные ампликоны теоретически могут возникнуть после отжига на нужном расстоянии следующих праймерных комбинаций – A|A, B|B, C|C, D|D, E|E, F|F; A|B, A|C, A|D, A|E, A|F; B|A, B|C, B|D, B|E, B|F; C|A, C|B, C|D, C|E, C|F; D|A, D|B, D|C, D|E, D|F; E|A. E|B, E|C, E|D, E|F; F|A, F|B, F|C, F|D, F|E.

При этом количество комбинаций праймеров в такой мультиплексной ПЦР подчиняется формуле $X = m^2$ где X – максимальное количество типов ампликонов, а m – число праймеров. Главное требование к таким праймерам – отсутствие образования между ними гомо- и/или гетеродимеров, что абсолютно не составляет какой-либо серьезной проблемы, и на этапе дизайна праймеров может быть легко достигнуто [Гарафутдинов и др. (Garafutdinov et al.), 2019] и уже описано нами выше, когда были приведены последовательности используемых в данной работе мультиплексных праймеров. Остается еще вопрос длины праймеров, от которой в том числе зависит и число вероятных ампликонов. Но слишком короткие праймеры будут приводить к неэффективной амплификации и плохой воспроизводимости результатов. Поэтому неким стандартом для RAPD-анализа являются декамерные праймеры, хотя они все же несколько коротки для ПЦР. По теории вероятности в любой геномной ДНК некие блоки, например, из десяти определенных нуклеотидов (способные служить местами отжига праймеров) могут встречаться неоднократно. Количество возможных комбинаций декануклеотидов определяется как число размещений с повторениями из 4 элементов по 10, что составляет 1048576 (4^{10}) вариантов (`AAAAAAAAAA AAAAAAAAAc AAAAAAAAAg AAAAAAAAAt AAAAAAAAcA AAAAAAAAgA AAAAAAAAtA AAAAAAAcAA ... tttttttttt`, где строчными буквами показаны отличающиеся нуклеотиды от таковых в условно первом 10-ти звенном олигонуклеотиде `AAAAAAAAAA`). То есть, в среднем любая последовательность из 10 нуклеотидов теоретически должна встречаться в геноме через приблизительно каждый миллион нуклеотидов. Данная закономерность применима к некоторому идеальному случаю, но здесь нет необходимости более детально разбирать какие особенности нуклеотидных последовательностей генома того или иного вида организма как будут влиять на этот процесс. Можно просто принять, что в ДНК растений с учетом размеров их геномов мест отжига таких праймеров (на обеих цепях) должно быть соответствующее пропорциональное количество. Однако также необходимо принимать во внимание неизбежную неравномерность распределения любых таких декануклеотидных блоков в ДНК, по которой согласно Гауссовому распределению определенное число таких виртуальных ампликонов действительно может иметь размер около миллиона пар нуклеотидов, но обязательно будут присутствовать и более крупные и более мелкие фрагменты ДНК, что будет во многом зависеть как от последовательностей тех или иных геномов, так и последовательностей конкретных 10-ти звенных олигонуклеотидов, но здесь мы сознательно от них абстрагируемся. В итоге в ходе уже настоящей амплификации с помощью ПЦР со случайными (до некоторой степени, конечно же) декануклеотидными праймерами будет образовываться некое количество мелких фрагментов ДНК с размерами, например до 500 пар нуклеотидов, которые с помощью капиллярного гель-электрофореза в денатурирующих условиях уже в виде одноцепочечной ДНК могут быть измерены с точностью до нуклеотида, что обеспечит получение истинно цифровых данных.

Поскольку RAPD-маркеры носят доминантный характер наследования, то получаемые при электрофорезе первичные данные легко перевести в двоичный формат. Так, если выбрать диапазон ампликонов, например от 51 до 500 нуклеотидов, обеспечивающего с помощью капиллярного гель-электрофореза установление точных (до нуклеотида) размеров ампликонов, то его можно представить в виде 450 воображаемых ячеек, в которых может присутствовать ДНК (и это будет ДНК[+]-ячейка) или отсутствовать ДНК (ДНК[-]-ячейка), что соответствует «1» и «0» в двоичной системе счисления соответственно. В качестве примера можно привести из этого диапазона небольшой участок от 141 до 150 нуклеотидов, где нахождение фрагментов ДНК может носить такой характер: … 141[+], 142[+], 143[-], 144[+], 145[-], 146[-], 147[+], 148[-], 149[+], 150[-] … , где цифрами обозначен размер фрагмента ДНК, [+] – наличие фрагмента ДНК, [–] – отсутствие. В двоичном формате запись будет следующей: … 1101001010 … . Такая бинарная запись может быть преобразована в генетический штрих-код, позволяющий визуально наблюдать сходство или различия исследуемых сортов растений. Причем нахождение в конкретной ДНК[+]-ячейке одного или большего числа разных ампликонов, имеющих одинаковый размер, не принципиально, поскольку ведется качественный, а не количественный анализ, и такая ДНК-ячейка все равно должна оцифровываться как «1». К тому же определение в каждой такой ячейке истинного числа разных ампликонов с одинаковыми размерами практически не представляется возможным. Хотя

нужно признать, что может оказаться, что подобранные праймер(ы) придутся на повторяющиеся элементы генома и тогда в «мокром» эксперименте будет образовываться чрезмерно большое количество такого ампликона, что будет отрицательно влиять на сбор первичных данных. Чтобы исключить такую возможность как раз и требуется предварительный *in silico* анализ полных геномов исследуемых видов.

Фактически информация об исследуемом образце (сорте, сортообразце, линии) может быть представлена в виде перемежающихся нулей и единиц в выбранном диапазоне длин берущихся в анализ ампликонов, являя собой пример удачной оцифровки и ухода от аналоговых или псевдоцифровых данных. Межлабораторное сравнение получаемых результатов при оперировании фактически «голыми» цифрами становится довольно легким делом, а формирование соответствующих баз данных во многом перестает быть проблематичным. В реальных экспериментах по выявлению полиморфизма ДНК (которым будет посвящена наша отдельная статья) самым важным становится точно измерить длину конкретных ампликонов после разделения продуктов ПЦР в секвенирующем геле и при этом отсечь некий «шум» в виде недостроенных ампликонов, имеющих случайные размеры.

Для однозначной ДНК-паспортизации сортов сельскохозяйственных растений следует обеспечить выявление их достоверных различий в виде необходимого (большего) числа ампликонов с отличающимися размерами, что достигается использованием мультиплексного RAPD-анализа. Теоретическое число комбинаций встречаемости в этих воображаемых ДНК-ячейках ампликонов разного размера рассчитывается как число сочетаний без повторений из *m* по *n* -

$$C_m^n = \frac{m!}{n!(m-n)!}$$

где *C* – общее число комбинаций встречаемости фрагментов ДНК выбранного размерного диапазона, *m* – число всех анализируемых в выбранном диапазоне воображаемых ДНК-ячеек, *n* – число ДНК[+]-ячеек, *(m-n)* – число ДНК[-]-ячеек.

Как можно видеть из таблицы 1 количество возможных комбинаций сильно зависит от числа нарабатываемых ампликонов разного размера. Причем наличие признака (ДНК[+]-ячейка) либо его отсутствие (ДНК[-]-ячейка) при подсчете количества комбинаций по сути равнозначны за исключением крайних значений в виде полного отсутствия ампликонов или занятия ими всего диапазона воображаемых ДНК-ячеек.

Таблица 1. Соотношение числа возможных комбинаций перестановок ампликонов в воображаемых ДНК-ячейках в зависимости от их количества / Table 1. The ratio of the number of possible combinations of permutations of amplicons in imaginary DNA-cells, depending on their number

Количество ампликонов (450 max) Number of amplicos	Количество возможных комбинаций Number of possible combinations
0	0
1	4,50e+2
2	1,01e+6
3	1,50e+8
4	1,68e+10
5	1,50e+11
10	8,48+e20
20	3,10e+34
30	5,53e+46
40	2,76e+57
50	8,90e+66
60	3,04e+75
70	1,53e+83
80	1,42e+90
90	2,92e+96
100	1,50e+102
110	2,13e+107
120	9,17e+111
130	1,26e+116
140	5,94e+119
150	9,91e+122
160	6,09e+125
170	1,42e+128
180	1,29e+130
190	4,67e+131
200	6,79e+132
210	4,02e+133
220	9,78e+133
225	1,09e+134
230	9,78e+133
240	4,02e+133
250	6,79e+132
260	4,67e+131
270	1,29e+130
280	1,42e+128
290	6,09e+125
300	9,91e+122
310	5,94e+119
320	1,26e+116
330	9,17e+111
340	2,13e+107
350	1,50e+102
360	2,92e+96
370	1,42e+90
380	1,53e+83
390	3,04e+75
400	8,90e+66
410	2,76e+57
420	5,53e+46
430	3,10e+34
440	8,48+e20
445	1,50e+11
446	1,68e+10
447	1,50e+8
448	1,01e+6
449	4,50e+2
450	1

Более наглядно изменение количества комбинаций в зависимости от числа образующихся ампликонов и занятия ими ДНК-ячеек, которые могут быть также отображены в виде генетических штрих-кодов, представлено на рис. 1.

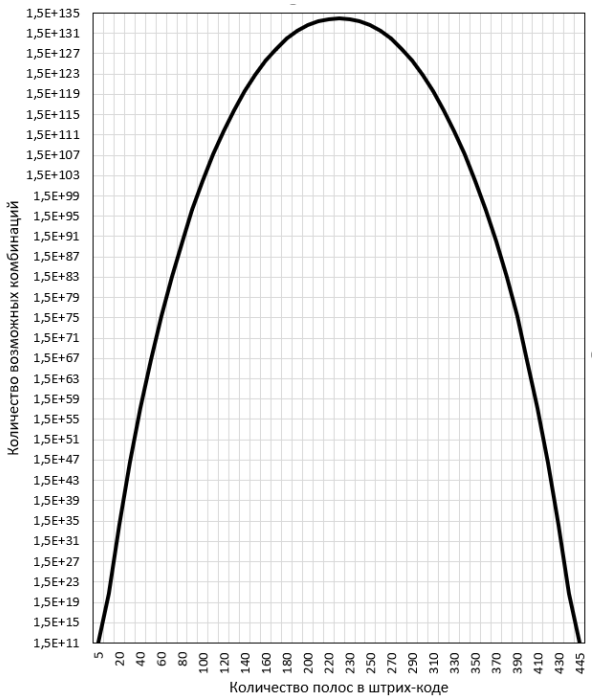


Рис. 1. Зависимость числа возможных штрих-кодов от количества полос в них. Применительно к генетическим штрих-кодам можно говорить о числе ампликонов в выбранном диапазоне длин. (Данные представлены в логарифмической шкале)
Fig. 1. Dependence of the number of possible bar-codes on the number of bars in them. In relation to genetical bar-codes, we can talk about the number of amplicons in the selected length range. (Data is presented in a logarithmic scale)

По теории вероятности самое большое количество комбинаций возникает при занятии ампликонами половины ячеек из числа, берущихся в анализ, что в данном случае (225 из 450) составит 1,09e+134. Занятие ампликонами 100 или 300 ячеек из 450 превысит гугол (10^{100}) комбинаций, составив 1,50e+102 вариантов их перебора. Однако даже формирование всего 10 ДНК[+]-ячеек из 450 возможных обеспечит секстиллионы (10^{20}) комбинаций. А если с какими-либо праймерами «наработается» 20 ампликонов разного размера, то это даст уже ундециллионы (10^{34}) комбинаций. При этом для однозначной ДНК-паспортизации сортов, сортообразцов или линий в ряде случаев и этого может оказаться мало, ввиду наличия большого числа видо- и даже родоспецифичных фрагментов. Тем более для установления родословных выводимых сортов желательно иметь дело с несколько большим количеством полиморфных (отличающихся у разных линий и сортообразцов) ампликонов нужного диапазона, в идеале приближающимся к 225 ДНК[+]-ячейкам для воображаемых 450 ДНК-ячеек. При необходимости диапазон анализируемых ампликонов может быть несколько расширен в обе стороны, например от 31 до 630 нуклеотидов, что дополнительно позволит выявлять еще более минорные различия между сортами и линиями.

Получаемые данные по наличию / отсутствию ампликонов в конкретной ДНК-ячейке преобразуются в бинарный код, служащий основой для формирования штрих-кода, способного за счет несметного числа комбинаций присваивать каждому сорту растений его уникальный генетический штрих-код, который позволит такой сорт каталогизировать и в дальнейшем однозначно идентифицировать. (Не следует путать с проектом по штрих-кодированию всего живого на основе гена цитохромоксидазы, развиваемого уже на протяжении ряда лет, который ориентирован на определение/каталогизацию таких таксонов как биологические виды/роды.) Формирование соответствующей базы данных по сортам сельскохозяйственных культур становится легко выполнимым, поскольку в ней должны храниться истинно цифровые данные, исключающие какие-либо неточности и допущения. Безусловно, при хорошей воспроизводимости данного метода, которую еще нужно добиваться.

Здесь можно также заметить, что нельзя исключать случаи, когда, например, произойдет буквально единичная мутация, которая повлечет за собой коренное изменение свойств сорта, выявляемое на фенотипическом или биохимическом уровне, но невидимое при избирательном анализе ДНК этого растения, поскольку замена такого нуклеотида окажется вне зоны отжига праймеров. При этом все остальные ныне используемые методы выявления полиморфизма ДНК растений с тем же «успехом» пропустят подобную мутацию, поскольку вероятность, что она окажется в анализируемом месте ничтожна и лишь секвенирование всего генома может теоретически позволить ее найти. Однако ДНК-паспортизация сортов и их ДНК-идентификация путем секвенирования и сравнения полных геномов вряд ли когда станет возможной и уместной, хотя бы потому что для большинства целей достаточно охарактеризовать часть генома. И только при наличии серьезной необходимости выявления всех возможных отличий можно (нужно) будет прибегать (в будущем) к секвенированию всего генома. Здесь можно провести некую аналогию с фотографией человека в

его паспорте, которая не дает всего представления о нем как личности (или биологической особи), но для целей первичной идентификации вполне пригодна.

Крестоцветные

Как уже говорилось выше, резуховидка Таля *A.thaliana* характеризуется довольно небольшим геномом в Царстве растений. К тому же выбор этого объекта продиктован тем, что полный геном размером около 135 млн. пар нуклеотидов этого растения представлен в базах данных в виде отдельных хромосом, которых у этого вида насчитывается 5. Более того, в рамках программы 1001 Genomes секвенированы еще несколько полных геномов арабидопсиса, что позволяет провести внутривидовое / внутригеномное сравнение.

Проведенный с помощью программы ABCDNA_GS *in silico* анализ геномов нескольких линий арабидопсиса на предмет наличия в них мест отжига для комбинации из 12 праймеров показал значительное сходство образующихся ампликонов, но при этом выявились и определенные различия, что было вполне ожидаемым (Табл. 2). Так, из 10 – 11 ампликонов, генерируемых с первой хромосомы арабидопсиса, у исследуемых образцов 7 совпали по размеру и для большей наглядности они выделены цветом. Для хромосомы 2 для всех линий все ампликоны были одинаковыми по размеру. Хромосома 3 показала наличие от 4 до 5 ампликонов, три из которых совпадают и также имеют свою цветовую гамму. Для хромосомы 4 из 5 – 6 ампликонов совпали 4. Хромосома 5 «дала» 8 – 9 ампликонов, из которых 7 были одинаковыми. В целом для проанализированных здесь секвенированных полных геномов *A.thaliana* выявлено от 32 до 35 ампликонов, из которых 25 совпали по размеру у всех исследованных линий резуховидки. Как уже говорилось выше, ни пластом, ни хондрионом арабидопсиса образования ампликонов с использованным комплектом праймеров не показал. Комплект же из 6 праймеров привел к образованию для этих линий резуховидки всего от 2 до 4 ампликонов, что крайне мало для ДНК-паспортизации растений (данные не приведены).

Таблица 2. Распределение мультиплексных ампликонов из диапазона 51 – 500 нуклеотидов по хромосомам у линий *A.thaliana* / Table 2. Distribution of multiplex amplicons from the range 51 – 500 nucleotides in chromosomes of different lines of *A.thaliana*

Species / Chromosome	1	2	3	4	5
Arabidopsis thaliana Col-0 (reference genome)	232*, 270, 271, 303, 331, 355, 364, 405, 467, 468, 499	367, 408, 450, 497	135, 145, 303, 458, 461	112, 332, 358, 468, 469, 495	146, 155, 229, 252, 281, 372, 383, 387, 470
A.thaliana An-1	232, 270, 271, 303, 346, 351, 364, 382, 405, 499	367, 408, 450, 497	145, 458, 461, 477	112, 332, 360, 468, 469, 495	146, 155, 158, 229, 252, 281, 372, 387
A.thaliana C24	232, 270, 271, 303, 330, 351, 355, 364, 405, 467, 499	367, 408, 450, 497	145, 303, 458, 461, 472	112, 332, 360, 468, 469	146, 155, 229, 252, 281, 372, 387, 469

* - здесь и в последующих подобных таблицах эти числа означают размеры ампликонов, выраженные в нуклеотидах.

Однако присутствие одинаковых по размеру фрагментов ДНК не гарантирует их принадлежность одинаковым или гомологичным участкам геномов этих видов. Для подтверждения этого были сравнены нуклеотидные последовательности некоторых ампликонов, которые для простоты будут здесь обозначаться по сокращенному названию образца и их размеру. Так, сравнение нуклеотидных последовательностей ампликонов Col145, An145 и C145, находящихся в составе хромосомы 3 с близкими геномными координатами – 10126045-10126190; 10086731-10086876; 10152132-10152277 соответственно, показало, что они гомологичны друг другу на 100%. Сопоставление ампликонов Col331 и C330, отличающихся по размеру на один нуклеотид выявило, что они также сильно схожи между собой, но несут три замены и одну делецию. Однако данные ампликоны Col331 и C330, несмотря на то, что они происходят из одинаковых участков генома для целей ДНК-паспортизации считать одинаковыми нельзя, поскольку этот наш подход с генетическим штрих-кодированием на основе размеров ампликонов не рассчитан на установление нуклеотидных последовательностей всех или даже части ампликонов, так как за счет большого числа комбинаций решает другую задачу (относительно) быстрой идентификации того или иного сорта растений.

На основе полученных размеров ампликонов были построены генетические штрих-коды проанализированных линий арабидопсиса, представленные на рис. 2.

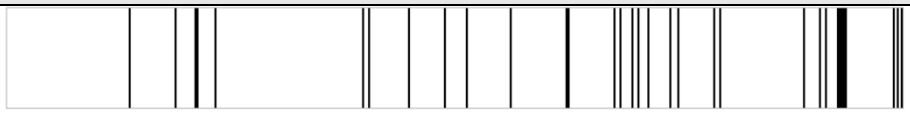
Species / varieties	Genetic bar-codes
Arabidopsis thaliana Col-0	
A.thaliana An-1	
A.thaliana C24	

Рис. 2. Генетические штрих-коды линий *Arabidopsis thaliana*
Fig. 2. Genetic bar-codes of several lines of *Arabidopsis thaliana*

Как можно видеть генетические штрих-коды трех линий резуховидок Таля довольно похожи друг на друга, но при этом имеют и определенные отличия. Данный результат позволяет прогнозировать, что у различных сортов культурных растений также могут выявляться достаточные отличия, позволяющие проводить их ДНК-паспортизацию. К тому же следует учесть, что у арабидопсиса геном весьма маленького размера, тогда как у большинства сельскохозяйственных культур размеры геномов крупнее и даже значительно крупнее, что должно обеспечивать большее количество ампликонов и среди них можно предполагать наличие и отличающихся по размеру, которые как раз и будут служить ДНК-идентификаторами сорта без какой-либо привязки к его генетическим особенностям в виде фенотипических характеристик.

Пасленовые

Из семейства Пасленовых в *in silico* анализ ядерной и плазмонной ДНК были взяты два вида сельскохозяйственных растений из одного рода, характеризующиеся близкими размерами своих геномов – картофель *S.tuberosum* (840 млн.п.н.) и томат *S.lycopersicum* (828 млн.п.н.). Обнаруженные в ядерных геномах этих видов ампликоны приведены в табл. 3. Соответствующие штрих-коды представлены на рис. 3.

Для картофеля обнаружен 41 ампликон, а для томата их оказалось несколько больше – 55. Ампликоны, имеющие у этих видов одинаковые размеры, выделены жирным шрифтом. При этом считать их все родоспецифичными можно только после того как будут выяснены их нуклеотидные последовательности и только в случае их совпадения (даже частичного) можно принять такое решение, хотя оно для задачи ДНК-паспортизации сортов в принципе излишне.

Таблица 3. Мультиплексные ампликоны из диапазона 51 – 500 нуклеотидов в ядерных геномах *S.tuberosum* и *S.lycopersicum* / Table 3. Multiplex amplicons from the range 51 – 500 nucleotides in nuclear genomes of *S.tuberosum* and *S.lycopersicum*

Виды / Species	Ампликоны / Amplicons
Solanum tuberosum	74, 94, 168, 180, **182**, 183, 184, **185**, 188, 195, 209, 254, 258, 265, 266, 268, 269, 274, 275, 298, 299, 302, 304, 305, 319, 345, 347, 348, 353, 361, 365, 399, 402, 404, 415, 427, 449, 465, 467, 472, **484**
Solanum lycopersicum	53, 73, 115, 128, 131, 134, 137, 142, 154, **182**, **185**, 196, 216, 242, 251, 259, 293, 314, 317, 331, 341, 344, 351, 357, 359, 362, 364, 366, 367, 368, 370, 371, 372, 373, 374, 376, 377, 378, 379, 380, 381, 382, 418, 423, 459, 472, 477, 481, 482, **484**, 490, 494, 498, 499, 500

Проведенный анализ нуклеотидных последовательностей трех ампликонов размерами 182, 185 и 484 пары нуклеотидов показал, что последние два у этих видов не имеют друг с другом ничего общего, тогда как ампликоны размером 182 пары нуклеотидов на удивление показывают гомологию на уровне выше 75%. При этом нужно отдельно коснуться этих ампликонов размерами 182 пары нуклеотидов, как выяснилось с помощью web-ресурса BLAST пришедшихся на повторяющийся перицентромерный гетерохроматиновый участок, имеющий гомологию с межгенным спейсером рДНК,

впервые обнаруженным у *S.bulbocastanum* [Stupar et al., 2002]. Оказалось, что этот ампликон, помимо этого вида и картофеля с томатом имеет еще довольно высокую гомологию с подобными участками диких видов картофеля *S.pennellii* и *S.pinnatisectum*. Однако все равно считать эти одинаковые по размеру ампликоны родоспецифичными нельзя, поскольку из-за имеющихся инделов у картофеля и томата совпадение по размеру оказалось случайным и у других видов этого рода такой ампликон с данными праймерами или вообще может не образоваться или иметь иные размеры.

Рис. 3. Генетические штрих-коды картофеля *Solanum tuberosum* и томата *Solanum lycopersicum*
Fig. 3. Genetic bar-codes of potato *Solanum tuberosum* and tomato *Solanum lycopersicum*

Из составленных генетических штрих-кодов картофеля и томата можно видеть, что эти виды имеют мало общего, и хотя сейчас они относятся к одному роду *Solanum* еще не так давно томат имел латинское название *Lycopersicum esculentum*, принадлежа к другому роду семейства Пасленовых. Возможно выделение томата в собственный род было более правильным.

Злаковые

К семейству Злаковых относятся основные хлебные культуры, среди которых ряд видов пшениц. Геномы нескольких представителей трибы пшеницевых были проанализированы *in silico* на предмет обнаружения ампликонов, ограниченных теми же шестью указанными выше праймерами.

Как известно мягкая пшеница *T.aestivum* несет гексаплоидный **BAD** геном, являющийся составным из трех субгеномов **B**, **A** и **D**, принадлежащих, как предполагается, одному из диплоидных видов эгилопсов из секции Sitopsis (наиболее вероятно, что *Ae.speltoides*), диплоидной пшенице *T.urartu* и диплоидному эгилопсу *Ae.tauschii*, выступившими их донорами, соответственно. Филогенетические взаимоотношения в пшенично-эгилопсном альянсе рассмотрены нами ранее [Кулуев и др., 2016] и здесь подробно на этих вопросах останавливаться не будем, указав лишь, что также взятые в данное исследование виды тетраплоидных пшениц *T.turgidum* и *T.dicoccoides*, имеющие геномы **BA**, образовались, вероятно, в результате гибридизации *Ae.speltoides* (ставшей материнской формой) и пшеницы *T.urartu*. Причем донор субгенома **B** так до конца и не выяснен, а что касается субгенома **A**, то на его роль предлагались многие виды, но ранее всех на *T.urartu* указали отечественные авторы [Конарев и др. (Konarev et al.), 1974]. Считается также, что один из видов тетраплоидных пшениц, близкий к *T.dicoccoides* и/или *T.turgidum*, вступив в гибридизацию с *Ae.tauschii*, привел к образованию гексаплоидной хлебной пшеницы *T.aestivum*. В этой связи интересно проследить наследование тех или иных ампликонов путем анализа *in silico* геномов этих видов, имеющих базовое число хромосом 7.

Проведенный по-хромосомно (1a, 1b, 1d; 2a, 2b, 2d; 3a, 3b, 3d; 4a, 4b, 4d; 5a, 5b, 5d; 6a, 6b, 6d; 7a, 7b, 7d) *in silico* анализ геномов *T.aestivum* (**BAD**), *T.turgidum* (**BA**), *T.dicoccoides* (**BA**), *Ae.tauschii* (**D**) и *T.urartu*[1] (**A**) с помощью компьютерной программы ABCDNA_GS показал разное количество ампликонов для них. Так, для мягкой гексаплоидной пшеницы выявлено в общей сложности 123 ампликона (субгеном **A** – 55, субгеном **B** – 44, субгеном **D** – 24). Тетраплоидные пшеницы *T.turgidum* и *T.dicoccoides* характеризуются 99 и 96 ампликонами соответственно. Для диплоидных пшеницы *T.urartu* и эгилопса *Ae.tauschii* обнаружено меньшее число ампликонов – 42 и 28 соответственно. При этом часть ампликонов у пшениц и эгилопса совпадает по размеру. В табл. 4 таковые выделены цветом – красным для генома и субгенома **A**, синим – для субгенома **B** и зеленым – для генома и субгенома **D**. Причем выделение производилось, когда, по крайней мере, в двух геномах или субгеномах имелись одинаковые по размеру ампликоны.

[1] Для данного вида по-хромосомный анализ невозможен, поскольку имеются только протяженные контиги.

Таблица 4. Распределение мультиплексных ампликонов из диапазона 51 – 500 нуклеотидов по хромосомам в трибе пшеницевых
Table 4. Distribution of mutiplex amplicons from the range 51 – 500 nucleotides in chromosomes in the Triticeae tribe

Species / Chromosome	1a	1b	1d	2a	2b	2d	3a	3b	3d	4a	4b	4d	5a	5b	5d	6a	6b	6d	7a	7b	7d
Triticum aestivum **BAD** genome	94 320 450 468 491	133 250 348 416 437 476 478 487	55 169 457 494	100 106 131 202 208 211 348 355 357 365 386	121 127 180 353 359 404 487	118 154 228 259 354 406	84 129 167 187 263 344 353 355 402 476 485 498	93 95 165 238 254 294	51 169 290 479	66 242 336 356 421	51 94 113 255 258 288 383 488	75 256	129 274 323 348 360 456	94 98 384 402 482	120 268 406	56 77 130 150 284 348 435	395 439 476 491	188	76 91 169 337 357 359 407 464 478	76 91 136 196 335 439	121 128 456 465
T.turgidum **BA** genome	94 320 468 491	91 248 250 270 359 416 437 476 478		100 106 202 208 355 357	127 180 246 353 358 406 487 489		84 128 167 187 238 264 355 402 498	93 95 165 172 294 319		66 129 242 313 336 356 433	94 113 288 366 488		67 129 247 274 323 348 360 456	94 100 133 381 384 402 481 482 485		77 127 150 284 312 348 435	51 255 359 395 439 476 491		76 91 169 337 357 385 407 464 478	76 100 130 160 439	
T.dicoccoides **BA** genome	94 320 348 468	94 133 172 181 248 249 250 270 277 348 416 476		94 100 106 202 208 211 348 355 357	180 391 487		84 187 238 247 264 342 402 453 486	71 91 93 95 165 238 294		284 313 336 348 411 419	51 94 113 288 488		65 129 264 323	98 100 254 384 481		56 77 150 298 312 348 434 440 442	92 94 123 282 333 395 439 475 491		76 91 169 330 348 357 408 464 478	76 130 160 439 489	
Aegilops tauschii **D** genome			55 129 169 457 495			51 118 121 156 259 354 406			51 169 290 479			75 259			174 268 406			188 189 465			128 398 456 465
T.urartu **A** genome (contigs)	65, 66, 68, 73, 76, 77, 91, 94, 127, 128, 129, 165, 169, 221, 226, 239, 245, 274, 282, 284, 310, 312, 330, 335, 341, 343, 355, 395, 397, 402, 407, 418, 439, 440, 441, 442, 451, 467, 478, 487, 491, 498																				

Однако присутствие одинаковых по размеру фрагментов ДНК не гарантирует их принадлежность одинаковым или гомологичным участкам геномов этих видов. Для подтверждения этого нами были сравнены нуклеотидные последовательности некоторых совпадающих по размеру ампликонов, которые для простоты будут здесь обозначаться по хромосоме/геному и размеру. Так, ампликон 1d55 оказался схож с ампликоном 1D55, отличаясь только заменами двух нуклеотидов, а ампликон 1d494 гомологичен ампликону 1D495, неся делецию одного нуклеотида, что не удивительно, поскольку, как уже говорилось выше, *Ae.tauschii* послужил донором субгенома **D** полиплоидных пшениц. При этом стоит заметить, что ампликоны 1d494 и 1D495, несмотря на то, что они принадлежат одинаковым участкам ДНК также как и для арабидопсиса ампликоны Col331 и C330 для целей ДНК-паспортизации считать одинаковыми не следует, поскольку предлагаемый подход с генетическим штрих-кодированием на основе размеров ампликонов не предполагает установление нуклеотидных последовательностей даже части ампликонов, так как рассчитан на относительно быструю и недорогостоящую ДНК-идентификацию того или иного сорта.

На основе размеров ампликонов, ограниченных указанными выше шестью мультиплексными праймерами, для исследованных видов пшениц и эгилопса построены генетические штрих-коды, приведенные на рис. 4, позволяющие визуально наблюдать существенные различия между ними. Для мягкой пшеницы был также построен генетический штрих-код на основе мультиплексного виртуального RAPD-анализа с тремя праймерами, образовавшими 54 ампликона (данные не приведены), что возможно и достаточно для выявления полиморфизма ДНК сортов мягкой пшеницы, но большее число ампликонов с шестью мультиплексными праймерами обеспечат их ДНК-паспортизацию увереннее.

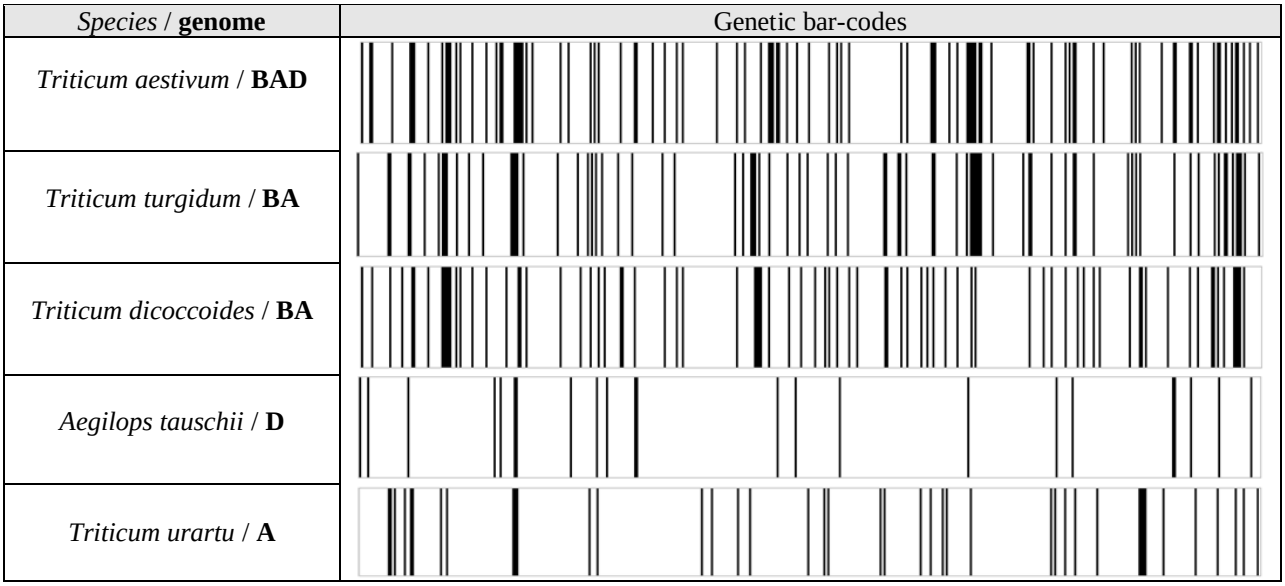


Рис. 4. Генетические штрих-коды ряда видов пшенично-эгилопсного альянса
Fig. 4. Genetic bar-codes of several species of wheat-aegilops alliance

Определенный интерес представляет выяснение филогенетических отношений исследованных видов пшениц и эгилопсов путем сравнения образовавшихся мультиплексных ампликонов, происходящих из фактически случайных мест их геномов, но этому вопросу будет посвящена отдельная статья.

Также из семейства Злаковых проведен *in silico* анализ двух подвидов риса – *O.sativa* japonica и *O.sativa* indica, представляющих собой диплоидные организмы с размерами геномов около 500 млн.п.н. Сравнение двух подвидов риса в большей степени подходит для прогнозирования имеющихся различий между отдельными сортами, чем аллополиплоидные геномы пшениц.

Для *O.sativa* indica обнаружены 33 ампликона, а для *O.sativa* japonica их оказалось немного меньше – 29, из которых с первым подвидом совпали 20. На основе полученных размеров ампликонов были построены генетические штрих-коды проанализированных двух подвидов риса, представленные на рис. 5.

Таблица 5. Мультиплексные ампликоны из диапазона 51 – 500 нуклеотидов в ядерных геномах двух подвидов риса *Oryza sativa* indica и *Oryza sativa* japonica

Table 5. Multiplex amplicons from the range 51 – 500 nucleotides in nuclear genomes of two subspecies of rice *Oryza sativa* indica and *Oryza sativa* japonica

Вид / Species	Ампликоны / Amplicons
Oryza sativa indica	55, 61, 74, 86, 93, 101, 103, 114, 133, 140, 149, 150, 151, 152, 179, 269, 274, 292, 293, 305, 381, 403, 407, 412, 413, 414, 415, 435, 454, 459, 460, 464, 478
Oryza sativa japonica	55, 61, 77, 93, 103, 114, 127, 133, 140, 150, 151, 152, 274, 292, 293, 305, 340, 341, 381, 399, 412, 413, 414, 415, 436, 454, 464, 480, 486

Species / varieties	Genetical bar-codes
Oryza sativa indica	
Oryza sativa japonica	

Рис. 5. Генетические штрих-коды подвидов риса *Oryza sativa* indica и *Oryza sativa* japonica
Fig. 5. Genetical barcodes of rice subspecies *Oryza sativa* indica and *Oryza sativa* japonica

Как можно видеть генетические штрих-коды двух подвидов риса схожи друг с другом, но при этом имеются и некоторые различия. Число совпадающих ампликонов оказалось несколько меньше, чем у отдельных разновидностей резуховидок, что в целом не противоречит нашей концепции использования случайных фрагментов генома (RAPD-ампликонов) для генетической паспортизации сортов, сортоообразцов и линий сельскохозяйственных растений.

Обсуждение результатов

Ранее нами был значительно усовершенствован предложенный для генотипирования бактерий метод RFEL – Restriction Fragment End Labeling [van Steenbergen et al., 1995], что позволило присваивать штаммам микроорганизмов уникальные генетические штрих-коды, формируемые на основе электрофоретического разделения рестрикционных фрагментов ДНК в секвенирующем геле в денатурирующих условиях [Баймиев и др. (Baymiev et al.), 1999]. С его помощью были успешно генотипированы различные штаммы клубеньковых бактерий из группы ризобий. Причем, близкородственные штаммы характеризовались большим числом совпадающих фрагментов ДНК, тогда как далеко отстоящие штаммы показывали лишь незначительное совпадение полос по их размерам. Однако ввиду того, что эукариотические организмы, включая растения, имеют геномы гораздо большего размера, то подход с Концевым Мечением Рестрикционных Фрагментов (КМРФ) для них не применим, даже при использовании рестрикционных эндонуклеаз с октануклеотидными сайтами узнавания, которые теоретически встречаются в среднем через каждые 65536 (4^8) нуклеотидов и таким образом практически все воображаемые ДНК-ячейки в этом случае будут заняты фрагментами ДНК. Заменой метода КМРФ для генотипирования (генетического штрих-кодирования / ДНК-паспортизации) растений можно считать мультиплексную ПЦР с подходящим числом праймеров небольшой длины – обычно декамерными, компьютерный анализ которой и был проведен в данной работе.

Так, *in silico* анализ ряда геномов родственных видов растений на предмет образования в виртуальной мультиплексной ПЦР достаточного числа различающихся ампликонов показал, что теоретические ожидания полностью подтвердились и путем экстраполяции результатов можно допустить, что выявляемый полиморфизм позволит создавать генетические штрих-коды сортов, сортообразцов и линий сельскохозяйственных культур для их ДНК-паспортизации и каталогизации и создания соответствующих баз данных по ним. Причем экстраполяция результатов здесь касается потенциального разнообразия геномов и не связана с фенотипическими проявлениями, характерными для тех или иных сортов, и поэтому вполне правомочна. Использование шести декамерных праймеров в мультиплексной виртуальной ПЦР и создание на основе выявляемых ампликонов генетических штрих-кодов убеждает в том, что такое количество праймеров можно успешно применять для геномов с размерами от одного млрд. пар нуклеотидов и выше. Для растений с небольшими геномами количество мультиплексных праймеров следует увеличивать. При этом для достижения более высокой эффективности амплификации и воспроизводимости результатов декамерные праймеры можно заменить 12-ти звенными, но их число в мультиплексной ПЦР также должно быть увеличено, по крайней мере, до 15-20, что, впрочем, зависит и от их нуклеотидных последовательностей. (Данные по нахождению таких праймеров в геномах исследуемых видов растений здесь не приведены.) Другим способом улучшить процесс амплификации является использование модифицированных нуклеотидов в праймерных последовательностях, например в виде их LNA-производных (Locked Nucleic Acid), повышающих температуру плавления по сравнению с обычной ДНК приблизительно на 2-6°C на одно модифицированное основание [Latorra et al., 2003].

Относительно недавно нами отмечалась возможность ДНК-паспортизации сортов растений с помощью не только RAPD-анализа, но и иных подходов, способных обеспечить точное (до нуклеотида) установление размеров ампликонов, например AFLP, SSR или ISSR амплификации с любыми праймерными комбинациями [Кулуев и др. (Kuluev et al.), 2018]. При этом обсуждалась возможность помещения информации о полиморфизме ДНК в единую базу данных, поскольку число комбинаций даже в выбранном диапазоне от 51 до 500 нуклеотидов настолько велико, что случайное совпадение генетических штрих-кодов разных образцов независимо от используемого метода и того или иного комплекта праймеров практически невозможно. Но в момент написания той нашей обзорной статьи подобных экспериментальных работ еще не было и они стали появляться позже. При этом для ДНК-идентификации внесенного в такую базу данных какого-либо сорта необходимо воспользоваться тем же комплектом праймеров и поэтому в базе данных обязательно должны указываться последовательности использованных для

создания генетического штрих-кода того или иного вида растений комплекта праймеров, что не составляет каких-либо трудностей и не должен относиться к секретной коммерческой информации. Недавно сообщено о создании специализированной базы данных Plant International DNA-fingerprinting System (PIDS) [Jiang et al., 2020], находящейся по адресу https://ssr.PIDS.online:8445/ssr. Они решили использовать для выявления полиморфизма ДНК такой метод как SSR (Simple Sequence Repeats), называемых также микросателлитами, и упомянули о том, что при выборе праймеров для формирования базы данных необходимо учитывать обеспечиваемый ими высокий уровень полиморфизма и что с такими праймерами мультиплексные варианты пока не разработаны. Но к сожалению, ни в данной статье, ни на организованном ими сервере мы не нашли нуклеотидных последовательностей используемых праймеров, при том, что это очень важно для последующих сравнений. Фактически в эту базу данных предлагается загружать изображения разделений ампликонов капиллярным гель-электрофорезом, генерируемые написанной этими авторами программой SSR Analyser для каждого праймера по отдельности. Из представленных результатов можно видеть, что в качестве генетического анализатора использовалась модель 3500 фирмы Applied Biosystems и диапазон разделяемых ампликонов составлял от 20 до 500 и более нуклеотидов. В цитируемой статье отмечается, что, помимо селекционеров, которым эта база данных в первую очередь предназначена, она может также использоваться и для других объектов, включая решение криминалистических задач.

Ранее для поиска *in silico* микросателлитных последовательностей в секвенированных геномах растений разработано web-приложение MISA-web [Beier et al., 2017]. По микросателлитным ДНК есть и другие базы данных, в частности PMDBase, содержащая информацию по 110 видам растений [Yu et al., 2017]. В другой работе было доложено о формировании базы данных для 73 образцов салата-латука *Lactuca sativa* capitate на основе анализа полиморфизма микросателлитных последовательностей, которые выявлялись с помощью ПЦР с использованием 23 пар праймеров [Zhou et al., 2019]. Причем в той статье приведены их последовательности, а в приложении для 19 пар праймеров даны точные размеры выявленных аллельных локусов, варьирующие в целом в пределах от 151 до 370 нуклеотидов, а также бинарное кодирование 73 образцов. При этом показано, что для однозначной идентификации такого количества сортов вполне достаточно даже 10 пар праймеров. Однако RAPD-анализ в предлагаемом нами варианте теоретически занимает более широкий диапазон, в котором могут с точностью до нуклеотида измеряться ампликоны, что обеспечивает некоторое преимущество перед SSR амплификацией, поскольку ввиду особенности организации SSR-локусов, коротких ампликонов для них просто быть не может.

Заключение

Вооружить ученых-аграриев, а также производственников удобным, хорошо воспроизводимым и при этом относительно недорогим методом ДНК-паспортизации (каталогизации) сортов и их ДНК-идентификации является актуальнейшей задачей, решение которой могло бы со временем поднять сельское хозяйство на небывалую высоту. При этом нынешняя реальность такова, что подобное случится, к сожалению, нескоро. К сожалению, в силу множества причин использование полиморфизма ДНК в аграрном комплексе встречается пока довольно редко. Тем не менее, довольно широкое распространение в настоящее время в научных исследованиях для выявления полиморфизма ДНК разных сортов/сортообразцов/линий получил обычный RAPD-анализ, причиной чему служит простота этого метода, не требующего сложного дорогостоящего оборудования. Но с его помощью невозможно получать истинно цифровые данные, которые можно уверенно сравнивать между собой и формировать на их основе современные базы данных. Использование мультиплексного RAPD-анализа с разделением коротких ампликонов (их одноцепочечных вариантов) с помощью капиллярного гель-электрофореза обеспечит получение истинно цифровых сведений, которые позволят однозначно ДНК-паспортизировать, а потом и ДНК-идентифицировать любые сорта любых сельскохозяйственных растений, включая анализ посевного материала. В этом случае итоговый результат перевешивает мнимую экономию средств, когда люди обходятся более простыми и дешевыми приборами. Тем более что генетические анализаторы капиллярного типа, требующиеся для точного установления размеров ампликонов, уже стали довольно рутинным оборудованием. К тому же в России ООО «Синтол» совместно с Институтом аналитического приборостроения РАН налажен серийный выпуск 8-ми капиллярного генетического анализатора модели «Нанофор-05», ничуть не уступающего зарубежным аналогам и при этом приблизительно вдвое их дешевле. За рабочий день с помощью одного такого прибора можно проанализировать до 24 сортов, сортообразцов, линий сельскохозяйственных растений. Себестоимость подобного анализа (ДНК-паспортизации / ДНК-идентификации) одного образца без учета зарплаты операторов, амортизации приборов и коммунальных платежей составит не более 100 рублей на образец.

Завершая данную статью, нужно заметить, что предлагаемый нами подход вполне применим и для ДНК-паспортизации пород животных, рас грибов, штаммов микроорганизмов. При этом лишь требуется учитывать размеры их геномов и корректировать число используемых мультиплексных праймеров. Причем база данных по ним всем может быть единой. Или разделенной на подбазы по царствам живых организмов. Здесь можно еще заметить, что в данном номере журнала есть другая наша статья [Гарафутдинов и др. (Garafutdinov et al.), 2020], где «проведена» виртуальная ПЦР с тем же мультиплексным комплектом из шести праймеров генома лошади, позволившая построить соответствующий генетический штрих-код этого вида сельскохозяйственных животных.

Благодарности

Данное исследование в рамках ГК №05.621.21.0033 выполнено с использованием оборудования РЦКП «Агидель» и УНУ «КОДИНК», поддержано грантом

РФФИ №17-44-020120, а также выполняемыми Госзаданиями по темам № АААА-А19-119021190011-0 и АААА-А16-116020350028-4 и АААА-А19-119021890026-7.

Литература

1. Баймиев Ал.Х., Чемерис А.В., Вахитов В.А. Анализ информативности некоторых современных методов идентификации полиморфизма ДНК микроорганизмов на примере симбиотических клубеньковых бактерий Rhizobium galegae // Генетика. 1999. Т.35. С.1613-1621.
2. Гарафутдинов Р.Р., Баймиев Ан.Х., Малеев Г.В., Алексеев Я.И., Зубов В.В., Чемерис Д.А., Кирьянова О.Ю., Губайдуллин И.М., Матниязов Р.Т., Сахабутдинова А.Р., Никоноров Ю.М., Кулуев Б.Р., Баймиев Ал.Х., Чемерис А.В. Разнообразие праймеров для ПЦР и принципы их подбора. *Биомика*. 2019. Т.11(1). С. 23 – 70. DOI: 10.31301/2221-6197.bmcs.2019-04
3. Гарафутдинов Р.Р., Гайнуллина К.П., Кирьянова О.Ю., Юрина А.В., Долматова И.Ю., Логинов О.Н., Чемерис А.В. Полиморфизм ДНК лошади *Equus caballus* и методы его выявления // Биомика. 2020. Т.12(2). С. 272-299. DOI: 10.31301/2221-6197.bmcs.2020-16
4. Кирьянова О.Ю., Кирьянов И.И., Кулуев Б.Р., Чемерис А.В., Гарафутдинов Р.Р., Губайдуллин И.М. Свидетельство о государственной регистрации программы для ЭВМ № 2020610703 ABCDNA_GS (Amplified Bar-Coded DNA Genome/Specimen) от 17.01.2020 г.
5. Конарев А.В., Гаврилюк И.П., Мигушова Э.Ф. Дифференциация диплоидных пшениц по данным иммунохимического анализа глиадина // Доклады ВАСХНИЛ. 1974. №6. С.12.
6. Кулуев А.Р., Матниязов Р.Т., Чемерис Д.А., Чемерис А.В. Современные представления о родственных взаимоотношениях в пшенично-эгилопсном альянсе (с краткой исторической справкой) // Биомика. 2016. Т. 8. № 4. С. 297-310.
7. Кулуев Б.Р. Методы ПЦР для выявления мультилокусного полиморфизма ДНК у эукариот, основанные на случайном праймировании / Кулуев Б.Р., Баймиев Ан.Х., Геращенков Г.А., Чемерис Д.А., Зубов В.В., Кулуев А.Р., Баймиев Ал.Х., Чемерис А.В. // Генетика. 2018. Т. 54. С. 495-511. DOI: 10.7868/S0016675818050016
8. Нигматуллина Н.В., Кулуев А.Р., Кулуев Б.Р. Молекулярные маркеры, применяемые для определения генетического разнообразия и видоидентификации дикорастущих растений. *Биомика*. 2018. Т10(3). С. 290-318. DOI: 10.31301/2221-6197.bmcs.2018-39
9. Сухарева А.С., Кулуев Б.Р. ДНК-маркеры для генетического анализа сортов культурных растений // *Biomics*. 2018. Т. 10. №1. С. 69–84. DOI: 10.31301/2221-6197.bmcs.2018-15
10. Beier S., Thiel T., Münch T., Scholz U., Mascher M. MISA-web: a web server for microsatellite prediction // *Bioinformatics*. 2017. V.33. P.2583–2585. doi: 10.1093/bioinformatics/btx198
11. Corley-Smith G.E., Lim C.J., Kalmar G.B., Brandhorst B.P. Efficient detection of DNA polymorphisms by fluorescent RAPD analysis // Biotechniques. 1997. V. 22. P. 690-699. doi: 10.2144/97224st04
12. Jiang B., Zhao Y., Yi H., Huo Y., Wu H., Ren J., Ge J., Zhao J., Wang F. PIDS: A user-friendly plant DNA fingerprint Database management system // Genes (Basel). 2020. V.11(4):373. doi: 10.3390/genes11040373
13. Latorra D., Campbell K., Wolter A., Hurley J.M. Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers // Hum Mutat. 2003. V. 22(1). P. 79-85. doi: 10.1002/humu.10228
14. Stupar R.M., Song J., Tek A.L., Cheng Z., Dong F., Jiang J. Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats // Genetics. 2002. V.162(3). P.1435-1444.
15. van Steenbergen T.J., Colloms S.D., Hermans P.W., de Graaff J., Plasterk R.H. Genomic DNA fingerprinting by Restriction Fragment End Labeling // Proc. Natl. Acad. Sci. USA. 1995. V.92(12). P.5572-5576. doi: 10.1073/pnas.92.12.5572
16. Williams J.G., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers // Nucleic Acids Res. 1990. V. 18. P. 6531-6535. doi: 10.1093/nar/18.22.6531
17. Yu J., Dossa K., Wang L., Zhang Y., Wei X., Liao B., Zhang X. PMDBase: A database for studying microsatellite DNA and marker development in plants // *Nucleic Acids Res*. 2017. V.45(D1). D1046-D1053. doi: 10.1093/nar/gkw906
18. Zhou, H., Zhang, P., Luo, J. et al. The establishment of a DNA fingerprinting database for 73 varieties of Lactuca sativa capitate L. using SSR molecular markers // Hortic. Environ. Biotechnol. 2019. V.60. P. 95–103. doi: 10.1007/s13580-018-0102-3

References

1. Baymiev Al.Kh., Chemeris A.V., Vakhitov V.A. Informative value of some modern methods for DNA polymorphism identification in microorganisms as exemplified by symbiotic root-nodule bacteria *Rhizobium galegae*. *Russian Journal of Genetics*. 1999. Т. 35. № 12. С. 1387-1393.
2. Beier S., Thiel T., Münch T., Scholz U., Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017. V.33. P.2583–2585. doi: 10.1093/bioinformatics/btx198
3. Corley-Smith G.E., Lim C.J., Kalmar G.B., Brandhorst B.P. Efficient detection of DNA polymorphisms by fluorescent RAPD analysis. *Biotechniques*. 1997. V. 22. P. 690-699. doi: 10.2144/97224st04
4. Garafutdinov R.R., Baymiev An.Kh., Maleev G.V., Alexeyev Ya.I., Zubov V.V., Chemeris D.A., Kiryanova J.Yu., Gubaydullin I.M., Matniyazov R.T., Sakhabutdinova A.R., Nikonorov Yu.M., Kuluev B.R., Baymiev Al.Kh., Chemeris A.V. Diversity of PCR primers and principles of their design. *Biomics*. 2019. V.11(1). P. 23 – 70. DOI: 10.31301/2221-6197.bmcs.2019-04
5. Garafutdinov R.R., Gainullina K.P., Kiryanova O.Yu., Yurina A.V., Dolmatova I.Yu., Loginov O.N., Chemeris A.V. DNA polymorphism of horse *Equus caballus* and

methods of its detection. *Biomics*. 2020. Vol. 12(2). P. 272-299. DOI: 10.31301/2221-6197.bmcs.2020-16

6. Jiang B., Zhao Y., Yi H., Huo Y., Wu H., Ren J., Ge J., Zhao J., Wang F. PIDS: A user-friendly plant DNA fingerprint Database management system. *Genes (Basel)*. 2020. V.11(4):373. doi: 10.3390/genes11040373
7. Kiryanova O. Yu., Kiryanov I. I., Kuluyev B. R., Chemeris A.V., Garafutdinov R. R., Gubaidullin I. M. Certificate of state registration of computer program no. 2020610703 ABCDNA_GS (Amplified Bar-Coded DNA Genome/Specimen) dated 17.01.2020
8. Konarev A.V., Gavriljuk I.P., Migushova E.F. Differentiation of diploid wheat as indicated by immunochemical analysis. *Proceeding of Soviet Agr. Sci. (USSR)*. 1974. No. 6. P.12.
9. Kuluev A.R., Matniyazov R.T., Chemeris D.A., Chemeris A.V. Modern concepts about relationships in the wheat-aegilops alliance (with a brief historical note). *Biomics*. 2016. V.8(4). P.297-310. (In Russian)
10. Kuluev B.R., Baymiev An.K., Gerashchenkov G.A., Chemeris D.A., Zubov V.V., Kuluev A.R., Baymiev Al.Kh., Chemeris A.V. Random priming PCR strategies for identification of multilocus DNA polymorphism in eukaryotes. *Russian Journal of Genetics*. 2018. V. 54(5). P. 499-513. DOI: 10.1134/S102279541805006X
11. Latorra D., Campbell K., Wolter A., Hurley J.M. Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. *Hum. Mutat*. 2003. V. 22(1). P. 79-85. doi: 10.1002/humu.10228
12. Nigmatullina N.V., Kuluev A.R., Kuluev B.R. Molecular markers used to determine the genetic diversity and species identification of wild plants. *Biomics*. 2018. V.10(3). P. 290-318. DOI: 10.31301/2221-6197.bmcs.2018-39
13. Stupar R.M., Song J., Tek A.L., Cheng Z., Dong F., Jiang J. Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. *Genetics*. 2002. V.162(3). P.1435-1444.
14. Sukhareva A.S., Kuluev B.R. DNA markers for genetic analysis of crops. *Biomics*. 2018. 10(1). P. 69-84. DOI: 10.31301/2221-6197.bmcs.2018-15 (In Russian).
15. van Steenbergen T.J., Colloms S.D., Hermans P.W., de Graaff J., Plasterk R.H. Genomic DNA fingerprinting by Restriction Fragment End Labeling. *Proc. Natl. Acad. Sci. USA*. 1995. V.92(12). P.5572-5576. doi: 10.1073/pnas.92.12.5572
16. Williams J.G., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res*. 1990. V. 18. P. 6531-6535. doi: 10.1093/nar/18.22.6531
17. Yu J., Dossa K., Wang L., Zhang Y., Wei X., Liao B., Zhang X. PMDBase: A database for studying microsatellite DNA and marker development in plants. *Nucleic Acids Res*. 2017. V.45(D1). D1046-D1053. doi: 10.1093/nar/gkw906
18. Zhou, H., Zhang, P., Luo, J. et al. The establishment of a DNA fingerprinting database for 73 varieties of Lactuca sativa capitate L. using SSR molecular markers. *Hortic. Environ. Biotechnol*. 2019. V.60. P. 95–103. doi: 10.1007/s13580-018-0102-3