

EDN: OIPJJV

УДК 519–7; 57.065; 581.1.577.21

Virtual *In Silico* PCR in Two-Dimensional Format as a Tool for Elucidating Phylogenetic Relationship in Allopolyploid Forms with Wheats and Their Wild Relatives *Aegilops* Used as an Example

Olga Yu. Kiryanova^{*a},

Azat R. Kuluev^b, Irek M. Gubaydullin^{a, c},

Bulat R. Kuluev^b, Alexey V. Chemeris^b

^a*Ufa State Petroleum Technological University*

^b*Institute of Biochemistry and Genetics –
Subdivision of the Ufa Federal Research Centre RAS*

^c*Institute of Petrochemistry and Catalysis RAS
Ufa, Russian Federation*

Received 16.12.2022, received in revised form 28.04.2023, accepted 09.07.2023

Abstract. Identification of natural donors of three bread wheat (*Triticum aestivum* L.) **BAD** subgenomes is of great importance for developing technologies aimed at improving this crop. The question about the species and subspecies of the wheat–*Aegilops* alliance serving as donors of the **B** and **A** subgenomes remains unanswered, while the *Aegilops tauschii* subspecies of *strangulata* is considered as a donor of the **D** subgenome. To identify the donor species of subgenomes of polyploid forms, the comparison of the nucleotide sequences of various genes, as well as fragments of repetitive DNA, is performed to construct phylogenetic trees. In that case, however, only one or a few genetic systems or loci are taken into analysis. The genomic barcoding proposed by the authors, which is not tied to any genetic system, has an advantage, since *in silico* RAPD analysis «finds» fragments matching in size in the entire genome, making, as it were, a complete «slice» of it. To estimate the phylogenetic relationship of different wheat species, we used the method of virtual multiplex RAPD analysis with 20 undecamer primers, which made it possible to create genomic barcodes of these species and compile two-dimensional maps for individual chromosomes of the analyzed wheat species. The proposed method of computer analysis of genomes showed that *T. aestivum* subgenome **D** and *Ae. tauschii* are very similar to each other, which may indicate their common origin. No such definite conclusions could be drawn about the donors of

© Siberian Federal University. All rights reserved

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

* Corresponding author E-mail address: olga.kiryanova27@gmail.com

ORCID: 0000-0002-8563-1244 (Kuluev A.); 0000-0002-9848-2882 (Gubaydullin I.); 0000-0002-1564-164X (Kuluev B.); 0000-0002-8917-0449 (Chemeris A.)

subgenomes **A** and **B**, which is probably due to the more ancient association of these subgenomes in tetraploid wheat and the accumulation of a larger number of mutations over this time.

Keywords: *Triticum aestivum*, *Aegilops tauschii*, *Triticum urartu*, *Aegilops splotoides*, *in silico* RAPD analysis, computer simulation of PCR, genomic barcoding.

Acknowledgements. The study was supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement no. 075–15–2021–1066, September 28, 2021) and State Assignment no. 122030200143–8.

Citation: Kiryanova O. Yu., Kuluev A. R., Gubaydullin I. M., Kuluev B. R., Chemeris A. V. Virtual *in silico* PCR in two-dimensional format as a tool for elucidating phylogenetic relationship in allopolyploid forms with wheats and their wild relatives *Aegilops* used as an example. *J. Sib. Fed. Univ. Biol.*, 2024, 17(1), 45–63. EDN: OIPJVV



Виртуальная *in silico* ПЦР в двумерном формате как инструмент для выяснения филогенетического родства у аллополиплоидных форм на примере пшениц и их диких сородичей эгилопсов

О. Ю. Кирьянова^а, А. Р. Кулуев^б,

И. М. Губайдуллин^{а, в}, Б. Р. Кулуев^б, А. В. Чемерис^б

^а*Уфимский государственный нефтяной технический университет*

^б*Институт биохимии и генетики –*

обособленное структурное подразделение ФГБНУ

Уфимского федерального исследовательского центра РАН

^в*Институт нефтехимии и катализа –*

обособленное структурное подразделение ФГБНУ

Уфимского федерального исследовательского центра РАН

Российская Федерация, Уфа

Аннотация. Определение природных доноров трех субгеномов **ВАД** мягкой пшеницы (*Triticum aestivum* L.) имеет большое значение в целях разработки технологий по дальнейшему совершенствованию данной культуры. До сегодняшнего дня остаются открытыми некоторые вопросы происхождения субгеномов **В** и **А** мягкой пшеницы, тогда как донором субгенома **Д** считается *Aegilops tauschii* подвид *strangulata*. Для установления видов-доноров субгеномов полиплоидных форм активно применяется сравнение нуклеотидных последовательностей различных генов, а также фрагментов повторяющейся ДНК, что позволяет строить филогенетические древа. Однако в этом случае в анализ берется лишь одна или несколько генетических систем или локусов. Предложенное нами геномное штрихкодирование, не привязанное к какой-либо генетической системе, имеет преимущество, поскольку *in silico*

RAPD-анализ «находит» одинаковые по размеру участки сразу всего генома, делая как бы его полный «срез». Для оценки филогенетического родства разных видов пшеницы использовали метод виртуального мультиплексного RAPD-анализа с 20 ундекамерными праймерами, что позволило создать геномные штрихкоды этих видов, сопровождаемые двумерными картами, составленными для отдельных хромосом анализируемой пшеницы. Предложенный метод компьютерного анализа геномов показал, что *T. aestivum* субгеном **D** и *Ae. tauschii* весьма схожи между собой, что подтверждает их общее происхождение. В случае с донорами субгеномов **A** и **B** такие однозначные выводы сделать не удалось, что, возможно, связано с более древним объединением этих субгеномов в тетраплоидной пшенице и накоплением за это время большего количества мутаций.

Ключевые слова: *Triticum aestivum*, *Aegilops tauschii*, *Triticum urartu*, *Aegilops speltoides*, *in silico* RAPD-анализ, компьютерное моделирование ПЦР, геномное штрихкодирование.

Благодарности. Работа выполнена при финансовой поддержке гранта Минобрнауки РФ (соглашение № 075–15–2021–1066 от 28 сентября 2021 г), а также в рамках государственного задания № 122030200143–8.

Цитирование: Кирьянова О. Ю. Виртуальная *in silico* ПЦР в двумерном формате как инструмент для выяснения филогенетического родства у аллополиплоидных форм на примере пшениц и их диких сородичей эгилопсов / О. Ю. Кирьянова, А. Р. Кулуев, И. М. Губайдуллин, Б. Р. Кулуев, А. В. Чемерис // Журн. Сиб. федер. ун-та. Биология, 2024. 17(1). С. 45–63. EDN: OIPJVV

Введение

Основной хлебный злак – мягкая пшеница *Triticum aestivum* L. представляет собой гексаплоидный вид с геномной формулой **BBAADD**, образовавшийся, как считается, всего около 10 тысяч лет назад. Однако какие именно виды пшениц и их ближайших сородичей – эгилопсов (*Aegilops* L.) послужили донорами трех субгеномов **B**, **A** и **D** до сих пор до конца неясно. Точно установленным считается, что последний субгеном **D** привнесен в некую тетраплоидную форму пшеницы эгилопсом *Aegilops tauschii* и, скорее всего, подвидом *stragulata*. При этом существует огромное количество экспериментальных и обзорных работ, в которых уделено внимание происхождению мягкой пшеницы и потенциальным донорам ее субгеномов (Мигушова, 1975; Goncharov, 2011; Haider, 2013 и др.). Нами тоже неоднократно уделялось внимание этим вопросам (Vakhitov et al., 2003;

Кулуев и др., 2016, 2020). Появившаяся возможность секвенировать полные геномы различных организмов привела к тому, что к настоящему времени таковые стали уже известны для многих видов организмов всех уровней генетической сложности. На сайтах <http://plants.ensembl.org/index.html> и <https://www.ncbi.nlm.nih.gov/genome/> в свободном доступе выложены полные геномы ряда видов пшенично-эгилопсного комплекса, включая саму гексаплоидную пшеницу *T. aestivum*, а также потенциальных доноров ее субгеномов, что позволяет проводить их всестороннее исследование.

С целью оценки филогенетического родства разных видов пшеницевых нами применен метод виртуального мультиплексного RAPD-анализа с 20 ундекамерными праймерами, что позволило создать геномные штрихкоды этих видов, сопроводив их двумерными картами, составленными для

отдельных хромосом анализируемых видов пшеницевых. Классический RAPD-анализ предполагает использование одного короткого обычно декамерного праймера, с помощью которого в ходе ПЦР образуется некоторое количество фрагментов ДНК, разделяемых гель-электрофорезом (Kuluev et al., 2018). При этом разделение в агарозном геле не обеспечивает определения размера ампликонов с точностью до нуклеотида, что очень важно для цифровизации данных, исключающих какие-либо неточности в виде \pm сколько-то нуклеотидов. Поэтому необходимо использовать секвенирующий гель-электрофорез в приборах капиллярного типа. Ранее нами был подробно описан принцип геномного штрихкодирования (Кирьянова и др., 2020), и поэтому здесь на нем останавливаться не будем, заметив лишь, что рассматриваемый нами диапазон длин потенциальных ампликонов от 51 до 500 п.н. обеспечивает гигантское число комбинаций, делая этот метод пригодным и для ДНК-паспортизации сортов любых видов растений, но для его использования на практике необходим этап анализа *in silico* на предмет подбора оптимального количества праймеров с конкретными нуклеотидными последовательностями. Возвращаясь к проблеме установления видов-доноров субгеномов полиплоидных форм пшениц, необходимо отметить, что с этой целью активно применяется сравнение нуклеотидных последовательностей различных генов, а также фрагментов повторяющейся ДНК, что позволяет строить филогенетические деревья. Однако в этом случае в анализ берется лишь одна или несколько генетических систем или локусов, которые могут эволюционировать по своим законам, не отражающим в полной мере реальную взаимосвязь исследуемых видов растений. Подход с геномным штрихкодированием,

не привязанный к какой-либо генетической системе или локусу, имеет в этом плане преимущество, поскольку *in silico* RAPD-анализ «находит» одинаковые по размеру участки сразу всего генома, делая как бы его полный «срез». При этом следующий этап в виде их двумерного *in silico* RAPD-анализа, примененного в данной работе, позволяет видеть не просто совпадение по размеру случайных фрагментов ДНК, но показывает их локализацию в хромосомах и позволяет принимать решение без установления (также *in silico*) их нуклеотидных последовательностей, что, впрочем, принципиально возможно, и нами ранее для контроля осуществлялось (Кирьянова и др., 2020).

Материалы и методы

Праймеры были сгенерированы с помощью разработанной нами программы генератора нуклеотидных последовательностей GATCGGenerator (Кирьянова и др., 2021). Общей особенностью этих праймеров является GC-состав, составляющий около 45 %, что делает их относительно AT-богатыми (55 %), что важно для растительных организмов и приблизительно совпадает с таковым для геномов пшеницевых. Также для исключения образования гетеродимеров в ходе реальной («мокрой») амплификации эти праймеры большей частью несут в себе лишь три нуклеотида из четырех возможных (Гарафутдинов и др., 2019).

Из всего множества сгенерированных праймеров вручную были отобраны 6 наборов тестовых праймеров: два набора из 6 праймеров, три набора из 12 праймеров. Было проведено *in silico* исследование нескольких наборов праймеров, в результате которого были отобраны праймеры, которые дают достаточное количество полос и в комплексе позволяют зрительно анализировать разницу

Таблица 1. Пример результатов поиска мест отжига ампликонов для генома *T. aestivum*Table 1. An example of results of the search for amplicon annealing sites for the *T. aestivum* genome

file	forward primer	forward primer position	reverse primer	reverse primer position	amplicon size
Triticum_aestivum.IWGSC.dna.chromosome.6B.fa	CGCAGAACCA	12886800	TGACTGCTGCT	12886858	58
Triticum_aestivum.IWGSC.dna.chromosome.7A.fa	AACGCCAAGA	311963094	TGGTGGTGGCT	311963153	59
Triticum_aestivum.IWGSC.dna.chromosome.1A.fa	AACGAACACA	83371490	TGTGTTCGGTT	83371557	67
Triticum_aestivum.IWGSC.dna.chromosome.7B.fa	AACGAACACA	348521709	TGTGTTCGGTT	348521776	67
Triticum_aestivum.IWGSC.dna.chromosome.1A.fa	AACGAACACA	439841303	TGTGTTCGGTT	439841371	68
Triticum_aestivum.IWGSC.dna.chromosome.3B.fa	AACGAACACA	525126456	TGTGTTCGGTT	525126524	68
Triticum_aestivum.IWGSC.dna.chromosome.5A.fa	AACGAACACA	507492160	TGTGTTCGGTT	507492228	68

ма был подготовлен отдельный файл с входными данными на основе найденных мест отжига праймеров в формате *.xlsx. Структура файла выглядит следующим образом: имя файла – название генома, каждый лист файла – информация об отдельной хромосоме. Лист именуется в виде порядкового номера хромосомы, или можно добавить букву, обозначающую субгеном. На каждом листе отображается информация в виде таблицы: Primer_position – позиции отжига прямого праймера (для определения локализации ампликона в хромосоме), Amplicon – размер ампликона.

В результате такого сравнения получается карта распределения ампликонов внутри хромосомы, на которой черным цветом обозначены совпадающие по размеру ампликоны, белым (прозрачный круг) и серым цветом – ампликоны, характерные только для одного из сравниваемых геномов. Пример такого сравнения показан на рис. 1.

Принцип разработанного нами подхода *in silico* анализа представлен на рис. 2. Здесь

наглядно показана последовательность действий, а также данные, которые получены в результате каждого шага.

После проведенного двумерного сравнения были определены ампликоны, совпадающие по размеру и наиболее близко расположенные друг к другу у сравниваемых геномов с учетом их коллинеарности (на рис. 1 они изображены как близко расположенные черные круги). Для этих ампликонов был проведен сравнительный анализ на уровне нуклеотидов с целью определения их гомологичности. То есть для примера, показанного на рис. 1, проводилось бы сравнение нуклеотидного состава ампликонов, изображенное на рис. 3.

Зная начальную и конечную позиции расположения прямого и обратного праймеров, можно найти непосредственно данный участок в FASTA-файле. Проведенное сравнение показало значительное сходство по нуклеотидному составу ампликонов при совпадении их размеров (табл. 2, 3, 4).

Проведенный *in silico* RAPD-анализ в виде виртуальной мультиплексной ПЦР

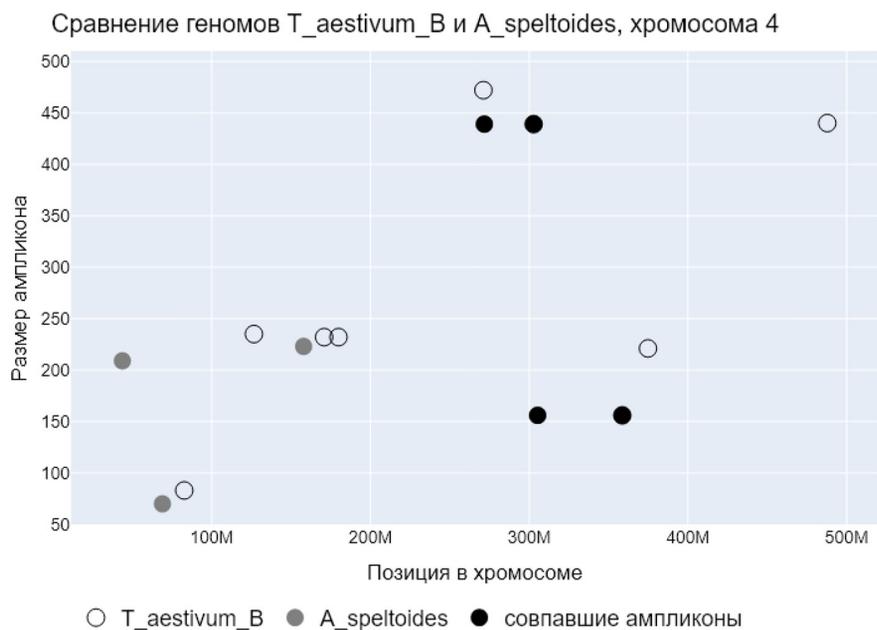


Рис. 1. Пример похромосомного двумерного анализа распределения ампликонов у мягкой пшеницы и видов – потенциальных доноров ее субгеномов (4 – номер гомеологичной хромосомы). Белый (прозрачный) круг – ампликоны *T. aestivum*; серый – ампликоны *Ae. speltoides*; черный – ампликоны *T. aestivum* и *Ae. speltoides*, совпавшие по размеру

Fig. 1. An example of chromosomal two-dimensional analysis of amplicon distribution in bread wheat and species that are potential donors of its subgenomes (4 is the number of homeologous chromosome). White (clear), *T. aestivum* amplicons; grey, *Ae. speltoides* amplicons; black – amplicons of *T. aestivum* and *Ae. speltoides* matching in size

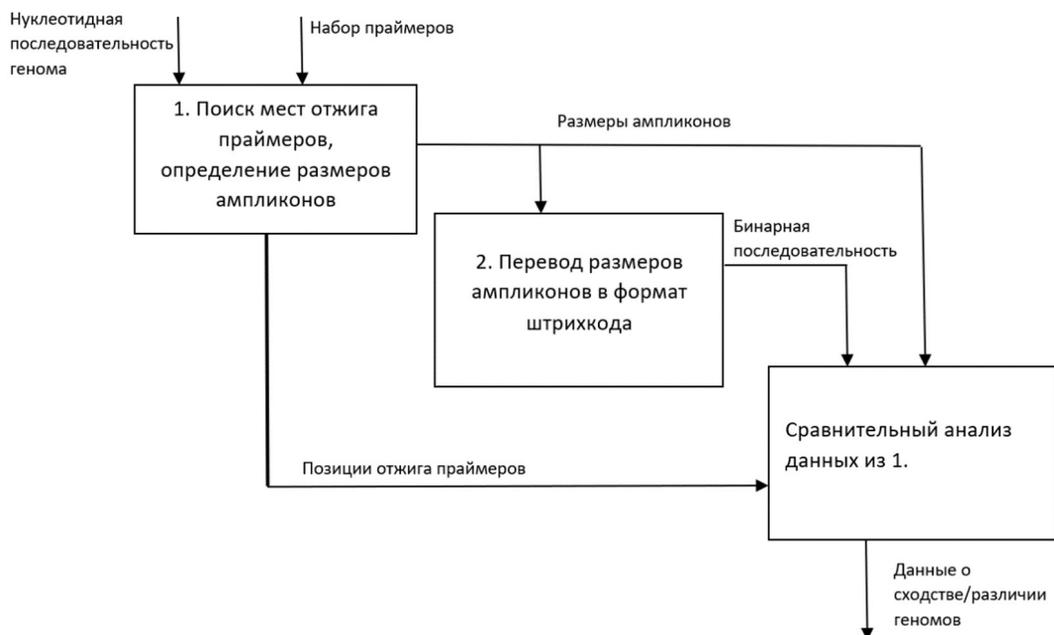


Рис. 2. Принцип предложенного *in silico* анализа геномов и последовательность обработки данных

Fig. 2. The principle of the proposed *in silico* analysis of genomes and steps of data processing

с подобранными нами 20 праймерами позволил обнаружить в выбранном диапазоне размеров от 51 до 500 п.н. около 40–50 ампликонов для субгеномов **B**, **A** и **D** мягкой пшеницы и видов – потенциальных доноров этих субгеномов. Как можно видеть на рис. 4 и 5, ряд ампликонов у этих видов и субгеномов со-

падают или очень близки по размеру, из чего можно предполагать, что это одинаковые или гомологичные участки их геномов. Как видно на рис. 4, геномные штрихкоды *T. aestivum* субгеном **D** и *Ae. tauschii* весьма схожи между собой, что еще раз подтверждает их общее происхождение. Стоит отметить, что при

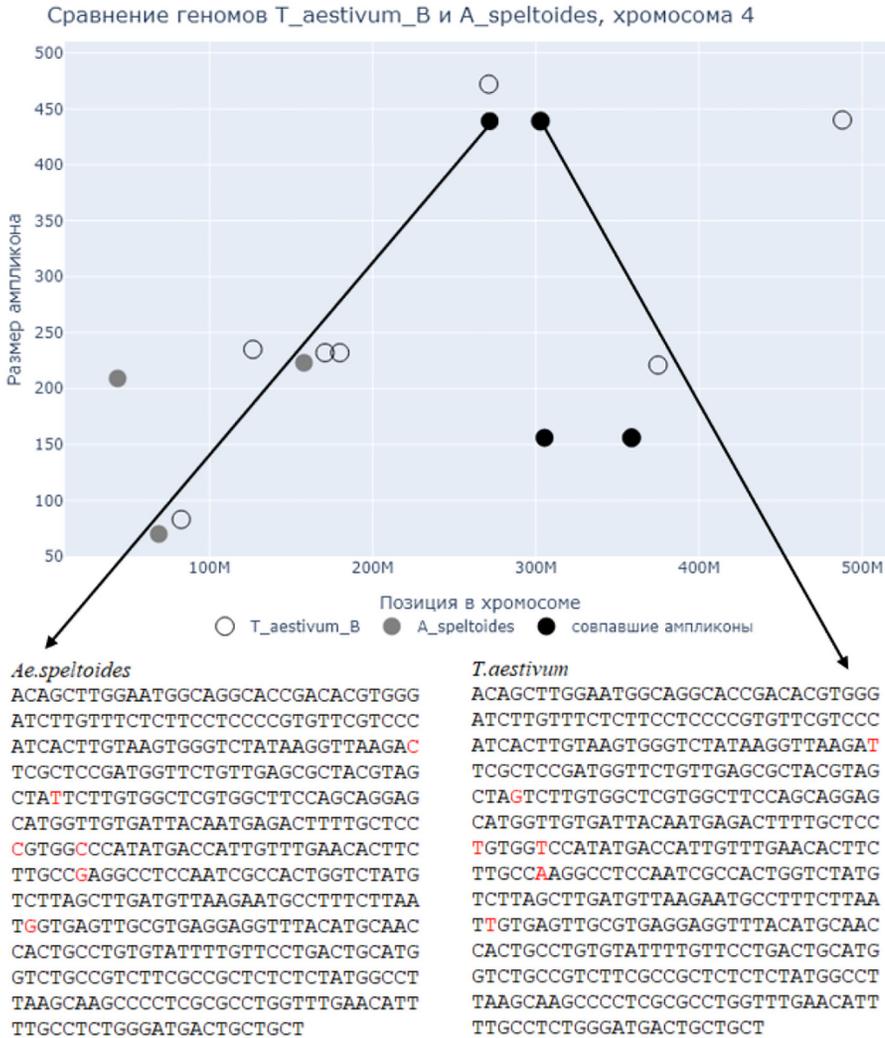


Рис. 3. Принцип выбора ампликонов для сравнения по нуклеотидным последовательностям. Справа показана нуклеотидная последовательность ампликона *T. aestivum*, а слева – *Ae. speltoides*. Из рисунка видно, что выделенные ампликоны содержат гомологичные нуклеотидные последовательности. Красным отмечены отличающиеся нуклеотиды. Данные последовательности сходны друг с другом на 98,6 %, что позволяет считать их родственными

Fig. 3. The principle of selection of amplicons for comparison by nucleotide sequences. The nucleotide amplicon sequence of the *T. aestivum* is shown on the right, and *Ae. speltoides* is shown on the left. The marked amplicons contain homologous nucleotide sequences. The different nucleotides are marked in red. These sequences are 98.6 % homologous to each other, which allows us to consider them related

Таблица 2. Сравнение совпадающих и близких по размеру (± 3 п.н.) ампликонов по нуклеотидному составу для субгенома **B** *T. aestivum* и *Ae. speltooides*Table 2. Comparison of identical and similar size (± 3 bp) amplicons by nucleotide composition of *T. aestivum* subgenome **B** and *Ae. speltooides*

<i>T. aestivum</i> субгеном B			<i>Ae. speltooides</i>			Процент идентичности нуклеотидов (%)
Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	
Хромосома 1						
439696211	439696440	229	414389062	414389291	229	70,5
Хромосома 2						
552164185	552164340	155	543878527	543878683	156	17,9
530456167	530456323	156	543878527	543878683	156	88,5
Хромосома 3						
138390601	138391090	489	86776651	86777140	489	94,1
Хромосома 4						
302854627	302855066	439	271785545	271785984	439	98,7
Хромосома 5						
63192550	63192733	183	43441847	43442030	183	99,5
Хромосома 6						
20465358	20465514	156	365979068	365979224	156	17,8
Хромосома 7						
263552300	263552456	156	212440595	212440751	156	90,4

формировании штрихкодов не учитывается кратность возникновения ампликонов одной длины, нас интересует только наличие того или иного ампликона. При визуальной оценке между *T. aestivum* субгеном **A** и *T. urartu* различий в штрихкодах выявляется уже гораздо больше (рис. 4). В случае сравнения субгенома **B** и генома *Ae. speltooides* совпадений было выявлено еще меньше. По результатам визуального сравнения штрихкодов субгенома **B** с геномами эгилопсов из секции *Sitopsis* видно, что с этим субгеномом одинаково схожи геномы не только *Ae. speltooides*, но и *Ae. bicornis*, *Ae. longissima* (рис. 5).

С учетом возможных делеций и инсерций можно предположить, что и несколько отличающиеся по размеру ампликоны, локализующиеся в сходных участках хромосом, могут принадлежать одинаковым или гомоло-

гичным участкам. Для подтверждения такой возможности нами проведен анализ образующихся ампликонов по размеру с учетом их локализации в конкретных хромосомах (рис. 6). В этом случае рассматривается распределение абсолютно всех ампликонов в хромосомах. Поэтому количество полос на штрихкоде и ампликонов на двумерных картах значительно отличается. При этом разным цветом обозначены ампликоны, свойственные или только субгеномам мягкой пшеницы (белый цвет), либо только видам – потенциальным донорам этих самых субгеномов (серый цвет). Черным цветом отображены ампликоны, совпадающие по размеру у мягкой пшеницы и видов – потенциальных доноров. При таком сопоставлении виртуальных ампликонов их точная локализация на хромосомах не столь принципиальна, как было важно показать об-

Таблица 3. Сравнение совпадающих и близких по размеру (± 3 п.н.) ампликонов по нуклеотидному составу для *T. aestivum* субгеном А и *T. urartu*Table 3. Comparison of identical and similar size (± 3 bp) amplicons by nucleotide composition of *T. aestivum* subgenome А and *T. urartu*

<i>T. aestivum</i> субгеном А			<i>T. urartu</i>			Процент идентичности нуклеотидов (%)
Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	
Хромосома 1						
83371490	83371557	67	95698487	95698554	67	100,0
439841303	439841371	68	440421903	440421971	68	100,0
439841303	439841371	68	440544506	440544574	68	100,0
494393892	494393963	71	490397728	490397799	71	100,0
Хромосома 2						
690910226	690910494	268	668800931	668801199	268	24,6
104671882	104672276	394	102352938	102353332	394	21,3
Хромосома 3						
519837733	519837880	147	521569961	521570108	147	96,6
Хромосома 4						
637114870	637115103	233	592870802	592871035	233	99,1
637114870	637115103	233	593219307	593219540	233	28,9
606921462	606921893	431	616666113	616666547	434	97,2
Хромосома 5						
384193781	384194005	224	371278975	371279199	224	100,0
445688045	445688409	364	422626585	422626950	365	94,8
Хромосома 6						
564509766	564509948	182	523401534	523401716	182	45,5
Хромосома 7						
442562219	442562451	232	435699634	435699866	232	39,6
551937862	551938289	427	547384499	547384925	426	96,7

щий вид распределения найденных ампликонов по всем субгеномам сразу.

В колонке «Субгеном В» проводилось похромосомное сравнение субгенома В мягкой пшеницы с *Ae. speltooides*, в колонке «Субгеном А» с *T. urartu* и в колонке «Субгеном Д» с *Ae. tauschii*. Проведенный анализ, представленный на рис. 6, наглядно показывает, что для субгенома Д и *Ae. tauschii* намного больше совпадающих ампликонов (129 по всем 7 хромосомам) и по размеру, и по их локализации в хромосомах, чем в случае сравнения субгеномов А и В с геномами соответствующих

эгилопсов. Сопоставление этого результата с подобными сравнениями субгеномов В и А с *Ae. speltooides* и *T. urartu*, демонстрирующая всего 37 и 36 совпадений соответственно, фактически подтверждает, что *Ae. tauschii* действительно был донором субгенома Д мягкой пшеницы, тогда как *Ae. speltooides* и *T. urartu* могут претендовать на виды, близкие к тем, что послужили донорами двух других субгеномов *T. aestivum*. Некая тетраплоидная пшеница с субгеномами В и А образовалась значительно раньше, нежели мягкая гексаплоидная пшеница, что все же позволяет

Таблица 4. Сравнение совпадающих и близких по размеру (± 3 п.н.) ампликонов по нуклеотидному составу для *T. aestivum* субгеном **D** и *Ae. tauschii*Table 4. Comparison of identical and similar size (± 3 bp) amplicons by nucleotide composition of *T. aestivum* subgenome **D** and *Ae. tauschii*

<i>T. aestivum</i> субгеном D			<i>Ae. tauschii</i>			Процент идентичности нуклеотидов (%)
Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	
Хромосома 1						
28872846	28873002	156	31342752	31342908	156	99,4
180301560	180301743	183	185649300	185649483	183	100,0
147104843	147105063	220	13610302	13610522	220	22,7
389120439	389120669	231	395539776	395540007	231	98,7
389747420	389747652	232	396094528	396094760	232	98,3
91569625	91569960	335	99752760	99753095	335	23,5
Хромосома 2						
380695918	380696062	144	379585379	379585523	144	100,0
327072075	327072287	212	325829056	325829268	212	100,0
342955210	342955441	231	341638015	341638246	231	100,0
20901974	20902206	232	22026349	22026581	232	100,0
296091638	296091870	232	294766590	294766822	232	100,0
346136110	346136342	232	344883552	344883784	232	99,1
349837392	349837624	232	348569370	348569602	232	99,6
346136110	346136342	232	348569370	348569602	232	93,1
349837392	349837624	232	344883552	344883784	232	94,4
510138042	510138274	232	508718054	508718286	232	100,0
584913162	584913394	232	583261185	583261417	232	98,7
465701901	465702133	232	465009093	465009325	232	100,0
465803101	465803333	232	465111081	465111313	232	100,0
548259100	548259368	268	546854889	546855157	268	99,6
442329630	442330034	404	441590658	441591062	404	99,5
Хромосома 3						
140869549	140869705	156	143215628	143215784	156	100,0
140823781	140823937	156	143170549	143170705	156	100,0
525401120	525401276	156	533932724	533932880	156	100,0
352801079	352801236	157	360184183	360184340	157	100,0
578168470	578168664	194	589015637	589015831	194	99,0
256153049	256153269	220	263700036	263700256	220	100,0
140869474	140869705	231	143215553	143215784	231	100,0
26549091	26549323	232	28749846	28750078	232	99,1
250172412	250172644	232	257598492	257598724	232	100
273916425	273916657	232	281410195	281410427	232	99,6
549679937	549680169	232	558167041	558167273	232	99,1
526997967	526998258	291	535619808	535620099	291	99,7

Продолжение табл. 4

Table 4 Continued

<i>T. aestivum</i> субгеном D			<i>Ae. tauschii</i>			Процент идентичности нуклеотидов (%)
Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	Позиция прямого праймера	Позиция обратного праймера	Размер ампликона	
285403802	285404242	440	292808722	292809162	440	99,5
Хромосома 4						
50009663	50009895	232	510018903	510019135	232	95,7
467557710	467557942	232	474225579	474225811	232	100,0
316999056	316999380	324	324545177	324545501	324	99,7
287554825	287555264	439	295090899	295091338	439	37,7
Хромосома 5						
342823059	342823215	156	350115450	350115606	156	100,0
300404853	300405084	231	307559257	307559488	231	99,6
33889733	33889965	232	36863145	36863377	232	99,6
281995744	281995976	232	289013408	289013640	232	99,6
537196318	537196550	232	549313870	549314102	232	100,0
300383556	300383788	232	307536829	307537061	232	100,0
262784322	262784730	408	269614209	269614617	408	99,5
Хромосома 6						
272111805	272112017	212	295735082	295735294	212	99,5
311661592	311661824	232	334717146	334717378	232	100,0
345211682	345211914	232	330655041	330655273	232	84,9
455146365	455146597	232	478193687	478193919	232	99,6
451684258	451684559	301	474150212	474150513	301	100,0
Хромосома 7						
308814473	308814701	228	314125286	314125514	228	100,0
595288430	595288661	231	601089913	601090144	231	100,0
51903609	51903841	232	50348563	50348795	232	99,6
323860155	323860387	232	328641564	328641796	232	99,5
411275737	411275969	232	410899236	410899468	232	99,6
467280096	467280328	232	467546667	467546899	232	100,0
619703348	619703580	232	624135569	624135801	232	100,0
293419829	293420298	469	295635832	295636301	469	99,1

думать о произошедших за долгие годы эволюции гораздо больших изменениях данных субгеномов в составе сначала тетраплоидного, а затем и гексаплоидного видов пшениц. Также представляет определенный интерес совпадение числа ампликонов в формате 2d для субгенома **B** мягкой пшеницы и эгилоп-

сов секции *Sitopsis* (табл. 5), из которого видно, что представители подсекции *Emarginata* гораздо сильнее удалены от некоего (может неизвестного) донора субгенома **B**, поскольку в этом случае выявлялось от 20 до 24 общих ампликонов. По результатам похромосомного двумерного анализа распределения амплико-

<i>Triticum aestivum</i> субгеном A	059 067 068 071 093 121 147 151 173 182 183 224 229 231 232 233 234 247 268 270 293 323 357 363 364 382 383 394 407 416 426 427 431 443 494	
<i>T. urartu</i>	067 068 071 093 121 147 156 182 183 219 220 223 224 230 232 233 252 255 268 272 290 301 357 358 362 363 364 365 381 383 394 411 416 426 427 428 429 434 443	
<i>T. aestivum</i> субгеном B	058 067 068 083 090 093 116 122 123 131 144 150 155 156 157 159 178 182 183 193 196 210 221 229 231 232 235 239 278 310 325 333 374 386 387 390 398 402 423 424 433 439 440 448 472 477 489 498	
<i>Aegilops speltoides</i>	058 066 070 093 144 155 156 157 158 173 183 196 197 209 223 229 232 305 320 368 385 390 439 446 460 468 489	
<i>T. aestivum</i> субгеном D	104 144 156 157 183 194 212 220 223 228 229 230 231 232 238 268 285 291 301 305 309 324 335 339 352 366 390 404 408 428 439 440 448 469	
<i>Ae. tauschii</i>	104 111 144 156 157 182 183 194 212 220 228 229 231 232 233 238 245 268 280 285 291 300 301 307 309 324 335 339 390 404 408 439 440 442 443 444 469	

Рис. 4. Геномные штрихкоды ряда видов пшенично-эгилопсного комплекса, полученные в результате моделирования мультиплексовой ПЦР с 20 праймерами. Во втором столбце приведены размеры потенциальных ампликонов в п.н.

Fig. 4. Genomic barcodes of several species of the wheat–*Aegilops* alliance obtained by modeling multiplex PCR with 20 primers. The second column shows the sizes of potential amplicons in bp.

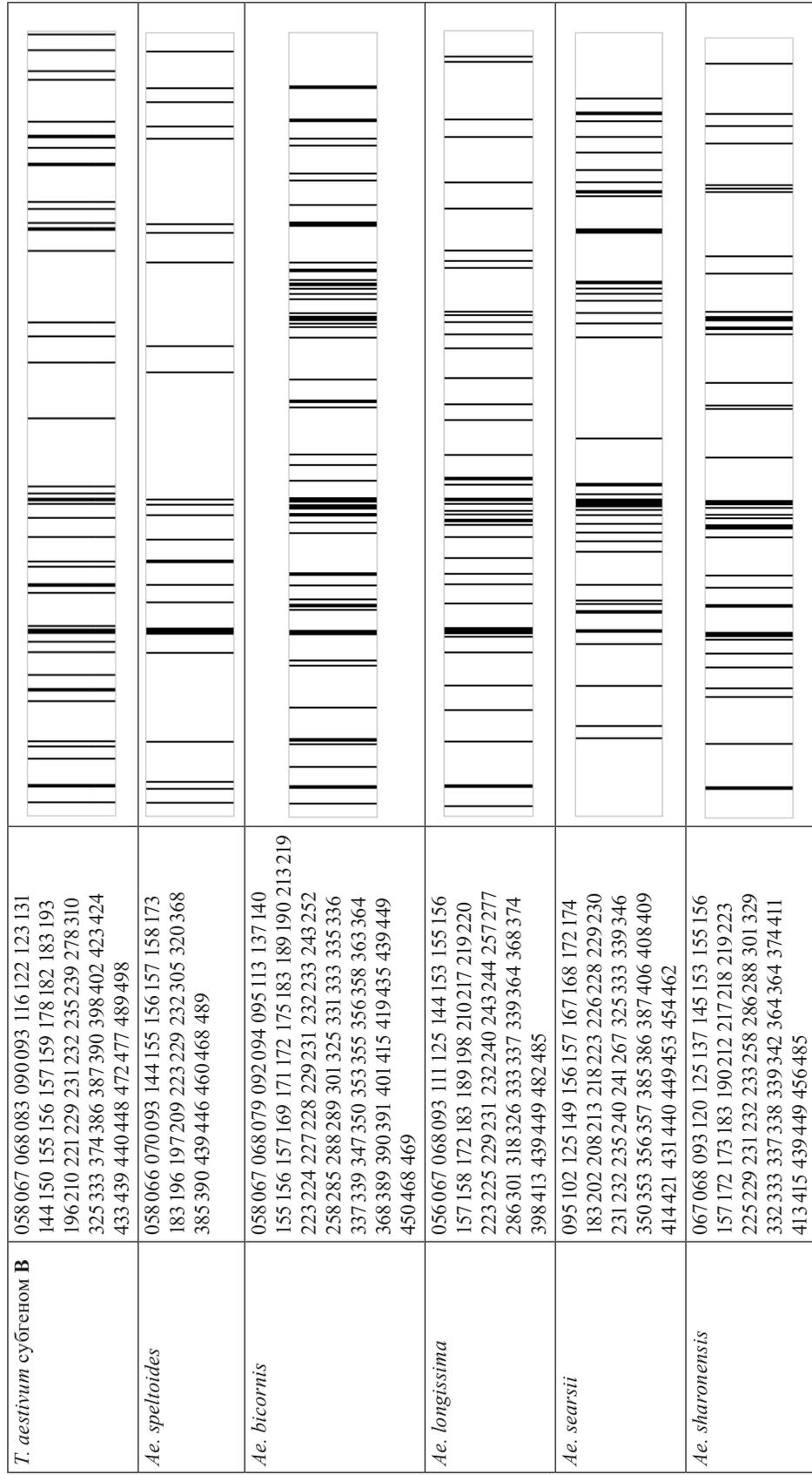


Рис. 5. Геномные штрихкоды субгенома **B** *T. aestivum* и видов рода *Aegilops* секции Sitopsis, полученные в результате моделирования мультиплексной ПЦР с 20 праймерами. Во втором столбце приведены размеры потенциальных ампликонов в п.н.

Fig. 5. Genomic barcodes of *T. aestivum* subgenome **B** and *Aegilops* species of the Sitopsis section obtained by modeling multiplex PCR with 20 primers. The second column shows the sizes of potential amplicons in bp.

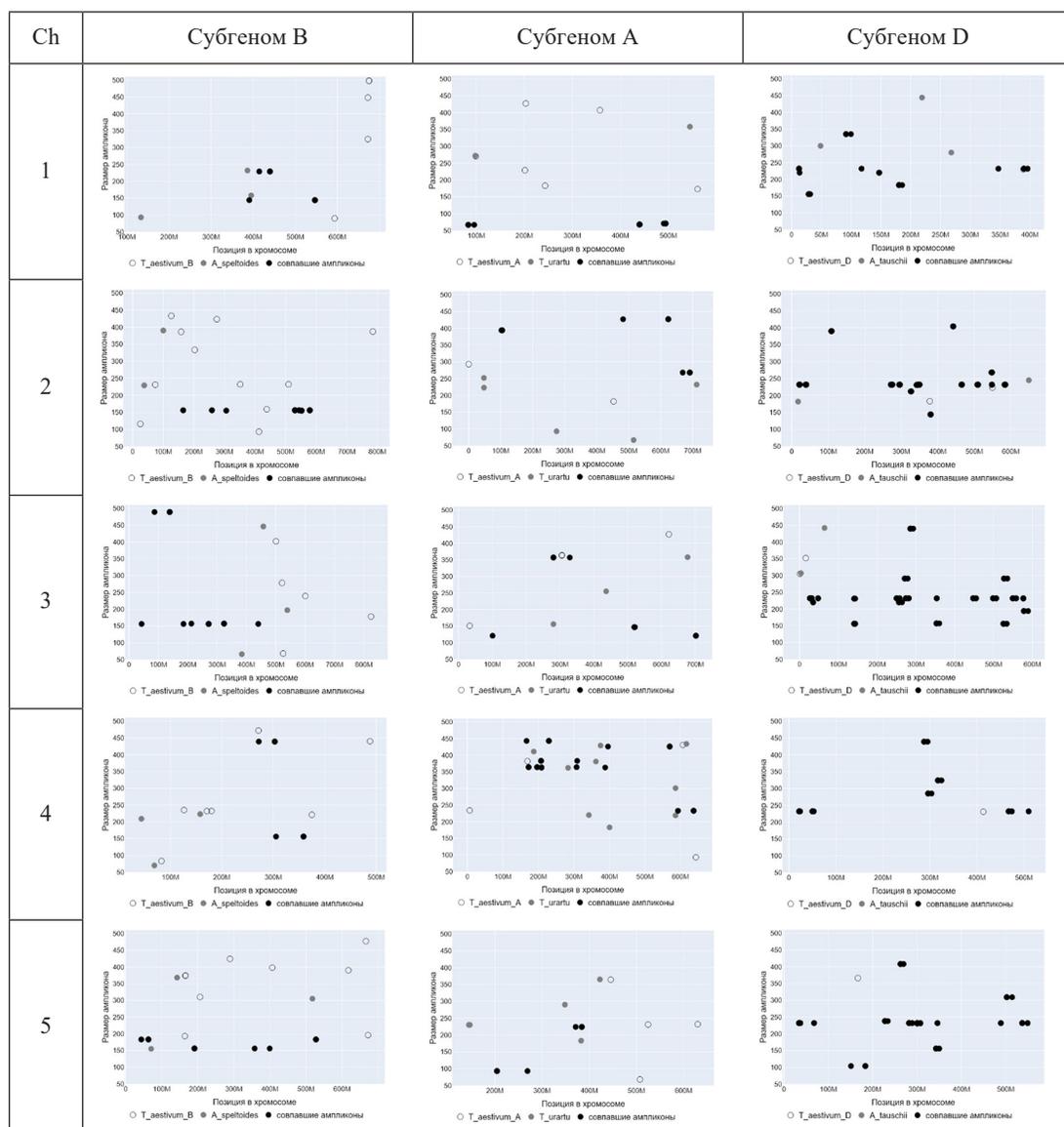
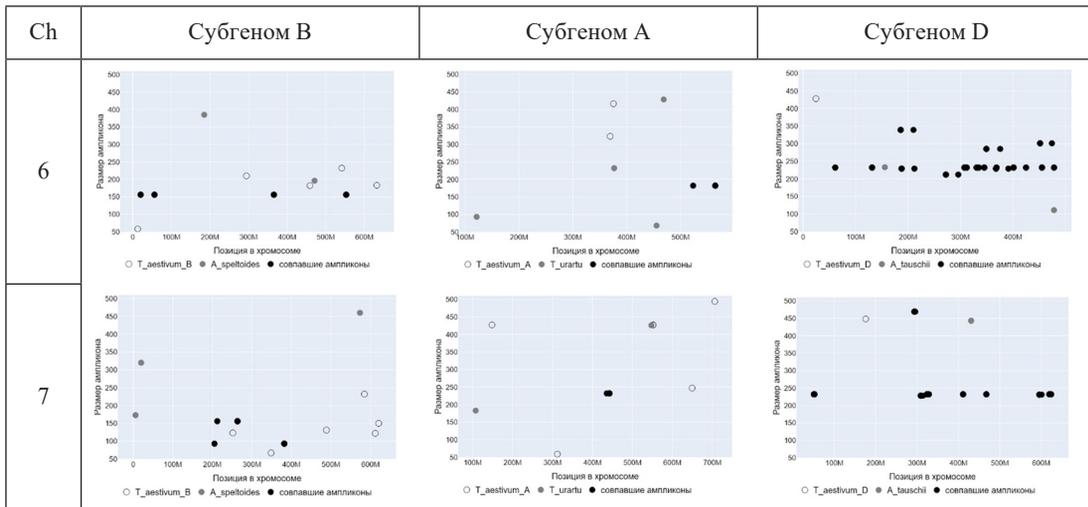


Рис. 6. Похромосомный двумерный анализ распределения ампликонов у мягкой пшеницы и видов *Ae. speletoides*, *T. urartu*, *Ae. tauschii* – потенциальных доноров ее субгеномов (1–7 – номера гомеологичных хромосом). Во втором столбце приведены результаты сравнения субгенома **В** мягкой пшеницы с геномом *Ae. speletoides*. В третьем столбце приведены результаты сравнения субгенома **А** мягкой пшеницы с геномом *T. urartu*. В четвертом столбце приведены результаты сравнения субгенома **Д** мягкой пшеницы с геномом *Ae. tauschii*

Fig. 6. Chromosomal two-dimensional analysis of amplicon distribution in bread wheat and species *Ae. speletoides*, *T. urartu*, *Ae. tauschii* – potential donors of its subgenomes (1–7 are the numbers of homeologous chromosomes). The second column shows the results of comparison of the **B** subgenome of bread wheat with the *Ae. speletoides* genome. The third column shows the results of comparison of the **A** subgenome of bread wheat with the *T. urartu* genome. The fourth column shows the results of comparison of the **D** subgenome of bread wheat with the *Ae. tauschii* genome



Продолжение рис. 6

Fig. 6 Continued

нов у субгенома **В** мягкой пшеницы и видов эгилопсов секции *Sitopsis* (табл. 5) можно видеть, что к данному субгеному более близок геном *Ae. speltoides*. На примерно таком же уровне схожести находятся субгеном **А** и геном *T. urartu*.

Нельзя было исключать того, что совпадающие по размеру ампликоны содержат не гомологичные нуклеотидные последовательности. С другой стороны, близкие по размеру ампликоны могли содержать родственные последовательности. Исходя из этого, нами было проведено выравнивание и сравнение совпадающих и наиболее близких по размеру (± 3 п.н.) нуклеотидных последовательностей, расположенных относительно близко друг от друга в рамках хромосомы. Из сравнения были исключены ампликоны, совпадающие по размеру, но локализованные в разных участках хромосомы, так как велика вероятность, что это разные нуклеотидные фрагменты. Действительно, к примеру, в случае с *T. aestivum* субгеном **В** и *Ae. speltoides* два совпадающих по размеру ампликона содержали не гомологичные нуклеотидные последовательности (табл. 2).

В случае с *T. aestivum* субгеном **А** и *T. urartu* 5 ампликонов оказались негомологичными (табл. 3). В то же время при сравнении субгенома **Д** с геномом *Ae. tauschii* выявлялось большое количество схожих ампликонов, для части из них было характерно 100 %-ное совпадение (табл. 4).

Из табл. 2–4 видно, что в большинстве случаев ампликоны одинакового размера, локализованные в одной области хромосомы, характеризуются очень схожим нуклеотидным составом, за исключением определенного количества нуклеотидных замен, что может говорить о родственности геномов, по крайней мере, при идентичности больше 50 %.

Таким образом, разработанный нами компьютерный метод анализа родственных геномов позволяет достаточно быстро провести сравнение геномов и даже спланировать натуральный эксперимент, подобрав оптимальный набор праймеров, а также дать общее представление того, какой результат может быть получен в ходе проведения ПЦР в «мокрых» экспериментах. Однако необходимо от-

Таблица 5. Число совпадающих по размеру ампликонов при сравнении субгеномов *T. aestivum* с представителями пшенично-эгилопсного комплексаTable 5. The number of amplicons with the same length as the result of comparison of the subgenomes of *T. aestivum* with the wheat-aegilops complex

Сравнение с субгеномом В								
Виды	Хромосомы							
	1	2	3	4	5	6	7	Σ
<i>Ae. speltoides</i>	4	7	8	4	6	4	4	37
<i>Ae. bicornis</i>	2	5	4	5	3	2	3	24
<i>Ae. longissima</i>	4	5	3	5	2	0	2	22
<i>Ae. searsii</i>	0	9	4	3	0	3	2	21
<i>Ae. sharonensis</i>	2	5	2	6	2	0	3	20
Сравнение с субгеномом А								
<i>T. urartu</i>	5	5	5	13	4	2	2	36
Сравнение с субгеномом D								
<i>Ae. tauschii</i>	13	16	32	11	21	24	12	129

метить, что для натуральных экспериментов необходимо более тщательно подойти к подбору набора праймеров. К тому же едва ли можно рассчитывать на то, что удастся наработать все теоретически предсказанные ампликоны, но при использовании автоматических капиллярных секвенаторов, по крайней мере, будут выявляться большинство амплифицированных фрагментов ДНК. Метод построения двумерных карт для исследуемых геномов проведен впервые и позволяет наглядно выявить потенциально схожие фрагменты ДНК, не прибегая к методу выравнивания нуклеотидных последовательностей геномов. Кроме того, полагаясь на созданные карты, можно провести более глубокий нуклеотидный анализ ампликонов, незначительно различающихся по размеру, но расположенных близко друг от друга на предмет выявления частичных изменений нуклеотидных последовательностей.

Проведённый нами геномный *in silico* RAPD-анализ очередной раз подтвердил, что донором субгенома **D** мягкой пшеницы действительно является *Ae. tauschii*. Наши

результаты согласуются с данными о том, что мягкая пшеница *T. aestivum* возникла в результате гибридизации между культурной полбой *T. turgidum* ($2n = 4x = 28$, **AABB**) и *Ae. tauschii* ($2n = 2x = 14$, **DD**) (Конарев и др., 1974; Dudnikov, 2017). В случае с донорами субгеномов **A** и **B** такие однозначные выводы с помощью предложенного нами подхода сделать не удалось, что, возможно, связано с более древним объединением этих субгеномов в тетраплоидной пшенице и накоплением за это время большего количества мутаций.

Заключение

Знание истинных доноров пшеничных субгеномов представляет далеко не праздный интерес, поскольку в пшенично-эгилопсном комплексе имеется по меньшей мере 12 диплоидных геномов, а Природа использовала при создании двух рядов полиплоидных пшениц (*aestivum-turgidum* и *timopheevii-zhukovskiyi*) не более шести, при том что человечество сейчас «кормят» всего три субгенома. К тому же Природа при гибридизации могла выбирать лишь из видов, произрас-

тающих совместно и поэтому для селекционеров остается еще большой простор для проведения различных скрещиваний с целью создания новых синтетических видов пшениц, которые потенциально могут иметь увеличенную урожайность, лучшую устойчивость к вредителям и прочие важные качества. Но для ведения подобных работ нужно точно знать, что уже было использовано Природой. Предлагаемый подход к анализу полных ядерных геномов при помощи виртуального RAPD-анализа *in silico* был испытан

на пшенично-эгилопном комплексе, и показано, что он может служить одним из методов поиска истинных доноров пшеничных субгеномов **В** и **А** для ряда *turgidum-aestivum* и субгеномов **G** и **A** для ряда *timopheevii-zhukovskyi* (когда их полные геномы будут секвенированы), что позволит вести селекцию в дальнейшем более осознанно. Разработанная технология может быть рекомендована к использованию и на других видах живых организмов, имеющих полиплоидное происхождение.

Список литературы / References

- Гарафутдинов Р.Р., Баймиев Ан. Х., Малеев Г.В., Алексеев Я.И., Zubov 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. (2019) Разнообразие праймеров для ПЦР и принципы их подбора. *Биомика*, 11(1): 23–70 [Garafutdinov R.R., Baymiev An. Kh., Maleev G.V., Alexeev Ya. I., Zubov V.V., Chemeris D.A., Kiryanova O. Yu., Gubaydullin I.M., Matniyazov R. T., Sakhabutdinova A.R., Nikonov Yu. M., Kuluev B.R., Baymiev Al. Kh., Chemeris A.V. (2019) Diversity of PCR primers and principles of their design. *Biomics* [Biomika], 11(1): 23–70 (in Russian)]
- Кирьянова О.Ю., Кирьянов И.И., Гарафутдинов Р.Р., Чемерис А.В., Губайдуллин И.М. (2021) GATCGGENERATOR. Свидетельство о регистрации программы для ЭВМ 2021667097, 25.10.2021. Заявка № 2021666409 от 20.10.2021 [Kiryanova O. Yu., Kiryanov I. I., Garafutdinov R. R., Chemeris A. V., Gubaydullin I. M. (2021) GATCGGENERATOR. Certificate of registration of a computer program 2021667097, 25.10.2021. Application No. 2021666409 dated 20.10.2021 (in Russian)]
- Кирьянова О.Ю., Кирьянов И.И., Кулуев Б.Р., Чемерис А.В., Гарафутдинов Р.Р., Губайдуллин И.М. (2020а) Свидетельство о государственной регистрации программы для ЭВМ № 2020610703 ABCDNA_GS (Amplified Bar-Coded DNA Genome/Specimen) от 17.01.2020 г. [Kiryanova O. Yu., Kiryanov I. I., Kuluyev B. R., Chemeris A. V., Garafutdinov R. R., Gubaydullin I. M. (2020a) Certificate of state registration of computer program no. 2020610703 ABCDNA_GS (Amplified Bar-Coded DNA Genome/Specimen) dated 17.01.2020 (in Russian)]
- Кирьянова О.Ю., Кирьянов И.И., Кулуев Б.Р., Чемерис А.В., Губайдуллин И.М. (2022) BARCODE DNA ANALYZER (BDANALYZER). Свидетельство о регистрации программы для ЭВМ № 2022663254, 13.07.2022 [Kiryanova O. Yu., Kiryanov I. I., Kuluyev B. R., Chemeris A. V., Gubaydullin I. M. (2022) Certificate of state registration of computer program no. 2022663254 BARCODE DNA ANALYZER (BDANALYZER) dated 13.07.2022 (in Russian)]
- Кирьянова О.Ю., Кулуев Б.Р., Кулуев А.Р., Марданшин И.С., Губайдуллин И.М., Чемерис А.В. (2020) Мультиплексный *in silico* RAPD-анализ ряда родственных растений с отличающимися размерами геномов и перспективы такого подхода для ДНК-паспортизации сортов сельскохозяйственных растений. *Биомика*, 12(2): 194–210 [Kiryanova O. Yu., Kuluev B. R., Kuluev A. R., Mardanshin I. S., Gubaydullin I. M., Chemeris A. V. (2020) 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* [Biomika], 12(2): 194–210 (in Russian)]

Конарев В. Г., Хакимова А. Г., Гаврилюк И. П., Мигушова Э. Ф. (1974) Дифференциация генома D по данным электрофоретического и иммунохимического анализа глиаина *Aegilops squarrosa* L. (*Ae. tauschii* Coss.). *Сельскохозяйственная биология*, 9(3): 352–358 [Konarev V. G., Hakimova A. G., Gavriilyuk I. P., Migushova E. F. (1974) Differentiation of the D genome according to electrophoretic and immunochemical analysis of *Aegilops squarrosa* L. (*Ae. tauschii* Coss.) gliadin. *Agricultural Biology* [Sel'skokhozyaistvennaya biologiya], 9(3): 352–358 (in Russian)]

Кулуев А. Р., Матниязов Р. Т., Чемерис Д. А., Чемерис А. В. (2016) Современные представления о родственных взаимоотношениях в пшенично-эгилопсном альянсе (с краткой исторической справкой). *Биомика*, 8(4): 297–310 [Kuluev A. R., Matniyazov R. T., Chemeris D. A., Chemeris A. V. (2016) Modern concepts about relationships in the wheat-aegilops alliance (with a brief historical note). *Biomics* [Biomika], 8(4): 297–310 (in Russian)]

Кулуев А. Р., Матниязов Р. Т., Чемерис Д. А., Юнусбаев У. Б., Кулуев Б. Р., Чемерис А. В. (2020) Филогенетические взаимоотношения в пшенично-эгилопсном альянсе через призму хлоропластного генома. *Биомика*, 12(4): 532–544 [Kuluev A. R., Matniyazov R. T., Chemeris D. A., Yunusbaev U. B., Kuluev B. R., Chemeris A. V. (2020) Phylogenetic relationships in the wheat-aegilops alliance through the prism of the chloroplast genome. *Biomics* [Biomika], 12(4): 532–544 (in Russian)]

Мигушова Э. Ф. (1975) К вопросу о происхождении геномов пшеницы. *Труды по прикладной ботанике, генетике и селекции*, 55(3): 3–26 [Migushova E. F. (1975) On the origin of wheat genomes. *Proceedings on Applied Botany, Genetics and Breeding* [Trudy po prikladnoi botanike, genetike i selektsii], 55(3): 3–26 (in Russian)]

Dudnikov A. J. (2017) Polymorphism of *Got2* DNA sequences sheds light on *Aegilops tauschii* Coss. intraspecies divergence and origin of *Triticum aestivum* L. *Genetic Resources and Crop Evolution*, 64(7): 1623–1640

Goncharov N. P. (2011) Genus *Triticum* L. taxonomy: the present and the future. *Plant Systematics and Evolution*, 295: 1–11

Haider N. (2013) The origin of the B-genome of bread wheat (*Triticum aestivum* L.). *Russian Journal of Genetics*, 49(3): 263–274

Kuluev B. R., Baymiev A. K., Gerashchenkov G. A., Chemeris D. A., Zubov V. V., Kuluev A. R., Baymiev A. K., Chemeris A. V. (2018) Random priming PCR strategies for identification of multilocus DNA polymorphism in eukaryotes. *Russian Journal of Genetics*, 54(5): 499–513

Levy A. A., Feldman M. (2022) Evolution and origin of bread wheat. *Plant Cell*, 34(7): 2549–2567

Vakhitov V. A., Chemeris A. V., Sabirzhanov B. E., Akhunov E. D., Kulikov A. M., Nikonorov Yu. M., Gimalov F. R., Bikbulatova S. M., Baymiev Al. Kh. (2003) The phylogeny of *Triticum* L. and *Aegilops* L. inferred from comparative analysis of nucleotide sequences in rDNA promoter regions. *Russian Journal of Genetics*, 39(1): 1–11