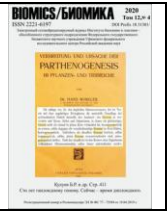




# BIOMICS/БИОМИКА



ISSN 2221-6197 <http://biomicsj.ru>

## СТО ЛЕТ ГАПЛОИДНЫМ ГЕНОМАМ. СЕЙЧАС НАСТУПАЕТ ВРЕМЯ ДИПЛОИДНЫХ

<sup>1</sup>Кулуев Б.Р., <sup>1</sup>Баймиев Ан.Х., <sup>1</sup>Герашенков Г.А., <sup>1</sup>Юнусбаев У.Б.,  
<sup>1</sup>Гарафутдинов Р.Р., <sup>2,3</sup>Алексеев Я.И., <sup>1</sup>Баймиев Ал.Х., <sup>1</sup>Чемерис А.В.

<sup>1</sup>Институт биохимии и генетики – обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, Россия, Уфа, пр. Октября 71, 450054, E-mail: [chemeris@amrb.ru](mailto:chemeris@amrb.ru)

<sup>2</sup>ООО «Синтол», Россия, 127550, Москва, ул. Тимирязевская, д. 42

<sup>3</sup>Институт аналитического приборостроения Российской академии наук,  
Россия, 198095, Санкт-Петербург, ул. Ивана Черных, д. 31/33

### Резюме

В 2020 году исполнилось 100 лет термину «геном», предложенному немецким ботаником Г.Винклером при описании партеногенеза в растительном и животном царствах. Этимология данного слова неясна, но нам представляется, что Винклер использовал игру слов, соединив морфемы «ген», «сома» и получив “genosom” (нем.), после чего решил удалить один слог, что и привело к краткому и емкому слову «геном», объединяющему гены и составляющем для них нечто целое, что также совпало с окончаниями «-ом» отдельных биологических терминов, обозначающих некое множество (в данном случае для генома – генов). Первоначально под геномом подразумевался гаплоидный набор хромосом, и таким образом хромосомы служили элементарной единицей генома. Сейчас же с появлением возможности секвенировать полный геном его элементарной единицей стала пара нуклеотидов, но и хромосомы оказались задействованными, поскольку нуклеотидные последовательности распределяются в базах данных по хромосомам, когда такое по результатам полногеномного секвенирования становится возможным. На протяжении столетия менялись взгляды ученых, что считать геномом, подразумевая под этим его гаплоидную или диплоидную природу. Сейчас принято считать, что геном - это нуклеотидные последовательности всей совокупности генов и прочих участков ДНК гаплоидного набора хромосом конкретного вида организмов, однако и габитус и функциональное состояние каждого организма определяется его полным диплоидным геномом, тогда как при сборке гаплоидного генома *a priori* происходит игнорирование тех или иных азотистых оснований в отличающихся аллелях, поскольку выбирается лишь один какой-то нуклеотид, что сразу снижает ценность такого генома, которая все равно очень высока, поскольку секвенирование полных гаплоидных геномов уже очень большого числа видов организмов различного уровня генетической сложности дало весьма важную информацию о Живом. Тем не менее, для персонализированной медицины будущего нужны знания об исключительно диплоидных геномах людей, получение которых пока представляет серьезную проблему, в отличие от нынешних квази-гаплоидных геномов, которые секвенируются почти массово. Диплоидные геномы растений также представляют интерес, в том числе для CRISPR/Cas геномного редактирования, когда необходимо произвести изменения в обоих аллелях парных хромосом, и для этого знать их возможно отличающиеся нуклеотидные последовательности крайне необходимо. Равно как и по завершении такого редактирования следует с помощью полногеномного диплоидного секвенирования выявить произведенные целевые и нецелевые мутации. Разрабатываемые подходы полногеномного секвенирования, в том числе протяженных участков ДНК вкупе с компьютерными программами, рассчитанными на фазировку данных, недавно вызвали к жизни новый термин «гаплотиг», представляющий собой гаплотип-специфичный контиг. Вне всякого сомнения, все это позволит со временем перейти к уверенному установлению полных диплоидных геномов, что придаст термину «геном» новый смысл, считая его для эукариотических организмов диплоидным, что будет к тому же отражать и саму сущность организации генома, имеющего в норме дигермическую природу.

**Ключевые слова:** геном, Г.Винклер, ген, гаплоидный набор хромосом, гаплоидный геном, квази-гаплоидный геном, диплоидный геном, плазмон, ДНК, полногеномное секвенирование, диплоидное секвенирование

**Цитирование:** Кулуев Б.Р., Баймиев Ан.Х., Геращенко Г.А., Юнусбаев У.Б., Гарафутдинов Р.Р., Алексеев Я.И., Баймиев Ал.Х., Чемерис А.В. Сто лет гаплоидным геномам. Сейчас наступает время диплоидных // Биомика. 2020. Т.12(4). С. 411-434. DOI: 10.31301/2221-6197.bmcs.2020-33

© Авторы

## ONE HUNDRED YEARS OF HAPLOID GENOMES. NOW TIME COMES FOR DIPLOID GENOMES

<sup>1</sup>Kuluev B.R., <sup>1</sup>Baymiev An.Kh., <sup>1</sup>Gerashchenkov G.A., <sup>1</sup>Yunusbaev U.B.,  
<sup>1</sup>Garafutdinov R.R., <sup>2,3</sup>Alekseev Ya.I., <sup>1</sup>Baymiev Al.Kh., <sup>1</sup>Chemeris A.V.

<sup>1</sup>Institute of Biochemistry and Genetics - Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, 71 Prospekt Oktyabrya, 450054, Ufa, Russia, E-mail: [chemeris@anrb.ru](mailto:chemeris@anrb.ru)

<sup>2</sup>Syntol Ltd, 42 Timiryazevskaya str., 127550, Moscow, Russia

<sup>3</sup>Institute for Analytical Instrumentation RAS, 31/33 Ivana Chernyh str., 198095, St. Petersburg, Russia

### Resume

The year 2020 marks the 100th anniversary of the term "genome", proposed by the German botanist H.Winkler when describing parthenogenesis in the plant and animal kingdoms. The etymology of this word is unclear, but it seems to us that Winkler used a play on words, combining the morphemes "gene", "soma" and getting "genosom"(German). Then he decided to delete one syllable, which led to a short and capacious word "genome", combining genes and making them something whole, which also coincided with the endings "-ome" of individual biological terms denoting a certain set (in this case, for the genome – genes). Initially, the genome was meant as a haploid set of chromosomes, and thus the chromosomes served as an elementary unit of the genome. Now, with the advent of the ability to sequence a entire genome, its elementary unit has become a pair of nucleotides, but the chromosomes are also involved, since the nucleotide sequences are distributed in databases by chromosome, when this is possible according to the results of whole genome sequencing. Over the course of a century, scientists have changing their views on what is considered a genome, meaning its haploid or diploid nature. Now it is considered that the genome is the nucleotide sequences of the entire set of genes and other DNA fragments of a haploid set of chromosomes of a particular species of organisms, but the habitus and functional state of each organism is determined by its complete diploid genome, whereas when assembling the haploid genome *a priori*, certain nitrogenous bases in different alleles are ignored, since only one nucleotide is selected, which immediately reduces value of such genome, which is still very high, since sequencing the complete haploid genomes of a very large number of species of organisms of various levels of genetic complexity has given very important information about Life. However, personalized medicine of the future requires knowledge exclusively diploid genomes of people, which is still a serious problem to obtain, in contrast to the current quasi-haploid genomes, which are sequenced almost en masse. Diploid genomes of plants are also of interest, including CRISPR/Cas genome editing, when it is necessary to make changes in both alleles of paired chromosomes and to do this, it is necessary to know their possibly different nucleotide sequences. As well as at the end of such editing, whole genome diploid sequencing should be used to identify the target and non-target mutations produced. The developed approaches to whole genome sequencing, including long fragments of DNA, together with computer programs designed for data phasing have recently given rise to the new term "haplotype", which is a haplotype-specific contig. All of this will undoubtedly allow to proceed to the establishment of complete diploid genomes, which will give the term "genome" a new meaning, considering it diploid for eukaryotic organisms, which will also reflect the very essence of the organization of the genome, which has a two-parent nature.

**Keywords:** genome, H.Winkler, gene, haploid set of chromosomes, haploid genome, quasi-haploid genome, diploid genome, plasmon, DNA, whole genome sequencing, diploid sequencing

**Citation:** Kuluev B.R., Baymiev An.Kh., Gerashchenkov G.A., Yunusbaev U.B., Garafutdinov R.R., Alekseev Ya.I., Baymiev Al.Kh., Chemeris A.V. One hundred years of haploid genomes. Now time comes for diploid genomes. *Biomics*. 2020. Vol. 12(4). P. 411-434. DOI: 10.31301/2221-6197.bmcs.2020-33 (In Russian)

© The Authors

### Введение

В 2020 г. исполнилось 100 лет термину «геном», но начнем с другого. В последние годы «на слуху» у населения целый ряд биологических терминов. Безусловным «лидером» 2020 г. стал «коронавирус», и это вполне объяснимо, поскольку этот инфекционный агент заставил весь мир себе подчиниться. Ранее мы уже отмечали небывалый рост числа научных публикаций, связанных с коронавирусом SARS-CoV-2 в первой половине 2020 г. [Гарафутдинов и др. (Garafutdinov et al.), 2020; 2020a]. В данном номере журнала также есть статья, посвященная проблемам молекулярной детекции SARS-CoV-2 [Мавзютов и др. (Mavzyutov et al.), 2020], что вполне оправданно, поскольку пандемия на убыль пока не идет. Но если брать не только 2020 год, а больший отрезок времени, то, наверное, к «популярным» биологическим терминам у неспециалистов можно отнести ПЦР, ГМО, ДНК и геном<sup>1</sup>. При этом все они находятся в сфере наших профессиональных интересов и тесно взаимосвязаны. Так, геном - это молекула(ы) ДНК<sup>2</sup>. ГМО представляют собой организмы, в геном которых добавлена некая чужеродная ДНК. А ПЦР - это реакция избирательной амплификации определенных участков ДНК, без которой сейчас невозможны ни создание и последующая детекция ГМО, ни прочие исследования ДНК, ни секвенирование полных геномов.

Основное внимание в данной статье будет уделено геномам высших организмов, включая человека, и поэтому следует коснуться единиц их размерности. Некоторое время для характеристики геномов использовали Дальтоны, пикограммы в виде так называемой величины C-value, даже некие условные единицы. Однако с появлением возможности определения нуклеотидных последовательностей путем их секвенирования в качестве основных количественных параметров геномов стали преимущественно оперировать миллионами и миллиардами пар нуклеотидов (п.н.). При этом усредненный вес одной п.н. составляет  $1,023 \times 10^{-9}$  пг ( $10^{-21}$  г); 1 млрд. п.н. соответственно будет «весить» 1,023 пг, а в 1 пг «вместится» 978 млн. п.н.

Пожалуй, здесь еще нужно заметить следующее. В 1991 г. N.E.Morton опубликовал

статью, посвященную характеристикам генома человека (размерам хромосом, определенных радиоавтографией и проточной цитофлуориметрией), начатую им со слов, что каждая наука имеет параметры точного определения чего-то важного для нее, и привел примеры: для астрономии - скорость света, а для химии - число Авогадро [Morton, 1991]. Он также обратил внимание, что для молекулярной биологии своих таких параметров до сих пор вроде как нет, поскольку она пока пользуется, например, тем же числом Авогадро. Нам представляется, что для молекулярной биологии, по крайней мере, для части, связанной с нуклеиновыми кислотами и геномами, важным параметром стало число нуклеотидов, определенное с точностью до одного азотистого основания, тем более, что это прямой путь к удобной оцифровке ДНК-данных, что нами недавно показано на примере огромного числа возможных комбинаций, обеспечиваемого некими участками ДНК, что позволяет использовать новый термин «ДНК-цифровизация» [Кирьянова и др. (Kiryanova et al.), 2020]. Причем, после того как получило развитие секвенирование ДНК, молекулярная биология, благодаря данному методическому (технологическому) прорыву, можно сказать, вошла этой своей частью в группу так называемых точных наук, тогда как экспериментальная биология<sup>3</sup> обычно лишь обнаруживает взаимосвязи тех или иных событий, воздействий и эффектов, где достоверность результатов из-за разброса получаемых данных иногда оставляет желать лучшего, а научные спекуляции являются нормой. Результатом же секвенирования с точностью не менее 99,99% и даже 99,999% стало возникновение многочисленных высокоточных генетических баз данных, и теперь работа с ними заключается в первую очередь в оперировании нуклеотидами, точнее их количеством и качеством<sup>4</sup>. Таким образом, нуклеотид (нт) или их пара (п.н.) в виде звеньев полимерной цепочки нуклеиновых кислот могут рассматриваться как важный (измерительный) параметр для молекулярной биологии и для анализа геномов, в частности.

Во избежание недоразумений, прежде чем перейти к изложению основного материала, возможно следует пояснить принципиальную разницу между гаплоидным, квази-гаплоидным и диплоидным

<sup>1</sup> Использованный порядок перечисления данных терминов не учитывает их «популярность» у населения.

<sup>2</sup> Значительная часть вирусов, в том числе упоминавшийся выше коронавирус, имеют геномы, представленные молекулами РНК, но вирусы – это особая форма существования материи, организмами считаться не могут и поэтому останутся за пределами нашего внимания.

<sup>3</sup> Что касается описательной биологии, то в ней, если и происходят какие-либо измерения, например, размеров листьев и стеблей, применяются большие допущения, поскольку подобные параметры могут весьма сильно отличаться, даже у одного растения.

<sup>4</sup> Здесь под качеством следует понимать типы азотистых оснований - А, С, G, Т (или U для РНК), а также модификации некоторых.

геномами. Любые технологии секвенирования ДНК предполагают фрагментацию этой молекулы на части. С помощью одних методов (разных поколений) происходит секвенирование фрагментов ДНК весьма малого размера в диапазоне (если очень округленно) от 100 до 1000 нуклеотидов. Другие методы новых поколений позволяют секвенировать участки ДНК длиной до 10 тысяч нуклеотидов и более, но их точность пока, к сожалению, уступает методам, обеспечивающим короткие прочтения. Но в любом случае, фрагменты генома длиной даже в 10 тысяч нуклеотидов и более - это ничтожно мало по сравнению с длиной даже самой маленькой 21-ой хромосомы человека, несущей ДНК размером в 46 млн.п.н.

После получения первичных данных в виде множества коротких (или длинных, или тех и других) прочтений (ридов - reads) с помощью компьютерных программ на основе совпадения последовательностей нуклеотидов их концевых участков производится сборка таких ридов в более протяженные фрагменты, называемые контигами, о которых нужно сказать отдельно. Так, впервые эта дефиниция была предложена R.Staden в 1980 г., решившим сократить использовавшееся при анализе нуклеотидных последовательностей выражение “contiguous consensus sequence” до “contig” [Staden, 1980]. Недавно введено в обиход новое понятие “haplotig”, обозначающее гаплотип-специфичный контиг [Koren et al., 2018], используемое при сборке диплоидных геномов. Что касается сборки гаплоидного генома и определения нахождения в том или ином месте ДНК конкретного нуклеотида, то таковой при формировании контигов определяется его преимущественным прочтением. При этом с одной стороны исключаются неизбежные ошибки метода, но с другой в случае гетерозиготной последовательности один из нуклеотидов будет проигнорирован, поскольку производится сборка одной (фактически гаплоидной) цепи ДНК. Точнее, такая цепь даже при секвенировании ДНК одного человека получается по сути квази-гаплоидной, поскольку не учитывает группы сцепления и объединяет фрагменты ДНК, гораздо меньшего размера, чем это происходит при кроссинговере. Таким образом, восстановленный с помощью такого подхода геном не будет соответствовать ни одной из гамет, реально образующихся в организме, и поэтому его действительно правильнее считать квази-гаплоидным. Конечно, по теории вероятности может произойти воссоздание действительно одного из истинных гаплоидных геномов, но вероятность этого события ничтожно мала, и ее даже, скорее всего, не подсчитать. А если для секвенирования брались молекулы ДНК разных людей, то такой геном тем более будет квази-гаплоидным, поскольку *a priori* в

нем не будет восстановлена ни одна из гамет. Теоретически можно секвенировать геномы, которые несут отдельные сперматозоиды или яйцеклетки, но какое их количество по отдельности должно браться в эксперимент и сколько одногаметных геномов, учитывая всевозможные группы сцепления, должно быть прочитано, чтобы получить представление о диплоидном геноме, никто, наверное, не ответит. Поэтому для большинства целей можно ограничиваться определением квази-гаплоидной последовательности генома, представляющей собой по сути нечто среднее из диплоидного генома. Однако для оценки функционирования генома для персонифицированной медицины (а также для других целей, о которых будет говориться дальше) необходимо иметь информацию об обоих аллельных вариантах в виде полноценного диплоидного генома, содержащего сведения о двух наборах хромосом, полученных индивидом от отца и от матери. Конечно, единичные гены, ответственные за те или иные болезни человека, могут быть направлены секвенированы и сейчас в виде обоих аллелей, причем не только кодирующих областей (экзома), но и регуляторных участков, так как и в них кроются причины некоторых заболеваний. Однако в будущем задача будет стоять сложнее и требовать анализа взаимодействия различных генов, что проще будет делать, располагая полным диплоидным геномом человека, состоящим из 6 млрд.п.н., а не 3 млрд.п.н., характерных для гаплоидного генома, и сейчас указываемых как истинный размер генома человека.

Полный геном человека – 6 млрд.п.н.! Но, к сожалению, нынешние технологии секвенирования, включая трио-секвенирование, компьютерный анализ для восстановления фазированных последовательностей и истинных гаплотипов, не позволяют легко определять полные диплоидные геномы, но к этому нужно стремиться.

Ранее мы коснулись истории появления в 1967 г. термина «гаплотип» [Чемерис и др. (Chemeris et al.), 2020], но считаем необходимым сделать это и в отношении термина «диплотип», представляющего собой диплоидную пару гаплотипов, и впервые использованного в 1975 г. [Riggio et al., 1975], а в следующий раз только в 1985 г. уже упоминавшимся N.E.Morton с коллегой [MacLean, Morton, 1985]. Лишь с начала 2000-х гг. диплотипы прочно вошли в лексикон генетиков, упоминаемые даже в заголовках статей [Kitamura et al., 2002]. При этом можно не сомневаться, что в связи с секвенированием диплоидных геномов популярность диплотипов заметно возрастет.

Все сказанное выше о человеке относится и к другим эукариотическим организмам, имеющим двуродительскую природу. Некоторым исключением

являются гаплоидные организмы, например, у всем известных пчел - это трутни. Также гаплоидный геном имеют специально созданные так называемые дигаплоидные формы растений, фактически являющиеся диплоидными, но несущими два одинаковых гаплоидных генома.

### ПЦР, ГМО, ДНК

Аббревиатура «ПЦР» и так была довольно широко известна, поскольку диагностика различных болезней уже много лет проводится с ее помощью, однако тот же коронавирус, добавив еще больше внимания к этой реакции, которая, к сожалению, может характеризоваться как ложноположительными, так и ложноотрицательными результатами. И это крайне плохо. Причем причин для того и другого исхода эксперимента весьма много, и ранее мы их достаточно подробно рассматривали [Чемерис и др. (Chemeris et al.), 2012; 2012a]. Также нами недавно опубликована статья, в которой описывается улучшение протекания ПЦР с помощью различных моно- и дисахаридов [Sakhabutdinova et al., 2020]. Готовится еще ряд статей на эту крайне важную тему, поскольку от успешности проведения ПЦР зависит эффективность и точность той же самой ДНК-диагностики. Несмотря на целую группу других реакций амплификации, в том числе изотермических, характеризующихся определенными преимуществами, ПЦР остается фактически методом номер 1 в молекулярной биологии и в целом ряде смежных дисциплин, даже таких далеких как ДНК-криминалистика. Вряд ли какой-либо метод амплификации специфичных участков нуклеиновых кислот сможет основательно потеснить ПЦР в будущем, хотя и исключать подобное нельзя.

Аббревиатура «ГМО», служившая неким «пугалом» на протяжении многих лет - возможно уходящий термин. Думать так позволяет «набирающее обороты» геномное CRISPR/Cas редактирование, и в одной из наших статей, посвященных проблеме ГМО [Вершинина и др. (Vershinina et al.), 2020], мы даже спрогнозировали появление нового термина - «ГРО» (Геном-Редактированные Организмы), но, насколько нам известно, в чиновничьих кругах зреет иное обозначение растений, созданных генно-инженерным путем, как «ГИМР» (Генно-Инженерно Модифицированные Растения). Возможно, как «уход» от ГМО. Благозвучность такого названия комментировать не будем, но вряд ли оно приживется. При этом ГРО<sup>5</sup> могут быть как трансгенными и, следовательно, подпадать под старое определение ГМО, так и представлять собой

имитации природных процессов, не неся никакой чужеродной ДНК, и тогда это уже не ГМО в его классическом понимании. Все это довольно подробно рассмотрено нами в целой серии статей, посвященных редактированию геномов растений с помощью передовой CRISPR/Cas-технологии [Баймиев и др. (Baymiev et al.), 2017; Кулуев и др. (Kuluev et al.), 2017; Чемерис и др. (Chemeris et al.), 2017; Кулуев и др. (Kuluev et al.), 2019; Геращенко и др. (Gerashchenkov et al.), 2020]. Возвращаясь к термину ГМО, еще одной причиной его возможно относительно скорого забвения, служит то, что в нашей стране стали уделять повышенное внимание созданию сортов сельскохозяйственных растений с помощью методов генной инженерии, и это становится политикой государства, тогда как некоторое время назад было наоборот. К тому же подобные растения не опасны ничуть, о чем мы уже неоднократно писали [Чемерис и др. (Chemeris et al.), 2014; 2015; Вершинина и др. (Vershinina et al.), 2020].

Что касается термина «ДНК», то это вещество «отпраздновало» в прошлом году 150-ти летний юбилей своего открытия выдающимся швейцарским ученым Ф.Мишером, назвавшим его «нуклеином», чему были посвящены несколько статей в одном из прошлогодних номеров «Биомики» [Гарафутдинов, Чемерис (Garafutdinov, Chemeris), 2019; Геращенко и др. (Gerashchenkov et al.), 2019; Byrne, Dahm, 2019]. Однако, нуклеин, переименованный в нуклеиновую кислоту в 1889 г. немецким ученым Р.Альтманом, стал называться ДНК только в начале 1930-х г. В то время еще не было известно, что ДНК – молекула наследственности, но термин «геном» уже существовал. Поскольку под геномом сейчас понимается вся совокупность ДНК в каком-либо организме, а хождение термина «геном» в народе будет только расти с приближением персонифицированной медицины, то и ДНК забыта не будет.

### Геном. История и смысл термина(ов)

В 1920 г. в монографии, посвященной вопросам распространения и причинам партеногенеза в растительном и животном царствах “*Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche*”, немецкий ботаник Ганс Винклер<sup>6</sup> впервые предложил термин «геном» [Winkler, 1920]. В главе про связь хромосом с партеногенезом на стр. 165 им высказана следующая мысль - «*Ich schlage vor, für den haploiden Chromosomensatz, der im Verein mit dem zugehörigen Protoplasma die materielle Grundlage der systematischen Einheit darstellt, den Ausdruck: das Genom...*», которая в переводе на русский могла бы

<sup>5</sup> ГРО – тоже не лучший вариант обозначения новых генно-инженерных растений.

<sup>6</sup> Ганс Винклер (нем. Hans Karl Albert Winkler) – профессор ботаники Университета в Гамбурге (23 апреля 1877 г. — 22 ноября 1945 г.).

звучать приблизительно так: «Я предлагаю использовать для гаплоидного набора хромосом, который вместе с прилежащей протоплазмой определяет материальные основы вида, выражение геном...». В этом же предложении Г.Винклер рассуждает далее о том, что если геном содержит более чем одну такую же единицу, то его следует называть гомогеноматическим, а если разные единицы, то тогда - гетерогеноматическим. Сейчас

подобные организмы принято обозначать как автополиплоиды и аллополиплоиды соответственно, а составные части их геномов – субгеномами, но к вопросу о них мы еще вернемся.

Как можно видеть из рисунка, на котором приведен фрагмент страницы 165 книги “Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche” Винклер в одном абзаце упомянул геном почти 10 раз.

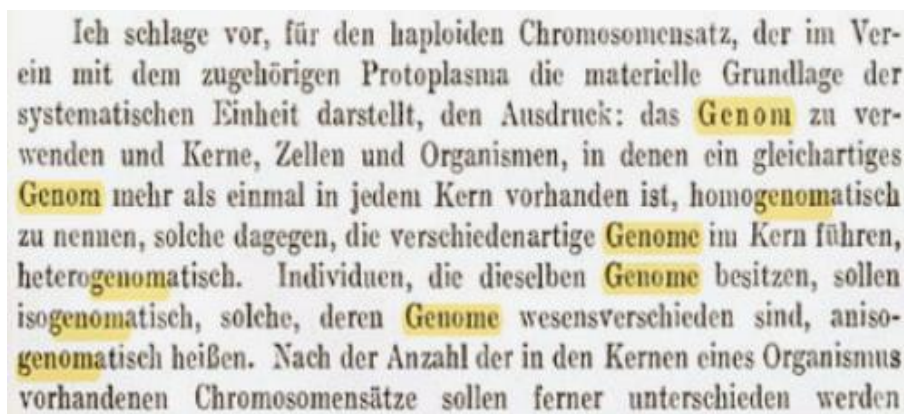


Рисунок. Фрагмент 165-ой страницы книги Г.Винклера “Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche”, изданной в 1920 г. в Йене (Германия)

Figure. Fragment of page 165 of the book by G.Winkler “Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche”, published in 1920 in Jena (Germany)

Oxford English Dictionary дает следующую трактовку происхождения термина “genome”<sup>7</sup>, возможно представляющего собой комбинацию объединенных Винклером слова “gene” и концовку слова “chromosome”. Также интересен взгляд известного ученого Дж.Ледерберга и его коллеги [Lederberg, McGraw, 2001] на происхождение термина “genome”. Они предположили (поскольку такую информацию Винклер не оставил), что он мог не заниматься соединением слов «ген» и «хромосома», а к слову “gene” добавить концовку в виде “-ome”, придав новому слову тем самым смысл некоего множества (генов) в хромосомах. Объяснение этому Ледерберг и МакГроу видят в том, что к 1920 г. в ходу уже были термины “biome”, “rhizome”, “phyllome”, “thallome” и др., обозначающие объединение неких единиц, которыми для генома явились гены. Но вполне вероятно, что с этимологией этого термина все проще и сложнее одновременно. Нам представляется, что Винклер мог использовать морфему в виде самостоятельного слова

«сома» (являющееся частью слова хромосома), которое на греческом обозначает «тело», а также морфему «ген» и, превратив последнюю в определение «генное», в итоге получил «генное тело» или «геносому» (genosom, нем.), которое счел нужным для удобства сократить, убрав один слог. Причем, нельзя исключать, что Винклер имел при этом ввиду и некие множества, которое дает окончание “-om” (“-ome”), допустив игру слов. Как бы то ни было, термин «геном» в виде объединения генов в нечто целое, оказался очень удачным, в том числе благодаря своей краткости, но при этом, превосходя по длине слово «ген», совокупность которых геном и объединяет.

Говоря о прилежащей цитоплазме, Винклер, похоже не думал о хлоропластах и митохондриях, которые, как мы сейчас знаем, также несут свою ДНК, и этот генетический материал (пластом и хондриом) иногда называют объединяющим термином «плазмон», который предложил в 1926 г. при обозначении генетического элемента цитоплазмы для его дифференциации от ядерного вещества наследственности другой немецкий ботаник F.R. von Wettstein (1895-1945 гг.) [Michaelis, 1954<sup>8</sup>]. Но

<sup>7</sup> Здесь и далее преимущественно используется английское написание терминов “gene”, “genome”, “chromosome” и прочих терминов, заканчивающихся на “-ome”, тогда как в немецком языке (на котором и был предложен термин “genom” все будет аналогично, только без окончания “e”.

<sup>8</sup> К сожалению, в разных источниках, включая толковые словари по биологии и биотехнологии, можно встретить указание о том, что «плазмон» введен



плазмоны в силу ряда причин, одна из которых - их мультикопийный характер, практически останутся за пределами рассмотрения в данной статье, поскольку нас будет интересовать в основном ядерный геном высших организмов и, в частности, человека.

Итак, термину «геном» в этом году - 100 лет, и в последние десятилетия в ненаучной среде он прочно ассоциируется лишь с человеком, главным образом по причине выполнения международной программы «Геном человека» и многочисленных публикаций на этот счет в обычной прессе, а также в интернете. При этом достаточно длительное время после его предложения «термин» геном использовали довольно редко и почти исключительно при описании хромосомных наборов растительных организмов. К тому же в 1920-ые и в 1930-ые годы были даже ярые противники термина «геном», среди которых были известные генетики и, в частности, Т.Добжанский (Т. Dobzhansky), который, впрочем, спустя много лет поменял свою точку зрения и в 1970 году в одной из своих работ использовал оба термина – геном и плазмон [Noguero-Solano et al., 2013]. При этом термин «геном» в первые десятилетия его существования использовался отечественными учеными. Так, свою статью, посвященную изучению одного из видов пшениц, В.В.Светозарова (Svetozarova) [1939] назвала «О втором геноме *T. timopheevii* Zhuk.». В заголовках статей зарубежных авторов прежних лет, посвященных растениям, также можно встретить слово “genome” [Smith et al., 1943; Stebbins, Pun, 1953].

Термин «субгеном» встречается при описании амфидиплоидного генома хлопчатника, по крайней мере, еще в 1954 г. [Menzel, 1954]. И для полиплоидных видов растений, особенно несущих разнородные геномы, использовать термин «субгеном» и правильно, и удобно, поскольку, пройдя через стадии интеграции и функциональной «диплоидизации», геномы (хромосомы) исходных родительских форм за счет как внутри-, так и межгеномных перестроек (транслокаций) в составе аллополиплоидных растений изменяются, представляя собой, по существу, уже элементы интегрального полиплоидного генома. К тому же это позволяет облегчить их дифференциацию от донорных геномов полиплоидных форм, например, самостоятельного генома **D** эгилопса *Aegilops tauschii* от его производного - субгенома **D**, являющегося частью составного гексаплоидного **VAD** генома мягкой пшеницы. Насколько нам известно, для обозначения геномных взаимоотношений в пшенично-эгилопсом альянсе этот термин впервые был употреблен лишь в 2000 г. [Stein et al., 2000]. При характеристике

полиплоидных пшениц мы также стали применять термин «субгеном» с начала нынешнего столетия [см. Кулуев и др. (see Kuluev et al.), 2016]. Однако регулярный характер использование понятия субгеномов при описании полиплоидных пшениц приобрело лишь с 2008 г. [Gupta et al., 2008]. Помимо полиплоидных растений понятие «субгеном» распространяют и на вирусы с сегментированными геномами, но, как уже говорилось выше, они останутся за пределами нашего рассмотрения. Кроме генома и субгенома сейчас в ходу также термины «метагеном» и пангеном», изначально предложенные на рубеже XX и XXI веков для микроорганизмов, и не могущие быть диплоидными. По сей причине мы касаться их не будем. Хотя нужно заметить, что для пангенома границы его применения заметно расширились с того времени.

В уже упоминавшейся статье [Noguero-Solano et al., 2013] довольно много внимания уделено меняющемуся смыслу, вкладывавшемуся на протяжении десятилетий в понятие «геном». Так, кроме классического определения, что геном - это совокупность гаплоидного набора хромосом, звучали предложения считать геномом все гены в клетке, т.е., переводя данное понятие на диплоидный уровень. Нужно заметить, что в основе восприятия генома в 1920 г. как гаплоидного набора хромосом лежали ограничения метода их визуализации путем окрашивания и наблюдения в световом микроскопе. Собственно, и позже технологические возможности влияли на смысл, вкладываемый в понятие генома. Известный генетик Г.Стент определил геном как совокупность всех генов индивида (диплоидный уровень), тогда как Дж. Уотсон в своем эпохальном труде «Молекулярная биология гена» указал, что геном - это гаплоидный набор хромосом, несущих ассоциированные с ними гены.

В другой своей статье эти же авторы [Noguero-Solano et al., 2017] решили вынести в заголовок новое слово «геномизация»<sup>9</sup>, посчитав, что этот процесс в биологической науке имеет место. По сути, они правы и в этом нет ничего дурного, поскольку именно понимание функционирования всего генома, опирающееся в первую очередь на информацию о его структурной организации, выявляемой сейчас главным образом секвенированием полных геномов методами новых поколений - это то, что требуется для расширения и углубления наших знаний о Живом. В качестве подтверждения действительно происходящей «геномизации» биологической науки можно сослаться на данные Clarivate Analytics, ранжирующих научные журналы по квартилям, из чего можно видеть, что

P. Michaelis в 1954 г., тогда как в этом своем труде сам Michaelis пишет о том, что он использует термин плазмон, предложенный в 1926 г. F.R. von Wettstein.

<sup>9</sup> Этот предложенный ими термин не прижился, хотя они и привели в своей статье данное слово в качестве ключевых слов.

свыше четырех десятков журналов (большая часть которых относится к Q1 и Q2) в своих названиях содержат слова “genome”, “genomic” или “genomics”. Причем среди них немало журналов, сменивших свои прежние названия. Так, первым это сделал издающийся с 1959 г. канадский журнал *Canadian Journal of Genetics and Cytology*, с 1987 г. получивший название “Genome”. Есть и другие примеры подобных переименований журналов с историей. Например, основанный еще в 1908 г. *MGG - Molecular and General Genetics* с 2002 г. стал *Molecular Genetics and Genomics*. И здесь о слове “Genomics” нужно сказать отдельно. Предложенное в 1987 г. название нового международного журнала *Genomics* (в настоящее время это журнал Q1 с импакт-фактором 6,205) стало самостоятельной интенсивно развивающейся наукой из группы наук о Жизни. В редакторской статье к первому номеру этого журнала V.A.McKusick и F.H.Ruddle [1987] объяснили свое решение тем, что «...логии» (по нашей несколько вольной и сокращенной трактовке их высказываний) очень академичны, тогда как «...мики», на их взгляд, предполагают некий атакующий стиль в исследованиях Живого. Однако данный термин быстро перерос границы журнального названия и это слово, став названием новой науки «геномики», дало спустя некоторое время жизнь и другим «...микам», среди которых первой была «протеомика», а потом уже и остальные – «транскриптомика», «метабомика» и т.д.<sup>10</sup> Но автором слова и затем названия науки геномики явился T.Roderick. о чем он рассказал в одном из своих интервью [Kuska, 1998]. Так, в 1986 г. во время симпозиума “The Human Genome” было проведено совещание по вопросу организации нового журнала и было предложено придумать его название. Часть совещавшихся пошла затем в близлежащий бар и за пивом предлагали разные варианты, склоняясь к тому, что в названии должно присутствовать слово “genome”, но назвать так коротко было нельзя, поскольку вышеупомянутый канадский журнал уже объявил о своем переименовании. Другие предлагавшиеся названия были слишком громоздкими, и тут T.Roderick предложил использовать новое слово “genomics” с окончанием от genetics, не думая о том, что это может послужить оформлению новой науки – геномики, главной задачей которой является всесторонне изучение геномов организмов.

Возвращаясь к статусу геномов, следует

<sup>10</sup> Название журнала «Биомика» фактически объединяет под своей обложкой разнообразные «-омики», поскольку в нем как раз преимущественно публикуются материалы подобной направленности, о чем говорилось в редакторской статье к первому номеру журнала [Чемерис и др. (Chemieris et al.), 2011].

заметить, что в настоящее время, к которому мы вскоре перейдем, под ядерным геномом понимается совокупность ДНК, находящейся в ядре клетки в ее половинчатом гаплоидном состоянии. При этом размеры гаплоидных геномов даже высших организмов могут варьировать в размерах, отличаясь на несколько порядков, от нескольких десятков млн.п.н. до десятков млрд.п.н. Причем четкой закономерности между размером генома и нахождением конкретного вида на эволюционной лестнице не наблюдается. В случае полиплоидных организмов размеры их геномов определяются с учетом всех субгеномов. Так, например, гаплоидный геном **VAD** мягкой пшеницы *Triticum aestivum*, оцениваемый в 17 млрд.п.н., включает в себя субгеном **V**, происходящий от некоего эгилопса из секции *Sitopsis*; субгеном **A**, донором которого стала одна из диплоидных пшениц, и субгеном **D**, привнесенный *Aegilops tauschii*. При этом полный гаплоидный геном **VAD** известен [Zimin et al., 2017], но состоящий из 42 хромосом диплоидный геном (**VBAADD**,  $2n = 6x$ ) этой пшеницы размером в 34 млрд.п.н. – нет, поскольку гаплоидный геном (для этого вида пшеницы –  $3x = 21$ ) не может быть просто механически умножен в два раза, ввиду того, что организмы, размножающиеся половым путем<sup>11</sup>, несут два отличающихся гаплоидных генома. И именно их комбинация во многом определяет фенотип организма, его функциональность.

#### Геном (гаплоидный) человека

Опубликовано не так много ранних статей, в которых в связи с человеком упоминался его геном. Более того, если по базе данных PubMed поискать подобные публикации с ключевыми словами “human” и “genome” с Булевым оператором AND, то таковые можно найти, но на проверку окажется, что слова “genome” в тексте статьи и нет. Так, в качестве примера можно привести найденную нами таким образом публикацию 1948 г., посвященную подсчету числа генов во всех хромосомах человека [Spuhler, 1948]. Нужно сказать, что количество генов на гаплоидный геном было определено довольно точно – в диапазоне от 219890 до 30420 генных локусов или, используя другой подсчет – от 20 до 42 тысяч. При этом нужно учесть, что как вообще организована ДНК, тогда не было известно.

В 1970-х гг. стали появляться статьи, в заголовках которых уже фигурировали слова “human genome” [Saunders et al., 1972; Lieberman, Poirer, 1974; Schmid, Deininger, 1975; Deininger, Schmid, 1976]. В

<sup>11</sup> Виды растений, размножающиеся в норме вегетативно или бесполосемянно, когда-то все равно произошли при объединении двух гаплоидных гамет и несут отличающиеся парные хромосомы.



них о секвенировании еще речь не шла, и описывались организация генома, исходя из соотношений фракций повторяющихся последовательностей, выявленных кинетикой реассоциации, а также из результатов электронной микроскопии. В те годы метод кинетики реассоциации являлся мощным инструментом анализа всего генома и позволил в одной из работ [Saunders et al., 1972] оценить размер гаплоидного генома человека с довольно высокой точностью, определив его общую длину в  $7 \times 10^9$  п.н., что приблизительно лишь в два раза превысило истинное значение. В качестве некоего подтверждения точности своих расчетов эти авторы сослались на статью разработчика данного метода Р.Бриттена и соавт. [Gelderman et al., 1971], в которой был «измерен» геном мыши, составивший по их оценкам чуть меньшую величину –  $3 \times 10^9$  п.н., что также недалеко от истины, как мы сейчас знаем.

Впервые всерьез о необходимости определения нуклеотидной последовательности всех 24 хромосом (22-х аутосом и 2-х половых X и Y хромосом) человека заговорили в середине 80-х годов прошлого столетия, после того, как произошла автоматизация процесса секвенирования ДНК ферментативным методом [Smith et al., 1986; Hood et al., 1987]. Так, вопросы о возможности секвенирования полного генома человека поднимались в 1986 г. на двух конференциях в Санта Фе и Колд-Спринг-Харборе [Palca, 1986; Robertson, 1986]. Обсуждались стратегии секвенирования и чью (мужскую или женскую) ДНК выбрать; в том числе не исключались разные люди для секвенирования различных хромосом. Были произведены предварительные подсчеты стоимости такого проекта, согласно которым при 10-ти кратном покрытии генома потребовалась бы работа в режиме 16/5 (часы/дни) 580 рабочих станций, состоящих из четырех (только готовящихся в тот момент к выводу на рынок) автоматических секвенаторов, обслуживаемых двумя лаборантами, что позволило бы получить необходимый объем данных за один год. На эту работу потребовалось бы 290 млн. долларов [Palca, 1986]. Собственно, такая сумма и ушла на непосредственно само секвенирование, которое затянулось на полтора десятилетия, о чем будет говориться ниже. При этом были и скептики, подсчитавшие, что для секвенирования полного генома человека потребуются приблизительно 30 тысяч человеко-лет [Robertson, 1986]. Существовали и опасения, что этот проект отнимет слишком много сил и денежных средств у других важных направлений исследований в молекулярной биологии. В еще одной краткой заметке о секвенировании генома человека того времени в журнале Nature [Walsh, Marks, 1986] задавались вопросы: «зачем секвенировать», «чей геном секвенировать» и «что потом делать с секвенированным геномом». При этом

воодушевившись первыми успехами при работе с древней ДНК (еще в доПЦРную эру), эти авторы посчитали, что следует секвенировать полный геном Ч.Дарвина, для чего потребовалась бы эксгумация его останков, погребенных в Вестминстерском аббатстве. Но обращаем внимание на эту публикацию главным образом потому, что в ней прозвучало, что необходимо секвенировать гаплоидный геном<sup>12</sup>, поскольку диплоидный может привести к неоднозначным данным ввиду различий между парными хромосомами. В том числе и поэтому предлагалось секвенировать геном мужчины для того, чтобы половые хромосомы хотя бы не были парными.

В 1988 г. в США был создан Национальный центр по исследованию генома человека, возглавил который тогда Дж.Уотсон. Немного позднее, в 1990 г. в США был разработан проект «Геном человека», практически сразу ставший международным, поскольку многие страны изъявили желание принять участие в его выполнении. Но прошло довольно много времени, прежде чем были опубликованы статьи, сообщившие о секвенировании в черномом варианте первых двух геномов человека с помощью ферментативного метода Сэнгера. Причем эти публикации, подготовленные Международным консорциумом и Институтом геномных исследований вместе с компанией Celera Genomics под руководством К.Вентера, вышли практически одновременно в журналах Nature и Science в 2001 г. [Lander et al., 2001; Venter et al., 2001]. Первый проект выполнялся 13 лет (вместо планировавшихся 15 и изначально даже 17 лет) и на него было потрачено приблизительно 3 млрд. долларов, из которых только около 300 млн. пошло непосредственно на проведение секвенирующих процедур, поскольку много сил, времени и денег Консорциум затратил на поиск различных маркеров и построение генетических карт<sup>13</sup>. Что касается проекта Вентера, то при его

<sup>12</sup> Здесь нужно заметить, что секвенирование гаплоидного генома в силу технологических возможностей предполагает по сути определение некоей половины диплоидного генома, причем в виде мозаики из отцовских и материнских хромосом. Секвенирование же полного диплоидного генома несравнимо сложнее и об этом будет говориться далее.

<sup>13</sup> Оказавшихся малонужными и как для непосредственно самого процесса секвенирования, и тем более после завершения проекта «Геном человека», поскольку эти карты практически полностью потеряли свою актуальность ввиду ставшей доступной принципиально иной новой более точной информации в виде полной последовательности нуклеотидов всех хромосом.

выполнении был применен передовой для того времени так называемый «shotgun» или иначе «случайный» подход к секвенированию больших геномов, что позволило справиться с секвенированием генома человека всего за три года, затратив лишь около 100 млн. долларов. Для этого им хватило 300 секвенаторов, получаемые результаты на которых обрабатывались одним мощным компьютером, тогда как в Международном проекте использовалось около 600 секвенаторов, разбросанных по всему миру, и множество компьютеров. Причем тогда для обеих групп это было секвенирование *de novo*, которое несравнимо сложнее последующих многочисленных ресеквенирований генома человека. При этом Международный консорциум использовал при секвенировании генома человека ДНК довольно большого числа людей, а под руководством Вентера было проведено секвенирование фрагментов геномов пяти человек (2 мужчин и 3 женщины), но, как выяснилось позже, значительная часть ДНК принадлежала самому Вентеру. Таким образом, оба секвенированных генома оказались однозначно квази-гаплоидными.

Наверное, самым удивительным результатом секвенирования полного генома человека стало весьма малое число генов, кодирующих белки. По различным оценкам, оно варьирует всего от 20 до 25 тысяч генов всего, тогда как ранее назывались цифры куда большие. Здесь уместно еще раз вспомнить статью 1948 г. [Spuhler, 1948], в которой число генов у человека было оценено в диапазоне от 20 до 42 тысяч. При этом незадолго до завершения секвенирования генома человека прогнозировавшиеся разными учеными количества генов варьировали от 27500 до 100000, 118000, 120000, 145000 и даже 153500 генов [Pennisi, 2000]. Видимо не все были знакомы с работой 1948 г.

Международный консорциум по секвенированию генома человека 14 апреля 2003 г. объявил о завершении работы над расшифровкой последовательности генома человека. В октябре 2004 г. ими была опубликована статья, в которой сообщалось о практически полном завершении (на >99%) секвенирования эухроматиновой части генома человека и Построения генома версии 35 [ISHGC, 2004]. Что касается генома человека, секвенированной компанией Celera, то в феврале 2004 г. опубликована статья [Istrail et al., 2004], в которой они сравнили результаты их улучшенного чернового варианта генома с существовавшим тогда IHGSC Построением 34 генома человека.

В США, начиная с 2004 г., через NHGRI - National Human Genome Research Institute в рамках Программы развития технологий секвенирования (Sequencing Technology Development Program) появилось множество проектов по разработке

технологий секвенирования ДНК новых поколений, которым оказывалась, в том числе, и государственная поддержка. Но еще в 2003 г. научный фонд К.Вентера объявил об учреждении премии за разработку быстрой и дешевой технологии секвенирования геномов размером 500 тысяч долларов. Однако спустя некоторое время был объявлен новый приз, о котором будет говориться дальше.

В последующие годы определенное внимание стало уделяться персональным геномам отдельных людей. Так, в мае 2007 г. при секвенировании генома Нобелевского лауреата Дж.Уотсона с помощью метода пиросеквенирования ДНК был достигнут исторический рубеж в один миллион долларов на геном [Wheeler et al., 2008]. Геном Уотсона стал первым, который был расшифрован с помощью метода секвенирования нового поколения – Next Generation Sequencing (NGS). Позже полупроводниковая технология секвенирования также из группы NGS была успешно использована для секвенирования индивидуального генома Г.Мура [Rothberg et al., 2011] – одного из основателей компании Intel и автора широко известного закона развития микроэлектроники [Moore, 1965]. Оба эти метода секвенирования ДНК новых поколений основаны на массовом параллелизме секвенирующих реакций. Еще один индивидуальный геном – одного из авторов работы профессора S.R.Quake был прочитан с помощью метода мономолекулярного секвенирования из группы NGS [Pushkarev et al., 2009].

С января 2008 г. начал выполняться международный проект 1000 Genomes, ставящий целью определить полные геномы не менее тысячи анонимных индивидуумов и создать подробный каталог генетических вариаций генома человека, а также улучшить референсную последовательность. По завершению пилотной стадии в 2010 г. [The 1000 Genomes Project Consortium, 2010], наряду с полными геномами с относительно низким покрытием 179 человек из четырех популяций и секвенирования экзотов у 697 человек из семи популяций, были (в среднем с 42-х кратным покрытием) также секвенированы трио геномов (мать, отец, ребенок) для двух семей. При этом такое трио-секвенирование было произведено для подсчета скорости мутаций в поколениях, составившей по этим данным около  $10^{-8}$  на нуклеотид. В ходе выполнения этого проекта 1000 геномов были секвенировано 2504 индивидуальных генома из 26 популяций [1000 Genomes Project Consortium, 2015; Sudmant et al., 2015].

Улучшение референсного генома GRCh38, который сконструирован на основе данных нуклеотидных последовательностей многих доноров, позже было продолжено и другими авторами [Guo et al., 2017]. Причем этот референсный геном GRCh38 секвенирован методом Сэнгера, являющимся

«золотым стандартом» в плане точности, превосходя приблизительно в 10 раз по этому показателю методы новых поколений. Фактически референсный геном GRCh38 также является квази-гаплоидным, поскольку в нем «перемешаны» последовательности парных хромосом разных индивидов. При этом ценность такого выверенного генома не вызывает сомнений, и его до некоторой степени можно считать типовым для вида *Homo sapiens*. Собственно, и персональные гаплоидные геномы, особенно секвенированные методами новых поколений с разбиением тотальной ДНК на мелкие участки, можно считать квази-гаплоидными в том смысле, что группы сцепления в них как таковые определить невозможно, и мозаицизм в виде участков из отцовских и материнских хромосом довольно высок. В том числе и по этой причине нужно определение нуклеотидных последовательностей диплоидных геномов.

#### Диплоидный геном человека

4 октября 2005 г. калифорнийской компанией Archon Genomics при участии Фонда Вентера [Venter, 2010] было объявлено о готовящемся состязании секвенирующих команд, главный приз за победу в котором должен был составить 10 млн. долларов, что, по мнению организаторов, должно было стимулировать развитие новых технологий секвенирования (<https://genomics.xprize.org/prizes/genomics>). Деньги были пожертвованы канадским филантропом геологом S.Blusson. В борьбе за приз Archon Genomics X Prize предстояло определить полные нуклеотидные последовательности 100 диплоидных геномов человека за 10 дней не более чем за 10 тысяч долларов каждый с точностью 99,999% (одна ошибка на 100 тысяч нуклеотидов). Тогда получила хождение фраза «геном медицинского качества», поскольку диплоидный геном мог быть использован для улучшенной оценки состояния здоровья человека и служить основой для рекомендаций по лечению конкретных заболеваний. Следует отметить, что в то время главенствующей технологией секвенирования служил метод Сэнгера, и на один гаплоидный геном уходило до полугода и требовалось около 100 тысяч долларов. Однако спустя некоторое время стало ясно, что условия конкурса необходимо откорректировать [Kedes, Liu, 2010]. Так, согласно новым требованиям, оглашенным в октябре 2011 г., в состязании (получившим название 100 на 100) предстояло секвенировать 100 фазированных по гаплотипам геномов долгожителей со всего мира в возрасте 100 лет и старше в течение 30 дней при стоимости в 1000 долларов за геном. Фонд Archon Genomics X Prize собрал образцы крови у нужного числа долгожителей и создал соответствующие клеточные линии. Конкурс должен был начаться 5 сентября 2013 г., но 22 августа был неожиданно отменен организаторами,

посчитавшими, что главную цель в виде развития новых технологий секвенирования он не достигнет, поскольку инновации и так уже опередили время. Что вызвало недовольство, по крайней мере, у двух уже зарегистрировавшихся команд в качестве участников. Столь небольшое число желающих посоревноваться также послужило дополнительным поводом денонсировать конкурс, поскольку было сочтено, что сумма в 10 млн. долларов большинству потенциальных конкурсантов не интересна.

Но секвенирование диплоидных геномов человека в то время понемногу шло. Так, в октябре 2007 г. была опубликована статья, в которой сообщалось о секвенировании индивидуального диплоидного генома человека, получившему обозначение HuRef и принадлежавшего К.Вентеру [Levy et al., 2007]. В рамках этой работы методом Сэнгера было определено 2810 млн.п.н. с 7,5-кратным покрытием и с помощью новой стратегии сборки геномов, учитывающей гаплотипы, 1,5 млрд.п.н. были фазированы сегментами по 200 тысяч п.н. Таким образом, геном Вентера оказался прочитанным приблизительно наполовину как диплоидный и первым в таком качестве. Следующий персональный диплоидный геном человека азиатского происхождения был секвенирован с 36-ти кратным покрытием и с помощью программы PHASE фазирован, что позволило идентифицировать в общей сложности 2434 гаплотипа размерами более 200 тысяч нуклеотидов, обеспечив 2,38 млрд.п.н. генома в диплоидном формате [Wang et al., 2008]. В данной работе при сравнительном анализе использовались персональные геномы Уотсона и Вентера. В последующие годы еще целый ряд диплоидных геномов человека, происходящих из разных мест Земного шара, были секвенированы и фазированы с той или иной степенью полноты [Ahn et al., 2009; Kitzman et al., 2011; Cao et al., 2015; English et al., 2015; Pendleton et al., 2015; Seo et al., 2016; Weisenfeldt et al., 2017; Zhang et al., 2019]. Для этого применялись различные подходы, как преимущественно использующие соответствующее программное обеспечение, так и секвенирование с большим покрытием, включая генерацию космидных клонотек с протяженными вставками, сборку *de novo* после комбинации длинных прочтений с помощью мономолекулярных методов секвенирования с высокоточными короткими прочтениями на основе флуоресцентного секвенирования, трио-секвенирование и пр.

Осенью 2020 г. появилась публикация [Soifer et al., 2020], в которой сообщено о секвенировании практически полного диплоидного генома человека. Для этого была использована комбинация инструментов и технологий секвенирования для *de novo* сборки диплоидного генома клеточной линии человека WI-38. Исходными данными послужили результаты длинных прочтений с помощью мономолекулярного SMRT-

секвенирования в реальном времени и оптического картирования вкупе с короткими фрагментами, получаемыми путем флуоресцентного (Illumina) высокоточного секвенирования метафазных хромосом, отсортированных проточной цитометрией, для получения фазированной информации. Окончательная сборка генома была почти полностью (на 94%) фазирована. Подробная информация об этом исследовании, включая фазированные нуклеотидные последовательности, доступные для скачивания и анализа с помощью специального WI-38 геномного браузера, выложены на созданном по этому случаю сайте <https://wi38.research.caliccolabs.com>.

Учитывая растущий интерес к диплоидным геномам человека, только за последние годы предложены новые подходы к получению таких данных, разработаны алгоритмы и написаны компьютерные программы, позволяющие вести *de novo* сборку фазированных гаплотипов [Koren et al., 2018; Roach et al., 2018; Satas, Raphael, 2018; Zhou et al., 2018; Mocci et al., 2019; Wang et al., 2019; Kajitani et al., 2020; Nguyen et al., 2020; Rhie et al., 2020]. Недавно сообщено о программе POLYTE (POLYploid filTEr), позволяющей генерировать гаплотипы для полипloidных организмов [Baaijens, Schönhuth, 2019].

Диплоидные геномы человека, которых пока крайне мало, представляют собой новое поколение геномных данных, необходимых для осуществления персонализированной медицины в будущем. Но нынешние технологии высокопроизводительного секвенирования все же не позволяют сделать такое секвенирование рутинным, как это уже имеет место для гаплоидных геномов. Требуются принципиально новые технологии секвенирования ДНК и, конечно же, полных геномов.

### Диплоидные геномы растений

Хотя для растений секвенирование диплоидных геномов не столь актуально как для человека, тем не менее, подобные работы с тем или иным успехом на протяжении ряда лет уже выполняются. Всего растительных геномов, которые могут считаться полными, уже секвенировано более сотни, но лишь для немногих из них нуклеотидные последовательности распределены по хромосомам и первым таким растением явился сорняк резуховидка Таля или арабидопсис *Arabidopsis thaliana*, являющийся в силу ряда причин (довольно маленький геном, короткое время вегетации, большое количество образующихся семян и пр.) модельным объектом. Неудивительно, что одним из первых фазированных растительных геномов стал именно геном этого растения. Помимо его относительно малого размера, в рамках Проекта 1001 геном таковых уже секвенировано значительное количество, большинство из которых доведены до стадии распределения

секвенированных нуклеотидных последовательностей по хромосомам. При составлении фазированного генома арабидопсиса использовали технологию трио-секвенирования гибрида и его родительских форм Col-0 и Cvi-0, а для сравнительного анализа использовали референсный геном TAIR10 [Chin et al., 2015]. Для достижения лучших результатов применяли несколько программ сборщиков геномов, в то числе, нацеленных на получение фазированных данных. В этой же работе был секвенирован диплоидный геном винограда *Vitis vinifera* сорта Каберне Совиньон, характеризующегося высоким уровнем гетерозиготности, что способствовало лучшей сборке фазированных последовательностей. С виноградом того же сорта работа по фазировке диплоидного генома затем была ими продолжена [Minio et al., 2016]. Позже этим авторским коллективом секвенирован диплоидный геном другого сорта винограда Carmenere [Minio et al., 2019]. В последние годы перечень культурных растений или их предковых форм, для которых осуществлены попытки определения их диплоидных геномов, пополнились картофелем [Zhao et al., 2020], грушей [Shi et al., 2019], яблоней [Sun et al., 2020], чайным деревом [Zhang et al., 2020]. При этом можно не сомневаться, что этот список будет расти. Тем более, если будет разработан принципиально новый высокопроизводительный метод секвенирования ДНК, позволяющий читать протяженные участки с минимальным числом ошибок.

Диплоидные геномы растений представляют интерес, в том числе в связи с CRISPR/Cas геномным редактированием, поскольку, как правило, бывает необходимо произвести изменения в обоих аллелях парных хромосом, и для этого нужно знать их нуклеотидные последовательности. Равно как и по завершении такого редактирования необходимо с помощью полногеномного диплоидного секвенирования выявить произведенные целевые и нецелевые мутации, что с дальнейшим развитием технологий секвенирования и массовым выводом растений с отредактированными геномами на поля, скорее всего, станет нормой.

### Заключение

На протяжении 100 лет существования термина «геном» концепция вкладываемого в него смысла менялась по двум направлениям сообразно развитию имевшихся знаний. Так, от понимания генома как гаплоидного набора хромосом, выполнявших вначале роль элементарных единиц, через совокупность генов (новых элементарных единиц) концепция пришла к восприятию генома как последовательности азотистых оснований, также являющихся сейчас элементарными наименьшими единицами любого генома. Вряд ли правомерно считать, что на уровне всего генома произойдет некое

уточнение статуса этих элементарных единиц в виде их эпигенетического состояния (например, наличия метильных групп у тех или иных цитозинового либо адениновых остатков), поскольку оно (такое состояние) способно меняться в процессе жизнедеятельности организма и не может быть константным. Почему сейчас понятие генома обходится без информации на белковом уровне в виде того же хроматина? Этому есть простое объяснение – именно в последовательности азотистых оснований заложены все нужные сведения. Геном – это как Свод законов, если применить юридический термин. При этом для понимания функционирования генома требуются знания и транскриптома, и протеома, и прочих «-омов», которые можно считать некими подзаконными актами и комментариями к Своду законов (Геному). Причем хромосомы опять стали являться важной составной частью генома, но уже не в виде элементарных единиц, а формируя более высокий уровень организации (как отдельные тома Свода законов), поскольку нуклеотидные последовательности полностью прочитанного генома компонуется по-хромосомно.

Другое важное направление восприятия генома, которое необходимо учитывать и которое время от времени в глазах разных ученых менялось от одного к другому, это какой уровень – гаплоидный или диплоидный – считать за (полный) геном организма. Как уже говорилось выше, в разные отрезки прошедшего столетия были на этот счет различные точки зрения, отчасти объясняемые уровнем технологических возможностей в конкретный момент истории. Видимо все же следует сейчас остановиться на том, что под геномом понимается последовательность всех азотистых оснований организма (где возможно, уточненная по-хромосомно) в его гаплоидном (для диплоидных организмов даже в некоем квази-гаплоидном состоянии), поскольку таковых полных уже очень много определено и внесено в базы данных. К тому же для прокариотических организмов иного (диплоидного) статуса и быть не может. Однако для понимания функционирования геномов высших организмов информация о них должна быть представлена в виде диплоидных геномов, получение которой пока сопряжено с большими трудностями. Но тот прогресс в секвенировании ДНК, достигнутый за последние полтора десятилетия, позволяет надеется, что и этот рубеж будет успешно преодолен в недалеком будущем. И тогда действительно только ВСЯ последовательность нуклеотидов в полном ДИПЛОИДНОМ геноме высших организмов будет считаться за ГЕНОМ. По крайней мере, к этому нужно стремиться. Применительно к человеку его геном будет считаться размером не 3 млрд. п.н. как сейчас, а 6 млрд. п.н. и это будет соответствовать

действительности, поскольку функциональное состояние генома зависит от нуклеотидных последовательностей в хромосомах, полученных как от отца, так и от матери; от матери, впрочем, человеку достается еще и митохондриальный геном, но его размер крайне мал и, к тому же он многокопийен.

Что касается информации о митохондриальном геноме, равно как и о пластидном (хондриом и пластом соответственно), то с учетом возможных случаев гетероплазмии, скорее всего, нужно допускать присутствие в базах данных их разнообразных вариантов (когда таковые есть) для каждого индивида или для вида / сорта растений.

Мы начали данную статью с хождения в народе ряда биологических терминов, не ранжировав их по популярности, так как для этого нужно было проводить социологический опрос, и в случае его проведения кем-то он теоретически может дать иной расклад, поставив нас впросак, поскольку это было бы наше абсолютно субъективное мнение, хотя и коллектива авторов. Однако можем взять на себя смелость представить, что подобный опрос будет проведен через три десятка лет и сейчас его не проверить.

Итак, через 30 лет можно ожидать следующие результаты. Безусловно, коронавирус будет побежден, хотя и может в виде рядовых сезонных инфекций себя периодически проявлять. К тому же будут созданы направленно действующие лекарства против разных вирусов, а не только против SARS-CoV-2. Поэтому термин «коронавирус» вряд ли будет интенсивно циркулировать в обществе, как и собственно его этиологический агент. Что касается термина «ГМО» (или каких-то новых его заместителей), то о нем (о них) никто и не вспомнит, поскольку практически все возделываемые культуры будут созданы генно-инженерным путем – ведь не задумываемся мы сейчас о том, что многие сорта сельскохозяйственных растений получены благодаря химическому или радиационному мутагенезу (то есть серьезному вмешательству в геном растения) с последующей их селекцией. Неактуально. Термин «ПЦР» может также исчезнуть, если будет найден дешевый способ идентификации единичных молекул ДНК или РНК, не прибегая к амплификации их отдельных участков. Впрочем, такая возможность есть и сейчас, но это пока очень трудоемкая работа, требующая весьма дорогостоящего оборудования, тогда как ПЦР в ее классическом простом варианте, в том числе проводимом в режиме реального времени, стала рутинной.

О будущем термина «геном» нужно поговорить отдельно. С учетом того, что человечество движется в сторону персонифицированной медицины, то для прогнозирования и лучшего лечения болезней необходимо знать функциональное состояние человека, которое во многом (помимо, конечно, средовых

факторов) определяется его ДИПЛОИДНЫМ геномом. Поэтому возможно, что через тридцать лет или даже раньше термин «геном» будет, что называется, «на языке» у каждого, по крайней мере, взрослого человека. Интересны прогнозы в вышедшей в конце октября 2020 г. статье в журнале Nature [Green et al., 2020]. Так, авторы привели, по их словам, 10 смелых предсказаний из области геномики человека, которые могут состояться к 2030 году. Пожалуй, стоит привести здесь часть этих прогнозов с нашими комментариями:

*1. Генерация и анализ полной последовательности генома человека будет обычным делом для любой исследовательской лаборатории, став таким же простым, как проведение очистки ДНК.*

Стоит отметить, что выделение и очистка ДНК могут быть очень разными, в том числе влияющими на сохранность очищенной ДНК и еще больше на размеры «обломков» этой гигантской молекулы, которые для определенных целей и задач могут быть критичными. Также и с секвенированием скорее всего будет дело обстоять, которое вряд ли будет совсем уж одинакового качества у всех в силу множества причин.

*2. Биологическая функция(и) каждого человеческого гена будет известна; для некодирующих элементов в геноме человека такое знание будет скорее правилом, чем исключением.*

Действительно вполне вероятно, и если не к 2030 г., то чуть позже.

*3. Общие особенности эпигенетического ландшафта и транскрипционных продуктов будут регулярно включаться в прогностические модели влияния генотипа на фенотип.*

Это можно ожидать только в случае секвенирования полных ДИПЛОИДНЫХ геномов людей, о чем в цитируемой статье абсолютно не говорится, хотя один раз применительно к секвенированию геномов при кратком описании результатов других авторов использовано слово “phased”. Не упоминаются ни гаплотипы, ни диплотипы. К тому же нужны знания происходящих модификаций азотистых оснований на отдельных этапах жизни человека, и это довольно непростая задача, поскольку нынешние технологии не позволяют определять это уверенно.

*5. Исследования, включающие анализ последовательностей генома и связанной с ними фенотипической информации, для миллионов людей-участников, будут регулярно демонстрироваться на школьных научных фестивалях.*

По крайней мере, для квази-гаплоидных геномов это вполне может быть.

*6. Регулярное использование геномной информации будет переходить от чего-то необычайного к обычной практике в клиниках, делая геномное*

*тестирование таким же рутинным, как и полный анализ крови.*

Опять-таки этого можно ожидать только в случае массового секвенирования полных ДИПЛОИДНЫХ геномов людей. Другое дело, анализ имеющих прямое отношение к тому или иному заболеванию конкретных генов, для которых и сейчас можно детектировать оба аллеля, поскольку нынешние технологии это позволяют, хотя по сложности анализ крови они пока превосходят, но в будущем действительно могут с ним сравняться.

*8. Полная последовательность генома человека вместе с информативными аннотациями при желании будет надежно и легко доступна на его смартфоне.*

Если судить по бурно развивающимся технологиям секвенирования полных квази-гаплоидных геномов и совершенствованию гаджетов, это вполне реально и даже ранее 2030 г., только зачем будет нужна эта информация, так как для персонализированной медицины требуются полные ДИПЛОИДНЫЕ геномы индивидов, да и диагноз ставить и рекомендации по лечению может давать только врач.

*10. Прорывные открытия приведут к передовым методам лечения, включающим геномные модификации для десятков генетических заболеваний.*

И этого можно ожидать только в случае массового секвенирования полных ДИПЛОИДНЫХ геномов людей.

Следует отметить, что авторы цитируемой статьи [Green et al., 2020] оговорились, что большинство из их предсказаний вряд ли будут в полной мере достигнуты к 2030 году, но достижение даже одного или нескольких из них приведет к стремлению к чему-то, что в настоящее время кажется недостижимым. И это правильный посыл. В этой статье также говорится, что данные предсказания были сделаны, чтобы в том числе, чтобы спровоцировать дискуссии о том, что будет возможно достижимо в будущем.

Наконец, нужно вспомнить и о термине «ДНК». С учетом того, что термин «геном» будет иметь широчайшее хождение в народе, а диплоидный геном человека представляет собой более 6 млрд. «буковок» из алфавита из четырех азотистых оснований в составе ДНК, то термин «ДНК», конечно же, забыт не будет. При этом хорошо бы, чтобы, по крайней мере, специалисты помнили о нуклеине и про того, кто его открыл. Знаем по собственному опыту и опыту своих коллег из других городов, что студенты биологических специальностей на вопрос «Кто открыл ДНК?», отвечают «Уотсон и Крик», тогда как они совершили ДРУГОЕ также крайне важное открытие, установив организацию ДНК в виде двойной спирали. Ф.Мишер же, возможно, сделал



самое фундаментальное открытие и, по всей видимости, не только в биологии, опередив свое время и прогресс науки больше, чем на столетие. Говорить нам так позволяет то, что считается, что любое крупное фундаментальное открытие рано или поздно дает выход в практику, но обычно этот латентный период занимает до двух-трех десятков лет до первых практических применений, или даже это происходит быстрее. В случае с нуклеином, открытым (напомним) в 1869 году, первые практические результаты были получены лишь в 1970-х и 1980-х годах, когда были созданы молекулы рекомбинантных ДНК и начата ДНК-диагностика. А сейчас без использования особенностей (полиморфизма) молекул ДНК биологическая жизнь человечества уже не представляется возможной.

Завершая статью о столетнем юбилее термина «геном», нужно признать, что новую жизнь и иную трактовку этот термин приобрел с начавшегося секвенирования ДНК, приведшего к определению нуклеотидных последовательностей всего генома, что дало новое понимание функционирования всего Живого и позволило восстановить, насколько это возможно, эволюционные процессы видообразования на Планете. Для этого, впрочем, вполне достаточно информации о гаплоидных геномах организмов. Для человека, для персонифицированной медицины будущего нужны сведения о диплоидных геномах. При этом гаплоидные геномы также вносят важный вклад в современную медицину, и их востребованность будет пока еще расти.

### Литература

1. Баймиев Ан.Х., Кулуев Б.Р., Вершинина З.Р., Князев А.В., Чемерис Д.А., Рожнова Н.А., Геращенко Г.А., Михайлова Е.В., Баймиев Ал.Х., Чемерис А.В. CRISPR/Cas редактирование геномов (растений) и общество // *Биомика*. 2017. Т.9. С.183-202.
2. Вершинина З.Р., Кулуев Б.Р., Максимов И.В., Михайлова Е.В., Гумерова Г.Р., Малеев Г.В., Князев А.В., Баймиев Ан.Х., Баймиев Ал.Х., Чемерис А.В. ГМО запретить невозможно разрешить! // *Биомика*. 2020. Т.12(1). С. 80-120. DOI: 10.31301/2221-6197.bmcs.2020-6
3. Гарафутдинов Р.Р., Чемерис А.В. «Российский след» в ранних исследованиях нуклеиновых кислот // *Биомика*. 2019. Т.11(3). С. 266-281. DOI: 10.31301/2221-6197.bmcs.2019-25
4. Геращенко Г.А., Гарафутдинов Р.Р., Баймиев Ан.Х., Кулуев Б.Р., Баймиев Ал.Х., Чемерис А.В. Два величайших открытия двух столетий - нуклеин и двойная спираль ДНК // *Биомика*. 2019. Т.11(3). С. 259-265. DOI: 10.31301/2221-6197.bmcs.2019-24
5. Геращенко Г.А., Рожнова Н.А., Кулуев Б.Р., Кирьянова О.Ю., Гумерова Г.Р., Князев А.В., Вершинина З.Р., Михайлова Е.В., Чемерис Д.А., Матниязов Р.Т., Баймиев Ан.Х., Губайдуллин И.М., Баймиев Ал.Х., Чемерис А.В. Дизайн РНК-гидов для CRISPR/CAS редактирования геномов растений // *Молекулярная биология*. 2020. Т.54(1). С. 29-50. DOI: 10.1134/S0026898420010061
6. 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
7. Кулуев А.Р., Матниязов Р.Т., Чемерис Д.А., Чемерис А.В. Современные представления о родственных взаимоотношениях в пшенично-эгилопном альянсе (с краткой исторической справкой) // *Биомика*. 2016. Т. 8. № 4. С. 297-310.
8. Кулуев Б.Р., Геращенко Г.А., Рожнова Н.А., Баймиев Ан.Х., Вершинина З.Р., Князев А.В., Матниязов Р.Т., Гумерова Г.Р., Михайлова Е.В., Никоноров Ю.М., Чемерис Д.А., Баймиев Ал.Х., Чемерис А.В. CRISPR/Cas редактирование геномов растений // *Биомика*. 2017. Т.9. С.155-182.
9. Кулуев Б.Р., Гумерова Г.Р., Михайлова Е.В., Геращенко Г.А., Рожнова Н.А., Вершинина З.Р., Князев А.В., Матниязов Р.Т., Баймиев Ан.Х., Баймиев Ал.Х., Чемерис А.В. Доставка CRISPR/CAS-компонентов в клетки высших растений для редактирования их геномов. // *Физиология растений*. 2019. Т.66(5). С.339-353. DOI: 10.1134/S0015330319050117
10. Светозарова В.В. О втором геноме *T. timopheevii* Zhuk. // Докл. АН СССР. 1939. Т. 23. № 5. С. 472-476.
11. Чемерис А.В., Бикбулатова С.М., Вахитов В.А. Об электронном журнале ИБГ УНЦ РАН «Биомика / Biomics» (редакторская статья) // *Биомика*. 2011. Т.1(1). С.1-7.
12. Чемерис А.В., Магданов Э.Г., Гарафутдинов Р.Р., Вахитов В.А. Как исключить появление ложнопозитивных результатов при проведении полимеразной цепной реакции // *Вестник биотехнологии и физико-химической биологии*. 2012. Т. 8. С. 34-45.
13. Чемерис А.В., Чемерис Д.А., Магданов Э.Г., Гарафутдинов Р.Р., Нагаев Н.Р., Вахитов В.А. Причины ложно-негативной ПЦР и недопущение некоторых из них // *Биомика*. 2012. Т. 4. № 1. С. 31-47.
14. Чемерис А.В., Бикбулатова С.М., Чемерис Д.А., Баймиев Ал.Х., Князев А.В., Кулуев Б.Р., Максимов И.В. Надо ли опасаться ГМО? Взгляд несторонних наблюдателей на истерию вокруг // *Биомика*. 2014. Т.6. С.77-138.
15. Чемерис А.В., Чемерис Д.А., Баймиев Ал.Х., Князев А.В., Кулуев Б.Р., Максимов И.В. Борьба с

- ГМО как неолысенковщина // *Биомика*. 2015. Т.7. С. 1-39.
16. Чемерис Д.А., Кирьянова О.Ю., Геращенко Г.А., Кулуев Б.Р., Рожнова Н.А., Матниязов Р.Т., Баймиев Ан.Х., Баймиев Ал.Х., Губайдуллин И.М., Чемерис А.В. Биоинформатические ресурсы для CRISPR/Cas редактирования геномов // *Биомика*. 2017. Т.9. С.203-228.
17. Чемерис Д.А., Гарафутдинов Р.Р., Сагитов А.М., Сагитова М.А., Михайленко К.И., Зубов В.В., Василов Р.Г., Сломинский П.А., Анисимов В.А., Хуснутдинова Э.К., Алексеев Я.И., Курочкин В.Е., Лавров Г.С., Воробьев А.А., Аминев Ф.Г., Чемерис А.В. Микродиплотипы как новые маркеры для ДНК-идентификации личности // *Биомика*. 2020. Т.12(2). С. 300-317. DOI: 10.31301/2221-6197.bmcscs.2020-17
18. 1000 Genomes Project Consortium, Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation // *Nature*. 2015. V.526(7571). P.68-74. doi: 10.1038/nature15393
19. Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group // *Genome Res*. 2009. V.19(9). P.1622-1629. doi: 10.1101/gr.092197.109
20. Baaijens J.A., Schönhuth A. Overlap graph-based generation of haplotigs for diploids and polyploids // *Bioinformatics*. 2019. V.35(21). P.4281-4289. doi: 10.1093/bioinformatics/btz255
21. Cao H., Wu H., Luo R., Huang S., Sun Y., Tong X., Xie Y., Liu B., Yang H., Zheng H., Li J., Li B., Wang Y., Yang F., Sun P., Liu S., Gao P., Huang H., Sun J., Chen D., He G., Huang W., Huang Z., Li Y., Tellier L.C., Liu X., Feng Q., Xu X., Zhang X., Bolund L., Krogh A., Kristiansen K., Drmanac R., Drmanac S., Nielsen R., Li S., Wang J., Yang H., Li Y., Wong G.K., Wang J. De novo assembly of a haplotype-resolved human genome // *Nat. Biotechnol*. 2015. V.33(6). P.617-622. doi: 10.1038/nbt.3200
22. Chin C.S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G.R., Delledonne M., Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing // *Nat. Methods*. 2016. V.13(12). P.1050-1054. doi: 10.1038/nmeth.4035
23. Deininger P.L., Schmid C.W. An electron microscope study of the DNA sequence organization of the human genome // *J. Mol. Biol*. 1976. V.106(3). P.773-790. doi: 10.1016/0022-2836(76)90264-3
24. English A.C., Salerno W.J., Hampton O.A., Gonzaga-Jauregui C., Ambreth S., Ritter D.I., Beck C.R., Davis C.F., Dahdouli M., Ma S., Carroll A., Veeraraghavan N., Bruestle J., Drees B., Hastie A., Lam E.T., White S., Mishra P., Wang M., Han Y., Zhang F., Stankiewicz P., Wheeler D.A., Reid J.G., Muzny D.M., Rogers J., Sabo A., Worley K.C., Lupski J.R., Boerwinkle E., Gibbs R.A. Assessing structural variation in a personal genome-towards a human reference diploid genome // *BMC Genomics*. 2015. V.16(1). P.286. doi: 10.1186/s12864-015-1479-3
25. Gelderman A.H., Rake A.V., Britten R.J. Transcription of nonrepeated DNA in neonatal and fetal mice // *Proc. Natl. Acad. Sci. USA*. 1971. V.68(1). P.172-176. doi: 10.1073/pnas.68.1.172
26. Green E.D., Gunter C., Biesecker L.G., Di Francesco V., Easter C.L., Feingold E.A., Felsenfeld A.L., Kaufman D.J., Ostrander E.A., Pavan W.J., Phillippy A.M., Wise A.L., Dayal J.G., Kish B.J., Mandich A., Wellington C.R., Wetterstrand K.A., Bates S.A., Leja D., Vasquez S., Gahl W.A., Graham B.J., Kastner D.L., Liu P., Rodriguez L.L., Solomon B.D., Bonham V.L., Brody L.C., Hutter C.M., Manolio T.A. Strategic vision for improving human health at The Forefront of Genomics // *Nature*. 2020. V.586(7831). P.683-692. doi: 10.1038/s41586-020-2817-4
27. Guo Y., Dai Y., Yu H., Zhao S., Samuels D.C., Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis // *Genomics*. 2017. V.109(2). P.83-90. doi: 10.1016/j.ygeno.2017.01.005
28. Gupta P.K., Mir R.R., Mohan A., Kumar J. Wheat genomics: present status and future prospects // *Int. J. Plant Genomics*. 2008. 2008:896451. doi: 10.1155/2008/896451
29. Hood L.E., Hunkapiller M.W., Smith L.M. Automated DNA sequencing and analysis of the human genome // *Genomics*. 1987. V.1(3). P.201-212. doi: 10.1016/0888-7543(87)90046-2
30. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome // *Nature*. 2001. V.409(6822). P.860-921. doi: 10.1038/35057062
31. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome // *Nature*. 2004. V.431(7011). P.931-945. doi: 10.1038/nature03001
32. Istrail S., Sutton G.G., Florea L., Halpern A.L., Mobarry C.M., Lippert R., Walenz B., Shatkay H., Dew I., Miller J.R., Flanigan M.J., Edwards N.J., Bolanos R., Fasulo D., Halldorsson B.V., Hannenhalli S., Turner R., Yooseph S., Lu F., Nusskern D.R., Shue B.C., Zheng X.H., Zhong F., Delcher A.L., Huson D.H., Kravitz S.A., Mouchard L., Reinert K., Remington, K.A. Clark A.G., Waterman M.S., Eichler E.E., Adams M.D., Hunkapiller M.W., Myers E.W., Venter J.C. Whole-

- genome shotgun assembly and comparison of human genome assemblies // *Proc. Natl. Acad. Sci. USA*. 2004. V.101(7). P.1916-1921. doi: 10.1073/pnas.0307971100
33. Kajitani R., Yoshimura D., Okuno M., Minakuchi Y., Kagoshima H., Fujiyama A., Kubokawa K., Kohara Y., Toyoda A., Itoh T. Platanus-alley is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions // *Nat. Commun.* 2019. V.10(1). P.1702. doi: 10.1038/s41467-019-09575-2
34. Kedes L., Liu E.T. The Archon Genomics X PRIZE for whole human genome sequencing // *Nat. Genet.* 2010. V.42(11). P.917-918. doi: 10.1038/ng1110-917
35. Kitamura Y., Moriguchi M., Kaneko H., Morisaki H., Morisaki T., Toyama K., Kamatani N. Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm. *Ann Hum Genet.* 2002. V.66(3). P.183-193. doi: 10.1017/S0003480002001124
36. Kitzman J.O., Mackenzie A.P., Adey A., Hiatt J.B., Patwardhan R.P., Sudmant P.H., Ng S.B., Alkan C., Qiu R., Eichler E.E., Shendure J. Haplotype-resolved genome sequencing of a Gujarati Indian individual // *Nat. Biotechnol.* 2011. V.29(1). P.59-63. doi: 10.1038/nbt.1740
37. Koren S., Rhie A., Walenz B.P., Diltthey A.T., Bickhart D.M., Kingan S.B., Hiendleder S., Williams J.L., Smith T.P.L., Phillippy A.M. De novo assembly of haplotype-resolved genomes with trio binning // *Nat. Biotechnol.* 2018. doi: 10.1038/nbt.4277.
38. Kuska B. Beer, Bethesda, and Biology: How "Genomics" Came Into Being // *J. Nat. Cancer Inst.* 1998. V.90(2). P.93. doi: 10.1093/jnci/90.2.93
39. Lederberg J., McCray A.T. 'Ome Sweet 'Omics - A Genealogical Treasury of Words // *Scientist*. 2001. V.15(7). P.8.
40. Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.H., Frazier M.E., Scherer S.W., Strausberg R.L., Venter J.C. The diploid genome sequence of an individual human // *PLoS Biol.* 2007. V.5(10). e254. doi: 10.1371/journal.pbio.0050254
41. Lieberman M.W., Poirier M.C. Distribution of deoxyribonucleic acid repair synthesis among repetitive and unique sequences in the human diploid genome // *Biochemistry*. 1974. V.13(15). P.3018-3023. doi: 10.1021/bi00712a003
42. MacLean C.J., Morton N.E. Estimation of myriad haplotype frequencies. // *Genet Epidemiol.* 1985. V.2(3). P.263-272. doi: 10.1002/gepi.1370020304
43. McKusick V.A., Ruddle F.H. A new discipline, a new name, a new journal // *Genomics*, 1987, V.1, P.1-2. doi: 10.1016/0888-7543(87)90098-X
44. Menzel M.Y. A Cytological Method for Genome Analysis in *Gossypium* // *Genetics*. 1955. V.40(2). P.214-223.
45. Minio A., Lin J., Gaut B.S., Cantu D. How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes // *Front Plant Sci.* 2017. V.8:826. doi: 10.3389/fpls.2017.00826
46. Minio A., Massonnet M., Figueroa-Balderas R., Castro A., Cantu D. Diploid Genome Assembly of the Wine Grape Carmenere // *G3 (Bethesda)*. 2019. V.9(5). P.1331-1337. doi: 10.1534/g3.119.400030
47. Mocci E., Debeljak M., Klein A.P., Eshleman J.R. A New Fast Phasing Method Based On Haplotype Subtraction // *J. Mol. Diagn.* 2019. V.21(3). P.427-436. doi: 10.1016/j.jmoldx.2018.12.004
48. Morton N.E. Parameters of the human genome // *Proc. Natl. Acad. Sci. USA*. 1991. V. 88. P. 7474-7476.
49. Michaelis P. Cytoplasmic inheritance in *Epilobium* and its theoretical significance // *Adv. Genet.* 1954. V.6. P.287-401.
50. Noguera-Solano R., Ruiz-Gutierrez R., Rodriguez-Caso J.M. Genome: twisting stories with DNA // *Endeavour*. 2013. V.37(4). P.213-219. doi: 10.1016/j.endeavour.2013.05.003
51. Noguera-Solano R., Ruiz-Gutierrez R., Rodriguez-Caso J.M. (2017) The Genomization of Biology: Counterbalancing Radical Reductionism. In: Petermann H., Harper P., Doetz S. (eds) *History of Human Genetics*. Springer, Cham. [http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-51783-4\\_8](http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-51783-4_8)
52. Palca J. Human genome: Department of Energy on the map // *Nature*. 1986. V.321. P.371. doi: 10.1038/321371a0
53. Pang A.W., MacDonald J.R., Pinto D., Wei J., Rafiq M.A., Conrad D.F., Park H., Hurles M.E., Lee C., Venter J.C., Kirkness E.F., Levy S., Feuk L., Scherer S.W. Towards a comprehensive structural variation map of an individual human genome // *Genome Biol.* 2010. V.11(5). P.R52. doi: 10.1186/gb-2010-11-5-r52
54. Pendleton M., Sebra R., Pang A.W., Ummat A., Franzen O., Rausch T., Stütz A.M., Stedman W., Anantharaman T., Hastie A., Dai H., Fritz M.H., Cao H., Cohain A., Deikus G., Durrett R.E., Blanchard S.C., Altman R., Chin C.S., Guo Y., Paxinos E.E., Korbel J.O., Darnell R.B., McCombie W.R., Kwok P.Y., Mason C.E., Schadt E.E., Bashir A. Assembly and diploid architecture of an individual human genome via single-molecule technologies // *Nat. Methods*. 2015. V.12(8). P.780-786. doi: 10.1038/nmeth.3454
55. Pennisi E. Human Genome Project. And the gene number is...? // *Science*. 2000. V.288(5469). P.1146-1147. doi: 10.1126/science.288.5469.1146

56. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome // *Nat. Biotechnol.* 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
57. Rhie A., Walenz B.P., Koren S., Phillippy A.M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies // *Genome Biol.* 2020. V.21(1). P.245. doi: 10.1186/s13059-020-02134-9
58. Riggio R.R., Saal S.D., Katz E.B., Tapia L., White R., Chami J., Sheigh J.S., Sullivan J.F., Stenzel K.H., Stubenbord W.T., Whitsell J.C., Rubin A.L. Improved renal allograft survival using the mixed lymphocyte culture for selection of nonidentical living related donors. // *Trans Am Soc Artif Intern Organs.* 1975. V.21. P.90-95.
59. Roach M.J., Schmidt S.A., Borneman A.R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies // *BMC Bioinformatics.* 2018. V.19(1). P.460. doi: 10.1186/s12859-018-2485-7
60. Robertson M. The proper study of mankind // *Nature.* 1986. V.322. P.11.
61. Sakhabutdinova A.R., Chemeris A.V., Garafutdinov R.R. Enhancement of PCR efficiency using mono- and disaccharides // *Anal. Biochem.* 2020. V.606:113858. doi: 10.1016/j.ab.2020.113858
62. Satas G., Raphael B.J. Haplotype phasing in single-cell DNA-sequencing data // *Bioinformatics.* 2018. V.34(13). P.i211-i217. doi: 10.1093/bioinformatics/bty286
63. Saunders G.F., Shirakawa S., Saunders P.P., Arrighi F.E., Hsu T.C. Populations of repeated DNA sequences in the human genome // *J. Mol. Biol.* 1972. V.63(3). P.323-334. doi: 10.1016/0022-2836(72)90430-5
64. Schmid C.W., Deininger P.L. Sequence organization of the human genome // *Cell.* 1975. V.6(3). P.345-358. doi: 10.1016/0092-8674(75)90184-1
65. Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, Kuk J, Park GH, Kim J, Ryu H, Kim J, Roh M, Baek J, Hunkapiller MW, Korlach J, Shin JY, Kim C. De novo assembly and phasing of a Korean human genome // *Nature.* 2016. V.538(7624). P.243-247. doi: 10.1038/nature20098
66. Shi D., Wu J., Tang H., Yin H., Wang H., Wang R., Wang R., Qian M., Wu J., Qi K., Xie Z., Wang Z., Zhao X., Zhang S. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants // *Genome Res.* 2019. V.29(11). P.1889-1899. doi: 10.1101/gr.251033.119
67. Smith H.H. Effects of Genome Balance, Polyploidy, and Single Extra Chromosomes on Size in *Nicotiana*. *Genetics.* 1943. V.28(3). P.227-236.
68. Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B., Hood L.E. Fluorescence detection in automated DNA sequence analysis // *Nature.* 1986. V.321(6071). P.674-679. doi: 10.1038/321674a0
69. Soifer L., Fong N.L., Yi N., Ireland A.T., Lam I., Sooknah M., Paw J.S., Peluso P., Concepcion G.T., Rank D., Hastie A.R., Jovic V., Ruby J.G., Botstein D., Roy M.A. Fully Phased Sequence of a Diploid Human Genome Determined de Novo from the DNA of a Single Individual // *G3 (Bethesda).* 2020. V.10(9). P.2911-2925. doi: 10.1534/g3.119.400995
70. Spuhler J.N. On the Number of Genes in Man // *Science.* 1948. V.108(2802). P.279-280. doi: 10.1126/science.108.2802.279-a
71. Staden R. A new computer method for the storage and manipulation of DNA gel reading data // *Nucleic Acids Res.* 1980. V.8(16). P.3673-94. doi: 10.1093/nar/8.16.3673
72. Stebbins G.L., Pun F.T. Artificial and Natural Hybrids in the Gramineae, Tribe Hordeae. VI. Chromosome Pairing in *Secale cereale* x *Agropyron* Intermedium and the Problem of Genome Homologies in the Triticinae // *Genetics.* 1953. V.38(6). P.600-608.
73. Stein N., Feuillet C., Wicker T., Schlegelhauf E., Keller B. Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the Lr10 resistance locus in hexaploid wheat (*Triticum aestivum* L.) // *Proc. Natl. Acad. Sci. USA.* 2000. V.97(24). P.13436-13441. doi: 10.1016/S0065-2660(08)60132-7
74. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkak MK, Malhotra A, Stütz AM, Shi X, Casale FP, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalina AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA; 1000 Genomes Project Consortium, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korb J.O. An integrated map of structural variation in 2,504 human genomes // *Nature.* 2015. V.526(7571). P.75-81. doi: 10.1038/nature15394
75. Sun X., Jiao C., Schwaninger H., Chao C.T., Ma Y., Duan N., Khan A., Ban S., Xu K., Cheng L., Zhong G.-Y., Fei Z. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication // *Nat Genet.* 2020. doi: 10.1038/s41588-020-00723-9
76. Venter J.C. Multiple personal genomes await // *Nature.* 2010. V.464(7289). P.676-677. doi: 10.1038/464676a
77. Venter J.C., Adams M.D., Myers E.W., Li PW,

- Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferreira S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome // *Science*. 2001. V.291(5507). P.1304-1351. doi: 10.1126/science.1058040
78. Walsh J.B., Marks J. Sequencing the human genome // *Nature*. 1986. V.322. P. 590.
79. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Li D, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang H, Wang J. The diploid genome sequence of an Asian individual // *Nature*. 2008. V.456(7218). P.60-65. doi: 10.1038/nature07484
80. Wang O., Chin R., Cheng X., Wu M.K.Y., Mao Q., Tang J., Sun Y., Anderson E., Lam H.K., Chen D., Zhou Y., Wang L., Fan F., Zou Y., Xie Y., Zhang R.Y., Drmanac S., Nguyen D., Xu C., Villarosa C., Gablenz S., Barua N., Nguyen S., Tian W., Liu J.S., Wang J., Liu X., Qi X., Chen A., Wang H., Dong Y., Zhang W., Alexeev A., Yang H., Wang J., Kristiansen K., Xu X., Drmanac R., Peters B.A. Efficient and unique cobarcode of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly // *Genome Res*. 2019. V.29(5). P.798-808. doi: 10.1101/gr.245126.118
81. Weisenfeld N.I., Kumar V., Shah P., Church D.M., Jaffe D.B. Direct determination of diploid genome sequences // *Genome Res*. 2017. V.27(5). P.757-767. doi: 10.1101/gr.214874.116
82. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.-J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X-z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. The complete genome of an individual by massively parallel DNA sequencing // *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
83. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. 1920. Jena: Gustav Fischer Verlag. 250 s.
84. Zhang L., Zhou X., Weng Z., Sidow A. Assessment of human diploid genome assembly with 10x Linked-Reads data // *Gigascience*. 2019. V.8(11). giz141. doi: 10.1093/gigascience/giz141
85. Zhang W., Luo C., Scossa F., Zhang Q., Usadel B., Fernie A.R., Mei H., Wen W. A phased genome based on single sperm sequencing reveals crossover pattern and complex relatedness in tea plants // *Plant J*. 2020. doi: 10.1111/tpj.15051
86. Zhou B., Arthur J.G., Ho S.S., Pattni R., Huang

Y., Wong W.H., Urban A.E. Extensive and deep sequencing of the Venter/HuRef genome for developing and benchmarking genome analysis tools // *Sci. Data*. 2018. V.5:180261. doi: 10.1038/sdata.2018.261

87. Zhou Q., Tang D., Huang W., Yang Z., Zhang Y., Hamilton J.P., Visser R.G.F., Bachem C.W.B., Robin Buell C., Zhang Z., Zhang C., Huang S. Haplotype-resolved genome analyses of a heterozygous diploid potato // *Nat. Genet.* 2020. V.52(10). P.1018-1023. doi: 10.1038/s41588-020-0699-x

88. Zimin A.V., Puiu D., Hall R., Kingan S., Clavijo B.J., Salzberg S.L. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum* // *Gigascience*. 2017. V.6(11). P.1-7. doi: 10.1093/gigascience/gix097

### References

- 1000 Genomes Project Consortium, Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature*. 2015. V.526(7571). P.68-74. doi: 10.1038/nature15393
- Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res*. 2009. V.19(9). P.1622-1629. doi: 10.1101/gr.092197.109
- Baaijens J.A., Schönhuth A. Overlap graph-based generation of haplotigs for diploids and polyploids. *Bioinformatics*. 2019. V.35(21). P.4281-4289. doi: 10.1093/bioinformatics/btz255
- Baymiev An.Kh., Kuluev B.R., Vershinina Z.R., Knyazev A.V., Chemeris D.A., Rozhnova N.A., Gerashchenkov G.A., Mikhailova E.V., Baymiev Al.Kh., Chemeris A.V. CRISPR/Cas genome editing (plants) and society. *Biomics*. 2017. V.9(3). P. 183-202. (In Russian)
- Cao H., Wu H., Luo R., Huang S., Sun Y., Tong X., Xie Y., Liu B., Yang H., Zheng H., Li J., Li B., Wang Y., Yang F., Sun P., Liu S., Gao P., Huang H., Sun J., Chen D., He G., Huang W., Huang Z., Li Y., Tellier L.C., Liu X., Feng Q., Xu X., Zhang X., Bolund L., Krogh A., Kristiansen K., Drmanac R., Drmanac S., Nielsen R., Li S., Wang J., Yang H., Li Y., Wong G.K., Wang J. De novo assembly of a haplotype-resolved human genome. *Nat. Biotechnol.* 2015. V.33(6). P.617-622. doi: 10.1038/nbt.3200
- Chemeris A.V., Bikbulatova S.M., Chemeris D.A., Baymiev Al.K., Knyazev A.V., Kuluev B.R., Maksimov I.V. Should beware of the GMOs? On-site observers view on the hysteria around. *Biomics*. 2014. V.6(2). P.77-138. (In Russian)
- Chemeris A.V., Bikbulatova S.M., Vakhitov V.A. About electronic journal of IBG USC RAS «Biomics» (Editorial paper). *Biomics*. 2011. V.1(1). P.1-7. (In Russian)
- Chemeris A.V., Chemeris D.A., Baymiev Al.K., Knyazev A.V., Kuluev B.R., Maksimov I.V. The fight against GMO is neolysenkoism. *Biomics*. 2015. V.7(1). P.1-39. (In Russian)
- Chemeris A.V., Chemeris D.A., Magdanov E.G., Garafutdinov R.R., Nagaev N.R., Vakhitov V.A. Causes of false-negative PCR and how to avoid some of them. *Biomics*. 2012. V.4. P.31-47. (In Russian)
- Chemeris A.V., Magdanov E.G., Garafutdinov R.R., Vakhitov V.A. How to avoid the appearance of false-positive results in a polymerase chain reaction? *Yu.A. Ovchinnikov Bulletin of Biotechnology and Physical and Chemical Biology*. 2012. V. 8(3). P. 34-45. (In Russian)
- Chemeris D.A., Kiryanova O.Yu., Gerashchenkov G.A., Kuluev B.R., Rozhnova N.A., Matniyazov R.T., Baymiev An.Kh., Baymiev Al.Kh., Gubaidullin I.M., Chemeris A.V. Bioinformatic resources for CRISPR/Cas genome editing. *Biomics*. 2017. V.9(3). P. 203-228. (In Russian)
- Chemeris D.A., Garafutdinov R.R., Sagitov A.M., Sagitova M.A., Mikhailenko K.I., Zubov V.V., Vasilov R.G., Slominsky P.A., Anisimov V.A., Khusnutdinova E.K., Alekseev Ya.I., Kurochkin V.A., Lavrov G.S., Vorobev A.A., Aminev F.G., Chemeris A.V. Microdiplotypes as a new markers for DNA identification. *Biomics*. 2020. V. 12(2). P. 300-317. DOI: 10.31301/2221-6197.bmcs.2020-17 (In Russian)
- Chin C.S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G.R., Delledonne M., Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods*. 2016. V.13(12). P.1050-1054. doi: 10.1038/nmeth.4035
- Deininger P.L., Schmid C.W. An electron microscope study of the DNA sequence organization of the human genome. *J. Mol. Biol.* 1976. V.106(3). P.773-790. doi: 10.1016/0022-2836(76)90264-3
- English A.C., Salerno W.J., Hampton O.A., Gonzaga-Jauregui C., Ambreth S., Ritter D.I., Beck C.R., Davis C.F., Dahdouli M., Ma S., Carroll A., Veeraraghavan N., Bruestle J., Drees B., Hastie A., Lam E.T., White S., Mishra P., Wang M., Han Y., Zhang F., Stankiewicz P., Wheeler D.A., Reid J.G., Muzny D.M., Rogers J., Sabo A., Worley K.C., Lupski J.R., Boerwinkle E., Gibbs R.A. Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genomics*. 2015. V.16(1). P.286. doi: 10.1186/s12864-015-1479-3
- Garafutdinov R.R., Chemeris A.V. "Russian



- traces" in early nucleic acids research. *Biomics*. 2019. V.11(3). P. 266-281. DOI: 10.31301/2221-6197.bmcs.2019-25 (In Russian)
17. Gelderman A.H., Rake A.V., Britten R.J. Transcription of nonrepeated DNA in neonatal and fetal mice. *Proc. Natl. Acad. Sci. USA*. 1971. V.68(1). P.172-176. doi: 10.1073/pnas.68.1.172
18. Gerashchenkov G.A., Garafutdinov R.R., Baymiev An.Kh., Kuluev B.R., Baymiev Al.Kh., Chemeris A.V. The two greatest discoveries of two centuries - the nuclein and the double helix of DNA. *Biomics*. 2019. V.11(3). P. 259-265. DOI: 10.31301/2221-6197.bmcs.2019-24 (In Russian)
19. Gerashchenkov G.A., Rozhnova N.A., Kuluev B.R., Kiryanova O.Yu., Gumerova G.R., Knyazev A.V., Vershinina Z.R., Mikhailova E.V., Chemeris D.A., Matniyazov R.T., Baimiev An.Kh., Gubaidullin I.M., Baimiev Al.Kh., Chemeris A.V. Design of guide RNA for CRISPR/Cas plant genome editing. *Molecular Biology*. 2020. V.54(1). P.29-50. DOI: 10.1134/S0026893320010069
20. Green E.D., Gunter C., Biesecker L.G., Di Francesco V., Easter C.L., Feingold E.A., Felsenfeld A.L., Kaufman D.J., Ostrander E.A., Pavan W.J., Phillippy A.M., Wise A.L., Dayal J.G., Kish B.J., Mandich A., Wellington C.R., Wetterstrand K.A., Bates S.A., Leja D., Vasquez S., Gahl W.A., Graham B.J., Kastner D.L., Liu P., Rodriguez L.L., Solomon B.D., Bonham V.L., Brody L.C., Hutter C.M., Manolio T.A. Strategic vision for improving human health at The Forefront of Genomics. *Nature*. 2020. V.586(7831). P.683-692. doi: 10.1038/s41586-020-2817-4
21. Guo Y., Dai Y., Yu H., Zhao S., Samuels D.C., Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics*. 2017. V.109(2). P.83-90. doi: 10.1016/j.ygeno.2017.01.005
22. Gupta P.K., Mir R.R., Mohan A., Kumar J. Wheat genomics: present status and future prospects. *Int. J. Plant Genomics*. 2008. 2008:896451. doi: 10.1155/2008/896451
23. Hood L.E., Hunkapiller M.W., Smith L.M. Automated DNA sequencing and analysis of the human genome. *Genomics*. 1987. V.1(3). P.201-212. doi: 10.1016/0888-7543(87)90046-2
24. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001. V.409(6822). P.860-921. doi: 10.1038/35057062
25. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004. V.431(7011). P.931-945. doi: 10.1038/nature03001
26. Istrail S., Sutton G.G., Florea L., Halpern A.L., Mobarry C.M., Lippert R., Walenz B., Shatkay H., Dew I., Miller J.R., Flanigan M.J., Edwards N.J., Bolanos R., Fasulo D., Halldorsson B.V., Hannerhalli S., Turner R., Yooseph S., Lu F., Nusskern D.R., Shue B.C., Zheng X.H., Zhong F., Delcher A.L., Huson D.H., Kravitz S.A., Mouchard L., Reinert K., Remington, K.A. Clark A.G., Waterman M.S., Eichler E.E., Adams M.D., Hunkapiller M.W., Myers E.W., Venter J.C. Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl. Acad. Sci. USA*. 2004. V.101(7). P.1916-1921. doi: 10.1073/pnas.0307971100
27. Kajitani R., Yoshimura D., Okuno M., Minakuchi Y., Kagoshima H., Fujiyama A., Kubokawa K., Kohara Y., Toyoda A., Itoh T. Platanus-alley is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat. Commun*. 2019. V.10. P.1702. doi: 10.1038/s41467-019-09575-2
28. Kedes L., Liu E.T. The Archon Genomics X PRIZE for whole human genome sequencing. *Nat. Genet*. 2010. V.42(11). P.917-918. doi: 10.1038/ng1110-917
29. 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)
30. Kitamura Y., Moriguchi M., Kaneko H., Morisaki H., Morisaki T., Toyama K., Kamatani N. Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm // *Ann Hum Genet*. 2002. V.66(3). P.183-193. doi: 10.1017/S0003480002001124
31. Kitzman J.O., Mackenzie A.P., Adey A., Hiatt J.B., Patwardhan R.P., Sudmant P.H., Ng S.B., Alkan C., Qiu R., Eichler E.E., Shendure J. Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nat. Biotechnol*. 2011. V.29. P.59-63. doi: 10.1038/nbt.1740
32. Koren S., Rhie A., Walenz B.P., Dilthey A.T., Bickhart D.M., Kingan S.B., Hiendleder S., Williams J.L., Smith T.P.L., Phillippy A.M. De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol*. 2018. doi: 10.1038/nbt.4277.
33. 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.
34. Kuluev B.R., Gerashchenkov G.A., Rozhnova N.A., Baymiev An.Kh., Vershinina Z.R., Knyazev A.V., Matniyazov R.T., Gumerova G.R., Mikhailova E.V., Nikonorov Yu.M., Chemeris D.A., Baymiev Al.Kh., Chemeris A.V. 2017. CRISPR/Cas editing of plant genomes. *Biomics*. 2017. V.9(3). P. 155-182. (In Russian)
35. Kuluev B.R., Gumerova G.R., Mikhaylova E.V.,

- Gerashchenkov G.A., Rozhnova N.A., Vershinina Z.R., Knyazev A.V., Matniyazov R.T., Baymiev An.Kh., Baymiev Al.Kh., Chemeris A.V. Delivery of CRISPR/Cas components into higher plant cells for genome editing. *Russ. J. Plant Physiol.* 2019. V.66(5). P. 694–706. doi 10.1134/S0015330319050117
36. Kuska B. Beer, Bethesda, and Biology: How “Genomics” Came Into Being. *J. Nat. Cancer Inst.* 1998. V.90(2). P.93. doi: 10.1093/jnci/90.2.93
37. Lederberg J., McCray A.T. 'Ome Sweet 'Omics - A Genealogical Treasury of Words. *Scientist.* 2001. V.15(7). P.8.
38. Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.H., Frazier M.E., Scherer S.W., Strausberg R.L., Venter J.C. The diploid genome sequence of an individual human. *PLoS Biol.* 2007. V.5(10). e254. doi: 10.1371/journal.pbio.0050254
39. Lieberman M.W., Poirier M.C. Distribution of deoxyribonucleic acid repair synthesis among repetitive and unique sequences in the human diploid genome. *Biochemistry.* 1974. V.13(15). P.3018-3023. doi: 10.1021/bi00712a003
40. MacLean C.J., Morton N.E. Estimation of myriad haplotype frequencies. // *Genet Epidemiol.* 1985. V.2(3). P.263-272. doi: 10.1002/gepi.1370020304
41. McKusick V.A., Ruddle F.H. A new discipline, a new name, a new journal. *Genomics.* 1987. V.1. P.1-2. doi: 10.1016/0888-7543(87)90098-X
42. Menzel M.Y. A Cytological Method for Genome Analysis in *Gossypium*. *Genetics.* 1955. V.40. P.214-223.
43. Minio A., Lin J., Gaut B.S., Cantu D. How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. *Front Plant Sci.* 2017. V.8:826. doi: 10.3389/fpls.2017.00826
44. Minio A., Massonnet M., Figueroa-Balderas R., Castro A., Cantu D. Diploid Genome Assembly of the Wine Grape Carmenere // G3 (Bethesda). 2019. V.9(5). P.1331-1337. doi: 10.1534/g3.119.400030
45. Mocchi E., Debeljak M., Klein A.P., Eshleman J.R. A New Fast Phasing Method Based On Haplotype Subtraction. *J. Mol. Diagn.* 2019. V.21(3). P.427-436. doi: 10.1016/j.jmoldx.2018.12.004
46. Morton N.E. Parameters of the human genome. *Proc. Natl. Acad. Sci. USA.* 1991. V. 88. P. 7474-7476.
47. Michaelis P. Cytoplasmic inheritance in *Epilobium* and its theoretical significance. *Adv. Genet.* 1954. V.6. P.287-401.
48. Noguera-Solano R., Ruiz-Gutierrez R., Rodriguez-Caso J.M. Genome: twisting stories with DNA. *Endeavour.* 2013. V.37(4). P.213-219. doi: 10.1016/j.endeavour.2013.05.003
49. Noguera-Solano R., Ruiz-Gutierrez R., Rodriguez-Caso J.M. (2017) The Genomization of Biology: Counterbalancing Radical Reductionism. In: Petermann H., Harper P., Doetz S. (eds) History of Human Genetics. Springer, Cham. [http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-51783-4\\_8](http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-51783-4_8)
50. Palca J. Human genome: Department of Energy on the map. *Nature.* 1986. V.321. P.371. doi: 10.1038/321371a0
51. Pang A.W., MacDonald J.R., Pinto D., Wei J., Rafiq M.A., Conrad D.F., Park H., Hurles M.E., Lee C., Venter J.C., Kirkness E.F., Levy S., Feuk L., Scherer S.W. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* 2010. V.11(5). P.R52. doi: 10.1186/gb-2010-11-5-r52
52. Pendleton M., Sebra R., Pang A.W., Ummat A., Franzen O., Rausch T., Stütz A.M., Stedman W., Anantharaman T., Hastie A., Dai H., Fritz M.H., Cao H., Cohain A., Deikus G., Durrett R.E., Blanchard S.C., Altman R., Chin C.S., Guo Y., Paxinos E.E., Korbel J.O., Darnell R.B., McCombie W.R., Kwok P.Y., Mason C.E., Schadt E.E., Bashir A. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods.* 2015. V.12(8). P.780-786. doi: 10.1038/nmeth.3454
53. Pennisi E. Human Genome Project. And the gene number is...? *Science.* 2000. V.288(5469). P.1146-1147. doi: 10.1126/science.288.5469.1146
54. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
55. Rhie A., Walenz B.P., Koren S., Phillippy A.M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020. V.21(1). P.245. doi: 10.1186/s13059-020-02134-9
56. Riggio R.R., Saal S.D., Katz E.B., Tapia L., White R., Chami J., Sheigh J.S., Sullivan J.F., Stenzel K.H., Stubenbord W.T., Whitsell J.C., Rubin A.L. Improved renal allograft survival using the mixed lymphocyte culture for selection of nonidentical living related donors // *Trans Am Soc Artif Intern Organs.* 1975. V.21. P.90-95.
57. Roach M.J., Schmidt S.A., Borneman A.R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics.* 2018. V.19(1). P.460. doi: 10.1186/s12859-018-2485-7
58. Robertson M. The proper study of mankind. *Nature.* 1986. V.322. P.11.
59. Sakhabutdinova A.R., Chemeris A.V., Garafutdinov R.R. Enhancement of PCR efficiency using mono- and disaccharides. *Anal. Biochem.* 2020. V.606:113858. doi: 10.1016/j.ab.2020.113858

60. Satas G., Raphael B.J. Haplotype phasing in single-cell DNA-sequencing data. *Bioinformatics*. 2018. V.34. P.i211-i217. doi: 10.1093/bioinformatics/bty286
61. Saunders G.F., Shirakawa S., Saunders P.P., Arrighi F.E., Hsu T.C. Populations of repeated DNA sequences in the human genome. *J. Mol. Biol.* 1972. V.63(3). P.323-334. doi: 10.1016/0022-2836(72)90430-5
62. Schmid C.W., Deininger P.L. Sequence organization of the human genome. *Cell*. 1975. V.6(3). P.345-358. doi: 10.1016/0092-8674(75)90184-1
63. Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, Kuk J, Park GH, Kim J, Ryu H, Kim J, Roh M, Baek J, Hunkapiller MW, Korlach J, Shin JY, Kim C. De novo assembly and phasing of a Korean human genome. *Nature*. 2016. V.538(7624). P.243-247. doi: 10.1038/nature20098
64. Shi D., Wu J., Tang H., Yin H., Wang H., Wang R., Wang R., Qian M., Wu J., Qi K., Xie Z., Wang Z., Zhao X., Zhang S. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res*. 2019. V.29(11). P.1889-1899. doi: 10.1101/gr.251033.119
65. Smith H.H. Effects of Genome Balance, Polyploidy, and Single Extra Chromosomes on Size in *Nicotiana*. *Genetics*. 1943. V.28(3). P.227-236.
66. Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B., Hood L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986. V.321(6071). P.674-679. doi: 10.1038/321674a0
67. Soifer L., Fong N.L., Yi N., Ireland A.T., Lam I., Sooknah M., Paw J.S., Peluso P., Concepcion G.T., Rank D., Hastie A.R., Jojic V., Ruby J.G., Botstein D., Roy M.A. Fully Phased Sequence of a Diploid Human Genome Determined de Novo from the DNA of a Single Individual. *G3 (Bethesda)*. 2020. V.10(9). P.2911-2925. doi: 10.1534/g3.119.400995
68. Spuhler J.N. On the Number of Genes in Man // *Science*. 1948. V.108(2802). P.279-280. doi: 10.1126/science.108.2802.279-a
69. Staden R. A new computer method for the storage and manipulation of DNA gel reading data. *Nucl. Acids Res*. 1980. V.8. P.3673-94. doi: 10.1093/nar/8.16.3673
70. Stebbins G.L., Pun F.T. Artificial and Natural Hybrids in the Gramineae, Tribe Hordeae. VI. Chromosome Pairing in *Secale cereale* x *Agropyron* Intermedium and the Problem of Genome Homologies in the Triticinae. *Genetics*. 1953. V.38(6). P.600-608.
71. Stein N., Feuillet C., Wicker T., Schlagenhauf E., Keller B. Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the Lr10 resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proc. Natl. Acad. Sci. USA*. 2000. V.97(24). P.13436-13441. doi: 10.1016/S0065-2660(08)60132-7
72. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkel MK, Malhotra A, Stütz AM, Shi X, Casale FP, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalin AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA; 1000 Genomes Project Consortium, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korb J.O. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015. V.526(7571). P.75-81. doi: 10.1038/nature15394
73. Svetozarova V.V. O vtorom genome *T. timopheevii* Zhuk. *Dokl. AN SSSR*. 1939. V. 23(5). P. 472-476. [On the second genome of *T. timopheevii* Zhuk.] (In Russian)
74. Sun X., Jiao C., Schwaninger H., Chao C.T., Ma Y., Duan N., Khan A., Ban S., Xu K., Cheng L., Zhong G-Y., Fei Z. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020. doi: 10.1038/s41588-020-00723-9
75. Venter J.C. Multiple personal genomes await. *Nature*. 2010. V.464. P.676-677. doi: 10.1038/464676a
76. Venter J.C., Adams M.D., Myers E.W., Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C,

- Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science*. 2001. V.291(5507). P.1304-1351. doi: 10.1126/science.1058040
77. Vershinina Z.R., Kuluev B.R., Maksimov I.V., Mikhaylova, E.V., Gumerova G.R., Maleev G.V., Knyazev A.V., Baymiev An.Kh., Baymiev Al.Kh., Chemeris A.V. To prohibit GMOs impossible to resolve! *Biomics*. 2020. V.12(1). P. 80-120. DOI: 10.31301/2221-6197.bmcs.2020-6 (In Russian)
78. Walsh J.B., Marks J. Sequencing the human genome. *Nature*. 1986. V.322. P. 590.
79. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Li D, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang H, Wang J. The diploid genome sequence of an Asian individual. *Nature*. 2008. V.456(7218). P.60-65. doi: 10.1038/nature07484
80. Wang O., Chin R., Cheng X., Wu M.K.Y., Mao Q., Tang J., Sun Y., Anderson E., Lam H.K., Chen D., Zhou Y., Wang L., Fan F., Zou Y., Xie Y., Zhang R.Y., Drmanac S., Nguyen D., Xu C., Villarosa C., Gablenz S., Barua N., Nguyen S., Tian W., Liu J.S., Wang J., Liu X., Qi X., Chen A., Wang H., Dong Y., Zhang W., Alexeev A., Yang H., Wang J., Kristiansen K., Xu X., Drmanac R., Peters B.A. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res*. 2019. V.29(5). P.798-808. doi: 10.1101/gr.245126.118
81. Weisenfeld N.I., Kumar V., Shah P., Church D.M., Jaffe D.B. Direct determination of diploid genome sequences. *Genome Res*. 2017. V.27(5). P.757-767. doi: 10.1101/gr.214874.116
82. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.-J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niaz F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X-z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
83. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. 1920. Jena: Gustav Fischer Verlag. 250 s.
84. Zhang L., Zhou X., Weng Z., Sidow A. Assessment of human diploid genome assembly with 10x Linked-Reads data. *Gigascience*. 2019. V.8(11). giz141. doi: 10.1093/gigascience/giz141
85. Zhang W., Luo C., Scossa F., Zhang Q., Usadel B., Fernie A.R., Mei H., Wen W. A phased genome based on single sperm sequencing reveals crossover pattern and complex relatedness in tea plants. *Plant J*. 2020. doi: 10.1111/tpj.15051
86. Zhou B., Arthur J.G., Ho S.S., Pattni R., Huang Y., Wong W.H., Urban A.E. Extensive and deep sequencing of the Venter/HuRef genome for developing and benchmarking genome analysis tools. *Sci. Data*. 2018. V.5:180261. doi: 10.1038/sdata.2018.261
87. Zhou Q., Tang D., Huang W., Yang Z., Zhang Y., Hamilton J.P., Visser R.G.F., Bachem C.W.B., Robin Buell C., Zhang Z., Zhang C., Huang S. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet*. 2020. V.52(10). P.1018-1023. doi: 10.1038/s41588-020-0699-x
88. Zimin A.V., Puiu D., Hall R., Kingan S., Clavijo B.J., Salzberg S.L. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience*. 2017. V.6(11). P.1-7. doi: 10.1093/gigascience/gix097