



РАЗНООБРАЗИЕ КОЛИЧЕСТВЕННЫХ ОЦЕНОК СОДЕРЖАНИЯ ДНК В ЯДРАХ РАСТЕНИЙ, ИХ РАЗБРОС, НЕКОТОРЫЕ ТЕРМИНЫ И ПОНЯТИЯ (ГЕНОМ, C-VALUE, ПАНГЕНОМ)

¹Баймиев Ал.Х., ¹Кулуев А.Р., ¹Матниязов Р.Т., ¹Гарафутдинов Р.Р.,
¹Баймиев Ан.Х., ¹Гималов Ф.Р., ²Чемерис Д.А., ¹Кулуев Б.Р., ¹Чемерис А.В.

¹Институт биохимии и генетики Уфимского федерального исследовательского центра
Российской академии наук, Россия, 450054, Уфа, проспект Октября, 71. Email: baymiev@mail.ru
²ООО «Максим Медикал», Россия, 123423, Москва, ул. Народного Ополчения, д. 34, стр. 1

Резюме

В качестве подтверждения огромного разнообразия содержания ДНК в ядрах растительных клеток приведены граничные значения минимальных и максимальных показателей таких параметров, как число хромосом, варьирующее у растений на гаплоидный набор от двух хромосом до 480 (или даже до 720) хромосом, а также «весовые» размеры геномов в пикограммах в виде 1C-value с диапазоном от 0,065 до 152,23 пг на гаплоидный набор хромосом. Отмечена важность знаний количества ДНК и числа хромосом для выполнения проектов по секвенированию полных еще неизвестных геномов растений. Приведены также примеры секвенированных полных геномов растений резко отличающихся размеров (менее 100 млн.п.н. и свыше 30 млрд.п.н.). Уделено внимание необходимости придания столетнему термину «геном» нового смысла, согласно которого под ним должно пониматься уже все количество ДНК во всех хромосомах, независимо от пloidности, а не гаплоидный набор, как прежде, поскольку только совокупность всей ДНК (всех аллелей, как гомологичных, так и гомеологичных хромосом) определяет жизненный статус растения, да и любого другого эукариотического организма. Обращено внимание на то, что в понятие «полного генома» вкладываются различные уровни их завершенности от контигов и прочих черновых квазигаплоидных геномов до так называемых T2T геномов, представляющих собой установленные нуклеотидные последовательности по-хромосомно «от теломеры до теломеры». Наиболее полную информацию об организме может дать диплоидный геном, при определении которого производится так называемая фазированная сборка по гаплотипам. При этом для растений, и особенно для сортов культурных растений, эталонным геномом должен служить пангеном, несущий в себе максимальную информацию о различиях нуклеотидных последовательностей, характерных для исследуемого образца. Использование в селекции данных по пангеномам растений ознаменует собой переход на новый уровень подобных работ. Возможно в будущем, когда начнется массовое секвенирование действительно полных диплоидных геномов в формате T2T, для них потребуется новый термин, каким мог бы стать «дигеном» или коротко «дином».

Ключевые слова: растение, ядро, хромосома, геном, размер генома, C-value, пикограмм, полиплоидия, полногеномное секвенирование, T2T секвенирование, квазигаплоидный геном, диплоидный геном, пангеном, T2T геном, дином

Цитирование: Баймиев Ал.Х., Кулуев А.Р., Матниязов Р.Т., Гарафутдинов Р.Р., Баймиев Ан.Х., Гималов Ф.Р., Чемерис Д.А., Кулуев Б.Р., Чемерис А.В. Разнообразие количественных оценок содержания ДНК в ядрах растений, их разброс, некоторые термины и понятия (геном, C-value, пангеном) // *Biomics*. 2022. Т.14(1). С.79-100. DOI: 10.31301/2221-6197.bmcs.2022-6

© Авторы

THE VARIETY OF QUANTITATIVE ESTIMATES OF THE DNA CONTENT IN PLANT NUCLEI AND THEIR DISPERSION, SOME TERMS AND CONCEPTS (GENOME, C-VALUE, PANGENOME)

¹Baymiev Al.Kh., ¹Kuluev A.R., ¹Matniyazov R.T., ¹Garafutdinov R.R.,

¹Bayimiev An.Kh., ¹Gimalov F.R., ²Chemeris D.A., ¹Kuluev B.R., ¹Chemeris A.V.

¹Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, 71 Pr. Oktyabrya, Ufa, 450054, Russia. Email: baymiev@mail.ru

²Maxim Medical LLC, 34-1 Narodnogo Opolcheniya str., Moscow, 123423, Russia

Resume

As confirmation of the huge diversity of DNA content in the nuclei of plant cells, the boundary values of minimum and maximum parameters such as the number of chromosomes, varying in plants per haploid genome from two chromosomes to 480 (or even up to 720) chromosomes, as well as the "weight" sizes of genomes in picograms in the form of 1C-value with a range from 0.065 to 152.23 pg per haploid set of chromosomes. The importance of knowledge of the amount of DNA and the number of chromosomes for the implementation of projects for sequencing complete yet unknown plant genomes was noted. Examples of sequenced complete genomes of plants with very different sizes (less than 100 million bp and more than 30 billion bp) are also given. Attention is paid to the need to give the century-old term "genome" a new meaning, according to which it should already mean the entire amount of DNA in all chromosomes, regardless of ploidy, but not a haploid set, as before, since only the totality of all DNA (all alleles, both homologous and homeologous chromosomes) determines the vital status of a plant, and of any other eukaryotic organism. Attention is drawn to the fact that the concept of a "complete genome" includes various levels of their completeness from contigs and other draft quasi-haploid genomes to the so-called T2T genomes, which represent established nucleotide sequences chromosomally from "telomere to telomere". The most complete information about the organism can be given by the diploid genome, which is determined by the so-called phased assembly by haplotypes. At the same time, for plants and especially for cultivars of agricultural plants, as the reference genome should serve a pangenome that carries maximum information about the differences in nucleotide sequences characteristic of the sample under study. The use of plant pangenomic data in breeding will mark the transition to a new level of such work. Perhaps in the future, when the mass sequencing of truly complete diploid genomes in the T2T format begins, a new term will be needed for them, which could become "digenome" or briefly "dinome".

Keywords: plant, nucleus, chromosome, genome, genome size, C-value, picogram, polyploidy, whole genome sequencing, T2T sequencing, quasi-haploid genome, diploid genome, pangenome, T2T genome, dinome

Citation: Baymiev Al.Kh., Kuluev A.R., Matniyazov R.T., Garafutdinov R.R., Bayimiev An.Kh., Gimalov F.R., Chemeris D.A., Kuluev B.R., Chemeris A.V. The variety of quantitative estimates of the DNA content in plant nuclei and their dispersion, some terms and concepts (genome, C-value, pangenome). *Biomics*. 2022. V.14(1). P. 79-100. doi: 10.31301/2221-6197.bmcs.2022-6 (In Russian)

© Authors

Введение

Как ни покажется удивительным, начнем статью о растениях с животных объектов. Причиной тому является уже прочно сложившаяся тенденция «отставания» растений¹ как объектов в биохимических, генетических и молекулярно-биологических исследованиях.

¹ Прокариотические организмы и вирусы, обычно «опережающие» в этом контексте и животные организмы, в этой статье нас интересовать не будут.

Как только было выяснено, что именно ДНК является тем биополимером, что несет наследственную информацию, то интерес к этой молекуле резко возрос, включая количественные аспекты. Так, на сайте функционирующей с 2001 г. базы данных Animal Genome Size Database (<http://www.genomesize.com/index.php>) в качестве эпиграфа приведена фраза из статьи 1950 г. французских ученых R.Vendrely и C.Vendrely, звучащая на русском так - «Несомненно, что систематическое исследование абсолютного содержания дезоксирибонуклеиновой кислоты в ядре многих видов

животных может дать интересные предложения в отношении проблемы эволюции». Спустя полвека стало ясно, что для той же цели намного актуальнее знания нуклеотидных последовательностей геномов организмов и в разделе FAQ этой базы данных на вопрос для чего нужно знать размеры геномов среди ответов есть такой - что это очень важно знать при намерении впервые секвенировать полный геном какого-либо организма. И это действительно так, поскольку нужно оценить масштабность намечаемого проекта, соответственно, прикинуть стоимость таких исследований и даже понять их принципиальную выполнимость в настоящее время с учетом имеющихся технологических возможностей, как у самих авторов, так и вообще.

Наиболее массовый характер сейчас носит оценка геномов разных организмов в виде так называемой C-value², измеряемой в пикограммах (10^{-12} г). Несмотря на то, что главный интерес для нас представляют здесь растения, все же будет нелишним обратить внимание на представленный на рис. 1 огромный разброс подобных значений среди всех групп живых организмов, причем не зависящий от их нахождения на эволюционной лестнице, что даже получило свое название как C-value paradox [Thomas, 1971].

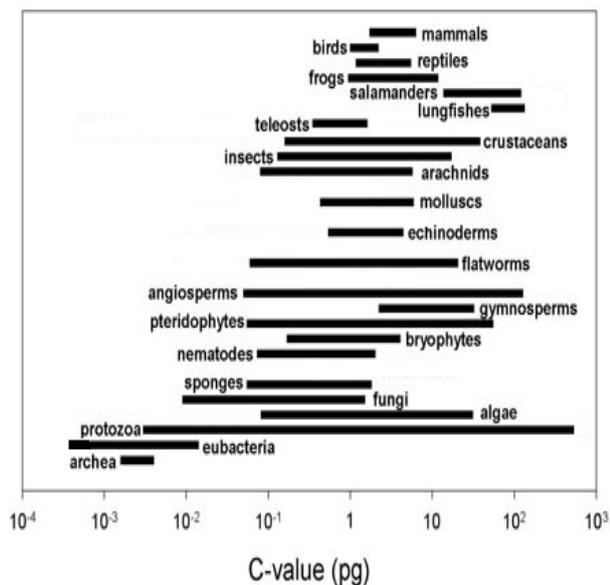


Рис. 1. «Разброс» значений C-value среди различных групп организмов (<http://www.genomesize.com/statistics.php> - модифицирован)

Fig. 1. Distribution of C-value values among different groups of organisms (<http://www.genomesize.com/statistics.php> - modified)

Для всех (свободноживущих) организмов в целом количества ДНК на (ядерный) геном сейчас укладываются в диапазон от 500 аттограмм (бактерии) до 700 пикограмм (простейшие), что составляет свыше шести порядков разницы и это огромный диапазон. Правда, в последние годы чрезмерно большие размеры геномов простейших стали ставить под сомнения, но это выходит за рамки темы данной статьи, и посему оставим без дальнейших комментариев. Но даже если не принимать во внимание простейших, то более пяти порядков разницы все равно имеются. Отметим лишь, что согласно этой базе данных самый крупный геном у животных организмов размером 132,83 пг выявлен у двоякодышащей рыбы *Protopterus aethiopicus* и к этой величине мы еще вернемся в Заключение. В целом можно сказать, что чем организмы более эволюционно продвинуты, тем в грубом приближении можно принять, что разброс C-value для их групп меньше. Исключение составляют археи.

В давние времена содержание ДНК в ядрах оценивалось химическим методом, экстрагируя ДНК и измеряя количество спектрофотометрически, так и цитологическим методом, окрашивая ДНК в ядре реактивом Фельгена с последующим микроденситометрированием, как это было описано в работе Ris, Mirsky [1949]. В той работе было оценено содержание ДНК в ядрах ряда организмов, среди которых была домашняя курица и у нее оказалось, как химическим так и цитохимическим методами определено одинаковое количество ДНК, составившее $2,4 \times 10^{-9}$ мг (то есть пикограмм), что вполне сопоставимо с представленным ныне в базе данных Animal Genome Size Database значениями C-value равными 1,25 / 1,28 пг при их правильном пересчете, поскольку в цитируемой статье это было содержание ДНК в ядре, а в базе данных – его количество указано на гаплоидное состояние. И в этом как раз кроются некоторая неоднозначность и непростые различия, к рассмотрению которых, на примере растений, дальше перейдем, причем ситуация с ними еще сложнее. Коснемся здесь лишь еще содержания ДНК в клетках человека, составляющее по этой базе данных 3,50 пг, что приблизительно соответствует так называемому гаплоидному геному, но при этом в соседней колонке приведено число хромосом, составляющее 46, что уже соответствует диплоидному геному человека, и у неподготовленных пользователей может возникать некоторая путаница с соответствием друг другу этих количественных оценок.

Ряд затрагиваемых нами вопросов очень подробно освещены в объемном обзоре Патрушева и Минкевича [2007], озаглавленном «Проблема размеров генома эукариот», и посему мы коснемся их коротко. Однако тот обзор вышел уже достаточно

² о которой будет говорить чуть ниже

давно, фактически до начала массового секвенирования полных геномов эукариотических организмов и за это время появились немало новых сведений, что подвигло нас на написание сего труда. К тому же мы поднимаем вопрос о нынешней неправильной трактовке самого понятия «геном» и необходимости придания ему фактически нового смысла, что становится насущным в результате новых технологических возможностей по секвенированию действительно полных геномов, включающих всю ДНК всего комплекта хромосом соматической клетки. При этом случаи эндополиплоидии мы здесь выводим за скобки, равно как и хлоропластный и митохондриальный геномы останутся за пределами рассмотрения.

Ядерный геном и связанные с ним термины и понятия

Раз уже неоднократно оказался упомянутым термин «геном», то ему необходимо уделить должное внимание. В столетний юбилей данного термина мы посвятили ему специальную статью [Кулуев и др. (Kuluev et al.), 2020] и потому остановимся на главном. Так, в монографии, посвященной вопросам распространения и причинам партеногенеза в растительном и животном царствах “*Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche*” Г.Винклером в 1920 г. был впервые предложен этот термин «геном» [Winkler, 1920]. В главе про связь хромосом с партеногенезом имеется высказывание, в переводе на русский звучащее «Я предлагаю использовать для гаплоидного набора хромосом, который вместе с прилегающей протоплазмой определяет материальные основы вида, выражение геном...». При этом далее в том же предложении следуют рассуждения о том, что если геном содержит более чем одну такую же единицу, то его следует называть гомогеноматическим, а если разные единицы, то тогда – гетерогеноматическим, что сегодня соответствует автополиплоидам и аллополиплоидам соответственно. Для составных частей геномов «гетерогеноматических» полиплоидов много позже был предложен свой термин «субгеном», впервые встречающийся при описании амфидиплоидного генома хлопчатника еще в 1954 г. [Menzel, 1954], но довольно долго остававшимся практически невостребованным. Для обозначения частей геномов полиплоидных видов в пшенично-эгилопном комплексе этот термин был употреблен лишь в 2000 г. [Stein et al., 2000], а сейчас практически ни одна статья, в которой рассматриваются филогенетически взаимоотношения этих видов, без этого термина уже не обходится, и мы к нему на примере пшениц в данной статье также обратимся.

Удивительно, но, похоже, кроме нас никто про юбилейную дату термина «геном» и не вспомнил. Причем важна не столько юбилейная дата (хотя она, безусловно, заслуживает внимания), сколько назревшее изменение вкладываемого в этот термин смысла. При этом нельзя обойти вниманием написанную несколько раньше статью отечественных авторов [Зеленин и др. (Zelenin et al.), 2016], посвященную возникновению различных генетических и молекулярно-биологических терминов, среди которых был и «геном». Однако в их работе были приведены классические определения, тогда как крайне необходимым становится пересмотр термина/понятия геном. Ранее зарубежными авторами [Noguero-Solano et al., 2013], наряду с происхождением термина «геном», было уделено значительное внимание вкладывавшемуся на протяжении десятилетий в понятие «геном» менявшемуся смыслу. Так, они отмечают, что в основе восприятия генома как гаплоидного набора хромосом в 1920 г. лежали тогдашние ограничения метода их визуализации путем окрашивания и наблюдения в световом микроскопе³. Технологические возможности и в дальнейшем сильно влияли на смысл, вкладываемый в понятие генома. В качестве примера этими авторами был приведен известный генетик Г.Стент, который определил геном как совокупность всех генов индивида⁴, тогда как Дж.Уотсон указывал, что геном – это гаплоидный набор хромосом, несущих ассоциированные с ними гены. В Заключении Noguero-Solano с соавторами [2013] пишут, что «концепция генома расширила область своего применения, объединив различные объекты, содержащие генетический материал от гаплоидного числа хромосом до диплоидного числа, а затем включив генетический материал, обнаруженный в органеллах, таких как митохондрии и хлоропласты». Мы с ними полностью солидарны, но поскольку рассмотрение генетического материала митохондрий и хлоропластов, названного еще в 1926 г. немецким ботаником F.R. von Wettstein [1926] «плазмоном», не входит в наши задачи, то оставим его в стороне и приведем лишь наше видение ядерного генома. Позволим себе с Нобелевским лауреатом, посчитавшим, что геном – это гаплоидный набор хромосом, не согласиться. Полный геном того или иного высшего организма – это не простое двукратное

³ К тому же что являлось непосредственным носителем генетической информации и насколько она в парных хромосомах отличается одна от другой тогда было абсолютно неизвестно, а именно это крайне важно.

⁴ т.е. придав понятию генома диплоидный уровень, хотя гены – далеко не весь геном

умножение одного, причем неодинакового гаплоидного набора, а сосуществование двух подчас довольно сильно различающихся гаплоидных наборов хромосом. И именно их сочетание определяет жизненную силу любого организма. Не говоря уже о том, что особи мужского пола несут две разные половые хромосомы, также являющиеся составной частью их гаплоидных геномов. Получается, что женщины и мужчины несут соответственно 23 и 24 хромосомы разных типов – какое же тут простое удвоение числа хромосом при переводе данных на диплоидный уровень для последних! При этом у собак, например, Y-хромосома намного меньше X-хромосомы (в 30 с лишним раз), что будет вполне заметно даже в пикограммах, измеряемых для C-value весьма округленно, не вспоминая уже про число нуклеотидов.

К сожалению, нынешние технологии полногеномного секвенирования фактически позволяют вести сборку лишь квазигаплоидных геномов, когда на одной цепочке ДНК непредсказуемым, по сути случайным образом могут чередоваться фрагменты материнской и отцовской гомологичных хромосом и для многих случаев это крайне плохо. Но в последние годы растет интерес именно к секвенированию действительно полных геномов, под которыми следует понимать всю ДНК всех хромосом организма - не важно какой он плоидности, на что мы уже делали акцент в цитируемой выше статье [Кулуев и др. (Kuluev et al.), 2020], говоря о том, что приходит время секвенирования полных диплоидных геномов, хотя прекрасно понимаем, что нынешнее секвенирование все же не очень для этого пригодно.

Для подсчетов C-value были предложены два новых термина – моноплоидный геном и холоплоидный геном (от греческого holos – весь, целый) [Greilhuber et al., 2005]. Первый должен был обозначать всю ДНК в одиночном наборе хромосом с числом x , а второй – ДНК всего хромосомного набора с числом хромосом n . Например, твердая пшеница *Triticum durum* представляет собой тетраплоидный организм, состоящий из $2n = 4x = 28$ хромосом с геномной формулой **ВВАА** с базовым числом хромосом 7, но принадлежащих двум родственным субгеномам с гомеологичными хромосомами, и если следовать «моно- и холоплоидным» обозначениям, то ее холоплоид будет представлен 14 хромосомами, а моноплоид – 7, что не есть верно, поскольку фактически *T. durum* – уже реальный диплоид, давно прошедший стадию функциональной диплоидизации.

Как можно видеть из рис. 2, полный геном тетраплоидной пшеницы состоит из 28 хромосом, среди которых есть как гомологичные, так и гомеологичные, образуемые двумя субгеномами **В** и

А, ведущими начало от двух диплоидных донорных видов – эгилопса и пшеницы, когда-то скрестившихся между собой, что было рассмотрено нами ранее [Кулуев и др. (Kuluev et al.), 2016].

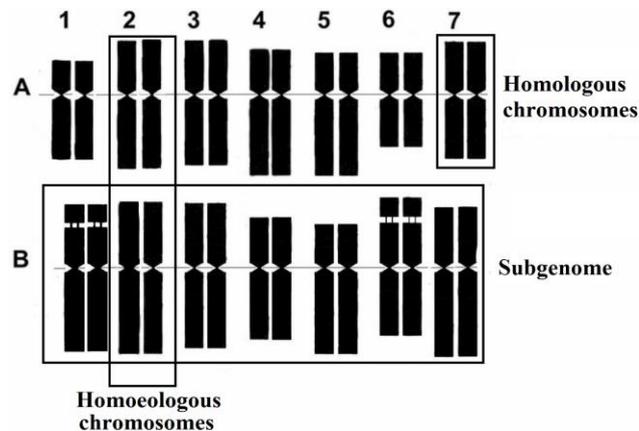


Рис. 2. Идиограмма хромосом тетраплоидной пшеницы. Пояснения в тексте.

Fig. 2. Idiogram of chromosomes of tetraploid wheat. Explanations in the text.

Поскольку полиплоидизация рассматривается как одно из приспособлений выживания и распространения растений [te Best et al., 2012; Soltis et al., 2015], то считается, что большинство видов растений – это древние палеоплоиды, давно ставшие по сути диплоидами [De Bodt et al., 2005] и для них моноплоидный статус установить не просто. С триплоидными видами растений, которых немало, ситуация еще запутаннее. Поэтому, исходя, в том числе и из подобных соображений, действительно проще и главное - точнее считать полным (ядерным) геномом всю совокупность ДНК всех хромосом в ядре. Что касается истинно гаплоидных геномов, то такие могли бы быть секвенированы при условии определения по отдельности комплектов парных хромосом, что для обычных организмов принципиально невозможно⁵. При этом даже в половых клетках единого гаплоидного генома, по сути, и нет, так как при кроссинговере происходят разнообразные обмены участками гомологичных хромосом с образованием множественных отличающихся гамет – почти каждая со своими уникальными геномами. Равно как не может быть единого транскрипта, поскольку они меняются от ткани к ткани, а также зависят от множества других факторов.

⁵ однако секвенирование ДНК, выделенной из культуры гомозиготных клеток, поддерживаемых *in vitro*, может позволить это сделать, о чем будет говориться дальше

Здесь нужно также заметить, что с появлением возможности секвенировать короткие фрагменты ДНК, а затем и полные геномы, у молекулярных биологов, кроме весовых характеристик содержащейся ДНК в ядре в виде соответствующих производных граммовых количеств, появилась новая единица измерения - нуклеотид, точнее пара нуклеотидов, поскольку ДНК в норме двухцепочечная молекула. Так, в 1991 г. была опубликована статья [Morton, 1991], посвященная различным характеристикам генома человека, в которой отмечалось, что каждая наука имеет параметры точного определения чего-то важного для нее, и приведены примеры: для астрономии - скорость света, а для химии - число Авогадро, при том, что для молекулярной биологии своих таких параметров нет, и она в ряде случаев пользуется тем же числом Авогадро. Нам представляется, что, по крайней мере, для той части молекулярной биологии, которая непосредственно связана с нуклеиновыми кислотами и геномами, важным параметром стало число нуклеотидов. В результате секвенирования с точностью не менее 99,99% и даже 99,9999% появилась возможность оперирования нуклеотидами, точнее их количеством и даже их неким качеством, под которым следует понимать типы азотистых оснований - А, С, G, Т (или U для РНК), а также пострепликационные модификации отдельных нуклеотидов. Таким образом, нуклеотид (нт) или их пара (п.н.) в виде звеньев полимерной цепочки нуклеиновых кислот могут рассматриваться как важный (измерительный) параметр для установления размеров геномов, в частности. В этой связи следует заметить, что появились новые базы данных⁶, где размеры полных геномов, определенные путем секвенирования, приводятся именно в нуклеотидах, а не в C-value, которые определяются сейчас с помощью проточной цитофлуориметрии и измеряются в пикограммах. При этом на самом деле легко пересчитать одни значения в другие. Так, 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 1 пг «составляют» 978 млн. п.н.

C-value растений

Наконец-то пришло время уделить внимание термину C-value и его возникновению. Согласно цитированной выше статье [Greilhuber et al., 2005] C-value обозначает содержание ДНК холоплоидного генома с числом хромосом n . То есть, это величина соответствует гаплоидному набору хромосом. Впервые термин C-value, обозначающий содержание ДНК на ядро, был предложен в 1950 г. при

исследовании количеств ДНК в разных типах тканей кукурузы и традесканции, подсчитанное тогда в неких условных единицах [Swift, 1950]. При этом в самой статье не дается пояснение - сокращением какого слова явилась буква “С” и у разных авторов были спекуляции на этот счет, но спустя много лет сам Swift в личной беседе пояснил, что “С” от “constant” [Bennett, Smith, 1976]. При этом само название той его статьи начинается со слова “Constancy” и автор [Swift, 1950] рассуждает в ней о постоянстве содержания ДНК на ядро в тканях разных типов у животных организмов, ставших известными к тому времени. Приблизительно тогда же вышла статья других авторов [Ogur et al., 1951], посвященная подсчету количества ДНК в ядрах *Lilium longiflorum* с помощью химического метода. Интерес эта работа вызывает тем, что приводимое количество ДНК указывается не в пикограммах, а как « μ ». Другой особенностью той статьи можно считать, что в ней РНК упоминается как пентозная нуклеиновая кислота с аббревиатурой “PNA”, что сейчас обозначает пептидно-нуклеиновую кислоту. Впрочем, в те годы, подобное обозначение РНК редкостью не было.

По сравнению с химическим методом более точные результаты оценки содержания ДНК в ядре обеспечивало окрашивание по Фельгену с последующим микроденситометрированием, и в различных вариациях такой подход просуществовал до начала 1970-х гг., когда ему на смену пришел способ определения количества ДНК в ядре с помощью проточной цитофлуориметрии. Эта технология заметно улучшила и упростила процедуру количественного измерения ДНК в ядрах и поставила определение C-value у различных организмов фактически на поток, что привело со временем к формированию соответствующей базы данных при ботаническом саде в Кью в Англии Plant DNA C-values Database (<https://cvalues.science.kew.org/>). Первый Release 1.0 состоялся в декабре 2001 г.⁷ В настоящее время последний Release 7.1 датирован апрелем 2019 г. и включает C-value для 12273 видов, из которых 10770 покрытосеменных, 421 голосеменных, 303 папоротников, 334 мохообразных и 445 водорослей [Pellicer, Leitch, 2020]. Притом, что считается, что различных видов в царстве растений, подцарстве Viridiplantae существует около 370 тысяч и поэтому в ней содержатся далеко не все виды, даже представляющие больший интерес, нежели некоторые остальные, в ней присутствующие. Чтобы не быть голословными, заметим, что нам не удалось найти в этой базе данных одуванчик осенний или крым-сагыз

⁶ например, Ensembl Plants - <http://plants.ensembl.org/species.html>

⁷ стоит отметить, что на самом сайте имеются серьезные расхождения с датировками выхода первого релиза

Taraxacum hybernum и одуванчик кок-сагыз *T.kok-saghyz* из семейства Астровых, являющиеся неплохими каучуконосами прошлого и, возможно, будущего [Гаршин и др. 2016; Кулуев и др., 2017; Гаршин, Кулуев, 2018; Bari et al., 2022]. Также мы не нашли информацию о размере генома первого каучуконоса, с которым познакомился Старый свет – кастилле эластичной *Castilla elastica* из семейства Тутовых [Кулуев и др. (Kuluev et al.), 2015], имеющей еще и другое название рода как *Castilloa* и этой практически полутравянистой путанице мы уделили специальное внимание в одной из своих статей [Сагитов и др. (Sagitov et al.), 2021]. Тем не менее, Plant DNA C-values Database выполняет очень важную роль для лучшего понимания мира растений и, вне всякого сомнения, востребована. При этом в Plant DNA C-values Database указывается, что приводятся значения содержания ДНК на ядро для 1C-value, тогда как в упоминаемой выше базе данных Animal Genome Size Database такого указания нет – в ней просто C value, хотя на самом деле имеется в виду именно - 1C. Понимая, что проблема с соответствием C-value разным пloidным статусам клеток существует, была предпринята попытка введения дополнительных обозначений в виде нижних и верхних индексов [Greilhuber, Dolezel, 2009], но до конца проблему это не решило.

Хромосомы растений

В упоминаемых выше базах данных по C-value не приводится информация о числе хромосом, в которых подсчитана ДНК и содержится, хотя это тоже довольно важная информация, в том числе для начала новых работ по полногеномному секвенированию неизвестных геномов. Впрочем, имеется специализированная база данных Chromosome Counts Database (<http://ccdb.tau.ac.il/>) [Rice et al., 2014], позволяющая проводить по ней поиск разных растений на предмет числа хромосом в них и, воспользовавшись этим, мы обнаружили, что у той же *Castilla*⁸ *elastica* число хромосом - 14⁹. Есть в этой базе данных и упоминаемые выше одуванчики с базовыми числами соответственно 16 и 8. Но поскольку *T.hybernum* – тетраплоид, то его базовое число хромосом также должно равняться 8. Однако с учетом его апомиктичности он скорее всего аллотетраплоид [Поддубная-Арнольди, Дианова, 1937], и тогда действительно его базовое число

хромосом будет 16, но ответить точно на этот вопрос может полногеномное секвенирование этого вида. Вообще, литературы по хромосомам растений довольно много, включая отечественные публикации, среди которых достаточно указать капитальный, почти тысячестраничный труд коллектива авторов [Болховских и др. (Bolkhovskikh et al.), 1969].

Поэтому мы здесь решили ограничиться лишь приведением крайних значений количества хромосом, обнаруженных ныне для растений, в том числе после публикации того глобального труда отечественных авторов. Так, сейчас известно шесть видов, базовое число хромосом у которых равно 2. Это отечественные злаки родов *Zingeria* и *Colpodium* [Цвелев, Жукова (Tsvelev, Zhukova), 1974], еще два однодольных растения *Ornitogalum tenifolium* и *Rhynchospora tenuis* из семейств Спаржевых и Осоковых соответственно, а также двудольные *Haplopappus gracilis* и *Brachyscome dichromosomatica* из Астровых [Castiglione, Cremonini, 2012]. Причем в этой последней статье приведены количества ДНК в пикограммах на ядро у всех этих шести видов, но указаны они все в 4C-value. Нет смысла приводить их конкретные значения по видам, лишь отметим, что варьируют они в довольно широких пределах от 1,56 до 8,20 пг, что означает, что геномы у них в основном немаленькие и составят от 1,2 до 4,0 млрд.п.н. на полный диплоидный геном. Растений с единственной хромосомой (в гаплоидном состоянии) пока не обнаружено и, возможно, это противоречит самой Природе.

Перейдем сейчас к другой крайности в виде громадного числа хромосом. Хотя в разных цитированных ниже статьях приводятся количества хромосом, как на гаплоидный, так и на диплоидный наборы (без учета их полиплоидных статусов) мы приведем их все, чтобы было удобнее сравнивать в пересчете на гаплоидный геном, как они фигурируют и в упоминаемой выше базе данных Chromosome Counts Database. Так, у папоротника уховника *Ophioglossum reticulatum* число хромосом на условный гаплоидный набор составляет 720 [Khandelwal, 1990] и это верхняя ныне известная граница, хотя в базе данных CCDB для этого вида приводится оценка 480 хромосом, что и в этом случае больше чем у остальных. У очитка *Sedum suaveolens* из семейства Толстянковых обнаружено 320 хромосом [Johnson et al., 1989]. Для одного из видов кокосовых пальм *Voanioala gerardii* сначала число хромосом было определено для гаплоидного набора как не менее чем 298 хромосом [Johnson et al., 1989], а позже уточнено – 303 хромосомы [Röser, 2015]. Новозеландский эндемик *Strasburgeria robusta* из семейства *Strasburgeriaceae* несет 250 хромосом в гаплоидном наборе [Oginuma et al., 2006]. Эндемик Австралии -

⁸ приводим здесь против правил название рода *Castilla* полностью, поскольку, как уже говорилось выше, имеются разночтения в его написании

⁹ здесь и далее будут приводиться числа хромосом в расчете на гаплоидное (n) состояние и где оно будет известно будут указываться базовые числа хромосом

древнее растение из порядка Psilotales *Tmesipteris obliqua* является, по всей видимости, октаплоидом и несет от 208 до 214 хромосом [Tindale, Roy, 2002].

Безусловно, все эти «мультихромосомные» виды являются полиплоидами, причем установить их кратность далеко не всегда представляется возможным, поскольку крайне трудно определиться с базовым числом их хромосом. Из этих видов согласно базе данных Plant DNA C-values Database лишь для нескольких определено содержание ДНК в виде их 1C-value. Так, для *S.suaveolens* оно составляет 9,1 пг, а для *V.gerardii* – 30 пг, а для *T.obliqua* – 150,61, что делает этот вид одним из рекордсменов по количеству ДНК в ядре, о чем дальше пойдет речь.

Весовые размеры геномов растений

Продолжим рассматривать C-value отдельных видов по 1C количеству также с акцентом на минимальные и максимальные граничные значения, известные к настоящему времени и представленные в том числе в базе данных Plant DNA C-values Database.

Долгое время считалось, что геном резуховидки Таля или арабидопсиса из семейства Капустных, характеризуясь значением 0,16 пг, является самым маленьким среди растений. Это количество ДНК у арабидопсиса распределено по 5 хромосомам. Еще несколько близких видов из этого рода также имеют относительно небольшие геномы. Поскольку резуховидка Талья обладает еще целым рядом удобных для экспериментаторов, в том числе для молекулярных биологов черт, то этот вид стал модельным растением и его геном первым среди растений был полностью секвенирован, о чем будет говориться ниже. Однако позже были найдены другие представители растительного царства с еще меньшими геномами. Ими оказались насекомоядные растения из семейства *Lentibulariaceae* – *Genlisea margaretae* (0,0648 пг), *G.aurea* (0,065 пг) и *Utricularia gibba* (0,0903 пг) [Greilhuber et al., 2006]. Позже эта же группа авторов произвела уточнения их размеров и установила, что наименьшие геномы имеют *G.tuberosa* и *G.aurea* – 0,61 и 0,65 пг, тогда как у *G.margaretae* он заметно крупнее [Fleischmann et al., 2014]. Также было показано, что для рода *Genlisea* для разных видов числа хромосом варьируют – $2n = 2x = 16$ и $2n = 4x = 32$. Другими авторами для представителей этих двух родов *Genlisea* и *Utricularia* были получены сходные данные [Veleva et al., 2014].

Есть и противоположная крайность размеров геномов в царстве растений. При этом по разным видам у различных авторов встречаются довольно сильно разнящиеся значения, что, впрочем, легко объяснимо тем, что при больших величинах и ошибки определения выходят значительнее. Приведем лишь несколько примеров, близких к некоей верхней

границе, о которой пойдет речь в Заключение. Нужно заметить, что порядок Лилейных из семейства однодольных в целом характеризуется весьма крупными геномами. Так, рекордсменом может считаться хорошо известное растение Вороний глаз японский *Paris japonica* ($2n = 8x = 40$) с его для 1C-value 152,23 пг [Pellicer et al., 2010]. *Trillium hageae* из семейства Melanthiaceae или Trilliaceae того же порядка Лилейных – 132,5 пг для 1C-value [Zonneveld, 2010], *Fritillaria assyriaca* из того же семейства – 77,6 пг для 1C-value [Leitch et al., 2007]. Из двудольных растений крупный геном у Омелы белой *Viscum album* – 102,9 пг [Zonneveld, 2010]. Упомянувшееся выше древнее растение из порядка Psilotales *Tmesipteris obliqua* имеет один из наиболее крупных известных геномов – 150,61 для 1C-value [Hidalgo et al., 2017].

Таким образом, можно констатировать, что по значениям 1C-value различные представители царства растений отличаются почти в 2500 раз. Однако крайности в виде крошечных и гигантских геномов все же не носят массовой характер. Так проведенный отечественными авторами анализ известных C-value растений показал, что около 80% геномов всех растений укладываются в относительно узкий диапазон от 0,6 до 16,7 пг [Шереметьев и др., 2011]. В одной из статей другие авторы задались вопросом – есть ли предел размерам геномов, которые как у растений, так и у животных организмов ограничиваются величинами около 150 пг [Hidalgo et al., 2017a]. Но мы вернемся к этому вопросу в Заключение.

Размеры геномов растений в п.н.

Первым секвенированным геномом стал геном арабидопсиса *Arabidopsis thaliana*, что не удивительно, учитывая его малый размер, казавшийся в то время наименьшим. Тогда удалось определить 115 млн.п.н. [Arabidopsis Genome Initiative, 2000]. Спустя пару лет были определены практически полные геномы двух подвидов риса – *Oryza sativa* ssp. *japonica* и ssp. *indica*, охватив более 90% их геномов размерами 420 и 462 млн.п.н. соответственно, статьи о чем появились в одном и том же номере журнала Science [Yu et al., 2002; Goff et al., 2002]. Здесь попутно можно заметить, что через полтора десятилетия с помощью секвенирования новых поколений в формате *de novo* повторно был собран геном *O.sativa* ssp. *indica* с 99% покрытием. За арабидопсисом и рисом последовали другие геномы, для части которых применялись комбинации методов секвенирования по Сэнгеру и полногеномного секвенирования с использованием разных платформ, а затем и только с помощью методов новых поколений.

Как говорилось выше, после арабидопсиса были найдены другие растений с еще меньшими, чем

у него геномами, одним из которых является вид *G. aurea*, имеющий ориентировочный размер около 63,6 млн.п.н., из которых 43,4 млн.п.н. было секвенировано отечественными авторами, что позволило составить черновой геном этого вида [Leushkin et al., 2013], размещенный в GenBank - https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_000441915.1/. Авторы обнаружили в секвенированной части генома около 18 тыс. генов. Для арабидопсиса подобный подсчет дал следующие числа – 25,5 тысяч генов оказались распределены по 115,5 млн.п.н. [Vevan et al., 2001]. Сопоставление этих соотношений показывает, что геном *G. aurea* в среднем чуть плотнее «упакован».

Также в 2013 г. был секвенирован миниатюрный геном другого насекомоядного растения – *Utricularia gibba*, размер которого составил 82 млн.п.н. с числом генов около 28,5 тысяч [Ibarga-Laclette et al., 2013]. Проведенный детальный анализ полного генома этого вида показал, что на долю повторяющейся ДНК приходится только около 3%. При этом авторы посчитали, что это древний палеоплоид, подвергнувшийся, по крайней мере, трем раундам полногеномной дупликации (WGD – Whole Genome Duplication) – по сути автополиплоидии, сопровождающейся потерей части ДНК, но (обычно) не генов, а их некоторым перераспределением между гомологичными хромосомами. Ими предложена схема этого процесса, посчитав, что подобные небольшие геномы возникают, когда скорость рекомбинационных делеций заметно выше по сравнению с пролиферацией различных последовательностей. Впоследствии был секвенирован также относительно небольшой геном другого вида этого рода – *U. reniformis*, имеющий размер около 304 млн.п.н. [Silva et al., 2020]. Для него было предсказано 42582 гена. Сравнение геномов этих двух видов *Utricularia* показало, что у *U. gibba* на долю генных последовательностей приходится около 51% всего генома, тогда как у *U. reniformis* – только 26%.

Стоит упомянуть, что из растений, имеющих хозяйственное значение, наименьший геном обнаружен и секвенирован у персика *Prunus persica* – всего 215 млн.п.н., распределенных по 8 хромосомам [Ahmad et al., 2011].

К секвенированию более крупных геномов приступили лишь после того, как стали доступны методы секвенирования новых поколений, поскольку до этого объем получаемых в виде отдельных прочтений данных не позволял составлять их черновые геномы, не говоря уже о более «высоких материях» в виде Т2Т и прочих гееномах. Причем наиболее эффективным оказалось совместное использование двух подходов в виде длинных

малочисленных прочтений с наложением на них коротких высокоточных прочтений.

К видам с такими крупными геномами безусловно нужно отнести мягкую пшеницу *Triticum aestivum*, гаплоидный размер генома которой предварительно оценивался в 17 млрд.п.н. Причем эта пшеница представляет собой гексаплоидный организм, хотя и прошедший стадию диплоидизации, тем не менее имеющий гомеологичные хромосомы из трех субгеномов со схожими нуклеотидными последовательностями, затрудняющими их сборку в дополнении к большому количеству повторяющихся элементов, также мешающими формированию протяженных последовательностей. Поэтому к геному мягкой пшеницы итоговым размером около 15 млрд.п.н. и отдельным сортам этого вида разные исследователи подступались неоднократно [Brenchley et al., 2012; Chapman et al., 2015; Zimin et al., 2017; IWGSC, 2018; Alonge et al., 2020; Athiyannan et al., 2022]. Если из культурных растений персик имеет наименьший геном, то у мягкой пшеницы он – наибольший.

Еще более крупные геномы несут представители некоторых голосеменных. В 2013 г. секвенирован геном ели европейской *Picea abies* размером 19,6 млрд.п.н. [Nystedt et al., 2013]. При этом подсчитанное число генов составило 28354, что ненамного больше, чем у арабидопсиса с геномом в 100 с лишним раз меньшим. Обращают на себя внимание размеры экзонов/интронов у этих видов. У арабидопсиса в среднем они составляют – 250/170 п.н., тогда как у ели всего лишь чуть-чуть больше – 312/1017 п.н. – несоразмерно с общими размерами их геномов. Также крупным секвенированным геномом оказался геном сосны ладанной *Pinus taeda* – 22 млрд.п.н. [Zimin et al., 2017a]. Еще крупнее геном у секвойи вечнозеленой *Sequoia sempervirens*, имеющий размер 26,5 млрд.п.н. [Neale et al., 2022]. Однако данный вид – гексаплоид. Лишь немного уступила секвоей китайская сосна *Pinus tabulaeformis*, имеющая геном 25,6 млрд.п.н. [Niu et al., 2022], распределенный по 12 гигантским хромосомам. При этом авторы отмечают, что не обнаружили следов недавнего WGD. Ранее, также довольно крупный геном лиственницы сибирской *Larix sibirica* размером 12,34 млрд.п.н. секвенирован отечественными авторами [Kuzmin et al., 2019]. Но «пальму первенства» удерживает сосна Ламберта, называемая еще сахарной, *P. lambertiana* с геномом около 31 млрд.п.н. [Stevens et al., 2016]. Размер экзонов и интронов у этой сосны в среднем составляет 1330 п.н. и 8039 п.н. соответственно.

В недавнем обзоре [Sun et al., 2022], посвященном 20-ти летию секвенирования первого растительного генома, говорится, что к моменту его написания в мире секвенированы более 1000 геномов,

принадлежащих 788 различным видам растений. И их количество постоянно растет. При этом свежие сведения можно почерпнуть из сайта полногеномного агрегатора **Published Plant Genomes** (https://plabipd.de/timeline_view.ep), во временной шкале которого отображается информация о всех секвенированных геномах растений в соответствии с датой их первой публикации. При наведении курсора мыши на тот или иной вид растения появляется соответствующее окно с указанием размера генома в млн.п.н. и ссылкой на первоисточник, кликнув по которому происходит переадресация на конкретную публикацию. Полные геномы около сотни видов растений собраны в базе данных **Ensembl Plants** (<http://plants.ensembl.org/species.html>), значительное количество полных геномов растений можно найти в **NCBI** (<https://www.ncbi.nlm.nih.gov>). Ряд геномов отдельных линий арабидопсиса содержатся в специализированной базе данных **1001 Genomes Project** (<https://1001genomes.org>). Развивается специальный проект **10,000 Plant Genomes Project** (<https://db.cngb.org/10kp/>), в ходе которого предполагается секвенировать полные геномы 10 тысяч растений из всех основных групп [Twyford, 2018]. Возможно, среди них окажутся виды как с мелкими, средними, так и с крупными геномами.

Ядерные геномы организмов – разнообразие вариаций их полноты секвенирования

Полные геномы организмов, включая растения - понятие довольно растяжимое, поскольку они имеют много разных уровней завершенности, которые здесь сначала просто перечислим, уделив больше внимания некоторым ниже. Так, по окончании процесса секвенирования в виде работы какого-либо полногеномного секвенатора образуется очень большое количество (обычно) коротких прочтений – ридов, представляющих собой сырые данные, характеризующиеся тем или иным показателем покрытия гаплоидного генома. Их компьютерная обработка и сборка превращает эти (короткие) последовательности в контиги¹⁰ различной (большей) протяженности, компиляции которых и составляют некий геном. При этом контиги могут покрывать далеко не весь геном, но если он секвенируется впервые, то этот уровень (пока еще) дает возможность говорить о «полном» прочтении данного генома. Большее покрытие позволяет из части контигов сформировать скаффолды и это, разумеется, также позволяет говорить о том, что полный геном

прочитан. Это так называемые «черновые» геномы. При этом, как уже говорилось выше, для высших организмов собираются по сути квазигаплоидные геномы, имеющие относительно мало общего с истинным распределением нуклеотидных последовательностей по конкретным хромосомам. Здесь нужно заметить, что если геном секвенируется впервые, то иного варианта сборки как *de novo* для него быть не может. Другое дело – повторное секвенирование уже известного генома у других образцов. В этом случае говорят о ресеквенировании и о сборке сырых данных, опираясь на известную последовательность.

Следующим уровнем секвенирования/сборки полных геномов нужно считать по-хромосомное представление данных, что, безусловно, несет больше важной информации, но при этом последовательности нуклеотидов в них не перестают быть квазигаплоидными. Приведение результатов секвенирования в формате «от теломеры до теломеры», называемое еще как «T2T» (*telomere-to-telomere*), стало использоваться лишь недавно и вызвало огромный рост интереса к этому варианту представления геномных данных, что является фактически более точной и полной версией по-хромосомного представления данных. Еще важнее секвенировать полные диплоидные геномы организмов, синонимами которых являются геномы с гаплотипной сборкой или фазированные, несущие информацию об обоих аллельных вариантах парных хромосом. В связи с этим подходом недавно появился новый термин – «гаплотиг»¹¹. Тем самым такие геномы перестают быть квазигаплоидными. При этом сборка подобных диплоидных геномов производится исключительно в формате *de novo*, поскольку важно найти именно отличающиеся участки.

«Повышенные» уровни завершенности составления полных геномов (по-хромосомно, T2T, фазированные/диплоидные) подразумевают затрачивание гораздо больше усилий и, кроме увеличенного покрытия, требуют применения длинных прочтений, Hi-C секвенирования, оптического картирования геномов и прочих ухищрений в виде, например трио-секвенирования, когда такое возможно. Также необходимо соответствующее программное обеспечение, которого уже разработано немало, но оно явится темой другой нашей статьи, и здесь на нем останавливаться не будем. Наконец, составляются уже и пангеномы, несущие, применительно непосредственно к нуклеотидным последовательностям, некую

¹⁰ впервые эта дефиниция была предложена R.Staden в 1980 г., решившим сократить использовавшееся при анализе нуклеотидных последовательностей выражение “*contiguous consensus sequence*” до “*contig*”

¹¹ термин “haplotig” обозначает гаплотип-специфичный контиг, используемый при сборке диплоидных геномов [Koren et al., 2018].

консолидированную информацию о разных аллелях. Фактически пангеномы можно считать неким аналогом множественных диплоидных геномов. Так, практически каждое растение любого вида/сорта несет несколько отличающийся в силу неизбежных мутаций геном, но определять диплоидные геномы отдельных растений не имеет смысла (в отличие от человека и некоторых животных организмов) и поэтому для них следует ограничиваться диплоидными геномами конкретного сорта, а также пангеномами вида.

В 2005 г. знание геномов нескольких образцов (изолятов) одного вида бактерий, слегка отличающихся друг от друга, привело к появлению термина «пангеном» [Tettelin et al., 2005], позже распространившегося и на другие организмы, включая растения. Сведения о разнообразии геномов одного вида продолжали накапливаться, и для растений секвенирование пангеномов началось в 2010 г., когда были ресеквенированы полные геномы 17 диких и 14 культурных образцов сои [Lam et al. 2010]. Стало ясно, что оперировать одной референсной или эталонной геномной последовательностью для одного вида нельзя. Так, пангеномы несут так называемые коровые гены, типичные для всех секвенированных образцов данного вида/сорта, а также некие дополнительные гены, присущие лишь отдельным образцам. Появилось немало обзоров по этой тематике [Golicz et al., 2016; Bayer et al., 2020; Della Coletta et al., 2021]. Можно встретить даже выражения, что аграрная наука вступила в пангеномную эру. Существуют специализированные базы данных по пангеномам растений и приведем в качестве примера подобную по мягкой пшенице <http://appliedbioinformatics.com.au/cgi-bin/gb2/gbrowse/WheatPan/>, изначально основанную на 18 сортах пшеницы [Montenegro et al., 2017], тогда как сейчас в ней содержатся сведения и о пангеномах капусты *Brassica oleraceae* и рапса *B.napus*. В недавней обзорной статье в виде временной шкалы приводятся сведения о почти трех десятках пангеномов сельскохозяйственных растений, упоминая в ряде случаев гаплотипную сборку [Li et al., 2022]. Ранее другие авторы прямо указали, что гаплотипное фазирование – это новые рубежи в сборке растительных пангеномов [Michael, VanBuren, 2020], требующиеся для селекционной работы нового качества. Частичным решением особенно для крупных геномов может служить полноэкзомное секвенирование, способное предоставить необходимую информацию об изменчивости кодирующих областей и присутствия в ней тех или иных генов, от которых зависит урожайность конкретного сорта и важности его вовлечения в селекционный процесс. Появится ли новый термин –

«панэкзом» – покажет время¹². Но с учетом того, что для селекции повторяющиеся последовательности не столь важны, то опираться на панэкзом может быть весьма продуктивно и намного экономнее. Однако здесь различных вариантов экзомного секвенирования касаться не будем, так как нами готовится специальная статья по этой теме.

Все же передовые рубежи в самом полногеномном секвенировании будущего и даже уже настоящего – это секвенирование фазированных (диплоидных) геномов, по-хромосомно, от теломеры до теломеры. И уже есть, по крайней мере, некоторые (частичные) примеры подобных геномов для растительных, да и для животных организмов. Для полноты картины набирающего сейчас силу фактически нового секвенирования начнем здесь с последних. Так, недавно сообщено о секвенировании (псевдо)диплоидного генома осла *Equus asinus* размером около 4 млрд.п.н., полнота данных которого в фазированном виде по-хромосомно составила больше 99,7% [Miao et al., 2022]. Просеквенировав обычные геномы 23 других ослов и объединив их с данными по 133 особям, авторам удалось найти более чем 39 млн. генетических вариаций. Другой подход был использован для генома человека в виде T2T секвенирования, для чего были взяты клетки гомозиготной линии с кариотипом 46,XX, что позволило собрать без пробелов все хромосомы в их действительно гаплоидном варианте¹³ общим размером 3,055 млрд.п.н. и исправить все ошибки предыдущих сборок [Aganezov et al., 2022; Nurk et al., 2022]. Последовавшие отклики коллег отмечают, что полученные данные представляют передовой уровень знаний о геноме человека и будут служить новым эталонным геномом, лишенным многочисленных ошибок прежних сборок [Commentary, 2022; Lupski, 2022]. Ранее другими авторами было продемонстрировано, что использование специального подхода, названного ими DipAsm (Diploid Assembly) позволяет (быстро) собирать диплоидные геномы (человека), что (как отмечается в цитируемой статье) крайне важно как для высококачественной медицины, так и для выяснения популяционного разнообразия [Garg et al., 2021].

Возвращаясь к растениям, нужно заметить, что в нашей предыдущей, упоминаемой уже несколько раз публикации [Кулуев и др. (Kuluev et

¹² Справедливости ради следует сказать, что термин “pan-exome” ранее в одной из статей уже был применен (причем лишь однократно и только в аннотации к ней), но использован он был в несколько ином контексте [Xu et al., 2017].

¹³ за небольшими исключениями, указываемыми самими авторами

al.), 2020], мы привели ряд примеров передовой сборки диплоидных геномов нескольких видов растений и, чтобы не повторяться, приведем геномы других растений, собранных после выхода той статьи, как в фазированном формате, так и от теломеры до теломеры, причем все они секвенировались повторно, поскольку их квазигапloidные геномы были давно известны. Так, собраны в фазированном формате геном хмеля *Humulus lupulus* размером около 2,8 млрд.п.н. [Padgitt-Cobb et al., 2020] и африканской кассавы или маниока *Manihot esculenta* размером около 700 млн.п.н. [Qi et al., 2022]. T2T геномы за последние годы секвенированы сразу у нескольких видов растений – кукурузы *Zea mays* [Liu et al., 2020], банана *Musa acuminata* [Belser et al., 2021], арбуза *Citrullus lanatus* [Deng et al., 2022], риса *Oryza sativa* [Zhang et al., 2022]. Ведется аналогичная сборка генома ячменя сорта Morex [Navratilova et al., 2022]. Нет сомнений, что вскоре перечень видов растений, у которых будут заново на подобных передовых уровнях секвенироваться их геномы, расширится.

Заключение

Как можно видеть из приведенных в статье примеров – для минимальных и максимальных количеств ДНК в ядрах растительных клеток существует огромный диапазон. Причем плохо объяснимый. Но сначала несколько слов стоит сказать про минимальные геномы и достаточное для растений число генов. С учетом имевших место процессов WGD и известных миниатюрных геномов ряда видов можно допустить, что теоретически могли бы существовать растения с геномом размером в 35 – 40 млн.п.н., что приблизительно на треть меньше ныне известных, и числом генов около 20 – 25 тысяч. Может быть, такие и будут когда-нибудь найдены с учетом того, что пока обследовано всего около 5% всех растений.

Что касается верхней границы, то она сейчас составляет около 150 пг или приблизительно столько же млрд.п.н. на гаплоидный геном и вряд ли увеличится даже на треть при анализе буквально всех организмов, в первую очередь на их C-value. Причин тому несколько, приведенных в одной из статей, авторы которой задались вопросом – есть ли предел размерам крупных геномов (растений) [Hidalgo et al., 2017a]. У большинства видов процессы, приводящие к расширению генома, обычно уравниваются механизмами, основанными на рекомбинации, приводящими к уменьшению размера генома, представляющими совместно два динамичных, но противоположных процесса. У видов с большими геномами, видимо, рекомбинация работает несколько иначе. В частности для растений рода *Fritillaria* семейства Лилейных с крупными геномами около 50

пг показано, что отсутствие процесса удаления ДНК ведет к увеличенным геномам [Kelly et al., 2015]. Здесь можно напомнить, что у животных организмов самый крупный геном, выявленный у двоякодышащей рыбы *P.aethiopicus* размером 132,83 пг лишь ненамного меньше чем максимальные у растений¹⁴. Вероятно, есть глубинные причины поддержания размеров крупных геномов на некоей величине, которая оказывается еще – 1) энергетически возможной для организма; 2) обеспечивающей точную репликацию при делениях клеток и репарацию постоянно происходящих мутаций; 3) пространственно допустимой; 4) способной поддерживать эффективность процессов транскрипции. Но чтобы лучше узнать про особенности функционирования таких гигантских геномов, нужны новые более эффективные методы секвенирования ДНК, к которым ниже перейдем.

К сожалению, в силу больших и даже огромных размеров геномов у некоторых растений, их подчас высокой ploидности, множественных повторов разной природы, а также сходства нуклеотидных последовательностей парных хромосом, существующие методы секвенирования новых поколений не соответствуют в полной мере предъявляемые к геномам высоких уровней сборки требованиям. Однако нужно заметить, что секвенирование ДНК по Сэнгеру¹⁵ для эффективного полногеномного секвенирования также не подходило (как мы сейчас это прекрасно понимаем), тем не менее, оно многие годы для этого использовалось, поскольку не было другого. Также и сейчас, нынешние методы крупномасштабного секвенирования ДНК¹⁶ мало пригодны для секвенирования диплоидных геномов (особенно крупных), но все же с тем или иным успехом использоваться могут. В этом даже усматривается некая параллель между методом Сэнгера и секвенированием нынешних квазигапloidных геномов и нынешними методами следующих поколений и передовым секвенированием диплоидных геномов, поскольку и в первом и во втором случаях присущие им технологические возможности удовлетворяли/ют не полностью существующим потребностям, но процессы шли/идут. При этом, однако, есть надежда и даже определенная

¹⁴ «огромные» геномы простейших все же вызывают сомнения и здесь про них опять-таки не вспоминаем – откуда они такие могли взяться или почему таковыми определены

¹⁵ рассмотренное нами когда-то очень детально – Чемерис и др. (Chemeris et al.), 1999

¹⁶ кратко рассмотренные в одной из наших недавних статей – Zubov и др. (Zubov et al.), 2021

уверенность в том, что рано или поздно появятся еще более производительные технологии секвенирования ДНК новейших поколений, которые выведут этот процесс на новый уровень, которым должно стать секвенирование действительно полных диплоидных геномов по-хромосомно от теломеры до теломеры.

Чтобы у читателей не сложилось впечатление, что так думаем только мы, приведем несколько высказываний из недавней статьи, посвященной перспективам изучения геномов растений [Kress et al., 2022]¹⁷. Так, авторы отмечают, что большинство усилий по сборке и аннотации геномов на сегодняшний день не соответствуют некоему эталонному геному, с которым можно было бы сравнивать геномы других секвенированных образцов того же вида. Причем концепция такого эталонного генома в последние годы заметно эволюционировала от компиляции контигов геномного пространства до хромосомной сборки, в которой имеет место уже хромосомная локализация генов. К тому же когда-то эталонный референсный геном принадлежал фактически отдельному организму (образцу), тогда как сегодня идеальным эталонным геномом должен быть пангеном. Также авторы обращают внимание на то, что черновые сборки геномов хотя и достаточны для оценки геномного пространства и содержания повторов, однако имеют ограниченную пригодность для исследований хромосомной организации. К тому же такие компиляции сборок с коротким чтением не следует вообще называть «геномами», однако исторический прецедент заставляет называть их именно «полными» (черновыми) геномами, при этом признавая, что они могут быть вполне информативными, в том числе для дальнейших исследований. В то же время эти авторы [Kress et al., 2022] призывают зарезервировать термин «геном» для сборок на уровне хромосом. Однако не уточняют, что это должны быть фазированные сборки диплоидных геномов, что на самом деле крайне важно. Хотя рассуждая дальше о новых открывающихся технологических возможностях в области геномного секвенирования, они отмечают, что «теперь можно создавать хромосомные и даже гаплотипные сборки очень сложных геномов».

Вне всякого сомнения, что по мере развития технологий полногеномного секвенирования, включая соответствующий биоинформатический анализ, эталонные геномы различных растений будут представлять собой сборки на уровне хромосом, несущие информацию о нескольких образцах или, иначе, конкретных пангеномах, которые будут служить множеству целей, включая эволюционные исследования, а также селекцию растений,

сопряженную с вопросами адаптации к условиям среды. В том числе, с использованием преимуществ машинного обучения для анализа пангеномов [Baeyer et al., 2021]. При этом число секвенированных действительно полных (в прямом смысле этого слова) диплоидных геномов будет непременно расти и в этом будут безусловно помогать опережающие сведения о C-value того или иного вида, поскольку их гораздо легче и быстрее получать.

Завершая статью о разнообразии количественного содержания ДНК в ядрах растений и их геномов, стоит уделить еще немного внимания используемым терминам. Так, в 1920 г. был предложен термин «геном», подразумевающий гаплоидный набор хромосом. Спустя годы, уже зная, что ДНК является веществом наследственности, появился термин «C-value», отражающий количество данного биополимера в ядре в пикограммах. Появившаяся возможность секвенирования протяженных фрагментов ДНК через некоторое время позволила определять полные квазигаплоидные геномы различных организмов, включая растения, используя специализированную единицу измерения их размеров – пары нуклеотидов или п.н. В 2005 г. в оборот был введен термин «пангеном», который был призван нести информацию не о геномах отдельных организмов (образцов), а их некоей совокупности, учитывая индивидуальные различия как частность, но отражая общность. Однако развитие технологий секвенирования ДНК позже привело к возможности определять полные геномы организмов по-хромосомно, в том числе в формате T2T (от теломеры до теломеры), а также устанавливать их фазированное состояние в виде гаплотипов или иначе диплоидных геномов, что пока довольно трудно достижимо. При этом нет сомнений, что дальнейшее совершенствование методов секвенирования ДНК обеспечит восстановление действительно полных геномов, которыми должны считаться только диплоидные, причем в формате T2T и для них потребуются новые обозначения, чтобы отличать от нынешних квазигаплоидных геномов, фактически лишь отражающих общую структурную организацию генома конкретного вида и позволяющих наблюдать синтению родственных геномов и их эволюцию. При этом они имеют весьма ограниченную ценность для селекционной работы, не говоря уже о медицине. Таким обозначением для истинно полных геномов, по нашему мнению, мог бы быть термин «дигеном» или более коротко «дином». Соответственно должен будет появиться и термин «пандином».

Еще одной причиной внедрения для высших организмов нового термина «дином», вместо конкретизации для них полностью секвенированного генома как диплоидного, служит еще и то, что,

¹⁷ хотя мы говорим об этом уже довольно давно

например, у бактерий и архей под термином «геном» понимается совокупность всей ДНК (кроме внехромосомных элементов) клетки и никакого другого статуса (гаплоидного, диплоидного) для них быть не может.

Работа выполнена в рамках государственных заданий №№ 122030200143-8, 122041400162-3, 122041400169-2, гранта Минобрнауки РФ (соглашение № 075-15-2021-1066 от 28 сентября 2021 г).

Литература

1. Болховских З.В., Гриф В.Г., Захарьева О.И., Матвеева Т.С. Хромосомные числа цветковых растений / Под ред. А.Н. Федорова. Л.: Наука, 1969. 920 с.
2. Гаршин М.В., Картуха А.И., Кулуев Б.Р. Коксагыз: особенности культивирования, перспективы возделывания и внедрения в современное производство // *Biomics*. 2016. Том 8, № 4. С. 323-333
3. Гаршин М.В., Кулуев Б.Р. Крым-сагыз: Особенности растения, перспективы возделывания и селекция (Обзор) // *Аграрная Россия*. 2018. №4, С.40-48
4. Зеленин А.В., Родионов А.В., Большева Н.Л. Бадаева Е.Д., Муравенко О.В. Истоки «генома»: происхождение и эволюция термина // *Молекулярная биология*. 2016. Т.50(4). С.611-620. DOI: 10.7868/S0026898416040170
5. 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.
6. 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
7. Кулуев Б.Р., Гарафутдинов Р.Р., Максимов И.В., Сагитов А.М., Чемерис Д.А., Князев А.В., Вершинина З.Р., Баймиев Ан.Х., Мулдашев А.А., Баймиев Ал.Х., Чемерис А.В. Натуральный каучук, его источники и составные части // *Биомика*. 2015. Т. 7(4). С. 224-283.
8. Кулуев Б.Р., Картуха А.И., Князев А.В., Фатерыга А.В., Чемерис А.В. Опыт выращивания *Taraxacum hybernum* (Asteraceae) // *Растительные ресурсы*. 2017. №4. С.543-554.
9. Поддубная-Арнольди В., Дианова В. Характер размножения некоторых каучуконосных и не каучуконосных видов рода *Taraxacum* L. // *Ботанический журнал СССР*. 1937. Т.22, №3. С.267-295
10. Патрушев Л.И., Минкевич И.Г. Проблема размера геномов эукариот // *Успехи биологической химии*. 2007. Т.47. С. 293–370.
11. Сагитов А.М., Золкин С.Ю., Кулуев Б.Р., Гималов Ф.Р., Князев А.В., Чемерис А.В. Кастилла эластичная (*Castilla elastica* Cerv.) – почти забытый каучуконос // *Biomics*. 2021. Т.13(2). С.106-137. DOI: 10.31301/2221-6197.bmcs.2021-9
12. Шереметьев С.Н., Гамалей Ю.В., Слезнев Н.Н. Направления эволюции генома покрытосеменных // *Цитология*. 2011. Т. 53. С. 295-311.
13. Цвелев Н.Н., Жукова П.Г. О наименьшем основном числе хромосом в сем. Роасеae // *Ботан. журнал*. 1974. Т. 59. С. 265–269.
14. Aganezov S., Yan S.M., Soto D.C. et al. A complete reference genome improves analysis of human genetic variation // *Science*. 2022. V.376(6588). eabl3533. doi: 10.1126/science.abl3533
15. Ahmad R., Parfitt D.E., Fass J., Ogundiwin E., Dhingra A., Gradziel T.M., Lin D., Joshi N.A., Martinez-Garcia P.J., Crisosto C.H. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection // *BMC Genomics*. 2011. V.12. P.569. doi: 10.1186/1471-2164-12-569
16. Alonge M., Shumate A., Puiu D., Zimin A.V., Salzberg S.L. Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies // *Genetics*. 2020. V.216(2). P.599-608. doi: 10.1534/genetics.120.303501
17. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* // *Nature*. 2000. V.408(6814). P.796-815. doi: 10.1038/35048692
18. Athiyannan N., Abrouk M., Boshoff W.H.P., Cauet S., Rodde N., Kudrna D., Mohammed N., Bettgenhaeuser J., Botha K.S., Derman S.S., Wing R.A., Prins R., Krattinger S.G. Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning // *Nat Genet*. 2022. V.54(3). P.227-231. doi: 10.1038/s41588-022-01022-1
19. Bari G., Gainullina K., Gumerova G., Uteulin K., Golovanov Ya., Chemeris A., Kuluev B. Multilocus DNA polymorphism of some rubber-bearing dandelions (*Taraxacum* spp.) of Russia and Kazakhstan // *Genet Resour Crop Evol*. 2022. V.69. No. 1. P.335.348. doi: 10.1007/s10722-021-01233-1
20. Bayer P.E., Golicz A.A., Scheben A., Batley J., Edwards D. Plant pan-genomes are the new reference // *Nat Plants*. 2020. V.6(8). P.914-920. doi: 10.1038/s41477-020-0733-0
21. Bayer P.E., Petereit J., Danilevicz M.F., Anderson R., Batley J., Edwards D. The application of pangonomics and machine learning in genomic selection

- in plants // *Plant Genome*. 2021. V.14(3). e20112. doi: 10.1002/tpg2.20112
22. Belser C., Baurens F.C., Noel B., Martin G., Cruaud C., Istace B., Yahiaoui N., Labadie K., Hřibová E., Doležel J., Lemainque A., Wincker P., D'Hont A., Aury J.M. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing // *Commun Biol*. 2021. V.4(1). P.1047. doi: 10.1038/s42003-021-02559-3
 23. Bennett M.D., Smith J.B. Nuclear DNA amounts in angiosperms // *Philos Trans R Soc Lond B Biol Sci*. 1976. V.274(933). P.227-274. doi: 10.1098/rstb.1976.0044
 24. Bevan M., Mayer K., White O., Eisen J.A., Preuss D., Bureau T., Salzberg S.L., Mewes H.W. Sequence and analysis of the Arabidopsis genome // *Curr Opin Plant Biol*. 2001. V.4(2). P.105-10. doi: 10.1016/s1369-5266(00)00144-8
 25. Brenchley R., Spannagl M., Pfeifer M., Barker G., D'Amore R., Allen A., et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing // *Nature*. 2012;491:705-710. DOI: 10.1038/nature11650
 26. Castiglione M.R., Cremonini R. A fascinating island: $2n = 4$ // *Plant Biosystems*. 2012. V.146(3). P.711-726. DOI: 10.1080/11263504.2012.714806
 27. Chapman V., Millert E., Riley R. Equivalence of the A genome of bread wheat and that of *Triticum urartu* // *Genet. Res.* 1976;27:69-76. DOI: 10.1017/S0016672300016244
 28. Commentary. Implications of the first complete human genome assembly // *Genome Res*. 2022. V32(4). P.595-598. doi: 10.1101/gr.276723.122
 29. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X, Wang J, Liu K, Qin P, Yang X, Zhu L, Li S, Liang C. Sequencing and de novo assembly of a near complete indica rice genome // *Nat Commun*. 2017. V.8. 15324. doi: 10.1038/ncomms15324
 30. De Bodt S., Maere S., Van de Peer Y. Genome duplication and the origin of angiosperms // *Trends Ecol Evol*. 2005. V.20(11). P.591-597. doi: 10.1016/j.tree.2005.07.008
 31. Della Coletta R., Qiu Y., Ou S., Hufford M.B., Hirsch C.N. How the pan-genome is changing crop genomics and improvement // *Genome Biol*. 2021. V.22(1). P.3. doi: 10.1186/s13059-020-02224-8
 32. Deng Y., Liu S., Zhang Y., Tan J., Li X., Chu X., Xu B., Tian Y., Sun Y., Li B., Xu Y., Deng X.W., He H., Zhang X. A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding // *Mol Plant*. 2022. V.22. S1674-2052(22)00192-7. doi: 10.1016/j.molp.2022.06.010
 33. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Müller KF, Heubl G. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms // *Ann Bot*. 2014 Dec;114(8):1651-63. doi: 10.1093/aob/mcu189
 34. Garg S, Functammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P, Hatas E, Ghurye J, Maguire J, Mahmoud M, Cheng H, Heller D, Zook JM, Moemke T, Marschall T, Sedlazeck FJ, Aach J, Chin CS, Church GM, Li H. Chromosome-scale, haplotype-resolved assembly of human genomes // *Nat Biotechnol*. 2021. V.39(3). P.309-312. doi: 10.1038/s41587-020-0711-0
 35. Goff S.A., Ricke D., Lan T.H. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*) // *Science*. 2002. V.296(5565). P.92-100. doi: 10.1126/science.1068275
 36. Golicz A.A., Batley J., Edwards D. Towards plant pangenomics // *Plant Biotechnol J*. 2016. V.14(4). P.1099-1105. doi: 10.1111/pbi.12499
 37. Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size // *Plant Biol (Stuttg)*. 2006. V.8(6). P.770-777. doi: 10.1055/s-2006-924101
 38. Greilhuber J., Dolezel J. 2C or not 2C: a closer look at cell nuclei and their DNA content // *Chromosoma*. 2009. V.118(3). P.391-400. doi: 10.1007/s00412-009-0205-9
 39. Greilhuber J., Dolezel J., Lysák M.A., Bennett M.D. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents // *Ann Bot*. 2005. V.95(1). P.255-260. doi: 10.1093/aob/mci019
 40. Hidalgo O., Pellicer J., Christenhusz M.J.M., Schneider H., Leitch I.J. Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica* // *Botanical Journal of the Linnean Society*. 2017. V.183(4). P.509–514. doi: 10.1093/botlinnean/box003
 41. Hidalgo O., Pellicer J., Christenhusz M., Schneider H., Leitch A.R., Leitch I.J. Is There an Upper Limit to Genome Size? // *Trends Plant Sci*. 2017. V.22(7). P.567-573. doi: 10.1016/j.tplants.2017.04.005
 42. Ibarra-Laclette E., Lyons E., Hernández-Guzmán G. et al. Architecture and evolution of a minute plant genome. *Nature*. 2013. V.498(7452). P.94-98. doi: 10.1038/nature12132
 43. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome // *Science*. 2018. V.361(6403). earr7191. doi: 10.1126/science.earr7191
 44. Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ. Evolution of genome size in Brassicaceae // *Ann Bot*. 2005. V.95(1). P.229-235. doi: 10.1093/aob/mci016

45. Johnson M.A.T., Kenton A.Y., Bennett M.D., Brandham P.E. *Voanioala gerardii* has the highest known chromosome number in the monocotyledons // *Genome*. 1989. V. 32(2). P.328-333. doi: 10.1139/g89-449
46. Khandelwal S. Chromosome evolution in the genus *Ophioglossum* L. // *Botanical Journal of the Linnean Society*. 1990. V.102(3). P 205–217. doi: 10.1111/j.1095-8339.1990.tb01876.x
47. Kelly L.J., Renny-Byfield S., Pellicer J., Macas J., Novák P., Neumann P., Lysak M.A., Day P.D., Berger M., Fay M.F., Nichols R.A., Leitch A.R., Leitch I.J. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size // *New Phytol*. 2015. V.208(2). P.596-607. doi: 10.1111/nph.13471
48. Koren S., Rhie A., Walenz B.P., Diltchey 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. V22:10.1038/nbt.4277. doi: 10.1038/nbt.4277
49. Kress W.J., Soltis D.E., Kersey P.J., Wegrzyn J.L., Leebens-Mack J.H., Gostel M.R., Liu X., Soltis P.S. Green plant genomes: What we know in an era of rapidly expanding opportunities // *Proc Natl Acad Sci USA*. 2022. V.119(4). e2115640118. doi: 10.1073/pnas.2115640118
50. Kuzmin D.A., Feranchuk S.I., Sharov V.V., Cybin A.N., Makolov S.V., Putintseva Y.A., Oreshkova N.V., Krutovsky K.V. Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb) // *BMC Bioinformatics*. 2019. V.20 (Suppl 1). P.37. doi: 10.1186/s12859-018-2570-y
51. Lam H.M., Xu X., Liu X. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection // *Nat Genet*. 2010. V.42. P.1053–1059. doi: 10.1038/ng.715
52. Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF. Punctuated genome size evolution in Liliaceae // *J Evol Biol*. 2007. V.20(6). P.2296-308. doi: 10.1111/j.1420-9101.2007.01416.x
53. Leushkin E.V., Sutormin R.A., Nabieva E.R., Penin A.A., Kondrashov A.S., Logacheva M.D. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences // *BMC Genomics*. 2013. V.14. P.476. doi: 10.1186/1471-2164-14-476
54. Li W, Liu J, Zhang H, Liu Z, Wang Y, Xing L, He Q, Du H. Plant pan-genomics: recent advances, new challenges, and roads ahead // *J Genet Genomics*. 2022. V.21. S1673-8527(22)00162-X. doi: 10.1016/j.jgg.2022.06.004
55. Liu J., Seetharam A.S., Chougule K. et al. Gapless assembly of maize chromosomes using long-read technologies // *Genome Biol*. 2020. V.21(1). P.121. doi: 10.1186/s13059-020-02029-9
56. Lupski J.R. Biology in balance: human diploid genome integrity, gene dosage, and genomic medicine // *Trends Genet*. 2022. V.38(6). P.554-571. doi: 10.1016/j.tig.2022.03.001
57. Menzel M.Y. A Cytological Method for Genome Analysis in *Gossypium*. // *Genetics*. 1955. V.40. P.214-223.
58. Miao X., Yu Y., Zhao Z., Wang Y., Qian X., Wang Y., Li S., Wang C. Chromosome-Level Haplotype Assembly for *Equus asinus* // *Front Genet*. 2022. V.13. 738105. doi: 10.3389/fgene.2022.738105
59. Michael T.P., VanBuren R. Building near-complete plant genomes // *Curr Opin Plant Biol*. 2020. V.54. P.26-33. doi: 10.1016/j.pbi.2019.12.009
60. Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. The pangenome of hexaploid bread wheat // *Plant J*. 2017. V.90(5). P.1007-1013. doi: 10.1111/tpj.13515
61. Morton N.E. Parameters of the human genome // *Proc. Natl. Acad. Sci. USA*. 1991. V. 88. P. 7474-7476. doi: 10.1073/pnas.88.17.7474
62. Navrátilová P., Toegelová H., Tulpová Z., Kuo Y.T., Stein N., Doležel J., Houben A., Šimková H., Mascher M. Prospects of telomere-to-telomere assembly in barley: Analysis of sequence gaps in the MorexV3 reference genome // *Plant Biotechnol J*. 2022. V.20(7). P.1373-1386. doi: 10.1111/pbi.13816
63. Neale D.B., Zimin A.V., Zaman S. et al. Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin // *G3 (Bethesda)*. 2022. V.12(1). jkab380. doi: 10.1093/g3journal/jkab380
64. Niu S., Li J., Bo W. et al. The Chinese pine genome and methylome unveil key features of conifer evolution // *Cell*. 2022. V.185(1). P.204-217.e14. doi: 10.1016/j.cell.2021.12.006.
65. 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
66. Nurk S., Koren S., Rhie A. et al. The complete sequence of a human genome // *Science*. 2022. V.376(6588). P.44-53. doi: 10.1126/science.abj6987
67. Nystedt B., Street N.R., Wetterbom A. et al. The Norway spruce genome sequence and conifer genome evolution // *Nature*. 2013. V.497(7451). P.579-584. doi: 10.1038/nature12211
68. Oginuma K., Munzinger J., Tobe H. Exceedingly high chromosome number in Strasburgeriaceae, a monotypic family endemic to New Caledonia // *Plant*

- Syst. Evol. 2006. V.262. P. 97–101. doi: 10.1007/s00606-006-0451-8
69. Ogur M., Erickson R.O., Rosen G.U., Sax K.B., Holden C. Nucleic acids in relation to cell division in *Lilium longiflorum* // Experimental Cell Research. 1951. V.2(1). P.73-89. doi: 10.1016/0014-4827(51)90007-9
70. Padgitt-Cobb L.K., Kingan S.B., Wells J., Elser J., Kronmiller B., Moore D., Concepcion G., Peluso P., Rank D., Jaiswal P., Henning J., Hendrix D.A. A draft phased assembly of the diploid Cascade hop (*Humulus lupulus*) genome // Plant Genome. 2021. V.14(1). e20072. doi: 10.1002/tpg2.20072
71. Pellicer J., Fay M.F., Leitch I.J. The largest eukaryotic genome of them all? // Botanical Journal of the Linnean Society. 2010. V.164(1). P.10–15. doi: 10.1111/j.1095-8339.2010.01072.x
72. Pellicer J, Leitch IJ. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies // New Phytol. 2020 Apr;226(2):301-305. doi: 10.1111/nph.16261
73. Qi W., Lim Y.W., Patrignani A., Schlöpfer P., Bratus-Neuenschwander A., Grüter S., Chanez C., Rodde N., Prat E., Vautrin S., Fustier M.A., Pratas D., Schlapbach R., Gruissem W. The haplotype-resolved chromosome pairs of a heterozygous diploid African cassava cultivar reveal novel pan-genome and allele-specific transcriptome features // Gigascience. 2022. V.11. giac028. doi: 10.1093/gigascience/giac028
74. Rice A., Glick L., Abadi S., Einhorn M., Kopelman N.M., Salman-Minkov A., Mayzel J., Chay O., Mayrose I. The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers // New Phytol. 2015. V.206(1). P.19-26. doi: 10.1111/nph.13191
75. Ris H., Mirsky A.E. Quantitative cytochemical determination of desoxyribonucleic acid with the Feulgen nuclear reaction // J Gen Physiol. 1949. V.33(2). P.125-146. doi: 10.1085/jgp.33.2.125
76. Röser M. Mitosis and Interphase of the Highly Polyploid Palm *Voanioala gerardii* (2n = 606 ± 3) // Cytogenet Genome Res. 2015;147(1):70-9. doi: 10.1159/000441677
77. Scott A.D., Zimin A.V., Puiu D., Workman R., Britton M., Zaman S., Caballero M., Read A.C., Bogdanove A.J., Burns E., Wegrzyn J., Timp W., Salzberg S.L., Neale D.B. A Reference Genome Sequence for Giant Sequoia // G3 (Bethesda). 2020. V.10(11). P.3907-3919. doi: 10.1534/g3.120.401612
78. Silva S.R., Moraes A.P., Penha H.A., Julião M.H.M., Domingues D.S., Michael T.P., Miranda V.F.O., Varani A.M. The Terrestrial Carnivorous Plant *Utricularia reniformis* Sheds Light on Environmental and Life-Form Genome Plasticity // Int J Mol Sci. 2019. V.21(1). P.3. doi: 10.3390/ijms21010003
79. Soltis P.S., Marchant D.B., Van de Peer Y., Soltis D.E. Polyploidy and genome evolution in plants // Curr Opin Genet Dev. 2015. V.35. P.119-125. doi: 10.1016/j.gde.2015.11.003
80. 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
81. Stevens K.A., Wegrzyn J.L., Zimin A., Puiu D., Crepeau M., Cardeno C., Paul R., Gonzalez-Ibeas D., Koriabine M., Holtz-Morris A.E., Martínez-García P.J., Sezen U.U., Marçais G., Jermstad K., McGuire P.E., Loopstra C.A., Davis J.M., Eckert A., de Jong P., Yorke J.A., Salzberg S.L., Neale D.B., Langley C.H. Sequence of the Sugar Pine Megagenome // Genetics. 2016. V.204(4). P.1613-1626. doi: 10.1534/genetics.116.193227
82. Swift H. The constancy of desoxyribose nucleic acid in plant nuclei // Proc. Natl. Acad. Sci. USA. 1950. V. 36. P.643–654. doi: 10.1073/pnas.36.11.643
83. Sun Y., Shang L., Zhu Q.H., Fan L., Guo L. Twenty years of plant genome sequencing: achievements and challenges // Trends Plant Sci. 2022. V.27(4). P.391-401. doi: 10.1016/j.tplants.2021.10.006
84. te Beest M., Le Roux J.J., Richardson D.M., Brysting A.K., Suda J., Kubesová M., Pysek P. The more the better? The role of polyploidy in facilitating plant invasions // Ann Bot. 2012. V.109(1). P.19-45. doi: 10.1093/aob/mcr277
85. Tettelin H., Massignani V., Cieslewicz M.J. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome" // Proc Natl Acad Sci USA. 2005. V.102(39). P.13950-13955. doi: 10.1073/pnas.0506758102
86. Thomas C.A, Jr. The genetic organization of chromosomes // Annual Rev. Genet. 1971. V. 5. P. 237-256. doi: 10.1146/annurev.ge.05.120171.001321
87. Tindale M.D., Roy S.K. A cytotoxic survey of the Pteridophyta of Australia // Australian Systematic Botany. 2002. V.15. P.839-937.
88. Twyford A.D. The road to 10,000 plant genomes // Nat Plants. 2018. V.4(6). P.312-313. doi: 10.1038/s41477-018-0165-2
89. Veleba A, Bureš P, Adamec L, Šmarda P, Lipnerová I, Horová L. Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae // New Phytol. 2014. V.203(1). P.22-28. doi: 10.1111/nph.12790
90. Wettstein Fr. von. Ueber plasmatische Vererbung, sowie Plasma- und Genwirkung. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen,

- Mathematisch-Physikalische Klasse. 1926. V.1926. P.250-281.
91. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. 1920. Jena: Gustav Fischer Verlag. 250 s.
92. Xu S., He Z., Zhang Z., Guo Z., Guo W., Lyu H., Li J., Yang M., Du Z., Huang Y., Zhou R., Zhong C., Boufford D.E., Lerdaun M., Wu C.I., Duke N.C.; International Mangrove Consortium, Shi S. The origin, diversification and adaptation of a major mangrove clade (Rhizophoraceae) revealed by whole-genome sequencing // *Natl. Sci. Rev.* 2017. V.4(5). P.721-734. doi: 10.1093/nsr/nwx065
93. Yu J., Hu S., Wang J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) // *Science*. 2002. V.296(5565). P.79-92. doi: 10.1126/science.1068037
94. Zhang Y., Fu J., Wang K., Han X., Yan T., Su Y., Li Y., Lin Z., Qin P., Fu C., Deng X.W., Zhou D., Yang Y., He H. The telomere-to-telomere gap-free genome of four rice parents reveals SV and PAV patterns in hybrid rice breeding // *Plant Biotechnol J.* 2022. doi: 10.1111/pbi.13880
95. Zimin A.V., Puiu D., Hall R., Kingan S., Clavijo B., Salzberg S. 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
96. Zimin A.V., Stevens K.A., Crepeau M.W., Puiu D., Wegrzyn J.L., Yorke J.A., Langley C.H., Neale D.B., Salzberg S.L. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing // *Gigascience*. 2017. V.6(1). P.1-4. doi: 10.1093/gigascience/giw016
97. Zonneveld B.J.M. New Record Holders for Maximum Genome Size in Eudicots and Monocots // *Journal of Botany*. 2010. V.2010. ID 527357. doi: 10.1155/2010/527357
- References**
1. Aganezov S., Yan S.M., Soto D.C. et al. A complete reference genome improves analysis of human genetic variation. *Science*. 2022. V.376(6588). eabl3533. doi: 10.1126/science.abl3533
2. Ahmad R., Parfitt D.E., Fass J., Ogundiwin E., Dhingra A., Gradziel T.M., Lin D., Joshi N.A., Martinez-Garcia P.J., Crisosto C.H. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics*. 2011. V.12. P.569. doi: 10.1186/1471-2164-12-569
3. Alonge M., Shumate A., Puiu D., Zimin A.V., Salzberg S.L. Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. *Genetics*. 2020. V.216(2). P.599-608. doi: 10.1534/genetics.120.303501
4. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* *Nature*. 2000. V.408(6814). P.796-815. doi: 10.1038/35048692
5. Athiyannan N., Abrouk M., Boshoff W.H.P., Cauet S., Rodde N., Kudrna D., Mohammed N., Bettgenhaeuser J., Botha K.S., Derman S.S., Wing R.A., Prins R., Krattinger S.G. Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. *Nat Genet.* 2022. V.54(3). P.227-231. doi: 10.1038/s41588-022-01022-1
6. Bari G., Gainullina K., Gumerova G., Uteulin K., Golovanov Ya., Chemeris A., Kuluev B. Multilocus DNA polymorphism of some rubber-bearing dandelions (*Taraxacum* spp.) of Russia and Kazakhstan. *Genet Resour Crop Evol.* 2022. V.69. No. 1. P.335-348. doi: 10.1007/s10722-021-01233-1
7. Bayer P.E., Golicz A.A., Scheben A., Batley J., Edwards D. Plant pan-genomes are the new reference. *Nat Plants*. 2020. V.6(8). P.914-920. doi: 10.1038/s41477-020-0733-0
8. Bayer P.E., Petereit J., Danilevicz M.F., Anderson R., Batley J., Edwards D. The application of pangenomics and machine learning in genomic selection in plants. *Plant Genome*. 2021. V.14(3). e20112. doi: 10.1002/tpg2.20112
9. Belser C., Baurens F.C., Noel B., Martin G., Cruaud C., Istace B., Yahiaoui N., Labadie K., Hřibová E., Doležel J., Lemainque A., Wincker P., D'Hont A., Aury J.M. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Commun Biol*. 2021. V.4(1). P.1047. doi: 10.1038/s42003-021-02559-3
10. Bennett M.D., Smith J.B. Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B Biol Sci.* 1976. V.274(933). P.227-274. doi: 10.1098/rstb.1976.0044
11. Bevan M., Mayer K., White O., Eisen J.A., Preuss D., Bureau T., Salzberg S.L., Mewes H.W. Sequence and analysis of the Arabidopsis genome. *Curr Opin Plant Biol.* 2001. V.4(2). P.105-110. doi: 10.1016/s1369-5266(00)00144-8
12. Bolhovskih Z.V., Grif V.G., Zahar'eva O.I., Matveeva T.S. Hromosomnye chisla cvetkovykh rastenij / Pod red. A.N. Fedorova. L.: Nauka, 1969. 920 s. [Chromosome numbers of flowering plants] (In Russian)
13. Brenchley R., Spannagl M., Pfeifer M., Barker G., D'Amore R., Allen A., et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*. 2012;491:705-710. DOI: 10.1038/nature11650
14. Castiglione M.R., Cremonini R. A fascinating island: $2n = 4$. *Plant Biosystems*. 2012. V.146(3). P.711-726. DOI: 10.1080/11263504.2012.714806
15. Chapman V., Millert E., Riley R. Equivalence of the A genome of bread wheat and that of *Triticum urartu*.

- Genet. Res.* 1976;27:69-76. DOI: 10.1017/S0016672300016244
16. Commentary. Implications of the first complete human genome assembly. *Genome Res.* 2022. V32(4). P.595-598. doi: 10.1101/gr.276723.122
 17. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X, Wang J, Liu K, Qin P, Yang X, Zhu L, Li S, Liang C. Sequencing and de novo assembly of a near complete indica rice genome. *Nat Commun.* 2017 May 4;8:15324. doi: 10.1038/ncomms15324
 18. De Bodt S., Maere S., Van de Peer Y. Genome duplication and the origin of angiosperms. *Trends Ecol Evol.* 2005. V.20(11). P.591-597. doi: 10.1016/j.tree.2005.07.008
 19. Della Coletta R., Qiu Y., Ou S., Hufford M.B., Hirsch C.N. How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 2021. V.22(1). P.3. doi: 10.1186/s13059-020-02224-8
 20. Deng Y., Liu S., Zhang Y., Tan J., Li X., Chu X., Xu B., Tian Y., Sun Y., Li B., Xu Y., Deng X.W., He H., Zhang X. A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol Plant.* 2022. V.22. S1674-2052(22)00192-7. doi: 10.1016/j.molp.2022.06.010
 21. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Tensch EM, Greilhuber J, Müller KF, Heubl G. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann Bot.* 2014 Dec;114(8):1651-63. doi: 10.1093/aob/mcu189
 22. Garg S, Functammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P, Hatas E, Ghurye J, Maguire J, Mahmoud M, Cheng H, Heller D, Zook JM, Moemke T, Marschall T, Sedlazeck FJ, Aach J, Chin CS, Church GM, Li H. Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol.* 2021 Mar;39(3):309-312. doi: 10.1038/s41587-020-0711-0. Epub 2020 Dec 7. PMID: 33288905; PMCID: PMC7954703.
 23. Garshin M.V., Kartuha A.I., Kuluev B.R. Kok-sagyz: osobennosti kul'tivirovaniya, perspektivy vozdeleyvaniya i vnedreniya v sovremennoe proizvodstvo. *Biomics.* 2016. T. 8(4). S. 323-333. [Kok-sagyz: features of cultivation, prospects of cultivation and introduction into modern production] (In Russian)
 24. Garshin M.V., Kuluev B.R. Krym-sagyz: Osobennosti rasteniya, perspektivy vozdeleyvaniya i selekciya (Obzor). *Agrarnaya Rossiya.* 2018. №4, S.40-48. [Krym-saghyz: Plant features, prospects of cultivation and selection (Review)] (In Russian)
 25. Goff S.A., Ricke D., Lan T.H. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science.* 2002. V.296(5565). P.92-100. doi: 10.1126/science.1068275
 26. Golicz A.A., Batley J., Edwards D. Towards plant pangenomics. *Plant Biotechnol J.* 2016. V.14(4). P.1099-1105. doi: 10.1111/pbi.12499
 27. Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biol (Stuttg).* 2006. V.8(6). P.770-777. doi: 10.1055/s-2006-924101
 28. Greilhuber J., Dolezel J. 2C or not 2C: a closer look at cell nuclei and their DNA content. *Chromosoma.* 2009. V.118(3). P.391-400. doi: 10.1007/s00412-009-0205-9
 29. Greilhuber J., Dolezel J., Lysák M.A., Bennett M.D. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Ann Bot.* 2005. V.95(1). P.255-260. doi: 10.1093/aob/mci019
 30. Hidalgo O., Pellicer J., Christenhusz M.J.M., Schneider H., Leitch I.J. Genomic gigantism in the whisker-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Botanical Journal of the Linnean Society.* 2017. V.183(4). P.509–514. doi: 10.1093/botlinnean/box003
 31. Hidalgo O., Pellicer J., Christenhusz M., Schneider H., Leitch A.R., Leitch I.J. Is There an Upper Limit to Genome Size? *Trends Plant Sci.* 2017. V.22(7). P.567-573. doi: 10.1016/j.tplants.2017.04.005
 32. Ibarra-Laclette E., Lyons E., Hernández-Guzmán G. et al. Architecture and evolution of a minute plant genome. *Nature.* 2013. V.498(7452). P.94-98. doi: 10.1038/nature12132
 33. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 2018. V.361(6403). eaar7191. doi: 10.1126/science.aar7191
 34. Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ. Evolution of genome size in Brassicaceae. *Ann Bot.* 2005. V.95(1). P.229-235. doi: 10.1093/aob/mci016
 35. Johnson M.A.T., Kenton A.Y., Bennett M.D., Brandham P.E. *Voanioala gerardii* has the highest known chromosome number in the monocotyledons // *Genome.* 1989. V. 32(2). P.328-333. doi: 10.1139/g89-449
 36. Khandelwal S. Chromosome evolution in the genus *Ophioglossum* L. *Botanical Journal of the Linnean Society.* 1990. V.102(3). P 205–217. doi: 10.1111/j.1095-8339.1990.tb01876.x
 37. Kelly L.J., Renny-Byfield S., Pellicer J., Macas J., Novák P., Neumann P., Lysak M.A., Day P.D., Berger M., Fay M.F., Nichols R.A., Leitch A.R., Leitch I.J. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes

- extreme expansions in genome size. *New Phytol.* 2015. V.208(2). P.596-607. doi: 10.1111/nph.13471
38. 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. V22:10.1038/nbt.4277. doi: 10.1038/nbt.4277
39. Kress W.J., Soltis D.E., Kersey P.J., Wegrzyn J.L., Leebens-Mack J.H., Gostel M.R., Liu X., Soltis P.S. Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proc Natl Acad Sci USA.* 2022. V.119(4). e2115640118. doi: 10.1073/pnas.2115640118
40. Kuluev A.R., Matniyazov R.T., Chemeris D.A., Chemeris A.V. Modern concepts about relationships in the wheat-aegilops alliance (with a brief historical note). *Biomics.* 2016. V.8(4). P. 297-310. (In Russian)
41. 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)
42. Kuluev B.R., Garafutdinov R.R., Maksimov I.V., Sagitov A.M., Chemeris D.A., Knyazev A.V., Vershinina Z.R., Baymiev An.K., Muldashev A.A., Baymiev Al.K. Chemeris A.V. Natural rubber, its sources and components. *Biomics.* 2017. V. 7. P. 224-283 (In Russian)
43. Kuluev B.R., Kartuha A.I., Knjazev A.V., Fateryga A.V., Chemeris A.V. Opyt vyrashhivaniya *Taraxacum hybernum* (Asteraceae). *Rastitel'nye resursy.* 2017. №4. S.543-554. [Kuluyev B.R., Kartukha A.I., Knyazev A.V., Fateryga A.V., Chemeris A.V. Experience of growing *Taraxacum hybernum* (Asteraceae)] (In Russian)
44. Kuzmin D.A., Feranchuk S.I., Sharov V.V., Cybin A.N., Makolov S.V., Putintseva Y.A., Oreshkova N.V., Krutovsky K.V. Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb). *BMC Bioinformatics.* 2019. V.20 (Suppl 1). P.37. doi: 10.1186/s12859-018-2570-y
45. Lam H.M., Xu X., Liu X. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet.* 2010. V.42. P.1053-1059. doi: 10.1038/ng.715
46. Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF. Punctuated genome size evolution in Liliaceae. *J Evol Biol.* 2007 Nov;20(6):2296-308. doi: 10.1111/j.1420-9101.2007.01416.x
47. Leushkin E.V., Sutormin R.A., Nabieva E.R., Penin A.A., Kondrashov A.S., Logacheva M.D. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics.* 2013. V.14. P.476. doi: 10.1186/1471-2164-14-476
48. Li W, Liu J, Zhang H, Liu Z, Wang Y, Xing L, He Q, Du H. Plant pan-genomics: recent advances, new challenges, and roads ahead. *J Genet Genomics.* 2022 Jun 21:S1673-8527(22)00162-X. doi: 10.1016/j.jgg.2022.06.004
49. Liu J., Seetharam A.S., Chougule K. et al. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol.* 2020. V.21(1). P.121. doi: 10.1186/s13059-020-02029-9
50. Lupski J.R. Biology in balance: human diploid genome integrity, gene dosage, and genomic medicine. *Trends Genet.* 2022. V.38(6). P.554-571. doi: 10.1016/j.tig.2022.03.001
51. Menzel M.Y. A Cytological Method for Genome Analysis in *Gossypium*. *Genetics.* 1955. V.40. P.214-223.
52. Miao X., Yu Y., Zhao Z., Wang Y., Qian X., Wang Y., Li S., Wang C. Chromosome-Level Haplotype Assembly for *Equus asinus*. *Front Genet.* 2022. V.13. 738105. doi: 10.3389/fgene.2022.738105
53. Michael T.P., VanBuren R. Building near-complete plant genomes. *Curr Opin Plant Biol.* 2020. V.54. P.26-33. doi: 10.1016/j.pbi.2019.12.009
54. Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. The pangenome of hexaploid bread wheat. *Plant J.* 2017. V.90(5). P.1007-1013. doi: 10.1111/tbj.13515
55. Morton N.E. Parameters of the human genome. *Proc. Natl. Acad. Sci. USA.* 1991. V. 88. P. 7474-7476. doi: 10.1073/pnas.88.17.7474
56. Navrátilová P., Toegelová H., Tulpová Z., Kuo Y.T., Stein N., Doležel J., Houben A., Šimková H., Mascher M. Prospects of telomere-to-telomere assembly in barley: Analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnol J.* 2022. V.20(7). P.1373-1386. doi: 10.1111/pbi.13816
57. Neale D.B., Zimin A.V., Zaman S. et al. Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. *G3 (Bethesda).* 2022. V.12(1). jkab380. doi: 10.1093/g3journal/jkab380
58. Niu S., Li J., Bo W. et al. The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell.* 2022. V.185(1). P.204-217.e14. doi: 10.1016/j.cell.2021.12.006.
59. Noguera-Solano R., Ruiz-Gutierrez R., Rodriguez-Caso J.M. Genome: twisting stories with DANN. *Endeavour.* 2013. V.37(4). P.213-219. doi: 10.1016/j.endeavour.2013.05.003
60. Nurk S., Koren S., Rhie A. et al. The complete sequence of a human genome. *Science.* 2022. V.376(6588). P.44-53. doi: 10.1126/science.abj6987

61. Nystedt B., Street N.R., Wetterbom A. et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013. V.497(7451). P.579-584. doi: 10.1038/nature12211
62. Oginuma K., Munzinger J., Tobe H. Exceedingly high chromosome number in Strasburgeriaceae, a monotypic family endemic to New Caledonia. *Plant Syst. Evol.* 2006. V.262. P. 97–101. doi: 10.1007/s00606-006-0451-8
63. Ogur M., Erickson R.O., Rosen G.U., Sax K.B., Holden C. Nucleic acids in relation to cell division in *Lilium longiflorum* // *Experimental Cell Research*. 1951. V.2(1). P.73-89. doi: 10.1016/0014-4827(51)90007-9
64. Padgitt-Cobb L.K., Kingan S.B., Wells J., Elser J., Kronmiller B., Moore D., Concepcion G., Peluso P., Rank D., Jaiswal P., Henning J., Hendrix D.A. A draft phased assembly of the diploid Cascade hop (*Humulus lupulus*) genome. *Plant Genome*. 2021. V.14(1). e20072. doi: 10.1002/tpg2.20072
65. Patrushev L.I., Minkevich I.G. Problema razmera genomov jeukariot. *Uspehi biologicheskoy himii*. 2007. T.47. S. 293–370. [The problem of the size of eukaryotic genomes] (In Russian)
66. Pellicer J., Fay M.F., Leitch I.J. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society*. 2010. V.164(1). P.10–15. doi: 10.1111/j.1095-8339.2010.01072.x
67. Pellicer J, Leitch IJ. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytol.* 2020 Apr;226(2):301-305. doi: 10.1111/nph.16261
68. Poddubnaja-Arnol'di V., Dianova V. Harakter razmnozhenija nekotoryh kauchukonosnyh i ne kauchukonosnyh vidov roda *Taraxacum* L. *Botanicheskij zhurnal SSSR*. 1937. T.22, №3. S.267-295 [The character reproduction of some rubber-bearing and non-rubber-bearing species of the genus *Taraxacum* L.] (In Russian)
69. Qi W., Lim Y.W., Patrignani A., Schläpfer P., Bratus-Neuenschwander A., Grüter S., Chanez C., Rodde N., Prat E., Vautrin S., Fustier M.A., Pratas D., Schlapbach R., Gruissem W. The haplotype-resolved chromosome pairs of a heterozygous diploid African cassava cultivar reveal novel pan-genome and allele-specific transcriptome features. *Gigascience*. 2022. V.11. giac028. doi: 10.1093/gigascience/giac028
70. Rice A., Glick L., Abadi S., Einhorn M., Kopelman N.M., Salman-Minkov A., Mayzel J., Chay O., Mayrose I. The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *New Phytol.* 2015. V.206(1). P.19-26. doi: 10.1111/nph.13191
71. Ris H., Mirsky A.E. Quantitative cytochemical determination of desoxyribonucleic acid with the Feulgen nuclear reaction. *J Gen Physiol.* 1949. V.33(2). P.125-146. doi: 10.1085/jgp.33.2.125
72. Röser M. Mitosis and Interphase of the Highly Polyploid Palm *Voanioala gerardii* (2n = 606 ± 3). *Cytogenet Genome Res.* 2015;147(1):70-9. doi: 10.1159/000441677
73. Sagitov A.M., Zolkin S.Yu., Kuluev B.R., Gimalov F.R., Knyazev A.V., Chemeris A.V. *Castilla elastica* Cerv. is almost forgotten rubber-bearing plant. *Biomics*. 2021. V.13(2). P. 106-137. DOI: 10.31301/2221-6197.bmcs.2021-9 (In Russian)
74. Scott A.D., Zimin A.V., Puiu D., Workman R., Britton M., Zaman S., Caballero M., Read A.C., Bogdanove A.J., Burns E., Wegrzyn J., Timp W., Salzberg S.L., Neale D.B. A Reference Genome Sequence for Giant Sequoia. *G3 (Bethesda)*. 2020. V.10(11). P.3907-3919. doi: 10.1534/g3.120.401612
75. Sheremet'ev S.N., Gamalej Ju.V., Sleznev N.N. Napravlenija jevoljucii genoma pokrytosemennyh. *Citologija*. 2011. T. 53. S. 295-311. [Directions of angiosperm genome evolution] (In Russian)
76. Silva S.R., Moraes A.P., Penha H.A., Julião M.H.M., Domingues D.S., Michael T.P., Miranda V.F.O., Varani A.M. The Terrestrial Carnivorous Plant *Utricularia reniformis* Sheds Light on Environmental and Life-Form Genome Plasticity. *Int J Mol Sci*. 2019. V.21(1). P.3. doi: 10.3390/ijms21010003
77. Soltis P.S., Marchant D.B., Van de Peer Y., Soltis D.E. Polyploidy and genome evolution in plants. *Curr Opin Genet Dev*. 2015. V.35. P.119-125. doi: 10.1016/j.gde.2015.11.003
78. 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
79. Stevens K.A., Wegrzyn J.L., Zimin A., Puiu D., Crepeau M., Cardeno C., Paul R., Gonzalez-Ibeas D., Koriabine M., Holtz-Morris A.E., Martínez-García P.J., Sezen U.U., Marçais G., Jermstad K., McGuire P.E., Loopstra C.A., Davis J.M., Eckert A., de Jong P., Yorke J.A., Salzberg S.L., Neale D.B., Langley C.H. Sequence of the Sugar Pine Megagenome. *Genetics*. 2016. V.204(4). P.1613-1626. doi: 10.1534/genetics.116.193227
80. Swift H. The constancy of desoxyribose nucleic acid in plant nuclei. *Proc. Natl. Acad. Sci. USA*. 1950. V. 36. P.643–654. doi: 10.1073/pnas.36.11.643
81. Sun Y., Shang L., Zhu Q.H., Fan L., Guo L. Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci*. 2022. V.27(4). P.391-401. doi: 10.1016/j.tplants.2021.10.006
82. te Beest M., Le Roux J.J., Richardson D.M., Brysting A.K., Suda J., Kubesová M., Pysek P. The more

- the better? The role of polyploidy in facilitating plant invasions. *Ann Bot.* 2012. V.109(1). P.19-45. doi: 10.1093/aob/mcr277
83. Tettelin H., Massignani V., Cieslewicz M.J. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA.* 2005. V.102(39). P.13950-13955. doi: 10.1073/pnas.0506758102
84. Thomas C.A., Jr. The genetic organization of chromosomes. *Annual Rev. Genet.* 1971. V. 5. P. 237-256. doi: 10.1146/annurev.ge.05.120171.001321
85. Tindale M.D., Roy S.K. A cytotoxic survey of the Pteridophyta of Australia. *Australian Systematic Botany.* 2002. V.15. P.839-937.
86. Tsvelev N.N., Zhukova P.G. On the minimal main chromosome number in the family Poaceae. *Botanich. Zhurnal SSSR.* 1974. V.59. P.265-269.
87. Twyford A.D. The road to 10,000 plant genomes. *Nat Plants.* 2018. V.4(6). P.312-313. doi: 10.1038/s41477-018-0165-2
88. Veleba A, Bureš P, Adamec L, Šmarda P, Lipnerová I, Horová L. Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae. *New Phytol.* 2014. V.203(1). P.22-28. doi: 10.1111/nph.12790
89. Wettstein Fr. von. Ueber plasmatische Vererbung, sowie Plasma- und Genwirkung. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse. 1926. V.1926. P.250-281.
90. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. 1920. Jena: Gustav Fischer Verlag. 250 s.
91. Xu S., He Z., Zhang Z., Guo Z., Guo W., Lyu H., Li J., Yang M., Du Z., Huang Y., Zhou R., Zhong C., Boufford D.E., Lerdau M., Wu C.I., Duke N.C.; International Mangrove Consortium, Shi S. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing // *Natl. Sci. Rev.* 2017. V.4(5). P.721-734. doi: 10.1093/nsr/nwx065
92. Yu J., Hu S., Wang J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science.* 2002. V.296(5565). P.79-92. doi: 10.1126/science.1068037
93. Zelenin AV, Rodionov AV, Bolsheva NL, Badaeva ED, Muravenko OV. Genome: Origins and evolution of the term. *Mol Biol (Mosk).* 2016. V.50(4). P.611-620. doi: 10.7868/S0026898416040170
94. Zhang Y., Fu J., Wang K., Han X., Yan T., Su Y., Li Y., Lin Z., Qin P., Fu C., Deng X.W., Zhou D., Yang Y., He H. The telomere-to-telomere gap-free genome of four rice parents reveals SV and PAV patterns in hybrid rice breeding. *Plant Biotechnol J.* 2022. doi: 10.1111/pbi.13880
95. Zimin A.V., Puiu D., Hall R., Kingan S., Clavijo B., Salzberg S. 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
96. Zimin A.V., Stevens K.A., Crepeau M.W., Puiu D., Wegrzyn J.L., Yorke J.A., Langley C.H., Neale D.B., Salzberg S.L. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience.* 2017. V.6(1). P.1-4. doi: 10.1093/gigascience/giw016
97. Zonneveld B.J.M. New Record Holders for Maximum Genome Size in Eudicots and Monocots. *Journal of Botany.* 2010. V.2010. ID 527357. doi: 10.1155/2010/527357