



**ПОЛВЕКА БЫСТРЫМ МЕТОДАМ СЕКВЕНИРОВАНИЯ ДНК.  
РЕТРОСПЕКТИВА И ПЕРСПЕКТИВЫ.**

<sup>1</sup>Зубов В.В.\*, <sup>2</sup>Чемерис Д.А., <sup>3,4</sup>Алексеев Я.И., <sup>4</sup>Курочкин В.Е., <sup>5</sup>Василов Р.Г., <sup>6</sup>Гарафутдинов Р.Р., <sup>6</sup>Чемерис А.В.

<sup>1</sup>Институт теоретической и экспериментальной биофизики Российской академии наук, Россия, Пущино

<sup>2</sup>ООО «ГЕНВЕД», Россия, Москва

<sup>3</sup>ООО «НПФ «Синтол», Россия, Москва

<sup>4</sup>Институт аналитического приборостроения Российской академии наук, Россия, Санкт-Петербург

НИЦ «Курчатовский институт», Россия, Москва

<sup>6</sup>Институт биохимии и генетики Уфимского федерального исследовательского центра

Российской академии наук, Россия, Уфа

\*E-mail: [genseq@mail.ru](mailto:genseq@mail.ru)

**Резюме**

Статья посвящена 50-летию быстрых методов секвенирования ДНК, начиная с пионерной разработки Ф.Сэнгера в 1975 г. Прослежена эволюция технологий секвенирования от методов первого поколения (ферментативный метод Сэнгера и метода химической дегградации Максама-Гильберта) до высокопроизводительных методов, среди которых методы второго поколения (NGS – next generation sequencing), рассчитанные на массовое параллельное секвенирование (МПС) - пиросеквенирование, полупроводниковое секвенирование, лигазное секвенирование, флуоресцентное секвенирование с короткими прочтениями, а также мономолекулярные методы секвенирования, включающие флуоресцентное секвенирование с длинными прочтениями, а также нанопоровое секвенирование с ультрадлинными прочтениями. Особое внимание уделено переходу от квазигеномов к полным диплоидным геномам с фазированной сборкой гаплотипов в формате T2T без промежутков, что обеспечивает точное установление связи генотипа с фенотипом. Рассматриваются современные достижения, включая снижение стоимости секвенирования в миллионы раз и увеличение длины прочтения до миллионов пар нуклеотидов. Перспективы развития связаны с созданием еще более производительных технологий пятого поколения и широким внедрением фазированной сборки геномов. Статья подчеркивает революционное значение секвенирования ДНК для наук о Жизни и необходимость отказа от устаревших подходов в виде сборки квазигеномов. В статье отмечается, что геномная революция продолжается, и ее потенциал еще далеко не исчерпан.

**Ключевые слова:** секвенирование ДНК, метод Сэнгера, метод Максама-Гильберта, МПС методы, флуоресцентное секвенирование, мономолекулярное секвенирование, нанопоровое секвенирование, геном, квазигеном, диплоидный геном, фазированный геном, T2T геном, гаплотипированная сборка, биологический геном

**Цитирование:** Зубов В.В., Чемерис Д.А., Алексеев Я.И., Курочкин В.Е., Василов Р.Г., Гарафутдинов Р.Р., Чемерис А.В. Полвека быстрым методам секвенирования ДНК. Ретроспектива и перспективы // *Biomics*. 2025. V.17(1). С. 103 - 120. DOI: 10.31301/2221-6197.bmcs.2025-8

© Авторы

**HALF A CENTURY OF FAST DNA SEQUENCING METHODS; A RETROSPECTIVE AND PROSPECTS**

<sup>1</sup>Zubov V.V.\*, <sup>2</sup>Chemeris D.A., <sup>3,4</sup>Alexeev Ya.I., <sup>4</sup>Kurochkin V.E., <sup>5</sup>Vasilov R.G., <sup>6</sup>Garafutdinov R.R., <sup>6</sup>Chemeris A.V.

<sup>1</sup>Institute of Theoretical and Experimental Biophysics RAS, Pushchino, Russia

<sup>2</sup>GENVED LLC, Moscow, Russia

<sup>3</sup>Syntol LLC, Moscow, Russia

<sup>4</sup>Institute for Analytical Instrumentation RAS, Saint-Petersburg, Russia

<sup>5</sup>National Research Center “Kurchatov Institute”, Moscow, Russia

<sup>6</sup>Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, Russia, Ufa

\*E-mail: [genseq@mail.ru](mailto:genseq@mail.ru)

### Resume

The article is dedicated to the 50th anniversary of rapid DNA sequencing methods, starting with the pioneering accomplishment of F.Sanger in 1975. The evolution of sequencing technologies is traced here from first-generation methods (named after Sanger and Maxham-Gilbert) to high-performance next-generation sequencing technologies, including second-generation methods designed for massive parallelism (pyrosequencing, semiconductor sequencing, ligase sequencing, and short-read fluorescent sequencing), and monomolecular sequencing approaches, including method in the form of long-read fluorescent sequencing as well as ultra-long-read nanopore sequencing. Special attention is given to the transition from quasi-genomes to complete diploid genomes with phased assembly of haplotypes in the T2T format without gaps, which ensures accurate determination of a genotype–phenotype relationship. Modern achievements are examined too, including millionfold reductions in the cost of sequencing and an increase in reading length to millions of nucleotide pairs. Prospective developments are related to the creation of even more productive fifth-generation technologies and widespread adoption of phased genome assembly. The article highlights revolutionary importance of DNA sequencing for life sciences and the need to abandon outdated approaches such as quasi-genome assembly. The article states that the genomic revolution continues, and its potential is far from exhausted.

**Keywords:** DNA sequencing, Sanger method, Maxam-Gilbert method, NGS methods, fluorescent sequencing, monomolecular sequencing, nanopore sequencing, genome, quasi-genome, diploid genome, phased genome, T2T genome, haplotyped assembly, biological genome

**Citation:** Zubov V.V., Chemeris D.A., Alexeev Ya.I., Kurochkin V.E., Vasilov R.G., Garafutdinov R.R., Chemeris A.V. Half a century of fast DNA sequencing methods; a retrospective and prospects. *Biomics*. 2025. V.17(1). P. 103 - 120. DOI: 10.31301/2221-6197.bmcs.2025-8

### © The Authors

#### Краткое содержание

Полвека назад появился первый быстрый «плюс/минус» метод секвенирования ДНК, вскоре после которого были разработаны сразу два еще более быстрых метода, ставших именными – метод Максама-Гильберта и метод Сэнгера. Общими для них стало использование радиоактивности и применение такого передового подхода как высоковольтный электрофореза в полиакриламидном геле в денатурирующих условиях, позволяющий разделять одноцепочечные фрагменты ДНК, отличающиеся на один нуклеотид. Метод Максама-Гильберта был рассчитан на химическую деградацию существующих цепей ДНК, тогда как метод Сэнгера предполагал ферментативное построение новой цепи ДНК с использованием дидезокситерминаторов. При этом метод Максама-Гильберта целое десятилетие оставался основным и сдал свои позиции после внедрения в метод Сэнгера флуоресцентной метки вместе с автоматизацией процесса и некоторыми другими новшествами. Именно с помощью метода Сэнгера были секвенированы два первых генома человека, но из-за мозаичной коненсусной сборки и использования ДНК разных людей их правильнее считать квазигеномами (геномы, представленные в виде единичной последовательности ДНК, при наличии в диплоидной клетке человека двух аллельных состояний, то есть двух отличающихся между собой последовательностей ДНК). После завершения проекта «Геном человека» стало ясно, что для продолжения подобных исследований по секвенированию полных геномов высших организмов метод Сэнгера с его электрофоретическим разделением продуктов реакций принципиально не подходит и нужны новые методы секвенирования ДНК, окрещенные за рубежом как методы NGS (Next Generation Sequencing). Если в середине 1970-х гг. прорыв в секвенировании ДНК был обеспечен, в том числе за счет использования высоковольтного гель-электрофореза, то для

осуществления нового прорыва в середине 2000-х гг. потребовалось от него «уйти». В результате появился целый сонм новых методов секвенирования ДНК, среди которых первым было пиросеквенирование, а также лигазное, полупроводниковое, флуоресцентное секвенирование и некоторые другие, рассчитанные на массовый параллельный анализ. Появились также мономолекулярные методы с длинными прочтениями, как основанные на детекции флуоресценции, так и детектирующее изменение силы ионного тока в виде нанопорового секвенирования с его ультрадлинными прочтениями. Причем идея нанопорового секвенирования родилась еще в период господства метода Сэнгера в середине 1990-х гг., но ее воплощение затянулось на долгих 20 лет. Однако проверку временем выдержали лишь три основных подхода: флуоресцентное секвенирование с короткими прочтениями, а также упомянутые выше два метода мономолекулярного секвенирования. При этом по сравнению с 2001 г. за два с лишним десятилетия секвенирование уже подешевело почти в миллион раз, что на несколько порядков превзошло действующий в электронике закон Мура и, по всей видимости, это еще не предел. Так, нельзя исключать появление новых методов секвенирования, которые потенциально могут еще больше увеличить производительность секвенирования ДНК и снизить его стоимость, вплоть до 100 долларов на геном, схожий по размеру человеческому.

Значительное внимание уделено «правильному» геномному секвенированию, в результате которого производится сборка фазированного генома по гаплотипам с распределением нуклеотидных последовательностей по хромосомам от теломеры до теломеры без промежутков. И это наивысший уровень сборки геномов в настоящее время, который только и обеспечивает установление взаимосвязи генотипа с фенотипом, тогда как в массе своей собираемые ныне геномы с мозаичной консенсусной сборкой правильнее считать квазигеномами, никак не претендующими на то, чтобы считаться биологическим геномом.

Отмечена важнейшая роль секвенирования ДНК, в том числе полных геномов в становлении и развитии наук о Жизни.

### Введение

В исследованиях нуклеиновых кислот и ДНК в частности, есть три наиболее важных события. Это само открытие Ф.Мишером в 1869 г. данного биополимера, названного им «нуклеин», установление Дж.Уотсоном и Ф.Криком в 1953 г. вторичной структуры ДНК в виде двойной спирали<sup>1</sup>, а также предложенные в середине 1970-х гг. двумя группами ученых два весьма быстрых для того времени метода определения последовательностей азотистых оснований у довольно протяженных фрагментов молекул ДНК причем принципиально разными способами. Все остальные важные и даже важнейшие результаты, включая химический синтез олигонуклеотидов, разработка метода ПЦР, очень сильно изменившего всю биологическую науку и проникшего даже в целый ряд смежных дисциплин, по сути, вторичны, поскольку без знания нуклеотидных последовательностей ДНК они все были бы невозможны или ограниченно применимы. Безусловно, серьезнейшее ускорение дальнейшему изучению живых организмов и в первую очередь эукариот дало появление в наступившем столетии высокопроизводительных методов секвенирования ДНК новых поколений, относительно легко «читающих» сейчас целые геномы, но всегда наиболее важны первоочередные результаты, дающие начальный толчок. Собственно так было и с нуклеином, про который лишь спустя 75 лет (в 1944

г.) стало ясно, что он<sup>2</sup> является веществом наследственности. И двойная спираль ДНК потребовала уточнения, в частности по числу водородных связей между комплементарными азотистыми основаниями, так как в первом варианте спиральной структуры ДНК считалось (и довольно долго), что АТ- и GC-пары образуют по две водородных связи, а также обнаружение существования не только Уотсон-Криковских пар, но и образование в некоторых условиях Хугстиновских пар, в которых, кстати, между нуклеотидами G и C формируются две водородные связи.

Признаёмся, что в этот юбилейный год для секвенирования ДНК взяли на себя смелость и взвалили огромную ответственность кратко описать насколько возможно в хронологическом порядке основные события в секвенировании ДНК, главным образом методического плана, происходившие за полвека, а также изложить свое видение о геномном секвенировании прошлого, настоящего и будущего.

Итак, в 2025 г. исполняется 50 лет с момента появления разработанного Ф.Сэнгером первого «быстрого»<sup>3</sup> метода секвенирования ДНК, основанного на ферментативном построении Кленовским фрагментом ДНК полимеразы I *E.coli*

<sup>1</sup> и мы этим вопросам ранее уделили значительное внимание [Gerashchenkov et al., 2019]

<sup>2</sup> который уже стал к тому времени известен как ДНК во многом благодаря нашим соотечественникам – Ф.А.Левене и Е.С.Лондону [Гарафутдинов, Чемерис, 2019]

<sup>3</sup> причем это определение “rapid” было вынесено в заголовок

вновь синтезированной радиоактивно-меченой цепи ДНК с использованием в качестве праймера короткого декалигонуклеотида и последующего гидролиза образовавшегося биополимера под действием 3'→5'-экзонуклеазной активности ДНК полимеразы фага Т4 с помощью так называемой «плюс/минус системы» [Sanger, Coulson, 1975]. Принцип метода заключался (после подготовительного этапа, который здесь опускаем) в проведении реакций с одним образцом ДНК в 8 пробирках, в 4 из которых происходила терминация полимеризации цепи ДНК в определенных местах (где встречался комплементарный нуклеотид отсутствующему в смеси дНТФ – минус-вариант), а в других четырех пробирках построенные в присутствии всех четырех дНТФ цепи ДНК подвергались экзонуклеазному разрушению с 3'-конца до тех пор пока в цепи не встречался нуклеотид, присутствующий в растворе в качестве того или иного дНТФ (плюс-вариант), который Т4 ДНК полимеразы обратно включала в состав цепи, благодаря большей скорости своей полимеризующей активности, фактически останавливая дальнейший гидролиз ДНК. После завершения этих реакций смесь подвергалась (что явилось тогда передовым подходом) высоковольтному электрофорезу в полиакриламидном геле в денатурирующих условиях, обеспечивающим уверенное разделение фрагментов ДНК, отличающихся на один нуклеотид, и после этапа радиоавтографии появлялась возможность восстановить нуклеотидную последовательность анализируемого участка ДНК. Несмотря на всю громоздкость проводимых процедур это был серьезный прорыв в определении нуклеотидных последовательностей ДНК. В том же году была опубликована еще одна статья этого коллектива авторов, в которой сообщалось о секвенировании этим методом двух участков ДНК фага φX174 длиной 60 и 90 п.н. [Air et al., 1975]. Спустя пару лет этим непростым «плюс/минус» методом была определена почти полная (из 5386 п.н.) нуклеотидная последовательность этого фага [Sanger et al., 1977a].

В том же 1975 г. в СССР вышла большая обзорная статья Е.Д.Свердлова [Свердлов (Sverdlov), 1975], в которой весьма детально были рассмотрены существовавшие подходы к определению первичной структуры ДНК, включая разработки автора и его коллег [Sverdlov et al., 1972; 1973], как путем химической дегградации, так и с использованием различных ферментов, среди которых были фосфодиэстераза, полинуклеотидкиназа, терминальная трансфераза и ряд других. Отдельного внимания

заслуживает упоминание<sup>4</sup>, что исследуемый (секвенируемый) участок ДНК можно получать в виде копии за счет запрограммированного копирования с включением в синтезируемый продукт неких модификаций, что затем явилось и остается основой многих современных методов ферментативного секвенирования ДНК, хотя тогда та фраза в первую очередь относилась к методам химической дегградации. Также в том обзоре были сделаны прогнозы относительно будущего развития подходов к определению нуклеотидных последовательностей и описаны задачи важные для понимания функционирования генных систем, которые могут быть тем самым решены. Нужно заметить, что о секвенировании геномов каких-либо организмов слов не было.

В 1975 же году в СССР был опубликован перевод книги С.Манделеса под названием «Установление первичной структуры нуклеиновых кислот» [Манделес, 1975]. Оригинальное издание “Nucleic Acid Sequence Analysis” датировано 1972 г. [Mandel, 1972] и в нем описывались весьма непроизводительные без использования высоковольтного гель-электрофореза методы определения нуклеотидных последовательностей ДНК, к которой ввиду очень крупного размера данного биополимера никак не могли тогда подступиться. Поэтому неудивительно, что за несколько лет до этого известный биохимик Э.Чаргафф в одной из своих статей написал, что чтение нуклеотидной последовательности ДНК - это задача следующего столетия [Chargaff, 1968]. И он, как показала жизнь, сильно ошибся, но одновременно оказался и прав, поскольку новые быстрые методы секвенирования ДНК появились через довольно непродолжительное время, но настоящий прорыв в этой области наступил лишь в середине 2000-х гг. и, как, ни странно, за счет отказа от одно время передового электрофоретического разделения продуктов реакции. И нужно сказать, что этот прорыв продолжается до сих пор. Однако детального рассмотрения различных методов секвенирования ДНК здесь касаться не будем, поскольку имеется множество публикаций на этот счет, включая юбилейную публикацию, посвященную 40-летию секвенирования ДНК [Shendure et al., 2017], а также наши работы [Чемерис и др. (Chemeris et al.), 1999; Зубов и др. (Zubov et al.), 2021]. Про ранние методы определения нуклеотидных последовательностей ДНК, помимо упомянутых изданий [Манделес, 1975;

<sup>4</sup> сделанное на стр. 32 обзора «Подходы к исследованию первичной структуры ДНК» [Свердлов, 1975] и приведенное нами почти что дословно

Свердлов, 1975], можно прочесть и в обзорной статье [Salsler, 1974], в которой также поднимался вопрос о нужности получения данных о нуклеотидных последовательностях молекул ДНК, но, как и следовало ожидать, ни о каких геномах не упоминалось.

В 1976 г. в майском номере журнале Science вышла небольшая статья под громким названием “DNA Sequencing: A New Era in Molecular Biology” [Kolata, 1976], что, конечно же, не являлось преувеличением. В ней кратко рассматривались достижения двух последних лет, сделанные разными коллективами ученых, включая упомянутый выше плюс/минус метод. При этом наибольшее внимание было уделено разработанному В.Гильбертом тогда еще неопубликованному методу секвенирования ДНК путем химической деградации, с помощью которого удалось прочесть около 100 нуклеотидов. Было отмечено, что Гильберт и Сэнгер независимо усовершенствовали гель-электрофорез нуклеиновых кислот, что позволило достичь разделения с точностью до нуклеотида в довольно большом размерном диапазоне и это оказалось краеугольным камнем для двух принципиально разных быстрых методов секвенирования ДНК, оригинальные статьи по которым затем вышли в следующем 1977 г. [Maxam, Gilbert, 1977; Sanger et al., 1977]. В той краткой статье [Kolata, 1976] были приведены подсчеты, сделанные А.Махам, согласно которым, чтобы прочесть 800 нуклеотидов - нужен гель длиной 5 метров и напряжение 20 тысяч вольт. Правда расчетная длительность такого электрофоретического разделения не приводилась.

\* \* \*

Одному из авторов данной статьи при секвенировании ДНК методом Максама-Гильберта во второй половине 1980-х гг. удавалось прочесть в приборе Macrohog шведской фирмы LKB более 600 нуклеотидов, разделяя их в течение 4 часов при напряжении 3000 V, но в 4%-ном (а не в 20%-ном как в оригинальном методе) полиакриламидном геле длиной 55 см клиновидной формы с целью торможения подвижности фрагментов ДНК меньшего размера, и с использованием низкоэнергетического <sup>33</sup>P, значительно улучшающего чтение в верхней части радиоавтографа, для которого применялась специальная рулонная рентгеновская пленка PA-1 отечественного производства. Для экспонирования на нее гель во время полимеризации приклеивался к одному стеклу и затем на нем высушивался, что также повышало четкость отображения радиоактивных полос ДНК. В частности, таким образом, был секвенирован участок рДНК, включающий ген 5,8S рРНК, фланкирующие его внутренние спейсеры (ITS 1 и ITS 2) между генами 18S рРНК и 26S рРНК и

соответственно края этих генов у дикого вида диплоидной пшеницы *Triticum urartu*, оказавшейся на тот момент третьим видом растений, у которого он был полностью прочитан [Чемерис, Вахитов (Chemeris, Vakhitov), 1989]. Спустя несколько лет определение нуклеотидных последовательностей этого участка рДНК растений стало носить по-настоящему массовый характер, поскольку ITS 1 и ITS 2 нашли очень широкое применение в филогенетических построениях и сейчас эта область секвенирована у десятков тысяч видов растений.

\* \* \*

Вспомнили здесь про метод Максама-Гильберта не случайно – он на протяжении десятка лет оставался основным методом определения нуклеотидных последовательностей ДНК. Для него даже был сконструирован робот, в автоматическом режиме выполнявший довольно трудоемкие операции химической модификации азотистых оснований и последующего расщепления цепей ДНК весьма опасными реагентами [Wada et al., 1983], но внедрение в метод Сэнгера фермента секвеназы вкупе с флуоресцентной меткой и автоматизацией процесса окончательно определило лидерство этого способа, предложенного, как уже говорилось выше, в том же 1977 г. [Sanger et al., 1977].

Так, в самом конце 1977 г. Сэнгером с соавторами было сообщено о разработке еще более быстрого и более простого нежели «плюс/минус» метод нового ферментативного метода секвенирования ДНК с так называемыми дидезокситерминаторами (2', 3'-ддНТФ), прерывающими синтез ДНК из-за отсутствия 3'-ОН группы в местах, где они включались в цепь, для чего оказалось достаточным использовать с соответствующими реакциями всего 4 пробирки (с ддАТФ, ддЦТФ, ддГТФ, ддТТФ по отдельности в определенной пропорции в дополнении ко всем четырем дНТФ в каждой) вместо 8 пробирок «плюс/минус» метода [Sanger et al., 1977]. Этот метод Сэнгера долгие годы, перехватив инициативу у метода Максама-Гильберта, служил основным методом для получения информации о нуклеотидных последовательностях ДНК и не сдает до конца свои позиции до сих пор, поскольку сейчас с его помощью решаются конкретные задачи, для которых не требуется высокопроизводительного секвенирования новых поколений, к которым позже перейдем. Серьезными улучшениями в определенные моменты первоначального метода Сэнгера можно считать использование фермента секвеназы, фага M13 (позднее фагмид), появление метода ПЦР, термосеквеназы (мутантный вариант ДНК полимеразы из *Thermus aquaticus*), а также произошедшие в середине 1980-х гг. автоматизация процесса и переход на использование для детекции флуоресцентных

красителей, а затем отказ от пластины геля и применение более производительного капиллярного гель-электрофореза.

Определенные успехи в развитии ферментативного метода секвенирования ДНК привели к созданию в 1988 г. в США Национального центра по исследованию генома человека (National Human Genome Research Institute – NHGRI), а немного позднее, в 1990 г. был разработан “Human Genome Project”, практически сразу ставший международным. Данный проект выполнялся 13 лет (вместо планировавшихся 15 и изначально даже 17 лет) и на него было потрачено приблизительно 3 млрд. долларов, из которых только около 300 млн. пошло непосредственно на проведение секвенирующих процедур, поскольку много денег, времени и сил было затрачено на поиск различных маркеров и построение генетических карт. При этом после завершения проекта «Геном человека» эти карты потеряли свою актуальность ввиду ставшей доступной принципиально иной более точной новой информации в виде последовательности нуклеотидов всего генома, пусть и в несовершенном виде<sup>5</sup>. Так, в 2001 г. одновременно в журналах Nature и Science вышли две статьи, в которых сообщалось о секвенировании с помощью метода Сэнгера первых двух геномов человека в черновом варианте [Lander et al., 2001; Venter et al., 2001]. Первая из этих статей была подготовлена Международным консорциумом, тогда как вторая – Институтом геномных исследований вместе с компанией Celera Genomics под руководством К.Вентера. При выполнении этого проекта коллегами Вентера был применен передовой для того времени так называемый «shotgun» подход к секвенированию больших геномов, что позволило справиться с секвенированием генома человека всего за три года, затратив лишь около 100 млн. долларов. Поэтому когда ведут отсчет снижения стоимости секвенирования генома человека с начала нынешнего столетия, то исходят из этой суммы. Возглавляемый Вентером коллектив, обошелся 300 производительными для того времени секвенаторами, получаемые результаты на которых обрабатывались одним мощным компьютером, тогда как в Международном проекте использовалось около 600 аналогичных секвенаторов, стоящих в нескольких странах мира, и множество компьютеров.

Международный консорциум по секвенированию генома человека 14 апреля 2003 г. в год 50-ти летнего юбилея открытия структуры ДНК в виде двойной спирали объявил о завершении работы над расшифровкой последовательности генома человека и в октябре 2004 г. ими была опубликована

соответствующая статья [ISHGC, 2004]. Аналогичная статья компании Celera вышла в феврале 2004 г. [Istrail et al., 2004]. При этом необходимо заметить, что ни тот, ни другой геном считать биологическим нельзя по ряду причин, среди которых использование для секвенирования тех геномов ДНК разных людей, но главным было то, что эти геномы представляли собой консенсусную сборку, в которой в мозаичном порядке перемежались участки материнских и отцовских геномов всех тех индивидов и поэтому его правильнее считать квазигеномом. Переходя на цифры, можно сказать, что велось секвенирование фрагментов находящейся в ядре клетки тотальной ДНК размером 6 млрд.п.н.<sup>6</sup>, а в итоге велась сборка из них более протяженных участков черновых геномов общим размером только 3 млрд.п.н., фактически сознательно отбрасывая половину генетической информации. То есть для того квазигенома  $3 + 3 \neq 6$ , а  $3 + 3 = 3$ , в результате чего по сути произошла потеря половина генетической информации, часть которой может оказаться критичной. Справедливости ради следует сказать, что подобное (неправильное) секвенирование (точнее сборка прочитанных последовательностей) имеет место до сих пор, за совсем небольшими исключениями, о чем речь пойдет особо.

Тут, пожалуй, стоит коснуться понятия геном, под которым, согласно действующего определения этого термина, подразумевается вся совокупность нуклеотидных последовательностей гаплоидного набора хромосом. Поэтому у конкретного человека не может быть вообще «генома человека» как такового, поскольку у него имеются определяющие его габитус и общее физиологическое состояние два достоящихся ему от родителей ядерных генома с общим размером около 6 млрд.п.н. И, как стало ясно позднее, довольно сильно отличающихся друг от друга, иногда называемых диплоидным геномом. Так, в одной из статей показано, что у одного человека два его генома (или две части диплоидного генома) отличаются на 3,3% [Jarvis et al., 2022], что довольно много так как долгое время считалось, что геном человека (точнее квазигеном) отличается от такого же генома нашего ближайшего сородича в дикой природе – шимпанзе на 0,5%. При этом раньше думалось, что парные хромосомы у диплоидных организмов практически одинаковы. В одной из недавних статей специально подчеркивается, что каждая человеческая клетка

<sup>5</sup> о чем будет говориться дальше

<sup>6</sup> а с учетом того, что при выполнении обоих проектов генома человека использовалась ДНК большого числа людей, то возможно и под 100 млрд.п.н. подвергалось секвенированию

несет два разных (ядерных) генома [Frazer, Schork, 2023].

Здесь нужно вернуться назад и понять, почему возникло решение секвенировать у человека лишь половинный набор хромосом. Впервые о необходимости определения нуклеотидной последовательности всех 24 хромосом (22-х аутосом и 2-х половых X- и Y-хромосом) человека заговорили в середине 80-х годов прошлого столетия. Так, в краткой заметке о секвенировании генома человека [Walsh, Marks, 1986] прозвучала мысль о необходимости секвенировать гаплоидный геном (но мужчины, чтобы не были парными хотя бы половые хромосомы), поскольку было опасение, что такой масштабный проект отвлечет слишком много денег от других проектов биологических исследований. На это Вентер указал в своей статье [Venter, 2010], написанной через десятилетие после обнародования в 2001 г. первых черновых геномов человека, посчитав, что решение о секвенировании (одного) генома человека из 3 млрд.п.н. вместо 6 млрд.п.н. послужило сдерживающим фактором для развития новых технологий секвенирования ДНК и было серьезной ошибкой<sup>7</sup>. В попытке ее хоть как-то исправить возглавляемый им коллектив секвенировал классическим методом Сэнгера персональный геном самого Вентера в диплоидном формате (частично) насколько позволяли тогдашние технологии [Levy et al., 2007]. Причем сборка велась в формате *de novo*, что является «краеугольным камнем» для исключения пропуска индивидуальных отличий персональных геномов. При этом оказалось, что «отцовский» геном Вентера отличался от доставшегося ему от матери на 0,5%. Вскоре было сообщено о секвенировании персонального диплоидного генома Нобелевского лауреата Дж. Уотсона [Wheeler et al., 2008]. Причем его геном стал первым, секвенированным методом новых поколений, а именно с помощью пиросеквенирования<sup>8</sup>, но для сборки использовался референсный геном человека, что практически исключало выявление всех уникальных особенностей генома Уотсона (в первую очередь инделов и прочих структурных вариаций, включая их копийность, так как именно они, а не однонуклеотидный полиморфизм вносят наибольший вклад в генетические различия высших организмов, и людей в частности) и поэтому он может считаться биологическим лишь с большой натяжкой.

<sup>7</sup> с чем сейчас, зная о множестве альтернативных высокопроизводительных методов секвенирования ДНК новых поколений, нельзя не согласиться, но уже и в 2010 г. к такому мнению прийти было можно

<sup>8</sup> о пиросеквенировании еще пойдет речь

Составленные в 2001 г. и уточненные в 2003 г. (квази)геномы человека, тем не менее явили собой гигантский прорыв в познании живого, но при этом привели к пониманию, что метод Сэнгера далее для решения подобных задач в массовом масштабе неприменим.

Сейчас несколько удивительно, но на протяжении четверти века интенсивного использования электрофоретических методов секвенирования ДНК буквально все экспериментаторы только и стремились, что улучшить ферментативный метод (главным образом) секвенирования ДНК. Лишь немногие работы были посвящены разработкам принципиально иных методов определения нуклеотидных последовательностей. В этой связи показательна статья известного специалиста J.Shendure из США, где он с некоторым удивлением (в нашем несколько вольном переводе) отмечает, что случилось так, что все исследователи настолько увлеклись улучшениями метода Сэнгера, что просто упустили из виду, что могут быть и какие-то иные способы секвенирования ДНК и что такие методы следует активно разрабатывать [Shendure, Ji, 2008]. И с этим мы полностью согласны, что легко заметить, в том числе из монографии «Секвенирование ДНК» (Чемерис и др. (Chemeris et al.), 1999), в которой про прочие методы секвенирования ДНК говорится как бы вскользь, поскольку таковых с относительно высокой производительностью почти и не было. Хотя справедливости ради следует сказать, что и в те годы все же появлялись некие альтернативные, тогда менее производительные методы секвенирования ДНК. Например, зарождающееся в те годы пиросеквенирование не казалось перспективным, поскольку по многим параметрам весьма сильно уступало методу Сэнгера. Но именно оно через 20 лет получило новое мощное развитие и «под него» был создан первый секвенатор нового поколения, позволяющий секвенировать бактериальные геномы за несколько часов вместо требовавшихся ранее на это многих месяцев упорного труда. И как уже говорилось выше, именно с помощью нового пиросеквенирования был прочитан геном Уотсона.

Однако в 2003 г. после завершения второго раунда секвенирования генома человека в умах исследователей словно произошел прорыв, что можно объяснить достижением желанной цели и получения большей свободы творчества, в результате чего стали один за другим разрабатываться методы секвенирования ДНК новых поколений на первый взгляд, казалось бы, малопроизводительные, но в том числе за счет массового параллельного анализа, в итоге обошедшие по общей скорости секвенирования классический метод секвенирования по Сэнгеру.

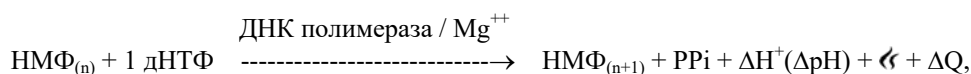
При этом практически единственной страной, оценившей перспективность разработки новых методов секвенирования ДНК, и то какую они сулят прибыль в будущем, в то время оказались США. Так, начиная с 2004 г., в США через NHGRI в рамках Программы развития технологий секвенирования (Sequencing Technology Development Program) множество проектов по разработке технологий секвенирования ДНК новых поколений получали финансовую поддержку, в том числе, и государственную, поскольку эта задача была определена в качестве приоритетной.

Проекты, выполняемые по той программе, были разделены на две основные категории – «сто тысяч долларов на геном» и «тысяча долларов на геном». На их разработку в 2004 г. было выделено 20 грантов (<https://www.genome.gov/12513210/2004-release-nhgri-seeks-next-generation-of-sequencing-technologies>). В 2005 г. было отобрано еще 12 проектов (<https://www.genome.gov/15015202/advanced-sequencing-technology-awards-2005>). Подобная поддержка продолжалась вплоть до 2014 г., что явствует из архивной web-страницы NHGRI (<https://www.genome.gov/Funded-Programs-Projects/Genome-Technology-Program/Archives>). Часть этих проектов были весьма экзотическими и

дальнейшего развития не получили, тогда как другие легли в основу технологий секвенирования ДНК новых поколений, в том числе доведенных до коммерческой реализации. Обобщенно многие методы нового секвенирования стали именовать NGS (Next Generation Sequencing), но с конкретными поколениями всех этих методов не все просто и нужно разбираться отдельно.

Помимо различных поколений методов секвенирования ДНК их еще можно подразделить по способам получения информации, два из которых уже упоминались – метод химической дегградации имеющихся молекул ДНК и ферментативного построения новых цепей ДНК с дидезокситерминаторами. Поскольку первый из них практически канул в лету, то главный интерес вызывает второй, хотя и сам он подразделился на целый ряд подходов, в том числе принципиально различающихся по используемым ферментам и «строительному» материалу.

Возникла новая аббревиатура SBS – Sequencing-by-Synthesis, но прежде чем рассматривать чуть более детально эту группу методов секвенирования стоит привести некую формулу происходящего при полимеризации ДНК под действием ДНК полимеразы -



где  $\text{НМФ}_{(n)}$  и  $\text{НМФ}_{(n+1)}$  представляют собой молекулы ДНК, в том числе удлинённую на один нуклеотид; образующиеся при полимеризации  $\text{PPi}$  и  $\text{H}^+$  – это соответственно пирофосфат и протон;  $\llcorner$  - обозначает выделяющееся тепло при разрыве макроэргической связи в дНТФ;  $\Delta Q$  – символизирует изменение импеданса реакционной смеси за счет увеличения общего заряда молекулы ДНК. Причем все эти изменения можно детектировать с той или иной степенью успешности и на них основаны разные методы секвенирования ДНК. Так, увеличение длины молекул ДНК регистрируется в классическом методе Сэнгера, выделение пирофосфата – в пиросеквенировании, выделение протона – в полупроводниковом секвенировании, выделение тепла и изменение импеданса в термосеквенировании и электронном секвенировании соответственно. Для двух последних подходов хотя и были созданы прототипы приборов, но на массовый рынок они не вышли, в отличие от детекции первых трех продуктов реакции полимеризации ДНК, для которых создано немало моделей коммерческих секвенаторов.

При пиросеквенировании для того, чтобы понять какой нуклеотид включился в растущую цепь ДНК, используется биолюминесценция, возникающая

в ходе каскада ферментативных реакций. Так, выделяющийся пирофосфат под действием АТФ-сульфуриказы превращает находящийся в реакционной смеси аденозинфосфосульфат в АТФ, способствующий затем окислению люциферина люциферазой с выделением регистрируемых квантов света. При этом из-за перекрестной реакции люциферазы с дАТФ приходится использовать его синтетический аналог с тио-группой в альфа-положении (дАТФ $\alpha$ S), что дополнительно приводит к удорожанию и без того недешевых реагентов. И это становится особенно серьезным недостатком пиросеквенирования при масштабировании процесса. Другой не менее серьезный недостаток заключается в отсутствии терминации из-за чего в растущую цепь ДНК может включиться сразу несколько одинаковых нуклеотидов, что с одной стороны отчасти ускоряет весь процесс, а с другой – ведет к искажению результатов. Так, количество таких подряд вставляемых нуклеотидов можно определить по уровню сигнала биолюминесценции, но при увеличении длины гомополимерных повторов точность определения количества одинаковых нуклеотидов в них быстро снижается, что приводит к ошибкам секвенирования таких повторов в виде

пропусков или вставок нуклеотидов. Собственно этот же недостаток в виде проблемы с точностью определения нуклеотидных последовательностей в гомополимерных участках присущ и гораздо более дешевому полупроводниковому секвенированию, а также так и не вышедшим на простор термосеквенированию и электронному секвенированию.

Наибольшее развитие среди NGS методов SBS секвенирования ДНК получил подход с регистрацией изменений флуоресценции, возникающих при присоединении очередных дНМФ, меченных соответствующими флуорохромами. Основной частью подобных секвенаторов является проточная ячейка, смонтированная на позиционируемом столике флуоресцентного микроскопа.

В свою очередь флуоресцентное SBS секвенирование подразделяется на ряд подходов, среди которых мономолекулярные методы и МПС методы. С последних и начнем рассмотрение флуоресцентного секвенирования. Наиболее массовым стало подобное секвенирование, развиваемое американской фирмой Illumina и основанное на проведении твердофазной мостиковой ПЦР (известной как технология получения молекулярных колоний, придумана, запатентована и экспериментально подтверждена в начале 90-х российским ученым А.Б.Четвериным из Института белка РАН в Пущино с использованием флуоресцентно меченых дНТФ с обратимым блокированием полимеразного синтеза, применяемое в большинстве современных флуоресцентных секвенаторов [Четверин, Четверина, 1995; Chetverin, Chetverina, 1997]. Необходимо отметить, что проведение ПЦР на поверхности проточных ячеек секвенатора – еще одна пионерная разработка российских ученых из Института молекулярной биологии им. В.А.Энгельгардта РАН под руководством А.Д.Мирзабекова [Khrapko et al., 1989], занимавшихся разработкой биологических чипов для секвенирования ДНК и выполненная на 10 лет раньше американских исследователей из Mosaic Technologies<sup>9</sup>. Субстратами, включаемыми в растущую цепь ДНК, служат модифицированные дНТФ, содержащие флуорофоры, связанные с азотистыми основаниями метилазидными линкерами. Важной особенностью 3'-О-азидометильных производных дНТФ<sup>10</sup> является сравнительно высокая стабильность, сочетающаяся с простотой и быстротой

деблокирования. Можно уверенно утверждать, что базовые элементы технологии секвенирования путем синтеза (СПС или SBS) были разработаны российскими учеными, но, что характерно и для многих других прорывных идей того времени, были подхвачены и реализованы в других странах, прежде всего США. Схожую технологию применяет китайская фирма BGI с тем отличием, что вместо мостиковой ПЦР она использует так называемые ДНК-наноболлы в виде клубков одноцепочной ДНК, формирующихся путем амплификации катящимся кольцом и содержащих сотни или даже тысячи tandemно повторяющихся секвенируемых участков ДНК, а также упорядоченное расположение таких ДНК-наноболлов на сканируемой поверхности проточных ячеек. При этом важным преимуществом данной технологии является то, что в каждом цикле (обороте ДНК полимеразы вокруг кольцевой матрицы) происходит копирование исходной матрицы ДНК, что уменьшает число возможных ошибок ДНК полимеразы, поскольку в отличие от ПЦР вновь образуемые цепи в дальнейшую амплификацию не вовлекаются.

В настоящее время многими фирмами в США и КНР производятся подобные флуоресцентные секвенаторы. В России Институтом аналитического приборостроения совместно с ООО «НПФ Синтол» и рядом других организаций<sup>11</sup> выпускается секвенатор «Нанофор СПС» чему в данном номере посвящена отдельная статья [Алексеев и др. 2025].

Помимо секвенирования ДНК за счет роста цепи с помощью ДНК полимеразы в МПС методах, одно время существовало еще лигазное секвенирование ДНК, в ходе которого происходили циклические процессы пришивки ДНК лигазой к 5'-концу субстратного комплекса флуоресцентно меченного короткого олигонуклеотида и после идентификации конкретной флуоресцентной метки, указывающей на тип определяемого нуклеотида, проводилось ее отщепление и подготовка субстратного комплекса, удлинённого на 5 нуклеотидов, к следующему этапу лигирования. Некоторое время американской фирмой Life Science Technologies производился соответствующий секвенатор модели SOLiD (Sequencing by Oligonucleotide Ligation and Detection), но длина чтения была крайне небольшой. Тем не менее, с помощью лигазного секвенирования вкупе с флуоресцентным полимеразным секвенированием, реализованным тогда в приборе модели GAII фирмы

<sup>9</sup> эту компанию позже купила фирма Illumina

<sup>10</sup> их синтез был разработан российскими учеными в ИБХ РАН под руководством А.И.Мирошникова [Zavgorodny et al., 2000]

<sup>11</sup> Здесь можно еще добавить, что данным коллективом производится также капиллярный секвенатор «Нанофор 05», рассчитанный на секвенирование ДНК по методу Сэнгера

Иllumina, впервые в России был секвенирован полный геном человека, принадлежащий мужчине русской национальности [Скрябин и др. (Skryabin et al.), 2009].

Что касается мономолекулярных методов секвенирования ДНК, основанных на флуоресценции, то два из них были доведены до коммерческого воплощения американскими фирмами Helicos и Pacific Biosciences. Однако первая, выпустив прибор Heliscope (оказавшийся очень громоздким – под тонну весом) и просеквенировав геном человека [Pushkarev et al., 2009], в 2010 г. обанкротилась. Серьезными причинами послужило, что чтение ДНК было очень неточным и характеризовалось слишком малой длиной прочтения, хотя неким преимуществом можно было считать единственный тип используемого флуорохрома, будучи фактически монокромной технологией. В данной технологии, названной tSMS (true Single Molecule Sequencing), проводилась временная терминация роста цепей ДНК, пришитых к стеклянной подложке.

В том же 2009 г. компанией Pacific Biosciences, Inc. была реализована другая технология мономолекулярного секвенирования ДНК - Single Molecule Real Time (SMRT) [Eid et al., 2009]. Этот подход также основан на использовании модифицированной фаговой ДНК полимеразы Phi29, но в отличие от технологии tSMS, в которой фиксировались олигонуклеотиды, в SMRT технологии к подложке (ко дну лунок цептолитрового объема, так называемых волноводов нулевой моды) пришивается ДНК полимеразы, а вместо дНТФ используются дезоксинуклеозидгексафосфаты дНМФ, в которых флуоресцентные красители введены по терминальному дзета-фосфату, что в результате присоединения дНМФ к цепи ДНК приводит к отщеплению пентафосфата, а с ним и флуоресцентной метки. Это обеспечивает рост по сути нативной цепи ДНК, что приводит к ее большей длине, достигающей десятка тысяч пар нуклеотидов. Причем такое секвенирование ведется в режиме реального времени, без смены реагентов и в варианте технологии кольцевой матрицы (Hi-Fi reading) позволяет достигать наивысших на сегодняшний день значений точности прочтения среди всех методов секвенирования ДНК – Q40 (одна ошибка на 10 тысяч нуклеотидов).

В 2010 году компания Pacific Biosciences выпустила на рынок NGS секвенатор PacBio RS. Прибор оказался также довольно громоздким (около одной тонны), но вскоре появилась его доработанная версия PacBio RSII, а в начале 2017 года в продаже появился секвенатор Sequel, отличающийся от своих предшественников повышенной производительностью и меньшим весом - 350 кг.

Затем ему на смену пришли Sequel II и Revio, производительность которых значительно увеличилась за счет существенного (с 200 тысяч для PacBio RS до 1 миллиона для Sequel и до 25 миллионов для Revio) увеличения количества волноводов нулевой моды на ячейке.

Наконец, последним из ныне применяемых методов высокопроизводительного секвенирования ДНК, на рынок было выведено нанопоровое секвенирование, хотя его идея возникла гораздо раньше остальных – еще в эпоху господствующего секвенирования по Сэнгеру в середине 1990-х гг. [Kasianowicz et al., 1996]. Но воплощение сильно затянулось, поскольку для реализации этой перспективной технологии потребовались длительные исследования, только подчеркивающие фундаментальность идеи. Связанные с этим события довольно подробно изложены в историческом обзоре, посвященном двум десятилетиям разработки нанопорового секвенирования ДНК [Deamer et al., 2016] и здесь их касаться не будем.

В основе нанопорового секвенирования лежит прохождение под действием электрического тока молекулы ДНК через подходящую белковую пору нанометрового размера – ионный канал в бислоистой липидной мембране. При этом, отличаясь по размеру, разные азотистые основания при прохождении по-разному влияют на величину ионного тока, что позволяет определить последовательность проходящих через пору нуклеотидов. Теоретически все представляется очень простым действием, но скорость такого прохождения ДНК составляет около миллиона нуклеотидов в секунду и подобное количество дискретных сигналов невозможно зарегистрировать за столь короткое время. Для торможения движения ДНК пытались использовать различные способы, но самым лучшим вариантом оказалось ферментативное торможение хеликазами – ферментами, расплетающими двухцепочечную ДНК с использованием энергии АТФ. Так что нанопоровое секвенирование до некоторой степени тоже можно считать ферментативным, хотя хеликаза выполняет, по сути, некую вспомогательную роль.

В 2005 г. в Великобритании была основана компания Oxford Nanopore Technologies Ltd., занявшаяся исключительно нанопоровой технологией секвенирования ДНК. Сейчас она производит целую линейку нанопоровых секвенаторов, представленных на их сайте <https://nanoporetech.com/> - MinION, GridION, PromethION, включая их различные модели, характеризующиеся разной производительностью. При этом, стоит обратить внимание на размер и вес первого из них – MinION умещается на ладони и весит всего 78 г против почти тонны у некоторых

старых моделей NGS секвенаторов при схожей производительности.

Учитывая многие факторы, среди которых дешевизна нанопорового секвенирования, большая длина прочтения, ограничиваемая, по сути, только длиной молекул ДНК, которыми удастся оперировать, интерес к этому методу растет. Причем проявление наибольшего интереса наблюдается в КНР, где сразу несколько фирм (Axbio Biotechnology, BGI/MGI Tech, Geneus Technologies, Genvida, Polyseq, Qitan Technology, RH Genetech) или уже производят нанопоровые секвенаторы разной производительности, или только готовятся к выпуску таковых, либо ведут оригинальные разработки. В России начат серийный выпуск подобного ДНК секвенатора модели «Нанопорус».

Таким образом, к настоящему времени в мире имеет место следующая ситуация. В США производятся флуоресцентные секвенаторы, как рассчитанные на МПС методы (фирма Illumina и ряд других), так и мономолекулярные (фирма PacBio). В Великобритании фирмой Oxford Nanopore Technologies (ONT) выпускается целая линейка нанопоровых секвенаторов. В КНР фирмой BGI производятся флуоресцентные секвенаторы нескольких моделей, а также ее дочерней фирмой MGI Tech ведутся разработки нанопорового секвенатора. Еще несколько нанопоровых секвенаторов разрабатывают другие китайские фирмы, которые вскоре, несомненно, смогут составить конкуренцию Oxford Nanopore Technologies. Россия является четвертой страной, в которой полногеномному секвенированию уделяется внимание, и уже массово производятся флуоресцентный секвенатор «Нанофор СПС» и нанопоровый секвенатор «Нанопорус». Имеются и дальнейшие планы по разработке более мощных систем, в том числе на основе флуоресцентной мономолекулярной технологии. В июле 2025 года в Институте аналитического приборостроения РАН стартует разработка нового отечественного полногеномного секвенатора «Нанофор СПС2» в 20 раз более производительного, чем «Нанофор СПС» (объем геномных данных за запуск не менее 120 млрд. нуклеотидов).

В связи с нанопоровым секвенированием нельзя не упомянуть о работах, в которых вместо белковых нанопор предлагается использовать так называемые твердотельные, что подробно рассмотрено в недавнем обзоре [Zhou et al., 2025], но внедрение этой технологии в практику еще нужно ждать. Иной оригинальный подход к нанопоровому секвенированию разрабатывает американская фирма

Roche Sequencing Solutions Inc.<sup>12</sup> [Kokoris et al., 2025]. Главная идея этого метода, названного Sequencing by Expansion (SBX), заключается в прохождении через нанопору некоего аналога ДНК, построенного энзиматически из дНТФ, содержащих разные объемные группы, что позволяет четче улавливать изменения ионного тока, что должно повысить точность прочтения. При этом такой аналог ДНК строится специально модифицированной ДНК полимеразой, «хорошо относящейся» к неприродным дНТФ (в том числе с крупными «довесками») при полимеризации цепей ДНК, однако длина таких молекул не может быть очень большой и тогда главное преимущество нанопорового секвенирования, как способного читать очень протяженные молекулы ДНК, теряется.

Готовясь завершить данную статью, еще раз повторим – секвенирование ДНК - это важнейший процесс, дающий исследователям ранее абсолютно недоступные сведения и обеспечивающий дальнейший прогресс в развитии наук о Жизни. Причем это имеет место до сих пор, когда впервые производится секвенирование полного генома организма, для которого он был ранее неизвестен. Впрочем, и ресеквенирование уже известных геномов позволяет составлять так называемые пангеномы<sup>13</sup>, которые на самом деле представляют очень важную информацию, так как уже давно стало ясно, что ни один референсный геном какого-либо вида не дает всей полноты информации о присущем этому виду организму геном пуле, поскольку имеются так называемые коровые гены, характерные для всех образцов конкретного вида, и варибельная часть генов, присущих не всем образцам. Если коровые гены обеспечивают преимущественно основной первичный метаболизм, то вторая группа генов отвечает за вторичный метаболизм и во многом определяет разнообразие форм, в том числе позволяя адаптироваться организму их несущему к меняющимся условиям внешней среды. И это особенно важно для селекционной работы, чтобы добиваться наилучшей комплементации разных генов, в том числе для создания гетерозисных гибридов с разнокачественными по сути гетерогеномами. Но обычные референсные, да даже и пангеномы плохи тем, что в большинстве своем представлены все теми

<sup>12</sup> ранее фирма Roche производила пиросеквенаторы серии GS FLX

<sup>13</sup> Пангеномике растений посвящена другая наша статья [Kuluev et al., 2025a]. Оказались не забыты нами и панпластомы [Samigullin et al., 2025], а также пан-митогеномы [Кулуев и др. (Kuluev et al.), 2025], но они имеют большее значение для геносистематических исследований.

же квазигеномами, в которых половина генетической информации попросту отсутствует.

Так, в настоящее время при выполнении большинства проектов по секвенированию геномов производится чтение нуклеотидных последовательностей ДНК из полного набора хромосом, а в итоге из них собирается только половинный набор, причем фрагменты материнского (М) и отцовского (Р) геномов располагаются в случайном мозаичном порядке. То есть при секвенировании тотальной ДНК определяются множественные фрагменты М генома в утрированном виде как  $m_n$  и Р генома как  $p_n$  и их сборка вместо двух истинных геномов - ... mmmmmmmmmmmmmmmmm ... etc и ... rrrrrrrrrrrrrrrrr ... etc соответственно производится в виде квазигенома – ... mrrrrrrrrrrrrrrrrrr ... etc, что далеко от истинного распределения таких фрагментов по материнским и отцовским хромосомам или иначе по гаплотипам. Причем при полногеномном секвенировании принято достигать некоего покрытия на геном, например для относительно коротких прочтений кратного 30, но для двух ядерных геномов с обязательно отличающимися последовательностями из полного набора хромосом реально оно оказывается только 15-ти кратным, что по теории вероятности недостаточно для того, чтобы уверенно прочитать все фрагменты двух ядерных геномов одного высшего организма.

В этой связи нужно сказать, что фактически геном геному рознь даже в том плане, что уровни сборок так называемых полных геномов сильно разнятся. Причем понятное дело они улучшались вслед за совершенствованием технологий секвенирования ДНК и развитием программного обеспечения вместе с ростом производительности вычислительных мощностей. Так, от черновых геномов в виде компиляции из контигов и скаффолдов стали затем собираться нуклеотидные последовательности, распределенные по хромосомам. Позже был достигнут уровень секвенирования/сборки по хромосомам от теломеры до теломеры, получивший название T2T (telomere-to-telomere). В частности не так давно собран геном «человека» в формате T2T для чего были взяты клетки гомозиготной линии с кариотипом 46,XX, что позволило собрать без пробелов все хромосомы в их действительно гаплоидном наборе общим размером 3,055 млрд.п.н. и исправить все ошибки предыдущих сборок [Aganezov et al., 2022; Nurk et al., 2022]. Однако считать эту сборку биологическим геномом вряд ли можно, поскольку живой организм с такими одинаковыми геномами в природе образоваться не может, да и его жизнеспособность тоже остается под вопросом. Впрочем, даже ранее научились собирать

диплоидные геномы с фазированной сборкой гаплотипов<sup>14</sup>, что реально ближе к биологическим геномам и о них дальше пойдет речь.

Еще в 2010 году Вентер в той своей, краткой, но по сути, программной статье [Venter, 2010], отвечая себе же на вопрос какой будет геномика через 10 лет, высказал предположение, что за счет улучшения технологий секвенирования для каждого человека будут секвенироваться его множественные геномы из сперматозоидов, яйцеклеток, стволовых клеток. При этом Вентер завершил статью словами, что «геномная революция только начинается» и это был, напомним 2010 г., когда уже несколько лет как появились методы NGS и счет секвенированным геномам эукариот пошел на десятки. Спустя полтора десятка лет, можно сказать, что геномная революция по-прежнему только начинается, несмотря на повышение общей производительности процесса секвенирования, которое можно действительно считать революционным, ввиду того, что она (революция) должна была произойти еще и в умах исследователей, но этого, к сожалению, в массе своей пока не случилось. Так, по-прежнему продолжают секвенироваться квазигеномы, ценность которых низка и принципиально новой информации для видов с известными референсными геномами они не несут, если не принимать во внимание пангеномы. Но и для них крайне важна фазированная сборка по гаплотипам, поскольку только в этом случае появляется возможность определить цис/транс-положения мутаций, которые могут приводить к заменам аминокислот со всеми вытекающими последствиями для их функциональной активности. Чтобы было понятнее стоит в качестве примера привести как результат мозаичной сборки отдельные аминокислотные остатки некоего белка в виде Cys...Arg...Asp, кодируемых триплетами TGT...AGG...AAC, возможно даже принадлежащих разным экзонам, тогда как в реальности два аллеля этого гена в диплоидном геноме кодируют два разных белка с аминокислотами - Cys...Ser...Asp (TGT...AGC...AAC) и Ser...Arg...Lys (AGC...AGG...AAG), что далеко не одно и то же и может оказаться критичным даже для жизнедеятельности. Поэтому, для понимания связи генотипа с фенотипом следует опираться только на нуклеотидные последовательности диплоидного генома с фазированной сборкой по гаплотипам – другого, что называется, не дано.

Нужно признать, что уже немало геномов собраны именно в таком T2T формате с фазированной сборкой по гаплотипам. Особенно много таковых собрано для растений, что

<sup>14</sup> и здесь можно вспомнить все тот же геном Вентера

рассмотрено в данном номере журнала [Baumiev et al., 2025]. И это нужно считать правильным секвенированием, о чем мы упоминали вначале. Еще в столетний юбилей термина геном в 2020 г. мы поднимали вопрос о необходимости прекращения секвенирования квазигеномов, как не отвечающих современным требованиям по установлению связи генотипа с фенотипом, который определяется диплоидным геномом из двух гаплотипов или двумя геномами [Кулуев и др. (Kuluev et al.), 2020]. Нужна переориентация на секвенирование ДНК (точнее сборку последовательности нуклеотидов) для полного набора хромосом любых высших организмов. Недавно в одной из статей международного коллектива авторов было отмечено, что применительно к человеку они больше не рассматривают сборку генома из 3 млрд.п.н. как современный уровень знаний, а рассматривают два генома для каждого диплоидного генома (т.е. 6 млрд.п.н. против 3 млрд.п.н.), где родительские гаплотипы собраны в фазированном формате [Porubsky et al., 2023]. И такой подход несомненно будет набирать силу. Так, недавно сообщено, что секвенированы 65 геномов человека и соответственно построены 130 фазированных сборок по гаплотипам в формате T2T [Longsdon et al., 2024]. И это теперь современный уровень правильных геномных данных.

Но что же мешает полностью перейти на секвенирование и сборку геномов любых высших организмов в таком формате? Ответ прост – слишком высокая схожесть парных хромосом, которые хотя и могут довольно сильно отличаться, но этого недостаточно, чтобы вести уверенную сборку по гаплотипам. Тем не менее, движение в эту сторону уже давно есть и оно должно шириться. Так, специально создано уже более полусотни программ-сборщиков таких геномов. Но им нужно предоставлять высокоточные данные и желательно подлиннее. Требуется применение и дополнительных подходов в виде Strand-seq, Hi-C seq, оптического картирования, трио-секвенирования (когда такое возможно), а также использовать прочие ухищрения. При этом можно не сомневаться, что уже в недалеком будущем под полногеномным секвенированием будет пониматься именно такое – в формате T2T без промежутков с фазированной сборкой по гаплотипам геномов организмов любого уровня плоидности.

Возвращаясь к полувековому юбилею относительно быстрого секвенирования ДНК, стоит заметить, что в отличие от упомянутых выше 60 и 90 п.н, которые чтобы прочитать в 1975 г. требовались значительные усилия и немалое время, сейчас почти играючи гораздо быстрее с использованием топовых моделей секвенаторов, причем рассчитанных на три основные платформы (Illumina, PacBio и ONT) можно

прочитать триллионы п.н. в одном приборе за один запуск. И верится, что это не предел. При этом и длина прочтения увеличилась заметно и даже во много раз. Так, в зависимости от используемого метода сейчас можно прочесть с разной точностью непрерывную последовательность длиной от 250 до 20 тысяч и даже свыше миллиона п.н.

Немаловажное значение имеет также стоимость секвенирования. Так, на сайте National Human Genome Research Institute на протяжении многих лет ведется отслеживание стоимости секвенирования генома человека (рис. 1) и стоимости секвенирования одного миллиона нуклеотидов (рис. 2), сопоставляя эти значения с законом Мура.<sup>15</sup> Как можно видеть из приведенных сравнений<sup>16</sup> темпы снижения стоимости секвенирования как генома человека, так и в пересчете на один миллион нуклеотидов с 2001 г. до конца 2007 г. были практически сопоставимы с тем, что происходит в микроэлектронике, но с начала 2008 г. кривая стоимости секвенирования пошла резко вниз и к концу 2015 г. вместо того, чтобы согласно закону Мура уменьшиться за этот период приблизительно в 8 раз, она снизилась на четыре (!) порядка. Справедливости ради следует отметить, что в последние годы темпы снижения стоимости секвенирования серьезно замедлились. Тем не менее, по сравнению с 2001 г. стоимость секвенирования упала на 5 – 6 порядков.

Как уже говорилось выше, сравнение с 2001 г. не случайно и объясняется не только (и даже не столько) началом нового столетия, поскольку это еще и дата завершения секвенирования черновых геномов человека. В этой связи можно вспомнить статью в Science [Collins et al., 1998], в которой сообщалось о ходе выполнения международного проекта “Human Genome Project” и о планах на следующую пятилетку. В частности, были приведены сведения, что по состоянию на 1998 г. достигнута скорость чтения нуклеотидных последовательностей генома человека в виде 90 млн. нуклеотидов в год (имелось в виду при выполнении того Проекта) с ценой за нуклеотид в 50 центов. При этом планировалось к 2003 г. достигнуть скорости 500 млн. нуклеотидов в год по цене 25 центов за нуклеотид. Однако такая стоимость в отличие от скорости была достигнута лишь к 2015 г. Сейчас же только один секвенатор (некоторых топовых моделей) практически за несколько суток производит Tb ( $10^{12}$ ) нуклеотидов при стоимости менее цента за нуклеотид.

<sup>15</sup> В середине 60-х годов прошлого столетия Г. Мур высказал предположение (получившее известность как «закон Мура»), что число транзисторов на кристалле будет удваиваться каждые два года, после того, как он подметил, что это происходит на протяжении нескольких лет [Moore, 1965].

<sup>16</sup> последние данные на сайте NHGRI приведены по состоянию на 2022 г.

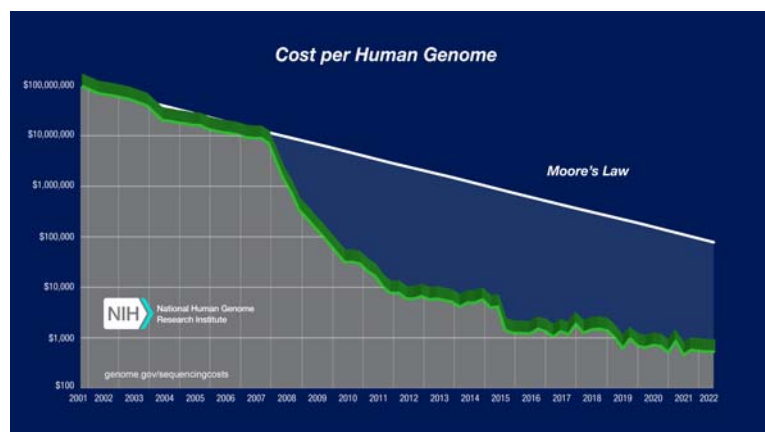


Рис. 1. Сопоставление снижения стоимости секвенирования генома человека с законом Мура  
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>  
 Fig. 1. Comparison of the reduction in the cost of sequencing the human genome with Moore's law

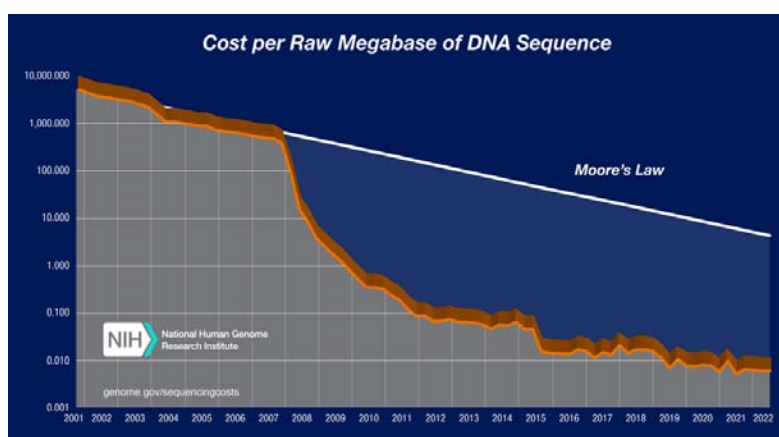


Рис. 2. Сопоставление снижения стоимости секвенирования одного миллиона нуклеотидов с законом Мура  
 Fig. 2. Comparison of the reduction in the cost of sequencing one million nucleotides with Moore's law  
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

И это все вместе, вне всякого сомнения, будет способствовать новой революции в геномном секвенировании, ожидая, в том числе, стоимости секвенирования в пределах 100 долларов на геном, равный по размеру человеческому.

Напоследок нужно коснуться нумерации поколений методов секвенирования ДНК. На веб-странице фирмы PacBio (<https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/>) приведены основные технологии секвенирования ДНК, разделенные по поколениям. Так, методам Сэнгера и Максама-Гилберта отведено первое поколение. Ко второму поколению были отнесены 454, Solexa, Ion Torrent, Illumina. Третье поколение представлено технологиями PacBio и

Oxford Nanopore. Одним из критериев, принимавшихся во внимание, были длины прочтения. Одно время NGS подходы относили к тем или иным поколениям, исходя из времени их появления, что не совсем корректно, поскольку важнее не даты, а производительность, в том числе с учетом разных параметров. Тем более, что идеи того или иного метода могли появляться задолго до их реального воплощения. В одном из обзоров [Feng et al., 2015] нанопоровое секвенирование сочли четвертым поколением методов секвенирования ДНК. Отчасти стоит с этим согласиться, основываясь и на длине прочтения и на принципиально ином способе детекции последовательности нуклеотидов, хотя его идея появилась намного раньше других прочих.

Учитывая крайнюю важность секвенирования ДНК и то, что нынешнее секвенирование все же пока не позволяет легко решать стоящие грандиозные задачи, то нужно ждать еще более производительных и высокоточных новых технологий секвенирования, и вполне вероятно, что они появятся. Мы сами также работаем над этим.

### Литература

1. Зубов В.В., Чемерис Д.А., Василев Р.Г. и др. Краткая история методов высокопроизводительного секвенирования нуклеиновых кислот // *Biomcs*. 2021. Т.13(1). С. 27- 46. DOI: 10.31301/2221-6197.bmcs.2021-4
2. Кулуев А.Р., Матниязов Р.Т., Березин А.А. и др. Пан-митогеномика // *Biomcs*. 2025. Т. 17(1). С.88-102 DOI: 10.31301/2221-6197.bmcs.2025-7
3. Кулуев Б.Р., Баймиев А.Х., Геращенко Г.А. и др. Сто лет гаплоидным геномам. Сейчас наступает время диплоидных // *Биомика*. 2020. Т.12(4). С. 411-434. DOI: 10.31301/2221-6197.bmcs.2020-33
4. Манделес С. Установление первичной структуры нуклеиновых кислот. М.: Мир. 1975. 319 с.
5. Свердлов Е.Д. Подходы к исследованию первичной структуры ДНК // *Итоги науки и техники*. М.: ВИНТИ. Сер. Молекулярная биология. Т.4. С.7-88.
6. Скрыбин К.Г., Прохорчук Е.Б., Мазур А.М. и др. Комбинирование двух технологических платформ для полногеномного секвенирования человека // *Acta Naturae*. 2009. Т.1(3). С. 113-119.
7. Чемерис А.В., Ахунов Э.Д., Вахитов В.А. Секвенирование ДНК. М., Наука, 1999. 429 с.
8. Чемерис А.В., Вахитов В.А. Первичная структура гена 5,8S рРНК и внутренних транскрибируемых спейсеров рДНК у диплоидной пшеницы *Triticum urartu* Thun. ex Gandil. // *Молекулярная биология*. 1989. Т.23. С.320-326.
9. Четверин А.Б., Четверина Е.В. 1995. Способ размножения нуклеиновых кислот, способ их экспрессии и среда для их осуществления. Патент РФ № 2048522
10. 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
11. Air GM, Blackburn EH, Sanger F, Coulson AR. The nucleotide and amino acid sequences of the N (5') terminal region of gene G of bacteriophage phiX 174 // *J Mol Biol*. 1975. V.96(4). P.703-719. doi: 10.1016/0022-2836(75)90147-3
12. Baymiev Al.Kh., Chemeris D.A., Sakhabutdinova A.R. et al. In higher plants as an example, one can see that the era of sequencing of their diploid genomes is coming // *Biomcs*. 2025. V.17(1). P. 17-41. DOI: 10.31301/2221-6197.bmcs.2025-3
13. Chargaff E. What really is DNA? Remarks on the changing aspects of a scientific concept // *Prog. Nucl. Acid Res. Mol. Biol*. 1968. V.8. P.297-333. 10.1016/s0079-6603(08)60549-8
14. Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing // *Nature Biotechnol*. 2016. V.34(5). P.518-524. doi: 10.1038/nbt.3423
15. Collins FS, Patrinos A, Jordan E et al. New goals for the U.S. Human Genome Project: 1998-2003 // *Science*. 1998. V.282(5389). P.682-689. doi: 10.1126/science.282.5389.682
16. Eid J., Fehr A., Gray J. et al. Real-time DNA sequencing from single polymerase molecules // *Science*. 2009. V.323. P.133-138. doi: 10.1126/science.1162986
17. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology // *Genomics Proteomics Bioinformatics*. 2015. V.13(1). P.4-16. doi: 10.1016/j.gpb.2015.01.009
18. Frazer KA, Schork NJ. The human pangenome reference anticipates equitable and fundamental genomic insights // *Cell Genom*. 2023. V.3(7). 100360. doi: 10.1016/j.xgen.2023.100360
19. Jarvis ED, Formenti G, Rhie A et al. Semi-automated assembly of high-quality diploid human reference genomes // *Nature*. 2022. V.611(7936). P.519-531. doi: 10.1038/s41586-022-05325-5
20. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004. V.431(7011). P.931-945. doi: 10.1038/nature03001
21. Istrail S., Sutton G.G., Florea L. et al. Whole-genome shotgun assembly and comparison of human genome assemblies // *Proc. Natl. Acad. Sci. USA*. 2004. V.101(7). P.1916-1921. doi: 10.1073/pnas.0307971100
22. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel // *Proc Natl Acad Sci USA*. 1996. V.93(24). P.13770-13773. doi: 10.1073/pnas.93.24.13770
23. Khrapko KR, Lysov YuP, Khorlyn AA et al. An oligonucleotide hybridization approach to DNA sequencing // *FEBS Lett*. 1989. V.256(1-2). P.118-122. doi: 10.1016/0014-5793(89)81730-2
24. Kolata GB. DNA sequencing: a new era in molecular biology // *Science*. 1976. V.192(4240). P.645-647. doi: 10.1126/science.192.4240.645
25. Kokoris M, McRuer R, Nabavi M et al. Sequencing by Expansion (SBX) - a novel, high-throughput single-molecule sequencing technology // *bioRxiv [Preprint]*. 2025. 2025.02.19.639056. doi: 10.1101/2025.02.19.639056

26. Lander ES., Linton LM, Birren B et al. Initial sequencing and analysis of the human genome // *Nature*. 2001. V.409(6822). P.860-921. doi: 10.1038/35057062
27. Levy S., Sutton G., Ng P.C. et al. The diploid genome sequence of an individual human // *PLoS Biol*. 2007. V.5(10). e254. doi: 10.1371/journal.pbio.0050254
28. Logsdon GA, Ebert P, Audano PA et al. Complex genetic variation in nearly complete human genomes // *bioRxiv* [Preprint]. 2024. 25:2024.09.24.614721. doi: 10.1101/2024.09.24.614721
29. Maxam A.M., Gilbert W. A new method for sequencing DNA // *Proc. Natl. Acad. Sci. USA*. 1977. V.74. P.560-564. doi: 10.1073/pnas.74.2.560
30. 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
31. Porubsky D, Vollger MR, Harvey WT et al. Gaps and complex structurally variant loci in phased genome assemblies // *Genome Res*. 2023. V.33(4). P.496-510. doi: 10.1101/gr.277334.122
32. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome // *Nat. Biotechnol*. 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
33. Salser WA. DNA sequencing techniques // *Annu Rev Biochem*. 1974. V.43. P.923-965. doi: 10.1146/annurev.bi.43.070174.004423
34. Sanger F, Air GM, Barrell BG et al. Nucleotide sequence of bacteriophage phi X174 DNA // *Nature*. 1977. V.265(5596). P.687-695. doi: 10.1038/265687a0
35. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase // *J Mol Biol*. 1975. V.94(3). P.441-448. doi: 10.1016/0022-2836(75)90213-2
36. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. USA*. 1977. V.74. P.5463-5467. doi: 10.1073/pnas.74.12.5463
37. Shendure J, Balasubramanian S, Church GM et al. DNA sequencing at 40: past, present and future // *Nature*. 2017. V.550(7676). P.345-353. doi: 10.1038/nature24286
38. Shendure J, Ji H. Next-generation DNA sequencing // *Nat Biotechnol*. 2008. V.26(10). P.1135-1145. doi: 10.1038/nbt1486
39. Sverdlov ED, Monastyrskaya GS, Budowsky EI, Grachev MA. A novel approach to structural analysis of oligonucleotides // *FEBS Lett*. 1972. V.28(2). P.231-235. doi: 10.1016/0014-5793(72)80719-1
40. Sverdlov ED, Monastyrskaya GS, Chestukhin AV, Budowsky EI. The primary structure of oligonucleotides. Partial apurination as a method to determine the positions of purine and pyrimidine residues // *FEBS Lett*. 1973. V.33(1). P.15-17. doi: 10.1016/0014-5793(73)80148-6
41. Venter JC. Multiple personal genomes await // *Nature*. 2010. V.464. P.676-677. doi: 10.1038/464676a
42. Venter J.C., Adams M.D., Myers E.W. et al. The sequence of the human genome // *Science*. 2001. V.291(5507). P.1304-1351. doi: 10.1126/science.1058040
43. Wada A, Yamamoto M, Soeda E. Automatic DNA sequencer: computer-programmed microchemical manipulator for the Maxam-Gilbert sequencing method // *Rev Sci Instrum*. 1983. V.54(11). P.1569-1572. doi: 10.1063/1.1137299
44. Walsh J.B., Marks J. Sequencing the human genome // *Nature*. 1986. V.322. P. 590. doi: 10.1038/322590a0
45. Wheeler D.A., Srinivasan M., Egholm M. et al. The complete genome of an individual by massively parallel DNA sequencing // *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
46. Zavgorodny SG, Pechenov AE, Shvets VI, Miroshnikov AI. S,X-acetals in nucleoside chemistry. III. Synthesis of 2'- and 3'-O-azidomethyl derivatives of ribonucleosides // *Nucleosides Nucleotides Nucleic Acids*. 2000. V.19(10-12). P.1977-1991. doi: 10.1080/15257770008045472

#### 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. Air GM, Blackburn EH, Sanger F, Coulson AR. The nucleotide and amino acid sequences of the N (5') terminal region of gene G of bacteriophage phiX174. *J Mol Biol*. 1975. V.96(4). P.703-719. doi: 10.1016/0022-2836(75)90147-3
3. Baymiev Al.Kh., Chemeris D.A., Sakhabutdinova A.R., Kuluev A.R., Vershinina Z.R., Mikhailova E.V., Matniyazov R.T., Baymiev An.Kh., Kuluev B.R., Garafutdinov R.R., Chemeris A.V. In higher plants as an example, one can see that the era of sequencing of their diploid genomes is coming. *Biomics*. 2025. V.17(1). P. 17-41. DOI: 10.31301/2221-6197.bmcs.2025-3
4. Chargaff E. What really is DNA? Remarks on the changing aspects of a scientific concept. *Prog. Nucl. Acid Res. Mol. Biol*. 1968. V.8. P.297-333. doi: 10.1016/s0079-6603(08)60549-8
5. Chemeris A.V., Akhunov E.D., Vakhitov V.A. DNA sequencing. Moscow. Nauka. 1999. 429 p. (In Russian)
6. Chemeris A.V., Vakhitov V.A. The primary structure of the 5.8S rRNA gene and the internal transcribed spacers of rDNA in the diploid wheat *Triticum urartu* Thum. ex Gandil. *Molecular Biology*. 1989. V.23. P.320-326.

7. Chetverin A.B., Chetverina H.V. 1997. Method for amplification of nucleic acids in solid media. U.S. Patent 5,616,478
8. Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing. *Nature Biotechnol.* 2016. V.34(5). P.518-524. doi: 10.1038/nbt.3423
9. Collins FS, Patrinos A, Jordan E et al. New goals for the U.S. Human Genome Project: 1998-2003. *Science.* 1998. V.282(5389). P.682-689. doi: 10.1126/science.282.5389.682
10. Eid J., Fehr A., Gray J. et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009. V.323. P.133-138. doi: 10.1126/science.1162986
11. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics.* 2015. V.13(1). P.4-16. doi: 10.1016/j.gpb.2015.01.009
12. Frazer KA, Schork NJ. The human pangenome reference anticipates equitable and fundamental genomic insights. *Cell Genom.* 2023. V.3(7). 100360. doi: 10.1016/j.xgen.2023.100360
13. Jarvis ED, Formenti G, Rhie A et al. Semi-automated assembly of high-quality diploid human reference genomes. *Nature.* 2022. V.611(7936). P.519-531. doi: 10.1038/s41586-022-05325-5
14. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004. V.431(7011). P.931-945. doi: 10.1038/nature03001
15. Istrail S., Sutton G.G., Florea L. et al. Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl. Acad. Sci. USA.* 2004. V.101(7). P.1916-1921. doi: 10.1073/pnas.0307971100
16. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA.* 1996. V.93(24). P.13770-13773. doi: 10.1073/pnas.93.24.13770
17. Khrapko KR, Lysov YuP, Khorlyn AA et al. An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* 1989. V.256(1-2). P.118-122. doi: 10.1016/0014-5793(89)81730-2
18. Kolata GB. DNA sequencing: a new era in molecular biology. *Science.* 1976. V.192(4240). P.645-647. doi: 10.1126/science.192.4240.645
19. Kokoris M, McRuer R, Nabavi M et al. Sequencing by Expansion (SBX) - a novel, high-throughput single-molecule sequencing technology. *bioRxiv* [Preprint]. 2025. 2025.02.19.639056. doi: 10.1101/2025.02.19.639056
20. Kuluev A.R., Matniyazov R.T., Berezin A.A. et al. Pan-mitogenomics. *Biomics.* 2025. V.17(1). P. 88 - 102. DOI: 10.31301/2221-6197.bmcs.2025-7 (In Russian)
21. Kuluev B.R., Baymiev An.Kh., Gerashchenkov G.A. et al. 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)
22. Lander ES., Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature.* 2001. V.409(6822). P.860-921. doi: 10.1038/35057062
23. Levy S., Sutton G., Ng P.C. et al. The diploid genome sequence of an individual human. *PLoS Biol.* 2007. V.5(10). e254. doi: 10.1371/journal.pbio.0050254
24. Logsdon GA, Ebert P, Audano PA et al. Complex genetic variation in nearly complete human genomes. *bioRxiv* [Preprint]. 2024. 25:2024.09.24.614721. doi: 10.1101/2024.09.24.614721
25. Mandeles S. Nucleic Acid Sequence Analysis. Columbia University Press. 1972. 282 P.
26. Maxam A.M., Gilbert W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 1977. V.74. P.560-564. doi: 10.1073/pnas.74.2.560
27. 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
28. Porubsky D, Vollger MR, Harvey WT et al. Gaps and complex structurally variant loci in phased genome assemblies. *Genome Res.* 2023. V.33(4). P.496-510. doi: 10.1101/gr.277334.122
29. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
30. Salser WA. DNA sequencing techniques. *Annu Rev Biochem.* 1974. V.43. P.923-965. doi: 10.1146/annurev.bi.43.070174.004423
31. Sanger F, Air GM, Barrell BG et. Al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature.* 1977. V.265(5596). P.687-695. doi: 10.1038/265687a0
32. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975. V.94(3). P.441-448. doi: 10.1016/0022-2836(75)90213-2
33. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 1977. V.74. P.5463-5467. doi: 10.1073/pnas.74.12.5463
34. Shendure J, Balasubramanian S, Church GM et al. DNA sequencing at 40: past, present and future. *Nature.* 2017. V.550(7676). P.345-353. doi: 10.1038/nature24286
35. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008. V.26(10). P.1135-1145. doi: 10.1038/nbt1486
36. Skryabin K.G., Prokhortchouk E.B., Mazur A.M. et al. Combining two technologies for full genome

- sequencing of human. *Acta Naturae*. 2009. V.1(3). P. 102-107.
37. Sverdlov E.D. Podhody k issledovaniju pervichnoj struktury DNK. Itogi nauki i tehniki. M.: VINITI. Ser. Molekuljarnaja biologija. T.4. S.7-88. [Approaches to the study of the primary structure of DNA] (In Russian)
38. Sverdlov ED, Monastyrskaya GS, Budowsky EI, Grachev MA. A novel approach to structural analysis of oligonucleotides. *FEBS Lett*. 1972. V.28(2). P.231-235. doi: 10.1016/0014-5793(72)80719-1
39. Sverdlov ED, Monastyrskaya GS, Chestukhin AV, Budowsky EI. The primary structure of oligonucleotides. Partial apurination as a method to determine the positions of purine and pyrimidine residues. *FEBS Lett*. 1973. V.33(1). P.15-17. doi: 10.1016/0014-5793(73)80148-6
40. Venter JC. Multiple personal genomes await. *Nature*. 2010. V.464. P.676-677. doi: 10.1038/464676a
41. Venter J.C., Adams M.D., Myers E.W. et al. The sequence of the human genome. *Science*. 2001. V.291(5507). P.1304-1351. doi: 10.1126/science.1058040
42. Wada A, Yamamoto M, Soeda E. Automatic DNA sequencer: computer-programmed microchemical manipulator for the Maxam-Gilbert sequencing method. *Rev Sci Instrum*. 1983. V.54(11). P.1569-1572. doi: 10.1063/1.1137299
43. Walsh J.B., Marks J. Sequencing the human genome. *Nature*. 1986. V.322. P. 590. doi: 10.1038/322590a0
44. Wheeler D.A., Srinivasan M., Egholm M. et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
45. Zavgorodny SG, Pechenov AE, Shvets VI, Miroshnikov AI. S,X-acetals in nucleoside chemistry. III. Synthesis of 2'- and 3'-O-azidomethyl derivatives of ribonucleosides. *Nucleosides Nucleotides Nucleic Acids*. 2000. V.19(10-12). P.1977-1991. doi: 10.1080/15257770008045472
46. Zubov V.V., Chemeris D.A., Vasilov R.G. et al. Brief history of high-throughput nucleic acid sequencing methods. *Biomics*. 2021. V.13(1). P. 27-46. DOI: 10.31301/2221-6197.bmcs.2021-4 (In Russian)