


doi 10.18699/vjgb-25-98

Отражение процессов повреждения ДНК в эволюции G-трактов в геномах

И.Р. Грин ¹, Д.О. Жарков ^{1, 2} 

¹ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 dzharkov@niboch.nsc.ru

Аннотация. Окисление ДНК представляет собой один из главных видов повреждения генетического материала живых организмов. Из многих десятков продуктов окислительного повреждения ДНК в наибольшем количестве встречается 8-оксогуанин (8-охоG) – предмутагенное основание, приводящее при репликации к трансверсиям G→T. Двухцепочечная ДНК обладает способностью к проводимости положительных зарядов, связанных с дефицитом электронов в π -системе азотистых оснований. Такие заряды в конечном итоге локализуются на 5'-концевом нуклеотиде полигуаниновых трактов (G-трактов). В связи с этим 5'-концевые нуклеотиды G-трактов служат характерными местами образования 8-охоG. Эти свойства G-трактов хорошо изучены *in vitro* на уровне реакционной способности, но остается неясным, насколько они могут отражаться в спектрах мутагенеза *in vivo*. В работе проанализирован нуклеотидный контекст G-трактов в репрезентативном наборе из 62 полных геномов прокариот и в геноме человека с покрытием «от теломеры до теломеры». Показано, что G-тракты в среднем короче полиадениновых трактов (A-трактов) и вероятность удлинения G-трактов на один нуклеотид ниже, чем в случае A-трактов. Установлено, что представленность T в положении, примыкающем к G-трактам с 5'-стороны, повышена, в особенности у организмов с аэробным метаболизмом, что согласуется с моделью преимущественных мутаций G→T в 5'-положении с 8-охоG как предшественником. В то же время в положении, примыкающем к A-трактам, повышена частота встречаемости G и C и снижена частота встречаемости T. В геноме человека наблюдается двухфазный характер разрастания G-трактов: начиная с длины 8–9 нуклеотидов вероятность их удлинения на один нуклеотид заметно увеличивается. Выявлена повышенная представленность C с 5'-стороны от длинных G-трактов и A при заменах в теломерных повторах, что может свидетельствовать о существовании мутагенных процессов, механизм которых пока не охарактеризован, но может быть связан с ошибками ДНК-полимераз при репликации продуктов дальнейшего окисления 8-охоG.

Ключевые слова: повреждение ДНК; мутагенез; 8-оксогуанин; G-тракты; теломеры

Для цитирования: Грин И.Р., Жарков Д.О. Отражение процессов повреждения ДНК в эволюции G-трактов в геномах. *Вавиловский журнал генетики и селекции*. 2025;29(7):913-924. doi 10.18699/vjgb-25-98


Финансирование. Исследование выполнено при финансовой поддержке РНФ в рамках проекта № 24-14-00285 (анализ генома человека). Часть работы, посвященная анализу геномов прокариот, поддержана государственным заданием ИХБФМ СО РАН 125012300657-2.

DNA damage reflected in the evolution of G-runs in genomes

I.R. Grin ¹, D.O. Zharkov ^{1, 2} 

¹ Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

 dzharkov@niboch.nsc.ru

Abstract. DNA oxidation is one of the main types of damage to the genetic material of living organisms. Of the many dozens of oxidative lesions, the most abundant is 8-oxoguanine (8-oxoG), a premutagenic base that leads to G→T transversions during replication. Double-stranded DNA can conduct holes through the π system of stacked nucleobases. Such electron vacancies are ultimately localized at the 5'-terminal nucleotides of polyguanine runs (G-runs), making these positions characteristic sites of 8-oxoG formation. While such properties of G-runs have been studied *in vitro* at the level of chemical reactivity, the extent to which they can influence mutagenesis spectra *in vivo* remains unclear. Here, we have analyzed the nucleotide context of G-runs in a representative set of 62 high-quality prokaryotic genomes and in the human telomere-to-telomere genome. G-runs were, on average, shorter than polyadenine runs (A-runs), and the probability of a G-run being elongated by one nucleotide is lower than in the case of A-runs. The representation of T in the position 5'-flanking G-runs is increased, especially in organisms with aerobic metabolism, which is consistent with the model of preferential G→T substitutions at the 5'-position with 8-oxoG as a precursor. Conversely, the frequency of G and C is increased and the frequency of T is decreased in the position 5'-flanking A-runs. A biphasic pattern of G-run expansion is observed in the human genome: the probability of sequences longer than 8–9 nucleotides being

elongated by one nucleotide increases significantly. An increased representation of C in the 5'-flanking position to long G-runs was found, together with an elevated frequency of 5'-G→A substitutions in telomere repeats. This may indicate the existence of mutagenic processes whose mechanism has not yet been characterized but may be associated with DNA polymerase errors during replication of the products of further oxidation of 8-oxoG.

Key words: DNA damage; mutagenesis; 8-oxoguanine; G-runs; telomeres

For citation: Grin I.R., Zharkov D.O. DNA damage reflected in the evolution of G-runs in genomes. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed*. 2025;29(7):913-924. doi 10.18699/vjgb-25-98

Введение

Окислительное повреждение ДНК – неизбежный спутник метаболизма, использующего окисление органических соединений молекулярным кислородом, который составляет основу энергетического обмена подавляющего большинства живых организмов уже более двух миллиардов лет (Prorok et al., 2021). Поврежденные нуклеотиды обычно достаточно быстро подвергаются репарации, однако некоторая их часть может оставаться в составе ДНК до момента репликации, что служит одной из главных причин мутаций (Liu et al., 2016; Chatterjee, Walker, 2017; Tubbs, Nussenzweig, 2017). Благодаря пониманию механизмов возникновения ошибок при работе ДНК-полимераз на молекулярном уровне в настоящее время стало возможным выделять характерные паттерны мутаций, вызванных тем или иным видом генотоксического стресса или даже конкретными поврежденными азотистыми основаниями (Alexandrov et al., 2013; Koh et al., 2021).

Из всех структурных элементов ДНК самым низким окислительно-восстановительным потенциалом характеризуется гуанин (Cadet et al., 2008, 2017; Fleming, Burrows, 2022). Самый распространенный продукт его окисления – 7,8-дигидро-8-оксогуанин (8-охоG) – встречается в ДНК на фоновом уровне $\sim 1/10^6$ гуанинов, и этот уровень значительно повышается при окислительном стрессе различного происхождения (ESCODD et al., 2005; Dizdaroglu et al., 2015; Chiorcea-Paquim, 2022; рис. 1, а, б). Ввиду того, что при атоме C8 в 8-охоG находится объемный атом кислорода, в соответствующем нуклеозиде, 8-оксо-2'-дезоксигуанине (8-oxodG), стандартная *анти*-конформация стерически затруднена, а *син*-конформация оказывается более выгодной (Cho et al., 1990; см. рис. 1, в, г). Это при-

водит к тому, что в отсутствие уотсон-криковских связей с цитозином, дополнительно стабилизирующих *анти*-конформацию, 8-оксо-2'-дезоксигуанин преимущественно переходит в *син*-конформацию, в которой может образовывать пару хугстеновского типа с аденином (Kouchakdjian et al., 1991; McAuley-Hecht et al., 1994; Lipscomb et al., 1995). Из-за этого ДНК-полимеразы с высокой частотой включают dAMP напротив 8-охоG в ДНК-матрице (Shibutani et al., 1991; Miller, Grollman, 1997; Maga et al., 2007; Yudkina et al., 2019).

В живых клетках на исход первичных событий окисления ДНК может оказывать влияние множество дополнительных факторов, и существуют системы репарации ДНК, удаляющие из генома поврежденные основания. Даже с учетом этого 8-охоG проявляет заметную мутагенность *in vivo*, которая характеризуется спектром с доминирующими трансверсиями G→T, слабо зависящими от окружающего однонуклеотидного контекста (Wood et al., 1992; Moriya, 1993). Такие мутации часто встречаются в опухолях человека и составляют основу мутационных подписей SBS18 и SBS36 (Alexandrov et al., 2013; Pilati et al., 2017; Viel et al., 2017; Kucab et al., 2019). Заметный вклад в мутагенез вносят и продукты дальнейшего окисления 8-охоG – гуанидиногидантоин и спироиминодигидантоин, которые преимущественно вызывают трансверсии G→C (Fleming, Burrows, 2017; Kino et al., 2020).

Сопряженная π -система ДНК обладает достаточно высокой дырочной проводимостью (Giese, 2002; Genereux, Barton, 2010). Многочисленные эксперименты и квантомеханические расчеты показывают, что при одноэлектронном окислении ДНК образовавшийся положительный заряд мигрирует по π -системе на значительные расстоя-

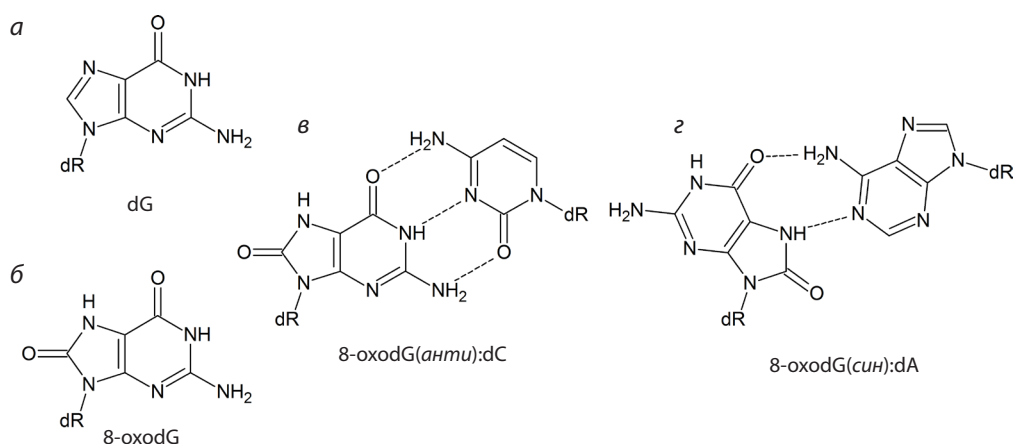


Рис. 1. Структуры 2'-дезоксигуанина (а), 8-оксо-2'-дезоксигуанина (б), пары уотсон-криковского типа 8-oxodG(*анти*):dC (в) и пары хугстеновского типа 8-oxodG(*син*):dA (г).

ния, а его конечными акцепторами служат основания G, которые главным образом окисляются до 8-охоG. При этом особенно чувствительны к окислению гуанины, находящиеся в крайнем 5'-положении в трактах из нескольких G (Sugiyama, Saito, 1996; Saito et al., 1998; Kurbanyan et al., 2003; Adhikary et al., 2009).

Несмотря на то что механизм миграции положительно заряда и преимущественного окисления гуанинов на 5'-конце G-трактов сегодня считается общепринятым, все экспериментальные данные в его поддержку получены в простых системах *in vitro*. Исследований спектров мутагенеза, вызванного появлением 8-охоG в таком контексте, до сих пор не проводилось. Если на 5'-конце G-трактов действительно происходит преимущественное превращение G в 8-охоG, можно ожидать, что из-за мутагенных свойств 8-охоG в таких позициях будет повышена частота мутаций G→T, что должно отражаться в увеличении встречаемости T перед трактами из нескольких G. В данной работе с целью проверки этой гипотезы проведен анализ встречаемости нуклеотидов, примыкающих с 5'-стороны к G-трактам (5'-фланговых нуклеотидов), в геномах прокариот и человека.

Материалы и методы

Для анализа были использованы сборка генома человека T2T-CHM13v2.0, включающая полноразмерные теломеры и высокоповторные участки (Nurk et al., 2022), и геномы прокариот, перечисленные в табл. 1.

Для работы с последовательностями геномов использовали программный пакет UGENE v37.0 (Okonechnikov et al., 2012) и специально написанные скрипты на языке bash. Ожидаемую частоту нуклеотидов во фланговых позициях перед повторами G_n (либо A_n) и после них в геномах прокариот рассчитывали, исходя из общего количества A, C и T (либо C, G и T) в данном геноме как $p_A = N_A / (N_A + N_C + N_T)$, где p_A – ожидаемая представленность (в данном случае для A), N_A , N_C и N_T – число A, C и T в обеих цепях генома соответственно. Для генома человека, ввиду известной недопредставленности динуклеотида CG, ожидаемую частоту рассчитывали аналогичным образом, исходя из представленности динуклеотидов AG, CG и TG. Статистический анализ проводили в программах SigmaPlot v11.0 (Grafiti, США), DATAPLOT (Национальный институт стандартов и технологий США) и RStudio v1.2 (Posit PBC, США). При всех множественных сравнениях и сериях тестов для корректировки уровня значимости использовали поправку Данн.

Результаты и обсуждение

Для анализа распределения нуклеотидов в геномах прокариот была составлена выборка из 54 геномов бактерий и 8 геномов архей, максимально отражающая таксономическое разнообразие в этих доменах живых организмов (см. табл. 1). В нее вошли геномы высокого качества, отнесенные в базе данных RefSeq (O'Leary et al., 2016) к референтным. Репрезентативность выборки составляла один геном на таксон уровня типа, за исключением типов Methanobacteriota и Thermoproteota для архей и Actinomycetota, Bacteroidota и Thermodesulfobacteriota для бактерий с репрезентативностью два генома из разных порядков на

тип, а также Bacillota и Pseudomonadota – три генома из разных порядков на тип. Содержание G+C в исследуемых геномах варьировало от 23.5 до 69 % (см. табл. 1). Геномы архей по своим параметрам достоверно не отличались от геномов бактерий, поэтому представители обоих доменов рассматривались как одна группа прокариот.

Поскольку большую часть прокариотических геномов составляют белок-кодирующие последовательности, мутации в которых могут подвергаться действию естественного отбора, предварительно было оценено возможное влияние всех 16 потенциальных замен аминокислотных остатков, происходящих в результате нуклеотидных замен G→A, G→C и G→T в первой позиции G-тракта (замены кодонов HNG→HHN, HGG→HHG, GGG→HGG, где H – A, C или T). Для этого использовали две независимые метрики: индекс консервативности C_n , рассчитанный на основе условных расстояний в наборе физико-химических свойств аминокислотных остатков (Taylor, 1986; Livingstone, Barton, 1993), и веса аминокислотных замен в матрице BLOSUM62, составленной по набору нескольких сотен групп гомологичных белков (Henikoff S., Henikoff J.G., 1992). Хотя замены G→A в целом вызывали меньшие изменения в свойствах и встречаемости аминокислотных остатков, как и следовало ожидать для точковых мутаций с сохранением класса нуклеотида, отличие от замен G→C и G→T не было статистически значимым (критерий Краскела–Уоллиса с поправкой Данн на множественные сравнения, $p > 0.05$).

Во всех геномных последовательностях был проведен поиск трактов HG_nH , BA_nB и соответствующих им по комплементарной цепи трактов DC_nD и VT_nV (H – A, C или T; B – C, G или T; D – A, G или T; V – A, C или G) с длиной $n \geq 2$. Встречаемость полипуриновых трактов в геномах была выше, чем можно ожидать из случайного распределения нуклеотидов при том же составе G+C (одновыборочный критерий Уилкоксона, $p < 0.001$), что говорит о функциональной важности таких последовательностей. Повышенная частота замен в первой позиции G-трактов должна постепенно приводить к их укорочению. Действительно, при сравнении длин G- и A-трактов в прокариотических геномах с поправкой на содержание соответствующих пуриновых нуклеотидов оказалось, что G-тракты в среднем короче (рис. 2, а). При этом тринуклеотиды HGG встречались чаще, чем BAA, однако при большей длине повтора встречаемость A-трактов была выше (см. рис. 2, б).

Для более подробного анализа распределения длин трактов была изучена вариабельность их длин в каждом геноме. Количество G- и A-трактов в составе каждого генома убывало практически строго экспоненциально в диапазоне длин от 2 до 5–6. При больших длинах в некоторых случаях наблюдались отклонения в любую сторону из-за небольшого числа таких повторов, в особенности для геномов малого размера (рис. 3, а, б). По линейной части зависимости логарифма количества повторов от длины тракта можно определить «коэффициент приращения» k_{inc} , который показывает, насколько легко в геноме данного нуклеотидного состава удлиннить тракт на один нуклеотид: чем выше k_{inc} , тем больше в геноме пропорция относительно более длинных трактов. При сравнении

Таблица 1. Геномы прокариот, использованные для анализа

Вид	Тип	Сборка генома	G+C, %	O ₂
Домен Archaea				
<i>Methanobacterium formicicum</i>	Methanobacteriota	GCF_001458655.1	41.0	AN
<i>Methanosarcina barkeri</i>		GCF_000970025.1	39.0	AN
<i>Nanobdella aerobiophila</i>	Nanobdellota	GCF_023169545.1	24.5	A
<i>Nitrososphaera viennensis</i>	Nitrososphaerota	GCF_000698785.1	52.5	A
<i>Promethearchaeum syntrophicum</i>	Promethearchaeota	GCF_008000775.2	31.0	AN
<i>Sulfolobus acidocaldarius</i>	Thermoproteota	GCF_000012285.1	36.5	A
<i>Thermoproteus tenax</i>		GCF_000253055.1	55.0	AN
<i>Cand. Nanoalobium constans</i>	Cand. Nanoalarchaeota	GCF_009617975.1	43.0	A
Домен Bacteria				
<i>Acidobacterium capsulatum</i>	Acidobacteriota	GCF_000022565.1	60.5	A
<i>Bifidobacterium longum</i>	Actinomycetota	GCF_000196555.1	60.5	AN
<i>Mycobacterium tuberculosis</i>		GCF_000195955.2	65.5	A
<i>Aquifex aeolicus</i>	Aquificota	GCF_000008625.1	43.5	A
<i>Fimbriimonas ginsengisoli</i>	Armatimonadota	GCF_000724625.1	61.0	A
<i>Atribacter laminatus</i>	Atribacterota	GCF_015775515.1	38.5	AN
<i>Bacillus subtilis</i>	Bacillota	GCF_000009045.1	43.5	A
<i>Clostridioides difficile</i>		GCF_018885085.1	28.5	AN
<i>Lactococcus lactis</i>		GCF_003176835.1	35.0	A
<i>Bacteroides fragilis</i>	Bacteroidota	GCF_000025985.1	43.0	AN
<i>Saprospira grandis</i>		GCF_000250635.1	46.5	A
<i>Cyclonatronum proteinivorum</i>	Balneolota	GCF_003353065.1	51.5	A
<i>Bdellovibrio bacteriovorus</i>	Bdellovibrionota	GCF_000196175.1	50.5	A
<i>Caldisericum exile</i>	Caldisericotota	GCF_000284335.1	35.5	AN
<i>Caldithrix abyssi</i>	Calditrichota	GCF_001886815.1	45.0	AN
<i>Campylobacter jejuni</i>	Campylobacterota	GCF_000009085.1	30.5	A
<i>Chlamydia trachomatis</i>	Chlamydiota	GCF_000008725.1	41.5	AN
<i>Chlorobium limicola</i>	Chlorobiota	GCF_000020465.1	51.5	AN
<i>Chloroflexus aurantiacus</i>	Chloroflexota	GCF_000018865.1	56.5	A
<i>Desulfurispirillum indicum</i>	Chrysiogenota	GCF_000177635.2	56.0	AN
<i>Coprothermobacter proteolyticus</i>	Coprothermobacterota	GCF_000020945.1	45.0	AN
<i>Synechococcus elongatus</i>	Cyanobacteriota	GCF_022984195.1	55.5	A
<i>Deferribacter thermophilus</i>	Deferribacterota	GCF_049472675.1	30.5	AN
<i>Deinococcus radiodurans</i>	Deinococcota	GCF_020546685.1	66.5	A
<i>Dictyoglomus thermophilum</i>	Dictyoglomota	GCF_000020965.1	33.5	AN
<i>Elusimicrobium minutum</i>	Elusimicrobiota	GCF_000020145.1	40.0	AN
<i>Fibrobacter succinogenes</i>	Fibrobacterota	GCF_000146505.1	48.0	AN
<i>Fidelibacter multiformis</i>	Fidelibacterota	GCF_041154365.1	45.5	AN
<i>Fusobacterium nucleatum</i>	Fusobacteriota	GCF_003019295.1	27.0	AN
<i>Gemmatimonas aurantiaca</i>	Gemmatimonadota	GCF_000010305.1	64.5	A
<i>Ignavibacterium album</i>	Ignavibacteriota	GCF_000258405.1	34.0	A
<i>Kiritimatiella glycovorans</i>	Kiritimatiellota	GCF_001017655.1	63.5	AN
<i>Lentisphaera profundii</i>	Lentisphaerota	GCF_028728065.1	40.5	A
<i>Mycoplasma mycoides</i>	Mycoplasmata	GCF_018389705.1	23.5	A
<i>Myxococcus xanthus</i>	Myxococcota	GCF_000012685.1	69.0	A
<i>Nitrospina watsonii</i>	Nitrospinota	GCF_946900835.1	57.0	A
<i>Nitrospira moscoviensis</i>	Nitrospirota	GCF_001273775.1	62.0	A
<i>Planctopirius limnophila</i>	Planctomycetota	GCF_000092105.1	53.5	A

Окончание табл. 1

Вид	Тип	Сборка генома	G+C, %	O ₂
<i>Escherichia coli</i>	Pseudomonadota	GCF_000005845.2	51.0	A
<i>Pseudomonas aeruginosa</i>		GCF_000006765.1	66.5	A
<i>Sphingomonas paucimobilis</i>		GCF_016027095.1	65.5	A
<i>Rhodothermus marinus</i>	Rhodothermota	GCF_000024845.1	64.5	A
<i>Spirochaeta thermophila</i>	Spirochaetota	GCF_000184345.1	61.0	AN
<i>Thermanaerovibrio acidaminovorans</i>	Synergistota	GCF_000024905.1	64.0	AN
<i>Desulfovibrio desulfuricans</i>	Thermodesulfobacteriota	GCF_017815575.1	57.0	AN
<i>Thermodesulfobacterium commune</i>		GCF_000734015.1	37.0	AN
<i>Thermodesulfobium narugense</i>	Thermodesulfobiota	GCF_000212395.1	34.0	AN
<i>Thermomicrobium roseum</i>	Thermomicrobiota	GCF_000021685.1	64.5	A
<i>Thermosulfidibacter takaii</i>	Thermosulfidibacterota	GCF_001547735.1	43.0	AN
<i>Thermotoga maritima</i>	Thermotogota	GCF_000230655.2	46.0	AN
<i>Verrucomicrobium spinosum</i>	Verrucomicrobiota	GCF_000172155.1	60.5	A
<i>Vulcanimicrobium alpinum</i>	Vulcanimicrobiota	GCF_027923555.1	68.5	A
Cand. <i>Cloacimonas acidaminovorans</i>	Cand. Cloacimonadota	GCF_000146065.2	38.0	AN
Cand. <i>Velamenicoccus archaeovorans</i>	Cand. Omnitrophota	GCF_004102945.1	53.0	AN

Примечание. Номер сборки генома взят из базы данных RefSeq (O'Leary et al., 2016). A – аэробы и факультативные анаэробы; AN – анаэробы.

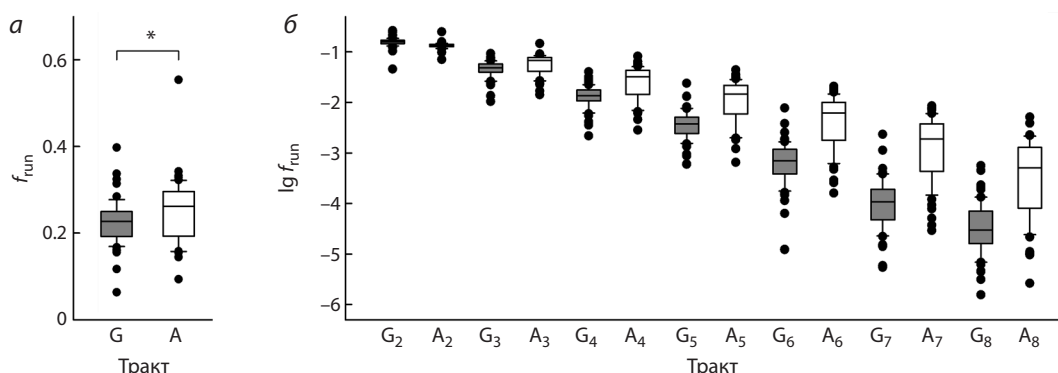


Рис. 2. Длина полипуриновых трактов в геномах прокариот. *a* – общая доля G или A в составе трактов любой длины среди числа соответствующих пуриновых нуклеотидов в геноме (f_{run}), * $p < 0.05$ (критерий Манна–Уитни); *б* – доля G или A в составе трактов длины 2–8 среди числа соответствующих пуриновых нуклеотидов в геноме. Во всех случаях разница между G- и A-трактами значима на уровне $p < 0.001$ (критерий Манна–Уитни).

Здесь и на последующих рисунках линия в «ящике» соответствует медиане, границы «ящика» – первому и третьему квартилям, «усы» – 10-му и 90-му перцентилям, точки – значениям выбросов.

зависимости значений k_{inc} для G- и A-трактов в геномах разного состава выяснилось, что G-тракты с возрастанием содержания G+C удлиняются медленнее, чем удлиняются A-тракты с возрастанием содержания A+T в геноме (см. рис. 3, *в*). Таким образом, баланс удлинения и укорочения G-трактов, определяемый многими факторами, в геномах прокариот смещен в сторону укорочения по сравнению с A-трактами.

Длины полипуриновых трактов могут изменяться в любую сторону из-за проскальзывания ДНК-полимераз при синтезе ДНК (Kunkel, Bebenek, 2000) или в результате отбора по физико-химическим свойствам полипуриновых участков (Bansal et al., 2022), однако эти процессы не зависят от окружающих тракт нуклеотидов. Напротив, укорочение G-трактов вследствие повреждения 5'-кон-

цевого основания должно сопровождаться характерным мутационным спектром, который определяется свойствами репликативных ДНК-полимераз. В связи с этим было интересно установить, насколько представленности 5'-фланговых нуклеотидов отличаются друг от друга и от их общей представленности по геному. Для количественной характеристики этих отличий был введен параметр $\Delta_{гер}$ – разница между наблюдаемой и ожидаемой представленностью каждого нуклеотида. Оказалось, что представленность нуклеотида T в первой позиции перед G-трактами статистически значимо повышена по сравнению с ожидаемой и по сравнению с представленностью A и C (рис. 4, *а*). Встречаемость нуклеотидов A и C в этой позиции была несколько ниже ожидаемой, но это отличие не достигало значимости; также по своей представленности

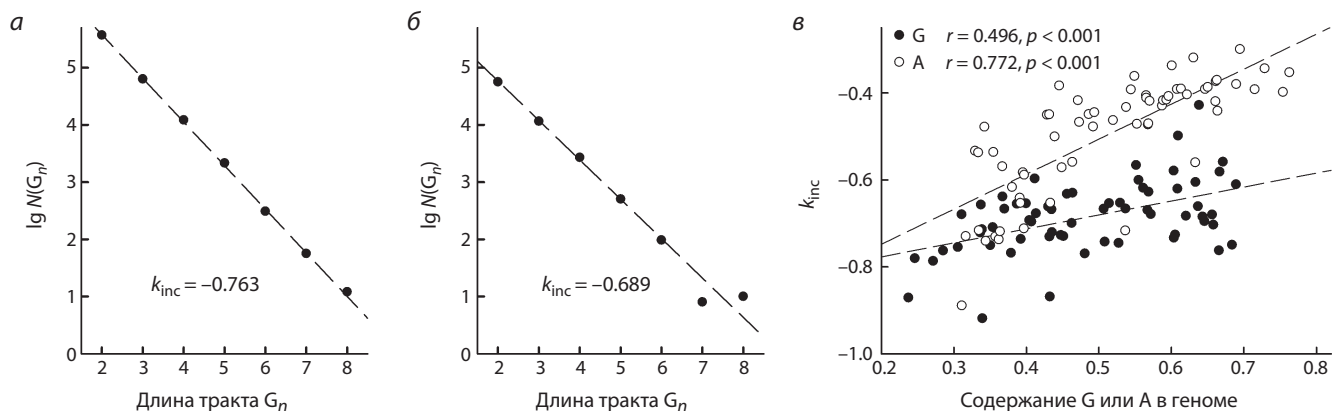


Рис. 3. Зависимость количества полипуриновых трактов в геномах прокариот от длины тракта и нуклеотидного состава генома.

a, б – примеры зависимости числа G-трактов $N(G_n)$ от их длины для геномов *E. coli* (*a*), размер генома 4.64×10^6 п.н., содержание G+C 51.0 %, и *Ch. trachomatis* (*б*), размер генома 1.04×10^6 п.н., содержание G+C 41.5 %; *в* – зависимость k_{inc} от нуклеотидного состава генома (содержания G+C для G-трактов, содержания A+T для A-трактов). Черные точки – G-тракты, белые точки – A-тракты; штриховые линии соответствуют линейной регрессии, коэффициенты регрессии приведены на рисунке.

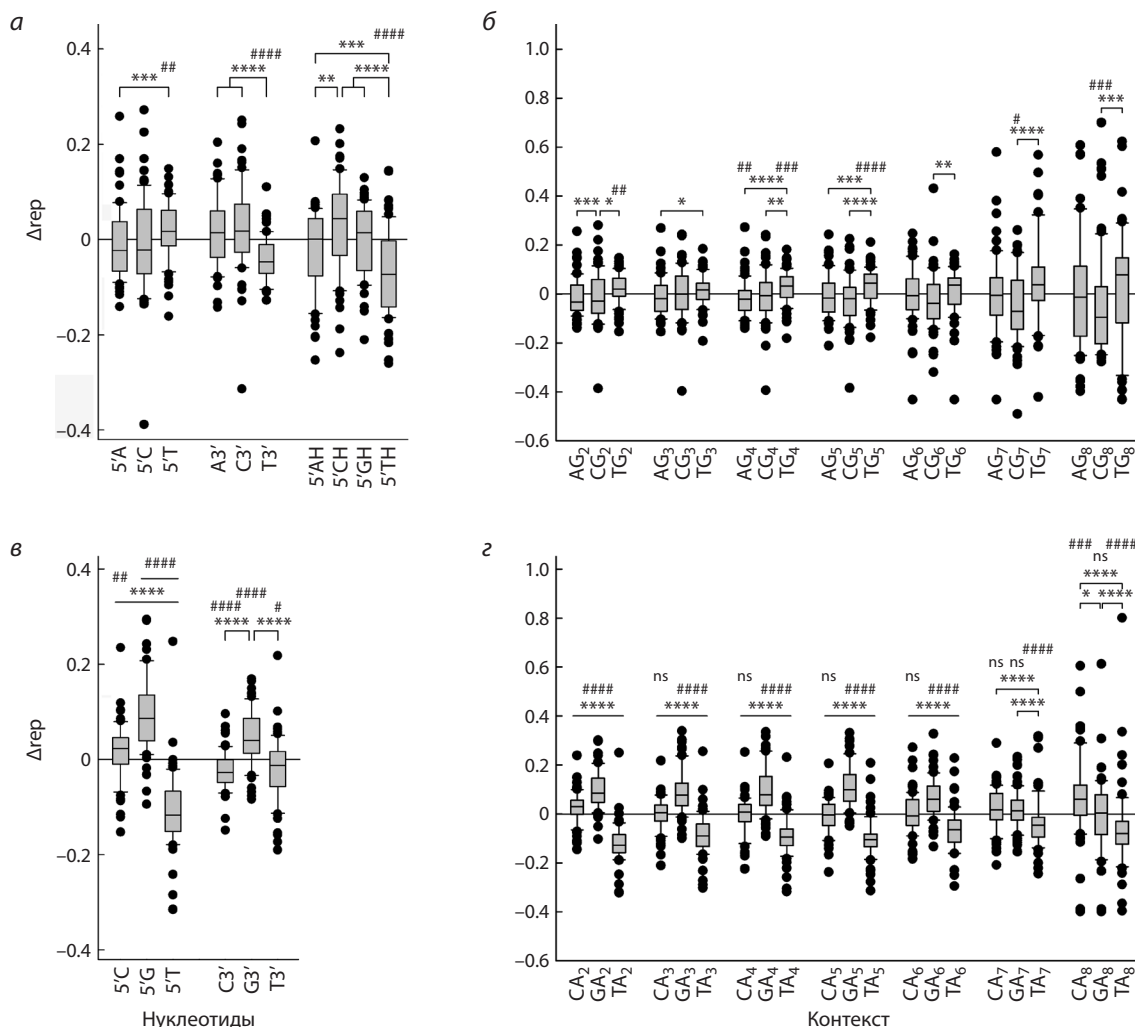


Рис. 4. Представленность разных 5'- и 3'-фланговых нуклеотидов при полипуриновых трактах.

a, в – отклонение от представленности 5'- и 3'-фланговых нуклеотидов для G-трактов (*a*) и A-трактов (*в*), ожидаемой на основе содержания соответствующего нуклеотида в геноме, в трактах любой длины. *б, г* – отклонение от представленности 5'-фланговых нуклеотидов в G-трактах (*б*) и A-трактах (*г*) длины 2–8. Отличие от ожидаемого: # $p < 0.05$, ## $p < 0.01$, ### $p < 0.005$, #### $p < 0.001$ (одно-выборочный критерий Уилкоксона с поправкой Данн на множественные сравнения); ns – нет достоверных отличий. Различия между группами: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$ (критерий Краскела–Уоллиса с поправкой Данн на множественные сравнения).

они не отличались друг от друга. Т встречался чаще, чем хотя бы один из нуклеотидов А или С, при любой длине G-тракта, при этом его представленность по сравнению с ожидаемой была выше перед трактами G₂, G₄, G₅ и G₆ (см. рис. 4, б). Нуклеотид А был недопредставлен в этой позиции лишь перед трактами G₄, а нуклеотид С – перед G-трактами длиной 5 и более. Напротив, как с 3'-стороны от G-трактов, так и во второй позиции с 5'-стороны от них Т был представлен меньше ожидаемого (см. рис. 4, а). В целом эти данные свидетельствуют в пользу модели преимущественного окисления первого G в трактах до 8-охоG с последующей трансверсией G→Т.

Достаточно неожиданным оказалось то, что перед А-трактами распределение нуклеотидов было неравномерным в еще большей степени, чем перед G-трактами. В этой позиции Т был недопредставлен, а С и G, напротив, перепредставлены (см. рис. 4, в). Для С данное отклонение объяснялось в основном перепредставленностью в тринуклеотидах САА, тогда как для G наблюдалась повышенная частота встречаемости вплоть до длины тракта в 6 нуклеотидов (см. рис. 4, з). Снижение доли Т также имело место в трактах любой длины (см. рис. 4, з). После А-трактов встречаемость С и Т была ниже ожидаемой, а G – выше ожидаемой (см. рис. 4, в). Не исключено, что эти отклонения тоже могут быть объяснены повреждением ДНК и последующими ошибками ДНК-полимераз, однако в настоящий момент физико-химическая природа таких событий остается неясной.

Количество возникающего в геноме 8-охоG напрямую зависит от наличия во внутриклеточной среде активных форм кислорода (Halliwell, Gutteridge, 2015). Прокариоты отличаются исключительным многообразием вариантов энергетического метаболизма: некоторые ведут строго анаэробный образ жизни, в то время как другие относятся к облигатным аэробам или факультативным анаэробам и подвержены окислительному стрессу более высокой интенсивности. Было проведено сравнение статистики встречаемости 5'-фланговых нуклеотидов в геномах двух рассматриваемых групп (см. табл. 1). В геномах аэробных прокариот в этой позиции при G-трактах с повышенной частотой по сравнению с ожидаемой встречался нуклеотид Т и с пониженной – А (рис. 5). Для анаэробных микроорганизмов значимой разницы во встречаемости 5'-фланговых нуклеотидов не обнаружено (см. рис. 5). Однако при сравнении представленности А, С и Т непосредственно между группами аэробов и анаэробов различия не достигали статистической значимости, что, скорее всего, связано с недостаточным объемом выборки. При А-трактах разница во встречаемости 5'-фланговых нуклеотидов в геномах аэробов и анаэробов была такой же, как и в объединенной группе (ср. рис. 4, в и рис. 5). Таким образом, сниженный уровень окислительного стресса у анаэробных микроорганизмов может быть ассоциирован с менее выраженным преобладанием Т в положении, примыкающем с 5'-стороны к G-трактам. Для более определенного ответа на этот вопрос требуются дополнительные исследования.

В отличие от прокариот, геномы эукариот характеризуются большим количеством повторяющихся элементов – транспозонов, сателлитной и теломерной ДНК, точная последовательность которых плохо определяется тра-

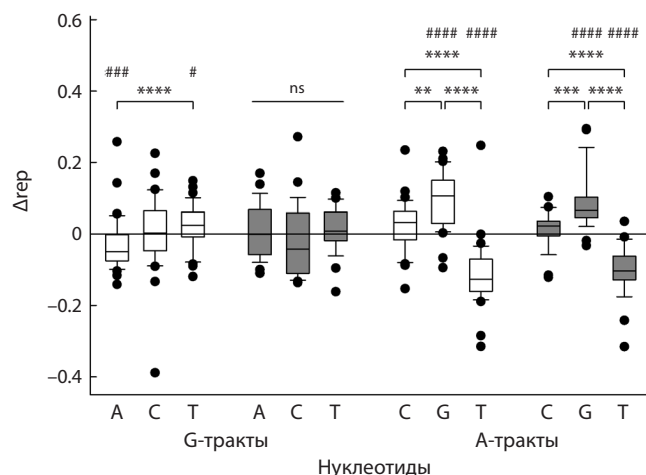


Рис. 5. Представленность 5'-фланговых нуклеотидов при полипуриновых трактах в геномах аэробных (белые прямоугольники) и анаэробных (серые прямоугольники) микроорганизмов.

Показано отклонение от ожидаемой представленности для G- и А-трактов любой длины. Отличие от ожидаемого: # $p < 0.05$, *** $p < 0.005$, **** $p < 0.001$ (одновыборочный критерий Уилкоксона с поправкой Данн на множественные сравнения). Различия между группами: ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$ (критерий Краскела–Уоллиса с поправкой Данн на множественные сравнения); ns – нет достоверных отличий.

диционными методами высокопроизводительного секвенирования (Richard et al., 2008; Liao et al., 2023). Появление методов сверхдлинного секвенирования (Oxford Nanopore, PacBio HiFi) позволило заполнить эти пробелы. Недавно опубликованный геном человека, прочитанный при помощи комбинации нескольких методов с покрытием «от теломеры до теломеры» (T2T) и высоким качеством (оценка частоты ошибок в теломерных областях $\sim 4 \times 10^{-8}$) (Nurk et al., 2022), предоставляет возможность анализа окружения G-трактов без искажений, вызванных более высокой представленностью уникальных последовательностей.

Значительно больший размер генома человека по сравнению с прокариотическими геномами позволил выявить интересные закономерности распределения размеров G_n-трактов. При $n = 2-8$ их число убывало экспоненциально и описывалось коэффициентом прирастания $k_{inc} = -0.674$, очень близким к центру распределения значений k_{inc} для G-трактов у прокариот (ср. рис. 6, а и рис. 3, в; $z = 0.141$). При $n = 9-16$ экспоненциальная зависимость сохранялась, но скорость уменьшения числа повторов замедлялась: значение k_{inc} возрастало до -0.198 , что лежит далеко за пределами диапазона значений k_{inc} для прокариотических геномов (ср. рис. 6, а и рис. 3, в; $z = 5.97$). Тракты такого размера в геномах прокариот отсутствовали или присутствовали в единичных случаях, поэтому выявить этот переход было невозможно. Дальнейший рост длины G-трактов сопровождался еще большим падением скорости уменьшения их числа (см. рис. 6, а). Очевидно, что в районе $n = 8-9$ (значение точки излома, определенное методом сегментированной регрессии: $n = 8.72 \pm 0.04$) происходит изменение баланса процессов укорочения и удлинения G-трактов в пользу последнего; рост тракта за счет проскальзывания ДНК-полимераз при репликации

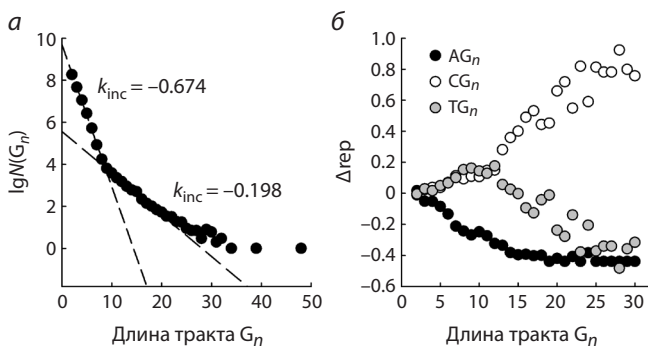


Рис. 6. Зависимость количества G-трактов и представленности 5'-флангового нуклеотида в геноме человека от длины тракта.

a – зависимость числа G-трактов $N(G_n)$ от их длины. Штриховые линии соответствуют линейной регрессии, значения k_{inc} для $n = 2-8$ и $n = 9-16$ приведены на фрагменте *b* – зависимость $\Delta гер$ 5'-флангового нуклеотида от длины тракта. Черные точки – фланговый нуклеотид А, белые точки – С, серые точки – Т.

или репарации становится «самоподдерживающимся», как в хорошо изученном случае трактов тринуклеотидных повторов (Mirkin, 2007; McMurray, 2010).

Еще более неожиданная картина выявилась при анализе представленности 5'-фланговых нуклеотидов. Поскольку хорошо известно, что количество динуклеотидов CG в геноме человека снижено из-за их роли в эпигенетической регуляции (Fazzari, Greally, 2004), ожидаемую представленность рассчитывали исходя не из общего содержания нуклеотидов, а из содержания динуклеотидов. При $n = 2$ представленность нуклеотидов близко соответствовала ожидаемой, но далее значения $\Delta гер$ для А стабильно падали, а представленность С и Т, напротив, росла практически с одной и той же скоростью (см. рис. 6, *b*). Однако начиная с $n = 8-11$ (значение точки излома для $\Delta гер(С) - \Delta гер(Т)$, определенное методом сегментированной регрессии: $n = 9.28 \pm 1.10$) зависимости для С и Т резко расходились: представленность Т падала, а представленность С росла. Одно из возможных объяснений такого явления может заключаться в том, что с ростом длины G-тракты служат более эффективными ловушками для мигрирующих по ДНК-дуплексу дырок, за счет чего 8-охоG в 5'-концевом положении доокисляется до гуанидиногидантоина и спироиминогидантоина с соответствующим переключением преимущественных нуклеотидных замен с $G \rightarrow T$ на $G \rightarrow C$.

Теломерную ДНК выделяют в отдельный класс высокоповторной части генома, представленный у человека множеством копий гексануклеотида TTAGGG. Известно, что теломерные повторы служат «горячими точками» окисления ДНК с образованием 8-охоG (Billard, Poncet, 2019; Opresko et al., 2025). Концы теломер в клетках зародышевой линии удлиняются за счет фермента теломеразы, который в качестве матрицы использует теломеразную РНК, поэтому изменения в этих областях никак не связаны с повреждением геномной ДНК. Однако большая часть теломер даже при наличии активной теломеразы реплицируется по обычному полуконсервативному механизму (Pfeiffer, Lingner, 2013; Higa et al., 2017; Bonnell et al., 2021), что может приводить к накоплению в них мутаций.

Таким образом, конкретная последовательность теломер в соматических клетках человека (в случае Т2Т-генома – immortalized клеточной линии хорионаденомного происхождения CHM13hTERT, экспрессирующей теломеразу) отражает как их недавнее удлинение теломеразой в клетках зародышевой линии, так и события мутагенеза в прошлых поколениях и в индивидуальном онтогенезе.

Распределение повторов TTAGGG в хромосомах, рассчитанное по обеим цепям ДНК, имело стандартный вид с пиками частот на концах хромосом и провалом в околоцентромерном участке (рис. 7, *a*). Единственным исключением была хромосома 8, для которой в районе центромеры наблюдался, напротив, небольшой подъем числа этих повторов. На хромосоме 2 был отчетливо виден пик частоты теломерных повторов в области слияния двух предковых хромосом гоминид, образовавших эволюционно молодую хромосому 2 человека (Ijdo et al., 1991; см. рис. 7, *a*). Но при более детальном анализе этого участка видно, что он уже достаточно сильно деградировал и повторов TTAGGG в нем осталось гораздо меньше, чем в истинных теломерах (см. рис. 7, *b*). Интересно, что аналогичные пики обнаруживались на хромосомах 15 и 22 в интронах активных белок-кодирующих генов *ATP10A* и *MICAL3*; возможно, они представляют собой остатки событий транслокации части теломеры.

Повторы TTAAGG, TTACGG и TTATGG были распределены по хромосомам без теломерных пиков. Общая частота повторов TTACGG при этом была заметно ниже, чем TTAAGG и TTATGG, что согласуется со сниженным количеством динуклеотидов CG в геноме человека (см. рис. 7, *a*). Отдельные пики частоты повторов наблюдались на хромосоме 2 для TTAAGG, хромосомах 8, 12, 17 и Y для TTACGG, хромосомах 4 и 22 для TTATGG (см. примеры на рис. 7, *a*). Характерная картина распределения повторов в околоцентромерной области с провалами во всех вариантах TTANGG имела место для хромосом 1–5, 7, 10–12, 16, 19 и 21. В других случаях в районе центромеры преобладал один тип повторов, а другие были обеднены, причем их суммарный недостаток компенсировал избыток преобладающего типа, как показано на рис. 7, *a* для хромосомы 6. В хромосомах 6, 13–15, 22 и X в качестве преобладающего повтора выступал TTATGG, в хромосоме 8 – TTAGGG, в хромосоме 17 – TTACGG. Хромосома 18 отличалась совпадающими пиками распределения двух повторов – TTACGG и TTATGG (см. рис. 7, *a*). В длинном плече хромосомы 9 в области конститутивного гетерохроматина, прилегающей к околоцентромерному участку с избытком TTATGG, находился длинный участок с преобладанием TTACGG.

Очевидно, что описанные выше случаи совместной или взаимно противофазной локализации повторов TTANGG в нетеломерных областях объясняются не точковыми мутациями в повторе TTAGGG, а наличием в этих локусах большого числа повторяющихся элементов, содержащих какой-то один или два из этих гексануклеотидов. Напротив, особенности точковых мутаций в первой позиции G₃-тракта теломерного повтора должны быть наиболее очевидны в областях, состоящих по большей части из TTAGGG, т. е. в самих теломерах и внутривнутрихромосомных блоках теломероподобных повторов. Для анализа частоты

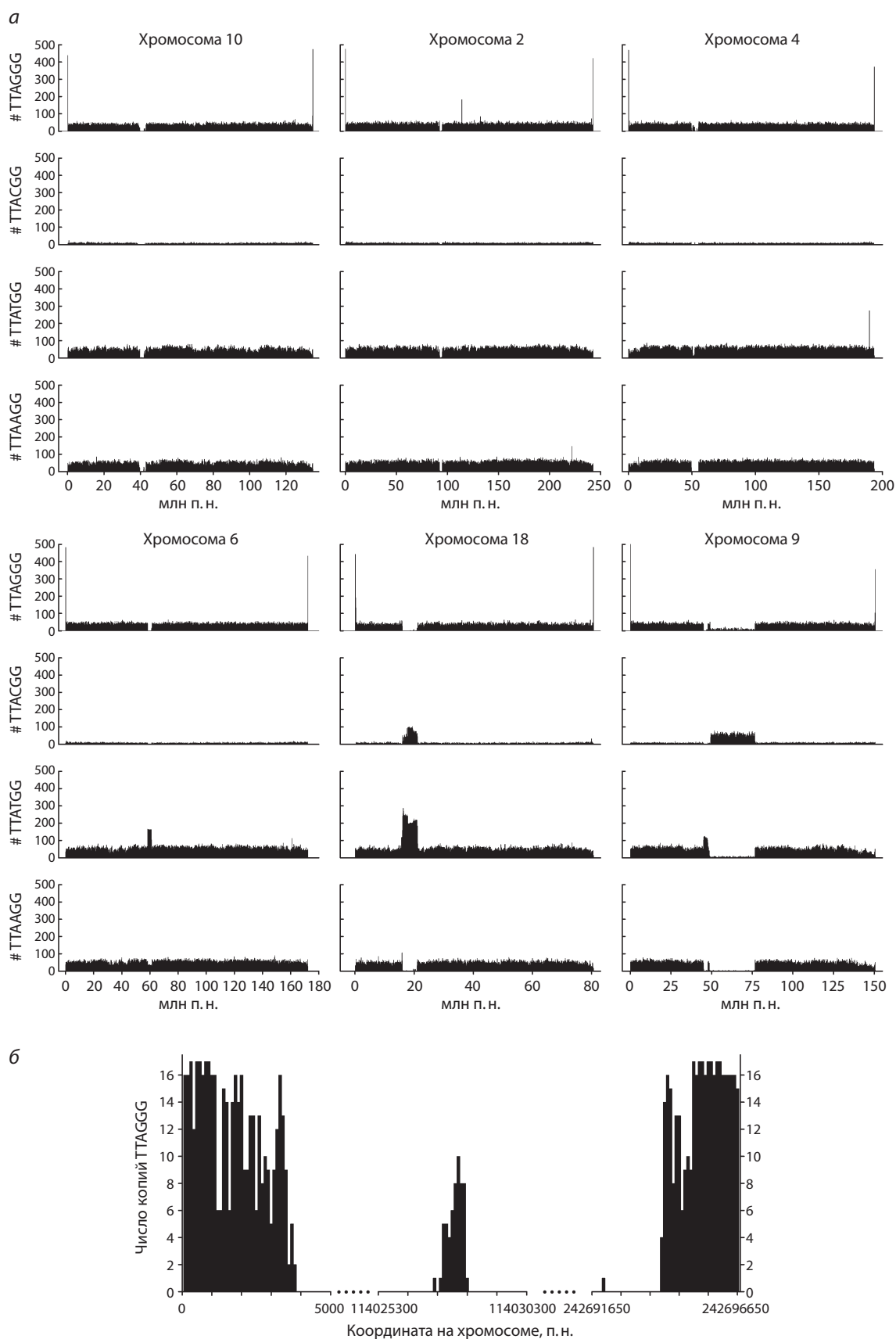


Рис. 7. Примеры распределения повторов TTAGGG, TTACGG, TTATGG и TTAAGG в хромосомах человека.

а – распределение повторов по всей длине хромосом 10, 2, 4, 6, 18 и 9. Количество повторов рассчитано в интервалах размером 100 т.п.н.; **б** – распределение повторов TTAGGG в теломерных участках и в области слияния предковых теломер хромосомы 2. Количество повторов рассчитано в интервалах размером 100 п.н.

Таблица 2. Представленность гексануклеотидов ТТАГГГ в теломерах и внутрихромосомных блоках теломероподобных повторов

Число копий ТТАГГГ в 100 нт	ТТАГГГ	ТТААГГ	ТТАСГГ	ТТАТГГ	$\chi^2_{(A=C=T)}$ *	$\chi^2_{(A=T)}$ **
Теломеры						
17	7361	0	0	0	–	–
16	6512	3	0	0	0.0498	0.0833
15	1935	9	1	1	0.00297	0.0114
14	980	26	0	3	8.12×10^{-10}	1.95×10^{-5}
13	715	23	1	2	1.85×10^{-8}	2.67×10^{-5}
12	468	14	2	2	3.35×10^{-4}	0.00270
11	341	6	0	0	0.00248	0.0143
10	490	16	1	0	6.97×10^{-7}	6.33×10^{-5}
9	327	8	0	0	3.35×10^{-4}	0.00468
Внутрихромосомные участки						
16–17	16	0	0	0	–	–
15	45	0	1	0	0.368	–
11–14	221	0	0	0	–	–
10	90	2	1	2	0.818	1.000
9	99	10	0	4	0.00439	0.109
Объединенная статистика						
17	7361	0	0	0	–	–
16	6528	3	0	0	0.0498	0.0833
15	1980	9	2	1	0.00865	0.0114
14	1022	26	0	3	8.12×10^{-10}	1.95×10^{-5}
13	767	23	1	2	1.85×10^{-8}	2.67×10^{-5}
12	540	14	2	2	3.35×10^{-4}	0.00270
11	396	6	0	0	0.00248	0.0143
10	580	18	2	2	8.84×10^{-6}	3.35×10^{-4}
9	423	18	0	4	5.12×10^{-6}	0.00284

* Значения критерия χ^2 для нулевой гипотезы равной представленности А, С и Т.
** Значения критерия χ^2 для нулевой гипотезы равной представленности А и Т.

замен в таких участках были выделены собственно теломерные области, а также все внутрихромосомные блоки, где в интервалах размером 100 т. п. н. встречалось не менее 100 копий повтора ТТАГГГ. Они были разбиты на более короткие интервалы длиной 100 п. н. Заполнение такого интервала только повторами ТТАГГГ соответствует 16 или 17 копиям (в зависимости от позиции первого полного гексануклеотида в интервале). Для анализа были взяты интервалы, в которых число копий ТТАГГГ было не менее 9, что составляет более половины от длины интервала. Подсчет встречаемости гексануклеотидов ТТААГГ, ТТАСГГ и ТТАТГГ в исследуемых участках выявил однозначную картину значительного обогащения по заменам G→А в первой позиции G-тракта по сравнению с заменами G→С и G→Т (табл. 2). Суммарное число последних было в пять раз ниже, а частоты достоверно не различались между собой. Таким образом, несмотря на то, что теломерные повторы служат преимущественными местами окисления гуанина, это не фиксируется в

повышенной частоте появления точковых мутаций G→Т. Разница между представленностью А и С+Т в 5'-фланговой позиции при динуклеotide GG между теломерными повторами и остальной частью генома может свидетельствовать о существовании в теломерах мутационного процесса, отличного от окисления G в 5'-положении GGG.

Заключение

Таким образом, анализ нуклеотидного контекста G-трактов в наборе из 62 полных геномов прокариот и в T2T-геноме человека показал, что представленность Т в положении, примыкающем к G-трактам, в целом повышена, что согласуется с моделью окисления G в 5'-положении трактов с последующими мутациями G→Т. Одновременно с этим выявлены и другие закономерности распределения 5'-фланговых нуклеотидов: неравномерность частоты нуклеотидов в положении, примыкающем к А-трактам, повышенная представленность С с 5'-стороны от длинных G-трактов в геноме человека, преобладание А при заменах

G в 5'-положении в теломерных повторах. Гипотеза о том, что удлинение G-тракта может вести к замене специфичности однонуклеотидных мутаций с G→T на G→C из-за изменения природы повреждения-предшественника может быть проверена экспериментально. Характерный спектр мутаций в теломерных повторах может быть вызван их тенденцией складываться в G-квадруплексные структуры, которые затрудняют движение ДНК-полимераз (Pfeiffer, Lingner, 2013; Higa et al., 2017; Bonnell et al., 2021), однако это предположение требует детального изучения точности действия репликативных ДНК-полимераз человека на неповрежденных и поврежденных матрицах такого строения. Для A-трактов неизвестно о существовании преимущественных положений повреждения ДНК; учитывая, что по сравнению с G-трактами A-тракты более длинные (см. рис. 2), разница в относительной представленности C, G и T в 5'-фланговом положении может быть не связана с мутационным процессом. Объяснение всех этих выявленных закономерностей требует дальнейших исследований.

Список литературы / References

- Adhikary A., Khanduri D., Sevilla M.D. Direct observation of the hole protonation state and hole localization site in DNA-oligomers. *J Am Chem Soc.* 2009;131(24):8614-8619. doi 10.1021/ja9014869
- Alexandrov L.B., Nik-Zainal S., Wedge D.C., Aparicio S.A.J.R., Behjati S., Biankin A.V., Bignell G.R., ... Campo E., Shibata T., Pfister S.M., Campbell P.J., Stratton M.R. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-421. doi 10.1038/nature12477
- Bansal A., Kaushik S., Kukreti S. Non-canonical DNA structures: diversity and disease association. *Front Genet.* 2022;13:959258. doi 10.3389/fgene.2022.959258
- Billard P., Poncet D.A. Replication stress at telomeric and mitochondrial DNA: common origins and consequences on ageing. *Int J Mol Sci.* 2019;20(19):4959. doi 10.3390/ijms20194959
- Bonnell E., Pasquier E., Wellinger R.J. Telomere replication: solving multiple end replication problems. *Front Cell Dev Biol.* 2021;9:668171. doi 10.3389/fcell.2021.668171
- Cadet J., Douki T., Ravanat J.-L. Oxidatively generated damage to the guanine moiety of DNA: mechanistic aspects and formation in cells. *Acc Chem Res.* 2008;41(8):1075-1083. doi 10.1021/ar700245e
- Cadet J., Davies K.J.A., Medeiros M.H.G., Di Mascio P., Wagner J.R. Formation and repair of oxidatively generated damage in cellular DNA. *Free Radic Biol Med.* 2017;107:13-34. doi 10.1016/j.freeradbiomed.2016.12.049
- Chatterjee N., Walker G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen.* 2017;58(5):235-263. doi 10.1002/em.22087
- Chiorcea-Paquim A.-M. 8-oxoguanine and 8-oxodeoxyguanosine biomarkers of oxidative DNA damage: a review on HPLC-ECD determination. *Molecules.* 2022;27(5):1620. doi 10.3390/molecules27051620
- Cho B.P., Kadlubar F.F., Culp S.J., Evans F.E. ¹⁵N nuclear magnetic resonance studies on the tautomerism of 8-hydroxy-2'-deoxyguanosine, 8-hydroxyguanosine, and other C8-substituted guanine nucleosides. *Chem Res Toxicol.* 1990;3(5):445-452. doi 10.1021/tx00017a010
- Dizdaroglu M., Coskun E., Jaruga P. Measurement of oxidatively induced DNA damage and its repair, by mass spectrometric techniques. *Free Radic Res.* 2015;49(5):525-548. doi 10.3109/10715762.2015.1014814
- ESCODD (European Standards Committee on Oxidative DNA Damage), Gedik C.M., Collins A. Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study. *FASEB J.* 2005;19(1):82-84. doi 10.1096/fj.04-1767fje
- Fazzari M.J., Greally J.M. Epigenomics: beyond CpG islands. *Nat Rev Genet.* 2004;5(6):446-455. doi 10.1038/nrg1349
- Fleming A.M., Burrows C.J. Formation and processing of DNA damage substrates for the hNEIL enzymes. *Free Radic Biol Med.* 2017;107:35-52. doi 10.1016/j.freeradbiomed.2016.11.030
- Fleming A.M., Burrows C.J. Chemistry of ROS-mediated oxidation to the guanine base in DNA and its biological consequences. *Int J Radiat Biol.* 2022;98(3):452-460. doi 10.1080/09553002.2021.2003464
- Genereux J.C., Barton J.K. Mechanisms for DNA charge transport. *Chem Rev.* 2010;110(3):1642-1662. doi 10.1021/cr900228f
- Giese B. Long-distance electron transfer through DNA. *Annu Rev Biochem.* 2002;71:51-70. doi 10.1146/annurev.biochem.71.083101.134037
- Halliwell B., Gutteridge J.M.C. Free Radicals in Biology and Medicine. Oxford Univ. Press, 2015
- Henikoff S., Henikoff J.G. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA.* 1992;89(22):10915-10919. doi 10.1073/pnas.89.22.10915
- Higa M., Fujita M., Yoshida K. DNA replication origins and fork progression at mammalian telomeres. *Genes.* 2017;8(4):112. doi 10.3390/genes8040112
- Ijdo J.W., Baldini A., Ward D.C., Reeders S.T., Wells R.A. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci USA.* 1991;88(20):9051-9055. doi 10.1073/pnas.88.20.9051
- Kino K., Kawada T., Hirao-Suzuki M., Morikawa M., Miyazawa H. Products of oxidative guanine damage form base pairs with guanine. *Int J Mol Sci.* 2020;21(20):7645. doi 10.3390/ijms21207645
- Koh G., Degasperis A., Zou X., Momen S., Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer.* 2021;21(10):619-637. doi 10.1038/s41568-021-00377-7
- Kouchakdjian M., Bodepudi V., Shibutani S., Eisenberg M., Johnson F., Grollman A.P., Patel D.J. NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn)-dA(anti) alignment at lesion site. *Biochemistry.* 1991;30(5):1403-1412. doi 10.1021/bi00219a034
- Kucab J.E., Zou X., Morganella S., Joel M., Nanda A.S., Nagy E., Gomez C., Degasperis A., Harris R., Jackson S.P., Arlt V.M., Phillips D.H., Nik-Zainal S. A compendium of mutational signatures of environmental agents. *Cell.* 2019;177(4):821-836.e816. doi 10.1016/j.cell.2019.03.001
- Kunkel T.A., Bebenek K. DNA replication fidelity. *Annu Rev Biochem.* 2000;69:497-529. doi 10.1146/annurev.biochem.69.1.497
- Kurbanyan K., Nguyen K.L., To P., Rivas E.V., Lueras A.M.K., Kosinski C., Steryo M., González A., Mah D.A., Stemp E.D.A. DNA-protein cross-linking via guanine oxidation: dependence upon protein and photosensitizer. *Biochemistry.* 2003;42(34):10269-10281. doi 10.1021/bi020713p
- Liao X., Zhu W., Zhou J., Li H., Xu X., Zhang B., Gao X. Repetitive DNA sequence detection and its role in the human genome. *Commun Biol.* 2023;6:954. doi 10.1038/s42003-023-05322-y
- Lipscomb L.A., Peek M.E., Morningstar M.L., Verghis S.M., Miller E.M., Rich A., Essigmann J.M., Williams L.D. X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine. *Proc Natl Acad Sci USA.* 1995;92(3):719-723. doi 10.1073/pnas.92.3.719
- Liu B., Xue Q., Tang Y., Cao J., Guengerich F.P., Zhang H. Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat Res.* 2016;768:53-67. doi 10.1016/j.mrrev.2016.03.006
- Livingstone C.D., Barton G.J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci.* 1993;9(6):745-756. doi 10.1093/bioinformatics/9.6.745
- Maga G., Villani G., Crespan E., Wimmer U., Ferrari E., Bertocci B., Hübscher U. 8-oxo-guanine bypass by human DNA polymerases in

- the presence of auxiliary proteins. *Nature*. 2007;447(7144):606-608. doi 10.1038/nature05843
- McAuley-Hecht K.E., Leonard G.A., Gibson N.J., Thomson J.B., Watson W.P., Hunter W.N., Brown T. Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs. *Biochemistry*. 1994;33(34):10266-10270. doi 10.1021/bi00200a006
- McMurray C.T. Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*. 2010;11(11):786-799. doi 10.1038/nrg2828
- Miller H., Grollman A.P. Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry*. 1997;36(49):15336-15342. doi 10.1021/bi971927n
- Mirkin S.M. Expandable DNA repeats and human disease. *Nature*. 2007;447(7147):932-940. doi 10.1038/nature05977
- Moriya M. Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G·C→T·A transversions in simian kidney cells. *Proc Natl Acad Sci USA*. 1993;90(3):1122-1126. doi 10.1073/pnas.90.3.1122
- Nurk S., Koren S., Rhie A., Rautiainen M., Bizikadze A.V., Mikheenko A., Vollger M.R., ... Zook J.M., Schatz M.C., Eichler E.E., Miga K.H., Phillippy A.M. The complete sequence of a human genome. *Science*. 2022;376(6588):44-53. doi 10.1126/science.abj6987
- Okonechnikov K., Golosova O., Fursov M.; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-1167. doi 10.1093/bioinformatics/bts091
- O'Leary N.A., Wright M.W., Brister J.R., Ciufu S., Haddad D., McVeigh R., Rajput B., ... Tatusova T., DiCuccio M., Kitts P., Murphy T.D., Pruitt K.D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-D745. doi 10.1093/nar/gkv1189
- Opreško P.L., Sanford S.L., De Rosa M. Oxidative stress and DNA damage at telomeres. *Cold Spring Harb Perspect Biol*. 2025;17(6):a041707. doi 10.1101/cshperspect.a041707
- Pfeiffer V., Lingner J. Replication of telomeres and the regulation of telomerase. *Cold Spring Harb Perspect Biol*. 2013;5(5):a010405. doi 10.1101/cshperspect.a010405
- Pilati C., Shinde J., Alexandrov L.B., Assié G., André T., Hélias-Rodzewicz Z., Ducoudray R., Le Corre D., Zucman-Rossi J., Emile J.-F., Bertherat J., Letouze E., Laurent-Puig P. Mutational signature analysis identifies *MUTYH* deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol*. 2017;242(1):10-15. doi 10.1002/path.4880
- Prorok P., Grin I.R., Matkarimov B.T., Ishchenko A.A., Laval J., Zharkov D.O., Saparbaev M. Evolutionary origins of DNA repair pathways: role of oxygen catastrophe in the emergence of DNA glycosylases. *Cells*. 2021;10(7):1591. doi 10.3390/cells10071591
- Richard G.-F., Kerrest A., Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*. 2008;72(4):686-727. doi 10.1128/MMBR.00011-08
- Saito I., Nakamura T., Nakatani K., Yoshioka Y., Yamaguchi K., Sugiyama H. Mapping of the hot spots for DNA damage by one-electron oxidation: efficacy of GG doublets and GGG triplets as a trap in long-range hole migration. *J Am Chem Soc*. 1998;120(48):12686-12687. doi 10.1021/ja981888i
- Shibutani S., Takeshita M., Grollman A.P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*. 1991;349(6308):431-434. doi 10.1038/349431a0
- Sugiyama H., Saito I. Theoretical studies of GG-specific photocleavage of DNA via electron transfer: significant lowering of ionization potential and 5'-localization of HOMO of stacked GG bases in B-form DNA. *J Am Chem Soc*. 1996;118(30):7063-7068. doi 10.1021/ja960982i
- Taylor W.R. The classification of amino acid conservation. *J Theor Biol*. 1986;119(2):205-218. doi 10.1016/S0022-5193(86)80075-3
- Tubbs A., Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell*. 2017;168(4):644-656. doi 10.1016/j.cell.2017.01.002
- Viel A., Bruselles A., Meccia E., Fornasari M., Quaia M., Canzonieri V., Policicchio E., ... Maestro R., Giannini G., Tartaglia M., Alexandrov L.B., Bignami M. A specific mutational signature associated with DNA 8-oxoguanine persistence in *MUTYH*-defective colorectal cancer. *EBioMedicine*. 2017;20:39-49. doi 10.1016/j.ebiom.2017.04.022
- Wood M.L., Esteve A., Morningstar M.L., Kuziemko G.M., Essigmann J.M. Genetic effects of oxidative DNA damage: comparative mutagenesis of 7,8-dihydro-8-oxoguanine and 7,8-dihydro-8-oxoadenine in *Escherichia coli*. *Nucleic Acids Res*. 1992;20(22):6023-6032. doi 10.1093/nar/20.22.6023
- Yudkina A.V., Shilkin E.S., Endutkin A.V., Makarova A.V., Zharkov D.O. Reading and misreading 8-oxoguanine, a paradigmatic ambiguous nucleobase. *Crystals*. 2019;9(5):269. doi 10.3390/cryst9050269

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 03.08.2025. После доработки 01.10.2025. Принята к публикации 17.10.2025.