

Введение В настоящее время множество тестов на мутагенность делается на бактериях, плодовых мушках, дафниях и, реже, на крысах и лабораторных мышах. Следовательно, встает вопрос о релевантности оценки мутагенного потенциала соединения для человека по данным экспериментов на низших организмах. Таким образом, вопрос о возможности экстраполяции результатов мутагенных тестов с организмов различного уровня развития, от низших до млекопитающих, неразрывно связан с преобразованиями ДНК этих организмов как носителя генетической информации в эволюционной ретроспективе [1, 2]. В настоящее время, несмотря на многочисленные исследования одним из сдерживающих факторов является ограниченное понимание механизмов эволюционных преобразований, а также их вклада и взаимодействия при формировании генетического аппарата высших организмов. Данная работа посвящена разработке статистической модели нарушения генетического кода за счет накопления нонсенс мутаций и мобильных генетических элементов. Результаты и обсуждение Негативное воздействие мутагенных факторов заключается в нуклеотидных заменах и вырезании отдельных нуклеотидов. Однако не любая мутация будет проявляться одинаковым образом в части влияния на функциональные процессы живой клетки. Максимальный негативный эффект будет наблюдаться в случае возникновения мутаций в последовательности ДНК, несущей информацию о структуре белка. Поэтому задача определения вероятности возникновения критичной мутации должна учитывать вероятность попадания в кодирующую область ДНК. Геномная ДНК любого организма включает гены (непосредственно несут информацию о структуру белков или РНК) и межгенные последовательности ДНК, включающие регуляторные участки, мобильные генетические элементы, и ДНК с неизвестной функцией. С эволюцией, размер межгенной ДНК возрастает на порядки. Если у кишечной палочки она составляет лишь 5% всего генома (3,5 млн нуклеотидов), то у эукариот она может составлять от 50% (у плодовой мушки) до 97% (у человека) [3-5]. Далее, у эукариот гены имеют мозаичную структуру и состоят из кодирующих и некодирующих области, называемых экзонами и интронами, соответственно. Оба типа последовательностей транскрибируются в пре-мРНК, но в то время как экзоны непосредственно составляют матричную мРНК и несут информацию о структуре белков, интроны вырезаются сразу после транскрипции в ходе сплайсинга [6]. Количество интронов неравномерно распределено между видами и возрастает с эволюцией [7,8]. Так, например, 287 интронов приходится на весь геном пекарских дрожжей (*Saccharomyces cerevisiae*) [9], 4760 у дрожжей *Schizosaccharomyces* и их количество достигает ~ 38 000 и ~ 140 000 в геномах плодовой мушки дрозофилы и человека, соответственно [10]. Вместе с их количеством, также и возрастает и их размер, который увеличивается в 10 раз. И, наконец, генетический код имеет избыточность кодирования. Так, замена третьего нуклеотида в триплете не

всегда приводит к замене аминокислоты. Наиболее чувствительные мутации, которые приводят к сдвигу рамки считывания и образования стоп-кодона. Трансляция обычно заканчивается, когда рибосома достигает нонсенс (или стоп) кодонов в конце кодирующей мРНК. Нонсенс-кодон, который расположен выше по течению от истинного может привести к преждевременной терминации трансляции и, таким образом, называется преждевременным терминирующим кодоном (ПТК) [11]. Таким образом, задача определения устойчивости ДНК к мутациям сводится к определению вероятности изменения триплетного генетического кода в кодирующей области ДНК и нарушения трехмерной структуры ее упаковки. Данная задача усложняется тем, что характер распределения структурных элементов в ДНК различно у индикаторных организмов, находящихся на более низких ступенях эволюционного развития по сравнению с человеком. Вторым ограничением является неизвестный характер и причины накопления некодирующих последовательностей ДНК в геноме с эволюцией [3–5]. В условиях ограниченной априорной информации о физических механизмах в теории сложных систем часто прибегают к подходу, основанному на построении математической модели процесса, отражающей его основные статистические свойства. Построение подобной математической модели, воспроизводящее ретроспективные эволюционные преобразования ДНК от низших до высших организмов на уровне основных статистических характеристик, влияющих на ее устойчивость к мутагенным воздействиям, позволило бы выполнять требуемые оценки устойчивости для высших организмов путем прямого математического моделирования с использованием эмпирических данных, полученных для низших организмов. Как уже было отмечено выше, существенными факторами, влияющими на устойчивость к мутагенным воздействиям, являются, во-первых, плотность и распределение генетической информации в геноме, и во-вторых, целостность пространственной структуры ДНК. Таким образом, эффективная модель (и соответственно алгоритм экстраполяции на ее основе) должны адекватно воспроизводить изменения как в объеме и положении кодирующей ДНК, так и участков (преимущественно некодирующей) ДНК, отвечающих за ее структурную устойчивость. Результаты многочисленных исследований указывают на то, что механизмы эволюционных преобразований ДНК носят вероятностный характер и могут быть лишь приближенно описаны с помощью алгоритмических схем. Одна из таких схем представлена в обзорной статье [11] специалистов из США, где предпринята попытка совместного схематического описания механизмов экзонизации и интронизации, а также роста некодирующей ДНК, вызванной вставками различных мобильных генетических элементов (транспозонов, вирусов и др.), копированием и переносом фрагментов некодирующей ДНК. По данным литературных исследований было установлено, что в предшествующих работах механизмы экзонизации и интронизации изучались в основном отдельно от

механизмов действия мобильных генетических элементов, что объясняет отсутствие разработанных комплексных математических моделей и алгоритмов на их основе, которые бы одновременно учитывали оба указанных механизма. Для восполнения данного пробела на основе схематической модели [11] нами был разработан алгоритм, моделирующий изменения исходной (кодирующей) ДНК в процессе интронизации и (возможной) последующей экзонизации, который был в дальнейшем дополнен алгоритмом, отражающим механизмы роста некодирующей ДНК. Последний алгоритм был построен с использованием математической модели, недавно предложенной совместно греческими и американскими учеными [12]. В качестве исходных данных для демонстрации работы алгоритма нами использовались последовательности ДНК архей, как предполагаемых эволюционных предшественников большинства высших организмов. Поскольку у низших организмов, включая архей, в последовательности исходной ДНК не более 5% является некодирующей, ее наличием пренебрегали. Также мы исключали из анализа оконечности хромосом (регионы 3' и 5'), для которых характерны несколько отличные от остальной части ДНК статистические распределения нуклеотидов. Выходом алгоритма являлось распределение кодирующей и некодирующей ДНК в виртуально эволюционировавшем (в соответствии со схематической моделью) геноме. Рис. 1 – Схема алгоритма модификации ДНК. Предложенная модель, схема которой представлена на рисунке 1, основывается на одном из аспектов возникновения негативных (летальных мутаций) путем образования нонсенс кодонов внутри кодирующей ДНК и возникновения интронов (некодирующей ДНК), или благодаря встраиванию мобильных генетических элементов. Результатом являются модельные распределения длин образовавшихся экзонов (рис. 2) и интронов (рис. 3) в результате от 103 до 106 итераций. При этом следует сделать оговорку, что в приведенной модели каждая итерация моделирует одно поколение клетки и в каждой генерации вероятность возникновения мутации равна 1. С учетом среднего числа мутаций в естественных условиях, равным 10^{-5} , происходит моделирование эволюции в течение $10^8 - 10^{11}$ поколений. Рис. 2 – Плотность распределения длин модельных экзонов. Рис. 3 – Плотность распределения длин модельных интронов. Полученное распределение кодирующей и некодирующей ДНК удобно задать в виде плотностей вероятностей длин экзонов и интронов, которые приведены на рис. 3. Из рис. 2 и 3 видно, что плотности вероятности длин экзонов и интронов при значениях длин существенно выше средних значений убывают по степенному закону, и могут быть асимптотически аппроксимированы обобщенным распределением Парето, при этом для длин экзонов скорость затухания плотности вероятности характеризуется показателем $(l/L)^{-3}$. Незначительное расширение распределений происходит с увеличением объема моделирования от 103 до 106 итераций алгоритма, что отражено на рисунке. Хорошее качество

аппроксимации для плотности вероятности длин интронов может быть достигнуто с использованием q -экспоненциального распределения, которое является частным случаем обобщенного распределения Парето и характеризуется асимптотически степенным убыванием плотности вероятности с показателем степени и перегибом к приближенно экспоненциальному поведению для средних и малых значений аргумента, причем крутизна перегиба определяется параметром. Следует отметить, что аналогичный вид функциональной зависимости отмечался ранее при анализе структуры ДНК различных организмов [12–15]. Аналогичные результаты были получены при различных начальных последовательностях ДНК (архей, червя, дрозофилы, рыбы, ящерицы, мыши, крысы и человека), что позволяет предположить, что механизм нарушений структуры ДНК за счет интронизации структуры ДНК имеет универсальную природу. Другие механизмы мутагенеза ДНК, а именно последствия ультрафиолетового облучения, внедрения мобильных генетических элементов, ошибок ДНК полимераз можно считать одинаково вероятным на единицу длины ДНК для всех живых организмов [3–5]. Таким образом мы предполагаем, что процесс интронизации можно использовать в качестве модели повреждений ДНК.