

МЕТОД РЕСТРИКЦИОННОГО АНАЛИЗА ГЕНОМОВ МЛЕКОПИТАЮЩИХ IN SILICO

М.А. АБДУРАШИТОВ, В.Н. ТОМИЛОВ*, В.А. ЧЕРНУХИН, Д.А. ГОНЧАР,
С.Х. ДЕГТЯРЕВ

НПО «СибЭнзим», Новосибирск

Предложен метод получения теоретических картин расщепления хромосомной ДНК млекопитающих эндонуклеазами рестрикции. Основываясь на опубликованных в последнее время данных по первичной структуре геномов, проведен расчет и построены диаграммы распределения фрагментов хромосомной ДНК мыши, крысы и человека, полученных при расщеплении по последовательностям 5'-GGCC-3', 5'-CCGG-3', 5'-GATC-3' и 5'-CC(A/T)GG-3'. Проведены эксперименты по гидролизу хромосомной ДНК эндонуклеазами рестрикции HaeIII, MspI, Kzo9I и Bst2UI, имеющими соответствующие сайты узнавания, и показано соответствие между рассчитанными диаграммами и экспериментально полученными картинками рестрикции ДНК.

Ключевые слова: метод, ДНК-повторы, рестрикционный анализ, эукариотический геном, человек, крыса, мышь.

Рестрикционный анализ ДНК широко используется в молекулярно-биологических исследованиях и прикладных работах и является одним из наиболее важных инструментов при изучении ДНК. Как правило, продукты расщепления ДНК анализируются с помощью гель-электрофореза в агарозном или акриламидном геле, а полученная таким образом картина разделения фрагментов ДНК в виде определенного, отличающегося для разных ферментов, набора полос и является результатом рестрикционного анализа той или иной ДНК [1].

Большое разнообразие существующих эндонуклеаз рестрикции (ЭР) позволяет проводить расщепление ДНК по более чем 150 сайтам узнавания [2]. Рестрикционный анализ проводится для самых различных ДНК, начиная от небольших фрагментов длиной несколько десятков нуклеотидных пар и вплоть до целых геномов эукариот размерами более 1 млрд. пар оснований.

Одним из существенных элементов рестрикционного анализа является получение теоретически рассчитанной картины разделения фрагментов ДНК, определяемой на основе известной первичной структуры

данной ДНК. В случае небольших размеров ДНК (до 100000–200000 п.о.) для этих целей обычно используют программное обеспечение Vector NTI, Lasergene, Dnasis Max и т.д., установленное на обычный персональный компьютер класса Pentium или Athlon. Однако для работы с протяженными эукариотическими ДНК, первичная структура которых была определена за последнее десятилетие, требуются наличие вычислительных центров и специальное программное обеспечение.

В данном исследовании предложен простой метод работы с данными по первичной структуре геномов млекопитающих, позволяющий строить диаграммы распределения фрагментов хромосомной ДНК после ее расщепления по определенным последовательностям и использующий обычный компьютер класса Athlon 64 и несложное программное обеспечение.

Целью данной работы явилось:

- 1) разработка несложного программного обеспечения для анализа протяженных последовательностей ДНК in silico;
- 2) построение теоретических диаграмм распределения фрагментов хромосомной ДНК после ее расщепления по сайтам узнавания ряда рестриктаз на примере геномов крысы, мыши и человека;
- 3) получение экспериментальных данных по гидролизу хромосомных ДНК соответствующими эндонуклеазами рестрикции и сравнение полученных результатов с теоретическими расчетами.

* Автор для переписки:

© 2006 г. Томилов Виктор Николаевич,
сотрудник НПО «СибЭнзим»,
630117 Новосибирск, ул. Ак. Тимакова, 2/12,
Тел/факс: 8-383-333-6853,
E-mail: vicont@sibenzyme.ru

Материалы и методы

Программное обеспечение. Для создания программного обеспечения были использованы: программная среда Borland Delphi (интерфейс, вывод, обработка) и Microsoft C++ compiler (модульная часть расчетов, распараллеливание). Программный продукт был оптимизирован для работы на системах с двумя и более процессорами (расчеты производились на двухъядерном процессоре AMD Athlon64).

Образцы ДНК. В экспериментах использовали крыс-самцов линии SD в возрасте 3–4 мес. и мышей-самцов линии А/Не в возрасте 5–6 мес. разведения вивария ИЦиГ СО РАН. Геномная ДНК из печени животных выделялась по методу [3]. Выделение геномной ДНК из лейкоцитов крови человека (женщин-доноров) проводили согласно методике [4]. Выборка полученных данных делалась не менее чем из трех экспериментов.

Перед проведением реакции гидролиза все препараты ДНК обрабатывали рибонуклеазой А (0,1 мг/мл) 10 мин. при комнатной температуре и диализовали в трубках DispoDialyzer MWCO 50,000 («Sigma», США) против 100 объемов буфера TE (10 мМ Трис-НСl, рН 8,0, 1 мМ ЭДТА) при 4 °С в течение 20 ч.

Гидролиз ДНК ферментами. В работе использовали ферменты производства НПО «СибЭнзим». Расщепление ДНК проводили в рекомендуемых производителем буферах для рестрикции ДНК при оптимальных температурах в течение 3 ч.

Гидролиз проводили в 40 мкл реакционной смеси, в которую добавляли 6 мкг ДНК и 3 мкл препарата фермента. Активности использованных препаратов ферментов: HaeIII — 20 ед/мкл, Kzo9I — 5 ед/мкл, MspI — 10 ед/мкл, Bst2UI — 10 ед/мкл.

Электрофорез. Для выявления фрагментов длиной от 40 до 500 п.о. использовали электрофорез в 8% полиакриламидном геле (на дорожку наносили 6 мкг гидролизованной ДНК); для разделения фрагментов в диапазоне 200–2000 п.о. (рис. 1а и 1б) использовали 1,5% легкоплавкую агарозу «Low melting point» («Sigma», USA). Для выявления более крупных продуктов гидролиза использовался электрофорез в 1% агарозе «Type I-A, Low EEO» («Sigma», USA); на агарозный гель наносили 3 мкг гидролизованной ДНК на дорожку. Во всех случаях для электрофореза применялся трис-ацетатный буфер. После проведения электрофореза ДНК визуализировали бромистым этидием и фотографировали в УФ-свете.

Результаты и обсуждение

Разработка метода компьютерного анализа и обработка данных. Благодаря улучшению методов определения первичной структуры ДНК за последнее десятилетие была установлена нуклеотидная последовательность всех хромосом целого ряда млекопитающих, включая мышь [5] и крысу [6]. Совсем недавно было в целом завершено определение нуклеотидной последовательности генома человека [7]. Эти данные по первичной структуре ДНК представлены на сайтах Entrez Genome (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>), EMBL Genomes Pages (<http://www.ebi.ac.uk/genomes/eukaryota.html>) и Ensemble Genome Browser (<http://www.ensembl.org>). На сегодняшний день структуры ДНК крысы, мыши и человека определена более чем на 95% и постоянно появляются новые данные, повышающие долю достоверно установленной первичной структуры. В настоящей работе мы использовали данные по последовательностям геномной ДНК крысы, мыши и человека с ресурса <ftp://ftp.ensembl.org/pub/> (версии от 2 июня 2006 года).

При разработке программы для поиска сайтов узнавания эндонуклеаз рестрикции на сверхпротяженных последовательностях ДНК и расчета длин фрагментов, получаемых при расщеплении ДНК по этим сайтам, учитывались следующие особенности:

1. Некоторые эндонуклеазы рестрикции узнают две или более последовательностей ДНК (непалиндромные и/или вырожденные сайты узнавания). В этих случаях поиск проводился для всех возможных сайтов узнавания данных эндонуклеаз рестрикции, включающих лишь основные символы, соответствующие обозначениям нуклеотидных остатков ('A', 'G', 'T', 'C').

2. Существующие в базах данных геномные последовательности представляют собой набор протяженных фрагментов хромосомных ДНК (так называемых «контигов»), длина которых в большинстве случаев не позволяет загружать их в память обычного персонального компьютера целиком. Поэтому последовательности подгружались в оперативную память последовательно, частями, в виде строковых переменных, в которых и проводился поиск сайтов эндонуклеаз рестрикции.

3. Нерасшифрованные до сих пор последовательности геномов, расположенные с обоих краев «контигов» (включая вариабельные теломерные участки хромосом), не позволяют провести определение всех фрагментов ДНК, получаемых при расщеплении по тем или иным сайтам. Для простоты и ускорения анализа все имею-

щиеся последовательности частей геномов объединялись в одну сверхпротяженную последовательность. В этом случае возможно появление дополнительных, несуществующих реально фрагментов, однако их незначительное количество позволяет предположить допустимость использования такого подхода.

Алгоритм компьютерного моделирования включал в себя следующие стадии:

1. *Анализ сайта узнавания и выбор оптимального метода поиска (в зависимости от вырожденности сайта, то есть количества вырожденных пар оснований и степени их вырожденности).* По результатам анализа сайта узнавания использовались следующие способы поиска: метод прямого поиска, метод подстановки или метод подстановки с обобщенной группировкой.

2. *Ввод данных и поиск сайтов узнавания.* Ввиду слишком большого размера всего генома данные из файла последовательности ДНК генома загружались в память последовательно, без служебных символов. Основной алгоритм: 1) загружал в буфер определенный сегмент данных, 2) проводил оптимизированный поиск в нем с учетом ограниченности пространства элементов ($\{'A', 'G', 'T', 'C'\}$ — прямой поиск и $\{'A', 'G', 'T', 'C', '*'\}$ — подстановка); 3) в случае успешного поиска (нахождение сайта узнавания в загруженном сегменте) проводил добавление результата и удаление обработанного фрагмента из буфера, или в случае неудачи (сайт не найден) — загружал новый сегмент данных из файла генома с учетом граничного эффекта, то есть участок старого фрагмента длиной $n - 1$ (где n — длина анализируемой последовательности сайта узнавания) добавлялся к новому сегменту.

3. *Расчет длин фрагментов между ближайшими сайтами узнавания, подсчет числа фрагментов для каждой длины в диапазоне 1–6000 п.о. и вычисление общего количества всех п.о. во всех фрагментах заданной длины.* Эти данные выводились в файл в формате CSV (comma separated values) для дальнейшей обработки.

Полученные данные импортировались в таблицы Microsoft Excel и использовались для построения графиков в виде диаграмм.

Таким образом, в процессе моделирования для каждого сайта узнавания рассчитывалось распределение количества фрагментов, получаемых при расщеплении хромосомной ДНК по этим сайтам, от размеров этих фрагментов. При построении графиков значение, откладываемое по оси ординат, определялось как число

пар оснований во всех фрагментах данной длины, и рассчитывалась по формуле

$$S_i = i n_i,$$

где S_i — суммарное число п.о., i — количество п.о. во фрагменте, n_i — количество фрагментов, содержащих i нуклеотидных пар. Поэтому функция S может рассматриваться как зависимость суммарной массы всех фрагментов длиной i п.о. от длины фрагментов i , выраженной в п.о., то есть как диаграмма распределения. При сравнении с экспериментальными данными значение этой функции пропорционально количеству молекул бромистого этидия, связавшихся с ДНК, и, таким образом, величина S коррелирует с интенсивностью полос на электрофоретическом геле.

На рисунках 1–4 представлены полученные диаграммы распределения значения S_i для фрагментов ДНК крысы (1а, 2а, 3а и 4а), мыши (1б, 2б, 3б и 4б) и человека (1с, 2с, 3с и 4с) при расщеплении по последовательностям 5'-GGCC-3' (рис. 1), 5'-CC(A/T)GG-3' (рис. 2), 5'-GATC-3' (рис. 3) и 5'-CCGG-3' (рис. 4). На диаграммах показаны расчетные данные для диапазона 1–1200 п.о., на врезках — расчетные данные для диапазона 1–6000 п.о. (в уменьшенном масштабе). По оси абсцисс отложены длины фрагментов (п.о.), по оси ординат — суммарное количество п.о. во всех фрагментах данной длины.

Как видно из рисунков, все полученные диаграммы распределения имеют отчетливые экстремумы для определенных длин фрагментов ДНК, размеры которых (в парах оснований) указаны числами над пиками. Размеры фрагментов, имеющих пиковые значения, варьируют от 40–50 п.о., получаемых при расщеплении ДНК человека по последовательностям 5'-CC(A/T)GG-3' и 5'-GGCC-3', и до 5538 п.о. — при расщеплении ДНК мыши по последовательности 5'-CCGG-3'. Величины получаемых пиков также сильно варьируют — от незначительных значений до максимальной высоты около 15000000 п.о., получаемой при расщеплении хромосомной ДНК человека по последовательностям 5'-GATC-3' (фрагменты длиной 174 и 175 п.о.) и 5'-CC(A/T)GG-3' (фрагменты длиной 49 и 50 п.о.). Наличие таких и более мелких пиков на диаграммах, вероятно, связано с существованием так называемых повторов в эукариотических ДНК [8]. Такие повторы имеют размеры от нескольких п.о. (микросателлиты) до 6–7 тыс. п.о. (повторы семейства LINE), и их количество в геноме может достигать 500 и более тысяч [5–7]. Расщепление большого количества таких повторов по коротким последовательностям ДНК будет приводить к появлению

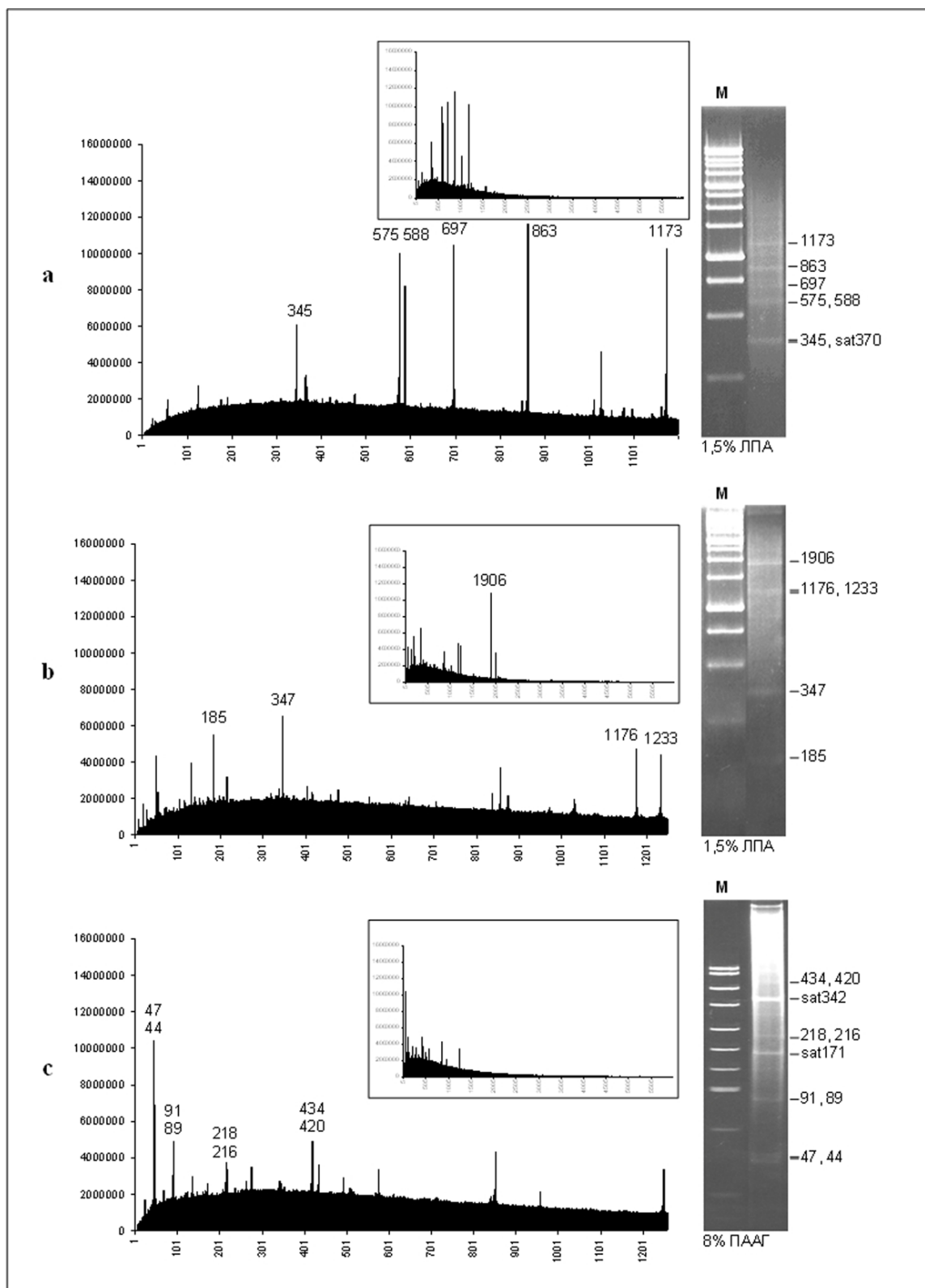


Рис. 1. Диаграмма распределения суммарного количества длин всех фрагментов (выраженное в парах оснований) в зависимости от их размера для расщепления ДНК крысы (а), мыши (б) и человека (с) по последовательности 5'-GGCC-3'. С правой стороны приведены экспериментально полученные электрофореграммы картин расщепления соответствующих ДНК эндонуклеазой рестрикции *NotI* (сайт узнавания 5'-GGCC-3'), снизу указан тип геля и его концентрация. М — маркеры длин фрагментов: для ЛПА использовался «SE 1 kb» (длины фрагментов — 10000, 8000, 6000, 5000, 4000, 2x3000, 2500, 2000, 1500, 2x1000, 750, 500 и 250 п.о.); для ПААГ использовался гидролизат ДНК *pUC19/MspI* (длины фрагментов — 501, 489, 404, 331, 242, 190, 147, 111, 110, 67, 2x34 п.о.)

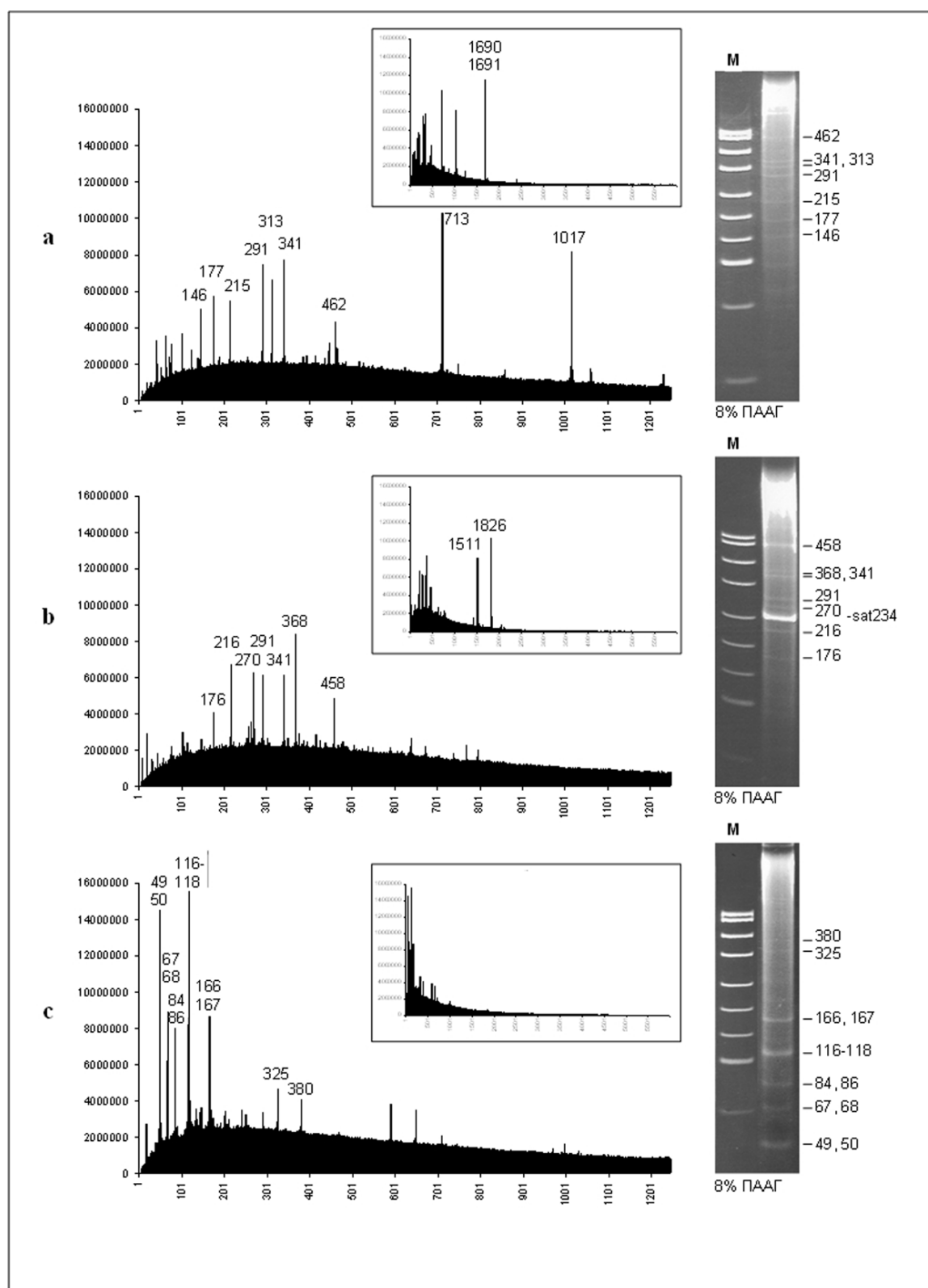


Рис. 2. Диаграмма распределения суммарного количества длин всех фрагментов (выраженное в парах оснований) в зависимости от их размера для расщепления ДНК крысы (а), мыши (б) и человека (с) по последовательности 5'-CC(A/T)GG-3'. С правой стороны приведены экспериментально полученные электрофореграммы картин расщепления соответствующих ДНК эндонуклеазой рестрикции Bst2UI (сайт узнавания 5'-CC(A/T)GG-3'). Остальные обозначения, как на рис. 1

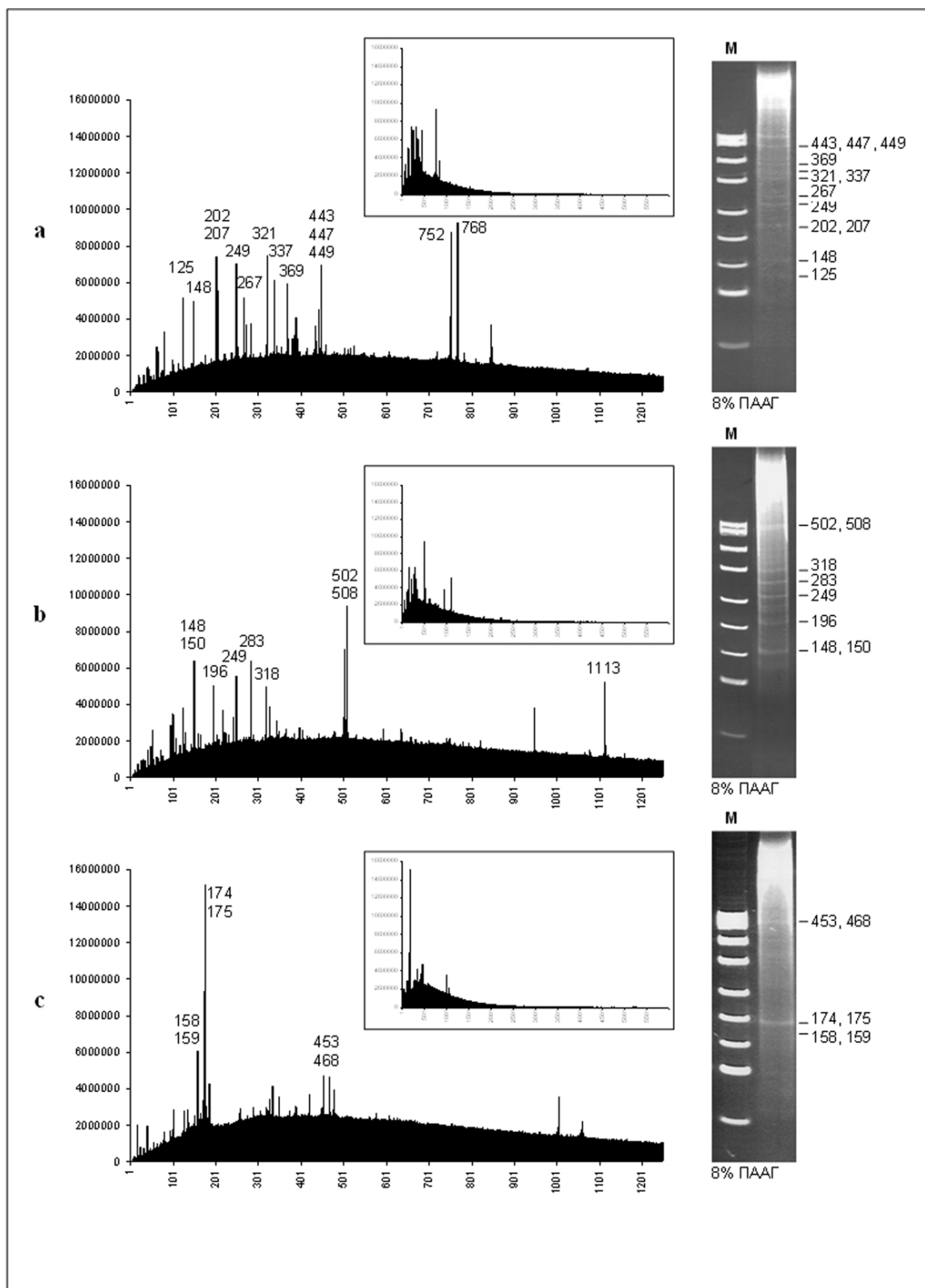


Рис. 3. Диаграмма распределения суммарного количества длин всех фрагментов (выраженное в парах оснований) в зависимости от их размера для расщепления ДНК крысы (а), мыши (б) и человека (с) по последовательности 5'-GATC-3'. С правой стороны приведены экспериментально полученные электрофореграммы картин расщепления соответствующих ДНК эндонуклеазой рестрикции Kzo9I (сайт узнавания 5'-GATC-3'). Остальные обозначения, как на рис. 1

множества фрагментов одинаковой длины. В результате при построении диаграммы мы получаем пики с высотой, пропорциональной количеству повторов того или иного фрагмента, присутствующих в хромосомной ДНК.

Получение диаграмм распределения фрагментов геномной ДНК, по сути, являющееся проведением рестрикционного анализа *in silico*, стало возможным благодаря тому, что нуклеотидные последовательности геномов эукариотических организмов, в частности, геномов крысы, мыши и человека, расшифрованы в значительной степени. Наличие брешей в уже определенной последовательности ДНК, а также возможные ошибки при установлении первичной структуры, не являются критическим моментом в работе, так как эти дефекты не превышают нескольких процентов от установленной структуры ДНК и не могут существенно изменить картину расщепления ДНК из-за большого количества существующих в ней повторов.

Получение экспериментальных картин расщепления хромосомных ДНК. Для изучения применимости предлагаемого метода расчета картин рестрикции нами были получены экспериментальные данные по расщеплению ДНК крысы, мыши и человека эндонуклеазами рестрикции с соответствующими сайтами узнавания.

На рисунке 5 представлены результаты электрофореза в 1% агарозном геле продуктов расщепления ДНК крысы, мыши и человека эндонуклеазами рестрикции Kzo9I (сайт узнавания GATC), HaeIII (GGCC), Bst2UI (CC(A/T)GG) и MspI (CCGG).

Электрофорез в 1% агарозном геле позволяет визуализировать большие фрагменты ДНК, которые видны на рисунке. Для представления картины гидролиза ДНК с фрагментами в диапазоне 200–2000 п.о. (диаграммы распределения на рис. 1а и 1б) мы использовали 1,5% легкоплавкую агарозу (ЛПА). Для получения остальных электрофореграмм и визуализации фрагментов меньших длин применялся электрофорез в 8% полиакриламидном геле (ПААГ). Полученные картины рестрикции ДНК приведены с правой стороны на рисунках 1–3. Как видно из приведенных данных на рисунках 1–3 и 5, в большинстве случаев при расщеплении хромосомных ДНК эндонуклеазами рестрикции образуются отчетливо видимые фрагменты определенной длины. Однако при этом на рисунках 1с и 2б на фоне типичной картины рестрикции ДНК существенно более ярко видны отдельные полосы. По размерам эти фрагменты соответствуют длинам основных повторов в сателлитных ДНК (234 п.о. — мыши [8], 342 п.о. и 171 п.о. — человека [9]). На электрофореграммах, приведенных на рисунках 1 и 2 эти полосы обозначены как «sat».

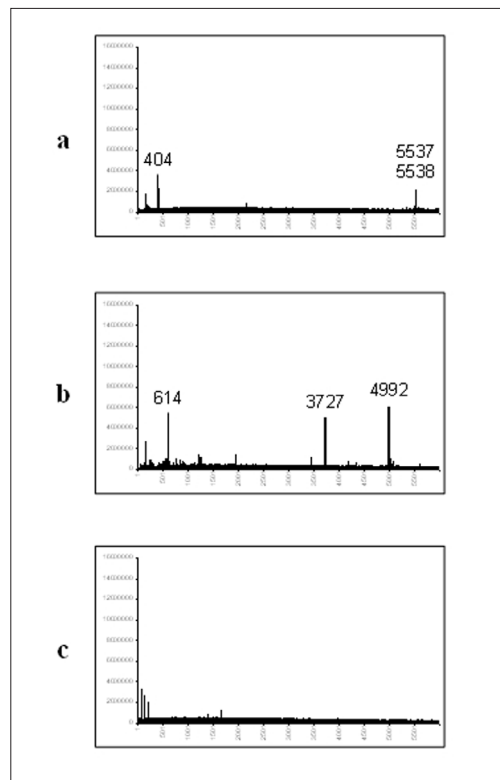


Рис. 4. Диаграмма распределения суммарного количества длин всех фрагментов (выраженное в парах оснований) в зависимости от их размера для расщепления ДНК крысы (а), мыши (б) и человека (с) по последовательности 5'-CCGG-3'

По-видимому, более высокая интенсивность этих полос связана с тем, что многократно повторенные tandemные последовательности, представляющие собой сателлитную ДНК, присутствуют в выделенных препаратах ДНК в количествах, значительно превышающих число повторов в протяженной хромосомной ДНК. При наличии в сателлитной ДНК сайта узнавания действие эндонуклеазы рестрикции приводит к выщеплению фрагментов, соответствующих по длине единичному повтору, а также кратных длине единичного повтора.

Известно, что очистка протяженной хромосомной ДНК больших размеров, в частности, ДНК эукариот, является сложной задачей из-за механического разрушения нитей молекул ДНК при использовании практически всех методик выделения. Тем не менее приведенные на рисунках 1–3 и 5 картины рестрикции показывают, что обычный фенол-хлороформный метод, с помощью которого были выделены все ДНК, вполне может быть использован для получения препаратов ДНК, пригодных для рестрикционного анализа *in vitro*. При этом частичная деградация ДНК, обычно приводящая к размыванию картины рестрикции, в данном случае компенсируется большим числом доминирующих повторов в геномах эукариот.

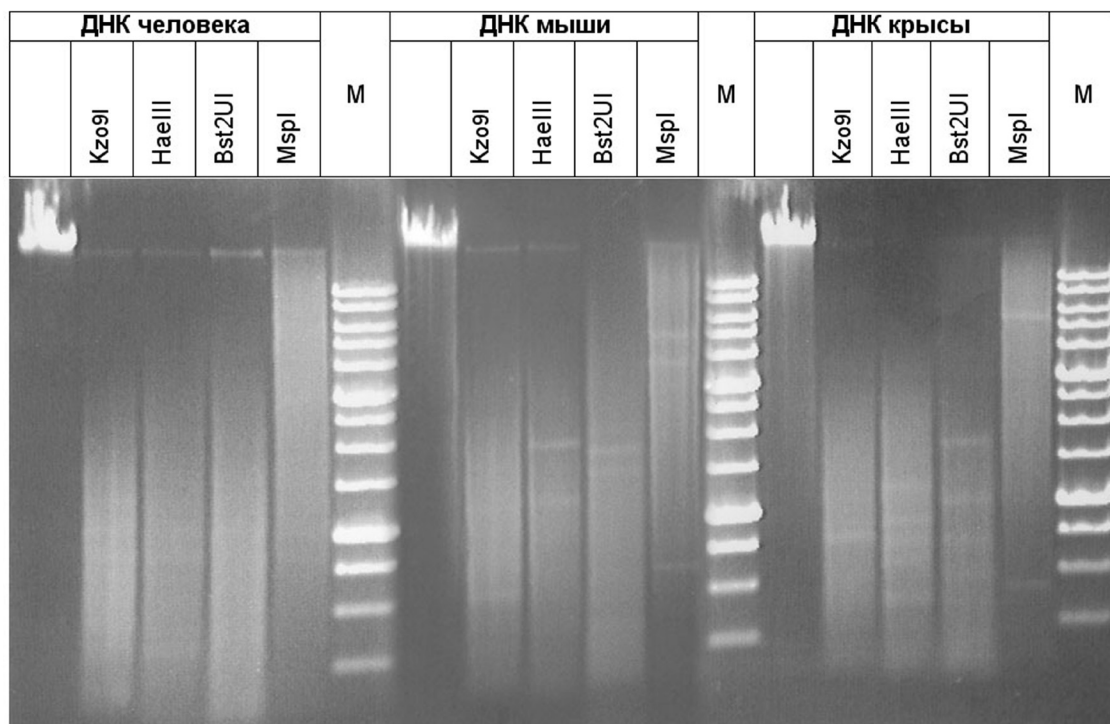


Рис. 5. Картины расщепления ДНК человека, мыши и крысы эндонуклеазами рестрикции Kzo9I (сайт узнавания GATC), HaeIII (GGCC), Bst2UI (CCWGG) и MspI (CCGG). Электрофорез в 1% агарозном геле. Дорожки М — маркер длин фрагментов «SE 1 kВ»

Сравнение теоретически рассчитанных и экспериментальных картин рестрикционного анализа. На рисунке 1 представлены расчетные данные по расщеплению хромосомных ДНК по сайту GGCC. Как видно из рисунков 1 и 5, наблюдается хорошая корреляция между теоретическими и экспериментальными данными. Фрагменты 345, 575–758, 697, 863 и 1173, получаемые при расщеплении хромосомной ДНК крысы, соответствуют фрагментам, видимым на геле, на рисунке 1а. Следует, однако, отметить, что хорошо видная полоса на геле, соответствующая фрагменту 345 п.о., видимо, перекрывается с фрагментом сателлитной ДНК крысы размером 370 п.о. [10].

Диаграмма расщепления мышинной ДНК содержит 5 фрагментов 185, 347, 1176–1233 и 1906 п.о., из которых только последний фрагмент имеет высокое значение Si и хорошо виден на геле. На фотографии геля также отчетливо видны фрагмент 347 п.о. и дубль фрагментов 1176 и 1223 п.о., тогда как фрагмент 185 п.о. едва различим. Более сложная картина наблюдается при расщеплении ДНК человека, где, помимо расчетных фрагментов с длинами 44–47, 89–91, 216–218, появляются более яркие фрагменты, представляющие собой, видимо, про-

дукты расщепления сателлитной ДНК с размерами 171 и 342 пар оснований (рис. 1с).

Представленные на рисунке 2 данные показывают, что расщепление хромосомной ДНК крысы по последовательности CC(A/T)GG дает более мелкие фрагменты, чем по сайту GGCC, однако все они до размера 462 п.о. видны на полиакриламидном геле. При этом более высокомолекулярные фрагменты (713, 1017 и 1690–1691) видны на агарозном геле на рисунке 5. Расщепление мышинной ДНК по сайту CC(A/T)GG также дает набор фрагментов, видимых на полиакриламидном геле, и два больших фрагмента 1511 и 1826 п.о., которые видны на рисунке 5. Следует также отметить наличие мощного фрагмента сателлитной ДНК размерами 234 п.о. (рис. 2б). Расщепление ДНК человека по сайту CC(A/T)GG дает уникальный набор из множества мелких фрагментов, которые представлены парами и поэтому особенно отчетливо видны на фотографии геля на рисунке 2с. Корреляция полученных экспериментальных и теоретических результатов наблюдается не только в соответствии количества и длин расчетных и наблюдаемых фрагментов, но и в их интенсивности. Как видно из диаграммы на рис. 2с, интенсивность пары

фрагментов 116, 118 п.о. существенно выше, чем пары фрагментов 166, 167 п.о., что соответствует данным электрофореза.

Представленные на рисунке 3 результаты по моделированию картины расщепления ДНК по сайту GATC дают большой набор мелких фрагментов в случае крысы и мыши, которые можно видеть на фотографии соответствующих гелей. Расщепление хромосомной ДНК человека дает только один существенный пик фрагментов ДНК размерами 174–175 п.о., который четко виден на картине геля.

На рисунке 4 представлены диаграммы расщепления трех ДНК по сайту CCGG, а на рисунке 5 приведена картина гидролиза этих ДНК ферментом MspI (сайт узнавания CCGG). На диаграмме расщепления ДНК крысы выявляются 2 пиковых значения для фрагментов с длинами 404 и 5537–5538 п.о., а соответствующие им полосы также отчетливо видны на снимке геля. Для ДНК мыши расчетные фрагменты наибольшей интенсивности для фермента MspI – 614, 3727 и 4992 п.о., и мы наблюдаем видимые полосы на электрофореграмме, отвечающие этим длинам. При расщеплении человеческой ДНК ферментом MspI фрагментов со значительной интенсивностью нет ни на диаграмме, ни на снимке геля.

Таким образом, как видно из рисунков 1–5, для большей части получаемых экспериментально фрагментов, за исключением фрагментов сателлитной ДНК, их длины соответствуют пиковым значениям на диаграммах, рассчитанным с помощью предложенного метода. В свою очередь, теоретически рассчитанные фрагменты с высоким значением функции Si видны на фотографиях гелей,

Очевидно, что существует некоторое пороговое значение величины функции Si, ниже которого полосы плохо визуализируются на геле. В условиях наших экспериментов пороговое значение для выявления фрагментов составило приблизительно 4 млн. п.о. (приблизительно 0,13–0,16% от всей длины генома). При этом необходимо учитывать, что интенсивность полос на геле, соответствующих фрагментам ДНК с примерно одинаковыми подвижностями, будет усиливаться благодаря эффекту наложения, что наблюдается в случае фрагментов, получаемых при гидролизе ДНК человека ферментом Bst2UI (сайт узнавания CC(A/T)GG). Аналогичный эффект наблюдается для случая гидролиза ДНК крысы рестриктазой MspI, когда высокомолекулярные фрагменты, имеющие на диаграмме значение функции S меньше 4 млн п.о., хорошо видны на геле.

Как видно из приведенных в настоящей работе фотографий гелей и теоретических диаграмм распределения фрагментов ДНК, наблюдается значительная корреляция расчетных данных и экспериментальных результатов. Дополнительные полосы с высокой интенсивностью, хорошо заметные при гидролизе ДНК мыши (рис. 2б, фрагмент 234 п.о.) и ДНК человека (рис. 1с, фрагменты 171 п.о. и 342 п.о.), видны значительно слабее при гидролизе ДНК крысы (рис. 1а, фрагмент 370 п.о.).

В последующих работах будут описаны отдельные исследования по расщеплению геномов крысы, мыши и человека по широкому спектру сайтов узнавания и сравнению полученных данных с экспериментальными результатами по гидролизу препаратов ДНК соответствующими эндонуклеазами рестрикции.

Заключение

В настоящей работе предложен простой метод поиска сайтов узнавания эндонуклеаз рестрикции в последовательностях геномов млекопитающих, позволяющий производить расчет длин и количества образуемых при расщеплении фрагментов ДНК, с последующим построением диаграмм их распределения. Используемое программное обеспечение не требует серверных мощностей и позволяет работать с протяженными нуклеотидными последовательностями длиной в несколько миллиардов п.о. на обычных персональных компьютерах. Мы полагаем, что данный метод окажется полезным в работах по изучению структуры геномов млекопитающих и приведет к существенному расширению списка сайт-специфических эндонуклеаз, применяемых в исследованиях ДНК эукариот. Хотя в своей работе мы ограничились анализом только трех геномов млекопитающих, очевидно, что использование предлагаемого метода рестрикционного анализа *in silico* возможно и в случаях геномов бактерий, грибов, растений и других организмов, для которых определена первичная структура ДНК.

Авторы благодарят к.б.н. Каледина В.И. за помощь в работе с животными и к.б.н. Васильева Г.В. за помощь в выделении ДНК.

Литература

1. Албертс Б., Брей Д., Льюис Дж., Рэфф М., Робертс К., Уотсон Дж. // Молекулярная биология клетки. — М.: Мир, 1994. — Т. 1. — С. 232–233.

2. Roberts R.J., Vincze T., Posfai J., Macelis D. REBASE – Restriction enzymes and DNA methyltransferases // *Nucleic Acids Res.* – 2005. – Vol. 33 (Database issue). – D230–D232.
3. Дашкевич В.С., Дымицкий Г.М., Салганик Р.И., Уланов Б.П. О дестабилизации вторичной структуры ДНК регенерирующей печени крыс в процессе репликации // *Молекулярная биология.* – 1972. – Т. 6 – № 5. – С. 689–697.
4. Lindblom B., Holmlund G. Rapid DNA purification for restriction fragment length polymorphism analysis // *Gene Anal. Tech.* – 1988. – Vol. 5. – P. 97–101.
5. *Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome* // *Nature.* – 2002. – Vol. 420. – P. 520–562.
6. *Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution* // *Nature.* – 2004. – Vol. 428. – P. 493–521.
7. *International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome* // *Nature.* – 2001. – Vol. 409. – P. 860–921.
8. Льюин Б. // *Гены.* – М.: Мир, 1987. – С. 298–308.
9. Willard H.F. Chromosome-specific organization of human alpha satellite DNA // *Am. J. Hum. Genet.* – 1985. – Vol. 37. – P. 524–532.
10. Pech M., Igo-Kemenes T., Zachau H.G. Nucleotide sequence of a highly repetitive component of rat DNA // *Nucleic Acids Res.* – 1979. – Vol. 7. – P. 417–432.

METHOD OF RESTRICTION ANALYSIS OF THE MAMMALS GENOMES IN SILICO

M.A. ABDURASHITOV, V.N. TOMILOV, V.A. CHERNUKHIN, D.A. GONCHAR,
S.Kh. DEGTAREV

SibEnzyme Ltd., Novosibirsk

A theoretical method to obtain the digestion patterns of mammalian chromosomal DNA cleavage with restriction endonucleases was proposed. Based on the recently published data on primary structures of genomes, a computative analysis was performed and diagrams of chromosomal DNA fragments distribution were obtained for DNA cleavages at 5'-GGCC-3', 5'-CCGG-3', 5'-GATC-3' and 5'-CC(A/T)GG-3' sequences. Experiments on chromosomal DNA digestion with HaeIII, MspI, Kzo9I and Bst2UI restriction endonucleases which recognize these sites were carried out. The correspondence of computed diagrams and experimentally obtained restriction patterns was shown.

Keywords: method, repetitive DNA elements, restriction analysis, eukaryotic genome, man, rat, mouse.