

ВЫЯВЛЕНИЕ ФРАКТАЛОПОДОБНЫХ СТРУКТУР В ДНК-ПОСЛЕДОВАТЕЛЬНОСТЯХ

Владимир Гусев, Любовь Мирошниченко, Надежда Чужанова

Аннотация: Разработан и реализован алгоритм выявления фракталоподобных структур в ДНК-последовательностях. Фрактальность трактуется как самоподобие, основанное на свойстве симметрии или комплементарной симметрии. Локальные фракталы интересны своей способностью аккумулировать множественные палиндромно-шпильчатые структуры с потенциально возможными регуляторными функциями. Выявлены реальные случаи проявления фрактальности в различных геномах: от вирусов до человека. Рассмотрена возможность использования фракталоподобных структур в качестве маркеров, различающих близкие классы последовательностей.

Keywords: DNA sequences, fractal-like structures, repeated fragments, palindrome, complementary palindrome.

ACM Classification Keywords: J. Computer Applications – J.3 Life and medical sciences – Biology and genetics; I. Computing Methodologies- I.5 Pattern recognition – I.5.2. – Design methodology –Feature evaluation and selection.

Conference: The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009

Введение

Отдельные фрагменты ДНК характеризуются проявлениями самоподобия, основанного на свойстве симметрии или комплементарной симметрии. По ассоциации с работой [Mandelbrot, 1992], будем называть их локальными фракталами (при отсутствии искажений) или фракталоподобными структурами (при их наличии). Такого рода объекты встречаются в участках аномально низкой сложности, содержащих повторяющиеся симметричные фрагменты (палиндромы) или комплементарные палиндромы [Gusev, 1999]. Значимость последних в регуляции генетических процессов не вызывает сомнений, тогда как роль обычных палиндромов не столь очевидна. Можно, тем не менее, указать на работу [Vacolla, 2004], в которой приведены примеры нестандартных структур, в образовании которых принимают участие близко расположенные симметричные фрагменты.

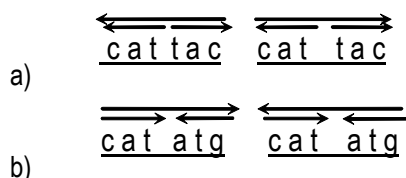
Реальные примеры фракталоподобных структур обнаружены нами в зоне начала репликации генома бактериофага λ [Гусев, 1989]. Фракталоподобную (в наших терминах) структуру образует сайт связывания trpR-репрессора E.coli, образованный повторением двух комплементарных палиндромов АСТАГТ со вставкой биграммы ТА между ними [Karlin, 2005]. В цитируемой работе этот пример приводится в связи с обсуждением аномалий в позиционном распределении комплементарного палиндрома СТАГ по длине генома. В недавней работе [Загоскин, 2008] по изучению диминуции хроматина в геноме S.kolensis (пресноводный рачок) было выявлено, что три из четырех исследовавшихся межмикросателлитных локуса длиной от 500 до 750п.н. остаются в геноме соматических клеток после прохождения диминуции, тогда как один из них устраняется в процессе диминуции. Именно в этом фрагменте ДНК обнаружен длинный комплементарный палиндром GGТАСГТGCАСГТАСС, который в двух из пяти вхождений повторяется тандемно. Возможно, возникающие при этом фракталоподобные структуры имеют отношение к объяснению механизма диминуции.

Приведённые примеры свидетельствуют об актуальности изучения фракталоподобных структур и их роли в регуляции основных генетических процессов. Поскольку ни один из известных методов (сложностные профили [Гусев, 1999], сканирующие статистики [Karlin, 1989], алгоритмы отыскания тандемных повторов [Stoichemore, 1994]) не гарантируют в общем случае выявления всех фракталоподобных структур с

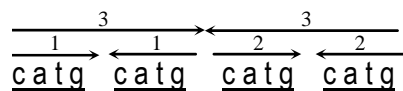
учетом возможного их *наложения* и заданными *ограничениями* на длину базового повтора ("периода"), размер гэпов между соседними его вхождениями и кратность повторений, требуется специальный алгоритм, удовлетворяющий этим ограничениям. *Целью работы* является разработка такого *алгоритма*, его апробация на различных текстах и характеристика выявляемых *фракталоподобных структур*.

Основные понятия и предпосылки

Обычный палиндром – это фрагмент, одинаково читаемый в обоих направлениях (например, cattac). Комплементарный палиндром удовлетворяет этому свойству лишь при переименовании элементов алфавита в одном из двух прочтений: $a \rightarrow t$, (a вместо t), $t \rightarrow a$, $c \rightarrow g$, $g \rightarrow c$, что соответствует известному в молекулярной биологии отношению комплементарности. Так, цепочка catatg, является комплементарным палиндромом, поскольку при прочтении ее справа налево с заменой g на c , t на a и т.д. получаем исходную последовательность символов. Элемент самоподобия проявляется в том, что повторение обычного симметричного палиндрама (случай "а") или комплементарного (случай "б") приводит к образованию нового палиндрама (соответственно комплементарного палиндрама) вдвое большей длины:



(здесь повторяющиеся фрагменты подчеркнуты; расходящиеся стрелки сверху обозначают палиндром, сходящиеся – комплементарный палиндром). При кратности повторений выше двух возникают множественные структуры (см. 1,2,3 и т.п. в примере с 4-кратным повторением комплементарного палиндрама catg):



Структуры, образуемые точными ("совершенными") тандемными повторами симметричных (в широком смысле) фрагментов, мы называем локальными фракталами. При наличии незначительных искажений внутри повторяющихся фрагментов, равно как и вставок между ними, используем термины "фракталоподобные структуры" или "несовершенные локальные фракталы". Предполагается, что размеры вставок сопоставимы с длинами повторяющихся фрагментов. Функциональная значимость при вставках не теряется: возникают "шпилечные" структуры, участвующие в регуляции многих генетических процессов.

Заметим, что множественные налагающиеся структуры возникают и при повторении левой (случай "а") или правой (случай "б") половинок симметричного в широком смысле фрагмента, даже если для этих половинок не выполняется свойство симметрии:



Однако свойство самоподобия в этом случае теряется.

Алгоритм обнаружения локальных фракталов

Поиск совершенных структур основан на вывлении всех тандемных повторов в тексте и отбору тех из них, в которых наименьший период (или его циклический сдвиг) является палиндромом или комплементарным палиндромом. Мы используем для выявления тандемных повторов технику сложностного анализа, а именно, ДНК-ориентированный вариант [Gusev, 1999] меры сложности Лемпеля и Зива [Lempel, 1976] реализованный для режима скользящего окна. Размер окна является естественным ограничением на

размеры выделяемых фракталов. Однако его увеличение практически не влияет на трудоемкость, поскольку при сдвиге окна на символ компоненты сложностного разложения не пересчитываются заново, а корректируются, причем далеко не все.

Пусть Σ – исходный алфавит; S – текст, составленный из элементов Σ ; $N = |S|$ – длина текста S ; $S[i]$ – элемент S , стоящий в i -й позиции ($1 \leq i \leq N$); $S[i : j]$ – фрагмент S , включающий элементы с i -го по j -й ($1 \leq i < j \leq N$). Сложностным разложением текста S назовём разбиение его на фрагменты ($S = v_1v_2\dots v_m$), где $v_1 = S[1]$, а v_k при $k > 1$ определяется следующим образом: если $|v_1v_2\dots v_{k-1}| = i$, то v_k – наибольший префикс u фрагмента $S[i + 1 : N]$, встречающийся хотя бы дважды в тексте $v_1v_2\dots v_{k-1}u$. Если такого u не существует, полагаем $v_k = S[i + 1]$. Следует иметь в виду, что v_k и его прототип нерасширяемы вправо, но могут допускать *расширение влево*, трактуемое как удлинение повтора. Число добавляемых слева символов строго меньше длины предыдущего компонента $|v_{k-1}|$. Общее число компонентов в разложении (сложность текста) не меняется, но компоненты лучше согласуются со структурой повторов в тексте. Данное определение сложностного разложения не запрещает наложения (со сдвигом) последнего вхождения u на предшествующее ему, что является индикатором наличия тандемной повторности.

Обозначим через $j(k)$ позицию, с которой начинается предпоследнее вхождение u в $v_1v_2\dots v_{k-1}u$ (в терминах [Lempel, 1976], $j(k)$ – это указатель ближайшего возможного прототипа для u). Если $j(k) + |v_k| \geq i + 1$ (*), т.е. прототип вплотную примыкает к порождаемому компоненту или накладывается на него, то имеет место тандемная повторность с длиной периода $t = i + 1 - j(k)$ и кратностью повторений не меньшей, чем $\text{entier}(|v_k| / t) + 1$. Можно показать, что проверка условия (*) значительно упрощается, если работать с *расширенными влево* компонентами разложения. Тогда следует проверять лишь компоненты со значением $j(k) = 1$. Если расположение тандемной структуры синхронизовано с началом окна, для всех значений $j(k) = 1$ выполняется и условие (*). Если тандемов нет, $j(k)$ может равняться 1, но условие (*) не выполняется. Если тандемы встречаются не в начале, а внутри окна, условие (*) может выполняться, но $j(k) \neq 1$ (эти тандемы будут выявлены при последующем движении окна).

Трудоемкость алгоритма в среднем составляет $O(N \log N)$, однако на текстах ограниченной длины он работает быстрее линейного алгоритма, описанного в [Crochemore, 1994]. Такое возможно при существенном различии в значениях мультипликативных констант в обоих алгоритмах (малое значение в нелинейном алгоритме и большое – в линейном).

Алгоритм обнаружения несовершенных фракталоподобных структур

Рассматриваются структуры, образованные повторяющимися палиндромами (обоих типов) при условии, что расстояние между соседними повторами не превышает заданного порога r . Искажение самих палиндромов не допускается. Фрагмент **agagaagactagattcaagatcaga**, например, при значении $r = 4$ относится к категории интересующих нас фракталоподобных структур с базовым симметричным повтором "aga". Фактически речь идет о выделении кластеров слов определенного типа с числом слов в кластере не меньшим 2 и расстоянием между соседями не большим r . Мы вновь используем идеологию *скользящего окна*, размер которого W ограничивает сверху размеры выделяемых структур.

Шаг 1 алгоритма связан с построением L -граммного дерева для фрагмента, выделяемого окном. Оно фиксирует полную совокупность L -грамм (связных цепочек из L символов), представленных в окне. У L -грамм "склеиваются" общие начала ("trie"-структура). Рёбра дерева помечены символами из L -граммных цепочек. В узле дерева содержится информация о местах вхождения в текст L -граммы, помечающей путь из корня в данный узел. Дерево строится для начального положения окна. Далее, при движении окна по тексту, оно лишь корректируется [Гусев, 2001]. Дерево используется для быстрого поиска палиндромов в пределах окна и проверки условий кластеризуемости. Параметр L примерно соответствует средней длине повторов (6–10 символов).

Шаг 2 алгоритма связан с попыткой обнаружения элементарного (минимального по длине) палиндрома в начале окна, которое рассматривается как центр симметрии. Проще всего воспользоваться "наивным"

алгоритмом, осуществляющим последовательное (до первого несовпадения) сравнение пар символов слева и справа от оси симметрии, которая может проходить через символ (палиндром нечётной длины) или между соседними символами (палиндром или комплементарный палиндром чётной длины). Если палиндром не найден, окно сдвигается на символ, корректируется L -граммное дерево (шаг 1) и повторяется шаг 2.

Шаг 3. Если на шаге 2 палиндром найден, с помощью L -граммного дерева проверяется наличие других его вхождений в окно анализа. По цепочке символов, образующих палиндром, двигаемся от корня до вершины, содержащей нужную информацию (при длине палиндрома большей L продолжение цепочек отслеживается по тексту). Если список вхождений не пуст, проверяются условия кластеризации и принимается решение о наличии (отсутствии) фракталоподобной структуры. Затем окно сдвигается на один символ и осуществляется поиск нового элементарного палиндroma. В случае успеха может быть выявлен новый кластер, наложенный на предыдущий, но с другим повторяющимся ядром (см. фрагмент генома вируса Эпштейна-Барр):

поз. 70490: $\overleftrightarrow{g g c c g g} \overleftrightarrow{g g c c g} \rightarrow \leftarrow \overleftrightarrow{g g c c g} \overleftrightarrow{g g c c}$

поз. 70491: $\overleftrightarrow{g c c g g g c c g} \overleftrightarrow{s a g a g} \overleftrightarrow{g c c g g g c c g}$

Здесь один и тот же фрагмент текста (сдвиг всего на символ) представлен вверху комбинацией комплементарных палиндромов с ядром $ggcc$, а внизу - комбинацией обычных палиндромов с ядром $gscg$.

Трудоемкость алгоритма составляет $O(N \cdot L)$, если считать среднюю (по всем позициям) длину выделяемых палиндромов, не зависящей от N , что, как правило, выполняется для ДНК-последовательностей.

Экспериментальные результаты

Апробация алгоритмов проводилась на вирусных и бактериальных геномах, отдельных генах человека (кодирующие последовательности) и хромосомах генома *Arabidopsis thaliana* (растение). Результаты можно суммировать следующим образом:

- невырожденные по НК-составу фракталоподобные структуры встречаются довольно редко. Характерные длины повторяющихся палиндромов – от 4 до 10, кратность повторения – от 2 до 8 (для совершенных структур – до 5). Исключение составляют участки микросателлитной ДНК с длиной периода 2 или 3 ($(ta)^n, (aga)^n$ и т.п., $n \sim 10$ и выше), обладающие специфическими свойствами ввиду вырожденности НК-состава (легкоплавкость и т.п.);

- не симметричные тандемные повторы радикально отличаются от локальных фракталов наличием длинных (десятки символов) периодов (рекордное наблюдавшееся нами значение для несовершенных фракталов – 14, для совершенных – не более 10);

- фракталоподобные структуры со вставками часто возникают в обычных тандемных повторах, содержащих симметричный фрагмент внутри периода (например, $(\overleftrightarrow{t g g a g g t g g c t a})^n$). Аналогичный пример, но уже с комплементарным палиндромом, демонстрирует фрагмент гена *Ecodermal dysplasia 1* у человека (accession number AF060999 в GenBank, поз. 599):

$\overleftrightarrow{g g a a t t c c a} \overleftrightarrow{g g g a t t c c t} \overleftrightarrow{G g a a t t c C a} g g a a \dots$
 ... (Gly Ile Pro) (Gly Ile Pro) (Gly Ile Pro)...

Здесь комплементарный палиндром (1) длины 8 входит в состав тандемных повторов длины 9 (см. круглые скобки). Повторение (1) "усиливает" конструкцию: возникает структура (2) шпильчатого типа. Нам неизвестна её функциональная нагрузка, но интересно отметить, что имевшие место мутации в выделенных заглавными буквами позициях (замена G на а и C на t), ослабляющие структуру и на нуклеотидном и на аминокислотном уровне, приводили к наследственным заболеваниям;

– наряду с проявлениями фрактальности выделяемые структуры порой демонстрируют другие уникальные свойства. Одна из таких структур обнаружена в четвертой хромосоме генома *Arabidopsis th.* Её ядро составляет комплементарный палиндром **tgatgatgaca**. Для (a,t)-богатой хромосомы этот идеально сбалансированный по НК-составу фрагмент встречается неожиданно часто – 194 раза. Его инвертированная копия встречается всего 1 раз (аномальная асимметрия!). Все вхождения сосредоточены в диапазоне от 1816917-й до 5664586-й позиции, что при длине хромосомы, равной 18585042нк, следует охарактеризовать как сильную позиционную аномалию. И, наконец, внутри указанного диапазона почти половина всех вхождений являются спаренными, т.е. отстоят друг от друга на 9-10 нк, образуя фракталоподобные структуры с гэпами. Две из них приведены для иллюстрации ниже (поз.1999210 и. 3931735):

$\overleftrightarrow{\text{tgatga}} \overleftrightarrow{\text{tcacatgag}} \text{tgatga} \text{tcacatgag} \overleftrightarrow{\text{tgatga}} \overleftrightarrow{\text{tcacatgag}}$
tgatga tcacatgag tgatga tcacatgag

$\overleftrightarrow{\text{tgatga}} \overleftrightarrow{\text{tcacatgag}} \text{gaggtagtaa} \overleftrightarrow{\text{tgatga}} \overleftrightarrow{\text{tcacatgag}}$
tgatga tcacatgag gaggtagtaa tgatga tcacatgag

Нетрудно видеть, что при строгой консервативности самих палиндромов и расстояний между ними разделяющие их вставки эволюционируют относительно свободно. Большая часть структур связана с ретротранспозонами ("gypsy-like retrotransposon family (Athila)", "non-LTR retrotransposon family (LINE)" и т.п.) Поскольку мобильные элементы рассматриваются как своего рода "батареи" регуляторных элементов, перемещение которых по геному меняет экспрессию генов, можно предполагать, что выделенная структура также имеет отношение к регуляции этого процесса;

– просматривается возможность использования фракталоподобных структур в качестве признаков (маркеров), разделяющих те или иные классы объектов. Рассматривались два класса объектов – геномы вирусов клещевого энцефалита (ВКЭ) и вирусов Повассан (ВП) – представители одного и того же рода флавивирусов. Эти геномы представлены молекулой одноцепочечной РНК длиной около 11 тыс. нуклеотидов, содержащей единственную открытую рамку считывания, в которой последовательно закодированы все структурные и неструктурные белки. Эти (индивидуальные) белки образуются в результате посттрансляционного расщепления единого полипротеина, длина которого у разных штаммов практически не меняется (~ 3414 аминокислотных остатков). Выравнивание кодирующих последовательностей у разных штаммов показывает высокий уровень гомологии (свыше 90%).

Нас интересовали два вопроса: 1) существуют ли на уровне РНК фракталоподобные структуры, разделяющие геномы ВКЭ и ВП? 2) существуют ли фракталоподобные структуры, разделяющие геномы одного класса (ВКЭ) на две группы: инаппарантные штаммы (выделенные от людей с диагнозом «укус клеща», но с бессимптомным течением заболевания) и высоко вирулентные (болезнетворные) штаммы?

Ответ на первый вопрос – положительный. В качестве примера можно указать на совершенный локальный фрактал **scatggccatgg**, выявленный в поз. 440 в геномах вируса Повассан: штаммы Nadezdinsk (номер доступа EU670438 в EMBL/Genbank), Spassk-9 (EU770575), Partizansk (EU643649) и LB (LO6436). Другой маркер – **tggccatggcca**, получающийся циклическим сдвигом из первого, выявлен в поз. 4223. У вируса клещевого энцефалита эти маркеры отсутствуют. Результаты носят предварительный характер ввиду ограниченности исходных данных

Ответ на второй вопрос требует более детального изучения. С одной стороны, можно указать на совершенные фракталы aggaaggaagga и gtgggtgggtg, первый из которых представлен только в инаппарантных штаммах в поз. 5995 (Primorye-212: EU816450, Primorye-253: EU816451 и др.), а второй – в высоковирулентных в поз. 933 (Primorye-94: EU816454, Dalnegorsk: FJ402886). С другой стороны, ввиду близости геномов чаще имеет место ситуация, когда фракталоподобная структура присутствует в варьированной форме как в инаппарантных, так и в высоковирулентных штаммах и требуется детально фиксировать различия между ними. Ниже приведено выравнивание фракталоподобных структур, выявленных в двух высоковирулентных штаммах Primorye-94 (длина структуры $l = 26$, поз. 439, базовый повтор gttg) и Kavalerovo (FJ402885, $l = 23$, поз. 441, базовый повтор tgg). В квадратные скобки заключён более короткий фрагмент, выявленный как фракталоподобная структура в инаппарантных штаммах Primorye-212, Primorye-253 и др. Видно, что зона относительной нестабильности расположена в конце и правее этого фрагмента: здесь имеются разночтения во всех сравниваемых геномах как на нуклеотидном, так и на аминокислотном уровне, прослеживается множественность структур. В штамме Kavalerovo чётко проявлена периодичность (ctgggt)³ – подчеркнута, в штамме Primorye-94 – симметрия (стрелки сверху).

	V V L L	
	←-----→	
поз. 439	[g t t g g t t g c t g g t t g t t g] t c c t g t t g	- Primorye-94
поз. 441	t g g t t g <u>c t g g t t c t g g t t c t g g t</u>	- Kavalerovo
...	W L L V L V L V...	

Наибольший интерес, однако представляет тот факт, что эти фракталоподобные структуры находятся на стыке капсидного белка С и preM – полипептида, что может свидетельствовать об их функциональной или эволюционной значимости*. В известных нам работах по структуре генома ВКЭ роль межпротеиновых интервалов никак не освещена. Добавим также, что на аминокислотном уровне обсуждаемая фракталоподобная структура представлена неслучайным кластером гидрофобных аминокислот, выделяемым с помощью позиционного анализа [Гусев, 2002]

Заключение

Введено понятие локального фрактала (совершенного и несовершенного) для фрагментов ДНК, характеризующихся проявлениями структурного самоподобия. Локальные фракталы содержат множество палиндромно-шпильчатых структур, налагающихся друг на друга. Возможно, они обладают различной функциональной нагрузкой и в этом смысле похожи на многозначные слова в естественном языке. Разработаны и реализованы алгоритмы выявления локальных фракталов в ДНК-последовательностях. Проведена их характеристика по результатам обработки ряда геномов. Указаны наиболее существенные отличия от проявлений обычной тандемной повторности. Приведены примеры наиболее интересных фракталоподобных структур.

Благодарности

Работа выполнена в рамках интеграционного проекта СО РАН № 63

* Наличие высокорегулярных фрагментов относительно небольшой длины на границах между крупными структурными единицами геномов отмечалось нами и другими авторами в связи с анализом алгоритмов сегментации. Такие фрагменты несут на себе следы эволюционных перестроек и могут являться элементами регуляторных структур. В качестве тестового объекта часто рассматривается геном бактериофага λ с достаточно хорошо изученной модульной структурой (см., например, [Гусев, 1989], [Бородовский, 1990], [Braun, 1998])

Библиография

- [Bacolla, 2004] Bacolla A, Jaworski A, Larson J.E., Jakupciak J.P., Chuzhanova N.A., Abeyasinghe S.S., O'Connell C.D., Cooper D.N., Wells R.D. Breakpoints of gross deletions coincide with non-B DNA conformations. In: Proc. Natl. Acad. Sci. USA, 2004, Vol.101, P.14162-14167.
- [Crochemore, 1994] Crochemore M., Rytter W. Text Algorithms. In: Text Algorithms, Oxford University Press, New York, Oxford, 1994.
- [Gusev, 1999] Gusev V.D., Nemytikova L.A., Chuzhanova N.A. On the complexity measures of genetic sequences. In: Bioinformatics, 1999, Vol.15, No 12, P.994-999.
- [Karlin, 1989] Karlin S. Statistical signals in bioinformatics. In: PNAS USA, 1989, Vol. 102, No 38, P.13355-13362
- [Lempel, 1976] Lempel A., Ziv J. On the complexity of finite sequences. In: IEEE Trans. Inform. Theory, 1976, Vol. IT-22, No.1, P.75-81.
- [Mandelbrot, 1992] Mandelbrot B. The Fractal Geometry of Nature. In: The Fractal Geometry of Nature. San Francisco: Freeman, 1992.
- [Гусев, 1989] Гусев В.Д., Куличков В.А., Чупахина О.М. Сложностной анализ генетических текстов (на примере фага λ) // Препринт Института математики СО РАН, Новосибирск, 1989.- 49 стр.
- [Гусев, 2001] Гусев В.Д., Немытикова Л.А. Учет проявлений повторности, симметрии и изоморфизма в символьных последовательностях // Вычислительные системы, вып. 167. Новосибирск, 2001. С. 11–33
- [Гусев, 2002] Гусев В.Д., Немытикова Л.А., Саломатина Н.В. Выявление аномалий в распределении слов или связанных цепочек символов по длине текста // Интеллектуальный анализ данных. – Новосибирск, 2002. – вып. 171: Вычислительные системы. С. 51—74.
- [Загоскин, 2008] Загоскин М.В., Гришанин А.К., Королёв А.Л., Паленко М.В., Муха Д.В. Характеристика межмикросателлитных последовательностей ДНК до и после диминуции хроматина у *Cyclus kolensis* // ДАН, 2008, т. 423, № 4. С. 551-555.
- [Бородовский, 1990] Бородовский М.Ю., Певзнер П.А. Зонная структура генома фага лямбда. In: Компьютерный анализ генетических текстов. М., Наука, 1990. С. 62-67.
- [Braun, 1998] Jerom V. Braun, Hans-Georg Müller. Statistical methods for DNA sequence segmentation. In: Statistical Science, 1998. Vol. 13, No. 2, P. 142-162.

Authors' Information

Владимир Гусев – Старший научный сотрудник, Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, Россия, 630090, Новосибирск, пр. ак. Коптюга., 4; e-mail: gusev@math.nsc.ru

Любовь Мирошниченко – научный сотрудник, Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, Россия, 630090, Новосибирск, пр. ак. Коптюга., 4; e-mail: luba@math.nsc.ru

Надежда Чужанова – Reader in Bioinformatics, School of Computing, Engineering and Physical Sciences, University of Central Lancashire, Preston, PR1 2HE Great Britain; e-mail: nchuzhanova@uclan.ac.uk