

Подсистема «Филогения»: филогенетический и сравнительный анализ геномных последовательностей.

Структура документа (оглавление).

1. Цель и задачи подсистемы «Филогения».....	1
Филогенетический анализ генетических последовательностей.....	1
2. Структура подсистемы «Филогения» и детальное руководство по ее применению.....	1
2.1. Описание работы с программными компонентами подсистемы «Филогения».....	2
2.2.1. Программная компонента ALN_COMP: множественное выравнивание последовательностей.	2
2.2.2. Программная компонента TREE_COMP: реконструкция филогенетического дерева последовательностей.	4
2.2.3. Программная компонента ADAPT_SITE: поиск районов генов, подверженных адаптивной эволюции.	6
2.2.4. Программная компонента ADAPT_SEQ: поиск ветвей филогенетических деревьев, на которых предположительно проходила адаптивная эволюция.....	9
2.2.5. Программная компонента CONS_COMP: оценка консервативности позиций множественного выравнивания аминокислотных последовательностей.....	11
2.2.6. Программная компонента SPECPOS: поиск позиций, определяющих специфичность функции белков.....	14
3. Полезные ссылки.....	19

1. Цель и задачи подсистемы «Филогения».

Филогенетический анализ генетических последовательностей

Филогенетический анализ последовательностей генетических макромолекул является важной частью исследований в системной биологии. Он позволяет строить парные и множественные выравнивания аминокислотных и нуклеотидных последовательностей, оценивать филогенетическое родство последовательностей, принадлежащих разным видам внутри белкового семейства, оценивать консервативность и вариабельность позиций белковых и нуклеотидных последовательностей.

Функции филогенетического и сравнительного анализа нуклеотидных и аминокислотных последовательностей реализованы в подсистеме «Филогения».

2. Структура подсистемы «Филогения» и детальное руководство по ее применению

Подсистема «Филогения» предназначена для выполнения следующих функций:

филогенетический и сравнительный анализ геномных последовательностей путем сравнительного анализа нуклеотидных и белковых последовательностей (парное и множественное выравнивание),

реконструкции филогенетических деревьев семейств кодирующих последовательностей, поиска адаптивно эволюционирующих генов и проекция адаптивно эволюционирующих позиций генов на структурно-функциональную организацию белков,

поиска участков белков с пониженной (повышенной) скоростью фиксации аминокислотных замен и оценки их функциональной и структурной значимости.

2.1. Описание работы с программными компонентами подсистемы «Филогения»¹

2.2.1. Программная компонента *ALN_COMP*: множественное выравнивание последовательностей.

В системе использован алгоритм множественного выравнивания – MUSCLE, который использует две меры измерения сходства для пар последовательностей: k -кортеж расстояния (для невыровненной пары) и расстояния, рассчитанные по Кимуре (для выровненной пары) (Edgar R.C., 2004). k -кортеж – непрерывная подпоследовательность длины k , также известная как слово. Гомологичные последовательности имеют тенденцию иметь больше k -кортежей расположенных вместе, чем случайные последовательности. Оценка этой величины не требует выравнивания, что дает существенное преимущество в скорости вычислений. Кратко, алгоритм MUSCLE состоит из трех этапов.

- Цель первой стадии состоит в наиболее быстром множественном выравнивании – вначале рассчитываются k -кортежи, расстояния для каждой пары последовательностей, что дает матрицу расстояний, по которой строится дерево, по порядку ветвления дерева производится прогрессивное множественное выравнивание.
- Имея выровненную пару последовательностей, можно вычислить парное сходство и преобразовать его в аддитивную оценку расстояния между последовательностями, применяя при этом коррекцию Кимуры для многократных замен. Матрица расстояний группируется, используя алгоритм UPGMA. Расстояния для каждой пары последовательностей подвергаются коррекции по Кимуре, что дает матрицу расстояний, по которой вновь строится дерево. Если топология дерева изменяется по сравнению с деревом, построенным на первой стадии, множественное выравнивание исправляется.
- Последней стадией является итеративная процедура улучшения выравнивания – выбирается узел из дерева, построенного на второй стадии, дерево делится на два поддерева, путем удаления этого узла. Затем производится выравнивание двух профилей, при отсутствии улучшения (регистрируется с помощью оценки взвешенной суммы пар) алгоритм оканчивает работу. Для расчета профиля выравнивания в алгоритме MUSCLE используется особая функция логарифмического ожидания аминокислоты принимающая во внимание вероятности замен аминокислот (РАМ матрица), их частоты в последовательности, а также частоты делеций. Применяется Байесов подход для определения вероятности определенной аминокислоты на определенном месте в последовательности.

2.2.1.1. Описание работы *ALN_COMP*

Для запуска программной компоненты для множественного выравнивания последовательностей генетических макромолекул необходимо в левой панели стартовой страницы модуля «Филогения» выбрать гиперссылку «Sequence Alignment». После этого браузер загрузит стартовую страницу программной компоненты для множественного выравнивания последовательностей (Рис. 1).

¹ Подробнее о работе с программными компонентами «Филогения» смотри в Руководстве пользователя АСНИ-01, 4.2.3.

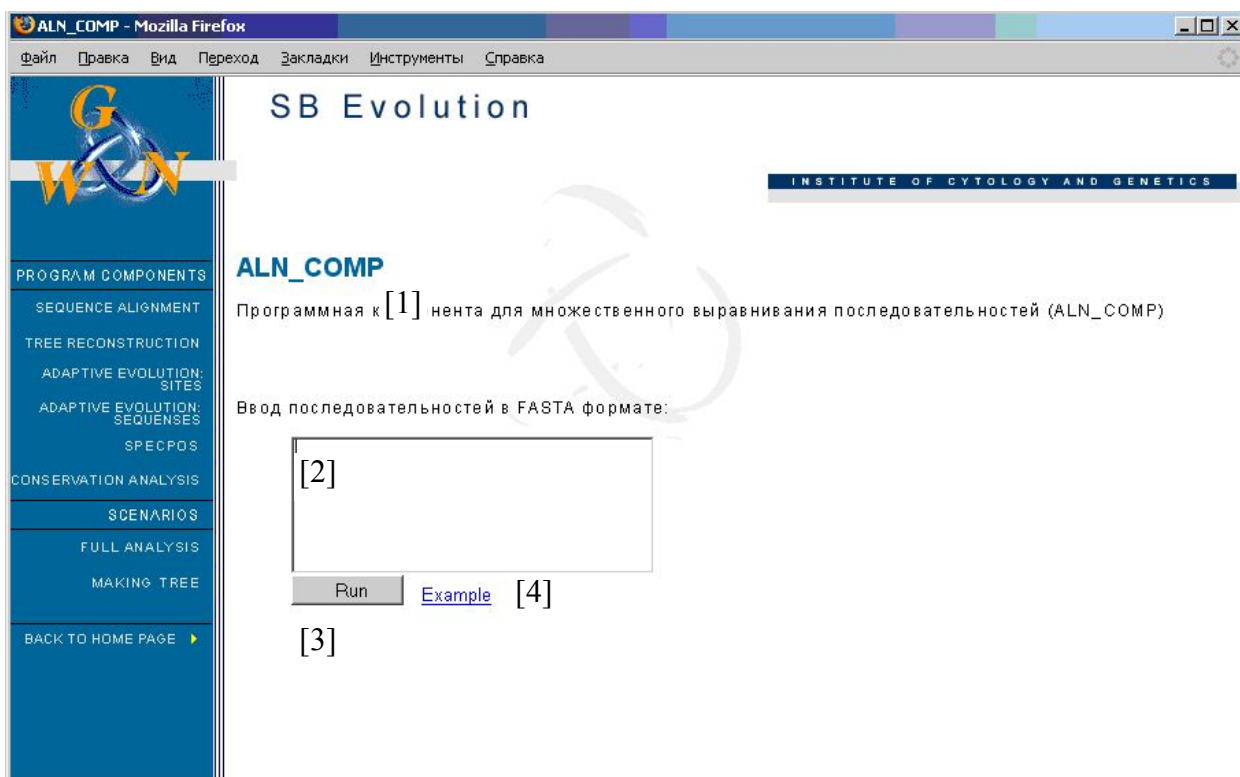


Рисунок 1. Стартовая страница для выполнения операции «Множественное выравнивание последовательностей».

Основные действия в требуемой последовательности

Для того чтобы ввести входные данные для анализа необходимо скопировать их в буфер обмена данными операционной системы в любом текстовом редакторе. После этого необходимо вернуться к странице ввода данных компоненты выравнивания последовательностей и поместить указатель мыши в поле текстового ввода [2] (Рис. 2). После этого необходимо скопировать данные из буфера в поле текстового ввода. Это делается нажатием комбинации клавиш <Ctrl-Ins> или <Ctrl-V>. После этого данные, помещенные в поле [2] (Рис. 2), отображаются в нем в текстовом виде.

После ввода данных в поле [2] (Рис. 2) для запуска программы необходимо нажать указателем мыши на кнопку «Run» [3] (Рис. 2). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

Для запуска контрольного примера необходимо перейти по гиперссылке «Example» [4] (Рис. 1).

Заключительные действия

После выполнения программы в левой панели появляется окно результата. Страница содержит три гиперссылки для различного способа отображения результатов расчета. При обращении к гиперссылке «Aligned sequences in text format» программа запустит новое окно браузера, в котором в текстовом формате будет представлено множественное выравнивание последовательностей в FASTA формате (Рис. 2).

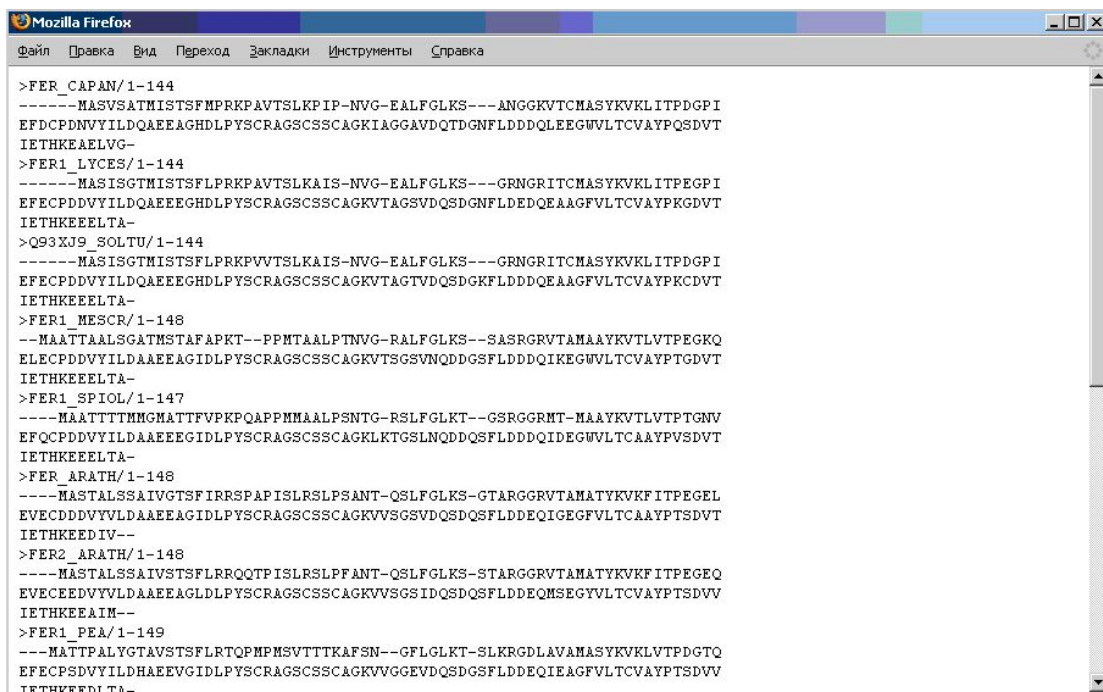


Рисунок 2. Страница визуализации множественного выравнивания в текстовом виде после выполнения операции «Множественное выравнивание последовательностей».

2.2.2. Программная компонента TREE_COMP: реконструкция филогенетического дерева последовательностей.

Для построения филогенетического дерева по матрице парных эволюционных расстояний используется метод объединения соседей, который является модифицированным методом минимальной эволюционной дистанции использующий особый алгоритм построения дерева - разложение звезды (Saitou N. and Nei M., 1987).

2.2.2.1. Описание работы TREE_COMP

Для запуска программной компоненты реконструкции филогенетического дерева необходимо в левой панели стартовой страницы выбрать гиперссылку «Tree reconstruction». После этого браузер загрузит стартовую страницу программной компоненты для реконструкции филогенетического дерева (Рис. 4). Панель ввода данных располагается справа и содержит название программной компоненты [1]. Панель содержит поле ввода множественного выравнивания [2] и кнопку запуска счета [3], а так же гиперссылку для перехода на страницу контрольных данных [4].

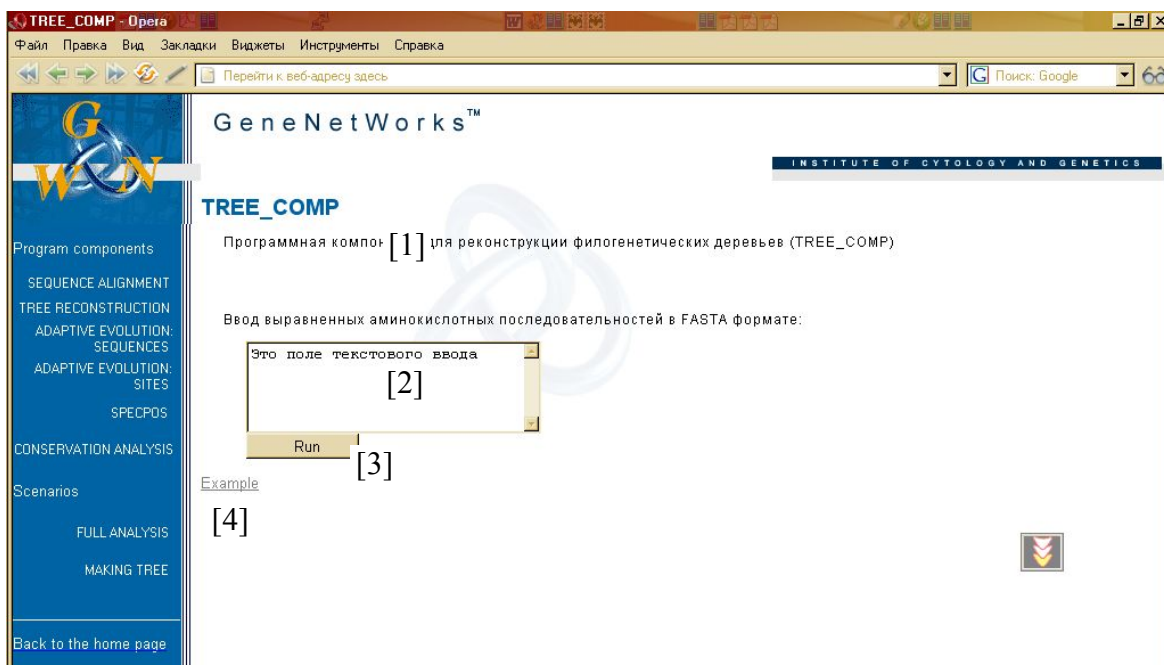


Рисунок 3. Стартовая страница для выполнения операции «Реконструкция филогенетического дерева по аминокислотным и нуклеотидным последовательностям».

Основные действия в требуемой последовательности

Для того чтобы ввести входные данные для анализа необходимо скопировать их в буфер обмена данными операционной системы в любом текстовом редакторе. После этого необходимо вернуться к странице ввода данных компоненты выравнивания последовательностей и поместить указатель мыши в поле текстового ввода [2] (Рис. 3). После этого необходимо скопировать данные из буфера в поле текстового ввода. Это делается нажатием комбинации клавиш <Ctrl-Ins> или <Ctrl-V>. После этого данные, помещенные в поле [2] (Рис. 3), отображаются в нем в текстовом виде.

После ввода данных в поле [2] для запуска программы необходимо нажать указателем мыши на кнопку «Run» [3] (Рис. 11). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

Для запуска контрольного примера необходимо перейти по гиперссылке “Example” [4] (Рис. 3)

Заключительные действия

После выполнения программы в левой панели появляется окно результата. Данная страница содержит гиперссылки для отображения результатов расчета.

При обращении к гиперссылке «Phylogenetic tree in text» программа запустит новое окно браузера, в котором будет представлено филогенетическое дерево в формате PHYLIP (Рис. 5).

3) Движущий отбор. В редких случаях на определенных этапах эволюции скорости фиксации несинонимических замен выше, чем синонимических, это означает, что происходит интенсивный процесс фиксации замен, который может отражать адаптивный режим эволюции белка связанный с изменением его функции. Такой режим может, например, иметь место после дубликации гена на стадии приобретения им новой функции. Интересно отметить, что на гены-паралоги стабилизирующий отбор действует слабее, чем на ортологи. Движущий отбор характерен также и для белков вирусов и для эпителий иммуноглобулинов, с ними взаимодействующих. В данном случае происходит постоянная коэволюция этих белков.

Для выявления движущего отбора связанного с процессом адаптивной эволюции, долгое время использовалось, и, к сожалению, используется и по ныне, отношение скоростей несинонимических и синонимических замен $\omega = K_a/K_s$, где K_a и K_s скорости фиксации синонимических и несинонимических замен на один кодон, или на один нуклеотид, соответственно. Если значение ω превышает единицу, предполагается, что можно утверждать о том, что имеет место движущий отбор.

Подходом для выявления возможных событий движущего отбора, который использован в нашей системе, является поиск участков подвергающихся быстрым изменениям на множественном выравнивании аминокислотных последовательностей (оценка константы ω для каждого кодона последовательности). Основная идея группы таких методов заключается в том, что медленная скорость эволюции поверхностных аминокислотных остатков на трехмерной структуре белков является результатом ограничений наложенных связыванием, например, с иными белками, лигандами или молекулой ДНК. В типичном случае алгоритм работы таких методов следующий: 1) построение филогенетического дерева; 2) использование этого дерева, для кластеризации эволюционно близких последовательности; 3) реконструкция консенсусной или предковой последовательности для каждого кластера и/или для каждого узла дерева; 4) сравнение консенсусов или рассчитанных предковых последовательностей и, например, определение каждой позиции как переменной или постоянной; 5) отображение состояния каждого сайта на трехмерной структуре белка. В нашей системе реализован метод, предложенный Пупко и сотр. - алгоритм RATE4SITE, который реализует метод максимального правдоподобия для описания процесса аминокислотных замен (Pupko T. et al., 2002). Для работы алгоритма RATE4SITE требуется множественное выравнивание белков, модель эволюции белков и филогенетическое дерево. Алгоритмом RATE4SITE рассчитывает с использованием информации о филогенетическом дереве вероятность замен для каждой позиции выравнивания.

2.2.3.1. Описание работы ADAPT_SITE

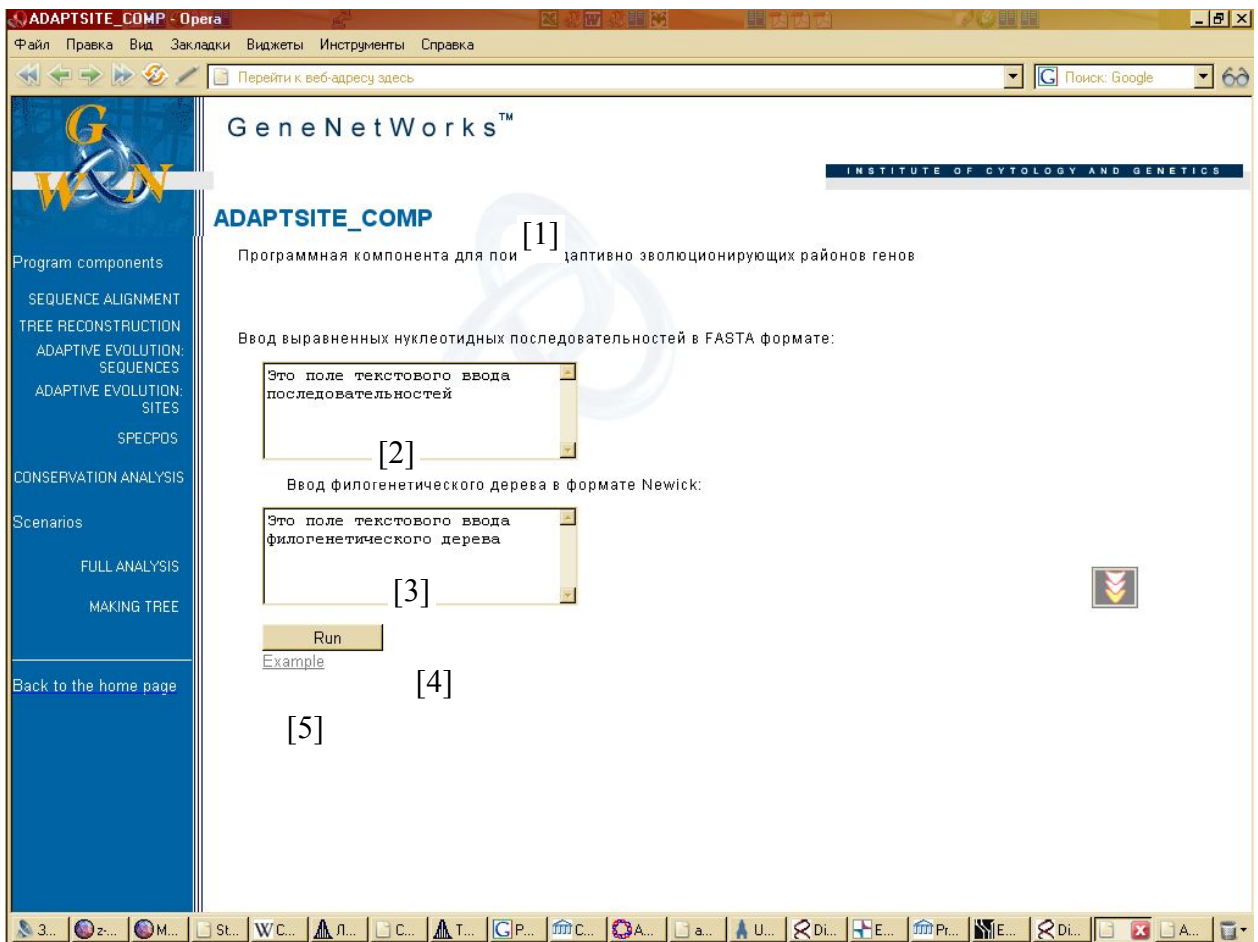


Рисунок 6. Стартовая страница для выполнения операции «Поиск адаптивно эволюционирующих районов генов».

Основные действия в требуемой последовательности

Для того чтобы ввести входные данные для анализа необходимо скопировать их в буфер обмена данными операционной системы в любом текстовом редакторе. После этого необходимо вернуться к странице ввода данных компоненты выравнивания последовательностей и поместить указатель мыши в поле текстового ввода [2] (Рис. 6). После этого необходимо скопировать данные из буфера в поле текстового ввода. Это делается нажатием комбинации клавиш <Ctrl-Ins> или <Ctrl-V>. После этого данные, помещенные в поле [2] (Рис. 6), отображаются в нем в текстовом виде.

После ввода данных в поля [2] и [3] (Рис. 6) для запуска программы необходимо нажать указателем мыши на кнопку «Run» [4] (Рис. 6). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

Для запуска контрольного примера необходимо перейти по гиперссылке “Example” [5] (Рис. 6)

Заключительные действия

После выполнения программы в левой панели появляется окно результата (Рис. 7).

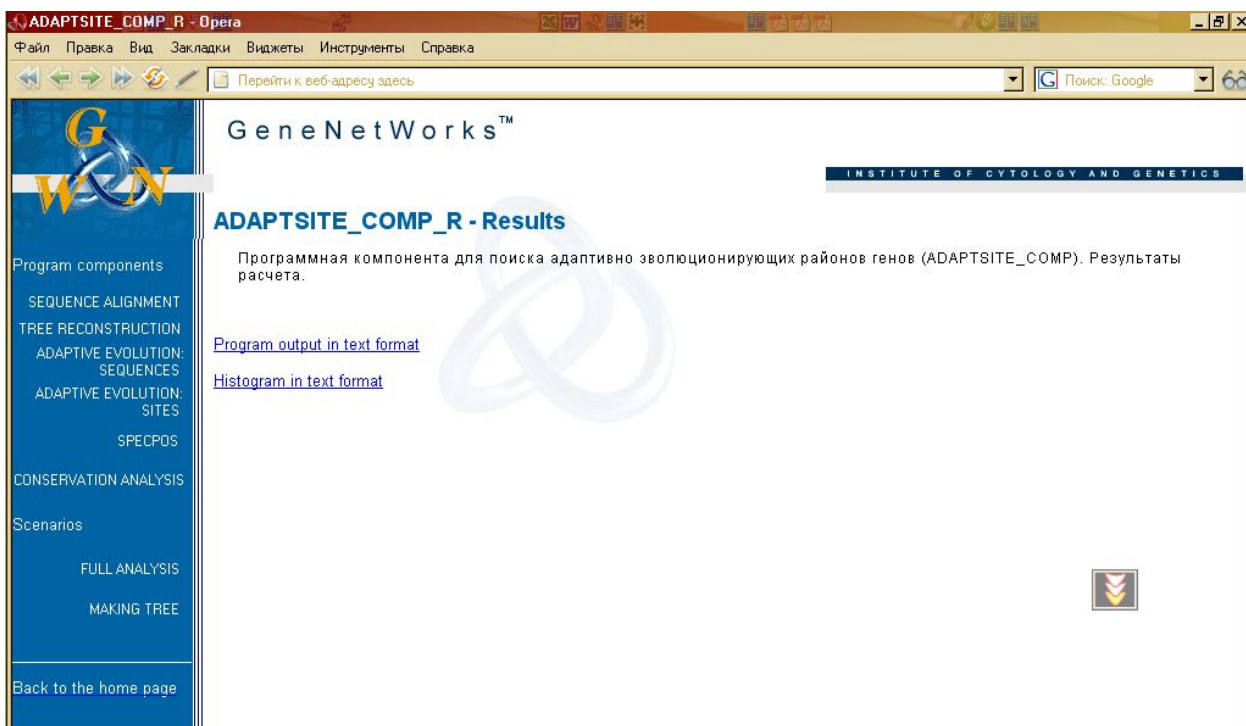


Рисунок 7. Страница результата после выполнения операции «Поиск адаптивно эволюционирующих районов генов».

Данная страница содержит гиперссылки для отображения результатов расчета.

При обращении к гиперссылке «Histogram in text format» [1] (Рис. 7) программа запросит сохранение файла матрицы попарных сравнений, который может быть открыт при помощи MS Excel. При обращении к гиперссылке «Program output in text format» (Рис. 7) программа откроет новое окно с текстовыми данными (Рис. 8).



Рисунок 8. Страница результата после выполнения операции «Поиск адаптивно эволюционирующих районов генов».

2.2.4. Программная компонента ADAPT_SEQ: поиск ветвей филогенетических деревьев, на которых предположительно проходила адаптивная эволюция.

Для того чтобы оценить, происходила ли на этапе, соответствующем одной из ветвей филогенетического дерева, адаптивная эволюция последовательностей, как правило, оценивают среднее отношение ω для всех кодонов последовательности между парами последовательностей, которые разошлись в ходе эволюции белкового семейства. В системе использован алгоритм для оценки адаптивного режима в парах последовательностей, предложенный Тангом и Ву (Tang H. and Wu C.I., 2006). Этот метод при оценке скоростей замен учитывает неравномерность частот замен кодонов, а так же физико-химическую природу аминокислот, которые он кодируют.

Для запуска программной компоненты поиска ветвей деревьев, на которых предположительно проходила адаптивная эволюция, необходимо в левой панели стартовой страницы (Рис. 1) выбрать гиперссылку «Adaptive evolution: sequences». После этого браузер загрузит стартовую страницу программной компоненты для реконструкции филогенетического дерева (Рис. 9). Панель ввода данных располагается справа и содержит название программной компоненты [1]. Панель содержит поле ввода множественного выравнивания [2] и кнопку запуска счета [3], а так же гиперссылку для перехода на страницу контрольных данных [4].

2.2.4.1. Описание работы ADAPT_SEQ

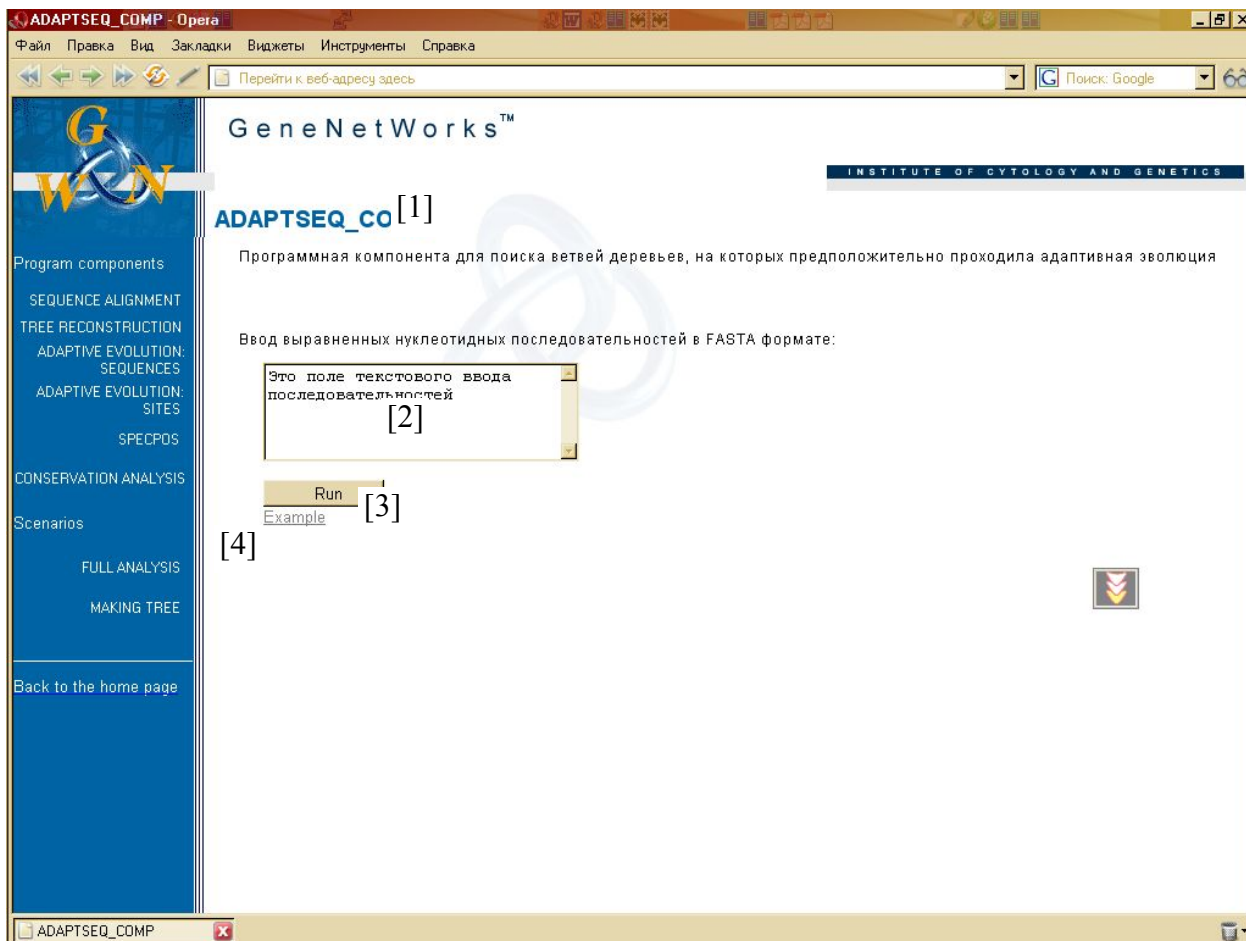


Рисунок 9. Стартовая страница для выполнения операции «Поиск ветвей деревьев, на которых предположительно проходила адаптивная эволюция».

Основные действия в требуемой последовательности

Для того чтобы ввести входные данные для анализа необходимо скопировать их в буфер обмена данными операционной системы в любом текстовом редакторе. После этого необходимо вернуться к странице ввода данных компоненты выравнивания последовательностей и поместить указатель мыши в поле текстового ввода [2] (Рис. 9). После этого необходимо скопировать данные из буфера в поле текстового ввода. Это делается нажатием комбинации клавиш <Ctrl-Ins> или <Ctrl-V>. После этого данные, помещенные в поле [2] (Рис. 9), отображаются в нем в текстовом виде.

После ввода данных в поле [2] (Рис. 9) для запуска программы необходимо нажать указателем мыши на кнопку «Run» [3] (Рис. 9). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

Для запуска контрольного примера необходимо перейти по гиперссылке “Example” [4] (Рис. 9)

Заключительные действия

После выполнения программы в левой панели появляется окно результата. Данная страница содержит гиперссылки для отображения результатов расчета.

При обращении к гиперссылке «Ka/Ks table in text format» [1] программа запросит сохранение файла матрицы попарных сравнений, который может быть открыт при помощи MS Excel. При обращении к гиперссылке «Kh/Ks table in text format» [1] программа запросит сохранение файла матрицы попарных сравнений, который может быть открыт при помощи MS Excel.

2.2.5. Программная компонента CONS_COMP: оценка консервативности позиций множественного выравнивания аминокислотных последовательностей.

Позиции аминокислотной последовательности, аминокислоты в которых слабо варьируют в ходе эволюции (т.е. являются консервативными), как правило, участвуют в выполнении важных белковых функций или играют важную роль в структуре белка. Поэтому их выявление является важной частью сравнительного анализа аминокислотных последовательностей.

В пакете реализован алгоритм выявления консервативных позиций, основанный на расчете энтропии столбцов множественного выравнивания (Durbin R. et al, 1998).

Для запуска программной компоненты для множественного выравнивания последовательностей генетических макромолекул необходимо в левой панели стартовой страницы выбрать гиперссылку «Conservation analysis». После этого браузер загрузит стартовую страницу программной компоненты для множественного выравнивания последовательностей (Рис. 10). Панель ввода данных располагается справа и содержит название программной компоненты [1]. Панель содержит поле ввода множественного выравнивания [2], поле ввода структурного или функционального профиля белка [3], кнопку запуска счета [4], а так же гиперссылку для перехода на страницу контрольных данных [5].

2.2.5.1. Описание работы CONS_COMP

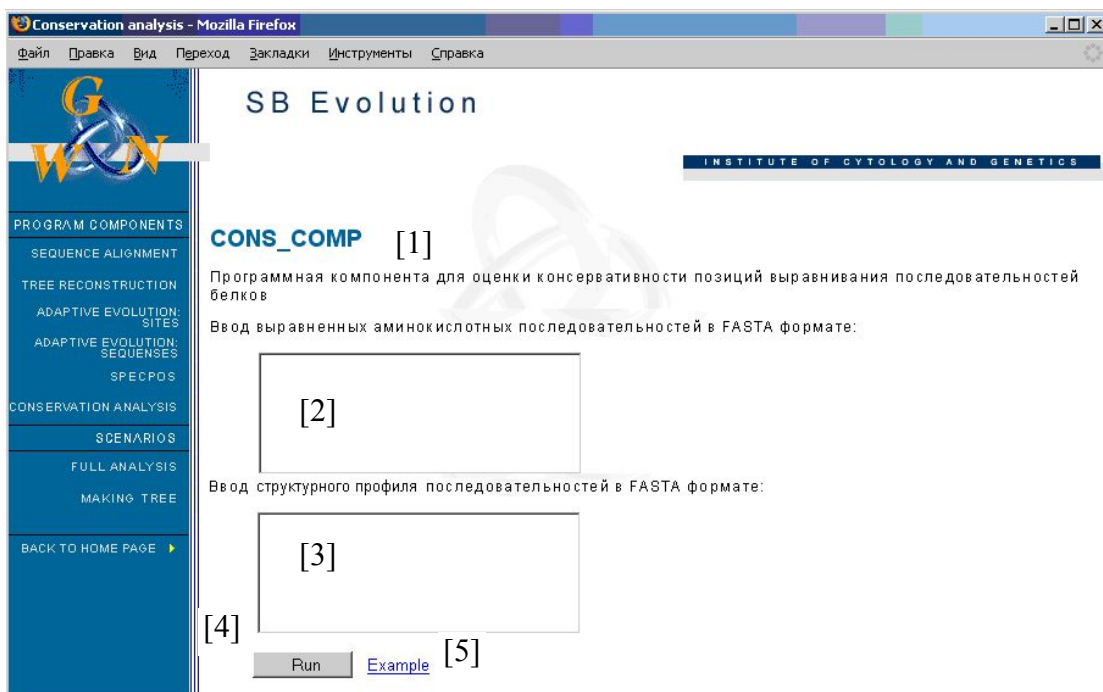


Рисунок 10. Стартовая страница для выполнения операции «Анализ консервативности позиций множественного выравнивания последовательностей».

Основные действия в требуемой последовательности

Для того чтобы ввести входные данные для анализа необходимо скопировать их в буфер обмена данными операционной системы в любом текстовом редакторе. После этого необходимо вернуться к странице ввода данных компоненты выравнивания последовательностей и поместить указатель мыши в поле текстового ввода [2] (Рис. 10). После этого необходимо скопировать данные из буфера в поле текстового ввода. Это делается нажатием комбинации клавиш <Ctrl-Ins> или <Ctrl-V>. После этого данные, помещенные в поле [2] (Рис. 10), отображаются в нем в текстовом виде. Аналогично вводится текстовый профиль структурных или функциональных свойств остатков в белках в поле [3] (Рис. 8), однако ввод структурного профиля не является обязательным.

После ввода данных в поля [2] и [3] (Рис. 10) для запуска программы необходимо нажать указателем мыши на кнопку «Run» [4] (Рис. 10). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

Для запуска контрольного примера необходимо перейти по гиперссылке “Example” [5] (Рис. 10)

После выполнения программы в левой панели появляется окно результата (Рис. 11).



Рисунок 11. Страница результата после выполнения операции «Анализ консервативности позиций множественного выравнивания последовательностей».

Данная страница содержит гиперссылки для отображения результатов расчета.

При обращении к гиперссылке «Conservation profile in html format» [1] (Рис. 11) программа запустит новое окно браузера, в котором в HTML формате будет представлен профиль консервативности множественного выравнивания последовательностей (Рис. 12).

CONS Positional data.

MULTIPLE ALIGNMENT OF "BACTERIOPHAGE T4 LYSOZYME (E.C.3.2.1.17) MUTANT" AND HOMOLOGOUS SEQUENCES.
NUMBER OF SEQUENCES: 361.
ALIGNMENT LENGTH: 162.

POSITIONAL INFORMATION:
1. Alignment position number.
2. Amino acid in selected sequence.
3. Residue's functional class [119L_0 secondary structure by DSSP program].
4. Number of gaps at this position.
5. Number of different amino acids (variability).
6. Entropy.

1	2	3	4	5	6
1	M-0	-	42	3	0.750
2	N-1	-	29	5	0.806
3	I-2	H	28	9	0.847
4	F-3	H	8	8	0.909
5	E-4	H	7	9	0.877
6	M-5	H	4	7	0.772
7	L-6	H	3	5	0.664
8	R-7	H	3	8	0.774
9	I-8	H	2	14	0.987
10	D-9	H	2	8	0.750
11	E-10	H	2	3	0.397
12	G-11	-	2	3	0.407
13	L-12	-	2	11	0.757
14	R-13	E	2	12	0.724
15	L-14	E	1	11	0.576
16	K-15	E	0	12	0.874
17	T-16	F	0	10	0.913

Рисунок 12. Страница визуализации результата после выполнения операции «Анализ консервативности позиций множественного выравнивания последовательностей».

2.2.6. Программная компонента SPECPOS: поиск позиций, определяющих специфичность функции белков.

Одной из задач при анализе набора гомологичных белковых последовательностей является поиск аминокислотных замен, которые фиксировались специфическим образом по отношению к определенной группе белков. Для определения таких позиций в системе предложен подход, основанный на алгоритме Калининой и соавторами (Kalinina, 2004). Данный метод позволяет решать задачу следующего типа. Пусть имеется множественное выравнивание белковых последовательностей, разбитых на несколько функциональных групп. Это могут гомологичные ферменты, имеющие различную специфичность по отношению к субстрату, транскрипционные факторы, распознающие различные варианты регуляторных сайтов ДНК, или гомологичные белки, формирующие несколько паралогических групп. В последовательностях этих белков необходимо определить позиции, в которых фиксация замен происходит специфически, по отношению к определенной группе последовательностей. Это означает, что в таких позициях в определенной группе белков встречаются преимущественно одна аминокислота, но при сравнении белков из разных групп, аминокислоты различаются. Исследования показывают, что, как правило, такие аминокислотные остатки модулируют специфичность белка по отношению к субстрату, а в пространственной структуре белка располагаются вблизи активных центров, участков связывания с ДНК или других белков.

Для оценки значимости позиции в данной программе используется значение коэффициента взаимной информации (mutual information) в позициях, как мера связи между типом аминокислоты в данной позиции и паралогичной группой, вычисляемое по формуле:

$$I_p = \sum_{i=1}^N \sum_{\alpha=1}^{20} f_p(\alpha, i) \log \frac{f_p(\alpha, i)}{f_p(\alpha) f(i)}, \quad (2)$$

где $f(\alpha)$ – частота аминокислоты типа α в позиции p множественного выравнивания, $f(i)$ – доля белков, принадлежащих к паралогичной группе i , $f(\alpha, i)$ – частота аминокислоты типа α в позиции p в белках паралогичной группы i . Оценка значимости коэффициента I_p проводится по методу Монте-Карло. Нумерация групп последовательностей для белков изменяется с помощью случайной перестановки большое количество раз. После каждой перестановки для позиции рассчитывается значение I_p . Распределение этой величины в выборках со случайным присваиванием группировки последовательностей аппроксимируется Гауссовым распределением, для которого рассчитывается среднее и стандартное отклонение. Если значение I_p в реальной выборке белков превосходит среднее в случайных выборках на величину более 4 стандартных отклонений, то зависимость типа аминокислоты в позиции от группировки последовательностей считается значимым.

2.2.6.1. Описание работы SPECPOS

Для запуска программной компоненты для определения специфических позиций аминокислотных последовательностей необходимо в левой панели стартовой страницы выбрать гиперссылку «Specific position». После этого браузер загрузит стартовую страницу программной компоненты для определения специфических позиций последовательностей (Рис. 13).

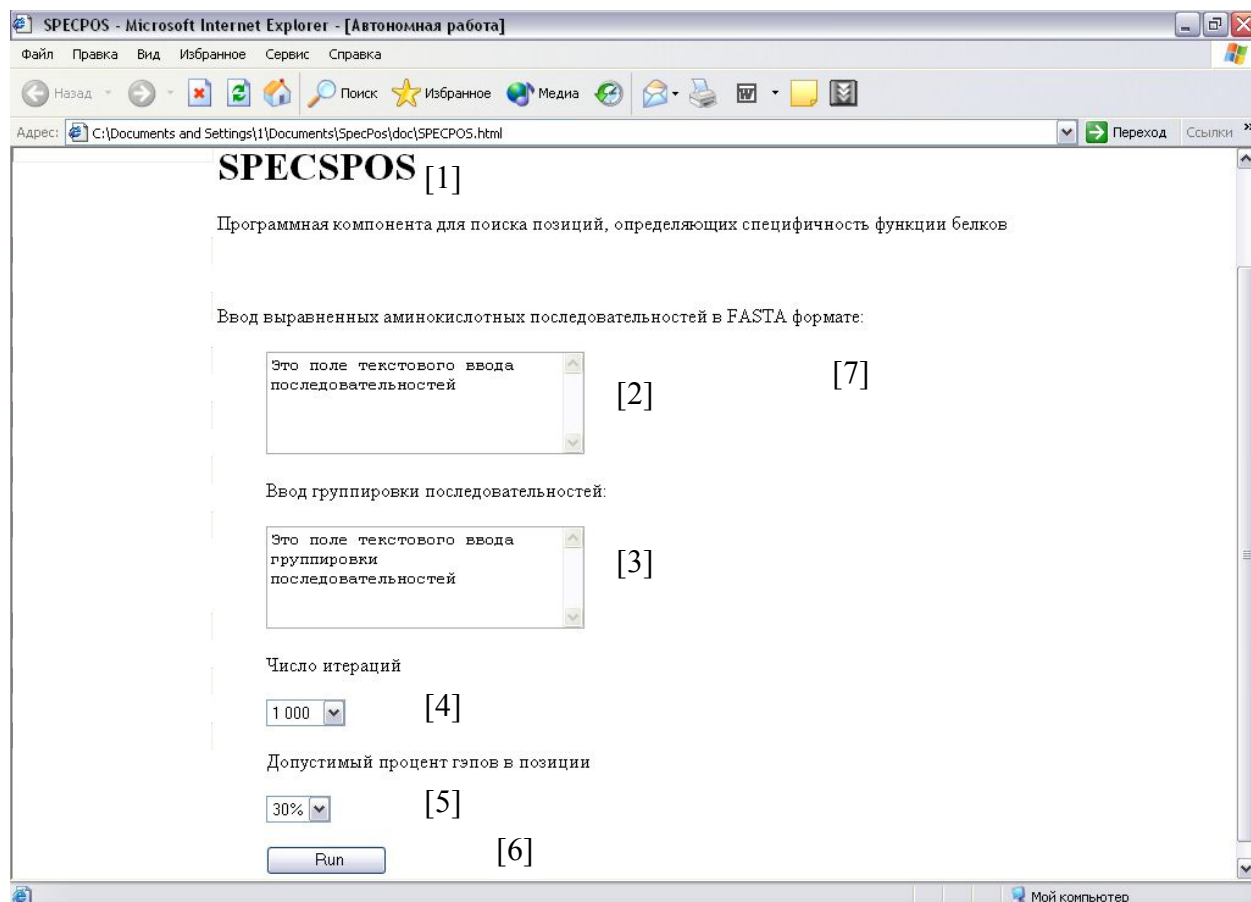


Рисунок 13. Стартовая страница для выполнения операции «Определение специфических позиций».

Панель ввода данных располагается справа и содержит название программной компоненты [1]. Панель содержит форму ввода данных и запуска счета.

Входные данные.

Входными данными для программной компоненты являются множественное выравнивание белковых последовательностей в FASTA формате и указания по разделению данной выборки на группы.

Также обязательным условием работы программы являются указания по разделению множественного выравнивания на группы. Данное описание должно начинаться с номера группы. После номера группы должна стоять точка и название группы. Следующей строкой является либо номера последовательностей, либо полные названия последовательностей в FASTA формате (начинается с символа ">"), принадлежащих данной группе.

Пользователь может определить число итераций (между 1000 и 10000) при подсчете статистической значимости данной позиции в выравнивании. Этот параметр влияет на время работы программы. Малое количество циклов позволит получить результат быстрее, но снизит достоверность данного результата. Последний параметр определяет допустимое процентное содержание гэпов в позиции множественного выравнивания. Обычно, он не превышает 30%, но в некоторых случаях требуется более высокое значение данного параметра. Однако при большом количестве гэпов программа может выдать некорректные результаты.

Основные действия в требуемой последовательности.

Для начала работы программы необходимо скопировать множественное выравнивание последовательностей в FASTA формате в окно [2] (Рис. 13) стартовой страницы и ввести данные по разделению выборки на группы в окно [3] (Рис. 13). Также выбрать количество итераций [4] (Рис. 2) и процентное содержание гэпов в позиции [5] (Рис. 13). По умолчанию значения данных параметров 1000 и 30, соответственно.

После ввода данных (Рис. 13) для запуска программы необходимо нажать указателем мыши на кнопку «Run» [3] (Рис. 13). После этого сервер системы АСНИ «СИСТЕМНАЯ БИОЛОГИЯ – 01» запустит задачу на счет.

После выполнения программы в левой панели появляется окно результата (Рис. 14).

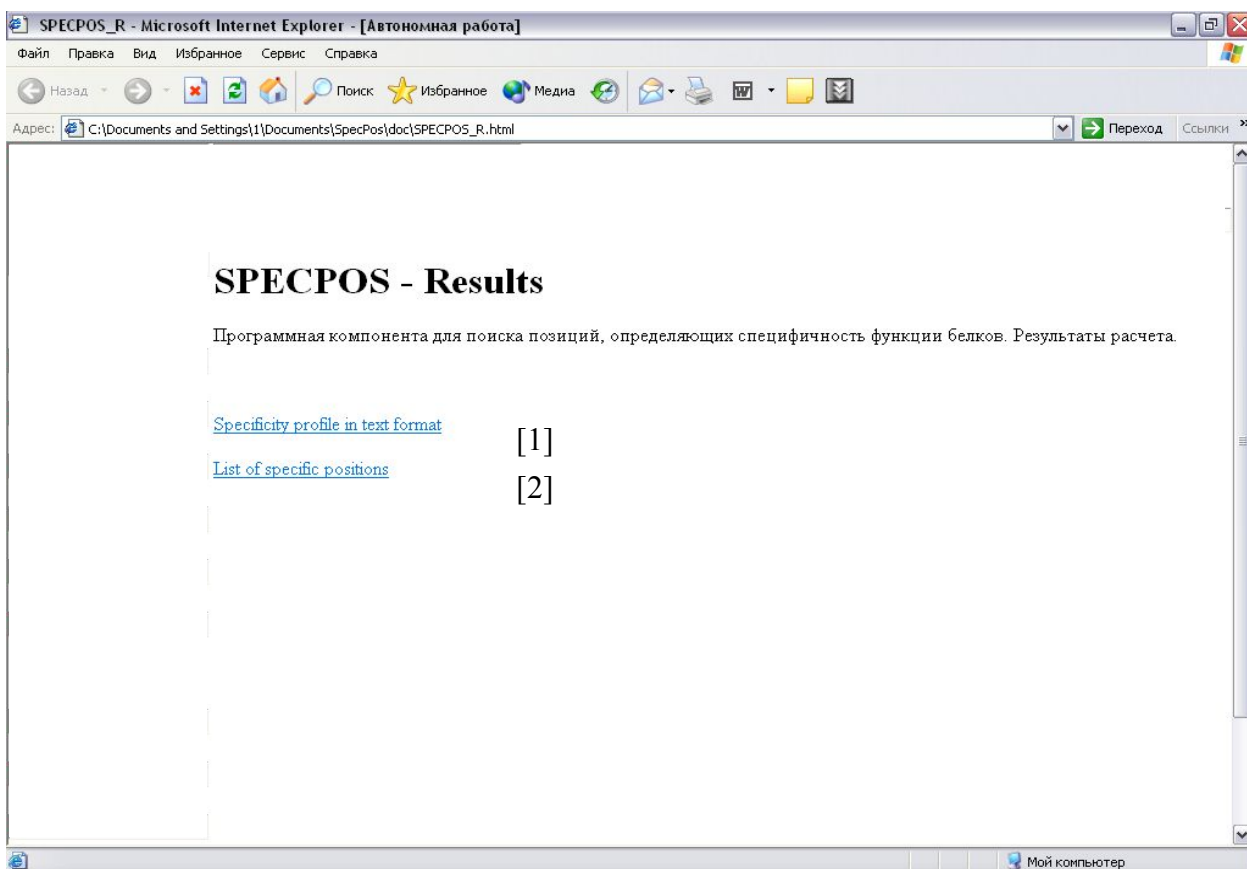


Рисунок 14. Страница результата после выполнения операции «Определение специфических позиций».

Данная страница содержит две гиперссылки для различного способа отображения результатов расчета.

При обращении к гиперссылке «Specificity profile in text format» [1] (Рис. 14) программа запустит новое окно браузера, в котором в текстовом формате будет представлено множественное выравнивание последовательностей в FASTA формате, разделенное на группы (Рис. 15). На этом выравнивании помечены позиции, определяющие специфическую функцию белка.

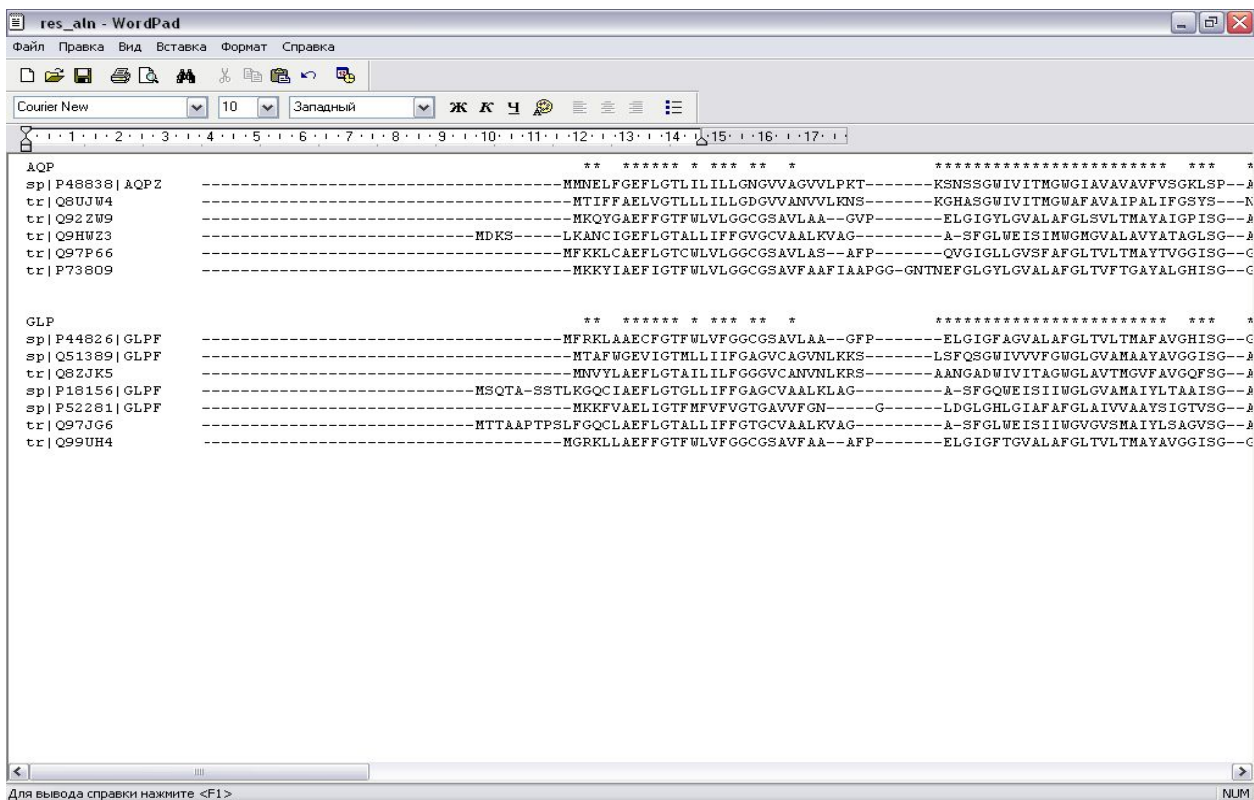


Рисунок 15. Результат работы программы SPECPOS – обозначение специфических позиций в множественном выравнивании.

При обращении к гиперссылке «List of specific position» [2] (Рис. 16) программа запустит новое окно браузера, в котором будет представлен список специфических позиций, а также значения коэффициента взаимной информации и значения Z-score (Рис. 16).

№	Alignment position	Mutual information	Z-score
1	60	2.264	2.264
2	335	2.071	2.071
3	364	2.860	2.860
4	83	2.034	2.034
5	238	2.628	2.628
6	331	2.503	2.503
7	319	2.310	2.310
8	325	2.182	2.182
9	311	2.635	2.635
10	206	2.502	2.502
11	315	2.403	2.403
12	54	2.057	2.057
13	62	1.741	1.741
14	318	2.106	2.106
15	208	2.504	2.504
16	310	1.906	1.906
17	361	1.870	1.870
18	203	1.834	1.834
19	58	2.017	2.017
20	317	1.954	1.954
21	324	2.323	2.323
22	52	2.710	2.710
23	88	3.809	3.809
24	165	2.357	2.357
25	87	1.903	1.903
26	220	1.687	1.687
27	61	1.652	1.652
28	234	1.791	1.791
29	50	2.848	2.848
30	339	1.813	1.813
31	79	1.870	1.870
32	194	1.599	1.599
33	39	1.549	1.549
34	49	1.875	1.875
35	207	1.933	1.933

Рисунок 16. Результат работы программы SPECPOS – список значимых специфических позиций.

3. Полезные ссылки.

1. Edgar R.C. Muscle: multiple sequence alignment with high accuracy and high throughput. // Nucleic Acids Research. 2004. V. 32. P. 1792-1797.
2. Pupko T., Bell R.E., Mayrose I., Glaser F., Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. // Bioinformatics. 2002. V. 18 P. S71-S77.
3. Jukes T.H. Silent nucleotide substitutions and the molecular evolutionary clock. // Science. 1980. V. 210. P. 973-978.
4. Durbin R., Eddy S.R., Krogh A., Mitchison G. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids //Cambridge University Press. 1998.
5. Tang H, Wu CI. A new method for estimating nonsynonymous substitutions and its applications to detecting positive selection. // Mol Biol Evol. 2006. V. 23. P. 372-379.
6. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. // Mol Biol Evol. 1987. V. 4 P. 406-425.
7. Kalinina O.V., Mironov A.A., Gelfand M.S., Rakhmaninova A.B. Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. // Protein Science. 2004. V. 13. P. 443–456.