

ISSN 0042-8809

©Коллектив авторов

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В ЦИТОГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

В.В.Руанет¹, А.К.Хетагурова¹, Е.Д.Бадаева²

¹Медицинский колледж РАМН,
Москва, 117105, Варшавское шоссе, 16А,
факс: (095) 952-55-81; эл.адрес: ruanet@mail.ru

²Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва.

Возможность решения многих сложных фундаментальных и прикладных задач биологии связана с развитием подходов к обработке и интерпретации информации, полученной в экспериментах. Цитогенетический анализ играет важную роль в генетических, молекулярно-биологических и филогенетических исследованиях, но формализация и интерпретация результатов цитогенетических исследований является сложной задачей. Это связано, во-первых, с процессом формализации графической информации и, во-вторых, с трудностями выбора конкретных программ и параметров их работы при анализе уже формализованных данных. В то же время, имеется новая информационная технология, ориентированная на анализ сложных не линейных задач, в частности, на работу с образной информацией - это искусственные нейронные сети (ИНС). Целью работы являлось доказать возможность использования ИНС для решения задач цитогенетики.

Ключевые слова: геном, идентификация, идиограмма, искусственные нейронные сети (ИНС), многослойный персептрон, сеть Кохонена

ВВЕДЕНИЕ. Возможность решения многих сложных фундаментальных и прикладных задач биологии связана с развитием подходов к обработке и интерпретации информации, полученной в экспериментах. Работа с данными медико-биологического профиля имеет свои особенности. Большинство данных в этой области знаний имеют описательный характер и выражаются с помощью формализмов, оценка которых нередко субъективна. Например, значительные сложности возникают при попытке формализации данных цитогенетических исследований.

Цитогенетический анализ играет важную роль в генетических, молекулярно-биологических и филогенетических исследованиях. Хромосомные болезни человека, с которыми связан ряд тяжелых патологий, также идентифицируются с помощью цитогенетических методов. Несмотря на то, что в результате дифференциального окрашивания и применения разных типов маркеров при гибридизации хромосом *in situ* мы получаем специфическую для каждого вида картину, расшифровка, т.е. определение видовой принадлежности или определенной патологии по количеству и расположению блоков, их интенсивности и другим параметрам, является сложной задачей. Это связано, во-первых, с процессом формализации графической информации и, во-вторых, с трудностями выбора конкретных программ и параметров их работы при анализе уже формализованных данных.

В то же время, имеется новая информационная технология, ориентированная на анализ сложных нелинейных задач, в частности, на работу с образной информацией - это искусственные нейронные сети (ИНС).

Искусственная нейронная сеть состоит из большого числа простых, параллельно работающих процессорных элементов - нейронов, соединенных адаптивными линиями передачи информации - синапсами, т.е. имеет структуру подобную биологическим нейронным сетям (нервной системы). Ниже представлена схема (рис. 1) стандартного

формального нейрона (далее по тексту нейрон), состоящего из входного сумматора, нелинейного преобразователя и точки ветвления на выходе

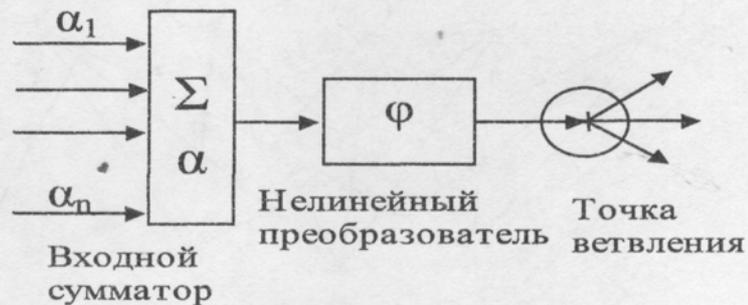


Рисунок 1
Схема стандартного формального нейрона

Подавая любые числа на входы сети, мы получаем какой-то набор чисел на выходах сети. Таким образом, работа нейросети состоит в преобразовании входного вектора в выходной вектор.

Принципиальное отличие нейросетевых технологий от традиционных способов обработки информации состоит: 1) в замене строго алгоритмированного пошагового анализа данных на параллельную обработку всего массива информации; 2) в замене программирования на обучение. Искусственные нейронные сети применяются для идентификации и классификации информации в случае ограниченных, неполных и нелинейных данных. Они отличаются универсальностью: одна и та же программа обеспечивает возможность работы в разных областях знаний.

В настоящее время разработано несколько видов искусственных нейронных сетей. Наиболее популярными из них являются многослойный перцептрон (МСП) и сеть Кохонена.

На рисунке 2 представлена схема строения МСП: Входной слой (слой 1) - группа связей, по которым ИНС получает информацию из внешнего мира. Промежуточный слой (слой 2) - группа связей, которая обеспечивает возможность моделирования нелинейных функций. Выходной слой (слой k) - группа выходных связей, с которых снимаются выдаваемые сетью сигналы. Сеть подобной структуры может моделировать функцию практически любой степени сложности. Число слоев и число элементов в каждом слое определяется сложностью функции и подбирается эмпирически в ходе решения конкретной задачи [1].

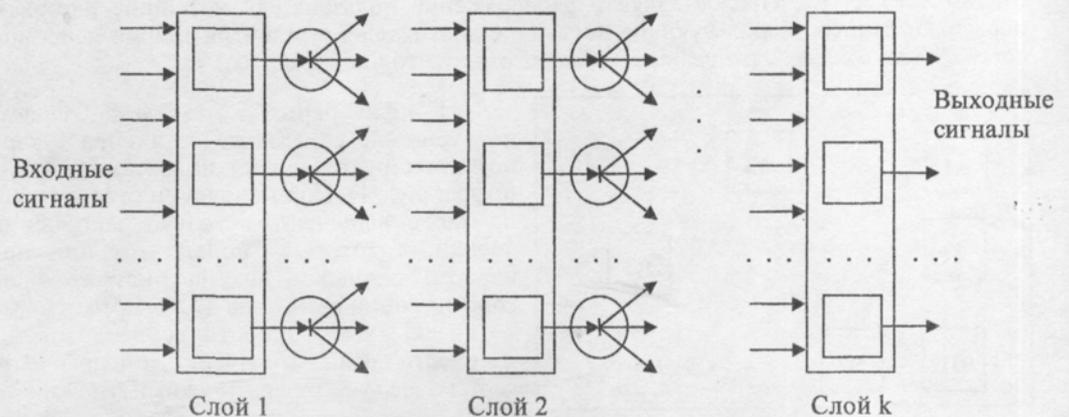


Рисунок 2.
Схема многослойного перцептрона.

На рисунке 3 представлена схема строения самоорганизующейся карты признаков - сети Кохонена: а - входные элементы (предназначены только для того, чтобы распределять входные данные между выходными элементами); б - выходные (кластерные) элементы выступают в роли карты признаков пространства входных данных.

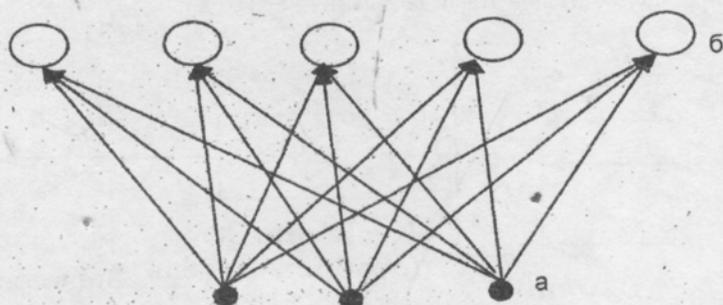


Рисунок 3.
Схема сети Кохонена

Сеть Кохонена, обучаясь на входных данных, способна построить двумерное отображение многомерной информации с минимально возможными искажениями. Такие самоорганизующиеся карты сохраняют локальную топологию данных - близость на карте подразумевает и близость в исходном пространстве. Сеть Кохонена может распознавать кластеры в данных, а также устанавливать близость классов, т.е. реализует одно из свойств ИНС - обобщения по подобию и одно из возможных применений таких сетей - разведочный анализ данных [2].

Характер обучения зависит от выбранного типа сети. Для работы с МСП используется алгоритм "обучение с учителем". При этом способе обучения готовится набор данных, представляющих собой ряд наблюдений, для которых указаны значения входных и выходных переменных (условия задачи → ответ).

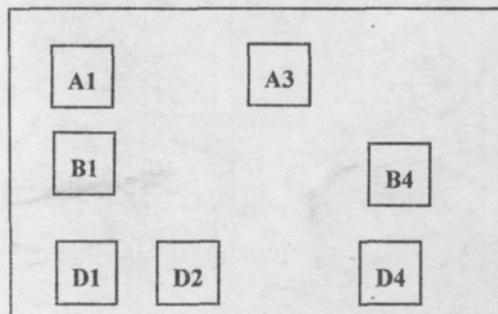
При работе с сетью Кохонена используется алгоритм "обучение без учителя". Процесс обучения, при котором на вход нейронной сети подаются данные, содержащие только значения входных переменных. Он предназначен для нахождения кластеров во входных данных.

Критерием достижения цели (обучение сети) считается результат тестирования набором примеров с известными ответами, не входящими в обучающую выборку.

Целью работы - доказать возможность использования ИНС для решения задач цитогенетики.

МЕТОДИКА. В работе были использованы следующие программные продукты, эмулирующие искусственные нейронные сети:

- Пакет программных продуктов "*Excel Neural Package*" расширяющий функциональные возможности широко распространенного средства работы с данными Microsoft Excel 97. Предоставляя в распоряжение пользователя новейшие алгоритмы обработки данных, использующие последние достижения в развитии теории нейронных сетей, "*Excel Neural Package*" представляет ответ сети в виде карты.



Конфигурация карты задается пользователем. Каждая ячейка карты позиционируется двумя индексами (буква + цифра, рис. 4). При получении ответа от сети на карте выделяются только заполненные ячейки, в которые "попал" тот или иной кластер данных. Так, на рисунке 4 при конфигурации сети "4 на 4" на карте отмечено 7 ячеек, хотя теоретически возможное количество равно 16. Ячейки топологической карты обозначаются латинскими буквами [1].

Рисунок 4.
Топологическая карта с конфигурацией матрицы "4 на 4"

- Программа "*NeuroPro 0.25*" (Царегородцев В.Г., ИВМ СО РАН). Данный программный продукт представляет собой менеджер обучаемых искусственных нейронных сетей, в данной программе реализованы только сети слоистой архитектуры (МСП), применение программного продукта возможно для решения любой задачи классификации или прогноза, которая решается при наличии выборки данных.

Выбор типа сети зависит от характера поставленной задачи. При решении задач прогнозирования и классификации чаще всего используются многослойные перцептроны; при решении задач категоризации данных чаще всего применяется сеть Кохонена [1].

Материалом для составления обучающих баз послужили идиограммы хромосом диплоидных видов *Aegilops*, содержащая информацию о расположении сайтов гибридизации 5S (круглые метки) и 18-26S (прямоугольные метки) рРНК генов - рисунок 5 (*Ae. speltoides* (S), *Ae. longissima* (S^l), *Ae. sharonensis* (S^{sh}), *Ae. searsii* (S^s), *Ae. bicornis* (S^b), *Ae. mutica* (M^t), *Ae. umbellulata* (U), *Ae. caudata* (C), *Ae. uniaristata* (Uⁿ), *Ae. comosa* (M), *Ae. heldreichii* (M^h), *Ae. squarrosa* (D).), всего двенадцать идиограмм.

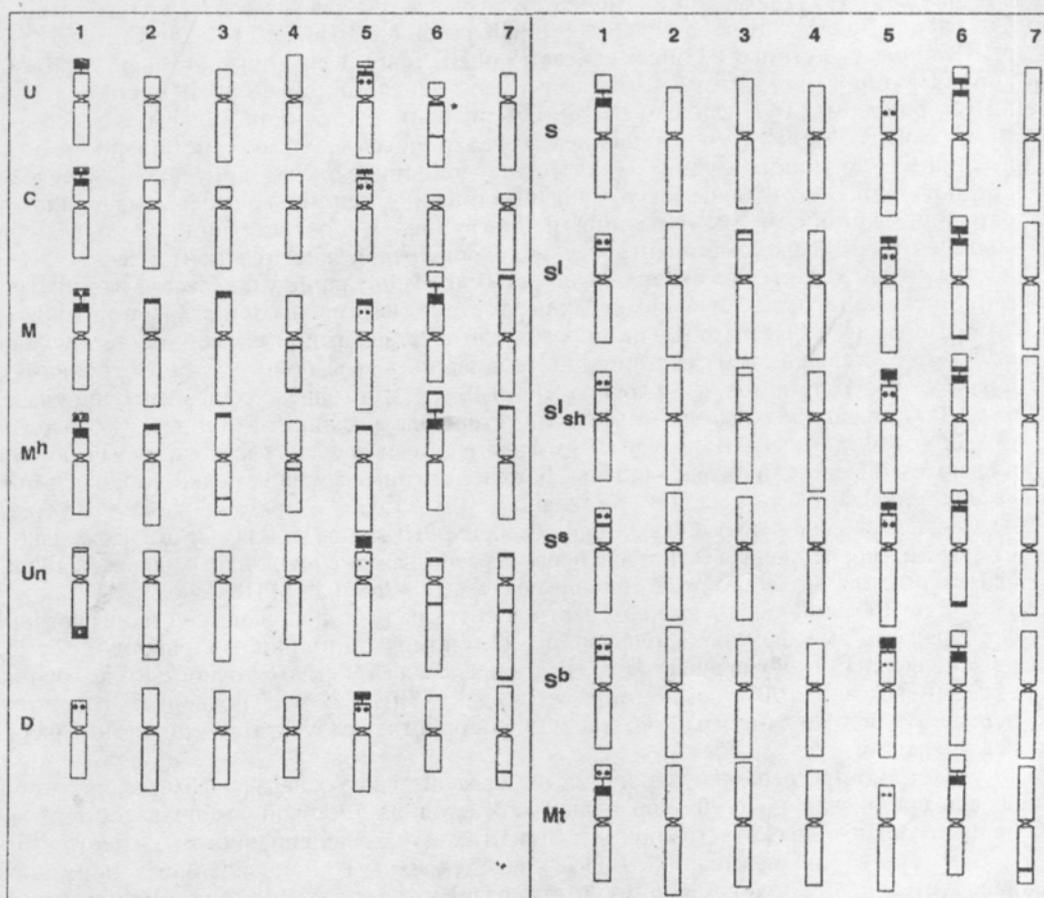


Рисунок 5

Идиограммы хромосом диплоидных видов *Aegilops*, содержащие информацию о расположении сайтов гибридизации 5S (круглые метки) и 18S - 26S (прямоугольные метки) рРНК-генов

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. Для оценки возможности создания экспертных систем на базе нейросетевых технологий для решения задач цитогенетики были выбраны представители рода *Aegilops*. Они имеют крупные хромосомы, и их филогения хорошо изучена. На первом этапе работы была поставлена задача определения групп геномов диплоидных видов *Aegilops* с помощью маркеров 18S-26S и 5S рРНК-генов с использованием искусственных нейронных сетей. Для решения данной задачи был использован многослойный перцептрон.

В качестве модельных объектов для поиска возможного подхода к решению этой задачи были взяты результаты исследования диплоидных видов *Aegilops*: *Ae. speltoides* (S), *Ae. longissima* (S^l), *Ae. sharonensis* (S^{sh}), *Ae. searsii* (S^s), *Ae. bicornis* (S^b), *Ae. mutica* (M^t), *Ae. umbellulata* (U), *Ae. caudata* (C), *Ae. uniaristata* (Un), *Ae. comosa* (M), *Ae. heldreichii* (M^h), *Ae. squarrosa* (D). Преимуществом данных объектов является то, что геномы этих видов были подробно охарактеризованы с помощью цитогенетических методов, и для каждого

был найден ряд маркеров, которые можно использовать для их характеристики. С точки зрения специалистов, одним из наиболее значимых параметров при классификации геномов эгилопсов является относительное расположение кластеров рибосомных генов (18S-26S и 5S рРНК-генов) на спутничных (SAT) хромосомах, определяющее их принадлежность к определенным гомеологичным группам [3].

Для составления обучающей базы были взяты обобщенные идиограммы хромосом диплоидных видов *Aegilops*, на которых приведено распределение локусов 5S и 18S-26S рРНК-генов. Обучающая база формировалась согласно разработанной методике [4].

Для характеристики генома мы использовали такие параметры, как общее число спутничных (SAT) хромосом в геноме (1 или 2), а затем для каждой из SAT хромосом: 1) Суммарное число сайтов 5S + 18S-26S рРНК-генов в спутнике; 2) Число сайтов 18S-26S рРНК-генов в спутнике; 3) Число сайтов 5S рРНК-генов в спутнике; 4) Число сайтов 18S-26S рРНК-генов в коротком плече хромосомы; 5) Число сайтов 5S рРНК-генов в коротком плече хромосомы; 6) Число сайтов 18S-26S рРНК-генов в длинном плече хромосомы; 7) Число сайтов 5S рРНК-генов в длинном плече хромосомы. Характеристики хромосом, не имеющих спутники, брали в качестве дополнительных критериев. Поскольку предполагается, что у пользователя нет информации о геноме, во избежание ошибок при описании хромосом их объединяли в группы в соответствии с морфологией (метацентрические, субметацентрические, акроцентрические, телоцентрические), и при последующей характеристике проводили суммирование данных по каждой из этих групп [5]. Конечные характеристики входных параметров системы представлены в таблице 1.

Проверку качества обучения проводили по стандартной схеме - на тестирующей выборке, составленной из образцов, не входивших в обучающую базу. Обученная сеть подразделяла 12 геномов на 10 групп (геномы S' , S'_{sh} и S^s , определялись как один класс).

Из литературных данных известно, что три вида, несущих эти геномы (*Ae. longissima* (S'), *Ae. sharonensis* (S'_{sh}) и *Ae. searsii* (S^s)) филогенетически очень близки, и многие ботаники в прошлом определяли их как один вид. Даже в настоящее время нет единой точки зрения о таксономическом статусе *Ae. sharonensis* [3]. Таким образом, сходство геномов перечисленных видов по расположению семейств рРНК-генов нельзя считать неожиданным. Следовательно, нейронная сеть точно провела распознавание диплоидных видов *Aegilops* на основании относительного расположения локусов 18S-26S и 5S рРНК-генов.

Тестирование полученной экспертной системы (ЭС) проводили на зашифрованных фотографиях метафазных пластинок D-геномов диплоидных видов *Aegilops*, представленных в обучающей базе. ИНС определяла геномы с точностью до группы с эффективностью 100%, а до вида - порядка 70-75%, что позволяет говорить о перспективности данного подхода при обработке результатов цитогенетических исследований.

Обычно при решении различных неформализованных задач в разных проблемных областях точность в 70-90% правильных ответов на тестовой выборке соответствует проценту правильных ответов при решении этих же задач специалистом-экспертом [6].

Вторым, и возможно, более перспективным направлением применения нейросетевых технологий может стать разработка подходов для оценки степени генетического родства между видами.

Для решения этой задачи мы воспользовались другим типом ИНС - сетью Кохонена. Сеть Кохонена может распознавать кластеры в данных, а также устанавливать близость классов, в данном случае, на основании относительного расположения локусов 18S-26S и 5S рРНК-генов на SAT хромосомах.

При бинарном шаге ветвления (на каждом следующем шаге ветвления анализируемый кластер делится на две части) сеть Кохонена разделила диплоидные виды *Aegilops* на 10 кластеров (рис. 6). Римскими цифрами на рисунке обозначен порядок ветвления. Поскольку в геномах некоторых видов были выявлены не только постоянные, но и полиморфные локусы 18S-26S рРНК-генов (рис. 5), то этот фактор следовало учитывать при составлении базы данных. Вариабельные геномы характеризовали двумя образцами: один с минимальным числом локусов (только постоянные), а второй - с максимальным (постоянные + все возможные полиморфные локусы).

Окончание ветвления определяется следующим образом. Если на следующем шаге сеть делила образцы на кластеры, но на топологической карте они представлены как единое целое, то это шаг считается моментом прекращения ветвления. В качестве примера можно рассмотреть деление кластера, содержащего M^a , M , M^b , M^c геномы на IV - V шаге ветвления. На IV шаге кластер, согласно заданному шагу ветвления, разделился на

Таблица 1. Характеристики входных и выходных полей обучающей базы данных для многослойного персептрона

| ВХОДНЫЕ ПОЛЯ | | | | | | | | | | | |
|-------------------------|-------|----------------|-------|----|-------|---------------|---------|-------|-------|-------|--|
| Спутничные хромосомы | | | | | | | | | | | |
| морфология спутника | | | | | | | | | | | |
| морфология хромосомы | | | | | | | | | | | |
| Тип хромосомы | Число | короткое плечо | | | | длинное плечо | | | | | |
| 6 | ГХ | маркеры | | | | ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 18-26S | число | 5S | число | | |
| | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| морфология спутника | | | | | | | | | | | |
| морфология хромосомы | | | | | | | | | | | |
| Тип хромосомы | Число | короткое плечо | | | | длинное плечо | | | | | |
| 22 | ГХ | маркеры | | | | к-во ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 28 | 18-26S | число | 5S | число | |
| | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | |
| Хромосомы без спутников | | | | | | | | | | | |
| Число М хромосом | Число | короткое плечо | | | | длинное плечо | | | | | |
| 33 | ГХ | маркеры | | | | ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 18-26S | число | 5S | число | | |
| | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | |
| Число SM хромосом | Число | короткое плечо | | | | длинное плечо | | | | | |
| 44 | ГХ | маркеры | | | | ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 18-26S | число | 5S | число | | |
| | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | |
| Число А хромосом | Число | короткое плечо | | | | длинное плечо | | | | | |
| 55 | ГХ | маркеры | | | | ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 18-26S | число | 5S | число | | |
| | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | |
| Число Т хромосом | Число | короткое плечо | | | | длинное плечо | | | | | |
| 66 | ГХ | маркеры | | | | ГХ | маркеры | | | | |
| | | 18-26S | число | 5S | число | 18-26S | число | 5S | число | | |
| | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | |

| ВЫХОДНОЕ ПОЛЕ | |
|---------------------|-------|
| 77 | |
| Коды выходного поля | |
| Класс | Геном |
| 1 | S |
| 2 | C |
| 3 | U |
| 4 | Mt |
| 5 | Sl |
| 6 | Slsh |
| 7 | Ss |
| 8 | Sb |
| 9 | M |
| 10 | Mh |
| 11 | D |
| 12 | Un |

два кластера. В состав одного вошел M^c геном, в состав другого - M , M , M^b , M^b геномы. При следующем шаге (V) сеть "докладывает" о наличии двух кластеров (M , M и M^b , M^b), что соответствует заданному шагу ветвления. Сеть "обязана" разделить образцы на два кластера, но усредненные значения входных параметров для принадлежащих им примеров настолько близки, что сеть не в силах разграничить области данных для кластеров, как на IV шаге ветвления. Это говорит о генетическом родстве образцов в данной системе параметров.

В качестве альтернативного подхода для проверки генетической близости объектов может быть предложено определение состава наиболее устойчивых кластеров. Он основан на том, что сети предлагается разделить образцы на кластеры с использованием различных конфигураций карты.

Согласно данным литературы, продолжение ветвления до "чистой" классификации (один образец в кластере) нерационально, т.к. не отвечает сути процесса кластеризации [2]. С нашей точки зрения, в проведении ветвления до "чистой" классификации есть своя логика. Увеличивая конфигурацию сети, мы можем определить сходство образцов по тем или иным признакам, определяя устойчивость состава кластера в процессе деления. На рисунке 7 представлена схема деления на кластеры диплоидных видов эгилопсов. В данном случае уже при конфигурации 2x2 сеть выделяет основные кластеры, состав которых, по мнению эксперта, соответствует морфологическим особенностям геномов. Сохранение при девяти различных конфигурациях карты в составе одного кластера - Sl и Slsh геномов, говорит об их генетической близости, что подтверждается литературными данными [3].

На рисунке 8 показаны топологические карты различных по конфигурации сетей Кохонена: 4x4, 5x5, 6x6 (латинскими буквами обозначены соответствующие геномы).

Состав кластеров представленных на топологической карте (рис. 8а) свидетельствует, например, о близости геномов: S^i и S_i^{sh} ; C и U; M и M^b . С другой стороны, с высокой степенью уверенности можно говорить, что родственные связи, например, между образцами, входящими в кластер A_1 и B_1 , больше, чем между образцами, входящими

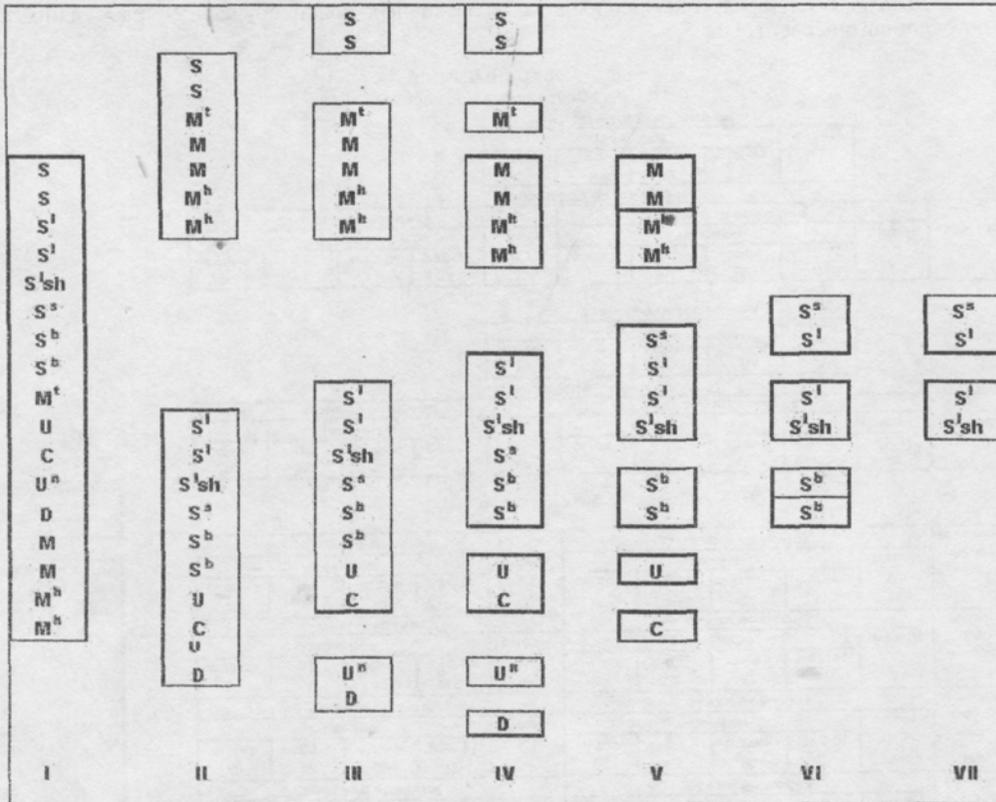


Рисунок 6
 Распределение на кластеры диплоидных видов *Aegilops* сетью Кохонена при бинарном шаге ветвления.

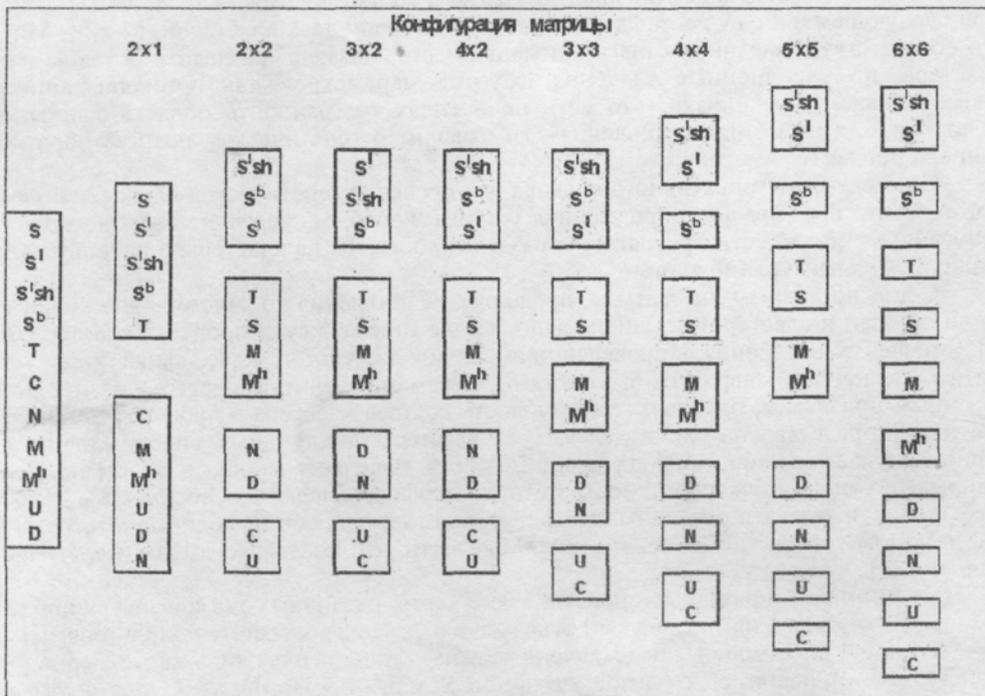


Рисунок 7.
 Схема деления на кластеры диплоидных видов *Aegilops* сетью Кохонена.

в кластеры A_1 и D_4 (рис 8а). Кроме того, рассматривая состав кластеров в процессе увеличения конфигурации матрицы (рис. 8 а,б,в), можно сделать вывод о том, что близость геномов S^i и $S^{i_{sh}}$ выше, чем у M^i и M^h , а их геномы в свою очередь ближе, чем у C и U . Данные выводы базируются на том, что сеть Кохонена сохраняет локальную топологию данных - близость на карте подразумевает и близость в исходном пространстве [2].

Известно, что сеть Кохонена может быть использована для обнаружения новых явлений. Сеть Кохонена распознает кластеры в обучающих данных и относит все данные к тем или иным кластерам. Если после этого сеть встретится с наблюдениями, непохожими ни на один из известных образцов, то она не сможет классифицировать такой набор (отнести представленные образцы к сформированным на стадии обучения кластерам) и тем самым выявит его новизну [2].

Мы решили проверить применимость данных подходов к цитогенетическим исследованиям. Задача классификации решалась следующим образом. Обученной сети Кохонена (рис. 8а) предлагалось определить принадлежность *Ae. searsii* (S^i) к определенному кластеру, т.е. отнести (классифицировать) данный геном к той или иной группе геномов и, таким образом, определить ближайших родственников. Сеть отнесла образец к кластеру A_1 (рис. 8а), что соответствует оценке эксперта.

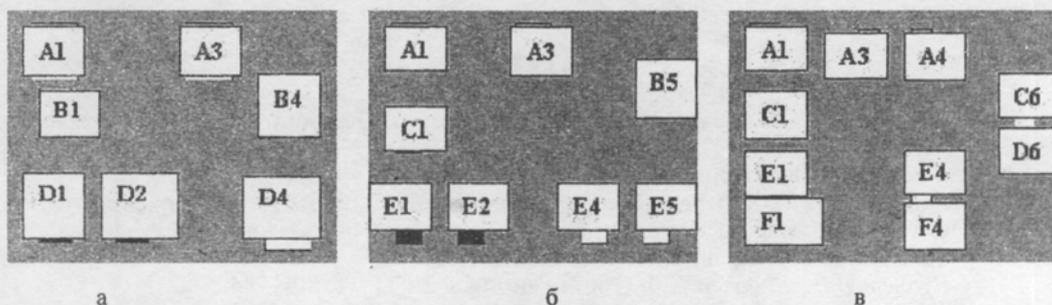


Рисунок 8.

Топологические карты

а - 4 x 4 ($A_1 - S^i, S^{i_{sh}}; A_3 - S, T; B_1 - S^b; B_4 - M, M^h; D_1 - D; D_2 - N; D_4 - C, U$); б - 5 x 5 ($A_1 - S^i, S^{i_{sh}}; A_3 - S, T; B_5 - M, M^h; C_1 - S^b; E_1 - D; E_2 - N; E_4 - C; E_5 - U$); в - 6 x 6 ($A_1 - S^i, S^{i_{sh}}; A_3 - S; A_4 - T; C_1 - S^b; C_6 - M; D_6 - M^h; E_1 - D; E_4 - U; F_1 - N; F_4 - C$).

Задача по обнаружению новых явлений решалась следующим образом. Была составлена обучающая база для сети, в состав которой вошли следующие геномы: $S^i, S^{i_{sh}}, S^b, S, T, M, M^h, C, U$. Была обучена сеть с конфигурацией 3x3. Сеть разделила обучающую базу на четыре кластера. Топологическая карта сети представлена на рисунке 9а. Состав кластеров на топологических картах (а): $A_1 - M, M^h; A_3 - C, U; C_1 - S, T; C_3 - S^i, S^{i_{sh}}, S^b$.

Затем обученной сети было предложено кластеризовать D и N геномы - тестирующая база. Топологическая карта ответа сети представлена на рисунке 9б. На карте появился новый кластер B_3 , в состав которого вошли D, N геномы. Следовательно, опираясь на свойства сети Кохонена, можно говорить о том, что мы имеем дело с новым, для данной сети (системы знаний), явлением. Это соответствует морфологическим различиям геномов, представленных в обучающей (имеют по две SAT хромосомы) и тестирующей (имеют по одной SAT хромосоме) базах [3].

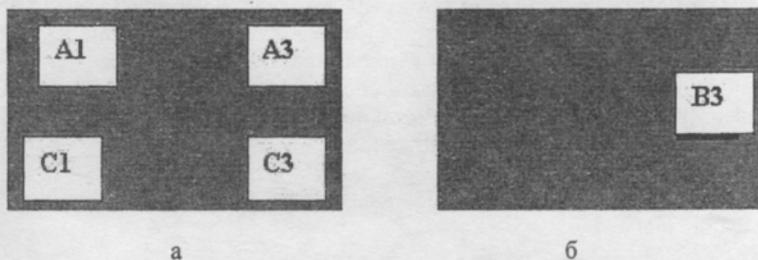


Рисунок 9.

Топологические карты: а) распределение на кластеры образцов обучающей базы; б) распределение на кластеры образцов тестирующей базы

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ. На наш взгляд, представленные результаты свидетельствуют о возможности и целесообразности применения искусственных нейронных сетей (МСП и сети Кохонена) для обработки информации, полученной в ходе цитогенетических исследований.

Для составления обучающей базы наиболее целесообразно использовать обобщенные видовые идиограммы хромосом, поскольку при таком подходе эффективное обучение сети можно проводить на минимальном числе примеров.

Бинарный тип ветвления можно считать наиболее субъективным, поскольку сети задаются строгие параметры кластеризации, и у нее нет "свободы выбора". Тем не менее, этот подход наиболее удобен при построении дендрограмм.

Применение различных типов ветвления и конфигурации карт позволяет более объективно подойти к вопросам оценки схожести образцов в выбранной системе параметров. Используя способ деления с большим числом степеней свободы (большим числом ячеек) для сети, мы повышаем объективность анализа.

Приведенные данные свидетельствуют, что использование сети Кохонена позволяет, на уровне предварительных оценок, определить генетическое родство образцов и оказать помощь в формализации и интерпретации полученной информации.

ЛИТЕРАТУРА

1. *Круглов В.В., Борисов В.В.* (2001) Искусственные нейронные сети М., Горячая линия - Телеком.
2. *Каллан Р.* (2001) Основные концепции нейронных сетей (пер. с англ), Вильямс, М.
3. *Бадаева Е.Д.* (2000) Эволюция геномов пшеницы и их дикорастущих сородичей: цитогенетическое исследование. Дисс. докт. биол. наук, Ин-т молекулярной биологии им. В.А. Энгельгардта РАН, Москва.
4. *Руанет В.В., Бадаева Е.Д.* (2002) Генетика, **38**(11), 1580-1584.
5. *Руанет В.В.* (2003) Использование искусственных нейронных сетей для решения частных задач генетики и селекции. Дисс. канд. биол. наук, Ин-т общей генетики им. Н.И. Вавилова РАН, Москва.
6. *Горбань А.Н., Россиев Д.А.* (1996) Нейронные сети на персональном компьютере. Наука, Новосибирск.

APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR CYTOGENETIC STUDIES

V. V.Ruanet¹, A.K.Khetagurova¹, E. D.Badaeva²

¹ Medical College, Russian Academy of Medical Sciences,
Varshavskoe shosse, 16A, Moscow, Russia, 117105K
fax: (095) 952-55-81; e-mail: ruanet@mail.ru

² Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

Species identification on the basis of cytogenetic results is an important issue of genetics, molecular biology, and phylogenetic studies. The use of various cytogenetic techniques allows obtaining species-specific sites of rRNA- genes regions of chromosome. However, the deciphering of this pattern, i.e., species identification based on the amount, distribution, and staining intensity of sites of rRNA - genes is a difficult problem. Perspectives of artificial neural networks (ANN) for processing and interpretation of the results of cytogenetic analysis are demonstrated.

Key words: genome, identification, idiogram, artificial neural network, multilayer perceptron, Self-Organizing feature map.