ГЕНЕТИКА, 2002, том 38, № 6, с. 1–11

УДК 577.112:595.773.4

СРАВНИТЕЛЬНАЯ ГЕНОМИКА И ПРОТЕОМИКА ДРОЗОФИЛЫ, НЕМАТОДЫ БРЕННЕРА И АРАБИДОПСИСА. ИДЕНТИФИКАЦИЯ ФУНКЦИОНАЛЬНО СХОДНЫХ ГЕНОВ И БЕЛКОВ СИНАПСИСА МЕЙОТИЧЕСКИХ ХРОМОСОМ

© 2002 г. Ю. Ф. Богданов, С. Я. Дадашев, Т. М. Гришаева

Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва 119991;
факс: (095)132-89-62, e-mail: bogdanov@vigg.ru

Поступила в редакцию 06.02.2002 г.

Изложены существующие в литературе принципы компьютерного анализа геномов и белковых наборов у таксономически далеких эукариот. Изложена разработанная авторами стратегия поиска в геномах таких организмов генов и белков, не имеющих гомологии по первичной структуре, но выполняющих одинаковые функции в клетках, делящихся путем мейоза. Стратегия основана на сочетании принципов геномики, протеомики и морфометрического анализа субклеточных структур. Изложены результаты выполненного авторами на основе этой стратегии компьютерного поиска в геномах насекомого *Drosophila melanogaster*, нематоды *Caenorhabditis elegans* и растения *Arabidopsis thaliana* генов, кодирующих белки синаптонемных комплексов. Эти белки функционально аналогичны белкам дрожжей Saccharomyces cerevisiae (белок Zip1p) и млекопитающих (белок SCP1).

В геномных базах данных к концу 2001 г. была собрана и обработана информация о полностью секвенированных геномах трех “модельных” эукариотических организмов и о геноме человека. Первым эукариотическим организмом, чей геном был секвенирован полностью, стали одноклеточные почкующиеся дрожжи Saccharomyces cerevisiae [1], затем наступила очередь нематоды *Caenorhabditis elegans* [2], мухи *Drosophila melanogaster* [3, 4] и “ботанической дрозофилы” – крестоцветного растения *Arabidopsis thaliana* [5]. Эти достижения дали богатый фактический материал для сравнительного анализа геномов и белковых наборов – протеомов. Сформировались специализированные дисциплины геномика и протеомика, обладающие всеми признаками сложившихся наук: собственными задачами и специфическими методами исследования. Сформировались также разделы этих наук – сравнительная геномика и сравнительная протеомика [6, 7]. Последние две дисциплины тесно связаны, ибо информация о полной первичной структуре ДНК генов служит материалом для компьютерного “расчета” первичной структуры белков. Решается и обратная задача: поиск генов по первичной структуре белков. Для этих целей созданы специальные программы, например, BLASTP и др. (см. [6, 7]). Важность сравнения геномов и протеомов у перечисленных выше организмов для общей теории эволюции не требует доказательств.

НЕКОТОРЫЕ ОБЩИЕ ПОЛОЖЕНИЯ СРАВНИТЕЛЬНОЙ ГЕНОМИКИ И ПРОТЕОМИКИ

На первом этапе развития сравнительной геномики и сравнительной протеомики исследователи разумно поставили простую задачу: найти сходные гены и белки у организмов разной эволюционной сложности. Были сформулированы понятия ортологии и паралогии генов и белков.

Ортологами называют гены и белки, выполняющие у разных организмов одинаковые функции и имеющие высокую степень сходства первичной структуры, что позволяет теоретически считать их произошедшими от общего предшественника. При сравнительном исследовании ортологи обычно группируются в кластеры или семейства (генов или белков), имеющие разветвления и образующие филогенетические дендрограммы. Ортологи – это, согласно теории эволюции, консервативные гены и белки. Для ортологичных генов, относящихся к разным семействам генов, характерна синтения – линейная сцепленность в хромосомах с другими генами, относящимися к другим семействам консервативных генов, по-видимому, имеющая физиологический смысл. Такие синтенные группы консервативны в пределах крупных таксонов, вплоть до таксономического типа. Иными словами, существует блочная организация геномов, и это может служить важным фактором эволюции.

В противоположность ортологам различают паралогов – последовательности (нуклеотидов,

Таблица 1. Консервативные аминокислотные последовательности белков у дрожжей и нематоды [6]

Вероятность случайности гомологии P	Группы последовательностей			Открытые рамки считывания (ORFs)	
	всего	по 1 члену	более 1 члена	дрожжи (n = 6217)	нематода (n = 19099)
1×10^{-100}	236	157	79	330 (5.3%)	370 (1.9%)
1×10^{-50}	552	322	230	888 (14.3%)	1094 (5.7%)
1×10^{-10}	1171	611	560	2497 (40%)	3653 (19%)

Примечание. Данные получены путем реципрокного анализа методом BLASTP каждой из ORFs дрожжей со всеми ORFs нематоды и наоборот и соединения числа совпадений при заданной вероятности P по данным [6].

аминокислот), которые выполняют разные биологические функции, но произошли, по-видимому, путем дупликаций и других перестроек генов и их дивергенции от общего предшественника (предка).

В ходе сравнительного анализа геномов и протеомов первых трех эукариотических организмов с полностью секвенированным геномом – дрожжей, нематоды и дрозофилы – возник естественный вопрос: сколько общих белков у трех видов? Сравнительный анализ предсказанных с помощью компьютерных программ белков показал,что примерно 35% генов дрозофилы имеют вероятных ортологов в геноме нематоды [7]. Поиск проводили по гомологии первичной структуры, при этом анализе соблюдали требование, чтобы первичная структура полной молекулы белка совпадала, по крайней мере, на 80% с первичной структурой предсказанного полного белка у другого организма. В таком случае их считали ортологами. Оценка 35% – минимальная потому, что из-за требований к размеру молекулы были исключены такие известные ортологи, как гомеодоменные белки, у которых мало сходства за пределами гомеодоменов. Число пар ортологичных белков заметно не уменьшалось при ужесточении требований сходства. Это позволило утверждать, что исследователи действительно идентифицировали ортологичные белки, у которых может быть общая функция.

Около 23% белков дрозофилы имеют аннотированных ортологов среди белков двух других организмов – дрожжей и нематоды [7]. Эти общие белки, скорее всего, имеют одинаковую функцию во всех клетках у всех эукариот.

Сравнение белков сразу у трех организмов дает картину, которую сложно анализировать и воспринимать без предварительного более простого опыта. Поэтому далее мы остановимся на кратком обзоре более раннего сравнительного исследования геномов двух организмов: одноклеточных дрожжей и просто устроенного многоклеточного организма, нематоды *C. elegans*. Тело этого червя состоит всего из 950 клеток, приблизительно 700 из которых являются соматическими, а остальные – клетками генеративной системы.

Тотальное компьютерное сравнение белков *S. cerevisiae* и *C. elegans* было проведено совместным авторским коллективом из Стенфордского и Бостонского университетов и Национального биотехнологического центра в Бетесде (США). Всего было исследовано 6217 открытых рамок считывания (ORFs) дрожжей и 19099 ORFs нематоды; информация доступна в Saccharimyces Data Base на веб-сайте http://genome-www.stanford.edu/ Saccharomyces/ help/worm/ W-Ycomparison.html

Это исследование принесло неожиданный и интересный результат. Оказалось, что 57% белков-ортологов у этих организмов содержит уникальные белки: по одному белку у каждого организма [7]. При этом выяснилось, что набор высококонсервативных белков у каждого из этих филогенетически далеких организмов кодируется малым числом ORFs: 40% у дрожжей и 20% у нематоды (табл. 1). Эти белки составили первую группу сравнения, проведенного этими авторами. Они кодируют базовые процессы в клетках, сходные у этих организмов, такие как : промежуточный метаболизм, метаболизм ДНК и РНК, “фолдинг” белков, внутриклеточные движения молекул, их дегенерацию [6].

Второй результат анализа групп ортологичных белков дрожжей и нематоды больше соответствовал ожиданиям авторов [6]. Дело в том, что у многоклеточных организмов система белков-регуляторов и проводников сигнала гораздо богаче, чем у одноклеточных. Сравнение таких белковых кластеров у дрожжей и нематоды показало, что хотя существуют общие белки-регуляторы и проводники сигнала, тем не менее у многоклеточного организма произошло увеличение их числа путем постепенного изменения внутримолекулярных доменов в результате их “перетасовки” (domain shuffling) [7]. Этот результат вполне объясним тем, что многоклеточность приводит к специализации клеток, появлению экспорта и импорта клеточных белков, требует создания системы межклеточной коммуникации. Авторы исследования, анализируя различия белков дрожжей и нематоды, предсказали определенные тенденции в эволюционном процессе от одноклеточности к многоклеточности, такие как 1) эволюцию новых регуляторных и сигнальных доменов, 2) эволюцию способа построения белков из предсуществующих белковых доменов и 3) экспансию некоторых семейств белковых доменов путем серии дупликаций.

Другая категория групп сравнения белков дрожжей и нематоды, состоящая из 560 групп белков (см. табл. 1), содержала более двух белков в группе, т.е. более одного ортолога хотя бы у одного из исследованных организмов (не обязательно у нематоды, обладающей большим числом белков вообще). К таким белкам относятся, например, РНК-полимеразы и С-субъединицы фактора репликации ДНК. Последние составляют кластеры, содержащие 12 белков. Но в целом, при уровне значимости 1×10^{-50}, который гарантирует 80%-ную гомологию первичной структуры, число белков, участвующих в метаболизме ДНК и РНК у нематоды, лишь на 10% больше, чем у дрожжей, и этот результат не выходит за рамки ожидавшегося, учитывая универсальность основных молекулярно-биологических явлений (генетический код, репликацию и т.д.). Белки-ортологи, участвующие в метаболизме ДНК и РНК у *S. cerevisiae* и *C. elegans*, составляют 18% всех белков, гомологичных не менее, чем на 80% (табл. 2).

Число белков, выполняющих сходные функции у дрожжей и нематоды, сильно зависит от типа этой функции. Среди структурных белков клетки, например среди белков цитоскелета, на каждый белок дрожжевой клетки приходится в среднем 2.1 ортологичных белка в клетках нематоды. Напоминаем, что речь идет не о массе белков, а о числе их наименований. В сумме ортологичные белки цитоскелета составляют лишь 5% от числа всех белков, гомологичных на уровне 80% у двух организмов, а белки с еще неизвестной функцией – 8% от числа всех белков при таком же уровне гомологии (табл. 2). Если сравнивать физическую массу белков цитоскелета и массу регуляторных белков метаболизма ДНК и РНК, то перевес, несомненно, будет на стороне первых, но это уже выходит за рамки протеомики и относится к области биологии клетки.

Этот качественный и количественный анализ, впервые проведенный на уровне сравнения целых геномов, подтвердил то, что было ранее показано другими методами: фундаментальные молекулярно-генетические процессы у филогенетически далеких организмов осуществляются с помощью примерно одинакового числа ортологичных белков. Кроме того, из этих наблюдений следует, что трехкратная разница общего количества белков, выполняющих сходные функции у дрожжей

Таблица 2. Доля семейств консервативных белков (ортологов) у дрожжей *S. cerevisiae* и нематоды *C. elegans*, имеющих не менее 80% гомологии первичной структуры при пороге вероятности сходства, $P = 1 \times 10^{-50}$ (по данным Червиса и др. [6])

Функция	Доля всех консервативных белков, %	Отношение числа белков у нематоды к их числу у дрожжей
Промежуточный метаболизм	28	1.0
Метаболизм ДНК и РНК	18	1.1
Упаковка и деградация белков	13	1.2
Транспорт и секреция	11	1.5
Проведение сигнала	11	1.2
Неидентифицированные белки	8	1.1
Белки рибосом	6	0.6
Цитоскелет	5	2.1

и нематоды, не может быть отнесена за счет изменчивости в белковых кластерах, хотя существенная изменчивость была обнаружена. Дело в том, что у дрожжей и нематоды есть белки, специфичные для каждого из этих видов. В частности, были выявлены наборы таких белков, связанных с внутриклеточной передачей сигнала (гуморального, ионного) [7] Подобные заключения тоже не новы, и по поводу отдельных клеточных функций их делали задолго до наступления эпохи геномики.

Итак, при анализе геномов установлено, что основные биохимические процессы у одноклеточного организма *S. cerevisiae* и многоклеточного *C. elegans* осуществляются посредством сравнимого числа близких по составу белков. Следовательно, выводы, сделанные из анализа белков у одного организма (а сначала такой анализ был выполнен в отношении дрожжей), могут быть распространены на другой организм. При этом ортологи у другого организма выявлены по первичной структуре. Данные, полученные при сравнении белков основного метаболизма у дрожжей и нематоды, согласуются с этими представлениями.

Наряду с поиском ортологов, основанным на гомологии первичной структуры, существует метод выявления структурных ортологов, основанный на идентификации белков с общими доменами и сходной доменной организацией. Этот подход, в отличие от первого (в основе которого лежат эволюционные представления), является структурно-функциональным и исходит из соображения, что одна и та же функция у разных организмов требует сходства структуры белка. Часто белки, состоящие из нескольких доменов, содержат два и более легко идентифицируемых домена, которые оказываются сходными или гомологичными у разных организмов. Наиболее простыми примерами таких общих доменов служат эукариотическая протеинкиназа (Accesse No. IPR 000719), домен которой встречается в 119 белках у *S. cerevisiae* и 388 белках у *C. elegans*, и домен "цинковых пальцев" типа C2H2 (Accesse No. IPR 001304), встречающийся соответственно 47 и 138 раз в разных белках этих организмов Для сравнительного компьютерного анализа таких белков с успехом использовали такие базы данных, как InterPro (http://www.ebi.ac.uk/interpro) и NCBI DART (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps).

Анализ с применением InterPro, дополненный "ручным" анализом, позволил выявить у трех организмов всего 1400 белковых семейств и отдельных доменов, среди которых 984 – у дрожжей, 1133 – у нематоды и 1177 – у дрозофилы. При этом 744 семейства или домена оказались общими у всех трех [7]. Сходство доменной организации наблюдается у сотен белков дрожжей, нематоды и дрозофилы, участвующих в формировании цитоскелета, регуляции клеточного цикла, клеточной адгезии и других внутриклеточных и межклеточных структурах и процессах.

Какова практическая ценность такого исследования? Наиболее впечатляющим результатом, особенно для правительств и организаций, финансирующих науку, является то, что из уже изученных и доступных для быстрого анализа 289 генов человека, связанных с его болезнями, 177 генов имеют ортологов у *D. melanogaster* и 150 – у *C. elegans* [7]. Это означает, что многие вопросы молекулярной биологии этих генов, их фенотипического проявления на молекулярном уровне и взаимодействия с другими генами, и т.п. могут решаться на более доступных и дешевых моделях – дрозофиле и нематоде и даже на дрожжах [8]. Первый пример такого рода был продемонстрирован еще в 1985 г. в экспериментах, в которых сигнальный белок RAS млекопитающих смог заменить ортологичный ему белок у RAS-дефицитных штаммов дрожжей (см. [7]).

Применимость результатов генетических и особенно феногенетических исследований на дрозофиле к распознанию генетических механизмов болезней человека была обоснована в ясной форме еще в 1926 г в классической работе Тимофеева-Ресовского и Фогта [9] и неоднократно обсуждалась классиком генетических исследований на дрозофиле Меллером [10]. Опубликованные в 2000 г. сравнительные исследования геномов и протеомов модельных организмов [6, 7] и человека положили начало полной каталогизации знаний, позволяющих использовать сравнительную геномику для решения медико-генетические проблем.

ФУНКЦИОНАЛЬНО СХОДНЫЕ, НО НЕГОМОЛОГИЧНЫЕ ГЕНЫ И БЕЛКИ У ФИЛОГЕНЕТИЧЕСКИ ДАЛЕКИХ ЭУКАРИОТ

На данном этапе развития биологических дисциплин, связанных с геномом и клеткой, можно сформулировать новую проблему, которая находится на границе геномики и протеомики, с одной стороны, и клеточной биологии – с другой. Суть ее в том, что в функционально одинаковых состояниях делящихся клеток у филогенетически далеких организмов существуют такие внутриклеточные структуры, которые, выполняя сходные функции, имеют частично или полностью различную ультраструктуру. К ним относятся кинетохоры хромосом, клеточные центры (центриоли животных, полярные тела веретена грибов, морфологически неоформленные клеточные центры растений) и, наконец, синаптонемные комплексы, играющие ключевую роль в синапсисе и генетической рекомбинации гомологичных хромосом в ходе мейоза. Все эти структуры обеспечивают тот или иной способ движения хромосом – их расхождения к полюсам делящейся клетки (кинетохоры и клеточные центры) или взаимного сближения и фиксацию их в состоянии контакта (синаптонемные комплексы).

Различие ультраструктуры этих органелл в разных царствах эукариот позволяет предположить, что в филогенетически очень далеких таксонах они могли формироваться из разных структурных белков. Отсутствие среди них ортологичных белков может восприниматься как вполне понятное явление, а обнаружение ортологов – как лишнее доказательство удивительного единства органического мира.

Авторы статьи столкнулись с этой проблемой в 2000 г., анализируя специфические гены мейоза у *D. melanogaster* и сравнивая их со специфическими генами мейоза у дрожжей *S. cerevisiae* и других организмов [11, 12]. Нас заинтересовали гены, кодирующие белки синаптонемного комплекса (СК).

СК, как "надстройка" хромосом, формируется только во время профазы I мейоза у подавляющего большинства эукариот и играет важную роль в мейозе [13]. Мутации генов, ответственных за формирование СК, приводят к частичной или полной неспособности гомологичных хромосом спариваться накануне (и в процессе) кроссинговера, к снижению частоты или полной элиминации кроссинговера, ахиазматии и нарушению сегрегации гомологичных хромосом в ходе мейоза I. Как следствие, наблюдается анеуплоидия или стерильность половых клеток.

СК всех эукариот сформированы по единой схеме [13]. Они состоят из специфичных для мейоза белковых осей хромосом (латеральные элементы СК), которые располагаются параллельно и соединяются поперечными белковыми филаментами (рис. 1). Пространство между латеральными элементами называется центральным пространством. Фибриллы хроматина гомологичных хромосом крепятся к латеральным элементам. Таким образом, СК служит "рамой", временно скрепляющей гомологичные хромосомы в точном порядке: гомологичные локусы выстраиваются один напротив другого. В центральном пространстве СК располагаются "рекомбинационные узелки" – конгломераты ферментов, необходимых для рекомбинации молекул ДНК [13]. При одинаковом общем плане строения детальная ультраструктура СК различается у сумчатых грибов, нематод, насекомых, высших растений, млекопитающих и в других таксонах. Структурные белки СК (белки латеральных элементов и поперечных филаментов) изучены лишь у четырех видов млекопитающих [14–17] и у дрожжей *S. cerevisiae* [18, 19] и описаны в обзорах [20, 21]. У дрожжей и млекопитающих не наблюдается гомологии первичной структуры этих функционально аналогичных белков. Возникают вопросы: каким образом функционально одинаковые клетки, а именно клетки, вступившие в профазу мейоза, у очень эволюционно далеких организмов создают функционально одинаковые субклеточные структуры из неодинаковых белков? Какие признаки белков, формирующих СК, оказываются решающими для построения СК по одинаковому плану в разных типах и даже царствах эукариот?

Наиболее простым типом белков для такого сравнительного исследования нам представились белки, формирующие поперечные филаменты в центральном пространстве СК. У дрожжей и млекопитающих поперечные филаменты сформированы из одного белка (около 700 аминокислотных остатков), состоящего из трех доменов: центрального, имеющего протяженную альфа-спираль, и двух концевых доменов, не формирующих альфа-спирали и имеющих подобие глобулярной формы (см. [18]).

Иммуноцитохимический анализ под электронным микроскопом [22] показал, что поперечные филаменты СК состоят из двух пар молекул таких белков (рис. 2). В каждой паре молекулы уложены параллельно и ориентированы одинаково, и эти пары располагаются в центральном пространстве СК перпендикулярно к латеральным элементам и тандемно по отношению одна к другой. При этом они ориентированы "голова к голове"

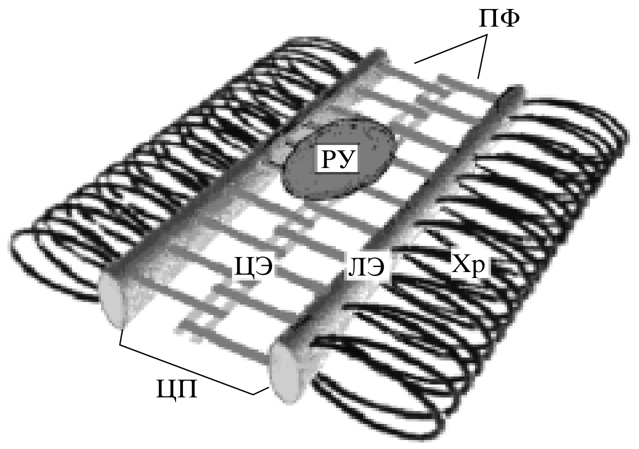

Рис. 1. Схема строения синаптонемного комплекса. ЛЭ – латеральные элементы, ЦП – центральное пространство, ЦЭ – центральный элемент, ПФ – поперечные филаменты, формирующие ЦП, Хр – петли хроматина, РУ – рекомбинационный узелок.

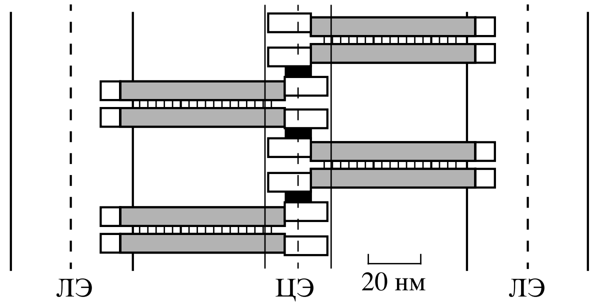

Рис. 2. Схема организации центрального пространства синаптонемного комплекса (СК). Пунктиром указана граница осевых элементов хроматид внутри латеральных элементов СК, белыми прямоугольниками – концевые (глобулярные) домены белка поперечных филаментов, серыми прямоугольниками – центральные (алфа-сиральные) домены этого белка, маленькими черными прямоугольниками – гипотетические белок-белковые контакты концевых доменов. Остальные бозначения те же, что на рис. 1.

(N-конец к N-концу). С-концы обладают ДНК-связывающими свойствами и погружены в латеральные элементы СК. Ширина центрального пространства СК зависит от размера центрального домена, имеющего палочковидную форму. Делеции в этом домене, обнаруженные у аллельных мутантов по гену *ZIP1* дрожжей, уменьшали ширину центрального пространства СК и сближали латеральные элементы СК [18, 19].

На основе данных литературы мы провели корреляционный анализ двух признаков: ширины центрального пространства СК и размера молекул белков, формирующих поперечные филаменты СК. Мы выявили высокую корреляцию ($r = 0.85$; $p < 0.001$) между длиной целой молекулы

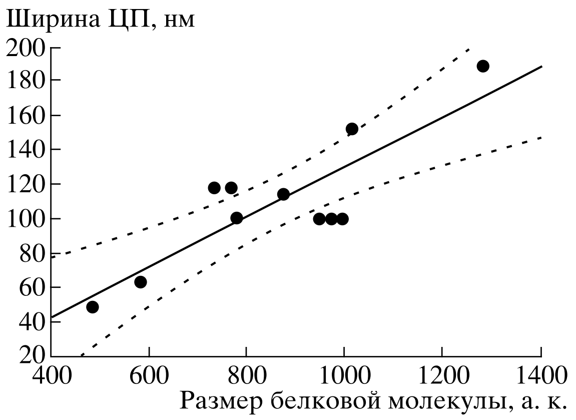

Рис. 3. Зависимость ширины центрального пространства СК от размеров белковой молекулы, его формирующей. ЦП – центральное пространство СК, а.к. – количество аминокислотных остатков в белке. Черные кружки соответствуют белкам SCP1 крысы, мыши и человека, и разным вариантам белка Zip1p у дрожжей *S. cerevisiae* дикого типа и мутантов с внутренними делециями и дупликациями разной протяженности в этом белке. Наклонная прямая – линия регрессии для указанных признаков. Пунктиром показан 95%-ной доверительный интервал линии регрессии.

белка SCP1 млекопитающих и белка Zip1p дрожжей и шириной центрального пространства СК у этих организмов (рис. 3). Еще более высокая корреляция была получена при сравнении длины центрального (альфа-спирального) домена белков с шириной центрального пространства СК ($r = 0.90$; $p < 0.001$). Столь высокий и достоверный коэффициент корреляции означает прямую функциональную связь, т.е. прямую зависимость ширины центрального пространства СК от длины центрального (линейного по форме) домена белков поперечных филаментов SCP1 и Zip1p [23].

Эти находки создали основу, которая позволила нам предпринять компьютерный поиск у *D. melanogaster* белка, аналогичного Zip1p дрожжей и SCP1 млекопитающих. Согласно нашему предположению (по аналогии с указанными белками), этот белок должен состоять из трех доменов, причем центральный домен должен содержать альфа-спираль, длина которой должна соответствовать половине ширины центрального пространства СК дрозофилы. Естественно, что в первую очередь мы искали среди тех белков, которые специфически экспрессируются в мейозе.

Компьютерная идентификация гена и белка СК у дрозофилы

Алгоритм автоматического поиска ортологов известных белков основан на сравнении их первичной структуры, даже если при этом используются данные о вторичной структуре (доменная организация). Так, например, программа автоматического подбора ортологов, SGD Worm-Yeast Protein Comparison (http://genome-www.stanford.edu/Saccharomyces/worm/), базируется на сравнении аминокислотных последовательностей с известными консенсусными последовательностями белковых семейств и белковых доменов. Однако это не позволяет выявлять структурно-функциональные аналоги у филогенетически далеких организмов, так как не учитывает роль белков в формировании субклеточных структур в мейотических клетках. Поэтому мы использовали метод, в котором результаты исследования *in silico* (т.е. в компьютерных базах данных) дополнялись данными морфологических и генетических экспериментов. Такой анализ пока еще слабо автоматизирован и основан на "ручном" сравнении вторичных структур (например, длины альфа-спирального участка) у интересующих нас белков.

Среди более чем 80 генов, контролирующих мейоз у *D. melanogaster* [12], лишь один ген – *c(3)G*, crossover suppressor on 3 of Gowen, локализованный в районе 89A2-5 цитологической карты хромосом [24], – несомненно и специфически затрагивает формирование СК [25]. Однако ни механизм его действия, ни соответствующий генный продукт до 2001 г. не были выявлены [26]. По одной из старых гипотез, ген *c(3)G* может кодировать синтез одного из компонентов СК [25]. Мы поставили перед собой задачу анализировать виртуальные продукты генов из района 3R хромосомы D. melanogaster, с избытком перекрывающего локус гена *c(3)G*. С этой целью мы исследовали все гены, локализованные в области 88E-89B цитологической карты Бриджеса. В частности, было решено сравнить доменную организацию и вторичную структуру доменов у предсказанных белков дрозофилы с таковой у известных белков СК дрожжей и млекопитающих. Мы проанализировали in silico 78 предсказанных генов из района 88E6-89B2, содержащего около 250 тпн (по базе данных NCBI). Белковые продукты этих генов ранее не были идентифицированы. Оказалось, что виртуальный белковый продукт только одного из этих генов по общему размеру и доменной организации аналогичен белкам SCP1 и Zip1p. Это продукт гена *CG17604* дрозофилы по номенклатуре NCBI. Он состоит из трех доменов: двух концевых и центрального, который имеет палочковидную альфа-спиральную структуру. Ген *CG17604* расположен на хромосоме 3R в позиции 36250–36253 тпн по молекулярной карте NCBI. Виртуальный продукт этого гена состоит из 744 аминокислот, а его центральный домен, состоящий, по уточненным данным, из 495 а.к., способен формировать альфа-спиральную структуру, сходную с таковой у центральных доменов белков Zip1p и SCP1. Ширина центрального пространства СК у *D.melanogaster* составляет 109 нм

Таблица 3. Характерика молекул экспериментально установленных [14–19] и предсказанных (данная работа) белков, формирующих поперечные филаменты СК, у организмов с полностью секвенированным геномом

Биологические виды, их нормальные и мутантные белки	Размеры белковых молекул и их доменов			pI белков и их доменов			
	общая длина молекулы	длина альфа-спирального участка	ширина ЦП СК	N-конце-вой домен	централь-ный домен	С-конце-вой домен	целая молекула
M. musculus SCP1	993	713	100	5.9	5.3	9.7	5.8
H. sapiens SCP1	973	677	100	5.0	5.4	9.7	5.7
R. norvegicus SCP1	946	717	100	4.2	5.33	9.8	5.6
S. cerevisiae Zip1	875	632	115	4.8	6.1	10.1	6.4
Zip1-m2*	583	285	63				
Zip1-mc1*	484	170	49				
Zip1-mc2*	776	634	101				
Zip1-n1*	732	578	118				
Zip1-nm1*	767	512	118				
Zip1-2XH2**	1012	799	153				
Zip1-3XH2**	1280	1067	189				
D. melanogaster CG17604	744	495	109	10.0	4.9	9.7	5.9
A. thaliana AAD10695	923	476	100–120	5.3	5.4	9.0	5.6
A. thaliana AAB70839	924	575	100–120	9.9	4.7	4.4	5.1
C. elegans Q11102	1132	938	70–85	11.9	5.1	11.0	5.5
C. elegans T26844	1083	536	70–85	8.1	6.9	5.3	6.6
C. elegans T27907	772	460	70–85	5.2	5.9	5.7	5.1
C. elegans Z81586	484	460	70–85	4.9	9.5	10.0	9.4
C. elegans WP:CE17456	213	50	70–85				

Примечание. ЦП – центральное пространство СК; pI – изоэлектрическая точка белка (домена).
* Белки Zip1 дрожжей с делециями; ** то же с дупликациями.

[27]. Эти данные мы нанесли на график (рис. 4), и полученная точка оказалась в зоне 95%-ного доверительного интервала линии регрессии двух признаков: размера альфа-спирального участка белковых молекул и ширины центрального пространства СК млекопитающих и дрожжей.

Оказалось, что по размеру альфа-спирального участка белок гена *CG17604* соответствует половине ширины центрального пространства СК у *D. melanogaster*, и этот результат удовлетворяет закономерности, наблюдавшейся при сравнении структуры СК у дрожжей и млекопитающих.

Согласно данным о первичной структуре белков [26], предсказанный продукт гена *CG17604* имеет гомологию с самыми разными белками, многие из которых способны формировать альфа-спирали, в том числе с белком полярного тельца веретена NUF1 (*S. cerevisiae*), тяжелой цепью миозина (*D. melanogaster, C. elegans, A. thaliana, Homo sapiens, Dugesia japonica*), белками ламины (*Mus musculus*). Подобная гомология выявлена также и для известных белков СК: Zip1p дрожжей [19], SCP1 крысы и мыши [14, 16]. Интересно, что по аминокислотной последовательности продукт гена *CG17604* не гомологичен ни SCP1, ни Zip1p, так же как Zip1p не гомологичен белкам поперечных филаментов СК млекопитающих [19]. Таким образом, несмотря на отсутствие гомологии по нуклеотидной последовательности с генами *ZIP1* и *SCP1*, виртуальный ген *CG17604 D. melanogaster* сходен с этими генами в том, что как он, так и они, имеет частичную гомологию с генами, кодирующими указанное выше большое семейство структурных белков.

Кроме такого сравнения мы проанализировали аннотированные компанией Selera Genomics белковые продукты генов из 9 районов локализации других мейотических мутаций D.melanogaster (*mei-9, mei-41, mei-217, mei-218, mei-P14, mei-P22, mei-P26, mei-w68, mei-S82*). Оказалось, что ни один из 200 аннотированных виртуальных белковых продуктов этих генов не способен формировать альфа-спираль необходимой длины для образования поперечных филаментов СК у дрозофилы.

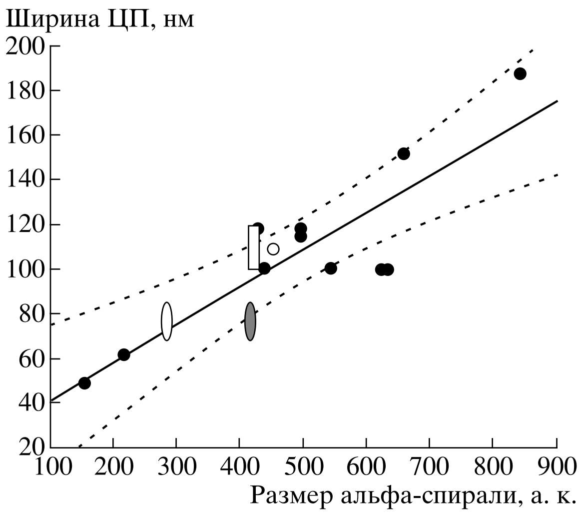

Рис. 4. Зависимость ширины центрального пространства СК от размера альфа-спирали белковой молекулы, формирующей поперечные филаменты.СК. Условные обозначения те же, что на рис. 3. Координаты белого кружка соответствуют ширине центрального пространства СК и длине альфа-спирали белка CG17604 *D. melanogaster*; координаты белого овала соответствуют ширине центрального пространства СК и длине альфа-спирали белка WP: CE17456 C elegans; координаты заштрихованного овала соответствуют ширине центрального пространства СК и половине длины альфа-спирали белка Q11102 C. elegans; координаты прямоугольника соответствуют ширине центрального пространства СК и длине альфа-спирали белка AAD10695 A. thaliana.

Для дополнительной проверки мы с помощью программы ProtParam оценили изоэлектрические точки (pI) известных белков СК дрожжей и млекопитающих и найденного нами продукта гена *CG17604*. Анализировали все молекулы целиком и отдельные домены (табл. 3). Для белка CG17604 все параметры, кроме pI N-концевого домена, сходны с параметрами белков Zip1 и SCP1/SYCP1. Как известно, С-концевые домены указанных белков имеют суммарный положительный заряд (основный, т.е.щелочной характер), погружены в латеральные элементы СК и поэтому контактируют с ДНК, а N-концевые домены имеют отрицательный заряд (кислый характер), направлены внутрь центрального пространства СК, где, перекрываясь, образуют центральный элемент СК [16, 19]. Возможно, что специфическая морфология центрального элемента СК дрозофилы (см. рис. 1), имеющего четкую ультраструктурную исчерченность и относительно большую ширину – 32 нм [27], связана с положительным зарядом N-концевого домена белка CG17604. Итак, по большинству исследованных параметров обнаруженный нами продукт гена *CG17604* подобен известным белкам поперечных филаментов СК у других организмов.

Ген *CG17604* локализован, по данным FlyBase, в районе 89A7-8. В то же время ген *c(3)G*, с которым, по нашему предположению, *CG17604* может быть идентичен, локализован в районе 89A2 (рис. 5). Это несовпадение, возможно, связано с тем, что картирование в данном районе третьей хромосомы *D. melanogaster* сопряжено с объективными трудностями. Подтверждением является имеющееся несоответствие в локализации генов на молекулярной и цитогенетической карте NCBI в данном районе в виде сдвига в положении генов, для которых были проведены оба типа картирования (рис. 6). Поэтому мы предполагаем, что именно ген *CG17604* кодирует белок поперечных филаментов центрального пространства синаптонемного комплекса *D. melanogaster* и, вполне вероятно, является известным геном *c(3)G*. Это предположение доступно экспериментальной проверке. Такой проверкой можно считать выполненный независимо от нас и одновременно с нами анализ молекулярной организации гена *c(3)G* и его продукта. В тезисах 17-й Европейской исследовательской конференции по дрозофиле (1–5 сентября 1001 г., Эдинбург) сотрудники лаборатории Р.С. Хоули сообщили, что ген *c(3)G* кодирует белок, который имеет структурное сходство с белками Zip1p и с SCP1 [28]. В этой работе были получены белковый продукт гена *c(3)G* и моноклональные антитела к нему. Методом иммунофлюоресцентной микроскопии было установлено, что эти антитела окрашивают центральную часть профазных бивалентов в ооцитах *D. melanogaster* [29]. Таким образом, локализация антител к белковому продукту гена *c(3)G* совпадает с предполагаемой нами локализацией белка

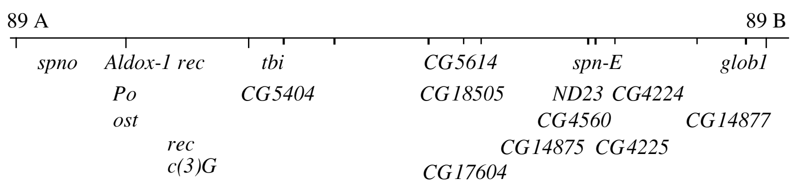

Рис. 5. Генетическое окружение гена *CG17604* (по данным FlyBase).Указано положение на цитологической карте хромосомы 3 дрозофилы, в районе 89A-89B, генов, выявленных компьютерными методами и аннотированных (в том числе гена *CG17604*), а также некоторых генов, установленных генетическимми методами, в том числе *c(3)G*.

CG17604 и это служит аргументом в пользу идентичности генов *c(3)G* и *CG17604*.

Компьютерный поиск генов и белков синапсиса у A. thaliana

Поиск конкретной последовательности гена *c(3)G* у дрозофилы был ограничен относительно небольшим районом, поэтому число проанализированных нами гипотетических генов было невелико (78 генов). В сравнении с этим задача выявления генов с функцией, аналогичной генам *SCP1* и *ZIP1*, у других организмов с полностью секвенированным геномом осложнена отсутствием у них неизвестных мутаций, нарушающих формирование синаптонемного комплекса. Для таких организмов область поиска расширяется до размеров целого генома. Для уменьшения числа анализируемых белков необходимо вводить еще один этап исследования, а именно, поиск в протеомах таких организмов белков, имеющих доменную организацию, характерную для SCP1 и Zip1p. Такой анализ был нами осуществлен с помощью программы DART (NCBI) (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps).

Проводя этот поиск, мы оценивали вторичную структуру выявленных данной программой белков и выбирали те из них, которые имеют доменную организацию, сходную с Zip1p или SCP1, и размеры, достаточные для формирования поперечных филаментов СК у *A. thaliana*.

Для *A. thaliana* нам не удалось найти в литературе данных о ширине центрального пространства СК. Тем не менее, анализ литературы показал, что важной особенностью общей организации СК у подавляющего большинства филогенетически далеких организмов, включая высшие растения, является то, что ширина центрального пространства, как правило, не выходит за пределы 90–110 нм. Руководствуясь этими соображениями, мы выявили в протеоме у *A. thaliana* виртуальный белок AAD10695, который по размеру и доменной организации аналогичен белкам Zip1p и SCP1 (табл. 3; рис. 4). Как сказано выше, важным параметром известныъх белков поперечных филаментов СК является щелочной характер (pI > 8) их С-концевого домена. Белок AAD10695 *A. thaliana* удовлетворял этому требованию.

Таким образом, так же как у D.melanogaster, у *A. thaliana* был обнаружен только один предполагаемый функциональный аналог SCP1 млекопитающих и Zip1p дрожжей, что совпадает с основными результатами сравнительной протеомики для высокоспециализированных белков.

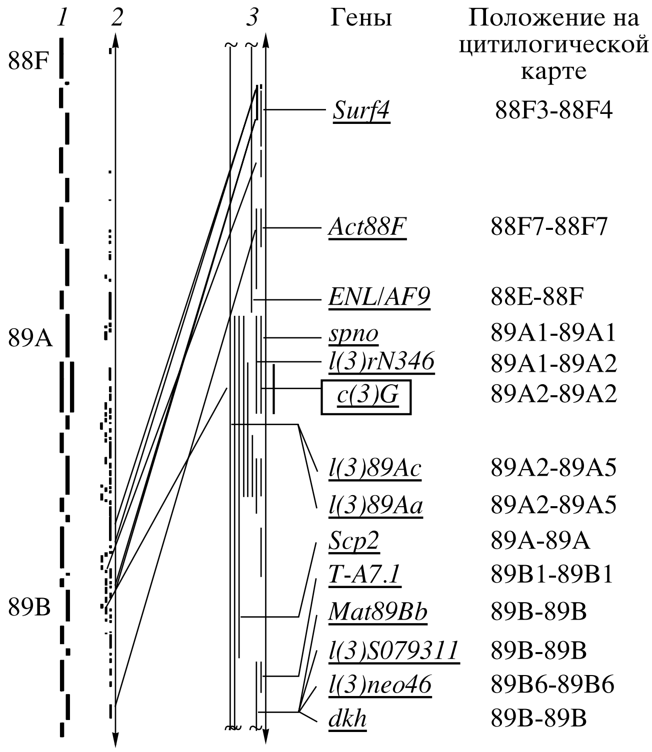

Рис. 6. Цитогенетическая и молекулярная карты участка 88F-89B хромосомы 3 D.melanogaster в районе локализации гена *c(3)G* (по данным NCBI). Указаны известные гены и их позиции на цитологической карте. 1 – цитогенетическая шкала, 2 – положение генов на молекулярной карте, 3 – положение генов на цитогенетической карте. Тонкие линии соединяют места локализации генов на двух картах. Несоответствие цитогенетической и молекулярной карт для рассматриваемого района выражается в сдвиге выявленных позиций генов на цитогенетической карте относительно молекулярной.

Особенности генов и белков СК у нематоды C. elegans

Особенностью поиска аналогов SCP1 и Zip1p у *C. elegans* было то, что для этой нематоды уже были найдены методами *in silico* несколько гипотетических генов, белки которых аннотированы как ортологи Zip1p. Так, по данным Worm Base (http://www.wormbase.org/), аналогом SCP1 и Zip1p у *C. elegans* является белок WP:CE17456 – продукт гена *syp2* [30]. По результатам анализа аминокислотной последовательности этого белка, проведенного нами с помощью программы ProtParam (http://www.expasy.ch/ tools/protparam.html), этот белок не может формировать альфа-спиральную структуру достаточной длины для того, чтобы сформировать поперечные филаменты СК, и на этом основании мы считаем, что он не может быть белком поперечных филаментов СК. По данным Proteome, Inc (http://www.proteome.com/index.html), аналогом SCP1 и Zip1 у *C. elegans* является белок Z81586. По всем параметрам – размеру белковой молекулы, ее доменной организации, размеру альфа-спирального центрального домена и физико-химическим характерис. тикам С-концевого домена – этот белок может считаться функциональным аналогом SCP1 и Zip1. Но морфология СК у *C. elegans* отличается от таковой у дрожжей, млекопитающих и дрозофилы. Это отличие состоит в отсутствии четко выраженного центрального элемента СК – электронноплотной полосы в центральном пространстве, идущей параллельно латеральным элементам СК [31]. На некоторых электронно-микроскопических фотографиях ультратонких срезов СК *C. elegans* видны поперечные филаменты, проходящие без перерыва от одного латерального элемента СК до другого и ни какого центрального элемента на их пути не пересекающие [32]. Это позволяет предполагать другую молекулярную организацию центрального пространства у *C. elegans*, а именно, формирование поперечных филаментов не двумя молекулами, направленными навстречу друг другу и контактирующими N-концами с образованием центрального элемента (модель 1), а одной длинной молекулой, протянутой от одного латерального элемента СК до другого (модель 2). О такой возможности строения центрального пространства СК у другого объекта – жука *Blaps cribrosa* писали Шмекель и Данехолт [33]. В этом случае оба концевых домена белка должны иметь щелочной (основный) характер.

Учитывая сказанное, мы провели собственный поиск среди белков *C. elegans* аналогов SCP1 и Zip1p. При этом мы отбирали белки, соответствующие обеим моделям: модели 1 из двух относительно коротких белков, расположенных “голова к голове”, и модели 2 из одного длинного белка . В геноме *C. elegans* нами были выявлены три гена, кодирующих такие белки (Q11102, T26844, T27907). Все три белка имеют альфа-спиральные участки требуемого размера. Для белка Q11102, удовлетворяющего модели 2, взята половина длины альфа-спирали, для двух остальных белков – полная длина (рис. 4). При этом критерию щелочной изоэлектрической точки (pI) концевых доменов удовлетворял только один белок – Q11102. Он соответствует модели 2 (длинный белок, у которого оба концевых домена основные). С другой стороны, существует белок Z81586, который по всем параметрам соответствует модели 1. Для того, чтобы сделать выбор в пользу одного из белков, мы программными средствами Worm Base провели анализ генетического окружения в районах локализации двух генов-кандидатов, а именно в районе –4 ± 1 генетической карты хромосомы 1 для гена, кодирующего Z81586, и в районе –10 ± 3.5 хромосомы X для гена, кодирующего Q11102. Ни одного известного гена, ответственного за нормальное протекание мейоза, в этих районах нами выявлено не было. Таким образом, на данном этапе еще нельзя выбрать единственно верный вариант, но, основываясь на необычной морфологии центрального пространства СК у *C. elegans,* мы считаем, что искомый функциональный аналог SCP1 и Zip1p – это белок Q11102 и что поперечные филаменты СК у этой нематоды формируются длинными молекулами этого белка, перекрывающими все центральное пространство в отличие от дрожжей и млекопитающих, у которых оно перекрывается двумя димерами белковых молекул, направленными навстречу друг другу.

ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Тотальное секвенирование геномов у трех модельных объектов – дрожжей, нематоды и арабидопсиса и создание специальных компьютерных программ позволило не только выявить открытые рамки считывания ранее неизвестных генов, но и прогнозировать структурно-функциональные характеристики белковых продуктов этих генов. Существующие ныне два основных метода автоматического поиска ортологов – по гомологии первичной структуры и по структурным доменам – не лишены недостатков, могут быть использованы как методы предварительного отбора (возможно, за исключением некоторых очевидных случаев). Для точной идентификации структурно-функциональных аналогов белков мы предлагаем дополнительный метод поиска in silico соответствия вероятных кандидатов выполнения предполагаемой функции и ультраструктуры клеточных органелл, сформированных этими белками. Этот подход оказался продуктивным в приложении к ключевому белку синаптонемного комплекса. Правильность нашего подхода нашла экспериментальное подтверждение при эксперментальном выявлении такого белка у дрозофилы. Удастся ли применить этот подход к другим белкам и построенным из них внутриклеточным структурам у других организмов, предстоит выяснить.

Интересная задача состоит в том, чтобы проследить способы формирования синаптонемных комплексов и других клеточных структур, построенных из белков, не имеющих гомологии первичной структуры, но являющихся структурными аналогами в разных царствах эукариот [34].

Работа выполнена при финансовой поддержке РФФИ (проект № 99-04-48182).

СПИСОК ЛИТЕРАТУРЫ

1. *Goffeau A., Barrell B.G., Bussey H. et al.* Life with 6000 genes // Science. 1996. V. 274. № 5287. P. 546, 563–7@.
2. *The C. elegans* Sequencing Consortium. Genome Sequence of the Nematode *C. elegans*: a platform for In-

vestigating biology // Science. 1998. V. 282. P. 2012–2020.

3. *Adams M.D., Celniker S.E., Holt R.A. et al.* The genome sequence of *Drosophila melanogaster* // Science. 2000. V. 287. P. 2185–2195.
4. *Myers E.W., Sutton G.G., Delcher A.T. et al.* // A whole-genome assembly of *Drosophila* // Science. 2000. V. 287. P. 2196–2204.
5. The Arabidopsis Genome Initiativa. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana // Nature. 2000. V. 408. P. 796–815.
6. *Chervitz S.A., Aravind L., Sherlock G. et al.* Comparison of the Complete Protein sets of worm and yeast: orthology and divergence // Science. 1998. V. 282. P. 2022–2028.
7. *Rubin M.G., Yandell M.D., Wortman J.R. et al.* Compararive genomics of the eukaryotes // Science. 2000. V. 287. P. 2204– 2215.
8. *Kornberg T.B., Krasnov M.A.* The Drosophila Genome sequence: implication for biology and meficine // Science. 2000. V. 287. P. 2218–2220.
9. *Timofeeff-Ressovsky N.W., Fogt O.* Uber idiosomatishe Variationsgruppen und ihre Bedeutung fur dise Klassifikation der Krankheiten // Naturwiss. 1926. B. 14. № 50–51. S. 1188–1190.
10. *Muller H.J.* Studies in genetics. Bloomington: Indiana Univ. Press, 1962. 618 p.
11. *Богданов Ю.Ф.* Молекулярная концепция мейоза выдерживает испытание. Итоги Четвертой Европейской конференции по мейозу // Генетика. 2000. Т. 36. № 4. С. 585–590.
12. *Гришаева Т.М., Богданов Ю.Ф.* Генетический контроль мейоза у дрозофилы // Генетика. 2000. Т. 36. № 10. С. 1301– 1321.
13. *Zickler D., Kleckner N.* Meiotic chromosomes: integrating structure and function // Ann. Rev. Genet. 1999. V. 33. P. 603–754.
14. *Meuwissen R.L.J., Offenberg H.H., Dietrich A.J.J. et al.* A coiled-coil related protein specific for the synapsed regions of meiotic prophase chromosomes // EMBO J. 1992. V. 11. P. 5091–5100.
15. *Dobson M.J., Pearlman R.E., Karaiskakis A. et al.* Synaptonemal complex proteins: occurence, epitope mapping and chromosome disjunction // J. Cell. Sci. 1994. V. 107. P. 2749–2760.
16. *Liu J.G., Yuan L., Brundell E. et al.* Localization of the N-terminus of SCP1 to the central element of the synaptonemal complex and evidence for direct interactions between the N-termini of SCP1 molecules organized head-to-head // Exptl. Cell Res. 1996. V. 226. P. 11–19.
17. *Meuwissen R.L.J., Meerts I., Hoovers J.M.N. et al.* Human synaptonemal complex protein 1 (SCP1): isolation and characterization of the cDNA and chromosomal localization of the gene // Genomics. 1997. V. 39. P. 377–384.
18. *Tung K.S., Roeder G.S.* Meiotic chromosome morphology and behavior in *zip1* mutants of *Saccharomyces cerevisiae* // Genetics. 1998. V. 149. P. 817–832.
19. *Dong H., Roeder G.S.* Organization of the yeast Zip1 protein within the central region of the synaptonemal complex // J. Cell Biol. 2000. V. 148. № 3. P. 417–426.
20. *Heyting C.* Synaptonemal complex: structure and function // Curr. Opin. Cell Biol. 1996. V. 8. P. 389–396.
21. *Пенкина М.В., Карпова О.И., Богданов Ю.Ф.* Белки синаптонемного комплекса – специфические белки мейотических хромосом // Молекуляр. биология. 2002. Т. 36. № 1. С. 1–11.
22. *Schmekel K., Meuwissen R.L., Dietrich A.J. et al.* Organization of SCP1 protein molecules within synaptonemal complexes of the rat // Exptl Cell Res. 1996. V. 226. № 1. P. 20–30.
23. *Богданов Ю.Ф., Гришаева Т.М., Дадашев С.Я.* Ген *CG17604 Drosophila melanogaster* – возможный функциональный гомолог генов *ZIP1* и *SCP1* (*SYCP1*), кодирующих белки синаптонемного комплекса // Генетика. 2002. Т. 38. № 1. С. 108–112
24. *Matsubayashi H., Yamamoto M.-T.* Dissection of chromosome region 89A of *Drosophila melanogaster* by local transposition of *P* elements // Genes Genet. Syst. 1998. V. 73. P. 95–103.
25. *Smith P.A., King R.C.* Genetic control of synaptonemal complexes in *Drosophila melanogaster* // Genetics. 1968. V. 60. P. 335–351.
26. FlyBase *Drosophila* database (http: // flybase.bio.indiana.edu).
27. *Carpenter A.T.C.* Electron microscopy of meiosis in *Drosophila melanogaster* females. I. Structure, arrangement and temporal change of the synaptonemal complex in wildtype // Chromosoma. 1975. V. 51. P. 157–182.
28. *Wayson S.M., Page S.L., Carey B.W. et al.* The role of *c(3)G* in promoting meiotic chromosome synapsis and exchange // 17th European *Drosophila* Research Conference (1–5 September 2001). Abstracts. Edinburg, 2001. P. 34.
29. *Page S.L., Hawley R.S. c(3)G* encodes a *Drosophila* synaptonemal complex protein // Genes Dev. 2001. V. 15. № 23. P. 3130–3143.
30. *MacQueen AJ, Colaiacovo MP, Villeneuve AM.* Molecules underlying meiotic nuclear reorganization and the homolog pairing process // International *C. elegans* Meeting. 2001. P. 291.
31. *Goldstein P.* The synaptonemal complex of Caenorhabditis elegans: pachytene karyotype analysis of the *Dp1* mutant and Disjunction Regulator Regions // Chromosoma. 1985. V. 93. P. 177–182.
32. *Dernburg A.F., McDonald K., Moulder G. et al.* Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis // Cell. 1998. V. 94. P. 387–309.
33. *Schmekel K., Daneholt B.* The central region of the synaptonemal complex revealed in three dimensions // Trends Cell Biol. 1995. V. 5. P. 239–242.
34. *Богданов Ю.Ф.* Гомологические ряды признаков мейоза: эволюция и консерватизм // Эволюция, экология, биоразнообразие: Матер. конф. памяти Николая Николаевича Воронцова (1934–2000) / Ред. В.А.Красилов. М. Университетский центр довузовского образования, 2001. С. 60–75.