

На правах рукописи



ГАЛАЧЬЯНЦ Юрий Павлович

**ОПРЕДЕЛЕНИЕ И АНАЛИЗ
НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
МИТОХОНДРИАЛЬНОГО И ХЛОРОПЛАСТНОГО ГЕНОМОВ
ДИАТОМОВОЙ ВОДОРΟΣЛИ *SYNEDRA ACUS***

03.01.03 – молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск – 2012

Работа выполнена в Лимнологическом институте СО РАН

Научный руководитель:

д.б.н., профессор **Лихошвай Елена Валентиновна**

Официальные оппоненты:

Демаков Сергей Анатольевич, д.б.н.

Институт молекулярной и клеточной биологии СО РАН,
заведующий лабораторией

Морозов Игорь Владимирович, к.б.н.

Институт химической биологии и фундаментальной
медицины СО РАН, старший научный сотрудник

Ведущая организация:

Институт биологии гена РАН

Защита состоится « 1 » июня 2012 г. в 11 часов 30 минут
на заседании диссертационного совета Д 003.045.01 на базе
Института химической биологии и фундаментальной медицины СО РАН
по адресу: 630090, Новосибирск, проспект академика Лаврентьева, 8

С диссертацией можно ознакомиться в библиотеке
Института химической биологии и фундаментальной медицины СО РАН

Автореферат разослан « 27 » апреля 2012 г.

Учёный секретарь диссертационного совета

к.х.н., доцент



Коваль В. В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Диатомовые водоросли – это одноклеточные автотрофные эукариотические организмы, одни из главных продуцентов органического углерода, формирующие около 40% первичной продукции Мирового океана, активные участники круговорота кремния, азота и фосфора (Nelson *et al.*, 1995; Stoermer, Smol, 1999). Они распространены в различных биотопах – от морских до ультра-пресных, от пакового льда Антарктики до горячих источников (Round, Crawford, Mann, 1990).

Традиционная систематика диатомей (отдел Bacillariophyta) базируется на симметрии и особенностях строения кремнистого панциря: радиально-центрические и биполярно-центрические, пеннатные с билатеральной симметрией панциря, имеющие или не имеющие шов (Round, Crawford, Mann, 1990). В результате молекулярно-филогенетического анализа нескольких генов (ядерных рРНК – Medlin, Williams, Sims, 1993; Medlin, Kooistra, Schmid, 2000; Theriot *et al.*, 2009; хлоропластного *rbcL* – Andersen, 2004; митохондриального *cox1* – Ehara *et al.*, 2000) систематика диатомей была пересмотрена. Оказалось, что только группа пеннатных шовных является монофилетичной.

Диатомовые водоросли относятся к царству Chromista (Cavalier-Smith, 1981; Cavalier-Smith, 2009). Согласно одной из наиболее распространенных гипотез, диатомовые водоросли возникли около 250 млн л.н. (Medlin, 2009) в результате эндосимбиоза автотрофной клетки, от которой они унаследовали хлоропласт, и гетеротрофной разножгутиковой клетки, от которой диатомеи унаследовали митохондрии (Gibbs, 1981; Sims, Mann, Medlin, 2006). В процессе эволюции разных линий Chromista эндосимбиоз происходил многократно (Cavalier-Smith, 1999; Cavalier-Smith, 2010; Dorrell, Smith, 2011). В результате эндосимбиогенеза и эволюции часть генов клеточных органелл была перенесена в ядерный геном (Keeling, 2009). Новые сведения для развития эволюционной теории может принести сравнительный анализ полных геномов диатомей и их органелл. К моменту начала настоящего исследования были определены последовательности только двух полных ядерных геномов и нескольких геномов органелл диатомовых водорослей. До недавнего времени получение новых данных было затруднено из-за отсутствия быстрых методов секвенирования.

Учитывая сложность эволюционной истории диатомей, необходимость определения последовательностей полных геномов их органелл становится актуальной задачей геносистематики и филогеномики.

Пеннатная бесшовная диатомея *Synedra acus* широко распространена в пресных водоемах. В озере Байкал подвид *S. acus* subsp. *radians* (Kütz.) Skabitsch. входит в состав доминирующих видов фитопланктона, активно развивается в весенний период, вносит существенный вклад в пищевые сети и круговорот кремния, а также является биостратиграфическим маркером байкальской палеолетописи плейстоцена и голоцена (Grachev *et al.*, 1998).

Цель и задачи исследования. Цель настоящей работы состояла в определении полных нуклеотидных последовательностей митохондриального и хлоропластного

геномов диатомовой водоросли *S. acus* для проведения сравнительного анализа геномов органелл диатомовых водорослей. В ходе исследования было необходимо решить следующие задачи:

1. Определить полные нуклеотидные последовательности геномов митохондрий и хлоропластов *S. acus*.
2. Провести сравнительно-геномный и молекулярно-филогенетический анализ мтДНК доступных геномов и отдельных митохондриальных генов диатомовых и других разножгутиковых (Heterokonta).
3. Провести сравнительно-геномный анализ доступных хлоропластных геномов и отдельных хлоропластных генов диатомовых водорослей.

Научная новизна и практическая ценность работы. Проведено определение нуклеотидных последовательностей полных геномов органелл диатомовой водоросли *S. acus*. До получения результатов настоящей работы были известны последовательности двух митохондриальных и шести хлоропластных геномов диатомей. *S. acus* является первой бесшовной пеннатной и первой из пресноводных диатомей, для которой определены последовательности геномов органелл. Впервые показано, что в мтДНК диатомовых водорослей существует блок с консервативным порядком генов и что интенсивность геномных перестановок в мтДНК диатомей выше, чем у бурых водорослей. Впервые показано, что разные интроны второй группы в мтДНК диатомей были приобретены независимо друг от друга. Впервые для диатомей обнаружено, что гены *atpF* и *atpD* не перекрываются в хлоропластном геноме *S. acus*. У *S. acus* выявлен ядерный ген *acpp*, кодирующий ацил-переносающий белок с хлоропластной локализацией. В рамках работы проведено освоение новой методики массового параллельного секвенирования с применением платформы Roche GS FLX и обработки данных, получаемых в результате пиросеквенирования.

Полученные данные имеют фундаментальное значение для выяснения вопросов филогении и эволюции диатомей и могут быть использованы в курсах лекций по молекулярной биологии для студентов университетов.

Апробация работы и публикации. По материалам диссертации опубликовано 7 печатных работ, из них 3 статьи в рецензируемых журналах. Результаты работы были представлены на международной конференции MolPhy2 (Москва, 2010) и на заседании Президиума ИНЦ (Иркутск, 2011).

Структура и объем работы. Диссертация состоит из введения, четырех глав, включающих обзор литературы, описание материалов и методов, изложение результатов и их обсуждение, выводов, списка цитированной литературы и четырех приложений. Работа изложена на 158 страницах, содержит 18 рисунков и 5 таблиц. Библиография включает 156 источников.

СОДЕРЖАНИЕ РАБОТЫ

Определение нуклеотидной последовательности и аннотация митохондриального генома *S. acus*

Предварительную сборку последовательности митохондриального генома получили при анализе библиотеки случайных фрагментов, приготовленной из

суммарной геномной ДНК ксеничной культуры *S. acus*, на приборе GS FLX (Roche Life Sciences, США) в Центре «Биоинженерия» (РАН, Москва). В результате запуска секвенатора было получено ~500 тыс. чтений со средней длиной 230 н.о. Сборку последовательностей чтений в контиги проводили с использованием программы GS De Novo Assembler v.2.1. Контиг протяженностью 43 384 п.н. был выявлен в сборке случайных фрагментов по значимым протяженным BLAST-совпадениям с известными мтДНК одноклеточных водорослей. Среднее значение степени перекрытия митохондриального контига составило 20. Оставшаяся область мтДНК размером 3,3 тыс. п.н., содержащая участок повторов, была амплифицирована с использованием праймеров, сайты гибридизации которых расположены во фланкирующих уникальных областях митохондриального контига, и секвенирована с помощью метода блуждающей затравки на приборе ABI 3730 («Applied Biosystems», США) в ООО «Евроген». Полученные контиги были объединены в кольцевую молекулу. Полная последовательность мтДНК *S. acus* доступна в GenBank (NC_013710).

Аннотацию белок-кодирующих генов проводили путем поиска в мтДНК открытых рамок считывания (ОРС) и их последующего BLAST-анализа по протоколу blastp среди аннотированных белков базы данных NR. Целостность белок-кодирующих генов проверяли вручную на основе сходства с близкородственными последовательностями. Поиск генов тРНК осуществляли с помощью программы tRNAscan-SE 1.23. Гены, кодирующие рРНК малой и большой субъединицы рибосомы, выявляли с помощью программы RNAmmer и на основе сходства *Synedra*-специфических последовательностей с рРНК диатомеи *Thalassiosira pseudonana* и других близкородственных организмов.

Основные свойства митохондриального генома *S. acus*

Митохондриальный геном *S. acus* (рис. 1А) является кольцевой молекулой и имеет размер 46 657 п.н. Он кодирует гены рРНК малой (*rns*) и большой (*rnl*) субъединиц рибосомы, 24 гена тРНК и 34 белок-кодирующих гена, а также два псевдогена рибосомных белков. В геноме имеется протяженная некодирующая область с повторяющимися элементами (рис. 1В). Набор тРНК *S. acus* достаточен для связывания всех природных аминокислот кроме треонина. В мтДНК *S. acus* обнаружены 34 белок-кодирующих гена, среди которых 11 генов рибосомных белков, три гена, кодирующих субъединицы циклооксигеназы, ген апоцитохрома В, 10 генов, кодирующих субъединицы NADH-дегидрогеназы, три гена, кодирующих субъединицы АТФ-азы и ген твин-аргинин транслокационной машины. Кроме того, мтДНК *S. acus* кодирует две ОРС внутри интронов гена *cox1* и три уникальных ОРС с неизвестными функциями.

Двумя наиболее протяженными некодирующими участками в мтДНК *S.acus* являются область размером 4 955 п.н., которая находится между 5'-концом гена *trnM(cau)f* и 3'-концом гена *nad2*, а также участок протяженностью 658 п.н. без выраженных структурных особенностей, который расположен между 5'-концом гена *cox2* и 3'-концом *trnM(cau)f* (рис. 1В). Наиболее крупный кластер (а) содержит два несовершенных прямых повтора протяженностью 1 450 и 1 316 п.н., разделенных 256-

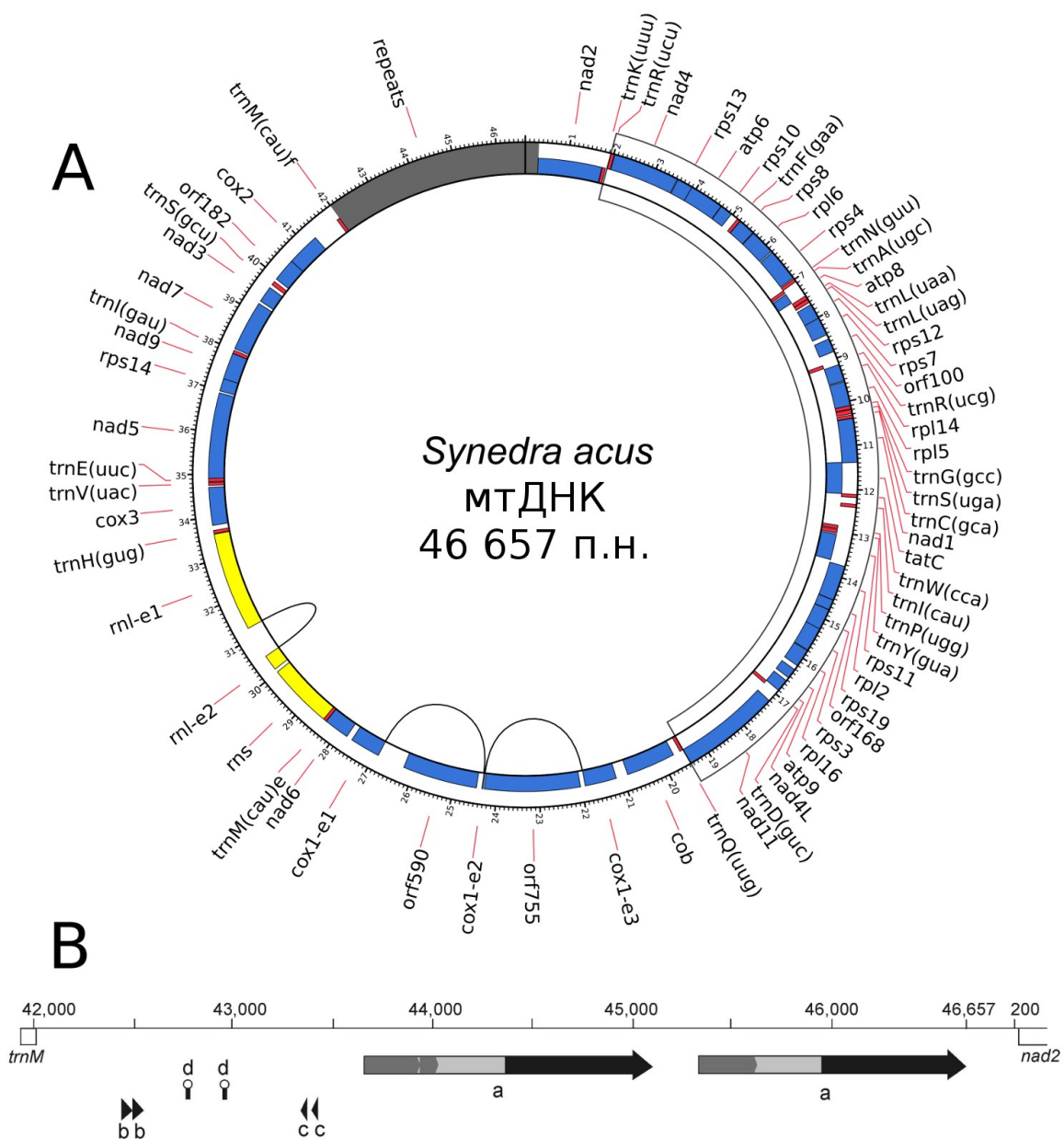


Рис. 1. Митохондриальный геном *S. acus*. **A** – геномная карта мтДНК *S. acus*. Гены рибосомных РНК обозначены желтыми блоками, гены транспортных РНК – красными и белков – синими. Гены транспортных РНК содержат в названии однобуквенный код аминокислоты и последовательность антикодона в скобках. Дуги соединяют экзоны интрон-содержащих генов. Гены, расположенные на внешней/внутренней окружностях геномной карты, транскрибируются по/против часовой стрелки, соответственно. Рамкой выделен блок генов, порядок в котором хорошо сохранился в мтДНК диатомей. **B** – схема некодирующей области мтДНК *S. acus* (42 005-301 п.н.). Стрелки и треугольники показывают относительное расположение повторяющихся участков последовательностей и их ориентацию. Сходство участков (*a*, *b*, *c*) обозначено интенсивностью заливки фигуры: *черный* – 100% сходство, *темно-серый* – сходство более 90%, *светло-серый* – сходство менее 90%. Последовательность из области *c* – входит в участок *b*. Два участка *d* несут инвертированные повторы и могут формировать шпилечные структуры.

нуклеотидной уникальной последовательностью. Оба повтора содержат последовательности, имеющие высокое сходство в терминальных областях. Тандемная дупликация участка протяженностью ~100 п.н. в первой копии повтора является причиной различия длины повторов. Кроме двух длинных повторов, межгенная

область *trnM-nad2* несет несколько относительно коротких прямых повторов, имеющих протяженность менее 50 п.н. (*b-b* и *c-c*), причем *b-b* – это инверсия *c-c*. Повторы *d* содержат короткие палиндромные последовательности, способные формировать шпилечные структуры.

Сравнительно-геномный анализ мтДНК диатомей

Отсутствие гена транспортной РНК для распознавания треонинового кодона характерно для всех мтДНК разножгутиковых организмов (Gray, Lang, Burger, 2004). В отличие от *T. pseudonana*, в мтДНК *P. tricornutum* и *S. acus* отсутствует ген *trnW(uca)*, который отвечает за трансляцию кодона TGA в триптофан во многих митохондриальных генах *T. pseudonana* и других водорослей. Набор белок-кодирующих генов у *S. acus* сходен с таковым в мтДНК других диатомей (табл. 1). Митохондриальные геномы диатомовых водорослей несут 59 общих генов, среди которых 32 белок-кодирующих гена, 24 гена транспортных РНК, два гена рибосомных РНК и одна ОРС в интроне гена *cox1*. Однако, в отличие от двух других диатомей, в мтДНК *S. acus* отсутствует ген *rps2*, а гены *rps3* и *rps7* присутствуют в виде псевдогенов, имеющих сдвиги рамки считывания и делеции. Вероятно, данные гены были перенесены в ядерный геном *S. acus*, и их митохондриальные версии утрачены или находятся в процессе утраты. Этот факт согласуется с тем, что для генов рибосомных белков обнаруживается самый высокий уровень элиминации из митохондриального генома у покрытосеменных растений (Adams, Palmer, 2003).

Судя по структуре, области повторов в мтДНК разножгутиковых организмов возникли в результате ошибок при репликации по механизму *slipped-strand mispairing*. Маловероятно, что области повторов были унаследованы диатомеями от общего предка или произошли в результате встройки мобильных элементов в мтДНК.

Табл. 1. Сравнение митохондриальных геномов диатомовых водорослей

порядок дивергенции



Характеристика	<i>T. pseudonana</i>	<i>S. acus</i>	<i>P. tricornutum</i>
Размер генома (тыс. п.н.)	43,8	46,6	77,4
Размер области повторов (тыс. п.н.)	3,8	4,9	36
Количество белок-кодирующих генов	35	34	32
Количество псевдогенов	не известно	2	не известно
Количество ОРС	1	3	не известно
Количество интронов	1 (в <i>cox1</i>)	3 (2 в <i>cox1</i> , 1 в <i>rnl</i>)	4 (2 в <i>cox1</i> , 1 в <i>rnl</i> , 1 в <i>rns</i>)
Количество ОРС в интронах	1	2	2
рРНК	<i>rns, rnl</i>	<i>rns, rnl</i>	<i>rns, rnl</i>
Количество генов тРНК	25	24	24
<i>trnT</i>	нет	нет	нет
<i>trnW(uca)</i>	есть	нет	нет

Анализ геномных перестановок в мтДНК диатомовых и бурых водорослей

Наборы генов в мтДНК трех диатомовых водорослей сходны (табл. 1), однако порядок генов отличается. Сравнение порядка генов показывает, что в мтДНК имеется два блока, которые отличаются по количеству геномных перестановок. Порядок генов в пределах первого блока *trnR*-*ucu* – *nad11* (рис. 1А) хорошо сохранен. Внутри второго блока *trnQ*(*uug*) – *trnM*(*cau*) гены в мтДНК диатомей переставлены. Кроме того, у *S. acus* и *P. tricorutum* гены в этой области являются котранскрибируемыми блоками, а у *T. pseudonana* они представлены двумя разнонаправленными транскрипционными единицами. Апоморфными признаками для мтДНК *S. acus* являются делеция *rps2*, транслокация *trnK*(*uuu*) и встройка *orf100* и *orf160*.

Генетическое расстояние, рассчитанное с помощью внесения разрывов в последовательность генов мтДНК, принимало значения в диапазоне 27-39 «разрывов» при попарном сравнении мтДНК диатомей (рис. 2А). В то же время, при попарном сравнении мтДНК бурых водорослей этот показатель находился в диапазоне 3-21 «разрывов» (рис. 2В), причем максимальное отличие наблюдалось для митохондриального генома *D. dichotoma* (20-21). Отличия остальных мтДНК друг относительно друга варьировало в пределах 0-6 «разрывов». Сходная картина имеет место при подсчете генетических расстояний между мтДНК на основании внесения инверсий в последовательность генетических маркеров (рис. 2).

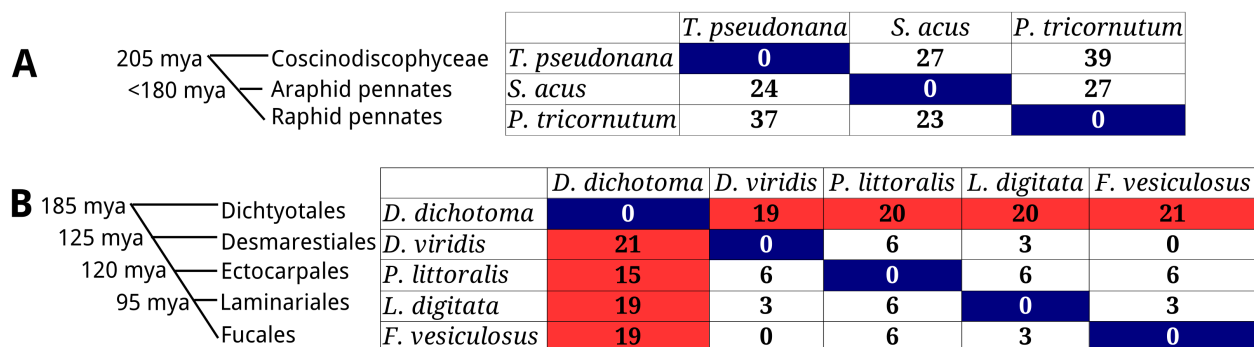


Рис. 2. Генетическая дистанция между мтДНК диатомовых (А) и бурых (В) водорослей. Значения над диагональю – дистанция, подсчитанная с помощью внесения разрывов, под диагональю – на основании инверсий. Филограммы слева от таблиц показывают порядок и время дивергенции диатомей (по Medlin, 2009) и бурых водорослей (по Silberfeld *et al.*, 2010).

Сопоставление геномных перестроек в мтДНК и времени дивергенции отдельных линий бурых водорослей позволяет предположить, что наибольшее количество перестановок генов произошло в линии *Dichtyotales* после ее дивергенции от основной линии *Phaeophyta*. У остальных бурых водорослей порядок генов в мтДНК является консервативным признаком. Раннее время дивергенции пеннатных диатомовых водорослей не дает возможности проверить, характерна ли для мтДНК диатомей сходная картина. Тем не менее, очевидно, что геномные перестройки протекали более интенсивно в мтДНК диатомей, чем у бурых водорослей (рис. 2). Включение в сравнительный анализ митохондриальных геномов других диатомовых водорослей позволит восстановить последовательность структурных перестроек митохондриальных геномов в ходе их эволюции.

Филогенетический анализ последовательностей ОТ-доменов из интронов группы II

В мтДНК *S. acus* обнаружены три интрона группы II (рис. 1А). Два интрона локализованы в гене *cox1* и имеют длину 2 564 и 2 416 п.н. Третий интрон протяженностью 778 п.н. находится в гене *rnl*, кодирующем рРНК большой субъединицы рибосомы. Внутри интронных ОРС с помощью BLAST-поиска по протоколу blastp и НММ-поиска в Pfam были локализованы участки, имеющие сходство с доменами, которые ассоциированы с активностями обратной транскриптазы и матуразы. В интроне гена *rnl* ОРС не обнаружено. BLAST-поиск показал, что *orf755* и *orf590* *S. acus* имеют наибольшее сходство с ОТ-белками из интронов *cox1* бурых водорослей *Pylaiella littoralis* и *Pavlova lutheri*, соответственно. Координаты интронов внутри генов *cox1* не совпадают в мтДНК разных диатомей. Последовательности ОТ-доменов из интронов диатомей не формируют отдельной ветви при филогенетическом анализе (рис. 3).

Vacillariophyta-специфические последовательности разбросаны среди ОТ-доменов разнообразных эукариот. Как и по результатам BLAST-поиска, *orf590* *S. acus* входит в кластер с ОТ-доменом из первого интрона *cox1* бурой водоросли *P. lutheri*, а *orf755* – с ОТ-доменом из второго интрона гена *cox1* *P. littoralis*.

Сборка и аннотация хлоропластного генома *S. acus*

На этапе анализа библиотеки случайных фрагментов была проведена черновая сборка последовательностей, полученных в двух запусках Roche GS FLX. С помощью BLAST-анализа по протоколу blastn в сборке последовательностей были выявлены два хлоропласт-специфических контига длиной 79 362 и 40 761 п.н., которые имели сходство с известными последовательностями хлоропластных геномов диатомей. Для улучшения качества полученных последовательностей на отобранные контиги в процессе референсной сборки были картированы короткие последовательности Illumina. При инспектировании полученной гибридной сборки оказалось, что результирующие контиги имели протяженное перекрытие с друг с другом. На основании этого они были объединены с образованием суперконтига длиной 111 170 п.н. Концы полученной последовательности имели протяженные участки перекрытия с фланкирующими областями одного из плеч обратного повтора, локализованного с помощью картирования генов рРНК. Это сделало возможным проведение окончательной сборки *in silico*. Результирующие файлы сборки хлоропласт-специфических контигов предоставляются по запросу (<mailto:yuri.galachyants@lin.irk.ru>).

Аннотацию белок-кодирующих генов, генов рРНК и тРНК осуществляли так же, как и для мтДНК *S. acus*. Поиск генов транспортно-информационных РНК (тиРНК) и генов частицы узнавания сигнала проводили с использованием программ ARAGORN и SRPscan. Полная последовательность хлоропластной ДНК *S. acus* доступна в GenBank (JQ088178).

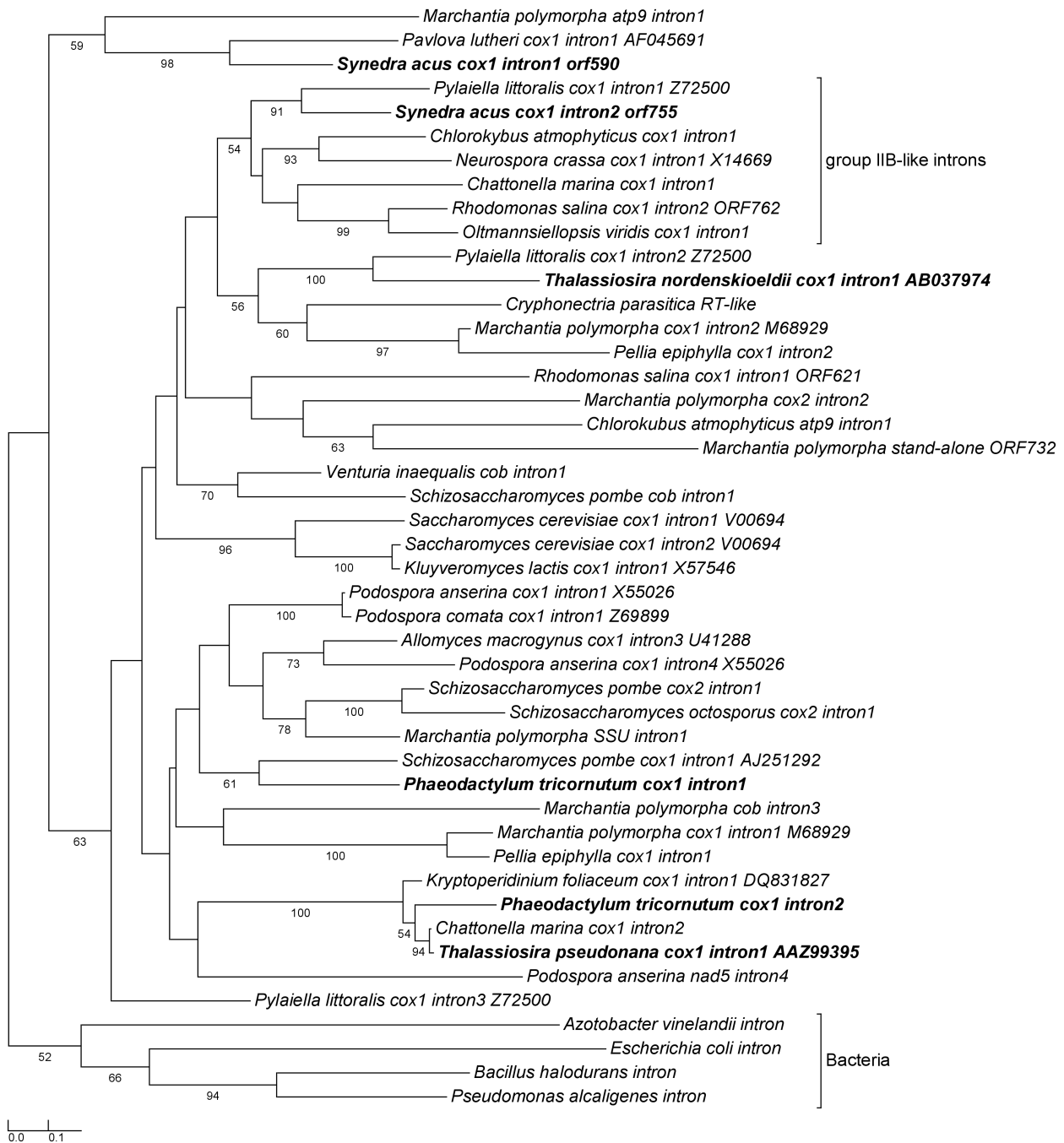


Рис. 3. Филогенетическое дерево, основанное на анализе аминокислотных последовательностей обратнo-транскриптазных доменов из интронов группы II, обнаруженных в мтДНК различных организмов. Модель LG+Г. Последовательности диатомовых водорослей выделены жирным шрифтом. Значения индексов поддержки ниже 50% не показаны.

Основные свойства хлоропластного генома *S. acus*

Хлоропластный геном *S. acus* является кольцевой молекулой и имеет размер 116 251 п.н. (рис. 4). Геном характеризуется как каноническая четырехсоставная структура, в которой два плеча инвертированного повтора IRa и IRb, несущих локусы генов рРНК, находятся между большой (LSC) и малой (SSC) уникальными областями. В хпДНК *S. acus* обнаружено 160 генов, среди которых пять генов с альтернативными

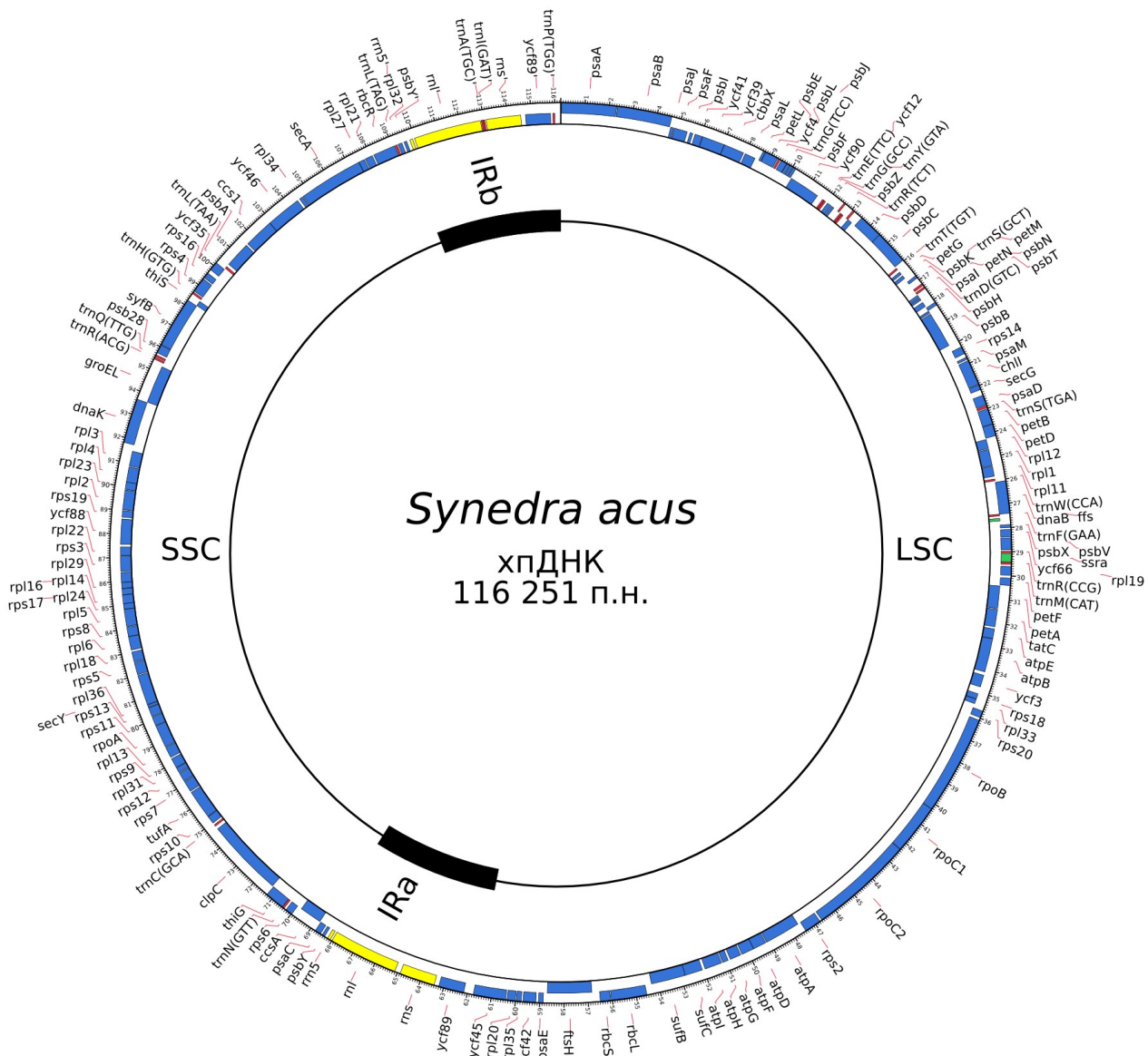


Рис. 4. Геномная карта хлоропластной ДНК *S. acus*. Гены рибосомных РНК обозначены желтыми блоками, гены транспортных РНК – красными, гены малых РНК – зелеными и гены белков – синими. Гены транспортных РНК содержат в названии однобуквенный код и последовательность антикодона в скобках. Гены, расположенные на внешней/внутренней окружностях геномной карты, транскрибируются по/против часовой стрелки, соответственно. На внутреннем круге схематически представлены элементы четырехсоставной структуры – плечи инвертированного повтора IRa и IRb, находящиеся между большой (LSC) и малой (SSC) уникальными областями хпДНК.

старт-кодонами: в *rbcS*, *ccsA*, *rps8* и *rpl23* используется GTG, а в *secY* – ATT (табл. 2). Терминирующий кодон TAA используется в 117 генах. В десяти генах стоп-кодоном является TTA, а в одном – TGA.

Сравнительно-геномный анализ хлоропластных геномов диатомовых водорослей

Хлоропластный геном *S. acus* имеет несколько меньший размер, чем другие хпДНК диатомей, тогда как GC-состав и кодирующая емкость геномов сходны. Все пластидные геномы диатомовых водорослей компактны, в них отсутствуют интроны. Как и у других диатомей, в хпДНК *S. acus* между генами *rpl14* и *rpl24* нет межгенного спейсера. Сходным образом, в геноме присутствуют три пары идентично

Табл. 2. Набор генов в хлоропластном геноме *S. acus*

Гены, не кодирующие белки (32)	
Рибосомные РНК (3)	<i>rnl¹, rns¹, rrn5¹</i>
Транспортные РНК (27)	<i>trnA(ucg)¹, trnC(gca), trnD(guc), trnE(uuc), trnF(gaa), trnG(gcc), trnG(ucc), trnH(gug), trnI(cau), trnI(gau)¹, trnK(uuu), trnL(caa), trnL(uaa), trnM(cau)f, trnM(cau)e, trnN(guu), trnP(ugg)¹, trnQ(uug), trnR(acg), trnR(ccg), trnR(ucu), trnS(gcu), trnS(uga), trnT(ugu), trnV(uac), trnW(cca), trnY(gua)</i>
Малые РНК (2)	<i>ffs, ssra</i>
Белок-кодирующие гены (128)	
Гены рибосомных белков (44)	
Малая субъединица (18)	<i>rps2-rps14, rps16-rps20</i>
Большая субъединица (26)	<i>rpl1-rpl6, rpl11-rpl14, rpl16, rpl18-rpl24, rpl27, rpl29, rpl31-rpl36</i>
Гены фотосинтеза (44)	
АТФ-синтаза (8)	<i>atpA, atpB, atpD, atpE, atpF, atpG, atpH, atpI</i>
Фотосистема I (10)	<i>psaA, psaB, psaC, psaD, psaE, psaF, psaI, psaJ, psaL, psaM</i>
Фотосистема II (18)	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT, psbV, psbX, psbY^{1,2}, psbZ (ycf9), psb28</i>
b6/f-комплекс (8)	<i>petA, petB, petD, petF², petG, petL (ycf7), petM (ycf31), petN (ycf6)</i>
Гены других процессов (40)	
Метаболизм (5)	<i>chlI, rbcL, rbcS, thiG², thiS² (ycf40)</i>
Фолдинг (5)	<i>clpC, dnaB, dnaK, fisH (ycf25), groEL</i>
Сборка, транслокация (8)	<i>ccs1 (ycf44), ccsA, secA, secG (ycf47), secY, sufB (ycf24), sufC (ycf16), tatC (ycf43)</i>
Транскрипция (6)	<i>cbbX, rbcR, rpoA, rpoB, rpoC1, rpoC2</i>
Трансляция (2)	<i>syfB², tufA</i>
ОРС (14)	<i>ycf3, ycf4, ycf12, ycf33, ycf35, ycf39, ycf41, ycf42², ycf45, ycf46, ycf66, ycf88, ycf89^{1,2}, ycf90</i>

перекрывающихся генов: *sufC-sufB* (1 п.н.), *rpl4-rpl23* (8 п.н.) и *psbD-psbC* (53 п.н.). В отличие от других диатомей, гены *atpD* и *atpF* не перекрываются, а располагаются друг за другом.

С точки зрения структурной организации, хпДНК диатомей не отличаются от хлоропластных геномов большинства других разножгутиковых организмов и характеризуются как четырехсоставные кольцевые молекулы ДНК. Такая структура характерна и для хлоропластных геномов, унаследованных от зеленых водорослей, в отличие от хпДНК красных водорослей, у которых отсутствуют области обратных повторов. По кодирующему потенциалу хпДНК диатомей имеют большее сходство с

¹ Гены находятся в инвертированных повторах.

² Данные гены не входят в набор 154 генов, общих для хпДНК диатомовых водорослей.

пластидными геномами красных водорослей, нежели с хлоропластами «зеленой» линии (Allen *et al.*, 2008) и несут около 160 генов, из них 154 гена являются общими для диатомей. В составе генов в хлоропластных геномах диатомей имеются различия (табл. 3).

Перестановки генов в хлоропластных геномах диатомовых водорослей

Визуальный анализ геномных перестановок показывает, что наибольшее сходство порядка генов в пределах LSC наблюдается в группе хпДНК диатомей *S. acus* и *P. tricornutum*, а также эндосимбионтов динофлагеллят *D. baltica* и *K. foliaceum*. По сравнению с геномами других пеннатных диатомей, в хпДНК *Fistulifera* sp. имеется инверсия протяженного участка генов *rbcS-rps14*. Пластидные геномы видов *Thalassiosira* отличаются друг от друга наличием нескольких крупных инверсий. Порядок генов в области SSC претерпел меньшие изменения. У пеннатных диатомей в SSC имеется два консервативных блока генов (*syfB-rbcS* и *clpC-dnaK*), в последнем из которых большинство генов находятся в пределах котранскрибируемой области. Интересно отметить, что в хпДНК *Thalassiosira* большая часть блока *syfB-rbcR* перенесена в область инвертированных повторов. При анализе геномных перестановок в хпДНК диатомей обнаруживается, что в терминальных областях однокопийных участков хпДНК геномные перестановки происходили менее интенсивно. Это можно объяснить структурными ограничениями, препятствующими обмену генами между LSC- и SSC-областями и связанными с образованием «перехвата» кольцевой молекулы хпДНК в области инвертированных повторов. Судя по наблюдаемой картине, геномные перестановки в хпДНК преимущественно происходили в результате инверсий в пределах уникальных областей.

Табл. 3. Различия наборов генов в хпДНК диатомовых водорослей и эндосимбионтов динофлагеллят *K. foliaceum* и *D. baltica*

порядок дивергенции 

Название	Описание	TP	TO	OS	SA	PT	Fsp	KF	DB
<i>acpp</i>	ацил-переносающий белок	-	-	+	-	+	-	-	-
<i>flrn</i>	<i>ffs</i> -подобная РНК	-	+	-	-	-	-	-	-
<i>orf127</i>	гипотетический белок	-	+	-	-	-	-	-	-
<i>petF</i>	VII субъединица комплекса b6/f	+	-	+	+	+	+	+	+
<i>psbI</i>	I белок фотосистемы II	+	+	+	+	+	-	+	+
<i>serc1</i>	предполагаемая сериновая рекомбиназа	-	-	-	-	-	-	+	-
<i>serc2</i>	предполагаемая сериновая рекомбиназа	-	-	-	-	-	+	+	-
<i>syfB</i>	β -субъединица фенилаланил-тРНК-синтазы	-	-	-	+	+	+	+	+
<i>thiG</i>	G-белок пути биосинтеза тиамин	+	+	+	+	+	-	+	+
<i>thiS</i>	S-белок пути биосинтеза тиамин	+	+	+	+	+	-	+	+
<i>tsf</i>	фактор элонгации Ts	-	-	-	-	+	+	-	-
<i>tyrC</i>	предполагаемая интеграз/рекомбиназа	-	-	-	-	-	-	+	-
<i>ycf42</i>	консервативная OPC; предположительно пероксиредоксин	+	+	+	+	+	+	-	-

В хпДНК диатомей и диатомовых водорослей имеется 154 общих гена (см. табл. 2). «+» – ген присутствует в хлоропластном геноме, «-» – отсутствует в хлоропластном геноме. TP – *T. pseudonana*, TO – *T. oceanica*, OS – *O. sinensis*, SA – *S. acus*, PT – *P. tricornutum*, Fsp – *Fistulifera* sp., KF – эндосимбионт *K. foliaceum*, DB – эндосимбионт *D. baltica*.

Топология дерева, основанная на анализе порядка генов в хпДНК диатомовых водорослей с использованием метода максимальной экономии (**рис. 5**), согласуется с порядком ветвления основных групп диатомей, выявленным с использованием общепринятых генетических маркеров. Это свидетельствует о применимости порядка генов в хпДНК диатомей для филогенетического анализа. Хотя дистанция между хлоропластными геномами эндосимбиотов динофлагеллят составляет 10 инверсий согласно анализу порядка генов, они дивергируют единым кластером от шовных пеннатных диатомей, что согласуется с исходными результатами сравнительно-геномного анализа (Imanian, Pombert, Keeling, 2010). Это дополнительно свидетельствует в пользу того, что эндосимбиоты *K. foliaceum* и *D. baltica* являются близкородственными организмами. Вероятно, данная ветвь динофлагеллят произошла в результате единичного события третичного эндосимбиоза представителя Dinophyceae с пеннатной шовной диатомовой водорослью.

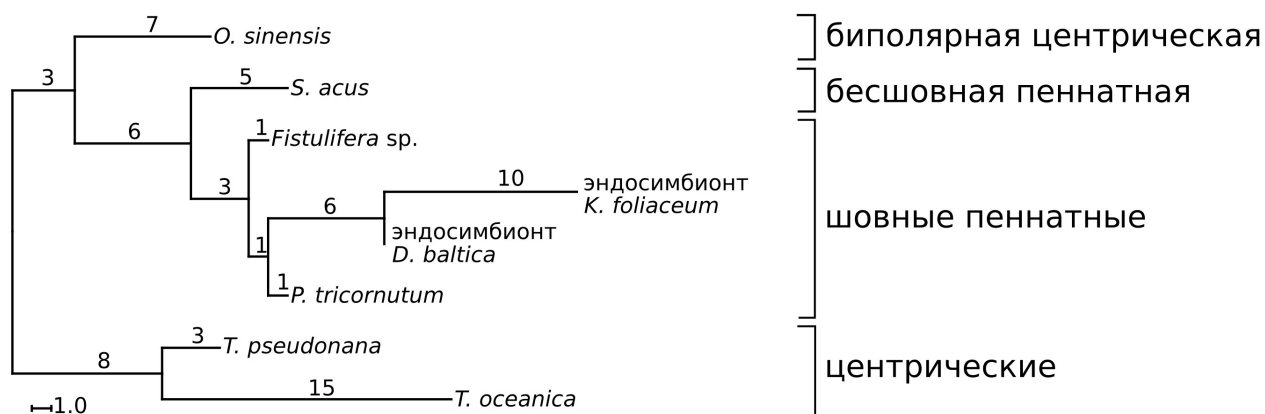


Рис. 5. Дерево, полученное на основе анализа порядка генов в хпДНК диатомовых водорослей и эндосимбиотов динофлагеллят *D. baltica* и *K. foliaceum*. Цифры над ветвями указывают количество инверсий, необходимое для получения последовательности генетических маркеров, реконструированной в проксимальном узле дерева из последовательности в терминальном узле.

Структурно-функциональный и филогенетический анализ гена *atpF*

Ген *atpF* кодирует б-субъединицу АТФ-синтазного комплекса, вторичная структура которой представляет собой протяженную α -спираль (Weber, 2006). Сравнительно-геномный анализ не обнаружил перекрытия хлоропластных генов *atpF* и *atpD* в геноме *S. acus*. Поскольку в других хпДНК диатомей эти гены перекрываются и протяженность перекрытия одинакова, то данный признак является автоапоморфным для хлоропластного генома *S. acus*. Объединение последовательностей диатомовых водорослей при анализе генетических расстояний (**рис. 6**) и формирование единой клады при филогенетическом анализе дает право предположить общее происхождение последовательностей *atpF* диатомей и дивергенцию *Synedra*-специфической последовательности *atpF*. Важно отметить, что при анализе *atpF* *S. acus* с помощью инструмента Pfam hmmsearch, обнаруживается значимое совпадение последовательности с выравниванием семейства АТВ-synt_B (E-значение 4.0×10^{-26}), которое наблюдается на всем протяжении ОРС гена.

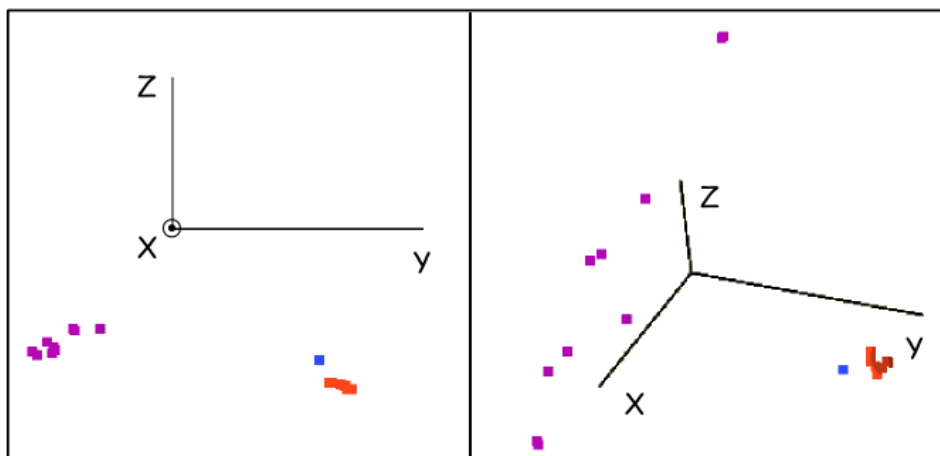


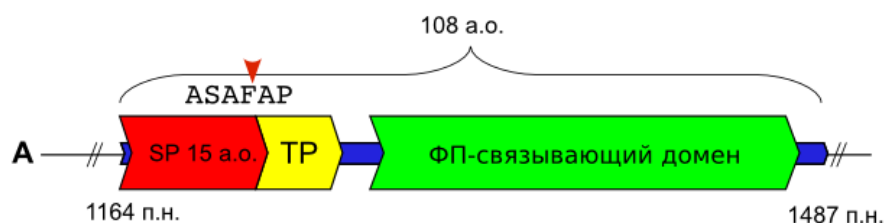
Рис. 6. Анализ выравнивания последовательностей *atpF* методом главных компонент, BLOSUM62; последовательность *S. acus* отмечена синим цветом, последовательности остальных диатомей – красным цветом, последовательности *atpF* Rhodophyta, Cryptophyta и Haptophyta – фиолетовым цветом.

В контигах ядерного генома *S. acus* последовательности, сходные с геном *atpF*, не обнаружены. Все эти факты свидетельствуют о том, что последовательность гена *atpF*, показывая значительную степень дивергенции, все-таки подвержена действию стабилизирующего отбора и, вероятнее всего, белковый продукт гена является функциональным. Тем не менее, в настоящий момент остаются непонятными причины дивергенции гена *atpF* *S. acus* и период времени, в течение которого происходили изменения последовательности.

Структурно-функциональный анализ гена *acpp*, продукт которого имеет хлоропластную локализацию

Ген *acpp* кодирует ацил-переносящий белок, простетическая группа которого служит переносчиком ацильных остатков с молекулы ацетил-коэнзима-А на растущую цепь жирной кислоты. В хлоропластном геноме *S. acus* отсутствует ген *acpp* (табл. 3). При поиске сходных последовательностей в сборке ядерного генома *S. acus* был обнаружен контиг с ОРС, имеющей значимое совпадение с последовательностью хлоропластного *acpp* *P. tricornutum*. Дальнейший анализ показал, что аминокислотная последовательность *acpp* *S. acus* несет двусоставной сигнал плазмидной локализации (рис. 7А), ответственный за транслокацию полипептида в строму хлоропласта. При анализе аминокислотной последовательности гена *acpp* *S. acus* с помощью приложения NESTAR N-концевая и транзитная часть сигнального пептида были предсказаны с высокой надежностью. Результаты выявления сигнального пептида с помощью программы SignalP также подтверждают наличие сигнального пептида и сайта его отщепления между Ala и Phe мотива ASAFAP в 15-ой позиции последовательности АПБ *S. acus*.

Сравнение последовательности *acpp* *S. acus* с выровненными последовательностями фосфоантетеин-связывающих доменов с использованием НММ Pfam выявило значимое совпадение профиля домена (рис. 7В). В центральной части последовательности присутствует консервативный мотив DSL, боковая группа



В Домен связывания фосфоантетеина

Семейство	Клан	Выравнивание		HMM		Bit-значение	E-значение
		Начало	Конец	От	До		
PP-binding	CL0314	38	104	2	67	41.3	1.3E-10

```

#HMM      kllleivaevlgldpdeildtdlfe.lglDSLlaveliarleeeefgveipvedlfeipptlaelaayl
#MATCH    +1++i+a++l +d  +  +++ + l  DSLl++ve++ +++efg+ei  e  +  +++++yl
#PP       699*****8889*****966*****
#SEQ      TLQOIADQLDVKAKVVKKEASFDdLDA DSL DVVEMVMSIEETFGIEIEDEAASSIKKVGDCYL
  
```

Рис. 7. Структурно-функциональный анализ последовательности гена *acpp*. **А** – схема структурных элементов *acpp* *S. acus* в пределах ОПС гена (показана синим цветом). Красный блок – сигнальный пептид (SP) для транслокации в шЭПР, желтый блок – транзитный пептид (TP) для транслокации в перипластидное пространство, зеленый блок – домен связывания фосфоантетеина. Стрелкой показан сайт отщепления сигнального пептида. **В** – сравнение последовательности *acpp* *S. acus* с выравниванием фосфоантетеин-связывающих доменов с использованием HMM Pfam. Рамкой выделен консервативный мотив DSL.

Сер которого служит для ковалентного связывания простетической группы фосфоантетеина при формировании активной молекулы голофермента.

Таким образом, структурно-функциональный анализ показывает, что АПБ, кодируемый последовательностью ядерного гена *acpp* *S. acus*, имеет все необходимые элементы для транслокации АПБ в строму хлоропласта и участия белка в биосинтезе жирных кислот.

Поиск сходных последовательностей в базе данных NR показал, что последовательность *acpp* *S. acus* не кластеризуется с последовательностями *P. tricornutum* и *O. sinensis* при использовании в качестве метрики баллов BLAST-совпадений. Напротив, для АПБ *S. acus* наиболее родственными являются гены δ -протеобактерий *Desulfovibrio* и *Syntrophobacter*, тогда как с хлоропластными последовательностями *acpp* диатомей *P. tricornutum* и *O. sinensis* наиболее сходны апикопластные последовательности, принадлежащие представителям группы гетеротрофных динофлагеллят *Aricomplexa*, у которых хлоропласт редуцирован и не выполняет функции фотосинтеза.

Можно предположить, что *Synedra*-специфический *acpp* был приобретен в результате горизонтального переноса гена от δ -протеобактерий, а хлоропластная копия была гена утрачена. С другой стороны, возможен сценарий, при котором сходство *acpp* *S. acus* с бактериальными последовательностями АПБ является конвергентным, и возникло в ходе интенсивной эволюции гена в процессе его интеграции в ядерный геном. Справедливость данных предположений можно будет проверить при определении последовательностей хлоропластных геномов, родственных *S. acus*. Учитывая современную парадигму дивергенции диатомовых водорослей, маловероятно, что перенос гена *acpp* в ядерный геном является следствием единичного события, которое имело место при эволюции древнего предка диатомовых водорослей.

ВЫВОДЫ

1. С применением методов массового параллельного секвенирования ДНК определены нуклеотидные последовательности митохондриального и хлоропластного геномов пресноводной пеннатной бесшовной диатомовой водоросли *Synedra acus* subsp. *radians*.
2. Митохондриальный геном *S. acus* subsp. *radians* имеет размер 46 657 п.н. и характеризуется наличием участка повторов протяженностью 4,2 тыс. п.н. В геноме присутствует 37 генов, отвечающих за транскрипцию и трансляцию митохондриальных белков (два гена рРНК, 24 гена тРНК и 11 генов рибосомных белков), 17 генов, продукты которых входят в дыхательную цепь, 6 генов с другими белок-кодирующими функциями и 2 псевдогена рибосомных белков.
3. Хлоропластный геном *S. acus* subsp. *radians* имеет размер 116 251 п.н. и содержит 160 генов, включая 83 гена, отвечающих за транскрипцию и трансляцию хлоропластных белков (три гена рРНК, 27 генов тРНК, 44 гена рибосомных белков, ген тиРНК и 8 генов, кодирующих субъединицы РНК-полимеразы, тРНК-фенилаланил-синтазу, а также кофакторы транскрипции и трансляции), 44 гена, кодирующих субъединицы АТФ-синтазы, цитохромового b6/f-комплекса и фотосистем I и II, 14 генов, продукты которых участвуют в транслокации и фолдинге хлоропластных белков и 19 других белок-кодирующих генов.
4. Сравнительный анализ митохондриальных геномов трех видов диатомей показал, что мтДНК сходны по кодирующей емкости и несут 59 общих генетических маркеров. Выявлены блок с консервативным порядком генов (*trnR*-уцу – *nad11*) и блок, в котором порядок генов различен (*trnQ*(ууг) – *trnM*(кау)). Показано, что геномные перестройки в митохондриальных геномах диатомовых водорослей происходили интенсивнее, чем в митохондриальных геномах бурых водорослей.
5. Впервые с помощью филогенетического анализа показано, что интроны группы II, обнаруженные в генах *cox1* мтДНК диатомовых водорослей, имеют независимое происхождение в каждой линии.
6. При сравнении геномов хлоропластов восьми видов диатомовых водорослей обнаружено 154 общих гена. Все хлоропластные геномы диатомей имеют характерную для разножгутиковых организмов четырехсоставную структуру. Дивергенция групп диатомовых водорослей, реконструированная при анализе порядка генов в хлоропластных геномах, согласуется с систематикой крупных таксонов диатомей. Ген *acpp*, который кодирует ацил-переносящий белок АСР, имеющий хлоропластную локализацию, у *S. acus* subsp. *radians* обнаружен в ядерном геноме.
7. На примере хлоропластного генома *S. acus* subsp. *radians* впервые для диатомовых водорослей показано отсутствие перекрытия генов *atpD* и *atpF*, кодирующих β - и δ -субъединицы АТФ-синтазы. Филогенетический анализ показал общее происхождение гена *atpF* у диатомовых водорослей и его дивергенцию у *S. acus* subsp. *radians*.

Основные результаты диссертации опубликованы в следующих работах:

1. Ravin N. V., Galachyants Y. P., Mardanov A. V., Beletsky A. V., Petrova D. P., Sherbakova T. A., Zakharova Y. R., Likhoshway Y. V., Skryabin K. G., Grachev M. A. Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes // Curr. Genet. – 2010. – V. 56. – P. 215-223.
2. Поздняк Е. И., Сидоров И. А., Галачьянц Ю. П. Генерализация алгоритма таксономического классификатора CARMA // Вестник ИрГТУ. – 2011. – Т. 9. – С. 11-15.
3. Galachyants Y. P., Morozov A. A., Mardanov A. V. Beletsky A. V., Ravin N. V., Petrova D. P., Likhoshway Y. V. Complete chloroplast genome sequence of freshwater araphid pennate diatom alga *Synedra acus* from Lake Baikal // Int. J. Biol. – 2012. – V. 4. – P. 27-35.

Благодарности

Автор благодарит директора ЛИН СО РАН, академика Михаила Александровича Грачева, д.б.н. Елену Валентиновну Лихошвай, коллег из отдела Ультраструктуры клетки ЛИН СО РАН, д.б.н. Николая Викторовича Равина и коллег из Центра «Биоинженерия» РАН за ценные замечания, обсуждение и помощь в подготовке работы, к.б.н. Александра Михайловича Мазура из ООО «Геноаналитика» за помощь в получении геномной библиотеки и секвенировании на приборе Illumina GAIIх, к.т.н. Ивана Александровича Сидорова и коллег из Иркутского суперкомпьютерного ЦКП ИДСТУ СО РАН за помощь при компьютерном обсчете данных, а также жену Агнию за терпение и поддержку.

Работа выполнена при поддержке Программы Президиума РАН «Молекулярная и клеточная биология», проект 6.3.