

## Основные результаты лаборатории за 2006-2010 годы по направлению «Информационные процессы на молекулярном и клеточном уровнях»

### 1. Конкуренция РНК-полимераз, транскрибирующих общий локус

Впервые изучена конкуренция РНК-полимераз в компактных геномах, включая пластомы растений и водорослей. Разработана и компьютерно реализована модель системы взаимодействующих случайных процессов, в частности, процессов конкуренции РНК-полимераз. Компьютерная программа <http://lab6.iitp.ru/ru/rivals/> свободно доступна на сайте лаборатории <http://lab6.iitp.ru>. Вычисления проводились в МСЦ РАН на суперкомпьютере МВС-100К с использованием 2048 процессоров.

**1.1** Получено объяснение сложного изменения уровней транскрипции генов в изолированных хлоропластах ячменя на основе конкуренции РНК-полимераз, которое наблюдалось в эксперименте.

**1.2** У *Arabidopsis thaliana* предсказаны интенсивности связывания с промоторами холоферментов РНК-полимераз бактериального типа с различными  $\sigma$ -субъединицами и получено хорошее согласие с экспериментами по нокауту  $\sigma$ -субъединиц Sig3 и Sig4.

Например, конкуренция наблюдается, в локусе *P-psbB-psbT-(psbN-P')-psbH-petB-petD* у *A. thaliana*, в котором ген *psbN* и промотор P' располагаются на комплементарной цепи относительно другого промотора P и других указанных генов. Уровни транскрипции в этом локусе существенно меняются после нокаута  $\sigma$ -субъединицы Sig3. На основе модели конкуренции здесь предсказаны даже терминаторы транскрипции. Аналогичное согласие с экспериментом получено при моделировании конкуренции и нокаута по Sig4 в локусе *P'-P''-P-ycf1-(ndhF-P\*)-rpl32* у *A. thaliana*, в котором ген *ndhF* и промотор P\* находятся на комплементарной цепи относительно других указанных промоторов и генов. Промотор P\* Sig4-зависимый; три промотора перед геном *ycf1* таковы: один – зависимый от РНК-полимеразы бактериального типа и два – от разных РНК-полимераз фагового типа.

**1.3** Описан новый механизм регуляции экспрессии генов, основанный на конкуренции РНК-полимераз. Например, в локусе дивергентно расположенных генов *glnB* и *rps20* в пластидах багрянков. А именно, фактор NtcA чувствителен к концентрации амония в пластиде и закрывает промотор гена *rps20*, тем самым, увеличивая транскрипцию *glnB*. Ген *glnB* кодирует сигнальный белок, взаимодействующий уже с различными белками, попадающими в пластиду из цитоплазмы.

**1.4** Найдено распределение числа длинных шпилек в межгенных трейлерных областях у актинобактерий. Обнаружены многочисленные длинные шпильки (до 31 н. в плече) после интенсивно транскрибируемых генов, в частности, генов многих тРНК. Вероятно, эти шпильки служат терминаторами транскрипции, предотвращающими конкуренцию РНК-полимераз.

## 2. Консервативные и переменные промоторы бактериального типа в пластидах

Разработан эффективный алгоритм распределенных вычислений <http://lab6.iitp.ru/ru/twobox/>, свободно доступный на сайте лаборатории, который впервые позволил выполнить систематический поиск в буквенных последовательностях сигнала с заданной сложной структурой.

**2.1** Проведен поиск промоторов бактериального типа в пластидах растений, водорослей и простейших. Получен результат принципиального характера: в пластидах число консервативных промоторов этого типа на порядок меньше числа консервативных генов. Консервативные (в масштабе Streptophyta) промоторы присутствуют только перед пятью генами *psaA*, *psbA*, *psbB*, *psbE*, *rbcL*. Для гена *psbA* консервативный промотор найден даже в хлоропласте у вторичного эндосимбионта *Bigeloviella natans* из таксона Chlorarachniophyceae. Однако у голосеменного *Cycas taitungensis* и мха *Anthoceros formosae* не удалось найти промотор перед геном *psbA*, а у *Chlorokybus atmophyticus* сохранился только -35 бокс промотора. Промотор перед геном *psbE* предсказан у многих Streptophyta, обладающих геном *psbE*. Среди них большинство цветковых, все доступные голосеменные, папоротникообразные и мохообразные, а также водоросли *Chaetosphaeridium globosum*, *Staurostrum punctulatum* и *Zygnema circumcarinatum*. У водорослей *Chara vulgaris*, *Ch. atmophyticus* и *Mesostigma viride* не найден промотор перед геном *psbE*. Промотор перед геном *rbcL* консервативен у всех наземных растений и у водоросли *Ch. vulgaris*. Консервативный промотор перед геном *psbB* найден у большинства семенных растений, включая голосеменные *C. taitungensis* и *Pinus* spp. Промотор перед геном *psaA* консервативен не только у наземных растений, но и у большинства водорослей таксона Streptophyta, за исключением рано отделившихся видов *Ch. atmophyticus* и *M. viride*.

**2.2** Показана быстрая эволюция промотора бактериального типа пластомного гена *ndhF* у цветковых растений. Пластом считается очень консервативной частью генома растений; мало что известно об эволюции пластомных промоторов. Показано, что ген *ndhF* находится в регионе, который у наземных растений редко подвергался крупным перестройкам. Нами показано однако, что лидерная область гена *ndhF* у цветковых растений имеет быстро эволюционирующий промотор, который представлен, по крайней мере, в четырех вариантах. Предок двух из четырех основных ветвей Magnoliophyta (magnoliids и eudicotyledons) имел один промотор *A*, но в филогенетической ветви rosids его заменил другой промотор *C*, который сохранился, в частности, у крестоцветных и винограда. Возможно, промотор *C* сформировался прежде отделения ветви rosids и в то время перед геном *ndhF* было два промотора *A* и *C*. В некоторых ветвях внутри группы rosids промотор *C* замещился другими промоторами *B* или *D*. Множественные выравнивания лидерных областей выявили, что

предсказанные промоторы часто подвергались мутациям, причем не только заменам нуклеотидов, но и небольшим инсерциям и делециям.

**2.3** Для высших растений набор  $\sigma$ -субъединиц различается даже в близких таксонах, например, у тополя и резушки Таля. Филогенетическое распределение промоторов фиксированного типа согласуется с распределением генов соответствующей  $\sigma$ -субъединицы; например, для промотора *C* перед *ndhF* и Sig4. Обнаружены псевдогены, которые кодировали  $\sigma$ -субъединицы, но как раз у растений, в которых соответствующие промоторы отсутствуют или значительно отличаются от своих консенсусов. Это указывает на чрезвычайную эволюционную пластичность системы, состоящей из  $\sigma$ -субъединицы и специфичного ей промотора.

### **3. Аттенуаторная регуляция транскрипции: классическая и неклассическая. Вэб-сервер для моделирования аттенуаторной регуляции**

*Разработана модель, включающая все этапы классической аттенуаторной регуляции экспрессии генов у бактерий.* Получены алгоритмические и программные решения, которые позволили быстрее, чем время протекания соответствующего внутриклеточного процесса, получать зависимость уровня транскрипции гена от концентрации аминокислоты в клетке. На этой основе создан сервер <http://lab6.iitp.ru/rnamodel/runmodel.php?lang=rus> на сайте лаборатории, который работает в режиме массового обслуживания.

**3.1** Найдены потенциальные аттенуаторные регуляторные структуры разных типов у многих бактерий из таксонов  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ -proteobacteria, Actinobacteria, Bacteroidetes/Chlorobi, Firmicutes и Thermotogae, Chloroflexi перед генами *hisG*, *hisZ*, *hisS*, *pheA*, *pheST*, *trpEG*, *trpA*, *trpB*, *trpE*, *trpS*, *thrA*, *thrS*, *leuA*, *leuS*, *ilvB*, *ilvI*, *ilvA*, *ilvC*, *ilvD*, *ilvG*. Поиск проведен по всем секвенированным бактериям из ГенБанка, NCBI. Не обнаружены аттенуаторные структуры ни у одного вида из таксонов Chlamydiae, Cyanobacteria, Mollicutes,  $\square$ -proteobacteria, Spirochaetales; а также – в хлоропластах водорослей, которые имеют гены синтеза аминокислот. Аттенуаторная регуляция наиболее часто встречается у протеобактерий и актинобактерий, классическая аттенуаторная регуляция наиболее часто встречается у  $\gamma$ -протеобактерий, у которых с ее помощью регулируются наибольшее число видов и разнообразие генов биосинтеза аминокислот и аминоацил-тРНК.

**3.2** Впервые найдены аттенуаторные регуляции новых типов, среди которых, например, случай, когда действие терминатора блокируется его расположением в петле псевдоузла. Впервые показана существенная роль РНК-триплексов и альтернативных цепей спиралей при формировании эффективных антитерминатора и терминатора. Приведем примеры. (1) Перед *hisG* у многих  $\gamma$ -протеобактерий (включая представителей семейств Enterobacteriales, Pasteurellales, Vibrionales, Alteromonadales, Aeromonadales) найден триплекс, который участвует в формировании ко-терминатора (структуры РНК, существующей одновременно с

терминатором). Третье плечо триплекса включает много остатков урацила, что обеспечивает высокую стабильность триплекса независимо от кислотности цитоплазмы. В большинстве случаев здесь складывается триплекс из Pu-Pu-Pu триад. Однако у *Alteromonadales bacterium* и *Pseudoalteromonas haloplanktis* складывается смешанный триплекс CUGU\*GAGG-CCUC, содержащий Pu-Pu-Pu и Pu-Pu-Pu триады. (2) Перед геном *hisZ* у нескольких бацилл (*Bacillus cereus*, *B. thuringiensis*, *B. anthracis*, *B. weihenstephanensis*) образуется слабый цитидилгуаниловый Pu-Pu-Pu триплекс, который участвует в формировании ко-терминатора. Перед геном *hisZ* у *Clostridium difficile* ко-терминатор складывается с участием Pu-Pu-Pu триплекса AAG\*AAG-CUU. (3) Перед геном *lysQ*, кодирующим пермеазу, у *Lactococcus lactis* (Firmicutes) наблюдается классическая аттенуаторная регуляция с гистидиновыми регуляторными кодонами и цепью спиралей, в которой ко-терминатор поддерживается Pu-Pu-Pu триплексом AGA\*AGA-UCU. (4) Для гена *ilvD* у родов *Staphylococcus* и *Listeria* (Firmicutes) в регуляторной области наблюдается цепь спиралей: четыре консервативные спирали с образованием триплекса из Pu-Pu-Pu триад при формировании ко-антитерминатора.

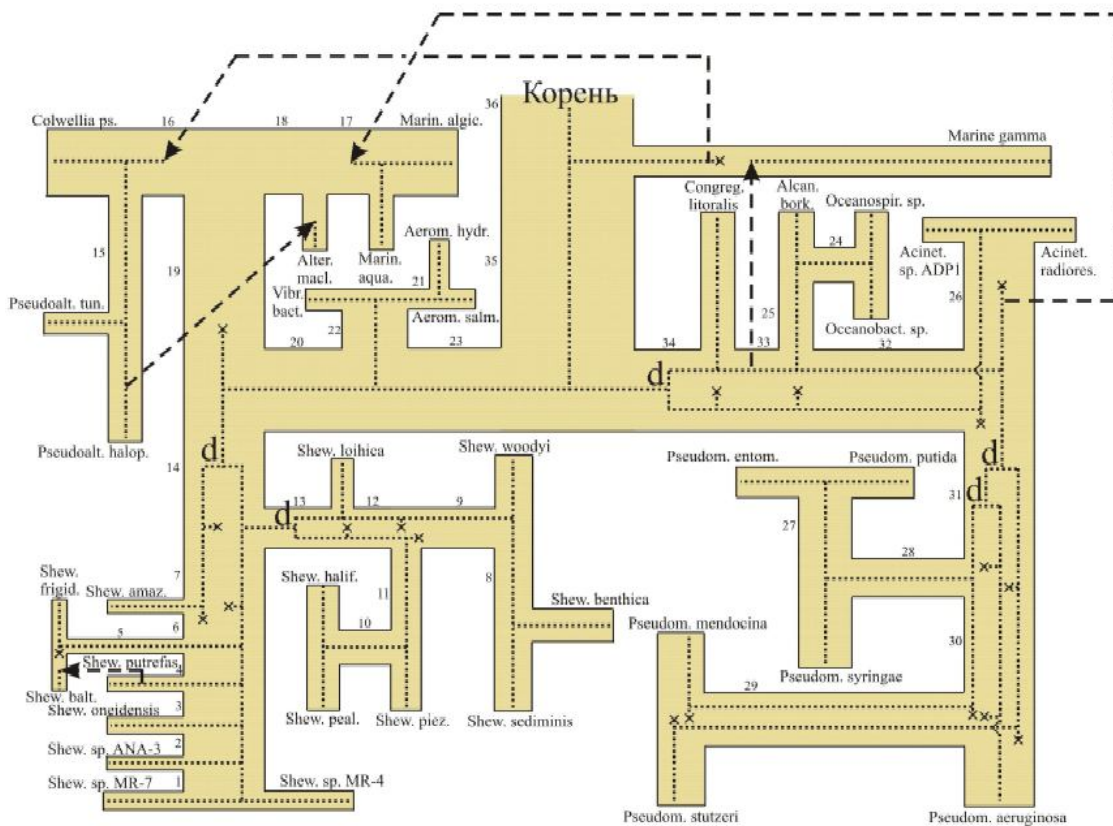
**3.3** Эти результаты получены на основе множественных выравниваний и подтверждены моделированием кинетики процессов. Например, для лидерной области *hisG* у  $\gamma$ -протеобактерий антитерминирующие структуры РНК не достаточно стабильны для обеспечения эффективной регуляции, если не учитывать триплексов. То же наблюдается для *ilvD* и родов *Staphylococcus* и *Listeria*. Для *C. difficile* и *hisZ* только увеличение энергии ко-терминатора, которое происходит при учете триплекса, приводит к отчетливой зависимости экспрессии гена от концентрации гистидил-тРНК.

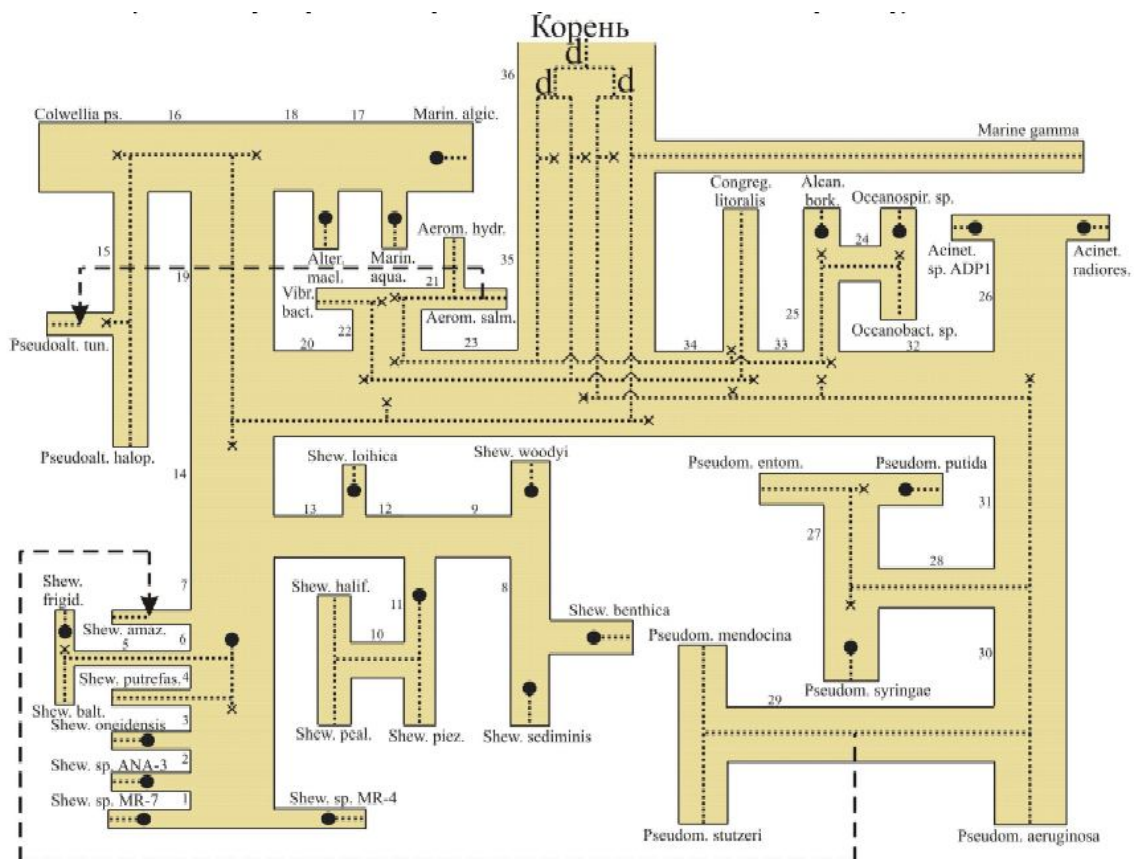
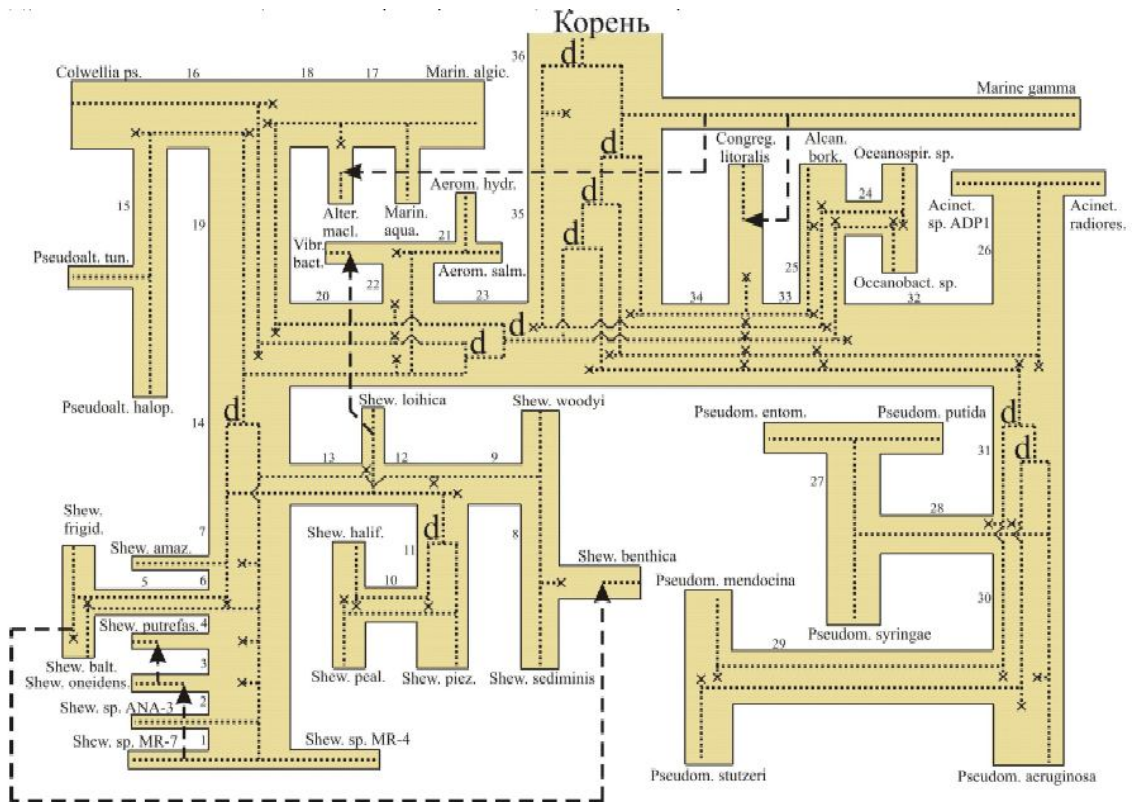
#### **4. Реконструкция и согласование эволюции регуляторных систем, генов и видов – новый подход**

Этот подход реализован в следующих программах <http://lab6.iitp.ru/ru/super3gl/>, <http://lab6.iitp.ru/ru/protfile/>, <http://lab6.iitp.ru/ru/artif/>, доступных на сайте лаборатории.

Разработан алгоритм построения супердерева видов по данному набору деревьев белков, алгоритм представляется адекватным большинству филогенетических задач и в тоже время имеет низкую, практически квадратичную, вычислительную сложность. Он основан на том, что супердерево ищется среди деревьев видов, большинство клад которых представлены среди клад исходных деревьев белков. Предложенный алгоритм тестировался на случайных и биологических наборах деревьев белков и показал практически полное восстановление исходного дерева видов. С другой стороны, предложено новое понятие сценария эволюции гена, основанное на разметке ребер дерева видов эволюционными событиями, которые на них произошли: дубликациями, потерями, возникновениями, горизонтальными переносами генов. Разработан эффективный алгоритм построения такого сценария сразу для многих

генов и регуляторных систем. Разработан полезный во многих задачах алгоритм поиска белка с заданным профилем. На этой основе проведены реконструкции совместной эволюции регуляции, гена и вида. Например, реконструированы эволюционные истории семейств NrdR, MntR, LacI, FNR, Irr, Fur и Rrf2 факторов транскрипции: в предковых вершинах дерева фактора реконструированы предковые сайты связывания и отмечены ребра, на которых произошли значимые эволюционные события. Другой пример. Получены сценарии эволюции генов *pro* синтеза пролина вместе с доменами регуляторного фактора и сайтами связывания, в такой же последовательности они показаны на следующих рисунках. На них ребра дерева видов показаны цветными полосами, дубликации – буквой *d*, потери – крестиками, переносы – стрелками, возникновения – черными кружочками.





## 5. Широкомасштабный поиск регуляции на уровне транскрипции у эубактерий и в пластидах растений и водорослей

На основе оригинальных методов поиска сигнала со сложной структурой и выравнивания вдоль филогенетического дерева проведен широкомасштабный поиск, результаты которого

представлены в пп. 5 и 6. Соответствующие программы <http://lab6.iitp.ru/ru/twobox/> и <http://lab6.iitp.ru/ru/treeal/> доступны на сайте лаборатории.

**5.1** Предсказана потенциальная Rho-зависимая аттенуаторная регуляция биосинтеза цистеина: для оперонов, включающих ген *cysK* цистеин синтазы, у *Propionibacterium acnes* и у всех *Mycobacterium* и для гена *cbs* цистатионин-β-синтазы у *Bifidobacterium longum*.

**5.2** Предсказана регуляция транскрипции генов биосинтеза пролина у протеобактерий: *proA* (γ-глутамилфосфат редуктаза) и *proB* (γ-глутамил киназа), транскрипционная регуляция гена *proC* не обнаружена. 5'-Лидерные области оперонов, содержащих гены *proA* или *proB*, исследованы во всех полностью секвенированных γ- и α-протеобактериях. Сайты с длиной в 8 н. и консенсусом МТАССААУ обнаружены, в основном, у *Pseudomonas* и *Shewanella*; слабые сайты с тем же консенсусом обнаружены у небольшого числа других γ-протеобактерий. У остальных γ-протеобактерий, включая энтеробактерии, такая регуляция отсутствует. Определен конкретный фактор из семейства TetR.

**5.3** Рассмотрены все доступные геномы цианобактерий (51 вид) и хлоропластов Rhodophyta (5 видов). Предсказано 477 потенциальных сайтов связывания фактора NtcA, из них 441 сайт – у видов, по которым нет экспериментальных данных. Предсказано 44 сайта связывания фактора NtcB, из них 40 – у видов, по которым нет экспериментальных данных. У цианобактерий сайты располагаются перед генами из 24-х кластеров ортологичных генов. Оригинальная методика применена к 84 кластерам генов и у 33 из них сайты отсутствуют. Показано, что NtcA-регулируемыми генами, в частности, являются: *glnA*, *glnN*, *glnB*, *icd*, *amtB* (*amt1*), *gifA*, *gifB*, *ntcA*, *nirA*, *nirB*, *narB*, *narK* (*nrtP*), *ntcB*, *nrtA*, *urtA*, *cynA*, *speB*, *mutS*, *rnc*, *apcF*, *som*, *psaI*, *petF* (*fdx*), *hupS*. Изучена эволюция NtcA-регулона. Состав регулона иногда значительно отличается даже у близких видов. Сайты искались только на длине 600 н. перед геном. У видов рода *Synechococcus* ген *icd* никогда не входит в NtcA-регулон, у небольшого числа видов ген *amtB* входит в состав регулона. Показано, что у Rhodophyta эволюционная консервативность фактора NtcA (т.е. присутствие домена из семейства cgr) коррелирует с присутствием потенциального сайта связывания этого фактора в лидерной области гена *glnB*; предложено объяснение этому на основе механизма конкуренции РНК-полимераз. Сайты найдены у родов *Cyanidium* и *Porphyra*. Сайты связывания фактора NtcB обнаружены, в частности, перед геном *nrtA* у видов порядка Chroococcales и перед *nirA* у порядков Chroococcales, Nostocales, Oscillatoriales, Gloeobacterales.

**5.4** Найдены тандемные повторы консервативных участков вблизи промоторов перед геном *chlL* в хлоропластах наземных растений и харовых водорослей, которые служат потенциальными сайтами связывания транскрипционного фактора с ДНК и имеют консенсус GATСТАТ-11N-GATСТАТ. У цветковых растений сам ген отсутствует. Вероятно, соответствующая регуляция является репрессией, так как промотор отчетливо выражен; фактор кодируется в ядре.

**5.5** Проведен массовый поиск и анализ Т-боксовой транскрипционной регуляции экспрессии генов бактерий.

## **6. Широкомасштабный поиск регуляции на уровнях процессинга и трансляции у эубактерий и в пластидах**

**6.1** Предсказана Т-боксовая трансляционная регуляция гена *ileS*, кодирующего изолейцил-тРНК синтазу, у многих актинобактерий. Ранее были известны только случаи Т-боксовой регуляции транскрипции.

**6.2** У большинства актинобактерий перед геном *leuA*, кодирующим 2-изопропилмалат синтазу, обнаружена новая потенциальная регуляторная система (названная LEU) с участием гена лидерного пептида и псевдоузла РНК, которая принимает альтернативные конформации. Предположено происхождение этого псевдоузла из регуляторного сигнала траспозаз. Идентифицирован белок, связывающий LEU-элемент. У многих  $\alpha$ -протеобактерий обнаружена другая РНКовая регуляция (названная LEU1) того же гена, обладающая сходным механизмом действия.

**6.3** Предложен новый специфичный для рода *Brucella* механизм регуляции генов, в частности, гена *mntH*, который кодирует транспортёр из семейства Nramp, и генов металлопротеинов. Предполагаемый механизм регуляции основан на деградации мРНК при участии длинных спиралей РНК, расположенных до или после регулируемого гена. А именно, если двухвалентных катионов марганца много, то марганец-зависимая РНКаза разрезает мРНК.

**6.4** Изучена регуляция экспрессии генов пластид у простейших паразитов типа Apicomplexa. Описаны консервативные некодирующие участки в апикопластах споровиков и в хлоропластах красных водорослей. Предполагается, что эти участки служат сайтами связывания регуляторных белков, перекрывающих сайт связывания рибосомы или обеспечивающих процессинг мРНК. (1) Консервативные участки обнаружены перед геном *ycf24*, необходимым для формирования FeS-центров, и перед некоторыми другими генами: в 5'-лидерных областях у красных водорослей родов *Porphyra* и *Gracilaria* и у многих видов Apicomplexa. Эти участки имеют консенсус UUnAnAUA-UnwUAUAwAwAAUUAUU. Таким образом у *Toxoplasma gondii* осуществляется совместная регуляция трех генов *ycf24*, *rps4* и *rpoB*. Предсказание согласуется с известными данными о том, что апикопласты из *T. gondii* нужны лишь в течение малого времени, когда они обеспечивают патогенность токсоплазмы. (2) Консервативные участки с консенсусом AnUAUUAAYUnUUnAAwnUnAnAAwUUknwAUwAAkkwkAUmU найдены перед геном *rpoB*, кодирующим  $\beta$ -субъединицу РНК-полимеразы, у красных водорослей рода *Porphyra* и у паразитов *Eimeria tenella* и *Theileria parva*. (3) Консервативный участок найден перед геном *ups8*, специфичным для апикопластов Piropiasmida. Этот ген расположен между генами *rpl14* и *rps8*, кодирующими



рибосомные белки. Участок имеет консенсус 5'-kATAGAm-3' и расположен на расстоянии от 170 до 100 н. от инициирующего кодона *ups8*. У *Babesia bovis* и *Babesia bigemina* участок располагается внутри кодирующей области гена *rpl14* вблизи С-конца и имеет вид TSYSIDDRNRFKD; он присутствует только у этих *Babesia* spp. и отсутствует у ортологичных белков L14.

**6.5** Показана корреляция светозависимой регуляции гена *chlL* с присутствием вставки внутри кодирующей области этого гена у зелёных водорослей *Chlamydomonas reinhardtii*, *Volvox carteri* и *Rychnococcus provasolii*. Такая вставка отсутствует у других зеленых водорослей ветви Chlorophyta и различается у этих видов. Поскольку аминокислотная последовательность белка ChlL чрезвычайно консервативна за исключением указанных вставок вблизи С-конца у перечисленных видов, можно думать, что вставки удаляются в ходе процессинга белка.

**6.6** Показано, что 5'-лидерные области генов хлоропластов *atpF*, *clpP*, *petB*, *psaA*, *psbA*, *psbB* содержат консервативные сайты РНК у высших растений и многих зелёных водорослей: для генов *atpF* и *petB* у высших растений, для гена *clpP* у растений и водорослей из ветви Streptophyta, для гена *psbA* у водорослей из ветви Chlorophyta, включая хламидомонаду. Сайты расположены вблизи инициирующих кодонов этих генов и предположительно участвуют в регуляции инициации трансляции. Наличие консервативных сайтов перед генами *atpF*, *clpP*, *petB* коррелирует с наличием интронов в кодирующей области. Возможно, некоторый белок связывает сайты до завершения сплайсинга, хотя здесь может играть роль и наблюдаемая шпилька мРНК. Такая регуляция экспериментально показана для гена *psbA* хламидомонады, для которого трансляция не идет в темноте.

**6.7** Найдены в 5'-лидерных областях генов *psbB* у растений и водорослей из ветви Streptophyta консервативные сайты разрезания мРНК. Экспериментально этот сайт показан у шпината для гена *psbB*.