

На правах рукописи



Жукова Ксения Владимировна

**Роль нуклеотидного состава 5'-области мРНК в эффективности
экспрессии генов в растениях**

1.5.21– физиология и биохимия растений

Автореферат диссертации
на соискание ученой степени
кандидата биологических наук

Москва – 2022

Работа выполнена в лаборатории функциональной геномики Федерального государственного бюджетного учреждения науки Института физиологии растений им. К.А. Тимирязева Российской академии наук, Москва.

Научный руководитель:

доктор биологических наук,
профессор

Голденкова-Павлова Ирина Васильевна

Официальные оппоненты:

Дейнеко Елена Викторовна

доктор биологических наук, профессор, главный научный сотрудник, заведующая Лабораторией биоинженерии растений Федерального государственного бюджетного научного учреждения "Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук

Ермошин Александр Анатольевич

кандидат биологических наук, доцент кафедры экспериментальной биологии и биотехнологий Института естественных наук и математики федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

Ведущая организация:

Институт биохимии и генетики - обособленное структурное подразделение Федерального государственного бюджетного учреждения науки Уфимского федерального исследовательского центра Российской академии наук, г. Уфа.

Защита диссертации состоится «15» декабря 2022 года в 11 часов на заседании Совета по защите докторских и кандидатских диссертаций 24.1.138.01 при Федеральном государственном бюджетном учреждении науки Институте физиологии растений им. К.А. Тимирязева Российской академии наук по адресу: 127276, Москва, Ботаническая ул., 35. Телефон/факс: (499) 678-54-20; e-mail: ifr@ippras.ru

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института физиологии растений им. К.А. Тимирязева Российской академии наук и на сайте www.ippras.ru

Автореферат разослан « » _____ 2022 года.

Ученый секретарь

диссертационного совета:
кандидат биологических наук



Азаркович Марина Ивановна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Согласно текущему мнению, реализация генетической информации — многостадийный процесс, включающий транскрипцию, процессинг пре-мРНК, трансляцию и обеспечение стабильности конечного белкового продукта. При этом следует отметить, что трансляция является фундаментальным процессом и важной контрольной точкой в регуляции экспрессии генов для всех живых клеток, поскольку в этом процессе раскрывается кодирующий потенциал мРНК через молекулу белка (Guerra et al., 2015). Из-за высокой энергозатратности (Rolfe and Brown, 1997) и катастрофических последствий неправильной регуляции - трансляция находится под строгим контролем (Akirtava, C. and McManus, C. J., 2021) посредством сложных взаимодействий между цис-действующими последовательностями и транс-действующими факторами.

Эукариотическая трансляция регулируется, прежде всего, на стадии инициации (Aylett and Van, 2017; Hinnebusch et al., 2016; Shirokikh and Preiss, 2018). Элементы последовательности, которые опосредуют регуляцию экспрессии генов на этапе трансляции, часто находятся в 5'- и 3'-нетранслируемых областях (5'-НТО и 3'-НТО) мРНК. Согласно текущему мнению, критическую роль в инициации трансляции играет 5'-НТО, которая не только задействована в иммобилизации малой субъединицы рибосомы, но и обеспечивает дальнейшее продвижение трансляционного комплекса вдоль мРНК до стартового AUG кодона, с которого и начнется синтез соответствующего белка. Подобный сценарий инициации трансляции характерен для подавляющего большинства мРНК эукариот, в целом, и растения — не исключение (Hinnebusch et al., 2016). В 5'-НТО мРНК имеется множество цис-регуляторных элементов последовательности, которые влияют на последующую судьбу мРНК - транспортировку, стабильность и эффективность трансляции (Vaughn, J. N. et al., 2012).

Среди характеристик, которые широко коррелируют с трансляцией мРНК, являются GC состав, вторичные структуры, uORF, нуклеотидный состав в 5'-НТО мРНК, а также нуклеотидный контекст вокруг стартового AUG кодона, которые действуют как контрольная точка для трансляции мРНК (Lei et al., 2015).

Подавляющее большинство исследований, посвященных изучению трансляции как важному этапу регуляции экспрессии генов, проводятся на клетках дрожжей, млекопитающих, в том числе и человека. При этом известно, что регуляторные элементы экспрессии у различных таксонов могут отличаться. Поэтому требуется детальное их изучение на примере растений, так как неправильная экспрессия регуляторных генов в растениях часто приводит к снижению их роста, и представляет собой серьезную проблему при выведении культур, устойчивых к биотическому или абиотическому стрессу (Silva et al., 2019).

Наша работа направлена на поиск регуляторных нуклеотидных контекстов в 5'-НТО мРНК растений *A. thaliana*. Мы осуществили это двумя разными способами.

В первом случае в качестве предмета исследования мы выбрали 5'-НТО с CpG островками. Так как известно, что промоторы с CpG островками не подвергаются

метилованию и остаются транскрипционно активными, в отличие от промоторов бедных CG, в которых цитозин метилируется. Поэтому было интересно изучить, как CG-богатая последовательность участвует в регуляции трансляции в качестве 5'-НТО.

Во втором случае для поиска мотивов 5'-НТО в активно транслирующихся мРНК мы применили метод профилирования полисом с последующим анализом полученного массива данных.

Цель исследования

Поиск регуляторных кодов 5'-области для генов растений на примере *A. thaliana* и оценка влияния ее на эффективность экспрессии гена в растительных клетках.

Задачи исследования

1. Разработать подходы *in silico* анализа больших выборок нуклеотидных последовательностей и поиска в них потенциальных регуляторных мотивов эффективной экспрессии генов у растений.
2. Выявить высоко представленные нуклеотидные мотивы с различным составом в 5'-области мРНК, рассматривая их как потенциальные регуляторные последовательности.
3. Выяснить функциональную роль нуклеотидного состава и регуляторных мотивов 5'-областей генов в эффективности экспрессии генов у растений.

Научная новизна исследования

Впервые сконструирована синтетическая последовательность, которая содержит характерные для 5'-области генов растений CG-богатые мотивы, выявленные на основании *in silico* анализа 5'-областей генов растений *A. thaliana*. На трансгенных растениях табака изучен вклад нуклеотидного состава 5'-области мРНК в эффективность экспрессии на уровне транскрипции и трансляции с использованием репортерного гена термостабильной лихеназы. Впервые показано, что синтетическая CG-богатая последовательность в 5'-области гена достоверно увеличивает уровень транскрипции репортерного гена и, по всей видимости, не оказывает негативного влияния на эффективность трансляции мРНК репортера. Разделение пулов мРНК *A. thaliana* ассоциированных с моносомной и полисомной фракцией методом профилирования полисом и анализ транскриптов (мРНК), ассоциированных с каждым пулом мРНК, за счет секвенирования РНК, позволили получить представление о трансляционной эффективности индивидуальных мРНК, а последующий *in silico* анализ – провести поиск регуляторных контекстов в 5'-области мРНК растений *A. thaliana*, которые могут быть потенциально важными для эффективной трансляции мРНК. Впервые на основе *in silico* анализа выборки стабильно транскрибируемых во время онтогенеза генов *A. thaliana* установлено, что пиримидиновые ди-нуклеотиды и мотивы характерны для 5'-НТО мРНК с высокой трансляционной эффективностью, тогда как пуриновые ди-нуклеотиды и мотивы ассоциированы с транскриптами, имеющими низкую трансляционную эффективность.

Теоретическая и практическая значимость работы

На основании проведенных теоретических и экспериментальных исследований определены консенсусные последовательности 5'-областей генов растений как новые

регуляторные элементы, которые потенциально могут обеспечить высокоэффективную экспрессию и синтез целевого продукта в растениях. Подход, основанный на поиске и экспериментальной верификации регуляторных кодов эффективной трансляции, позволит разрабатывать оптимальные системы экспрессии генов в растениях и создать оптимальные экспериментальные модели трансгенных растений, в том числе и с использованием современных методов переноса гетерологичных генов в растения. На основании метода профилирования полисом проведен поиск регуляторных контекстов в 5'-области мРНК растений *A. thaliana*, которые могут быть потенциально важными для эффективной трансляции мРНК.

Методология и методы научного исследования.

Для дизайна регуляторной последовательности использованы последовательности 5'-областей мРНК *A. thaliana*, выборки которых получали с использованием публичной базы данных *JetGene* и ее программного обеспечения. Экспрессионные векторы, несущие синтетическую CG-богатую последовательность в качестве 5'-области репортерного гена термостабильной лихеназы, конструировали стандартными приемами молекулярного клонирования. Трансгенные растения табака получали методом ко-культивирования листовых дисков с суспензией агробактерий, несущих соответствующие растительные векторы. Для оценки функциональной роли CG-богатой 5'- области гена определяли уровень мРНК репортерного гена и уровень его белкового продукта в линии трансгенных растений. Для поиска регуляторных контекстов в 5'- области мРНК *A. thaliana* с разной трансляционной эффективностью применили метод профилирования полисом, с последующим секвенированием полученных фракций мРНК ассоциированных с полисомами и моносомами.

Положения, выносимые на защиту.

1. CpG динуклеотиды, локализованные в 5'-области гена, увеличивают активность транскрипции, и, по всей видимости, не оказывает негативного влияния на эффективность трансляции мРНК репортера.
2. Пиримидиновые ди-нуклеотиды и мотивы характерны для 5'-НТО мРНК с высокой трансляционной эффективностью характерны, тогда как пуриновые ди-нуклеотиды и мотивы ассоциированы с транскриптами, имеющими низкую трансляционную эффективность.

Степень достоверности и апробация результатов работы

Достоверность полученных результатов подтверждается воспроизводимостью результатов, их статистической обработкой и публикацией в рецензируемых журналах. Для исследований использовано современное, сертифицированное оборудование и реагенты. При проведении экспериментов использовались классические и современные молекулярно-биологические, биохимические методы, а также методы анализа экспериментального материала, которые подтверждают обоснованность и достоверность полученных экспериментальных результатов. Результаты исследований были представлены на конференциях: Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2016» (Москва, 2016); 18-я научная конференция молодых ученых «Биотехнология в растениеводстве,

животноводстве и ветеринарии» (Москва, 2018); II международная научно-практическая конференция «Клеточная биология и биотехнология растений» (Минск, 2018); IX Съезд общества физиологов растений России «Физиология растений – основа создания растений будущего» (Казань, 18–24 сентября 2019 г.)

Публикации

По материалам диссертации опубликовано 13 научных работ, в том числе 4 статей в журналах, рекомендованных ВАК РФ, 1 патент, и 8 тезисов конференций.

Структура и объем научно-квалификационной работы.

Диссертационная работа состоит из введения, обзора литературы, описания методов исследований, результатов исследований и их обсуждения, заключения, выводов, списка используемой литературы и приложения. Работа изложена на 125 страницах машинописного текста, содержит 7 таблиц и 20 рисунков. Библиографический список включает 238 источников, из них 235 на иностранном языке.

СОДЕРЖАНИЕ РАБОТЫ

Материалы и методы исследования

In silico анализ

В исследовании использованы последовательности 5'-НТО мРНК *A. thaliana*, выборки которых получали с использованием публичной базы данных *JetGene* (<https://jetgene.bioset.org/>) и ее программного обеспечения.

Конструирование растительных экспрессионных векторов

В работе использовали стандартные процедуры молекулярного клонирования и протоколы ПЦР (Maniatis et al., 1985). Последовательность нуклеотидов с высоким содержанием CpG динуклеотидов (обозначена нами CG) и лидерная омега (Ω) последовательность вируса мозаики табака (TMV) были синтезированы из олигонуклеотидов методом ПЦР (ЗАО «Евроген», Россия).

Растительный материал и бактериальные штаммы

В работе использовали растения табака *Nicotiana tabacum* (сорт *Petit Habana*). Контрольные и трансгенные растения табака поддерживали как стерильную культуру на агаризованной MS среде при 23°C и 16-час освещении 100 мкмоль квантов/(м² с). pVIG-T-LicB, pVIG-T-CG-LicB, pVIG-T- Ω -LicB и pVIG-T- Ω -CG-LicB векторами трансформировали штамм агробактерий GV3101. Трансформацию *Nicotiana tabacum* (сорт *Petit Habana*) проводили методом ко-культивирования листовых дисков с суспензией агробактерий, несущих соответствующие векторы. Для молекулярного и биохимического анализа использовали растения, которые культивировали *in vitro* 4-5 недель.

Полисомное профилирование проводили на растениях *A. thaliana*. Растения выращивали при 22 ± 1°C, 12-часовом фотопериоде, освещенности 100 мкмоль квантов/(м² с).

Анализ трансгенных растений

Присутствие последовательностей целевого гетерологичного гена в геноме трансформантов растений табака подтверждали методом мультиплексной ПЦР.

Размер белковых продуктов, синтезирующихся в трансгенных растениях, был подтвержден методом зимограмм. Уровень накопления репортерного белка и гибридных белков, сконструированных на основе репортерного белка, рассчитан по активности термостабильной лихеназы в экстрактах, полученных из трансгенных растений. Активность лихеназы определяли, используя в качестве субстрата лихенан. Определение восстанавливающих сахаров, освобождающихся из субстрата, проводили по методу Вуда и Бхат.

Для оценки уровня транскрипции репортерного гена в полученных линиях трансгенных растений определяли уровень мРНК методом количественной ПЦР в реальном времени.

Получение моносомных и полисомных фракций, выделение РНК из фракций и тотальной РНК, секвенирование, сборка транскриптома и картирование

Разделение моносомных и полисомных фракций проводили дифференциальным центрифугированием в градиенте сахарозы согласно методу, описанному ранее (Mustroph et al., 2009). Качество и количество препаратов суммарной РНК и РНК из полисомных и моносомных фракций растений оценивали на Agilent Bioanalyzer 2100. Секвенирование библиотек проводили на 1 дорожке Illumina HiSeq 4000 (101 цикл с двух сторон фрагментов) с использованием реактивов HiSeq 4000 sequencing kit version 1 (ЗАО “Евроген”, Россия). Относительный уровень транскрипции (logCPM) рассчитывали, как логарифмическое отношение количества прочтений на последовательности аннотированного гена, которые нормализованы на миллион (counts per million - CPM) согласно протоколу (Qi et al., 2017). Показатель трансляционной эффективности (logFC ПФ/МФ) рассчитывали, как логарифмическое отношение количества прочтений для отдельных транскрипционных единиц, представленных в полисомной и моносомной фракции (Larsson et al., 2010). Для сравнения логарифмических соотношений использован t-критерий Стьюдента (двусторонняя неравная дисперсия) (Larsson et al., 2010).

Классификация генов по эффективности трансляции подтверждена методом количественной ПЦР в реальном времени. Количественную ПЦР проводили, используя StepOnePlus Real-Time PCR System (“Applied Biosystems”, США) и SYBR Green как флуоресцентный краситель согласно протоколу производителя.

Основные результаты

In silico анализ 5'- областей генов растений и дизайн CG богатой регуляторной последовательности

Для того, чтобы получить представление о влиянии CG богатого контекста 5'- области на экспрессию гетерологичного гена в растениях, первоначально определились с критериями размера и нуклеотидного состава целевой регуляторной последовательности: (1) размер не менее 200 п.н. (минимальный размер CpG-островка); (2) GC содержание выше 50% (одна из характеристик CpG-остроек); (3) наличие мотивов, характерных для 5'-областей мРНК генов *A.thaliana* с GC содержанием не менее 50%; (4) нуклеотиды, окружающие стартовый кодон, в положении -3 и +4, согласно последовательности Козак (Kozak, 1986b); (5) отсутствие

альтернативных стартовых и терминирующих кодонов. Характеристики выбранных последовательностей суммированы в таблице 1.

Таблица 1. Характеристики 5'-областей генов *A.thaliana*, использованных для поиска мотивов, и синтетической регуляторной последовательности

	Характеристики последовательностей Среднее значение (интервал значений)			
	Размер, п.н.	CG, %	Доля CpG динуклеотидов *	Мотивы****
5'-НТО	330 (от 200 до 982)	52,1 (от 50,0 до 60,1)	0,06 (от 0,1 до 0,03)	CAAGGT, GGTGCC, GCCGGT, GCTGCT,
5'-область кодирующей последовательности**	400	57,8 (от 50,5 до 69,5)	0,06 (от 0,09 до 0,05)	GGTGGT, GGTGCT, GCCGCT
5'-область кДНК***	400	55,2 (от 51,0 до 65,3)	0,06 (от 0,09 до 0,05)	CAAGGT GGTGCC, GGTGGT, GGTGCT, GCCGCT
Синтетическая регуляторная последовательность	405	62,5	0,06	CAAGGT, GGTGCC, GGTGGT, GCCGGT, GCTGCT, GGTGCT, GCCGCT

* Рассчитана в пересчете на длину последовательности в п.н.;

** - группа генов, с отсутствием 5'-НТО, взяты первые 400 п.н. кодирующей последовательности;

*** - группа генов с 5'-НТО до 200 п.н., взяты первые 400 п.н. кДНК;

**** мотивы, которые встречаются в 5'-области у 30% последовательностей

При дизайне регуляторной последовательности учитывали пять критериев, описанных выше, а также вторичную структуру, которая будет непосредственно прилегать к началу стартового AUG кодона. Анализ синтетической регуляторной последовательности алгоритмом *centroidfold* (Sato et al., 2009) не показал наличие стабильной вторичной структуры (данные в автореферате не приводятся).

Экспрессионные векторы для трансформации растений

Для того, что выяснить функциональную роль синтезированной последовательности в регуляции экспрессии генов у растений были сконструированы

две серии экспрессионных векторов. Серия векторов А включает вектор pVIG-T-LicB, в котором экспрессия репортерного гена термостабильной лихеназы контролируется 35S РНК СаMV промотором, а также вектора pVIG-T-CG-LicB и pVIG-T-Ω-LicB, в которых между последовательностью 35S РНК СаMV промотора и стартовым кодоном трансляции (ATG) репортерного гена интегрированы: синтетическая CG-богатая последовательность (405 п.н.) и последовательность омега (Ω) TMV, соответственно, которые в векторах будут функционировать в качестве 5'-НТО мРНК репортерного гена (Рисунок 1). Последовательность омега (Ω) TMV использована как контроль, поскольку ее функция как трансляционного энхансера убедительно доказана (Wever et al., 2010). Серия Б включает вектор pVIG-T-Ω-CG-LicB, который содержит гибридный ген, кодирующий слияние синтетической CG-богатой последовательности (405 п.н.) с последовательностью репортерного гена термостабильной лихеназы за счет введения в 5'-область гибридных генов стартовых ATG кодонов (Рисунок 1). В этой серии векторов последовательность омега (Ω) TMV функционирует в качестве 5'-НТО гибридных генов, а синтетическая CG-богатая последовательность как 5'-область кодирующей последовательности гибридного гена.



Рисунок 1. Схема целевой части Т-ДНК экспрессионных векторов серии А pVIG-T-LicB, pVIG-T-CG-LicB, pVIG-T-Ω-LicB и серии Б pVIG-T-Ω-CG-LicB для трансформации растений. P 35 - промотор 35S РНК СаMV; LicB – последовательность репортерного гена термостабильной лихеназы; CG – синтетическая регуляторная последовательность, сконструированная из характерных для 5'-области растительных генов мотивов; Ω – трансляционный энхансер омега из TMV; ATG – стартовый кодон.

Получение и ПЦР-анализ трансгенных растений

С использованием сконструированных экспрессионных векторов методом *Agrobacterium*-опосредованной трансформации были получены трансгенные растения *N. tabacum*. Линии регенерантов-растений, полученных с векторами pVIG-T-LicB, pVIG-T-CG-LicB, pVIG-T-Ω-LicB и pVIG-T-Ω-CG-LicB были обозначены как LicB, CG-LicB, Ω-LicB и Ω-CG-LicB, соответственно. Регенерированные растения, устойчивые к селективному агенту канамицину, были проанализированы методом мультиплексной ПЦР (Рисунок 2А и Б). В результате анализа для каждой линии трансгенных растений было отобрано 8-12 независимых трансформантов *N. tabacum*.

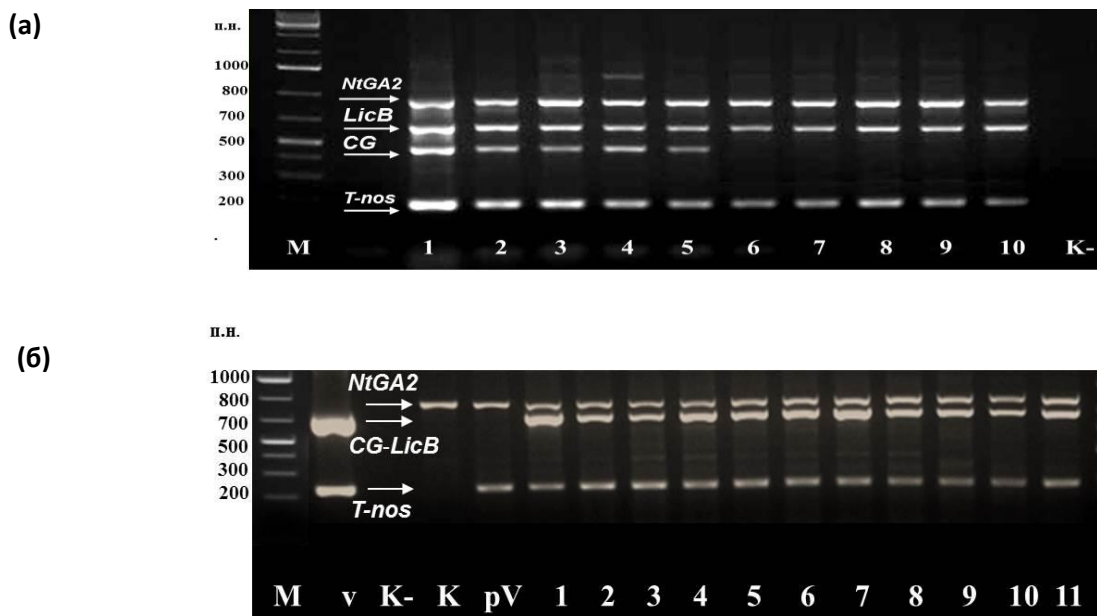


Рисунок 2. Анализ геномной ДНК трансгенных линий *N. tabacum* для обнаружения перенесенных последовательностей методом мультиплексной ПЦР.

Дорожки на (а): 1–5 – ДНК трансгенных растений линии CG-LicB; 6–8 - линии Ω-LicB; 9, 10 - линии LicB; LicB (642 п.н.) – ампликоны репортерного гена термостабильной лихеназы; CG (455 п.н.) – ампликоны части синтетической CG-богатой последовательности и части репортерного гена; дорожки на (б): v – ДНК экспрессионного вектора; K – ДНК контрольного растения; pV – ДНК трансгенного растения табака, трансформированного пустым вектором; 1–11 – ДНК трансгенных растений табака линии Ω-CG-LicB; CG-LicB (682 п.н.) - ампликоны гибридного гена, кодирующего слияние синтетической CG-богатой последовательности с последовательностью репортерного гена термостабильной лихеназы; на (а) и (б): М – ДНК маркер молекулярного веса; К- - отрицательный контроль, ПЦР без добавления геномной ДНК; T-nos - ампликоны терминатора T-nos (188 п.н.); амплификация фрагмента гена актина (NtGA2) (795 п.н.) служит внутренним положительным контролем; отсутствие ампликона 432 п.н., соответствующего гену *virD1 A. tumefaciens*, указывает на отсутствие агробактериальной контаминации.

Отбор и анализ трансформантов растений, экспрессирующих репортерный ген термостабильной лихеназы

Трансгенные линии *N. tabacum*, для которых наличие перенесенных генов было доказано методом мультиплексной ПЦР, первоначально проанализированы с помощью качественного теста активности лихеназы. Размер белковых продуктов, синтезирующихся в трансгенных растениях, был подтвержден методом зимограмм (Рисунок 3). Как видно из представленных результатов, в растениях происходит синтез белковых продуктов с молекулярной массой, соответствующей теоретически рассчитанной – около 25 кДа – для термостабильной лихеназы и около 36 кДа - для гибридного белка, в котором лихеназа слита с CG-богатой последовательностью.

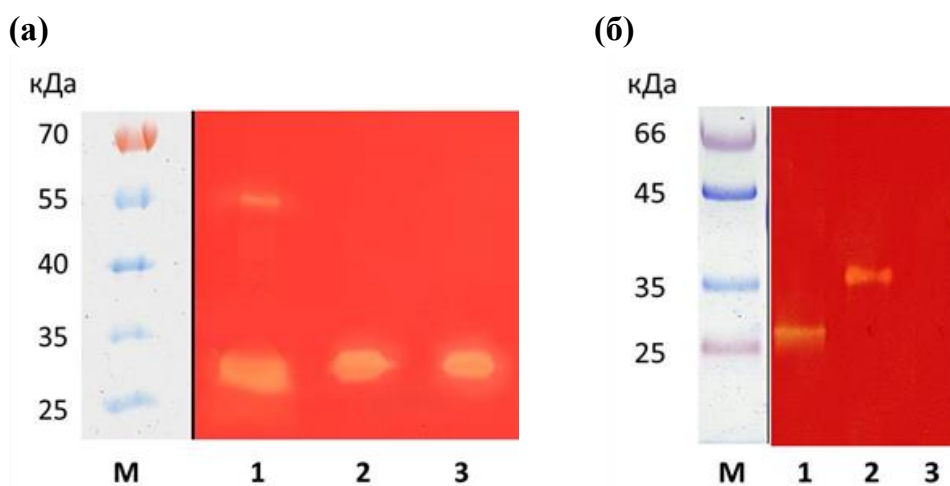


Рисунок 3. Энзимограмма белковых лизатов, полученных из трансгенных растений *N. tabacum*. а - размер белковых продуктов репортерного гена, синтезирующихся в трансгенных растениях линии LicV (1), линии CG-LicV (2) и линии Ω-LicV (3); б - размер белковых продуктов репортерного (1) и гибридного (2) гена, синтезирующихся в трансгенных растениях (линия CG-LicV и линия Ω-CG-LicV соответственно), (3) – белковый экстракт трансгенного растения, трансформированного пустым вектором; М – маркер молекулярной массы.

Оценка эффективности регуляторных последовательностей в экспрессию репортерного гена в трансгенных растений

Для оценки эффективности экспрессии репортерного гена за счет синтетической регуляторной последовательности определили уровни накопления репортерного белка и мРНК репортерного гена у трансгенных растений. Уровень накопления репортерного белка и гибридных белков, сконструированных на основе репортерного белка, рассчитан по активности термостабильной лихеназы в экстрактах, полученных из трансгенных растений (Рисунок 4А). Полученные результаты продемонстрировали, что в линии CG-LicV уровень накопления белка увеличился в 14,1 раз, в линии Ω-CG-LicV - в 19,5 раз и в линии Ω-LicV в 5 раз, соответственно, по сравнению с уровнем накопления лихеназы в линии растений LicV. Эти данные свидетельствуют о положительном функциональном вкладе CG-богатой последовательности, локализованной в 5'-области репортерного гена, в эффективность его экспрессии. Для оценки уровня транскрипции репортерного гена в полученных линиях трансгенных растений определяли уровень мРНК методом количественной ПЦР в реальном времени. Полученные результаты продемонстрировали, что относительное количество (RQ) транскрипта репортерного гена достоверно выше у линий трансгенных растений CG-LicV и Ω-CG-LicV – в 4,1 и 3,5 раза, соответственно, по сравнению с уровнем мРНК репортерного гена в линии растений LicV (Рисунок 4Б). При этом уровень транскрипции репортерного гена между линиями LicV и Ω-LicV достоверно не отличается (Рисунок 4Б).

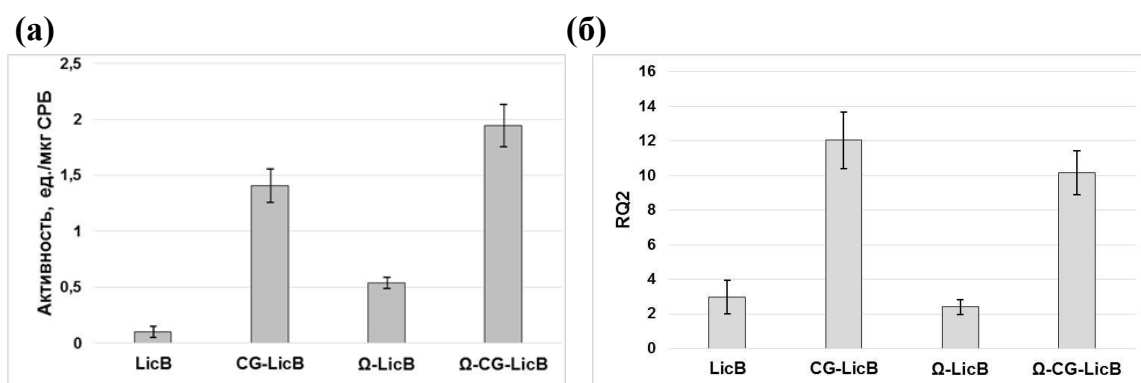


Рисунок 4. Количественный анализ активности репортерного белка термостабильной лихеназы (а) и уровень мРНК репортерного гена (б) в линиях трансгенных растений. Представлены средние значения для 7 (линия Ω-CG-LicB) и 8 (линии LicB, CG-LicB и Ω-LicB) независимых трансформантов в каждой линии растений. На графиках указаны стандартные отклонения результатов трех независимых измерений.

Эти результаты свидетельствуют о том, что синтетическая CG-богатая последовательность достоверно увеличивает уровень транскрипции репортерного гена и, по всей видимости, не оказывает негативного влияния на эффективность трансляции мРНК репортера.

Поиск и функциональная роль нуклеотидных мотивов в 5'-НТО мРНК генов растений

Полисомное профилирование мРНК растений *A.thaliana* на стадии развития 3.70, и анализ транскриптов за счет секвенирования РНК

Во второй части исследования мы применили метод профилирования полисомом для разделения мРНК в зависимости от их загрузки рибосомами, а именно на полисомные и моносомные фракции. Разделение пулов таких мРНК и анализ транскриптов (мРНК), ассоциированных с каждым пулом мРНК, за счет секвенирования РНК, позволили получить представление о трансляционной эффективности индивидуальных мРНК, а последующий *in silico* анализ – провести поиск регуляторных контекстов в 5'-области мРНК растений *A. thaliana*, которые могут быть потенциально важными для эффективной трансляции мРНК.

Для проведения полисомного профилирования, а также для выделения суммарной РНК и РНК из полисомных фракций получали растения *A. thaliana* на стадии развития 3.70 (Boyes et al., 2001). На стадии развития 3.70 у растений *A. thaliana*, которую они достигают примерно за 28 дней, размер розетки составляет примерно 70% от своего окончательного размера и начинают формироваться бутоны.

Полисомный профиль определяли путем измерения содержания РНК в каждой фракции с помощью прибора Nanodrop ND-1000 (“LabTech International”, Великобритания). Фракции РНК объединяли следующим образом: с 3 по 7 (суммарная моносомная фракция), с 8 по 12 (суммарная полисомная фракция) (Рисунок 5).

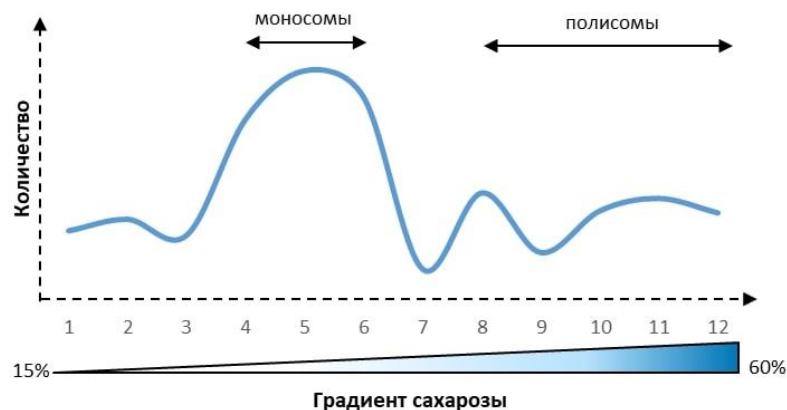


Рисунок 5. Профиль полисом, оцененный по содержанию РНК во фракциях.

Качество и количество препаратов суммарной РНК и РНК из полисомных и моносомных фракций растений оценивали электрофоретическим анализом (Рисунок 7) и на Agilent Bioanalyzer 2100.

В результате RNA-Seq получены высококачественные прочтения ($Q30 > 84\%$) в общей сложности: для моносомной фракции (МФ) – $1,57 \times 10^8$; для полисомной фракции (ПФ) – $1,43 \times 10^8$; и для образцов тотальной мРНК (ТФ) - $1,68 \times 10^8$. Полученные для трех наборов образцов прочтения с удаленными адаптерами картированы на геном *A. thaliana* (TAIR10), оценены профиль охвата прочтений вдоль хромосом и профиль покрытия прочтений по длине отдельных транскрипционных единиц - ТФ, МФ и ПФ.

Характеристика транскриптов из моносомных и полисомных фракций, экспрессируемых на выбранной стадии развития, по уровню их транскрипции и относительной трансляционной эффективности и проведение их кластеризацию на группы в зависимости от эффективности трансляции

Для проведения последующего *in silico* анализа формирование выборки генов проведено, исходя из следующего критерия: выбранные гены характеризуются постоянным уровнем транскрипции в широком диапазоне абсолютных значений на всех этапах онтогенеза растений (Czechowski et al., 2005). Этот выбор базируется на следующем соображении – у таких генов уровень мРНК каждого индивидуального гена сохраняется на постоянном уровне во время онтогенеза, что может нивелировать модуляцию эффективности трансляции, потенциально возможную за счет изменения количества индивидуального транскрипта. В выборку включено 130 последовательностей с низким коэффициентом вариации транскрипции (показатель SD/MV), который характеризует стабильность транскрипции на всех изученных ранее стадиях развития *A. thaliana* (Crooks et al., 2004). Для каждого гена из сформированной выборки из 130 стабильно транскрибируемых при разных стадиях роста и развития генов *A. thaliana* оценивали: (1) относительный уровень транскрипции – показатель logCPM (см. Материалы и методы) (Qi et al., 2017); и (2) трансляционную эффективность – показатель logFC ПФ/МФ.

Используя показатели относительного уровня транскрипции (logCPM) и трансляционной эффективности (logFC ПФ/МФ), построена эмпирическая функция распределения мРНК для каждого транскрипта (Рисунок 6).

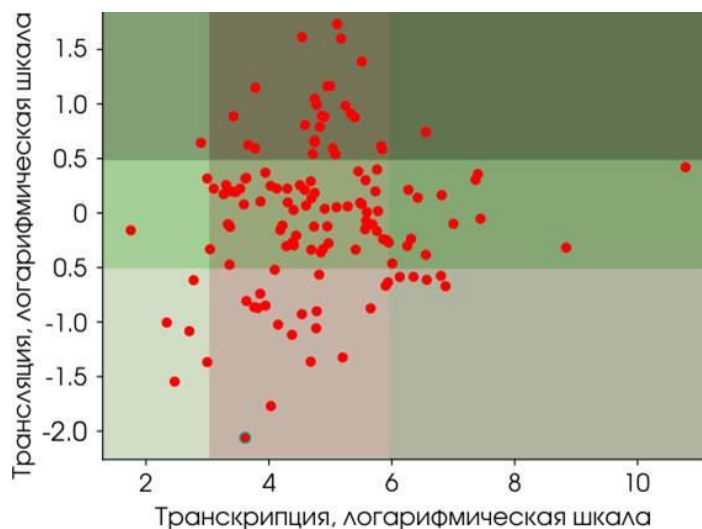


Рисунок 6. Распределения мРНК в зависимости от уровня транскрипции и трансляции.

Как видно из представленных данных корреляция между уровнем транскрипции генов и трансляционной эффективностью их мРНК у сформированной выборки крайне скромная.

Далее, все транскрипты в зависимости эффективности трансляции разделяли на кластеры, используя следующие критерии для классификации: $\text{Log}_2\text{FC ПФ/МФ} > 0,5$ – высокая (кластер 1), $- 0,6 > \text{Log}_2\text{FC ПФ/МФ} < 0,5$ – средняя (кластер 2) и $\text{Log}_2\text{FC ПФ/МФ} < - 0,6$ – низкая (кластер 3).

Классификация генов по эффективности трансляции подтверждена методом количественной ПЦР в реальном времени (Рисунок 7).

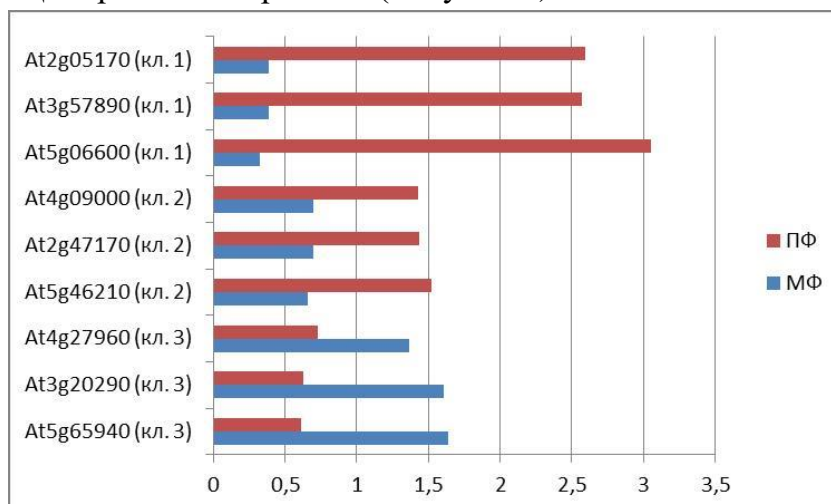


Рисунок 7. Относительные количества (RQ) транскриптов генов из кластеров с разной трансляционной эффективностью. ПФ и МФ - суммарная полисомная и моносомная фракция, соответственно; At2g05170, At3g57890, At5g06600, At4g09000, At2g47170, At5g46210, At4g27960, At3g20290 и At5g65940 – AGI генов из разных кластеров транскриптов; кл.1, кл.2 и кл.3 – обозначение кластера 1, 2 и 3, соответственно.

Поиск регуляторных контекстов в кластерах генов с разной трансляционной эффективностью

Для того, чтобы выяснить какие характеристики 5'-области мРНК, в том числе и их регуляторные контексты характерны для генов с разной трансляционной эффективностью, проведен сравнительный *in silico* анализ последовательностей 5'-областей из разных кластеров генов.

Длина 5'-НТО и эффективность трансляции

Результаты нашего сравнительного исследования длин 5'-НТО из разных кластеров генов демонстрируют отсутствие столь явной корреляции длины этой области и эффективности трансляции (Рисунок 8).

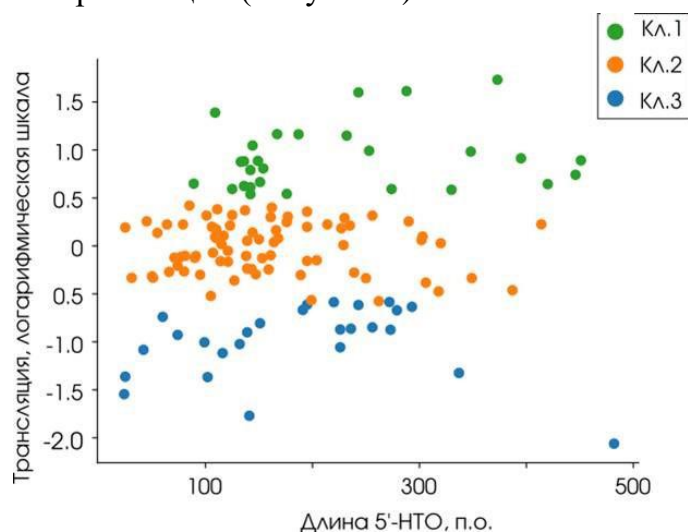


Рисунок 8. Распределение генов из разных кластеров в зависимости от длины 5'-НТО и эффективности трансляции. Кл.1, Кл.2 и Кл.3 – обозначение кластера 1, 2 и 3, соответственно.

Наличие стартовых кодонов, локализованных выше основного (uAUG) в 5'-НТО и эффективность трансляции

Мы провели поиск представленности uAUGs в 5'-НТО мРНК в разных кластерах генов, с учетом их нуклеотидного окружения, согласно классической последовательности Козак – присутствие пурина (A/G) в положении -3 и гуанина в +4. Сравнительный анализ показал, что у генов из кластера 3 (низкая трансляционная эффективность, оцененная по загрузке мРНК рибосомами) достоверно выше количество uAUGs в 5'-НТО мРНК – 20% последовательностей, которые имеют uAUG с оптимальным нуклеотидным окружением согласно последовательности Козак, по сравнению с кластером 1 (высокая трансляционная эффективность) - 3% и кластером 2 (средняя трансляционная эффективность) - 6% (Таблица 2). Таким образом, нами продемонстрировано, что uAUG с оптимальным для трансляции нуклеотидным окружением потенциально могут быть важны для эффективной трансляции.

Нуклеотидный контекст в 5'-НТО мРНК и его роль в эффективности трансляции

Результаты нашего сравнительного *in silico* исследования композиции моно- и динуклеотидов 5'-НТО из разных кластеров транскриптов позволили установить следующие закономерности: (1) отмечена достоверно высокая представленность

нуклеотидов **C** и **U** и ди-нуклеотидов **CU** и **UC** в дополнении к известному **AU** ди-нуклеотиду в 5'-НТО мРНК с высокой трансляционной эффективностью (кластер 1); (2) установлена достоверно высокая представленность нуклеотида **G** и ди-нуклеотида **GG** в 5'-НТО мРНК с низкой трансляционной эффективностью (кластер 3); (3) выявлена достоверно высокая представленность нуклеотида **A** и ди-нуклеотида **AA** в 5'-НТО мРНК со средней трансляционной эффективностью (кластер 2) (Таблица 2).

Таблица 2. Сравнительный анализ нуклеотидных композиций 5'-НТО из кластеров транскриптов с разной трансляционной эффективностью.

Параметр	Кластеры мРНК		
	1	2	3
Наличие uAUG (%)	3	6	20*
Мононуклеотиды (%)			
A	27.6	30.4*	28.3
C	22.8*	21	21.8
G	17.7	18	19.3*
U	31.9*	30.6	30.6
Динуклеотиды (%)			
AA	10.7	12.5*	10.7
AC	4.1	4.3	5.2*
AG	5.1	6.3	6.2
AU	7.7*	7.2	6.3
CA	5.4	5.8	5.4
CC	4.4	4.1	4.8
CG	4.9	4.1	4.7
CU	8.2*	7	6.9
GA	7.1	7.2	7.4
GC	3	3	3.1
GG	3.1	3.5	4*
GU	4.4	4.4	4.6
UA	4.4	4.9	4.9
UC	11.3*	9.7	8.7
UG	4.6	4.1	4.4
UU	11.6	11.9	12.7

В результате дальнейшего *in silico* анализа получены данные о нуклеотидных мотивах в 5'-НТО мРНК *A. thaliana*, которые встречаются не менее, чем в 50% последовательностях для каждого кластера транскриптов. Основываясь на данных о количестве последовательностей 5'-НТО с индивидуальным мотивом, а также максимальном количестве мотивов в единичной последовательности 5'-НТО, в каждом кластере выделены наиболее значимые мотивы: для кластера 1 (транскрипты с высокой трансляционной эффективностью) – это **TCTCT** (этот мотив имеют двадцать одна 5'-НТО, с максимальным количеством в одной последовательности - 8), и **CTCTC** (имеют шестнадцать 5'-НТО, с максимальным количеством 7); **AAGAA** (шестнадцать 16 5'-НТО, с максимальным количеством 4) и **GAAGA** (имеют пятнадцать 5'-НТО, с максимальным количеством 4) для кластера 3 (транскрипты с низкой трансляционной

эффективностью). Один мотив, который является общим для транскриптов всех кластеров – САААА.

Таким образом, результаты проведенного анализа позволяют предположить, что пиримидины, в том числе пиримидиновые ди-нуклеотиды и мотивы длиной 5 нуклеотидов характерны для 5'-НТО мРНК с высокой трансляционной эффективностью (кластер 1), тогда как пурины, пуриновые ди-нуклеотиды, а также пуриновые мотивы длиной 5 нуклеотидов являются характерным параметром для 5'-НТО мРНК со средней (кластер 2) и низкой (кластер 3) трансляционной эффективностью. Высокая представленность этих мотивов у транскриптов из разных кластеров, может указывать на то, что такие мотивы могут быть потенциально задействованы в контроле эффективности трансляции.

Окружение стартового кодона

Нуклеотидная композиция вокруг стартового кодона у транскриптов из разных кластеров оценена за счет *in silico* анализа нуклеотидов в позициях -3 и +4, а консенсусные последовательности окружения стартового кодона выведены с использованием программы *Logo* (Crooks et al., 2004). Весьма примечательно, что для транскриптов всех кластеров оптимальное окружение стартового кодона, согласно последовательности Козак присуще только для 30% транскриптов. При этом сигналы консенсусных последовательностей окружения стартового кодона для каждого кластера транскриптов были следующими: ААНАУГННГ, ААСАУГННН и ААНАУГНСН для кластера 1, 2 и 3, соответственно.

Выводы:

1. Сконструирована синтетическая последовательность, которая содержит характерные для 5'-области генов растений СG-богатые мотивы, выявленные на основании *in silico* анализа больших выборок нуклеотидных последовательностей 5'-областей генов растений.
2. Доказано, что синтетическая СG-богатая последовательность в 5'-области гена достоверно увеличивает уровень транскрипции репортерного гена и, по всей видимости, не оказывает негативного влияния на эффективность трансляции мРНК репортера.
3. С помощью *in silico* анализа проведен поиск регуляторных контекстов в 5'-области мРНК растений *A. thaliana*, ассоциированных с полисомными и моносомными фракциями, которые могут быть потенциально важными для эффективной трансляции мРНК.
4. Установлено, что пиримидиновые ди-нуклеотиды и мотивы характерны для 5'-НТО мРНК с высокой трансляционной эффективностью, тогда как пуриновые ди-нуклеотиды и мотивы ассоциированы с транскриптами, имеющими низкую трансляционную эффективность.
5. Отмечено достоверно увеличение количества uAUG (стартовые кодоны трансляции, локализованные выше основного) в 5'-НТО мРНК – 20% последовательностей, которые имеют uAUG с оптимальным нуклеотидным окружением согласно

последовательности Козак, по сравнению с кластером с высокой трансляционной эффективностью - 3% и кластером со средней трансляционной эффективностью - 6%.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи, опубликованные в изданиях, рекомендованных ВАК РФ

1. Тюрин, А.А., Кабардаева, К.В., Гра, О.А., Мустафаев, О.Н., Садовская, Н.С., Павленко, О.С., Голденкова-Павлова, И.В., 2016. Эффективность экспрессии гетерологичного гена в растениях зависит от нуклеотидного состава 5'-области мРНК. Физиология растений 63(4), 546-558.
2. Goldenkova-Pavlova, I.V., Pavlenko, O.S., Mustafaev, O.N., Deyneko, I.V., Kabardaeva, K.V., Tyurin, A.A., 2018. Computational and Experimental Tools to Monitor the Changes in Translation Efficiency of Plant mRNA on a Genome-Wide Scale: Advantages, Limitations, and Solutions. *Int J Mol Sci* 20(1).
3. Kabardaeva, K.V., Tyurin, A.A., Pavlenko, O.S., Gra, O.A., Deyneko, I.V., Kouchoro, F., Mustafaev, O.N., Goldenkova-Pavlova, I.V., 2019. Fine Tuning of Translation: A Complex Web of Mechanisms and Its Relevance to Plant Functional Genomics and Biotechnology. *Russian Journal of Plant Physiology* 66(6), 835-849.
4. Кабардаева, К. В., Тюрин, А. А., Мустафаев, О. Н., Дейнеко, И. В., Фадеев, В. С., Голденкова-Павлова, И. В., 2020. Регуляторные контексты в 5'-области мРНК растений *Arabidopsis thaliana* и их роль в эффективности трансляции. Физиология растений 67(3), 259-270.

Публикации в других изданиях, тезисы докладов

1. Тюрин, А.А., Кабардаева, К.В., Гра, О.А., Павленко, О.С., Садовская, Н.С., Голденкова-Павлова, И.В., Алиева Г.П. Способ получения синтетической CG-богатой генетической последовательности и ее использование в растениях. Патент РФ 2630650 С. Дата выдачи август 2017 г.
2. Кабардаева, К.В., Тюрин, А.А., Кимиссе, М.Г., Гра О.А., Мустафаев, О.Н., Садовская, Н.С., Голденкова-Павлова И.В. Эффективность экспрессии гетерологичного гена в растениях зависит от нуклеотидного состава 5'-области мРНК. Всероссийская научная конференция с международным участием и школа для молодых ученых "Фундаментальные и прикладные проблемы современной экспериментальной биологии растений". Москва, 23-27 ноября 2015, 316-320.
3. Кабардаева, К.В., Кимиссе, М.Г., Тюрин, А.А., Гра, О.А., Фадеев, В.С., Голденкова-Павлова, И.В. Влияние синтетической последовательности 5'-НТО с высоким содержанием динуклеотидов CpG на эффективность экспрессии гетерологичного гена в растениях. Международная научная конференция студентов, аспирантов и молодых ученых. ЛОМОНОСОВ – 2016. Секция «БИОЛОГИЯ». Москва, 11-15 апреля 2016, 280.
4. Кабардаева, К.В., Тюрин, А.А., Мустафаев, О.Н., Павленко, О.С., Гра, О.А., Садовская, Н.С., Голденкова-Павлова, И.В. Эффективность экспрессии гетерологичного гена в растениях зависит от нуклеотидного состава 5'-области мРНК. VI Всероссийский симпозиум «Трансгенные растения: технологии создания,

биологические свойства, применение, биобезопасность». Москва, 16-21 ноября 2016, 69-72.

5. Кабардаева, К.В. Влияние контекста 5'-НТО на экспрессию гетерологичного гена в растениях. В сборнике статей Международной научно-практической конференции, посвященной 129-й годовщине со дня рождения Н.И. Вавилова. Саратов, 24-25 ноября, 2016, 103.
6. Кабардаева, К.В., Мустафаев О., Садовская, Н.С., Фадеев, В.С., Бердичевец, И.Н., Тюрин, А.А. Вклад консенсусной 5'-нетранслируемой области в эффективности трансляции гетерологичных генов в растительных клетках. В материалах 21-й Международной Пущинской школы-конференции молодых ученых «Биология – наука XXI века». Пущино, 17-21 апреля. 2017, 98-99.
7. Кабардаева, К.В., Тюрин, А.А., Гра, О.А., Фадеев, В.С., Мустафаев, О., Голденкова-Павлова, И.В. Анализ мотивов 5'-НТО *A. thaliana*. В сборнике: Механизмы устойчивости растений и микроорганизмов к неблагоприятным условиям среды. Сборник материалов Годичного собрания Общества физиологов растений России, Всероссийской научной конференции с международным участием и школы молодых ученых. В 2-х частях. Иркутск, 10-15 июля 2018, 1253-1254.
8. Кабардаева, К.В., Мустафаев, О., Павленко, О.С., Тюрин, А.А., Голденкова-Павлова, И.В. 5'-НТО - важный регуляторный элемент на этапе инициации трансляции гетерологичных генов в растительных клетках. В книге: Клеточная биология и биотехнология растений. Тезисы докладов II Международной научно-практической конференции. Белорусский государственный университет, Институт леса НАН Беларуси; Редколлегия: И.И. Смолич (отв. ред.), В.В. Демидчик, В.Е. Падутов. 2018, 39-40.
9. Кабардаева, К.В., Мустафаев, О., Тюрин, А.А., Гра, О.А., Фадеев, В.С., Голденкова-Павлова, И.В. Регуляторные коды эффективной трансляции у растений: взаимосвязь локальных детерминант 5'-области мРНК растений и их комбинаций с трансляционной эффективностью транскриптов. В книге: IX Съезд общества физиологов растений России «Физиология растений - основа создания растений будущего». тезисы докладов. Казань, 2019, 195.