



# Дифференциальная экспрессия генов *Lavandula angustifolia* Mill. при адаптации *ex vitro*



Цюпка В.А., Кулешова О.Н., Эмирсалиев А.О., Жданова И.В., Гребенникова О.А., Митрофанова И.В.

Федеральное государственное бюджетное учреждение науки «Ордена Трудового Красного Знамени Никитский ботанический сад - Национальный научный центр РАН», Российская Федерация, Ялта,  
E-mail: valentina.brailko@yandex.ru

## Введение

Размножение сортов лаванды (*Lavandula angustifolia* Mill.) в культуре **in vitro** имеет ряд преимуществ по сравнению с семенным размножением или черенкованием: получение однородного, генетически-стабильного посадочного материала. При этом эффективность данного процесса зависит также от последующей адаптации растений к их культивированию в условиях открытого грунта.



Рис.1. Лаванда сорт 'Прима' ex situ



Рис.2. Лаванда сорт 'Прима' in vitro (3 пассаж)

**Материалы и Методы** Исследования проведены на сорте 'Прима' коллекции Никитского ботанического сада (Рис.1). Листья отбирали у растений **in vitro** (Рис.2) перед переносом их в условия **in vivo** и растений после 7 суточной адаптации (Рис.3) проходила в многоуровневой камере для роста растений Conviron (Канада). Культивирование растений лаванды **in vitro** проходило на питательной среде МС, дополненной 0,5 mg/L БАП или 0,5 mg/L кинетина и 0,025 mg/L НУК при температуре 22-25°C, 14-16-часовом фотопериоде, интенсивности освещения 25,0-37,5  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Для индукции ризогенеза (после 3 пассажа) использовали среду МС с различным содержанием ауксинов ИМК или НУК (0,5-1,0 mg/L). Для анализа возможных молекулярных механизмов, лежащих в основе адаптационных процессов растений при переходе от роста в культуре **in vitro** к **росту in vivo** из листьев были выделены образцы общей РНК по протоколу NucleoSpin RNA Plant (NucleoSpin, Germany), созданы 24 баркодированных RNA-Seq библиотек с использованием Illumina TruSeq Stranded mRNA Library Prep Kit. Секвенирование полученных библиотек выполнялось на высокопроизводительном секвенаторе Illumina NextSeq 550 набором NextSeq 500 High Output v2 Kit с расчетной производительностью не менее 90 млн. чтений. В результате секвенирования было получено по 3 библиотеки транскриптомных данных для каждого из состояний (**in vitro** и **in vivo**). Для полученных данных проведены подготовка, **de novo** сборка транскриптома, оценка уровней экспрессии транскриптов, поиск белок кодирующих регионов и их аннотация по базе данных SwissProt Viridiplantae. В ходе оценки дифференциальной экспрессии выявлены транскрипты с положительным (up-regulation) и отрицательным (down-regulation) регулированием. Использованы следующие программные пакеты: FastP v. 0.19.5, Trinity v.2.11.0, Kallisto v.0.46.1, TransDecoder v.5.5.0, BLAST v.2.11.0, edgeR v.3.32.1.



Рис.3. Лаванда сорт 'Прима' in vivo (7 суток адаптации под изолятором) (а), растения in vitro перед высадкой в субстрат (б)

## Результаты

Выполнены следующие этапы первичной биоинформационной обработки: разделение последовательностей по баркодам, конвертация полученных данных в формат FASTQ, оценка качества данных. Общее количество полученных данных составило **22,01 Gb**, средняя доля высококачественных последовательностей (Mean Quality Score > 30) составила **90,7%** от всех данных для девяти экспериментов.

Для последующего анализа использованы прочтения, подготовленные в программе **Fastp**. Статистический анализ данных до и после предобработки (Табл. 1, Рис. 4) позволяет заключить, что суммарная длина прочтений составляет **21,9G п.н.** при длине прочтенных последовательностей **35-76 п.н.** до предподготовки и **20,6G п.н.** при длине прочтенных последовательностей **15-76 п.н.** после. Большинство прочтений нуклеотидов имеют достаточное качество.

Таблица 1. Результаты сборки общего транскриптома

Total trinity 'genes'	Total trinity transcripts	Percent GC	Stats based on ALL transcript contigs				Stats based on ONLY LONGEST ISOFORM per 'GENE'			
			Contig N50	Median contig length	Average contig	Total assembled bases	Contig N50	Median contig length	Average contig	Total assembled bases
91157	224883	42.37	1781	828	1154.22	259563766	1398	410	777.85	70906692

\*Значение N50 – адекватно (~ 1000 – 2000 bases) для транскриптомных сборок

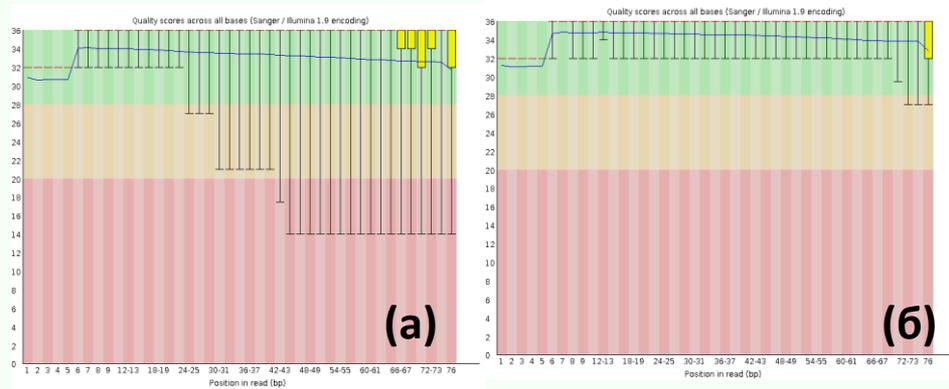
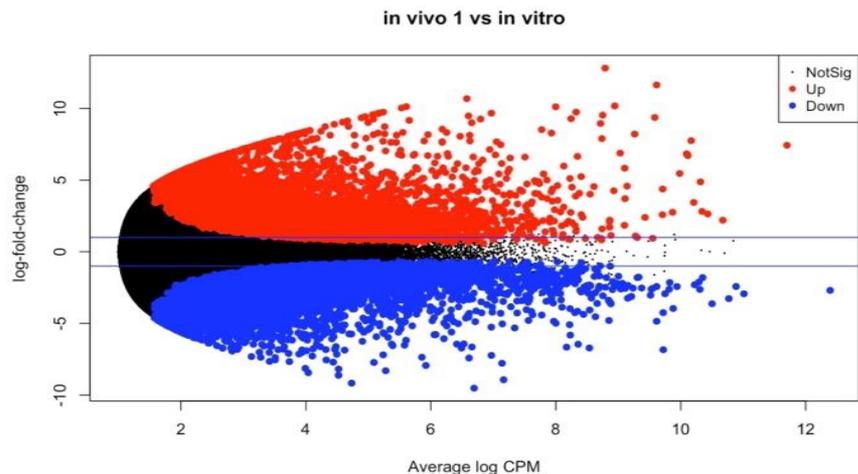


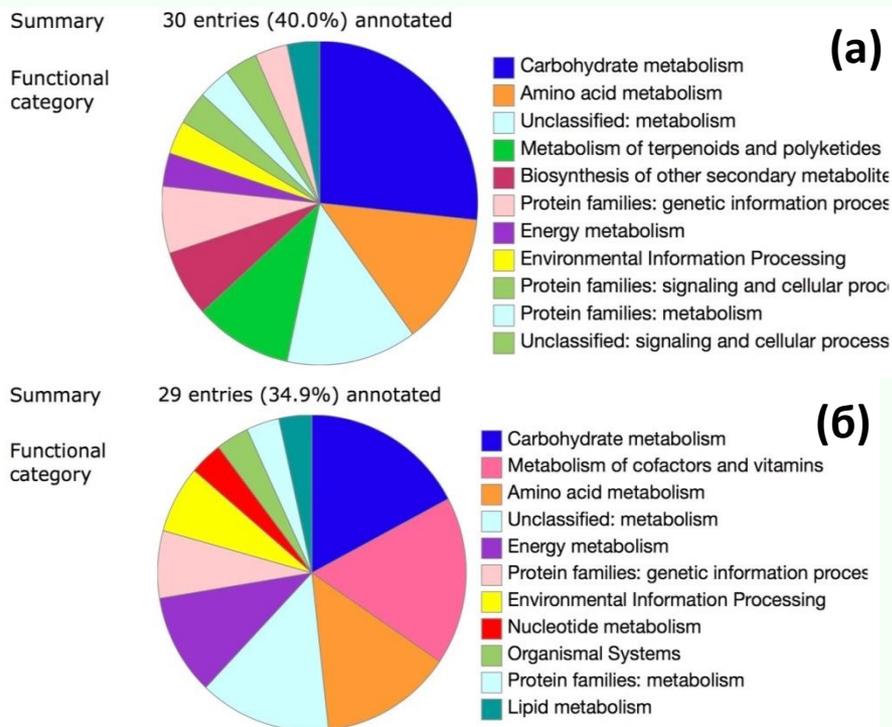
Рис.4 Агрегированная статистика показателей качества в каждой позиции для каждого нуклеотида последовательности по всем прочтениям, полученная в программе **FastQC** для библиотеки L8 прямого прочтения до (а) и после предподготовки (б) в программе **Fastp**.

\*Ось абсцисс это шкала позиции нуклеотида в прочтении, ось ординат – показатель качества (Q), обратно пропорциональный вероятности ошибки чтения нуклеотида в процессе секвенирования, выраженный в десятичной логарифмической шкале. Значения Q > 20, соответствует более чем 99% точности чтения нуклеотида (Q >30 – более 99,9% точности чтения). Синяя линия показывает среднее значение Q для каждого нуклеотида по всем прочтениям, красная – соответствует медиане Q для выделенных диапазонов на шкале положения нуклеотида.

Для последующего анализа использованы библиотеки подготовленные в программе **Fastp**.



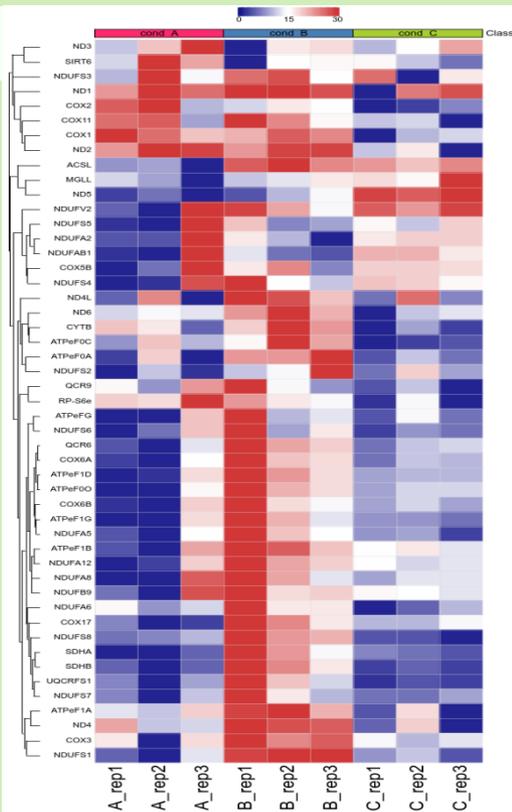
**Рис. 5 – Распределение дифференциально-экспрессируемых генов (Mean-Difference) *in vitro* - *in vivo***



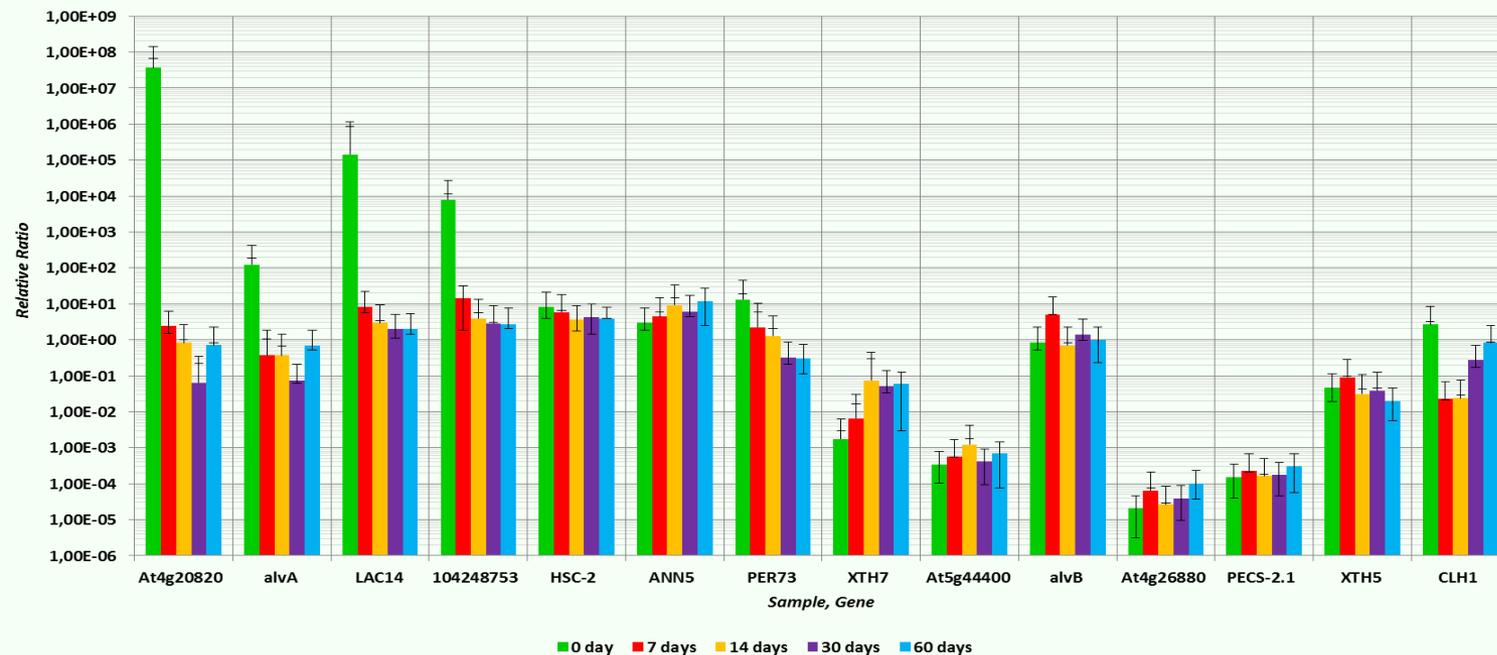
**Рис. 6 – Функциональные категории проаннотированных по базе данных KEGG Plant предсказанных в *up-regulation* (a) и *down-regulation* (b) белок-кодирующих генов состояния *in vitro* – *in vivo*.**

На основе полученных значений составлена матрица, которая проанализирована в программном пакете **edgeR v.3.32.1** со следующими настройками: **Data normalization method – TPM, Test Method - Exact Tests, P adjustment method – False discovery rate (cutoff 0,1)**. Общее количество выявленных значимых генов: **13309, Up-regulation: 5610, Down-regulation: 7699** (Puc.5). Далее был сформирован список транскриптов, наиболее значимых для Up-regulation и Down-regulation в каждой из стадий. Для полученных транскриптов проведены автоматические предсказание белок кодирующих регионов и их аннотация. Предсказание функций и метаболических путей по базе данных KEGG проведено с использованием **web-сервиса BlastKOALA** с настройками: taxonomy group – Plant, KEGG GENES database – family\_eukaryotes (Puc.6). Для всех стадий культивирования и адаптации определены и аннотированы белки, участвующие в определенных метаболических путях и функциональных иерархиях (Puc.6, Puc.7).

Благодаря полученным транскрипционным данным, для наиболее значимых транскриптов предсказанных белок кодирующих последовательностей в **Up-regulation** и **Down-regulation**, связанных с адаптационными процессами, были разработаны праймеры для последующей оценки сравнительной экспрессии нескольких генов в состояниях ***in vitro* – *in vivo* 7 суток адаптации** под перфорированным изолятором, **14 суток адаптации** под перфорированным изолятором, **30 суток адаптации** и **60 суток адаптации**. Относительный уровень экспрессии генов определяли методом qPCR, анализировали на фоне референсного гена актина (ACT2). Для отбора рабочих праймеров провели скрининг, проведя амплификацию на кДНК образцов 7 суточных адаптируемых растений.



**Рис. 7 – Heat мар ко-экспрессии генов лаванды, ассоциированных с гидротермическим стрессом при *in vitro* адаптации: A – состояние *in vitro*, B – адаптация *in vivo* 7 суток, C - адаптация *in vivo* 60 суток**



**Рис.8 – Относительная экспрессия изученных генов, нормировано к экспрессии актина (Relative Ratio в логарифмическом выражении по данным Rel Quant анализа)**

Результаты детекции позволили отобрать пары праймеров, образующие ожидаемый продукт амплификации: 3.1.a (для гена **ACT2**), 1.1.u (ген **At4g20820**), 3.2.u (не имеющий аннотации, **alvA**), 5.1.u (**LAC14**), 6.2.u (**At5g44400**), 9.1.u (**alvB**), 11.2.u (Gene ID **104248753**), 17.2.u (**At4g26880**), 20.1.u (**PECS-2.1**), 36.1.u (**HSC-2**), 44.2.u. (**ANN5**), 45.1.u. (**PER73**), 1.2.d (**XTH7**), 4.2.d (**XTH5**), 5.2.d (**CLH1**).

В результате анализа определено, что наиболее **высокоэкспрессируемыми генами растений в состоянии in vitro** являются: **At4g20820** (его экспрессия в состоянии in vitro в 10<sup>7</sup> раз выше, чем при адаптации), **LAC 14** (его экспрессия в состоянии in vitro в 10<sup>5</sup> раз выше, чем при адаптации; в течение адаптации постепенно снижается в 4 раз от начальной адаптации к 1 месяцу культивирования ex vitro), **104248753** (относительная экспрессия in vitro в 10<sup>3</sup> раз выше, в 100 раз снижается в начальный период адаптации), **alvA** (в 25 раз выше экспрессируется в условиях in vitro, чем в других состояниях), в около 10 раз повышена экспрессия генов **PER73**, **HSC-2**, **CLH1**, **ANN5**. В **начальных этапах in vivo адаптации** повышена относительная экспрессия генов **alvB**, **104248753**, **LAC 14**, **At4g20820**, **PER73** и **HSC-2**. Линейное повышение относительной экспрессии генов **при постепенной адаптации in vivo** определено для **ANN5**, **XTH7**. Для генов At5g44400, At4g26880, PECS-2.1, XTH5 отмечен низкий (следовой) относительный уровень экспрессии. Обнаружено 2 дифференциально-экспрессируемых гена, не имеющих в базах данных аннотации (однако представленных и у других растений); в связи с их экспрессией у растений лаванды при in vivo адаптации сделано предположение об их роли в указанных процессах антиоксидантной защиты, сигнальных функциях; им присвоены рабочие названия **alvA**, **alvB** (от «adaptation lavandula in vitro»).

## Выводы:

Показано, что в условиях in vitro растения лаванды экспрессируют гены, ответственные за биосинтез клеточных стенок, рост, окислительный и температурный стресс. Регуляция процессов адаптации происходит за счет работы пероксидаз, оксидоредуктаз, белков телового шока, уплотнения клеточных стенок, работы устьиц. Учитывая сравнительно низкий уровень экспрессии генов сигнальных систем во время 30- и 60-суточной адаптации, можем обоснованно предположить, что процессы адаптации завершены, растения готовы к высадке и успешному культивированию в открытом грунте.