

## Поиск генов устойчивости картофеля к золотистой картофельной нематоды с помощью технологии RNA-seq

Состав коллектива:

1. Шмаков Николай Александрович, аспирант ИЦиГ. [shmakov@bionet.nsc.ru](mailto:shmakov@bionet.nsc.ru)
2. Афонников Д. А., зав. Лаб. ИЦиГ, к.б.н. [ada@bionet.nsc.ru](mailto:ada@bionet.nsc.ru)
3. Кочетов Алексей Владимирович, зав. Лаб. ИЦиГ. Д.б.н. [ak@bionet.nsc.ru](mailto:ak@bionet.nsc.ru)
4. Егорова Анастасия Александровна, магистрант ИЦиГ

Работа выполняется по гранту РФФИ 16–16-04073 – Генетические маркеры хозяйственно-ценных признаков картофеля.

Содержание:

1. Постановка задачи  
Изучить генетические механизмы устойчивости картофеля к поражению золотистой картофельной нематодой, провести сравнительный анализ транскриптома сортов картофеля, контрастных по устойчивости к ЗКН, выделить гены, связанные с устойчивостью к ЗКН
2. Современное состояние проблемы  
Картофель – один из наиболее важных для человека видов растений из-за её повсеместного выращивания для употребления в пищу людей, корма скота и использования в промышленности. В то же время, вредители картофеля наносят огромный урон сельскому хозяйству. Один из основных паразитов картофеля – нематода *Globodera rostochiensis*, поражающая корни, уменьшая в результате массу клубней и понижая жизнеспособность растения. На сегодняшний день существуют сорта с повышенной устойчивостью к поражению этой нематодой, однако, их устойчивость недостаточна для сельского хозяйства. Из-за постоянных скрещиваний для выведения сортов генетическое разнообразие среди культурных растений сравнительно невелико. Дикие виды растений считаются важным источником генетических ресурсов для выведения новых сортов с/х растений, имеющих необходимые признаки. Одним из примеров – дикорастущая линия картофеля k-11,291 из Перу, отличающийся повышенной устойчивостью к ЗКН. Изучение механизмов устойчивости этой линии к поражению ЗКН открывает широкие перспективы для создания сельскохозяйственно важных сортов картофеля, устойчивых к ЗКН.  
Семейство генов NBS-LRR (nucleotide binding site – leucine-rich repeat) содержит около 400 известных генов, участвующих в создании растительного иммунитета. Поскольку это семейство очень обширно и отвечает за такой важный для растения признак как иммунитет к патогенам и паразитам. Биоинформатический анализ геномов и транскриптомов – перспективный подход для обнаружения NBS-LRR генов, участвующих в формировании иммунитета к конкретным патогенам и паразитам. Таким образом, изучение дикорастущей линии картофеля, устойчивой к ЗКН, может определить новые NBS-LRR гены, которые в дальнейшем можно использовать для создания устойчивых к ЗКН с/х сортов картофеля.
3. Методы  
В работе была использована технология RNA-seq для исследования транскриптомов изучаемой и контрольной линий картофеля. Секвенирование проводилось с помощью платформы IonProton компании IonTorrent. Полученные библиотеки были проанализированы по ряду параметров, таких как распределение длин ридов, распределение среднего качества по позициям внутри рида, GC-состав и другие, при помощи программы prinseq-lite. Также при помощи prinseq-lite осуществили фильтрацию библиотек по качеству и длине ридов – были удалены все риды с длиной менее 50 нуклеотидов и/или средним качеством прочтения Phred quality score < 20.

Библиотеки, прошедшие фильтрацию, были картированы на референсный геном *S. tuberosum*. Последовательность генома и разметка были взяты из базы данных Ensembl plants. Далее референсная последовательность генома была индексирована при помощи Bowtie2, и после этого было проведено картирование библиотек при помощи TopHat2. Для поиска дифференциальной экспрессии использовали линейку программ cufflinks. Программа cufflinks провела сборку транскриптов из картирования каждой библиотеки, при помощи программы cuffmerge отдельные транскрипты были объединены в транскриптом, содержащий все изформы, экспрессирующиеся в данных библиотеках, программа cuffquant провела нормализацию уровней экспрессии генов и изоформ, получив значения RPKM. Гены, имеющие суммарное значение RPKM среди всех библиотек меньше 12, были удалены из рассмотрения. После этого, при помощи программы cuffdiff для каждого из генов и изоформ были оценены уровни изменения экспрессии. Дифференциально экспрессирующимися генами (ДЭГ) были названы гены, имеющие изменение уровня экспрессии не менее двухкратного между рассматриваемыми линиями, и для которых р-значение после поправки на множественное выравнивание не превышало 0,05. Среди ДЭГ отдельно проводился поиск NB-LRR генов.

Списки генов, имеющих повышенную и пониженную экспрессию между исследуемыми линиями, далее были оценены при помощи сервиса Singular Enrichment Analysis базы данных AgriGO для поиска наиболее представленных терминов генной онтологии (ГО). При помощи сервиса BioMart базы данных Ensembl plants для ДЭГ были приведены списки идентификаторов генов NCBI, после чего было применено средство Search&Color базы данных KEGG.

Библиотеки, относящиеся к каждой из линий, были слиты вместе, после чего результирующие массивы последовательностей были использованы для сборки транскриптома de novo программой Trinity. Полученные контиги были картированы на геном *S. tuberosum* при помощи gmap, в результате чего были получены списки контигов, не имеющих гомологии с последовательностями референсного генома, которые в дальнейшем будем называть «новые контиги». Далее, новые контиги, относящиеся к устойчивой линии, были картированы на новые контиги из сборки уязвимой линии при помощи gmap, при этом транскрипты считали совпадающими при наличии перекрытия не менее 80%. Новые транскрипты из сборки устойчивой линии, не имеющие гомологии к транскриптам уязвимой линии были картированы на последовательности NB-LRR генов из публикации Jure, 2013, при помощи blastx. Порог обнаружения гомологии был выставлен как 'e-value < 10<sup>-10</sup>'.

#### 4. Результаты

Двенадцать библиотек, содержащих суммарно 48059222 коротких ридов, были получены как сырые данные секвенирования. В результате фильтрации было удалено примерно 1,5% ридов, и оставшиеся 47310018 ридов поступили в дальнейший анализ. Поиск ДЭГ осуществлялся в четырёх сочетаниях, для которых в таблице 2 приведено количество обнаруженных ДЭГ с разными направлениями изменения экспрессии.

Таблица 2. Количество ДЭГ между двумя линиями картофеля, обработанными водой и инокулированными цистами и личинками ЗКН.

Тип изменения экспрессии	Сравниваемые образцы			
	Sus_cont против Res_cont	Sus_nem против Sus_cont	Res_nem против Res_cont	Sus_nem против Res_nem
Повышение	592	208	240	278
Понижение	424	188	306	164

Для 327 генов, имеющих NB-LRR домен, была показана экспрессия в исследуемых линиях картофеля. В общей сложности для 17 из этих генов обнаружена дифференциальная

экспрессия между растениями исследуемых линий, прошедших обработку водой или суспензией цист нематоды.

Анализ терминов генной онтологии показал, что для генов, имеющих повышенную экспрессию в линии i-0144787, устойчивой к поражению ЗКН, характерны термины ГО 'ответ на окислительный стресс' ( $p=5,72 \cdot 10^{-16}$ ) и 'пероксидазная активность' ( $p=3,05 \cdot 10^{-16}$ ). Для генов, имеющих пониженную экспрессию в этой линии, перепредставлены оказались термины ГО 'трансляция' ( $p=0,0011$ ), 'сборка нуклеосомы' ( $p=9,17 \cdot 10^{-4}$ ), 'нуклеосома' ( $p=5,6 \cdot 10^{-4}$ ) и 'структурные составляющие рибосомы' ( $p=2,4 \cdot 10^{-4}$ ). Термины генной онтологии представлены на рисунке 1

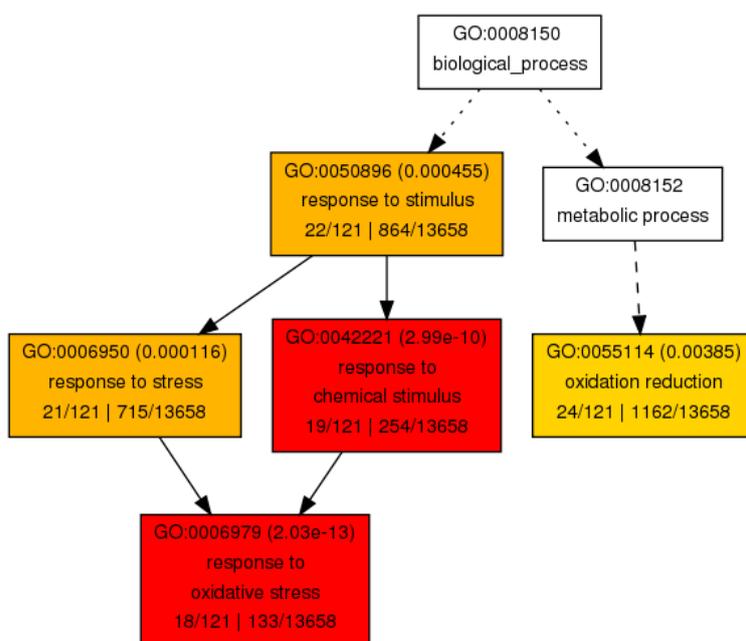


Рис 1. Термины онтологии для генов с пониженной экспрессией при поражении нематодой. При помощи базы данных KEGG было обнаружено 10 метаболических и регуляторных путей, в каждом из которых задействовано более пяти генов, имеющих различия в уровнях экспрессии между исследуемыми линиями картофеля.

### Эффект от использования ресурсов ИВЦ НГУ

Расчёты для оценки уровней экспрессии генов картофеля проводились на кластере НГУ. При этом использовались программы для очистки и анализа качества библиотек коротких ридов: CutAdapt, Prinseq-lite, FastQC; программы для картирования библиотек на референсный геном: TopHat2, Bowtie2; программы из линейки cufflinks для оценки картирования и поиска дифференциальной экспрессии генов и изоформ.

Также ресурсы кластера были использованы для реконструкции транскриптома *de novo* с помощью программы Trinity, и анализ качества сборки с использованием сопутствующих скриптов Trinity. Кроме того, были использованы программы Gmap и локальная версия Blast для поиска гомологии между полученными контигами и геномом картофеля. Таким образом, весь биоинформатический анализ был проведён на кластере ИВЦ НГУ.

Публикации, посвящённые работе

Kochetov, Alex V et al. "Differential expression of NBS-LRR-encoding genes in the root transcriptomes of two *Solanum phureja* genotypes with contrasting resistance to *Globodera rostochiensis*." *BMC plant biology* vol. 17,Suppl 2 251. 28 Dec. 2017, doi:10.1186/s12870-017-1193-1