

Тема работы

СБОРКА И АНАЛИЗ ГЕНОМОВ СОРТОВ КАРТОФЕЛЯ, ВЫРАЩИВАЕМЫХ В
РОССИИ.

Состав коллектива

Афонников Дмитрий Аркадьевич, к.б.н., ИЦиГ СО РАН, Курчатowski геномный центр
ИЦиГ СО РАН, НГУ (в.н.с.)

Каретников Дмитрий Игоревич, ИЦиГ СО РАН, Курчатowski геномный центр ИЦиГ СО
РАН, НГУ (студент НГУ, аспирант ИЦиГ СО РАН)

Генаев Михаил Александрович, к.б.н., ИЦиГ СО РАН, Курчатowski геномный центр
ИЦиГ СО РАН (с.н.с.)

Аннотация работы

Картофель *Solanum tuberosum* является одной из важнейших сельскохозяйственных культур и используется для употребления в пищу, а также для использования в промышленности (получение спирта и крахмала). Культурный картофель является тетраплоидом, размножается вегетативно (клубнями), сильно подвержен заболеваниям и воздействию вредителей. Для создания новых устойчивых сортов требуется изучение генетики картофеля на уровне генома и пан-генома. При этом сборка последовательностей до уровня хромосом требует больших затрат на секвенирование (необходимо реализовать высокое покрытие за счет длинных прочтений, что высокзатратно). Использование коротких парных прочтений не позволяет реконструировать полный геном, однако позволяет получить фрагменты, содержащие последовательности белок-кодирующих генов, провести на этой основе аннотацию функций генов и сравнительный геномный анализ. В проекте мы провели сборку и анализ геномов картофеля на основе коротких прочтений для 15 сортов картофеля, произрастающих в России. Проведена реконструкция и анализ генетической изменчивости пластидного генома картофеля, на основе *in silico* анализа проведена классификация полученных пластидных ДНК на несколько типов.

Научное содержание работы

Картофель (*Solanum tuberosum* L.) является наиболее важной сельскохозяйственной культурой, выращиваемой во всем мире. Сорта картофеля насчитывается более четырех тысяч, однако в мире до сих пор существует потребность для выведения новых сортов с желаемыми и более эффективными свойствами. Прогресс в этом направлении может быть достигнут за счет изучения генома различных видов, сортов, линий картофеля,

произрастающих в разных экологических условиях. Такой анализ открывает новые возможности для направленной селекции картофеля в будущем.

В ходе работы проведена реконструкция на основе коротких прочтений геномных последовательностей 15 сортов картофеля, широко культивируемых в России (14 сортов российского происхождения и 1 голландского). На основе полученных данных мы выявляем консервативную и переменчивую части пан-генома, оцениваем функции, связанные с ними. Проводим оценку разнообразия генов NBS-LRR. С использованием дополнительных данных о последовательностях 12 геномов южноамериканских видов и сортов картофеля мы проводим сравнительный анализ CNV, который позволил выявить структурные вариации генов, отличающиеся по частотам встречаемости у южноамериканских и российских сортов. Проведен функциональный анализ этих генов и показано, что они ассоциированы с ответом на абиотический и биотический стрессы. Выявлен ряд генов из этого пула, которые ассоциированы с туберизацией и контролем циркадных ритмов.

Проведена реконструкция и анализ генетической изменчивости пластидного генома картофеля, на основе *in silico* анализа проведена классификация полученных пластидных ДНК на несколько типов.

Методы анализа

Данные по секвенированным фрагментам 15 геномов картофеля, выращиваемого в России были получены в ИЦиГ СО РАН и Курчатовском институте (г. Москва). Данные по фрагментам ДНК образцов картофеля из Южной Америки взяты из работы (Kiriakidou et al., 2020).

Биоинформатическая обработка геномных данных включала:

1. Сборку генома картофеля до уровня контигов с помощью пакета MaSuRCA v3.4.2;
2. Оценку качества сборок геномных фрагментов посредством пакета QUAST v5.2.0;
3. Маскирование TEs с помощью полученных нами библиотек и программы RepeatMasker;
4. Идентификацию открытых рамок считывания с помощью AUGUSTUS v3.4.0;
5. Фильтрацию коротких контигов (<1000 п.н.) не содержащих ORF;
6. Оценку полноты сборки генома с помощью BUSCO v5.3.0) на основе БД Solanales DB10 (2020.08.05).

С помощью данного конвейера были обработаны последовательности сортов картофеля, культивируемые в России. В связи с тем, что аминокислотные

последовательности южноамериканских сортов картофеля не были представлены в базе данных NCBI, для этих геномов открытые рамки считывания были выявлены нами с помощью AUGUSTUS v3.4.0.

Для выявления ортологических групп и построения филогении был использован OrthoFinder v2.5.2. Функциональная аннотация каждого сорта производилась с помощью InterProScan v5.51.85. поиск доменов NBS-LRR белков, ассоциированных с растительным иммунитетом. Для этого была использована программа NLR-Parser.

Оценка вариаций числа копий генов (CNV) проводилась программой CNVpytor v0.4.1.

Филогенетические деревья визуализировались с помощью iTOL v6.

Для сборки пластидных геномов препроцессинг сырых прочтений выполняли с помощью программы fastp v0.20.1. Контроль фильтрации по качеству производился с помощью FastQC v0.11.9. Полученные фильтрованные прочтения были собраны с использованием программы GetOrganelle v1.7.6.1. Полученные сборки были аннотированы с использованием программы GeSeq. Множественное выравнивание нуклеотидных последовательностей пластов было получено с помощью программы MAFFT v7.

Результаты

Основные характеристики наших сборок (N50~11000, L50~19000, BUSCO score ~ 65-70% complete single and duplicated), согласуются с таковыми, полученными на основе сборок для коротких прочтений различных геномов картофеля. Количество идентифицированных нами белок-кодирующих генов примерно в два раза меньше, чем для полногеномной сборки тетраплоидного картофеля с учетом его гаплотипов.

Согласно нашим данным ~60% генов относятся к консервативной части пан-генома (core+softcore), 38% приходится на оболочку и 2% - на облачную часть. В целом, наши данные недооценивают содержание ORF в тетраплоидном геноме картофеля, однако функциональная классификация коровой и варибельной части пангенома согласуется с общепринятой структурой пангенома: основная часть отвечает за функцию базального гена, варибельная часть тесно связана с иммунным ответом и гены, ответственные за адаптацию к окружающей среде. Наши данные позволили оценить варибельность генов NBS-LRR в проанализированных образцах: в среднем мы идентифицировали 227 полных последовательностей NBS-LRR на геном; доля генов, классифицированных по классам CNL/TNL, составляет около 40%.

Результаты реконструкции дерева сходства между исследованными геномами картофеля представлено на рисунке 1.

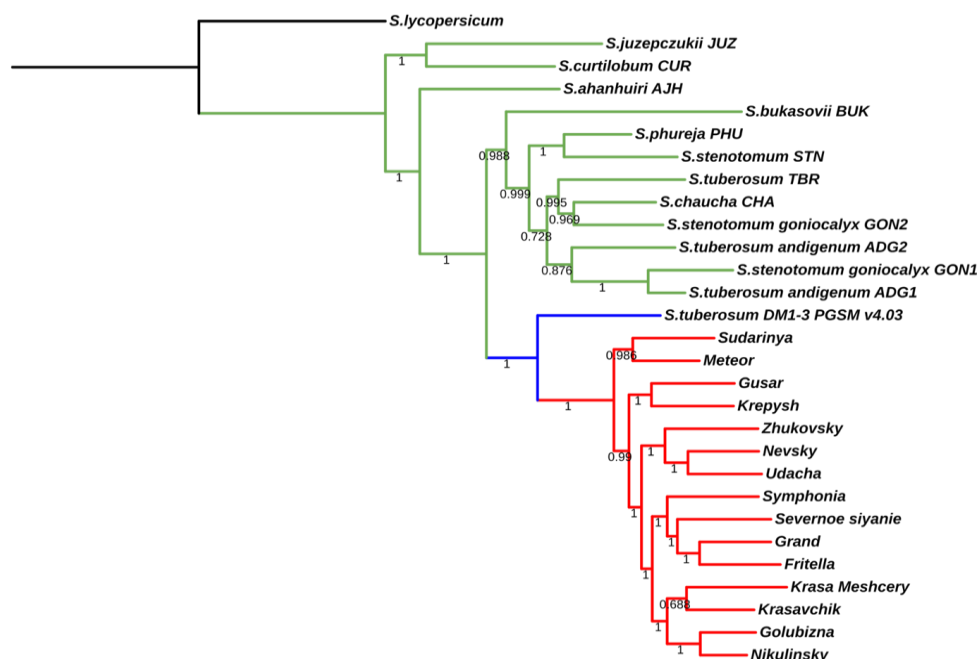


Рисунок 1. Дерево сходства между исследованными ядерными геномами картофеля. Зеленым цветом представлены южноамериканские сорта, красным – сорта, произрастающие в России, синим – референсный геном. У вершин приведены уровни достоверности кластеризации.

Гены, связанные с туберизацией и контролем фотопериода у картофеля, активно изучаются в связи с его адаптацией к длинному дню, характерному для Европейских широт. Чтобы оценить влияние CNV на эти гены у российских и южноамериканских сортов мы провели поиск генов, ассоциированных с туберизацией и контролем фотопериода среди списка, имеющих значимые различия по представленности CNV у двух этих популяций. Оказалось, что для 4 генов такие различия были нами обнаружены (Таблица 1).

Таблица 1. Список генов, ассоциированных с процессами туберизации и контролем фотопериода, для которых представленность CNV в популяциях российских и южноамериканских сортов значимо отличаются. Приведены идентификаторы гена, число значимых CNV типов делеция/дупликация в популяции российских и южноамериканских сортов, *p*-value достоверности различий и название гена.

GeneID	RU del/dup	SA del/dup	p-value	Gene description
PGSC0003DMG40000678 ¹	0/2	4/1	0.005	Metalloprotease inhibitor
PGSC0003DMG400012838 ¹	4/0	7/0	0.009	Non-specific lipid-transfer protein
PGSC0003DMG400023272 ¹	0/0	4/0	0.008	Elongation factor 1-alpha
PGSC0003DMG400015766 ²	1/8	4/0	0.010	Phytochrome A
PGSC0003DMG400029361 ³	0/15	0/3	0.002	Poly(ADP-ribose) glycohydrolase

В дополнение к известным генам, мы обнаружили еще один ген Poly(ADP-ribose) glycohydrolase, гомолог которого PARG у арабидопсиса вовлечен в функционирование циркадных ритмов.

Проведена сборка пластидных геномов для сортов картофеля, выращиваемого в России. На основе множественного выравнивания последовательностей пластидных геномов сортов, возделываемых в России, с референсным геномом cv. Desiree нами были выявлены наборы SNP. Всего идентифицировано 128 уникальных полиморфизмов, из которых 67 являются транзициями, а 61 - трансверсиями. Нуклеотидное разнообразие π , рассчитанное при сравнении пластома по всей длине, составило 0.0004. Расчет π для скользящих окон вдоль генома показывает, что варибельность последовательностей неравномерна. Существуют участки с низкой варибельностью, близкой к 0 (рисунок 3).

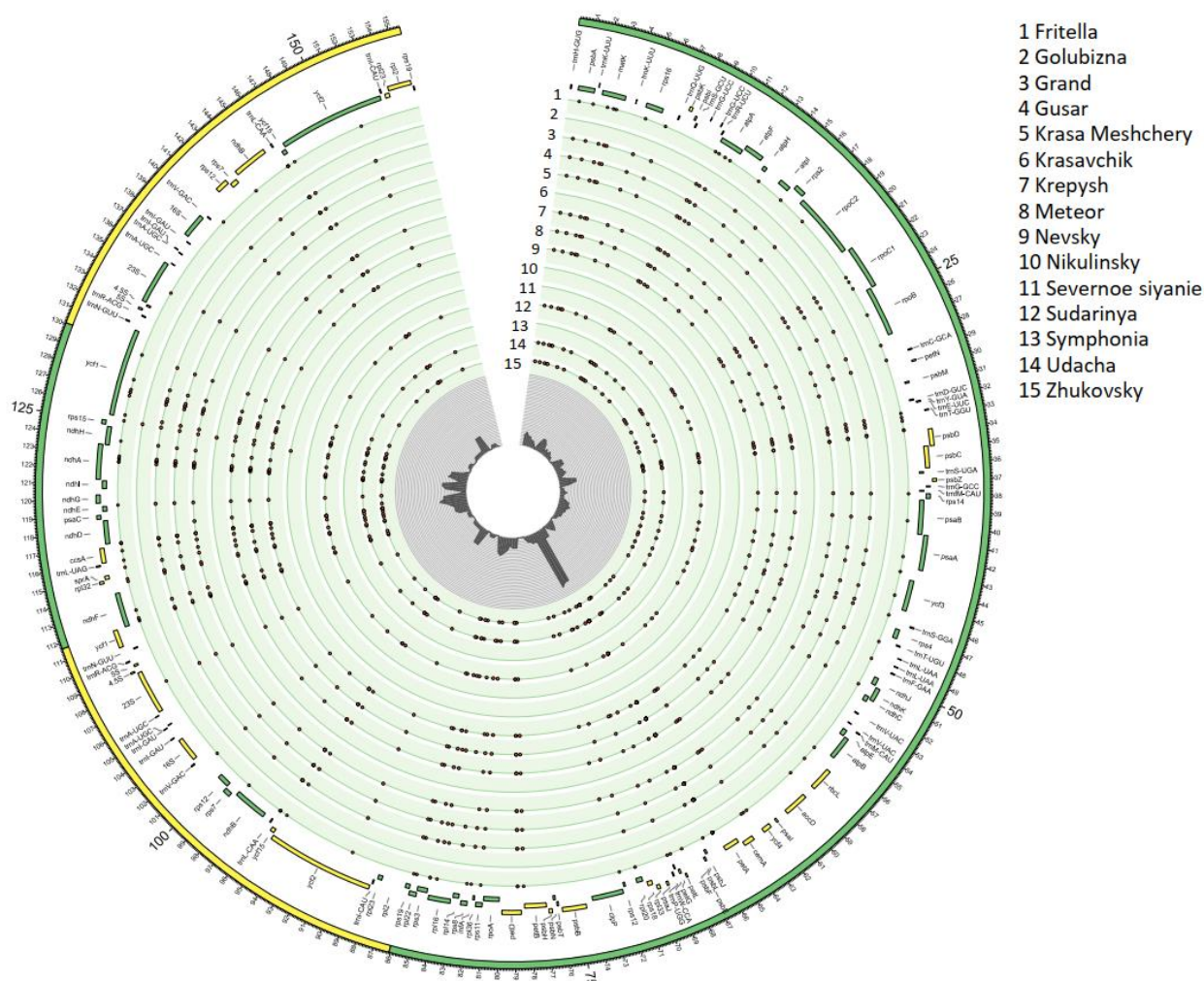


Рисунок 2. Круговая диаграмма Circos, показывающая распределение нуклеотидных мутаций для 15 пластома (светло-зеленая дорожка) и референсного пластома. На самой внешней дорожке светло-зеленый цвет отвечает за инвертированные повторы, а зеленый - за области с единственной копией. Желтыми прямоугольниками отмечены гены, лежащие

в комплементарной цепи, зелеными - в основной цепи. SNP отмечены красным цветом по одному на каждой зеленой дорожке. Серая диаграмма в центре показывает распределение нуклеотидного разнообразия, где ось X показывает координаты множественного выравнивания, а ось Y показывает нуклеотидное разнообразие.

Анализ позволил выделить наиболее вариабельные участки пластомов картофеля. Это область генов *rpl32*, *sprA*, *trnL*, *ccsA*, *ndhD*, *psaC* (позиции 113445-119462, $0.00079 < \pi < 0.00174$) и область генов *semA*, *petA*, *psbJ*, *psbL*, *psbE*, *psbF* участка LSC в положении 62364-67431, для которой значение π достигает пика ($0.00287 < \pi < 0.00304$).

Данные результаты могут быть использованы для подбора маркеров в пластидном геноме, специфичных к определенному геному картофеля.

Публикации по итогам работы

1. Karetnikov DI, Vasiliev GV, Toshchakov SV, Shmakov NA, Genaev MA, Nesterov MA, Ibragimova SM, Rybakov DA, Gavrilenko TA, Salina EA, Patrushev MV, Kochetov AV, Afonnikov DA. Analysis of Genome Structure and Its Variations in Potato Cultivars Grown in Russia. *Int J Mol Sci.* 2023 Mar 16;24(6):5713. doi: 10.3390/ijms24065713

2. Karetnikov, D.I.; Salina, E.A.; Kochetov, A.V.; Afonnikov, D.A. Assembly and Analysis of Plastomes for 15 Potato Cultivars Grown in Russia. *Agronomy* 2023, 13, 1454. <https://doi.org/10.3390/agronomy13061454>

Финансовая поддержка работы

Курчатовский геномный центр Федерального исследовательского центра ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662.

Бюджетный проект № FWNR-2022-0020 «Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции генных сетей человека, животных, растений и микроорганизмов»