



Анализ пангенома выявил влияние мобильных элементов и пloidности на эволюцию видов картофеля.

Илайда Бозана,¹ Сай Редди Ачаккагари,¹ Ноэль Л. Энглиндо нашей эры,¹ Дэвид Эллис, Хелен Х. Тай,¹ и Мартина В. Стрёмвика,²

Под редакцией Рональда Седероффа, Университет штата Северная Каролина, Роли, Северная Каролина; получено 28 июня 2022 г.; принят 9 июня 2023 г.

Картофель (*Соланум сп.*, семейство Пасленовые) — важная незерновая продовольственная культура во всем мире. У него более 100 диких родственников.*Соланумраздел Петота*, в котором представлены виды как с половым, так и с бесполом размножением и различным уровнем пloidности. Пангеном *Соланумраздел Петота* состоит из 296 образцов, включая диплоиды и полиплоиды, сравниваемые по вариации присутствия/отсутствия (PAV). *Петота* ядро (гены, общие как минимум для 97% образцов) и геномы оболочки (общие для 3–97%), обогащены основными молекулярными и клеточными функциями, в то время как облачный геном (гены присутствуют менее чем в 3% образцов-членов) показало обогащение мобильными элементами (TE). Было проведено сравнение PAV у одомашненных и диких образцов, и на основе PAV было построено филогенетическое дерево, группирующее образцы в разные клады, аналогично предыдущим филогениям, полученным с использованием ДНК-маркеров. Подход к кладовому пангеному выявил абиотический стрессовый ответ среди основных генов в кладе 1+2 и кладе 3, а также цветение/клубнеобразование среди основных генов в кладе 4. Содержание TE различалось между кладами, при этом клада 1+2 состоит из видов из Северной и Центральной Америки с репродуктивной изоляцией от видов других клад, имеющих значительно более низкое содержание TE по сравнению с другими кладами. Напротив, образцы с историей размножения *in vitro* были идентифицированы и имели высокие уровни TE. Результаты указывают на роль TE в адаптации к новой среде, как естественной, так и искусственной. *Соланумраздел Петота*.

Петота пангеном | транспозоны | адаптация | полиплоидия | Соланум

Картофель (*Паслен туберозный* L.) является третьей по важности культурой и самой важной незерновой культурой для продовольственной безопасности в мире (1). Это часть *Соланумраздел Петота*, который насчитывает более 100 видов, адаптированных к разнообразной сезонности и осадкам (2, 3). Самое большое разнообразие диких видов *Соланумраздел Петота* встречается в высокогорных, адаптированных к холоду горных местообитаниях в центральной Мексике и центральных Андах (2–4), а культивируемые виды и их местные сорта происходят от западной Венесуэлы до южного Чили, с центром происхождения в озере Титикака, которое находится между Перу и Боливия (5).

Одомашнивание произошло примерно 10 000 лет назад в высокогорьях Анд (ныне юг Перу), где различные местные сорта картофеля до сих пор выращиваются на большой высоте (от 3 000 до 4 000 м) (5, 6). Одомашнивание произошло в результате человеческого и естественного отбора клубней большего размера, более низкого содержания гликоалкалоидов, а также выбора клубней привлекательной формы и цвета. Многие виды картофеля подверглись дальнейшей гибридизации или пloidным изменениям, что привело к образованию обширного генофонда (7), состоящего из более чем 4500 местных сортов картофеля (в Перу называемых «папас нативас») (8).

Уровни пloidности являются основным барьером для селекции многих сельскохозяйственных культур при проникновении генетического материала диких видов. Основное (гаплоидное) число хромосом картофеля составляет 12, хотя пloidность варьируется от диплоидной до гексаплоидной, при этом большинство видов классифицируются как диплоидные (9). Местные сорта варьируются от диплоидных до пентаплоидных. Полиплоидизация (а затем и видообразование) усиливается в более суровых условиях окружающей среды, особенно в холодных и засушливых, а также при биотическом стрессе (10). Большая часть современного выращиваемого в коммерческих целях культурного картофеля, особенно в Северной Америке и Европе, является ауотетраплоидной ($2n=4x=48$), хотя в селекции для придания определенных признаков используются как диплоиды, так и полиплоиды. Иногда диплоидные родственники культивируемого и дикого картофеля производят 2n гамет, что позволяет им гибридизоваться с тетраплоидами. Считается, что триплоидные виды возникают в результате интерплоидной гибридизации между n и 2n гаметами (11). Однако в Андах культивировались тетраплоиды (*C. Tuberosum* группа Андигена; $2n=4x=48$) являются результатом автополиплоидизации диплоидов (11).

Нигде биоразнообразие не используется так широко, как в сельскохозяйственных системах по всей Южной Америке, чтобы максимально эффективно использовать возделываемые ресурсы для обеспечения продовольственной безопасности. Культивируемый и дикое картофеля представляет большой глобальный интерес, и многие геномы картофеля были секвенированы и собраны (12–23). И все же, поскольку *Соланумраздел Петота* является сложным, поскольку виды демонстрируют переменные уровни пloidности, как бесполого, так и полового размножения, самонесовместимости, презиготических и постзиготических репродуктивных барьеров (24), а также потому, что последовательности полиплоидного генома

Значение

The *Соланумраздел Петота* включает сорта картофеля, местные сорта и дикие виды. Последние являются источником генетического разнообразия, обеспечивающего устойчивость к биотическому и абиотическому стрессу, питательные и адаптивные свойства, имеющие решающее значение для адаптации к изменению климата, повышения устойчивости растениеводства и улучшения качества питания для здоровья и продовольственной безопасности. Эта работа сравнивает диплоидные и полиплоидные виды по секциям *Петота* рассматривая присутствующие и отсутствующие гены в масштабе пангенома, что хорошо отражает таксономическое дерево. Репродуктивно изолированные клады демонстрируют различия в мобильных элементах (TE), что доказывает роль TE в видообразовании растений и открывает новый взгляд на сложность таксономии картофеля.

Принадлежность автора: Департамент растениеводства, Университет Макгилла, Сент-Анн-де-Бельвю, QC H9X 3V9, Канада; «Международный центр картофеля, Лима 15023, Перу; Министерство сельского хозяйства США – Служба сельскохозяйственных исследований (USDA ARS) Исследование зародышевой плазмы мелкого зерна и картофеля, Абердин, ID 16915 2700W; и Центр исследований и разработок Фредериктона, Сельское хозяйство и агропродовольствие Канады, Фредериктон, NB E3B 4Z7, Канада

Вклад авторов: исследования, разработанные IB, SRA, ННТ и MVS; IB, SRA и MVS провели исследование; MVS предоставила аналитические инструменты; IB, SRA, NLA, DE, ННТ и MVS проанализировали данные; MVS курировал исследование в области биоинформатики; и IB, SRA, NLA, DE, ННТ и MVS написали статью.

Авторы заявляют об отсутствии конкурирующих интересов.

Эта статья представляет собой прямую публикацию PNAS.

Copyright © 2023 Автор(ы). Опубликовано PNAS. Эта статья в открытом доступе распространяется под лицензией Creative Commons с указанием авторства 4.0 (CC BY).

1IB и SRA внесли равный вклад в эту работу.

2Кому может быть адресована корреспонденция. Электронная почта: martina.stromvik@mcgill.ca.

Эта статья содержит вспомогательную информацию в Интернете по адресу <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.221117120/-/DCSupplemental>.

Опубликовано 24 июля 2023 г.

трудно решить (25), в более крупных исследованиях генома картофеля на сегодняшний день использовались либо только диплоиды (23), либо тетраплоиды (21), и не существует сборки пангенома, которая интегрировала бы и синтезировала огромное генетическое разнообразие и различные уровни пloidности этой секции. Пангеном может ответить на более широкие вопросы родства и адаптации и предоставить более глубокие знания о разнообразии интересующей группы.

Здесь мы представляем *Летотапангеном*, основанный на последовательностях 296 диплоидных и полиплоидных образцов различных видов, местных сортов, сортов и коммерческих сортов, представляет собой наиболее таксономизированный пангеном картофеля на сегодняшний день. Он был построен с использованием *S. Tuberosum* эталонный геном группы *Tuberosum* RH89-039-16 v3 (далее называемый RH) (19), к которому было добавлено ~7,3 Гб дополнительных последовательностей и 56 843 дополнительных гена, что в общей сложности составило 132 355 пангенов. Данные о наличии/отсутствии вариации (PAV) были использованы для исследования генома ядра, оболочки и облака. *Летотапангеном* наряду с кластическими различиями в членских образцах. Было обнаружено, что мобильные элементы (TE) различаются между кладами и особенно преобладают в образцах, размножаемых *in vitro*. Кластеризация на основе PAV использовалась для построения филогенетического дерева, которое поддерживает предыдущие таксономические группировки и раскрывает функции генов, участвующие в адаптации к изменениям окружающей среды и климата, методам выращивания, а также дает представление о TE и видообразовании в разделе *Летота*.

Полученные результаты

Пангеном *Соланум* Раздел *Летота*. Всего 296 разнообразных

Образцы картофеля были использованы для создания суперпангенома *Соланум* раздел *Летота* использовать обширную информацию, содержащуюся в текущих данных о последовательностях генома, и выявить вариации, отсутствующие в существующих эталонных геномах. Эти образцы представляют широкий спектр диких видов, местных сортов и сортов с разными уровнями пloidности (т.е. 2x, 3x, 4x и 5x) и разным географическим происхождением. Общедоступные сборки генома 33 образцов были использованы непосредственно при конструировании пангенома, тогда как остальные образцы были собраны *de novo*, чтобы создать в общей сложности 154 Гбит контигов (*Приложение СИ, Таблица S1 и Рис. S1*). Сборки короткого чтения имеют значения N50 в диапазоне от 1393 до 16809 п.н., тогда как сборки длительного чтения имеют лучшую непрерывность и полноту (*Приложение СИ, рис. С2*). Эти сборки вместе с общедоступными сборками генома были сопоставлены с эталонным геномом (RH) для получения неререференсных последовательностей. Эти неререферентные последовательности были подвергнуты множеству циклов удаления избыточности для получения окончательного набора избыточных неререферентных последовательностей. Новые невыровненные последовательности, добавленные к эталону, состояли из 5 146 040 контигов, которые имели <90% идентичности с эталоном, с длиной не менее 500 п.н. и общим размером ~7,3 Гбп, из которых ~5,9 Гбп (79,5%) состоят из повторяющихся элементов. Основная часть неререферентного генома состоит из последовательностей диких видов клады 1+2 и клады 3 (*Приложение СИ, рис. S3*). Все неререферентные последовательности были добавлены к эталонному геному, чтобы сделать раздел *Летотапангеном*. Пангеном состоит из 12 хромосом (гаплотипически разрешенных, 24 гаплотипа) плюс 5 148 884 контигов общим размером ~9,05 Гбп. В пангеноме имеется 37 083 контига размером не менее 10 kb и 176 938 контигов размером не менее 5 kb. Существует 2 957 143 (~ 57,4%) контигов с >90% повторяющихся последовательностей. Пангеном состоит на 100% из ортологов BUSCO из *Виридипланта* набора данных, а также более 99% ортологов из *Эмбриофитум Соланалес* набора данных (*Приложение СИ, рис. S4*). Аннотация неререферентных последовательностей позволила получить в общей сложности 56 843 генные модели, представляющие 54 135 избыточных генов. Среди неререферентных генов в общей сложности 42 779 генам были присвоены функции и/или домены Pfam посредством функциональных аннотаций. Эти

неререферентные гены были добавлены к справочным аннотациям RH, в результате чего в пангеноме образовалось в общей сложности 132 355 генов.

Вариация PAV в генах, кодирующих белки. PAV может идентифицировать ген разнообразие и вариации среди отдельных образцов (26). В нескольких пангеномных исследованиях были успешно использованы данные PAV для понимания вариаций содержания генов и их влияния между популяциями (26–29). Эффективный вызов PAV может быть выполнен с использованием последовательностей малой глубины и подхода «карта-панорама», который, как было показано, оказывает минимальное влияние на количество последовательностей, используемых при вызове PAV (30). В этом исследовании мы провели PAV-анализ, чтобы включить образцы с низкой глубиной чтения. PAV-анализ показал, что количество генов в каждом образце варьировало от самого низкого у диплоидного дикого образца. *S. полиадиениум* (PI 347770) с 54 867 генами, наивысшими из которых являются триплоид и пентаплоид, *Паслен юзепчукий* (CIP 706050) и *Паслен куртилубум* (CIP 702937), с 81 418 и 81 246 генами соответственно. Частоты присутствия были рассчитаны для каждого гена, которые были разделены на три группы (ядро, оболочка и облако) на основе их общих частот (рис. 1). В большинстве образцов (>97% образцов) имеется 23 055 коровых генов (рис. 1).А). Аналогично, 25 776 облачных и 83 524 панцирных генов присутствуют у <3% и от 3 до 97% образцов соответственно (рис. 1).А). Больше количество вариабельных генов и очень низкое количество коровых генов наблюдалось у *Летотапангеном*, демонстрируя изменчивость, существующую в *Соланум*, как было также ранее показано в пантранскриптомном исследовании, в котором было зарегистрировано в общей сложности 96 886 избыточных репрезентативных транскриптов, а количество генов, специфичных для сорта, было больше, чем количество основных генов в сборке только из четырех генотипов картофеля - *S. Tuberosum* группа *пуреджа* (DM) и три тетраплоидной группы *Tuberosum* (сорт Désirée, сорт Ruwal и селекционный клон PW363) (31).

Этот пангеном состоит из гораздо большего количества видов (особенно диких видов), чем любой предыдущий пангеном растений (23, 26–28, 32), и поэтому его можно рассматривать как суперпангеном или супергеном. По мере добавления большего количества генов количество основных генов уменьшалось, а общий пангеном быстро увеличивался, пока после добавления около 80 геномов не было достигнуто плато (рис. 1).Б). Основной геном функционально обогащен фундаментальными молекулярными и клеточными процессами, включая процессы модификации клеточных белков (GO: 0006464) и процессы метаболизма фосфора (GO: 0006793), демонстрирующие наибольшее количество генов с низкими значениями *p* (*Приложение СИ, рис. S5*). В то время как геном оболочки обогащен регуляцией биосинтеза РНК (GO: 2001141), реакции на стресс (GO: 0006950) и трансляции (GO: 0006412) функциями, связанными с реакцией на патогены, реакцией на питательные вещества и развитием, а также с тритерпеноидами. биосинтез (GO:0016104) (гены, участвующие в синтезе гликоалкалоидов — типичных защитных метаболитов *Соланум* растений), облачный геном заметно обогащается в ответ на воду (GO:0009415), ДНК-опосредованную транспозицию (GO:0006313), развитие эмбриона (GO:0009793), интеграцию ДНК (GO:0015074) и регуляцию биосинтеза РНК. (GO:2001141), с функциями, связанными с функциями TE (транспозиции), регуляцией экспрессии генов, реакциями на стресс и фотосинтезом.

Среднее количество генов у диких видов значительно ниже, чем у местных сортов и культурных сортов: 63 567, 65 856 и 71 950 соответственно (рис. 2).А). Более того, у полиплоидов больше генов по сравнению с диплоидами, а у диплоидных сортов больше генов, чем у диплоидных диких видов (рис. 2).Б) (*Приложение СИ, рис. S6*), показывая, что увеличение содержания генов в картофеле в основном обусловлено пloidностью, но одомашнивание и селекция также играют свою роль. Любопытно, что диплоидные выбросы — те, которые имели гораздо более высокое содержание генов в дикой группе — все сохранились в течение многих лет в культуре тканей по сравнению с остальными диплоидными видами, которые произошли от односемянного происхождения (рис. 2).А).

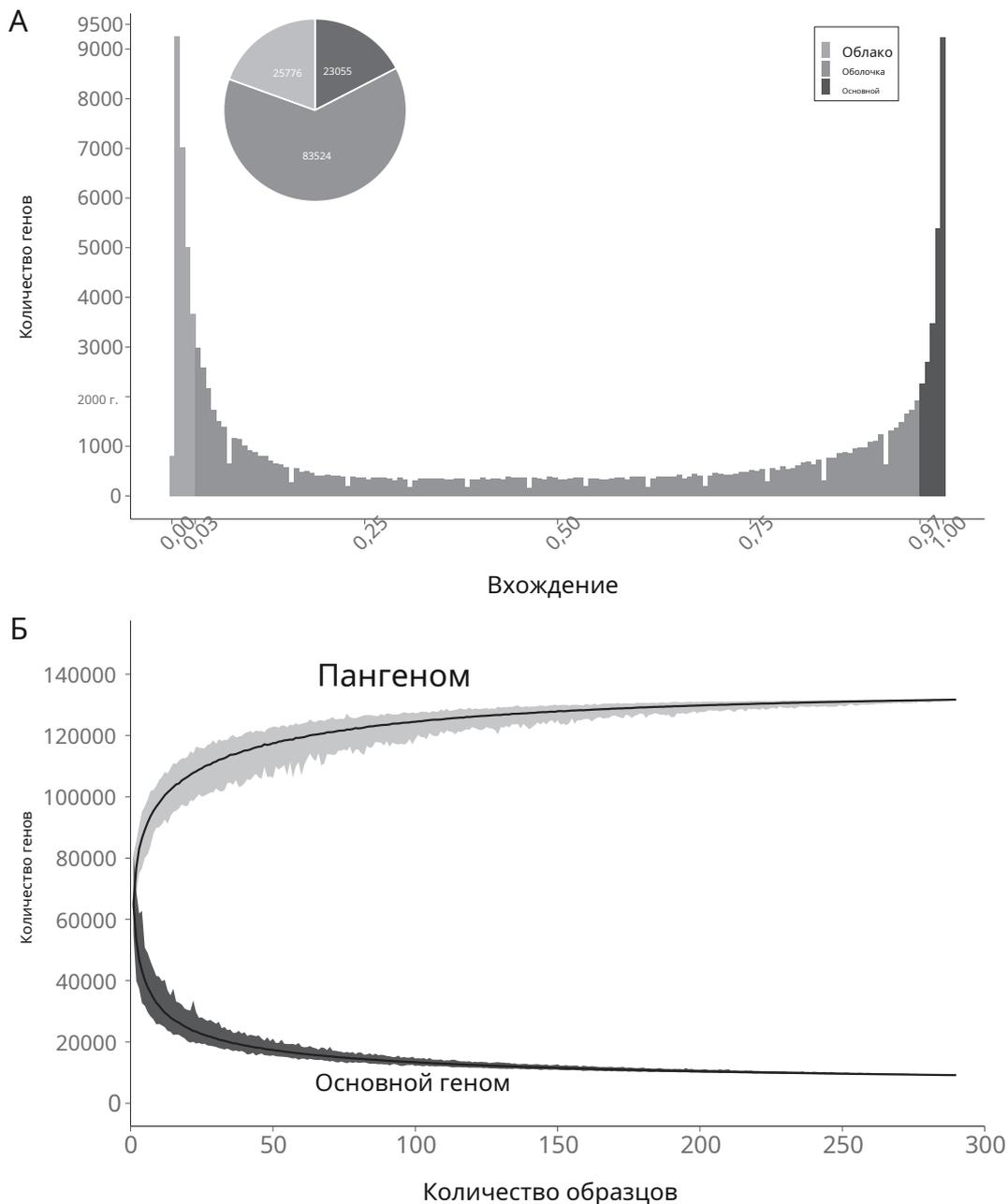


Рисунок 1. The *Solanum* урздел *Петота* пангеном. (А) Гистограмма генов ядра, оболочки и облака в *Solanum* урздел *Петота* пангеном с круговой диаграммой, показывающей количество генов, обнаруженных в каждой категории. (Б) Кривые разрежения показывают увеличение размера пангенома и уменьшение размера ядра генома по мере добавления новых образцов. Пангеном достигает плато около 80 образцов.

Чтобы выявить сигналы отбора в PAV для местных сортов и сортов, сравнивали частоты генов между дикими видами, местными сортами и сортами. Были идентифицированы гены со значительными различиями PAV, причем гены с более высокой частотой встречаемости у местных сортов или сортов по сравнению с дикими видами были отнесены к категории «благоприятных» генов, а гены с более низкой частотой, чем у диких видов, были отнесены к категории «неблагоприятных» генов (27, 33). Всего среди сортов и местных сортов выявлено 8684 благоприятных гена и 4814 неблагоприятных генов. Кроме того, было обнаружено 3617 благоприятных генов, специфичных для местных сортов (где частота присутствия местных сортов > частота присутствия сортов) и 4717 благоприятных генов, специфичных для сортов (частота присутствия сортов > частота присутствия местных сортов) (Приложение СИ, рис. S7). Анализ обогащения выявил функции, связанные с реакцией на стресс, развитием эмбриона, заканчивающимся покоем семян, развитием корней, клеточным гомеостазом глюкозы,

и фотосинтез в генах, благоприятных для сортов. Репарация ошибочных спарений, фотосинтетическая цепь переноса электронов, регуляция точности трансляции и аутофагия пероксисом были обогащены генами, благоприятными для местных сортов (Приложение СИ, рис. S8). Эти результаты позволяют предположить, что одомашнивание включало изменение генома в результате процессов дупликации и потери генов. Функциональный анализ генов с PAV предполагает улучшение фотосинтеза, модификацию метаболизма и развитие местных сортов и сортов по сравнению с дикими видами.

Кластеризация образцов на основе содержания генов. Данные ПАВ

были использованы для построения филогенетического дерева максимального правдоподобия (рис. 3) для сравнения группировок на основе PAV с существующими концепциями клад картофеля. Это филогенетическое дерево согласуется с предыдущими филогенетическими исследованиями (2, 14, 15, 23, 34). Четкое разделение

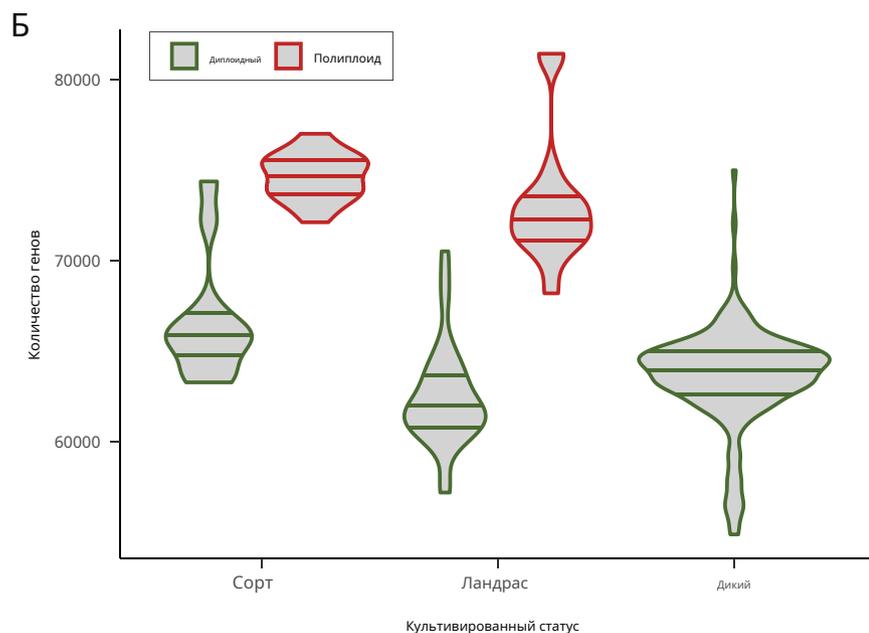
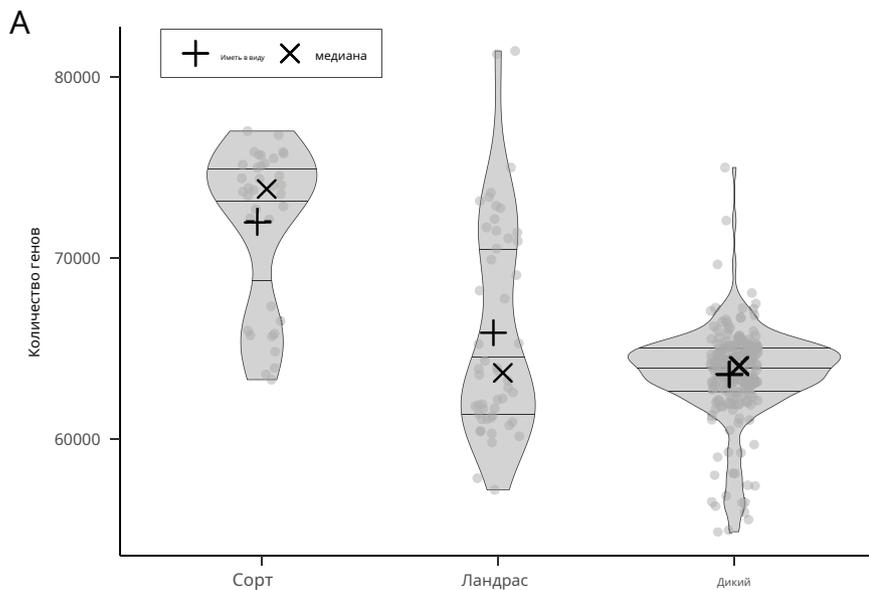


Рис. 2. Генный состав образцов, включенных в *Solanum* раздел *Петотапангеном* на основе PAV. (A) Распределение генов у диких видов, местных сортов и сортов. *Петота* присоединения. Линии на графике скрипки показывают расположение верхнего квартиля, медианы и нижнего квартиля. Знак «+» — это средневзвешенное значение, а «X» — медиана. (B) Распределение генов у диких видов, местных сортов и сортовых образцов с разным уровнем пloidности.

В этой классификации наблюдается присутствие разных клад, как также видно из ссылки. 15. Образцы в филогении на основе PAV следовали за ранее классифицированными кладами 1+2, 3 и 4. Клада 4 была подразделена на кладу 4 на север и юг. Все образцы, ранее отнесенные к видам клады 3, сгруппированы вместе, за исключением неоднозначных видов, *Паслен кахамаркенский* и *Паслен согарандинум*. Последний ранее считался видом клады 4 (2) или обозначался как кладу 3 (15), как указано в *Solanaceae Source* (35); однако настоящее дерево поместило его вне клады 3 и кладу 4 в отдельную группу, аналогично тому, что сообщалось в предыдущем исследовании (23). Это конкретное присоединение *С. Кахамаркенсе* (PI 230522) помечен как «гибридное семя» и сгруппирован с кладой 4 на юге в данных пластоме (34), тогда как ядерные данные зафиксировали разные места для этого образца с помощью квартетов разложения сингулярных значений (SVD) и анализа максимального правдоподобия (ML) (15).

Виды кладу 4 в дальнейшем разделились на группы кладу 4 южная и кладу 4 северной, как также показано в других исследованиях (2, 15, 34). Южная группа кладу 4 состоит из видов южного региона.

включая Боливию и Аргентину, а также *Паслен бородавчатый* вид из Мексики (*Приложение СИ, рис. S9*). Северная группа состоит из диких видов из Перу, а также местных сортов и сортов. Очень интересны группировки, наблюдаемые в северной группе.

Диплоидные местные сорта (*Solanum phureja*, *Solanum stenotomum*, и *Паслен гониокаликс*) сгруппированы вместе и имеют схожие потери/прибыли генов с *Паслен кандоллеанский* разновидность (*Solanum bukasovii*, *Solanum multidissectum*, *Solanum abancayense*, *Solanum canasense*, *Solanum marinasense*, *Solanum ambosinum*, и *Паслен пампасный*). Все образцы Andigena и Tuberosum тесно сгруппированы друг с другом и с горьким картофелем (*С. куртилобум*, *Пасленум аянхуири*, и *С. юзепчукий*).

Анализ главных компонент (PCA) ПАВ также показал разделение, аналогичное таковому на филогенетическом дереве (рис. 4).A). Наблюдалось четкое разделение видов кладу 1+2, кладу 3 и кладу 4, причем *С. согарандинум* и *С. Кахамаркенсе* образцы выделены в отдельную группу. Южные дикие виды образовали отдельный кластер, отделенный от остальных видов кладу 4. A

Клада 4 юг



Рис. 3. Филогенетическое дерево максимального правдоподобия, построенное с использованием данных PAV для *Solanum* из раздел *Петотопангеном*. Образцы сгруппированы в три основные клады с сильной бутстреп-поддержкой (100): клада 1+2, клада 3 и клада 4. Клада 4 далее делится на кладу 4 южную и кладу 4 северную. Значения начальной загрузки более 90, за исключением значений в основных узлах, были удалены, чтобы сделать рисунок более понятным.

Наблюдается тесное скопление, включающее дикие виды северной клады 4 и культивируемые виды. Последующие анализы исключают *S. согарандинум* и *S. Кахамаркенсе* образцы из клады 3.

Вариация содержания генов позволяет различать клады и подгруппы. Чтобы понять различия в общем количестве генов среди образцов, была создана тепловая карта, показывающая количество генов, общих для образца с другими образцами (рис. 4).Б.

Для каждой из трех клад образцы внутри клады имеют больше общих генов по сравнению с образцами между кладами (*Приложение СИ, рис. S10A*), при этом среднее количество общих генов в кладе 1+2, кладе 3 и кладе 4 составляет 49 939, 52 105 и 54 953 соответственно. Данные PAV также использовались для сравнения содержания генов между тремя кладами: кладой 1+2, кладой 3 и кладой 4. Существует 68 492 гена, которые имеет по крайней мере один член этих клад, а 6 251, 6 424 и 25 239 генов однозначно встречаются в кладе 1+2, кладе 3 и кладе 4 соответственно (рис. 4).С. Сходным образом,

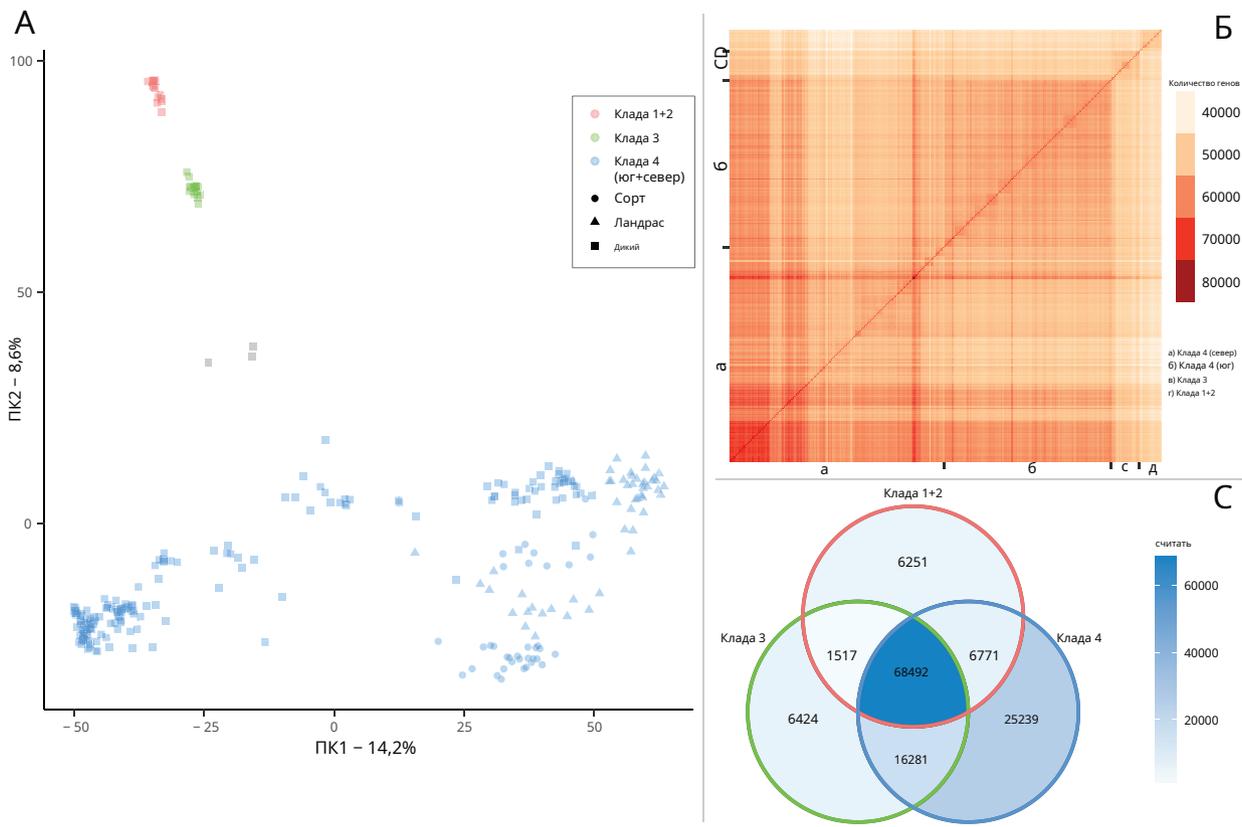


Рис. 4. Анализ различных клад и подгрупп в пангеноме *Соланум* из раздела *Петота*. (А) PCA на основе PAV показывает кластеризацию образцов. (Б) Тепловая карта, показывающая общее количество генов среди всех образцов. Образцы упорядочены на основе филогенетического дерева. (С) Диаграмма Венна генов, присутствующих в трех кладах: кладе 1+2, кладе 3 и кладе 4. Ген считается присутствующим в кладе, если он есть хотя бы у одного образца из этой клады. (Д) Блок-график сборок генома de novo и количества TE между указанными группами.

Сравнение содержания основных генов, специфичных для каждой клады, показало, что существуют схожие функции, общие для основных геномов каждой клады, и что в кладе 1+2, кладе 3 и кладе 4 присутствует 5612, 3880 и 2892 уникальных коровых гена соответственно. (Приложение СИ, рис. S10Б). Термины GO (GeneOntology) уникального основного генома клады 1+2 указывают на обогащение при холодной акклиматизации.

(GO:0009631) наряду с функциями регуляции абсцизовой кислоты (GO:0009787) и метаболизма инозитола (GO:0019310), которые связаны с холодной реакцией (GO:0009631) Приложение СИ, рис. S11). Уникальный основной геном клады 3 обогащен терминами GO, участвующими в ультрафиолете (УФ) (GO:0009411), реакции на холод (GO:0009409) и транспорте полиаминов (GO:0015846), связанных с реакцией на стресс.

в дополнение к биосинтезу тритерпеноидов (GO:0016104), который производит защитные метаболиты, включая гликоалкалоиды. Уникальный основной геном клады 4 также обогащен генами, участвующими в биосинтезе тритерпеноидов, и генами, участвующими в положительной регуляции транскрипции. Кроме того, уникальные основные гены клады 4 отмечены как регулирующие развитие цветков (GO: 0009909), что также является аннотацией генов, участвующих в регуляции клубнеобразования.

Содержание гена PAV также было использовано для выявления различий между подгруппами: дикие виды в кладе 4 юг (South Wild), дикие виды в кладе 4 север (North Wild), группа с образцами с историей размножения *in vitro* (в *in vitro*), а также группу с оставшимися культивируемыми образцами (Культивируемые/ландрасы) без истории размножения *in vitro*. Приложение СИ, рис. S12А показывает, что существует 3947, 316, 3037 и 2547 генов, которые являются уникальными для группы South Wild, North Wild, культивируемых/ландрасов и *in vitro* соответственно (по крайней мере, у одной особи из этой группы они есть). Существует 950 основных генов, которые являются общими для всех групп, за исключением образцов, консервативных *in vitro*, и 768 генов, которые присутствуют только в основном геноме подгруппы *in vitro*. Приложение СИ, рис. S12Б. Функциональное обогащение уникальных коровых генов *in vitro* выявило функции, связанные с транспозицией, интеграцией ДНК, модификацией РНК и рекомбинацией ДНК (Приложение СИ, рис. S13А). Гены, отсутствующие в группе *in vitro*, но присутствующие в остальных группах, были обогащены функциями, связанными с биосинтезом рибофлавина, метаболическим процессом ДНК, регуляцией формы клеток и секрецией белка (Приложение СИ, рис. S13Б).

содержание в пангеноме. Пангеном состоит из

75,5% TEs, а анализ содержания генов показал функциональное обогащение генами метилирования ДНК и TE в облачном геноме и подгруппе *in vitro* (Приложение СИ, рис. C5 и C13А). Сборки участников были просканированы на предмет содержания TE и наличия генов, связанных с транспозицией (рис. 4). Никаких отклонений от качества сборки в количестве идентифицированных TE не наблюдалось (Приложение СИ, рис. S14). Образцы клады 1+2 имеют очень низкое количество TE и генов, связанных с TE (рис. 4). Ди Приложение СИ, рис. S15) по сравнению с образцами из других групп. В частности, образцы подгруппы *in vitro* имеют большее количество TE. Большинство TE в пангеноме представляют собой ретротранспозоны, особенно Gypsy/DIRS1. Количество ретроэлементов и ДНК-транспозонов в подгруппе *in vitro* значительно выше, чем в группе клады 1+2 (Приложение СИ, рис. S16). С другой стороны, в группе клады 1+2 больше LINE по сравнению с другими группами, хотя LINE представляют лишь небольшую часть TE во всех образцах. Малые РНК участвуют в регуляции генов, включая замалчивание TE, а также они более распространены в кладе 1+2 по сравнению с другими группами. Кроме того, *C. окадаж* (ОКА15), *C. коммерции* (СОМ) и *C. чаконсе* Образцы (М6), которые были размножены *in vitro*, имеют более высокое количество генов TE по сравнению с остальной частью южной группы клады 4.

Обсуждение

Пангеномы сельскохозяйственных культур повлияли на улучшение сортов благодаря углублению понимания важности структурных вариаций генома (36). Недавний тетраплоидный пангеном картофеля позволил пролить свет на одомашнивание картофеля и влияние диких видов посредством обширной интрогрессии к культурному картофелю (21), а недавний диплоидный пангеном картофеля, который включал 44 вида, выходящих на плато с точки зрения содержания генов примерно в 40 геномах (23). Текущее исследование расширяет пангеном до более крупных *Solanum* раздел *Петота*, в состав которого входят клубнеобразные

местные сорта и дикие виды всех уровней пloidности. Этот более крупный панпloidный пангеном насчитывает около 80 геномов и демонстрирует огромное генетическое разнообразие в разделе, которое включает виды с бесполом и половым размножением, самонесовместимостью, интрогрессией и межвидовой гибридизацией, осложненной пре- и постзиготическим репродуктивным барьером (24). *Solanum* раздел *Петота* Созданный пангеном включает общедоступные и новые геномы, выбранные из наших предыдущих исследований, которые охватывают широкий географический диапазон и таксономическое разнообразие (2, 3). Этот пангеном является полным по оценке BUSCO (99% и 100% для *Solanum* и *Виридиплантае*, соответственно) по сравнению с эталонными геномами одного картофеля, которые варьируются от 83,9 до 97,9%. Использование пангенома для полногеномного сравнительного анализа, обнаружения структурных вариаций и функционального анализа значительно облегчает понимание видообразования у животных. *Петота* предоставляет геномные ресурсы для улучшения урожая картофеля за счет интрогрессии полезных диких видов. Анализ PAV культивируемых и диких образцов показал, что одомашнивание связано с изменениями в фотосинтезе, защитной реакции и метаболических процессах. В биологических функциях, связанных с развитием корней и периодом покоя, также были отмечены различия.

Мы сообщаем о филогенетическом дереве на основе PAV и наблюдаемых группировках, основанных на общей потере/прибыли генов. Построенное филогенетическое дерево пангенома использует полногеномный PAV генов, включая TE, для определения генетического расстояния для *Solanum* раздел *Петота* и сравнили самую большую на сегодняшний день коллекцию ядерных геномов из этого раздела. Дубликация генов с помощью PAV включает процессы репарации ДНК, а также активацию TE, которые при мобилизации влияют не только на количество их собственных копий, но и на другие гены (37). PAV являются важным источником генетических вариаций, истории эволюции и межвидовых различий. Настоящее исследование показывает увеличение содержания генов в местных сортах и сортах по сравнению с дикими видами, что указывает на тенденцию увеличения количества генов во время одомашнивания картофеля, в отличие от других видов растений (27, 33). Более того, увеличение содержания генов среди полиплоидов показывает разнообразие, существующее внутри секции *Петота*.

Процессы, приводящие к PAV, также могут влиять на SNP и индели. Следовательно, существовали общие черты этой филогенетической структуры PAV с другими, которые произошли от SNP и инделей (14, 15, 38), и это поддерживает кладистические группировки Spooner (2). Включение различных уровней пloidности в анализ на основе PAV показало, что диплоидные местные сорта (*C. пуреджа*, *C. стенотомум*, и *C. гониокаликс*) имеют аналогичные генные PAV, что и сосредоточенные *C. кандоллеанум* вид (3), тогда как *C. Tuberosum* сорта (в том числе «Андигена» и «Хилотанум» группы) виды имеют разнообразную генетическую основу и имеют сходство с южными и северными дикими видами. Тетраплоид *C. Tuberosum* сорта показали интрогрессию от большого числа диких видов (14, 21). Еще одним дополнительным преимуществом информации PAV была возможность функционального анализа группировок, обнаруженных в филогении, что позволило выявить данные о видообразовании в сложном комплексе. *Solanum* раздел *Петота*.

Клады, обнаруженные здесь и в предыдущих исследованиях, имеют сильную географическую идентичность: виды клады 1+2 встречаются преимущественно в Северной и Центральной Америке, клада 3 — в Эквадоре и северном Перу, а клада 4 — преимущественно в Южной Америке (2). Было обнаружено, что гены, уникальные для каждой клады, обладают функциями, связанными с реакциями на абиотический стресс, особенно на холод, что позволяет предположить участие PAV в адаптации к окружающей среде и вклад в разделение видов в кладах. Кроме того, гены, участвующие в биосинтезе тритерпеноидов, которые производят защитные соединения, включая предшественники стероидных гликоалкалоидов, также оказались уникальными для клады 3 и 4. Эти результаты согласуются с вариациями в стероидных гликоалкалоидах, наблюдаемыми у *Solanum* раздел

Петота в других исследованиях (39). Клада 4 также имела уникальные гены, отвечающие за функции развития цветков. Регуляторы клубнеобразования картофеля являются гомологами регуляторов цветения и также несут признаки развития цветка (40, 41). Результаты показывают, что PAV генов, участвующих в цветении/клубнеобразовании, способствуют группировке в кладе 4. Наши результаты подтверждаются другими исследованиями, показывающими отбор генов, участвующих в производстве гликоалкалоидов и клубнеобразовании при доместикации (14, 15, 21).

Основной геном *Петота* Пангеном всех клад демонстрирует фундаментальные биологические, молекулярные и клеточные функции. Геном раковины особенно богат генами биосинтеза вторичных метаболитов, участвующих в производстве стероидных гликоалкалоидов, которые являются характерными защитными метаболитами этого рода. *Соланум*. Облачные гены демонстрируют функции в ответ на стресс и ТЕ-активность. Хупс и др. (21) также отметили ТЕ для генов облака и большого числа копий тетраплоидного пангенома картофеля (21). Высокое содержание ТЕ в *Петота* Пангеном указывает на то, что ТЕ являются важным фактором, который следует учитывать в эволюции генома секции и, вероятно, всего рода. Известно, что ТЕ влияют на структуру и функцию генома, приводя к нарушениям генов, геномным перестройкам, таким как транслокации, делеции, дубликации генетических элементов, а также к изменениям в экспрессии генов, таким как инактивация генов, дисфункция генов и дифференциальная экспрессия (42, 43). Эти элементы обычно подавляются сложным регуляторным механизмом защитных механизмов генома (43, 44). Например, метилирование ДНК может подавлять активность ТЕ, чтобы избежать потенциальных негативных последствий для нормального развития растения (45). Мутации, вызванные ТЕ, могут быть устранены в популяции или распространиться по ней в зависимости от эффективного размера популяции (43). Оценка скорости транспозиций в геноме является сложной задачей; однако считается, что они на несколько порядков выше, чем мутации нуклеотидных оснований, что делает ТЕ важным игроком в эволюционных изменениях (42). Растения сталкиваются с многочисленными биотическими и абиотическими стрессами, и ТЕ могут играть роль в запуске генетической пластичности, изменяя экспрессию генов (43) и вызывая всплески активности ТЕ (44).

ТЕ, влияющие на картофель, ранее были продемонстрированы в литературе. Удаление ТЕ из гена, кодирующего флавоноид 3',5'-гидроксилаза, зарегистрировано, что гидроксилаза приводит к изменению цвета кожицы клубней. Это было вызвано безбилетным миниатюрным мобильным элементом с инвертированными повторениями (MITE), который был удален из генома картофеля в условиях культуры ткани, производя соматические варианты, что привело к реверсии флавоноида 3',5'-гидроксилазная функция и клубень с пурпурной кожицей (46). Гены картофеля, контролирующие самонесовместимость, *Сли47*, (48) и фотопериодзависимая клубнеобразование, *СтКД1*(41), функционально нарушаются ТЕ, что приводит к появлению новых черт: первые обеспечивают способность к самоопылению, а вторые обеспечивают естественное расширение секции. *Петота* дальше на юг, в Южной Америке, в регионы с длинной продолжительностью светового дня и распространением картофеля для выращивания в северных зонах с умеренным климатом. Настоящее исследование предоставляет дополнительные доказательства роли ТЕ в видообразовании в разделе *Петота*. ТЕ-мобилизация связана с ограничением потока генов между возникающими видами (49). В разделе *Петота*, который является одним из наиболее быстро диверсифицирующихся или излучающих разделов внутри *Соланум*(50), и там, где многие молодые виды возникли в результате прямой гибридизации двух видов (2), мобилизация ТЕ, возможно, способствовала разнообразию и большому количеству наблюдаемых более молодых видов.

Виды клады 1+2 имели меньшее количество ТЕ и различия в классах ТЕ по сравнению с другими кладами (*Приложение СИ, рис. S17*). Основываясь на таксономии, Хоукс (3) предположил, что дикие виды в *Петота* зародились в Северной и Центральной Америке. Далее было высказано предположение, что экспансия продолжалась на юг, в сторону Южной Америки, при этом виды распространялись как дальше на юг, так и обратно на север. *Петота* пангеном предполагает, что активация ТЕ

возможно, внесли свой вклад в геномное и таксономическое разнообразие, наблюдаемое у *Соланум* раздел *Петота* вида, которые адаптировались к новой среде обитания, расширяясь на юг из Северной и Центральной Америки, хотя этому разнообразию возможны и другие объяснения, такие как генетический дрейф.

Виды клады 1+2 также имеют гибридационную несовместимость с другими кладами. *Петота*(51–53). Барьеры гибридизации в *Соланум* может быть объяснено с помощью концепции баланса числа эндосперма (EBN) (54) с отношением 1EBN к видам клады 1+2. Они несовместимы с видами клады 3 и 4, то есть 2EBN и 4EBN. Метилирование ДНК играет важную роль в барьерах гибридизации видов у *Арабидопсис Талиана*(55–61) и дикие *Соланум* родственники томата (59), где он контролирует молчание однородительских генов в эндосперме при геномном импринтинге. Потеря метилирования ДНК и геномного импринтинга нарушает регуляцию однородительского молчания генов, вызывая неподходящие дозы генов, ведущие к недоразвитию эндосперма и аборт семян. Результаты текущего исследования показывают низкие уровни ТЕ для видов клады 1+2, все из которых имеют 1EBN, по сравнению с кладами 3 и 4, которые имеют виды 2EBN и 4EBN. Учитывая, что метилирование ДНК участвует в молчании ТЕ (49), интересно рассмотреть возможную роль ТЕ, метилирования ДНК и геномного импринтинга в разделе *Петота* репродуктивные барьеры ЭБН; однако для выяснения связи потребуются дальнейшие исследования.

Из-за высокого уровня гетерозиготности и способности к аутокроссированию, а также из-за необходимости обеспечения постоянства и единообразия качественных признаков картофель обычно размножают вегетативным путем (62). Много *Соланум* раздел *Петота* вида также легко адаптируются к размножению *in vitro* посредством сохранения в генных банках (63), поскольку это один из наиболее эффективных методов сохранения картофеля, позволяющий гарантировать фиксирование аллельных состояний и защитить его от вирусов; тем не менее, их также хранят в виде семян, чтобы помочь сохранить разнообразие, содержащееся внутри вида. Сохранение как настоящих семян, в дополнение к тканевой культуре, является предметом рассмотрения для селекционных материалов и генных банков, учитывая вариации ТЕ, наблюдаемые в текущем исследовании. На основании истории исходного материала, включающей сохранение и размножение культур тканей *in vitro*, образцы были подгруппированы в подгруппу образцов *in vitro*, которая состояла из культивированных образцов *С. Tuberosum* сорта и селекционную зародышевую плазму, а также местные сорта и образцы диких видов картофеля *in vitro*. В этой подгруппе *in vitro* обнаружено наибольшее количество генов, связанных с ТЕ, по сравнению с остальными образцами пангенома. Поскольку общим знаменателем в этой группе является исходный материал из культуры *in vitro*, а не родство, результаты показывают, что культура *in vitro* запускает активацию ТЕ, что также было обнаружено другими (64). Настоящее исследование доказывает, что эти изменения происходят в *Соланум* раздел *Петота* предоставляет дополнительные доказательства активации ТЕ при адаптации к новой среде. Таким образом, транспозиция является источником генетических вариаций, которые можно использовать для селекции. *Петота* Суперпангеном обеспечивает основу для создания инструментов для дальнейшего исследования и применения ТЕ для климатически оптимизированной селекции картофеля.

Выводы

Будущее выживание картофеля перед лицом изменения климата зависит от сохранения биоразнообразия и нашего понимания того, как внести разнообразие в существующие и развивающиеся сорта. Здесь мы представляем комплексный пангеном, построенный из 296 образцов, представляющих 60 видов в *Соланум* раздел *Петота*. Анализ пангенома предоставил доказательства роли PAV, особенно вариации ТЕ, в видообразовании у *Соланум* раздел *Петота* обнаружили тенденцию к увеличению содержания ТЕ в материале, размножаемом *in vitro*, что сродни увеличению активности ТЕ при естественной адаптации из-за стресса.

Материалы и методы

Геномные последовательности образцов картофеля. Всего 296 образцов картофеля.

были использованы при строительстве *Соланум*раздел *Летотанген* генома. Они представляют 60 видов, основанных на первоисточнике: 63 образца взяты из разнообразной группы, включающей дикие виды, местные сорта и культурные сорта (14); 199 образцов включают преимущественно дикие *Соланум*виды и сорта картофеля широкого географического происхождения (15); 13 образцов из Международного центра картофеля (CIP), Лима, Перу (17, 18, 65, 66); десять образцов от Министерства сельского хозяйства и агропродовольственной промышленности Канады (AAFC) (67, 68); шесть тетраплоидных сортов картофеля из Европы и Северной Америки (21); и пять опубликованных эталонных последовательностей генома (13, 16, 19, 20, 69). Подробный список этих образцов представлен в [Приложение СИ, Таблица S2](#). Таксономия соответствует соответствующему первоисточнику.

Поскольку образец обычно определяется как группа родственного растительного материала, некоторая генетическая изменчивость может существовать внутри образца семян, состоящего из семян от скрещивания, создавшего новую комбинацию аллелей. У картофеля образцы семян, хранящиеся в генных банках, обычно представляют собой гетерогенный материал, собранный из его естественной среды или регенерированный контролируемым образом. В материале, размноженном клоном, образец представляет собой отдельную особь или генотип, в котором комбинации аллелей зафиксированы посредством вегетативного размножения. Геномы в этом исследовании включали дикий семенной материал, в котором для анализа было выбрано одно семя образца, образцы, размноженные *in vitro*, в которых образец представлял собой одно растение (клон), а также образцы, размноженные клубнями, которые включают сорта и селекционные линии с предыдущая история размножения *in vitro* для поддержания клонов и/или сертифицированного производства семян. Во всех случаях образец представляет собой один генотип.

Строительство Пангенома. Эталонный геном диплоидного картофеля с разрешением гаплотипов, *C. Tuberosum* группа *Tuberosum* RH89-039-16 v3 сборки (19), использовалась в качестве эталона при построении *Летотанген* генома. В различных предыдущих исследованиях сборки генома de novo были созданы для 22 образцов, которые использовались непосредственно при построении пангенома, а остальные образцы собирались de novo с использованием необработанных считываний секвенирования. 11 образцов романа *C. Букасовий* (БУК2), *C. окая* (ОКА15) и девять диплоидных клонов из коллекции зародышевой плазмы AAFC (H412-1, W5281.2, 10908-06, DW.84-1457, 12625-02, 12120-03, 08675-21, 07506-01 и 11379). -03) были секвенированы и собраны, как описано (66, 68). Необработанные данные секвенирования остальных 262 образцов были загружены с портала SRA Национального центра биотехнологической информации (NCBI) и обработаны с использованием *Тримматика v0.39*

(70) и собраны заново с использованием *Мегахит v1.2.971*. Другой значения использовались для наборов данных с разной длиной чтения, '-к-мин 21 --к-макс 85 --к-шаг 8' была установлена для длины чтения 100 п.н., '-к-мин 25 --к-макс 95 --к-шаг 10' был установлен на длину чтения 125 п.н., '-к-мин 31 --к-макс 111 --к-шаг 10' был установлен на длину чтения 150 п.н., и '-к-мин 35 --к-макс 125 --к-шаг 10' был установлен на длину чтения 350 п.н. Полученные контиги длиной <500 п.н. были отфильтрованы из сборок.

Чтобы удалить из сборок последовательности органеллярных геномов и сохранить только ядерные последовательности, сборку генома RH объединили с общедоступными органеллярными геномами картофеля (67, 72–74). Каждую из загруженных и вновь собранных последовательностей сравнивали с RHO (RH плюс органеллы) с помощью *нумералайнер изржженный v4.0.Obeta2* с '-maxmatch' возможность получить максимальное количество совпадений (75) ([Приложение СИ, рис. S18](#)). Совпадения были отфильтрованы с длиной выравнивания минимум 300 п.н. и идентичностью 90% с использованием *дельта-фильтр*. Последовательности, которые не совпадали с RHO, сохранялись как невыровненные последовательности. К невыровненным последовательностям добавляли любые выровненные последовательности, содержащие непрерывную невыровненную область длиной не менее 500 п.н.

Невыровненные контиги были отфильтрованы на наличие примесей путем запроса базы данных избыточных нуклеотидов NCBI (загруженной 21 июля) с минимальной идентичностью 90% с использованием *взрыв+ v2.11.076*. Контиги, попавшие только на зеленые растения, назывались надежными и сохраняемыми, тогда как контиги, попавшие только на зеленые растения, назывались загрязненными и удаленными. Контиги с совпадениями как с зелеными растениями, так и с другими сохранялись, когда по крайней мере 90% охватываемой запросом области соответствовали зеленым растениям. Кроме того, невыровненные контиги также проверялись на наличие адаптерных и оставшихся органеллярных последовательностей путем запроса на соответствие *Юнивеб* базы данных и органеллярные последовательности из полных *Соланум* таксоны. После струйной фильтрации *нумерис* использовался итеративно для сравнения отдельных образцов и удаления дополнительной избыточности в невыровненных контигах для получения единого файла невыровненных последовательностей. Чистые невыровненные последовательности были

далее проверяется на избыточность путем выполнения выравнивания «все против всех» с использованием *взрыв+* с последующим *сд-хит v4.8.1(77)*. Для вышеуказанного использовались порог идентичности последовательности 95% и охват запроса 95%. *нумер, взрыв+, и CД-хит*шаги. Любые органеллярные последовательности, присутствующие в сборке RH, были удалены, и только ядерные последовательности были объединены с невыровненными избыточными последовательностями для создания *Летотанген* генома. Качество и полноту пангенома оценивали с помощью *КВАСТ v5.0.2(78)* и *БУСКО v5.2.2(79)*.

Аннотация Пангенома. Для пангенома была создана пользовательская библиотека повторов (CRL) с использованием *повторитель модели v2.0.2(80)*, а неизвестные повторы были аннотированы с использованием базы данных транскриптов. Также фрагменты генов из неизвестных повторов были удалены с помощью *Протеклюдер(81)* для создания окончательного списка отзыва сертификатов. Повторы в невыровненных последовательностях пангенома были мягко замаскированы с использованием *ПовторМаскер v4.1.1(82)*. Окончательный CRL пангенома и DMv6.1 (69) использовались в качестве входных библиотек повторов для RepeatMasker. Невыровненные последовательности с мягкой маской были аннотированы с помощью конвейера BRAKER (83, 84) с использованием эталонных белковых последовательностей и данных последовательности рибонуклеиновой кислоты (RNA-seq). Для первого запуска белковые последовательности *Соланалес* из OrthoDB были объединены с белковыми последовательностями DMv6.1, RH и пантранскриптома картофеля (31). Предсказание генов ab initio было выполнено *АВГУСТ(85)* и *ГенеМарк-ЭП(86)* с белковыми последовательностями в качестве внешнего доказательства. Для второго запуска данные РНК-секвенирования из 102 библиотек были загружены из NCBI ([Приложение СИ, Таблица S3](#)), обрезано с помощью *TrimGalore v0.6.7(87)* и совмещены с невыровненным пангеномом с помощью *hisat2 v2.2.1(88)*. Полученные файлы карты двоичного выравнивания (BAM) были объединены с помощью *Самтулс v1.1.3(89)*. Предсказание генов было выполнено *АВГУСТ/ГенеМарк-ЭП(90)* с объединенным файлом BAM в качестве доказательства секвенирования РНК.

Результаты обоих запусков BRAKER были объединены с помощью *ЦЕБПав1.0.3(91)*. Полученные аннотации были отфильтрованы с использованием *gFAC v1.1.2(92)* со следующими параметрами '-rem-all-incompletes --rem-genes-without-start-codon --rem-genes-without-stop-codon --разрешить-альтернативные-старты --min-exon-size20 --min-intron-size20 --min-CDS-size74 --statistics --create-gff3 --allowed-inframe-stop-codons 0'. Модели генов с >50% повторов в кодирующей области удаляли с помощью *gffcompare v0.12.0(93)*. Ошибки формата в файле gff3 были исправлены с помощью *Набор инструментов GFF3 v2.1.0(94)*, чтобы создать окончательный набор невыровненных аннотаций пангенома. Гены также были функционально аннотированы *ахрд v3.3.3* (доступны на <https://github.com/groupschoof/AHRD>), чтобы получить удобочитаемые описания аннотаций путем запроса к базам данных TrEMBL, UniProt и Araprot11 (загружено 23 декабря 2021 г.). *Лфамдомены* и термины GO были идентифицированы в этих генах путем поиска в базе данных InterPro с использованием *интерпроскан v5.52-86.0(95)*. Наконец, аннотация невыровненного пангенома была объединена с аннотацией генома RH как полная аннотация пангенома.

ПАВ-анализ. Отфильтрованные данные короткого считывания из 285 образцов были использованы для анализа PAV. Одиннадцать образцов были исключены из этого анализа либо из-за короткой длины чтения (<100 бит/с), либо из-за гибридного происхождения. Считывания были картированы индивидуально на пангеноме с использованием *бва v0.7.17(96)*. Только сопоставленные и правильно спаренные выравнивания (прямые и обратные чтения, сопоставленные на ожидаемом расстоянии и ориентации) были сохранены для дальнейшего анализа с использованием *Самтулс* флагом '-f2.SGSGeneLoss' использовался для вызова PAV из трасс (97). Гены считались отсутствующими, если менее 20% областей экзонов были покрыты как минимум *n* чтениями, где *n* определялось на основе средней глубины пангенома (*n* = 2, 4, 6, 8 или 10, когда глубина = 0). до 10, от 11 до 20, от 21 до 40, от 41 до 80 или 80+ соответственно). Эти настройки были выбраны для уменьшения потенциальной систематической ошибки при вызове PAV после тестирования различных значений *n* для различной глубины секвенирования. Все гены пангенома были разделены на три категории на основе результатов PAV. Если ген присутствует в 97% образцов, его называют основным геном, аналогично гены оболочки и облака были классифицированы, если они присутствуют в от 3 до 97% и <3% образцов соответственно. Данные PAV были преобразованы в двоичные данные и использованы при построении филогенетического дерева с использованием *IQ-TREE2 v2.1.3(98)*. Наиболее подходящая модель замещения была определена с помощью *IQ-ДЕФЕБО2*, и филогенетическое дерево максимального правдоподобия было построено с использованием модели замещения GTR2 + FO + R10 и 1000 бутстрап-репликаций. PCA PAV проводился с использованием *R v4.1.0* *прком* команда с параметрами по умолчанию. Анализ обогащения ГО был проведен для определения обогащенных термов ГО и их биологического процесса с использованием *топГО v3.14(99)*.

Анализ различий в частоте генов. Для идентификации генов, подлежащих селекции, были проведены следующие сравнения: дикая клада 4 с местными сортами и дикая клада 4 с культурными сортами для анализа различий в частоте присутствия генов. Таблица непредвиденных обстоятельств 2x2 была построена для каждого гена в каждой группе на основе их статуса PAV. Фишера

Точный тест использовался для расчета значимости различий в частотах присутствия генов между сравниваемыми группами. *p*-значения, полученные с помощью теста Фишера, были скорректированы с использованием коэффициента ложного обнаружения (FDR). Скорректированный Рузвельтом *p*-значение <0,001 при кратном изменении частоты встречаемости генов >2 использовалось в качестве критерия значительного обогащения генов между сравниваемыми группами (27, 33). Гены с более высокой частотой встречаемости у местных сортов или сортов, чем у диких видов, считались «благоприятными генами», а гены с более низкой частотой, чем у диких видов, считались «неблагоприятными генами». Более того, благоприятные гены с более высокой частотой встречаемости у местных сортов, чем у сортов, были названы генами, благоприятными для местных сортов, а более высокие частоты у сортов, чем у местных сортов, были названы генами, благоприятными для местных сортов.

Доступность данных, материалов и программного обеспечения. Необработанные геномные чтения и де сборки пово, используемые в этом исследовании, общедоступны, и каждый из инвентарных номеров указан в соответствующих таблицах (*Приложение СИ, Таблицы S2 и S3*).

1. Продовольственная и сельскохозяйственная организация Объединенных Наций, «FAOSTAT: Статистические базы данных ФАО» (Зерноводство и продукция животноводства, Лима, Перу, 2018 г.). <https://www.fao.org/faostat/en/#data/QCL>. По состоянию на 28 мая 2022 г.
2. Спунер Д.М., Гислен М., Саймон Р., Янский С.Х., Гавриленко Т. Систематика, разнообразие, генетика и эволюция дикого и культурного картофеля. *Бот. Преподобный* 80, 283–383 (2014).
3. Дж. Дж. Хоукс. *Картофель: эволюция, биоразнообразие и генетические ресурсы* (Белхейвен Пресс, 1990).
4. Р. Дж. Хиджманс, Д. М. Спунер, Географическое распространение диких видов картофеля. *Являясь*, Дж. Бот. 88, 2101–2112 (2001).
5. Д.М. Спунер, К. Маклин, Г. Рамзи, Р. Во, Г. Дж. Брайан, Единичное приручение картофеля на основе генотипирования полиморфизма длины мультикопированного амплифицированного фрагмента. *Учеб. Натл. акад. наук. США* 102, 14694–14699 (2005).
6. Овчинникова А. и другие., Систематика культурного картофеля (*Solanum* секция *Petota*: *Solanaceae*). *Бот. Дж. Линн. Соц.* 165, 107–155 (2011).
7. Гавриленко Т. и другие., Генетическое разнообразие и происхождение культурного картофеля на основе пластидного микросателлитного полиморфизма. *Жене. Рез. Кроп Эвол.* 60, 1997–2015 (2013).
8. Международный Центр Картофеля, «Родной картофель», Лима (Перу) СР 2 стр. (2017).
9. Ватанабе К. Генетика, геномика и приложения картофеля. *Порода. наук.* 65, 53–68 (2015).
10. Ю. Ван де Пер, Т.-Л. Эшман, П.С. Солтис, Д.Е. Солтис, Полиплоидия: эволюционная и экологическая сила в стрессовые времена. *Растительная клетка* 33, 2899 (2021).
11. К. Ватанабе, С. Дж. Пелокин, Встречаемость пыльцы 2n и частоты генов *rs* в культурных группах и родственных им диких видах в клубненосных пасленах. *Теор. Прил. Жене.* 78, 329–336 (1989).
12. Консорциум по секвенированию генома картофеля. Секвенирование генома и анализ клубнеплодов картофеля. *Природа* 475, 189–195 (2011).
13. Р. Аврсонан и другие. Последовательность генома *Solanum commersonii* дает представление об адаптации к стрессовым условиям и эволюции генома родственников дикого картофеля. *Растительная клетка* 27, 954–968 (2015).
14. М. А. Хардиган и другие. Разнообразие генома клубненосного *Solanum* раскрывает сложную эволюционную историю и цели одомашнивания культивируемого картофеля. *Учеб. Натл. акад. наук. США* 114, E9999–E10008 (2017).
15. Ю. Ли и другие. Геномный анализ дает маркеры для идентификации агрономически важных генов картофеля. *Мол. Растение* 11, 473–484 (2018).
16. К.П. Лейснери и другие. Последовательность генома М6, диплоидного инбредного клона клубненосного вида картофеля *Solanum chacoense* с высоким содержанием гликоалкалоидов, демонстрирует остаточную гетерозиготность. *Плант Дж.* 94, 562–570 (2018).
17. М. Кириакиду и другие., Структурный геномный анализ таксонов культивируемого картофеля. *Теор. Прил. Жене.* 133, 951–966 (2020).
18. М. Кириакиду, Н. Л. Энглин, Д. Эллис, Х. Х. Тай, М. В. Стрёмвик, Сборка геномов шести полиплоидных геномов картофеля. *наук. Данные* 7, 88 (2020).
19. Ц. Чжао и другие., Анализ генома гетерозиготного диплоидного картофеля с разрешением по гаплотипам. *Нат. Жене.* 52, 1018–1023 (2020).
20. Н. ван Лисхаут и другие., *Solyntus*, новый высоконепрерывный эталонный геном картофеля (*Solanum tuberosum*). *G3 (Бетесда)* 10, 3489–3495 (2020).
21. Г. Хупси и другие. Потатные сборки генома тетраплоидного картофеля в масштабе хромосом обнаруживают сложный геном, транскриптом и прогнозируемый протеомный ландшафт, лежащий в основе генетического разнообразия. *Мол. Растение* 15, 520–536 (2022).
22. Х. Суни и другие. Сборка генома тетраплоидного сорта картофеля в масштабе хромосом и с разрешением гаплотипов. *Нат. Жене.* 54, 342–348 (2022).
23. Д. Тани и другие., Эволюция генома и разнообразие дикого и культурного картофеля. *Природа* 606, 535–541 (2022).
24. Э. Л. Камадро, Л. Е. Эраццу, Ж. Ф. Мауне, М. К. Бедогни, Генетический подход к проблеме видов дикого картофеля. *Растительная биол.* 14, 543–554 (2012).
25. М. Кириакиду, Х. Х. Тай, Н. Л. Англин, Д. Эллис, М. В. Стрёмвик, Современные стратегии сборки последовательности генома полиплоидных растений. *Передний. Наука о растениях* 9, 1660 (2018).
26. Ц. Чжао и другие. Пангеномный анализ подтверждает степень геномной изменчивости культивируемого и дикого риса. *Нат. Жене.* 50, 278–284 (2018).
27. Л. Гао и другие., Пангеном томата обнаруживает новые гены и редкую аллель, регулирующую вкус фруктов. *Нат. Жене.* 51, 1044–1051 (2019).
28. С. П. Гордони и другие. Обширные вариации содержания генов в пангеноме *Brachypodium distachyon* коррелируют со структурой популяции. *Нат. Коммун.* 8, 2184 (2017).
29. Ж.-Х. Ли и другие. Высококачественные геномы в масштабе хромосом способствуют эффективной идентификации крупных структурных различий в остром и сладком перце. *Хортик. Рес.* 9, ihaa210 (2022 г.), 10.1093/hortres/iaa210.
30. Глик Л., Мэйроу И. Пангеномы растений весьма уязвимы к методологическим соображениям. *Исследовательская площадка* [Препринты] (2022). <https://doi.org/10.21203/rs.3.rs-1415605/v1> (по состоянию на 10 марта 2022 г.).
31. М. Петеки и другие., Реконструкция транскриптома, специфичного для сорта, и пан-транскриптома тетраплоидного картофеля. *Научные данные* 11, 1–15 (2020).
32. Ю. Лю и другие., Пангеном дикой и культурной сои. *Клетка* 182, 162–176.e13 (2020).

Все данные исследования включены в статью и/или *Приложение СИ*. Вся сборка пангенома доступна через DRYAD (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.cfxpvnxbn>) (100).

благодарности. Мы хотели бы выразить признательность командам по селекции картофеля и генным банкам в AAFС и СР, Перу, за поддержание и распространение клонов, использованных в исследовании. Финансирование исследования было предоставлено Инициативой исследований и разработок в области геномики AAFС, совместной наградой AAFС-Génome Québec в области геномики, управляемой управлением, «Революционное создание сортов картофеля для климатически умного картофеля» для ННТ и MVS (J-002367/GQ-AAC-2019-2); награда Compute Canada RPP (Исследовательские порталы и платформы: «Портал разнообразия генома картофеля») и награда Compute Canada Resources for Research Groups («Анализ структурных вариаций сложных геномов растений в поисках климатической адаптации») для MVS

33. Дж. Ли и другие. Пангеном хлопка восстанавливает утраченные последовательности и гены во время одомашнивания и селекции. *Геном Биол.* 22, 119 (2021).
34. Б. Хуанг, Х. Рюсс, К. Лян, К. Коллеони, Д.М. Спунер, Анализ 202 пластидных геномов проясняет филогению *Solanum* секция *Petota*. *наук. Представитель* 9, 4454 (2019).
35. Проект PBI *Solanum*, источник пасленовых. (2022). <http://www.solanaceae.org/>. По состоянию на 1 ноября 2022 г.
36. Р. Дела Колетта, Ю. Цю, С. Оу, М.Б. Хаффорд, К.Н. Хирш, Как пангеном меняет геномику и улучшение сельскохозяйственных культур. *Геном Биол.* 22, 3 (2021).
37. Кривалева К.В. Роль мобильных элементов и механизмов репарации поврежденной ДНК в дупликациях и слияниях генов в геномах растений. *Курс. Мнение. Растительная биол.* 48, 18–25 (2019).
38. М. А. Хардиган, Дж. Бамберг, CR Buell, DS Douches, Таксономия и генетическая дифференциация среди дикой и культивируемой зародышевой плазмы *solanum* sect. *Petota*. *Геном растения* 8, eplantgenome2014.06.0025 (2015).
39. X. Х. Тай, К. Уорролл, Ю. Пелетье, Д. Де Койер, Л. А. Кэлхун, Сравнительное профилирование метаболитов *Solanum tuberosum* по сравнению с шестью дикими видами *Solanum*, устойчивыми к колорадскому жуку. *Дж. Агрик. Пищевая хим.* 62, 9043–9055 (2014).
40. К. Наварро и другие., Контроль цветения и образования запасовых органов картофеля с помощью FLOWERING LOCUS T. *Природа* 478, 119–122 (2011).
41. Б. Клоостерман и другие., Встречающиеся в природе разнообразие аллелей позволяет выращивать картофель в северных широтах. *Природа* 495, 246–250 (2013).
42. К. Витте, М.-А. Фустье, К. Аликс, М.И. Тейнаон, Светлая сторона транспозонов в эволюции сельскохозяйственных культур. *Функция «Брифинг». Геномика* 13, 276–295 (2014).
43. М. Сахеби и другие., Вклад мобильных элементов в геном растения. *Ген* 665, 155–166 (2018).
44. М. Дж. Дубин, О. М. Шейд, К. Беккер, Транспозоны: благословение и проклятие. *Курс. Мнение. Растительная биол.* 42, 23–29 (2018).
45. К. Бротигам, К. Кронк, Метилирование ДНК и эволюция сложности развития растений. *Передний. Наука о растениях* 9, 1447 (2018).
46. Момосе М., Абз Ю., Оззи Ю. Миниатюрные перевернуто-повторяющиеся мобильные элементы безбилетника активны в картофеле. *Генетика* 186, 59–66 (2010).
47. Э.-Ж. Эггерс и другие. Нефункционализация гена *Sli* приводит к самосовместимости и облегчает точную селекцию картофеля. *Нат. Коммун.* 12, 4141 (2021).
48. Л. Ман и другие. Ген F-box, не являющийся S-полюсом, устраняет самосовместимость у диплоидного картофеля. *Нат. Коммун.* 12, 4142 (2021).
49. А. Серрато-Капучина, Д. Р. Матуге, Роль мобильных элементов в видообразовании. *Ген* 9, 254 (2018).
50. Эчевериа-Лондосо С., Саркинен Т., Фенон И.С., Первис А., Кнапп С. Динамизм и контекстная зависимость в диверсификации мегаэволюционного рода растений *Solanum* (*Solanaceae*). *Дж. Сист. Эвол.* 58, 767–782 (2020).
51. С. И. Лара-Кабрера, Д. М. Спунер, Таксономия мексиканского диплоидного дикого картофеля (*Solanum* sect. *Petota*): Морфологические и микросателлитные данные. *Монография Сист. Бот.* 104, 199 (2005).
52. Р. Ортис, П. Саймон, С. Янский, Д. Стелли, Манипулирование плоидностью гаметофита, эндосперма и спорофита в природе и для улучшения сельскохозяйственных культур: дань уважения профессору Стэни Дж. Пелокину (1921–2008). *Анна. Бот.* 104, 795–807 (2009).
53. ДМ Спунери и другие., Дикий картофель (*Solanum* секция *Petota*; *Solanaceae*) Северной и Центральной Америки. *Сист. Бот. Монорг.* 68, 1–209 (2004).
54. С.А. Джонстон, Р.Э. Ханнеман младший, Манипуляции с балансовым числом эндосперма преодолевают барьеры пересечения между диплоидными видами *Solanum*. *Наука* 217, 446–448 (1982).
55. Р.А. Батиста, К. Келер, Геномный импринтинг в растениях – пересмотр существующих моделей. *Генс Дев.* 34, 24–36 (2020).
56. Хендерсон И.Р., Якобсен Е.С. Эпигенетическая наследственность у растений. *Природа* 447, 418–424 (2007).
57. Искива Р., Киносита Т. Эпигенетическое программирование: проблема гибридизации видов. *Мол. Растение* 2, 589–599 (2009).
58. К. Йозефссон, Б. Дилкс, Л. Комай, Зависимая от родителей потеря молчания генов во время межвидовой гибридизации. *Курс. Биол.* 16, 1322–1328 (2006).
59. М. Рот, А. М. Флорес-Руада, М. Пэрис, Т. Стедлер, Транскриптомы эндосперма диких томатов обнаруживают общие роли геномного импринтинга как в ядре, так и в клеточном эндосперме. *Плант Дж.* 95, 1084–1101 (2018).
60. Блевинс Т., Ван Дж., Флигер Д., Понтвиан Ф., Пикард К.С. Гибридная несовместимость, вызванная эпиплюидностью. *Учеб. Натл. акад. наук. США* 114, 3702–3707 (2017).
61. Н. Вайд, Р.Э. Лайтинен, Разнообразные пути к гибридной несовместимости арабидопсиса. *Плант Дж.* 97, 199–213 (2019).
62. Дж. М. Бредин, К. Г. Хейнс, «Введение в картофель» в *Генетика, геномика и селекция картофеля*, Дж. М. Бредин, К. Коле, ред. (Тейлор и Фрэнсис Бока-Ратон, 2011), стр. 1–19.
63. Д. Эллис, А. Салас, О. Чавес, Р. Гомес, Н. Энглин, «Сохранение ex situ генетических ресурсов картофеля [*Solanum* секция *Petota* (*Solanaceae*)] в генных банках» в *Урожай картофеля*, Х. Кампос, О. Ортис, ред. (Спрингер, Чам, 2020 г.), стр. 109–138.

64. П.Т. Беднарек, Р. Орловска. Среда культуры тканей растений как переключатель (эпи)генетических изменений. *Культ органов и тканей растений*.140,245–257 (2020).
65. М. Кириакиду, *Сборка генома и обнаружение структурной изменчивости таксонов культивируемого картофеля* (Университет Макгилла, 2020 г.).
66. И. Бозан, *Анализ генома диплоидного дикого картофеля Solanum tuberosum* (Университет Макгилла, 2021).
67. С. Р. Ачаккагари, Х. Х. Тай, К. Дэвидсон, Х. Де Йонг, М. В. Стрёмвик. Полные сборки митогенома 10 диплоидных клонов картофеля обнаруживают рекомбинационные и перекрывающиеся варианты. *ДНК Рез.* 28,dsab009 (2021).
68. С.Р. Ачаккагари и другие. Секвенирование генома адаптированных диплоидных клонов картофеля. *Передний. Наука о растениях*.13, 954933 (2022 г.).
69. ГМ Фаи и другие. Создание сборки длинночитываемого эталонного генома картофеля в масштабе хромосомы. *Гигасайенс*.9, гиа100 (2020).
70. AM Bolger, M. Lohse, B. Usadel, Trimmomatic: гибкий триммер для данных последовательности Illumina. *Биоинформатика*30,2114–2120 (2014).
71. Ли Д., К.-М. Лю, Р. Луо, К. Садакане, Т.-В. Лам, МEGAHIT: сверхбыстрое одноузловое решение для большой и сложной сборки метагеномики с помощью краткого графа де Брейна. *Биоинформатика*31,1674–1676 (2015).
72. С.Р. Ачаккагари и другие. Полные сборки пластома из панели, состоящей из 13 различных таксонов картофеля. *ПЛОС Один*15,e0240124 (2020).
73. SR Achakagari, HN Tai, C. Davidson, HD Jong, MV Strömviik, Полные пластоменные последовательности девяти диплоидных клонов картофеля. *Митохондриальная ДНК В Рез.*6,811–813 (2021).
74. С.Р. Ачаккагари и другие. Полные сборки митогенома из панели из 13 различных таксонов картофеля. *Митохондриальная ДНК В Рез.*6,894–897 (2021).
75. Ж. Марсеи и другие. MUMmer4: Быстрая и универсальная система выравнивания генома. *ПЛОС Компьютер. Биол.*14, e1005944 (2018).
76. К. Камачо и другие. BLAST+: Архитектура и приложения. *БМЦ Биоинф.*10,421 (2009).
77. В. Ли, А. Годзик. Cd-hit: быстрая программа для кластеризации и сравнения больших наборов белковых или нуклеотидных последовательностей. *Биоинформатика*22,1658–1659 (2006).
78. Гуревич А., Савельев В., Вяхи Н., Теслер Г., QUAST: Инструмент оценки качества сборок генома. *Биоинформатика*29,1072–1075 (2013).
79. Ф.А. Симау, Р.М. Уотерхаус, П. Иоаннидис, Е.В. Кривенцева, Е.М. Здобнов, BUSCO: Оценка сборки генома и полноты аннотаций с помощью однокопийных ортологов. *Биоинформатика*31, 3210–3212 (2015).
80. Дж. М. Финли и другие. RepeatModeler2 для автоматического геномного обнаружения семейств мобильных элементов. *Учеб. Натл. акад. наук. США*117,9451–9457 (2020).
81. М. С. Кэмпбелли и другие. MAKER-P: Набор инструментов для быстрого создания, управления и контроля качества аннотаций генома растений. *Физиол. растений*.164,513–524 (2014).
82. А. Ф. Смит, Р. Хабли, П. Грин, Repeat-Masker open-4.0 (2013-2015). (февраль 2022 г.). <https://www.repeatmasker.org/faq.html>.
83. Т. Бруна, К.Дж. Хофф, А. Ломсадзе, М. Станке, М. Бородовский, BRAKER2: Автоматическая аннотация генома эукариот с помощью GeneMark-EP+ и AUGUSTUS, поддерживаемая базой данных белков. *НАР Геном. Биоинф.*3, lqaa108 (2021 г.).
84. К.Дж. Хофф, А. Ломсадзе, М. Бородовский, М. Станке, Полногеномная аннотация с помощью BRAKER. *Методы Мол. Биол.*1962,65–95 (2019).
85. Станке М., Шеффман О., Моргенштерн Б., Ваак С. Предсказание генов у эукариот с помощью обобщенной скрытой марковской модели, использующей подсказки из внешних источников. *БМЦ Биоинф.*7,62 (2006).
86. Бруна Т., Ломсадзе А., Бородовский М., GeneMark-EP+: Предсказание генов эукариот с самообучением в пространстве генов и белков. *НАР Геном. Биоинф.*2, lqaa026 (2020).
87. Ф. Крюгер, Ф. Джеймс, П. Юэлс, Э. Афонян, Б. Шустер-Беклер, Феликс Крюгер/TrimGalore: v0.6.7 — DOI через Zenodo (2021), 10.5281/zenodo.5127899.
88. Д. Ким, Дж. М. Пагги, К. Парк, К. Беннетт, С. Л. Зальцберг, Выравнивание генома на основе графов и генотипирование с помощью HISAT2 и HISAT-генотипа. *Нат. Биотехнология*.37,907–915 (2019).
89. Х. Ли и другие. Формат выравнивания/карты последовательностей и SAMtools. *Биоинформатика*25,2078–2079 (2009).
90. Ломсадзе А., Бернс П.Д., Бородовский М. Интеграция картированных считываний RNA-Seq в автоматическое обучение алгоритма поиска генов эукариот. *Нуклеиновые кислоты Рез.*42,e119 (2014).
91. Л. Габриэль, К. Дж. Хофф, Т. Брюна, М. Бородовский, М. Станке, TSEBRA: Селектор транскрипта для BRAKER. *БМЦ Биоинф.*22,566 (2021).
92. М. Кабальеро, Дж. Вегжин, gFACS: Фильтрация, анализ и преобразование генов для унификации аннотаций генома в рамках систем выравнивания и прогнозирования генов. *Геномика Протеомика Биоинформатика* 17,305–310 (2019).
93. Пертеа Г., Пертеа М., Утилиты GFF: GffRead и GffCompare. *F1000Pec.*9,304 (2020).
94. М.-Дж.М. Чен, Х. Лин, Л.-М. Чан, К. П. Чайлдс, М. Ф. Поулчан, Набор инструментов GFF3: контроль качества и конвейер слияния для аннотации генома. *Методы Мол. Биол.*1858,75–87 (2019).
95. Э. Кевийони и другие. InterProScan: идентификатор белковых доменов. *Нуклеиновые кислоты Рез.*33,W116–W120 (2005 г.).
96. Х. Ли, Р. Дурбин, Быстрое и точное выравнивание короткого считывания с помощью преобразования Берроуза-Уиллера. *Биоинформатика*25,1754–1760 (2009).
97. А.А. Голич и другие. Потеря генов грибового возбудителя канолы Leptosphaeria maculans. *Функт. Интегр. Геном.*15,189–196 (2015).
98. БК Минь и другие. IQ-TREE 2: Новые модели и эффективные методы филогенетического вывода в эпоху генома. *Мол. Биол. Эвол.*37,1530–1534 (2020).
99. А. Алекса, Дж. Раненфюрер, topGO: Анализ обогащения онтологии генов (2022). Версия пакета R 2.46.0. <http://bioconductor.org/packages/release/bioc/html/topGO.html>.
100. С.Р. Ачаккагари и другие. Картофельный суперпангеном v1.0. триада. <https://datadryad.org/stash/dataset/doi:10.5061/dryad.cfxpvnxbn>. Депонировано 5 июля 2023 г.