

doi 10.18699/vjgb-25-93

Последовательности ДНК агробактериального происхождения в филогенетических исследованиях растений

Т.В. Матвеева , П.М. Журбенко , Г.В. Хафизова , А.Д. Шапошников , Р.Р. Жидкин , А.В. Родионов ¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия² Ботанический институт им. В.Л. Комарова Российской академии наук, Санкт-Петербург, Россия³ Хьюстонский университет, Хьюстон, США radishlet@gmail.com

Аннотация. Агробактериальная трансформация – наиболее распространенный метод получения трансгенных растений. Метод основан на способности определенных почвенных бактерий родов *Agrobacterium* и *Rhizobium* переносить и интегрировать в хромосому растения-реципиента фрагмент своей плазмиды. Этот фрагмент получил название Т-ДНК (transferred DNA – переносимая ДНК). В лабораторных условиях было показано, что из трансгенных клеток можно регенерировать целые растения. Вскоре стало ясно, что подобные процессы происходят и в природе, приводя к появлению природно-трансгенных растений или природных генно-инженерно-модифицированных организмов (ГМО). Таким образом, природно-трансгенными называются растения, у которых в геномах присутствуют гомологи генов Т-ДНК агробактерий (клеточные Т-ДНК, клТ-ДНК). Эти последовательности наследуются в ряду половых поколений и сохраняют свою функциональность. Кроме того, продемонстрирована возможность использования новоприобретенных растениями последовательностей в филогенетических исследованиях, поскольку клТ-ДНК являются четко определенными, высокоспецифичными и узнаваемыми фрагментами ДНК, не похожими на последовательности ДНК растений. Они не встречаются у нетрансформированных предков, и их интеграция в определенном хромосомном сайте маркирует монофилетическую группу видов. В представленном обзоре освещены вопросы разнообразия клТ-ДНК, возможности их применения как филогенетических маркеров, в том числе описаны основные методические подходы таких работ, на конкретных примерах рассмотрены возможности уточнения спорных моментов в филогении родов *Nicotiana*, *Camellia*, *Vaccinium* и *Arachis*. Важным моментом филогенетического анализа на основе клТ-ДНК является реконструкция отдельных аллелей. Это дает возможность отслеживать факты межвидовой гибридизации. Именно этот подход позволил продемонстрировать незавершенный процесс видообразования в пределах секции *Thea* рода *Camellia*, а также подтвердил использование межвидовой гибридизации в ходе селекции североамериканских голубик. В обзоре рассмотрены вопросы датировки событий трансформации на основе клТ-ДНК, организованных в виде несовершенных повторов, а также использования филогенетических исследований при изучении вопросов биоразнообразия генов Т-ДНК агробактериального происхождения.

Ключевые слова: агробактериальная трансформация; клеточная Т-ДНК; филогенетика; *Nicotiana*; *Camellia*; *Vaccinium*; *Arachis*

Для цитирования: Матвеева Т.В., Журбенко П.М., Хафизова Г.В., Шапошников А.Д., Жидкин Р.Р., Родионов А.В. Последовательности ДНК агробактериального происхождения в филогенетических исследованиях растений. Вавиловский журнал генетики и селекции. 2025;29(6):856-867. doi 10.18699/vjgb-25-93

Финансирование. Работа выполнена при поддержке Российского научного фонда, грант 25-26-00123.

Agrobacterium-derived DNA sequences in phylogenetic studies of plants

T.V. Matveeva , P.M. Zhurbenko , G.V. Khafizova , A.D. Shaposhnikov , R.R. Zhidkin , A.V. Rodionov ¹ Saint-Petersburg State University, St. Petersburg, Russia² Komarov Botanical Institute of the Russian Academy of Sciences, St. Petersburg, Russia³ University of Houston, Houston, United States radishlet@gmail.com

Abstract. One of the main methods for obtaining transgenic plants is *Agrobacterium*-mediated transformation. This process relies on the ability of certain soil bacteria, specifically from the genera *Agrobacterium* and *Rhizobium*, to transfer and integrate a fragment of their plasmid into the chromosome of the recipient plant. This transferred DNA is referred to as T-DNA. Laboratory studies have demonstrated that whole plants can be regenerated from transgenic cells. It soon became evident that similar processes occur in nature, leading to the emergence of naturally transgenic plants, or natural GMOs. Thus, naturally transgenic plants possess homologues of the T-DNA genes from agrobacteria

in their genomes (cellular T-DNA, or cT-DNA). These sequences are inherited through multiple sexual generations and retain their functionality. Furthermore, the potential for using newly acquired plant sequences in phylogenetic studies has been established, as cT-DNAs are clearly defined, highly specific, and recognizable DNA fragments that differ from typical plant DNA sequences. They are not found in untransformed ancestors, and their integration at specific chromosomal sites marks a monophyletic group of species. This review highlights the diversity of cellular T-DNAs and their potential use as phylogenetic markers. It includes a description of the main methodological approaches to such studies and discusses specific examples that clarify controversial points in the phylogeny of the genera *Nicotiana*, *Camellia*, *Vaccinium*, and *Arachis*. An important aspect of phylogenetic analysis based on cT-DNA is the assembly of individual alleles, which enables the tracking of interspecific hybridization events. This approach demonstrated the incomplete process of speciation within the *Thea* section of the genus *Camellia* and confirmed the role of interspecific hybridization in the breeding of North American blueberries. The review also addresses the dating of transformation events based on cT-DNA, which are organized in the form of imperfect repeats, as well as the application of phylogenetic studies to investigate the biodiversity of agrobacterial T-DNA genes.

Key words: agrobacterium-mediated transformation; cellular T-DNA; phylogenetics; *Nicotiana*; *Camellia*; *Vaccinium*; *Arachis*

For citation: Matveeva T.V., Zhurbenko P.M., Khafizova G.V., Shaposhnikov A.D., Zhidkin R.R., Rodionov A.V. Agrobacterium-derived DNA sequences in phylogenetic studies of plants. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2025;29(6):856-867. doi 10.18699/vjgb-25-93

Введение

Агробактериальная трансформация – наиболее распространенный на сегодняшний день метод получения трансгенных растений для нужд сельского хозяйства, медицины, ветеринарии и других отраслей народного хозяйства. Метод основан на способности почвенных бактерий *Agrobacterium tumefaciens* (Smith and Townsend 1907) Conn 1942 (Approved Lists 1980) и *Rhizobium rhizogenes* (Riker et al. 1930) Young et al. 2001 переносить фрагмент своей плазмиды и интегрировать его в ДНК растения (Schell, Van Montagu, 1977; Bahramnejad et al., 2019). Эти последовательности получили название Т-ДНК (T-DNA, transferred DNA – переносимая ДНК).

В природных условиях процесс переноса и интеграции Т-ДНК в хромосому растения-хозяина обычно вызывает разрастание трансгенной ткани на нормальном растении по причине экспрессии генов Т-ДНК, ответственных за регуляцию морфогенеза (Nester, 2015). Однако описана возможность регенерации полностью трансгенных растений из таких тканей (Terfer, 1990; Christey, 2001). Множество фактов указывает на то, что этот процесс происходит и в природе, без участия человека, в результате чего появляются формы, содержащие в геномах последовательности бактериального происхождения и передающие их в ряду половых поколений (Матвеева, 2021). Первые природно-трансгенные растения были обнаружены среди представителей рода *Nicotiana* в 1983 г. (White et al., 1982, 1983), а к настоящему времени известны десятки родов двудольных, в пределах которых описаны природно-трансгенные виды (Matveeva, Otten, 2019; Matveeva, 2021). Т-ДНК в составе геномов природных ГМО (пГМО) получила название клеточной Т-ДНК (клТ-ДНК) (Матвеева, 2021). Сопоставляя количество видов с секвенированными геномами и количество пГМО среди них, можно сказать, что следы агробактериальной трансформации сохранились у семи процентов видов двудольных растений (Matveeva, Otten, 2019). Таким образом, в руки ученых попал обширный материал для изучения функций бактериальных генов в растениях, их эволюционной судьбы.

Не менее интересные данные были получены с использованием Т-ДНК как филогенетического маркера,

поскольку новоприобретенные последовательности имеют несколько преимуществ для изучения происхождения и эволюции видов пГМО (Матвеева и др., 2011; Chen et al., 2022; Zhidkin et al., 2023; Bogomaz et al., 2024). Так, клТ-ДНК являются четко определенными, высоко-специфичными и узнаваемыми фрагментами ДНК, не похожими на последовательности ДНК растений. Они не встречаются у нетрансформированных предков, и их интеграция в определенном хромосомном сайте маркирует монофилетическую группу видов. Как правило, клТ-ДНК представлены единичными вставками, что служит значительным преимуществом по сравнению с классическими ядерными маркерами. Наконец, клТ-ДНК могут быть достаточно протяженными и древними, накапливающими однонуклеотидные замены (ОНЗ), что позволяет строить филогенетические деревья. Вероятность того, что в двух независимых филогенетических ветвях одна и та же последовательность Т-ДНК встроится в геном в одну и ту же мишень, с идентичными границами, кажется нам событием не исключенным, но маловероятным. Это делает инсерции Т-ДНК, так же как инсерции в геном транспозонов, чрезвычайно важными синапоморфиями (Shedlock, Okada, 2000; Doronina et al., 2022) и, следовательно, мощными инструментами для систематики. Их анализ в масштабах генома также может помочь выявить причины конфликта филогенетического сигнала в случае его возникновения (Kuritzin et al., 2016). Наконец, повторяющиеся последовательности в клТ-ДНК могут использоваться для оценки времени с момента трансформации или в качестве относительных временных маркеров (Chen et al., 2022).

Структуры Т-ДНК

Физическая структура Т-ДНК различается между плазмидами из разных штаммов агробактерий и может быть классифицирована в зависимости от количества переносимых в геном растения фрагментов, закодированных в одной плазмиде. Т-ДНК может быть непрерывной, как, например, у манныпиновых, микомпиновых и кукумопиновых штаммов (Jouanin, 1984; Hansen et al., 1991). Она представляет собой непрерывный участок ДНК, фланкированный пограничными последовательностями на обоих

концах (как в плазмиде pRi8196). В плазмидах другого типа, например в pRiA4, Т-ДНК разделена на два сегмента – TL-ДНК и TR-ДНК (White et al., 1985), между которыми находится в качестве спейсера не переносимый в растения участок ДНК размером около 15 kb. Независимо от различий их структуры, во всех упомянутых выше Т-ДНК содержатся гены синтеза опинов, используемых агробактериями как источник питания, и так называемые онкогены, продукты которых заставляют растительные клетки делиться.

Структуры клТ-ДНК пГМО более разнообразны. Среди них преобладают те, что содержат только гены синтеза опинов. Этот феномен можно объяснить, по крайней мере, тремя способами. Во-первых, в известных к настоящему времени Т-ДНК гены синтеза опинов находятся ближе к правой границе. При вырезании одноцепочечной Т-ДНК из плазмиды в ходе трансформации белок VirD2 ковалентно связывается с 5'-концом Т-ДНК у ее правой границы, а затем направляет Т-цепь через систему секреции белка типа IV в растительную клетку (Gelvin, 2021). Поскольку делеции происходят при интеграции Т-ДНК в хромосому растения, то они более обширны на 3'-, чем на 5'-конце, который пространственно защищен белком VirD2 (Gelvin, 2021). Во-вторых, можно ожидать наличие штаммов агробактерий, в Т-ДНК которых присутствуют только гены опин-синтаз. В-третьих, нельзя исключать возможность потери большей части Т-ДНК уже после ее интеграции в хромосому в процессе эволюции потомков природного трансформанта с сохранением только генов синтеза опинов, поскольку они не влияют на морфогенез в той степени, как онкогены, иными словами, не приводят к аномалиям роста и развития (Matveeva, Otten, 2021).

На втором месте по распространенности находятся протяженные фрагменты Т-ДНК с онкогенами и генами синтеза опинов. Эти клТ-ДНК, как правило, представлены инвертированными несовершенными повторами. Данную особенность можно использовать при датировке актов трансформации. Наименее распространенными являются клТ-ДНК, содержащие только онкогены (Matveeva, Otten, 2019; Matveeva, 2021).

Рассмотрим подробнее вопрос датировки актов возникновения пГМО в ходе эволюции.

Подходы к определению времени событий трансформаций у пГМО

Датировка событий трансформации у пГМО может быть проведена на основе метода молекулярных часов. Особенности механизма интеграции Т-ДНК в геном инфицированного растения приводят к образованию повторяющихся последовательностей Т-ДНК или ее участков (Tzfira et al., 2004; Singer, 2018). Эти повторы являются длинными клТ-ДНК, которые представлены несовершенными инвертированными повторами или их частями, образовавшимися в ходе делеции участка или его встраивания (Matveeva, Otten, 2019; Matveeva, 2021). Используя различия последовательности нуклеотидов таких повторов, можно определить с помощью коэффициента нуклеотидных замен приблизительное время расхождения данных повторов и, как следствие, время интеграции Т-ДНК в геном будущего пГМО (Gaut et al., 1996).

Расчет времени трансформации можно провести по формуле

$$T = \frac{d}{2r},$$

где T – приблизительное время интеграции клТ-ДНК; d – отношение нуклеотидных различий двух повторов; r – средний коэффициент нуклеотидных замен, 6.5×10^{-9} замены на сайт в год (Gaut et al., 1996; Lynch, Conery, 2000). На основе этого подхода была проведена датировка интеграции Т-ДНК у представителей рода *Camellia*, которые содержат различные структурные типы клТ-ДНК (Chen et al., 2022, 2023). Согласно этой датировке, наиболее древняя из вставок была интегрирована в геном около 7.5 млн лет назад, тогда как наиболее молодая была интегрирована 0.04 млн лет назад (Chen et al., 2023)¹.

Однако этот подход имеет ряд ограничений, которые не позволяют использовать его для определения времени интеграции многих клТ-ДНК. Первое из ограничений – невозможность обнаружения инвертированных повторов у всех клТ-ДНК, так как большая часть видов пГМО не содержит протяженных клТ-ДНК, где можно найти повторяющиеся прямые и инвертированные последовательности (Matveeva, 2021). Вторым является то, что даже не все протяженные клТ-ДНК имеют повторы. Так, только 6 из 12 типов клТ-ДНК у представителей рода *Camellia* содержат повторы, которые позволяют определить время интеграции (Chen et al., 2023). Еще одно важное условие при выборе последовательностей для датировки – необходимость исключить, что эти повторы могли образоваться в процессе других геномных перестроек, например в результате активности мобильных элементов генома.

Метод датировки клТ-ДНК, несомненно, следует использовать вместе с другими подходами, которые позволяют точно оценить скорость нуклеотидных замен именно в этом виде, сравнить время дивергенции с другими близкородственными таксонами для валидации полученных данных. Хотя данный метод датировки времени трансформации пГМО еще требует дальнейшей доработки и верификации, он может позволить лучше понять эволюционный процесс, который претерпевали пГМО после агробактериальной трансформации.

Множественные акты трансформации в эволюции рода *Nicotiana*

Изучение феномена генетической трансформации растений в природе без участия человека началось с вида *Nicotiana glauca* Graham, в ядерном геноме которого были обнаружены последовательности, гомологичные агробактериальным Т-ДНК (White et al., 1983). Первая найденная клТ-ДНК была названа gT. Она была организована как несовершенный инвертированный повтор и состояла из одной копии гомолога *rolB* и двух копий гомологов *rolC*, *ORF13*, *ORF1* и *mis* (Suzuki et al., 2002). Последовательность gT послужила отправной точкой для проведения поиска клТ-ДНК в геномах других видов рода *Nicotiana* L. (Furner et al., 1986; Intrieri, Buiatti, 2001). На сегодняшний

¹ В результате общения с авторами статьи было уточнено, что приведенные в статье данные о времени вставок рассчитаны по ошибочной формуле, а именно завышены в два раза. В нашем обзоре приводятся корректные значения.

день известно о 16 природно-трансгенных видах данного рода (Otten, 2020). Род *Nicotiana* включает двенадцать секций (Knapp et al., 2004), в шести из которых описаны природно-трансгенные представители.

Поиск и анализ клТ-ДНК в растительных геномах долгое время осуществлялся на уровне отдельных генов (Furber et al., 1986; Meyer et al., 1995; Intrieri, Buiatti, 2001). В тот же период были предприняты попытки реконструкции филогенетических связей между клТ-ДНК разных видов табака и Т-ДНК агробактерий. Однако дальнейшие исследования поставили под сомнение некоторые более ранние эволюционные построения. Переход к полнотемным данным позволил не только искать клТ-ДНК и изучать их состав, но также оценивать их количество и сайты локализации в растительном геноме. Первый такой анализ был проведен для *N. tomentosiformis* Goodsp. (Chen et al., 2014), в геноме которого были найдены четыре разные по составу клТ-ДНК (ТА, ТВ, ТС и ТД) (см. таблицу), отличные от ранее исследованной гТ в *N. glauca* (Suzuki et al., 2002). Таким образом, стало ясно, что использование участков клТ-ДНК в качестве филогенетического маркера необходимо начинать с оценки количества клТ-ДНК в геноме и их сайтов локализации, маркирующих потомков конкретных событий трансформации.

Множественные протяженные клТ-ДНК у видов табака были использованы для датировки событий трансформации в эволюции. Было показано, что у *N. tomentosiformis* все четыре клТ-ДНК являются результатами независимых актов агробактериальной трансформации, произошедших в разное время. Клеточная Т-ДНК ТС была определена как самая древняя. По приблизительной оценке ее возраст составил 1 млн лет (Chen et al., 2014). Позже в геноме вида *N. otophora* Griseb., филогенетически близкого к *N. tomentosiformis*, были найдены две копии ТС, различающиеся между собой на 4 % и расположенные в том же сайте, что и ТС *N. tomentosiformis*. Общий сайт локализации указывает на то, что клТ-ДНК ТС, скорее всего, была получена неким общим для *N. tomentosiformis* и *N. otophora* предковым видом (рис. 1).

В ходе эволюции данного вида, вероятно, случилась дупликация ТС, после чего образовались вид *N. otophora*, несущий в геноме две копии ТС, и *N. tomentosiformis*, утративший одну копию ТС в ходе видообразования (Chen et al., 2018). Попытки восстановить ход эволюционных событий, в том числе видообразования, в роду *Nicotiana* осложняет свойственный ему ретикулярный, или сетчатый, тип эволюции, для которого характерно частичное слияние предковых ветвей с образованием гибридных форм (Knapp et al., 2004). Так, например, известный представитель рода *Nicotiana*, *N. tabacum* L. (культурный табак), является межвидовым гибридом *N. tomentosiformis* и *N. sylvestris* Speg. (Yukawa et al., 2006). Геном *N. tabacum* содержит три из четырех клТ-ДНК, найденных у *N. tomentosiformis* (ТА, ТВ, ТД), на основании чего можно предположить, что события агробактериальной трансформации *N. tomentosiformis* предшествовали видообразованию *N. tabacum* (Chen et al., 2014). В пользу данного предположения свидетельствует оценка времени видообразования *N. tabacum*, проведенная методами про-

Структуры клТ-ДНК у *N. tomentosiformis*

Название	Перечень генов (в скобках отмечены плечи несовершенных инвертированных повторов)
ТА	(orf8, orf3, rolA, rolB, rolC, orf13, orf14, mis) (mis, orf14, orf13, rolC, rolB, rolA, orf8)
ТВ	(mis, orf14) (orf14, mis, ags-like, mas1', mas2')
ТС	(ocl, orf2, orf3, orf8, rolA, rolB) (rolB, rolA, orf8, orf3, orf2, c)
ТД	(orf14), orf55, orf 15, orf511, (orf1)

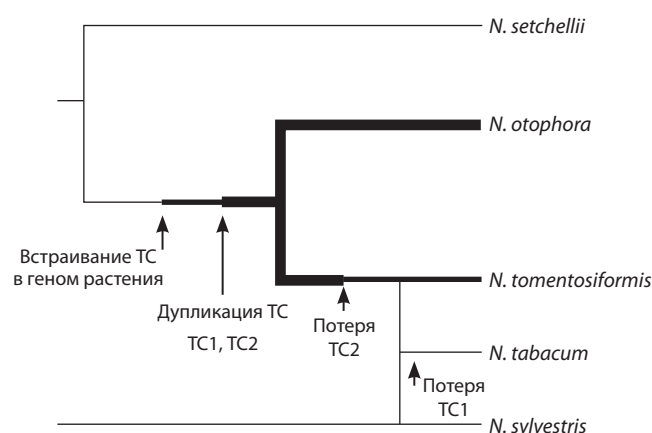


Рис. 1. Модель эволюции клТ-ДНК ТС в геномах *Nicotiana*.

Толщина линий отражает присутствие в геноме вида одной или двух копий ТС либо ее отсутствие, по (Chen et al., 2018).

точной цитометрии и геномной гибридизации, согласно которой возраст вида составляет менее 600 лет (Leitch et al., 2008). КлТ-ДНК ТС, копия которой содержится в геноме предкового вида *N. tomentosiformis*, вероятно, была утрачена в ходе видообразования *N. tabacum* (Chen et al., 2014) (см. рис. 1). Используя данные датировок клТ-ДНК и результаты оценки времени видообразования, можно проследить путь ТС от первого переноса ее в растительный геном до утраты либо закрепления в различных видах, произошедших от предковой формы на разных этапах эволюции рода.

Все вышеописанные представители относятся к секциям *Tomentosae* и *Nicotiana*. Для представителей секции *Noctiflorae* (*N. glauca*, *N. noctiflora* Hook.) тоже описано присутствие трех различающихся клТ-ДНК, полученных в ходе независимых актов трансформации (одна вставка в геноме *N. glauca*, две другие – в *N. noctiflora*) (Khafizova et al., 2023), однако для восстановления филогенетических отношений в данной ветви необходимо больше природно-трансгенных представителей с секвенированными геномами.

Таким образом, степень изученности таксона напрямую связана с доступностью материала и его хозяйственной значимостью. Внутривидовое разнообразие культурного табака также представляет интерес, поэтому в следующем разделе мы остановимся на использовании клТ-ДНК в исследовании этого вопроса.

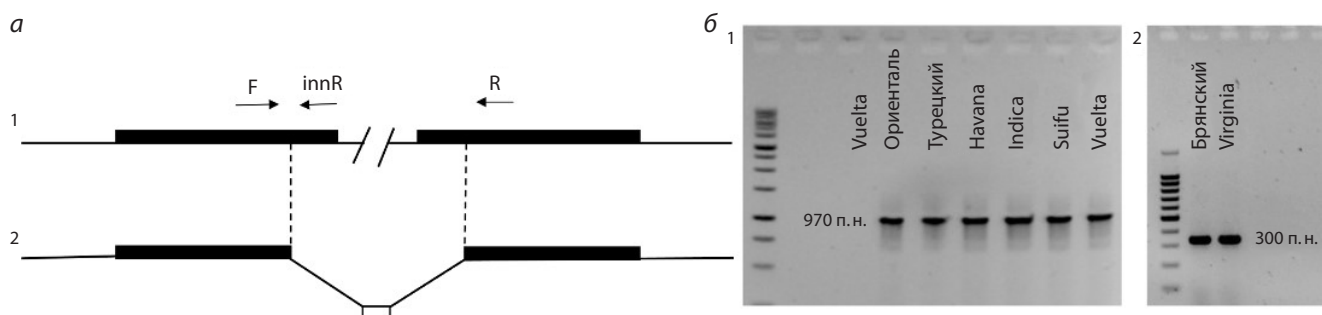


Рис. 2. КлТ-ДНК ТА в геномах *N. tabacum*.

a – схема ТА; центральная часть не показана, черными прямоугольниками отмечен ген *orf13*, штриховыми линиями – границы делеции. Отмечены праймеры, подобранные в работе (Хафизова, Матвеева, 2020); 1 – клТ-ДНК ТА без делеции; 2 – клТ-ДНК ТА с делецией, белым прямоугольником показана последовательность неизвестного происхождения длиной 42 п.н.; *б* – фрагменты, полученные с использованием праймеров F и innR (1), F и R (2) при ПЦР-анализе восьми сортов *N. tabacum*, по (Хафизова, Матвеева, 2020).

Внутривидовая изменчивость клТ-ДНК у табака

Nicotiana tabacum, или культурный табак, – аллотетраплоид, образовавшийся в ходе межвидовой гибридизации видов *N. tomentosiformis* и *N. sylvestris* (Yukawa et al., 2006). Количество существующих сортов и разновидностей *N. tabacum* насчитывает десятки тысяч, при этом родственные отношения между ними установлены далеко не во всех случаях (Moon et al., 2009; Fricano et al., 2012; Sierro et al., 2014). В основе самой широко используемой сегодня внутривидовой классификации культурного табака лежат различия в морфологии растений, а также в количественном и качественном составе их вторичных метаболитов. Данные показатели определяют ключевые характеристики табачного сырья, а саму классификацию называют рыночной (Lewis, Nicholson, 2007). Всего выделяют восемь рыночных классов: Берли, наполнитель для сигар, табак для закручивания сигар, темный табак воздушной сушки, темный табак паровой сушки, табак дымовой сушки, Мэриленд и восточный табак (Moon et al., 2009). При этом, что среди сортов культурного табака наблюдается высокий уровень фенотипической изменчивости (Lewis, Nicholson 2007), уровень нуклеотидной изменчивости, выявленный методами полиморфизма длины рестрикционного фрагмента (RFLP), случайно амплифицированной полиморфной ДНК (RAPD), маркерами полиморфизма длины амплифицированного фрагмента (AFLP) и полногеномным поиском ассоциаций (GWAS), сравнительно невысок (Brandle, Bai, 1999; Ren, Timko, 2001; Rossi et al., 2001; Tong et al., 2020).

Для проведения филогенетического анализа и построения карты, отражающей истинные родственные отношения внутри вида, необходимо использовать дополнительные молекулярные маркеры. Одним из таких маркеров может служить клТ-ДНК. Изучение клТ-ДНК в полногеномных данных трех сортов *N. tabacum* позволило выявить протяженную делецию в центральной части клТ-ДНК ТА у сорта Basma/Xanti. При этом в сортах K326 и TN90 была обнаружена последовательность ТА без делеции (Chen et al., 2014). Различия в структуре ТА уже были показаны ранее методом ПЦР для сортов *N. tabacum* Basma Drama 2, Samsoun и Xanthi, клТ-ДНК которых содержат неполные последовательности гомолога гена *orf13*, в отличие от клТ-ДНК в сортах Wisconsin 38 и Havana 425, а также в

виде *N. tomentosiformis* (Mohajjel-Shoja et al., 2011). Позже с помощью метода ПЦР были проанализированы еще восемь сортов культурного табака: Vuelta abajo, Suifu, Black Indian, Havana 307, Турецкий, Ориенталь, Брянский 91 и Virginia × Berley 38 (рис. 2). В результате проведенного анализа описанная ранее делеция была обнаружена в сортах Брянский 91 и Virginia × Berley 38, при этом сайт локализации делеции совпал с сайтами делеции в геномах сортов Basma Drama 2, Samsoun, Xanthi и Basma/Xanti с точностью до нуклеотида (Хафизова, Матвеева, 2020).

Таким образом, использование структурного варианта клТ-ДНК в качестве молекулярного маркера позволило отделить сорта, принадлежащие к рыночному классу восточных табаков, в отдельную группу по признаку делеции в ТА. Можно предположить, что центральная часть ТА была утрачена предковой формой *N. tabacum*, которая дала начало данному классу. К настоящему времени делеция в ТА описана только у пяти сортов культурного табака. На основании анализа подобной выборки рано делать вывод о том, является ли наличие данной делеции маркером, по которому сорт может быть отнесен к восточным табакам. При этом результаты исследований указывают на то, что структурный полиморфизм клТ-ДНК в сортах *N. tabacum* может быть использован в качестве одного из молекулярных маркеров для изучения внутривидовых родственных отношений сортов. С расширением списка секвенированных клТ-ДНК различных сортов культурного табака в перспективе, возможно, будут найдены новые структурные различия в последовательностях клТ-ДНК, на основании которых могут быть разработаны новые маркеры для решения поставленной задачи.

Тонкий полиморфизм клТ-ДНК в филогенетических исследованиях

Тонкий полиморфизм клТ-ДНК общего происхождения в геноме (с общим сайтом локализации) может быть использован для изучения межвидовой изменчивости и реконструкции филогенетических связей внутри монофилитической группы видов-потомков древнего трансформанта. Важно понимать, что для корректной оценки завершенности процесса видообразования у растений имеет смысл оценить и сопоставить как внутри-, так и межвидовую изменчивость по изучаемым маркерам. С этой целью

было предложено реконструировать отдельные аллели изучаемых маркеров, особенно для перекрестно-опыляемых видов (Chen et al., 2022). Для начала остановимся на методологии разделения аллелей.

Методы исследования

В научной литературе разделением аллелей или гаплотипов (англ. allele phasing/haplotype phasing) называют получение последовательностей фрагментов ДНК, находящихся на одной хромосоме среди пары гомологичных или, в случае полиплоидов, еще и среди гомеологичных хромосом. При этом об аллелях чаще говорят, имея в виду короткие фрагменты, тогда как про гаплотипы – если речь идет о более протяженных участках.

Разделение гаплотипов дает дополнительную информацию по сравнению с информацией, полученной только на основании консенсусной последовательности генома. Точные последовательности гаплотипов используются в самых различных исследованиях, включая филогенетическую реконструкцию (Tiley et al., 2024) и изучение гибридов (Sun et al., 2020).

Существуют разные подходы для разделения гаплотипов (Snyder et al., 2015). Так, можно провести физическое разделение фрагментов гомологичных хромосом с помощью методов молекулярной биологии с дальнейшим секвенированием фрагментов по отдельности. Для этого можно использовать, например, бактериальное клонирование. Другим подходом является разделение гаплотипов на основании данных генотипирования о частотах нуклеотидов в полиморфных позициях в популяции. В этом случае используют различные статистические методы (Browning S.R., Browning B.L., 2011). Однако наиболее активно развивается подход, использующий данные высокопроизводительного секвенирования. Внутри этого подхода можно выделить методы, основанные на сборке коротких прочтений или на картировании коротких прочтений на референсный геном (Zhang et al., 2020). Рассмотрим подробнее последний случай.

Если две полиморфные позиции находятся на одном секвенированном фрагменте ДНК (прочитанном с одного или обоих коротких прочтений), то они принадлежат одному гаплотипу. Последовательность гаплотипа можно продолжать реконструировать, пока находятся фрагменты, соединяющие соседние полиморфные позиции. Соответственно, чем больше плотность полиморфных позиций, длиннее перекрытия между короткими прочтениями и больше глубина прочтения, тем надежнее и длиннее последовательность гаплотипов, которую можно реконструировать. Эти параметры особенно важны при анализе полиплоидных геномов, где необходимо разделить более двух гаплотипов, некоторые из которых могут мало различаться на достаточно протяженных участках (Schinnerer et al., 2020).

Перечисленные выше показатели будут зависеть и от технологии секвенирования. Лучших результатов можно достичь с технологией Hi-Fi, которая позволяет получать длинные прочтения протяженностью несколько десятков тысяч пар нуклеотидов высокого качества с точностью прочтения более 99 % (Wenger et al., 2019). С помощью секвенирования Hi-Fi можно получать сборки протяжен-

ных участков гаплотипов, вплоть до длины всего генома (Tanaka et al., 2023). С использованием технологии Oxford Nanopore можно получать очень длинные прочтения, длиной до нескольких миллионов пар нуклеотидов, но с более низким качеством прочтения – точность около 90 % (Wang Y. et al., 2021). С помощью этих длинных прочтений при достаточной глубине секвенирования можно также успешно разделять на гаплотипы протяженные участки генома. Помимо длинных прочтений, для разделения гаплотипов можно использовать и короткие, получаемые с применением технологии Illumina. Однако такие фрагменты будут ограничены консервативными регионами, где перекрытия прочтений недостаточно для соединения соседних полиморфных позиций.

Одной из популярных программ для разделения гаплотипов является WhatsHap (Martin et al., 2016). Она умеет работать с прочтениями от всех перечисленных методов секвенирования как для диплоидных, так и для полиплоидных организмов. На вход программа требует последовательность референсного генома, файл в формате bam с картированием прочтений на референсный геном и файл в формате vcf, содержащий информацию о полиморфных позициях, отличающих картированные прочтения от референса. Результатом разделения будет модифицированный vcf файл, содержащий информацию о принадлежности полиморфных позиций к гаплотипам. На основании этого файла можно получить информацию о длинах и координатах разделенных фрагментов (haplotype blocks), последовательности гаплотипов в формате fasta, добавить в визуализацию картирования прочтений информацию об их принадлежности к гаплотипам и многое другое.

Разделять аллели можно и путем анализа результатов секвенирования по Сэнгеру. При секвенировании фрагментов ДНК гетерозигот на хроматограммах в одной и той же позиции появляются два «пика», соответствующих определенным нуклеотидам (Carr et al., 2009; Dehairs et al., 2016; Xie et al., 2019). Чтобы определить по этим данным последовательности аллелей изучаемого гена, полученные сиквенсы от каждого образца можно представить в виде вектора, каждая ячейка которого соответствует полиморфной позиции в гене для изучаемого вида. Затем каждую ячейку заполняют по правилу: «11» – если у данного образца в данной позиции наиболее распространенный нуклеотид; «00» – если наименее распространенный нуклеотид; «10» – если в этой позиции в геноме объекта встречаются два разных нуклеотида (предполагаемая гетерозигота). Полученные таким образом вектора впоследствии можно расписать по всевозможным комбинациям гетерозиготных позиций «10», оставив нетронутыми однозначные позиции «11» и «00». Поскольку среди образцов имеются гомозиготы и образцы с одной ОНЗ, то их последовательности формируют первичный пул аллелей, которые в дальнейшем могут быть обнаружены в остальных образцах. Каждая из этих аллелей в диплоиде должна соответствовать аллели с альтернативными значениями в полиморфных позициях, поэтому гомологичную пару можно найти для первичной аллели (Zhidkin et al., 2023). Перебирая таким образом возможные комбинации аллелей и выбирая те, чья частота встречаемости выше, можно определить генотипы образцов.

Описанные выше подходы были использованы для изучения внутри- и межвидовой изменчивости растений родов *Camellia*, *Vaccinium*, *Arachis*. Остановимся подробнее на этих примерах.

Полиморфизм клТ-ДНК у видов рода *Camellia*

Род *Camellia* L. принадлежит к семейству Theaceae и содержит несколько экономически и культурно значимых видов. *Camellia sinensis* (L.) Kuntze состоит из двух основных разновидностей: var. *sinensis* (обычно используется для производства зеленого чая) и var. *assamica* (в основном используется для производства черного чая). В работе (Chen et al., 2023) проанализировано 72 вида, относящихся к 12 из 14 секций рода *Camellia*, где описано не менее 12 различных вставок клТ-ДНК. Суммарно эти последовательности занимают 374 kb и содержат 47 открытых рамок считываний. Обнаруженные гены можно разделить на четыре типа: к первому относятся 19 *Plast* генов, ко второму – 6 генов синтеза опинов, к третьему – 4 гена, кодирующих триптофан монооксигеназы, к четвертому – все остальные гены с неизвестными функциями. Белковые последовательности генов клТ-ДНК обладают различным уровнем сходства с известными агробактериальными последовательностями. Средний уровень сходства равен 73.8 % (стандартное отклонение 12.8 %), тогда как минимальное и максимальное значения составляют 46 и 92 % соответственно.

Любопытно, что некоторые гены, гомологичные найденным у камелий, были обнаружены также в нескольких видах грибов, причем как у аско-, так и у базидиомицетов. В 7 из 12 вставок присутствуют внутренние инвертированные повторы, возникшие, вероятно, из-за встраивания одновременно нескольких копий Т-ДНК. Различия между повторами варьировали для разных фрагментов от 0.05 до 10 %. Учитывая, что на момент встраивания повторы были идентичны, и используя универсальное значение частоты замен на позицию в год 6.5×10^{-9} (Gaut et al., 1996) и количество замен между повторами, можно рассчитать примерное время вставки фрагментов (Haubold, Wiehe, 2001). Оно составляет от 0.04 до 7.5 млн лет назад (м. л. н.) (Chen et al., 2023).

Поскольку на разных этапах эволюции рода фрагменты встраивались в разные филогенетические линии, то они оказываются специфичными для современных таксонов различного уровня. Показано, что часть фрагментов была утрачена в определенных линиях и что самые молодые фрагменты еще не успели окончательно зафиксироваться в популяциях своих видов, присутствуя только в ряде особей.

При известных последовательностях клТ-ДНК их наличие/отсутствие в разных филогенетических линиях исследуемого таксона тоже является удобным признаком, состояние которого можно легко оценить. Так, последовательность дивергенции секций внутри рода, полученная на основании данных о наличии различных клТ-ДНК, согласуется с филогенетическим деревом, построенным по результатам полногеномного секвенирования (Wu et al., 2022). Различия между повторами также могут быть использованы для уточнения относительного и абсолютного времени дивергенции.

У представителей секции *Thea* описана только одна разновидность клТ-ДНК, которая была использована для уточнения филогенетических отношений видов этой секции. Данная секция содержит одиннадцать видов (Min, Bartholomew, 2007), включая чайный куст *C. sinensis* (L.) O. Kuntze и дикие виды чая. *C. sinensis* далее подразделяется на разновидности *sinensis*, *assamica*, *pubilimba* и *dehungensis*. Для девяти видов секции *Thea* рода *Camellia* была построена филогения на основании одной из вставок, описанных выше (Chen et al., 2022). Вставка имеет длину 5.5 kb, включает три гена, *acs*-like, *sus*-like и фрагмент *rolB*-like, и имеет на концах инвертированные повторы длиной 1 kb каждый. Анализ последовательностей показал, что гены *acs*-like и *sus*-like содержат стоп-кодоны, однако было отмечено, что у части видов они возникли независимо в разных позициях. Примерное время вставки фрагмента составляет 7.5 м. л. н., что примерно совпадает со временем возникновения секции *Thea* – 6.7 м. л. н. (Wu et al., 2022). Таким образом, данная вставка, от которой мы ожидаем, что она лишена возможных искажений, связанных с неполной сортировкой аллелей (incomplete lineage sorting), является подходящим маркером для этой группы.

В процессе сборки фрагментов из коротких прочтений обнаружено, что многие образцы содержат два аллеля, при этом от 2 до 7 % позиций являются полиморфными. В результате разделения 142 образцов было получено 225 аллелей, на основании которых построено филогенетическое дерево (рис. 3). В большинстве случаев полученные клады не совпадали с границами видов, а у многих гетерозиготных образцов присутствовали аллели из разных отдаленных кладов. Особенно заметно присутствие в части образцов аллелей из клада, удаленных на значительное эволюционное расстояние. Максимальное количество замен между аллелями из разных кладов, встречающихся в одном образце, составило около 4 %. Скорее всего, такая дивергенция аллелей была бы невозможна в популяции одного вида, поэтому выдвинуто предположение, что эволюция аллелей из основных кладов происходила в границах отдельных предковых видов секции *Thea*, после чего современные виды, с высоким разнообразием аллелей в популяциях, образовались в результате интрогрессивных скрещиваний. Если предположить, что количество основных кладов совпадает с количеством предковых видов, то по топологии дерева можно оценить степень родства этих видов, по количеству замен между аллелями – время дивергенции, а по частоте аллелей – вклад в современные популяции. Так, например, можно выделить две основные пары филогенетически отдаленных кладов, между которыми происходила гибридизация. Соответственно, в результате одних скрещиваний сформировался вид *C. sinensis*, причем от одного из предполагаемых предковых видов присутствует в два раза больше аллелей, чем от второго. Расстояние между аллелями в 2.5 % позволяет предположить, что эти виды дивергировали 1.9 м. л. н. Тогда как в результате скрещиваний с равным участием внутри второй пары, дивергировавшей 2.9 м. л. н., возник вид *C. tchangensis*.

Наша работа позволила поставить вопрос о пересмотре границ видов, интенсивности межвидовых скрещиваний и других особенностях эволюции видов секции.

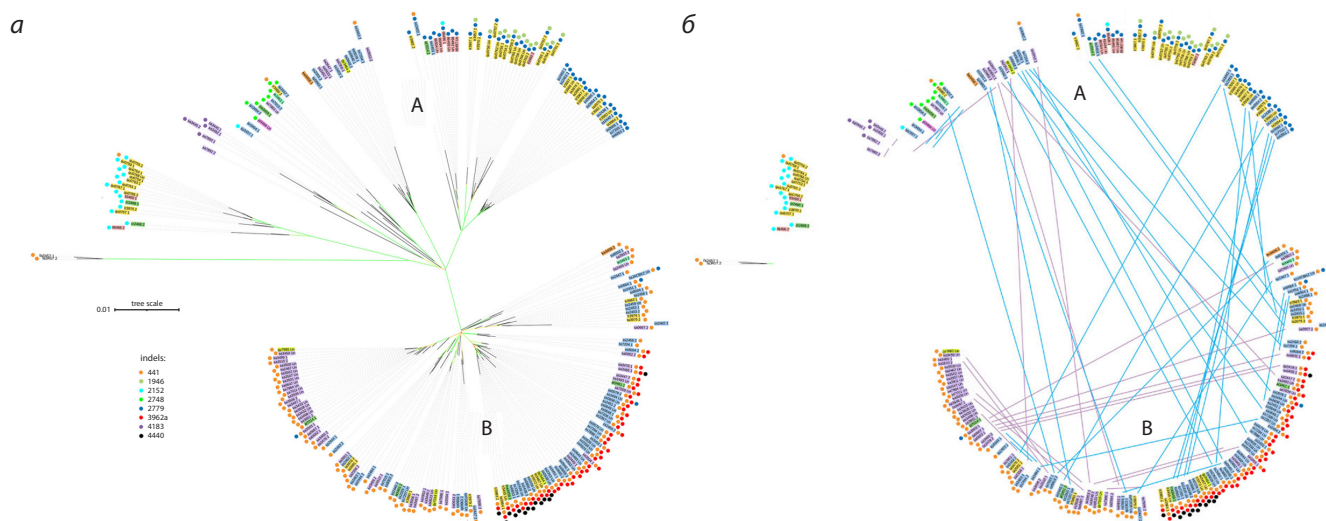


Рис. 3. Филогенетический анализ секции *Thea* с использованием кЛТ-ДНК: а – филогенетическое дерево 225 аллелей кЛТ-ДНК девяти видов рода *Camellia* секции *Thea*; б – линиями соединены аллели, находящиеся в одном образце *C. sinensis*.

Две основные клады на обоих рисунках обозначены латинскими буквами А и В. Цветовые обозначения видов: *C. sinensis* (включает *C. sinensis* var. *sinensis*, *C. pubilimba*, *C. angustifolia*) – голубой, *C. sinensis* var. *assamica* – фиолетовый, *C. tachangensis* – темно-желтый, *C. crassicolumna* – зеленый, *C. gymnogyna* – желтый, *C. taliensis* – розовый, *C. leptophylla* – темно-зеленый, *C. kwangsiensis* – оранжевый, *C. ptilophylla* – красный, *C. fangchengensis* – без цвета. Инделы обозначены цветными точками. Значения бутстреп показаны цветом: зеленый – 85–100 %, желтый – 50–85 %, красный – менее 20 %. Рисунок создан на основе иллюстраций из (Chen et al., 2022), опубликованных под лицензией CC BY 4.0.

Полиморфизм кЛТ-ДНК у видов рода *Vaccinium*

Первым описанным в литературе природно-трансгенным видом из рода *Vaccinium* L. была клюква крупноплодная, *V. macrocarpon* Aiton (Matveeva, Otten, 2019). При помощи BLAST-алгоритма в полном геноме клюквы была найдена последовательность кЛТ-ДНК, представленная единичной, интактной копией *rolB/C*-подобного гена. Данный ген относится к *plast*-генам, и кодируемая им аминокислотная последовательность более схожа с *Plast*-белками таких организмов, как *Laccaria bicolor* (Matre) Orton, *Nyssa sinensis* Oliver и *Ensifer* sp., чем с аналогичными последовательностями из геномов известных видов агробактерий. Это может говорить о происхождении указанных последовательностей в ходе различных актов трансформации одним и тем же неизвестным видом агробактерий (Matveeva, Otten, 2019). Вторым после клюквы крупноплодной видом *Vaccinium*, содержащим гомолог изучаемого гена, была голубика высокорослая, *V. corymbosum* (Matveeva, 2021). Позже нам удалось показать наличие *rolB/C*-подобного гена еще у 26 видов *Vaccinium*, а также у вида *Agapetes serpens* (Wight) Sleumer, относящегося к тому же семейству Вересковые – Ericaceae (Zhidkin et al., 2023). Характерным является и общий сайт локализации обнаруженных последовательностей у видов *Vaccinium* и *A. serpens*. Такое широкое распространение трансгена по роду и общий сайт локализации указывают на трансформацию общего предка изучаемых видов. А значит, последовательность *rolB/C*-подобного гена можно использовать в качестве маркера для реконструкции филогенетических взаимоотношений между ними. Данная задача весьма актуальна, поскольку филогения рода *Vaccinium* остается противоречивой (Becker et al., 2024).

У всех изученных видов *rolB/C*-подобный ген представлен полноразмерной последовательностью. Исключение составила клюква обыкновенная *V. oxycoccos* L., у

большинства образцов которой имелись крупные делеции различной длины в центральной части гена, но в то же время существовали и образцы с полноразмерной последовательностью *rolB/C*-подобного гена. У остальных видов полиморфизм *rolB/C*-подобного гена представлен однонуклеотидными заменами и инделами, кратными трем, т. е. не нарушающими открытую рамку считывания. Паттерн таких нуклеотидных различий был видоспецифичен.

Несмотря на то что такие представители рода, как клюква, голубика, черника и брусника, издавна использовались человеком в пищу и в медицинских целях, селекция данных культур отсчитывается с начала XX в. (Wang H. et al., 2017; Vorsa, Zalapa, 2019; Sultana et al., 2020). Именно с началом селекционной работы связана разработка системы рода, первая из которых была предложена в 1945 г. (Camp, 1945). Представленная классическая система рода опиралась на различные морфологические признаки и делила род на секции. В дальнейшем система регулярно дополнялась и изменялась, поскольку в роду распространены межвидовая гибридизация и полиплоидизация (Camp, Gilly, 1943; Hancock et al., 2008). Эти особенности не позволяли однозначно определить филогенетические отношения между видами, что побудило с развитием молекулярной филогенетики применить для решения проблем методы ДНК-штрихкодирования (Kron et al., 2002; Powell, Kron, 2003). Полученная по последовательностям ITS (internal transcribed spacer – внутренний транскрибируемый спейсер) и *matK* (maturase K – пластидный ген матуразы К) дендрограмма шла вразрез с традиционными представлениями о разделении рода на секции, а также указывала на полифилетичность рода *Vaccinium*. Однако использование классических филогенетических маркеров в кладистическом анализе для видов, в эволюции которых существенную роль играли процессы гибридизации и

полиплоидизации, может приводить к ошибкам (Soltis et al., 2008). В то же время данные, полученные с помощью SSR-маркеров (simple sequence repeats – микросателлитная ДНК) (Zalapa et al., 2015; Schlautman et al., 2017), филогеномики (Diaz-Garcia et al., 2019; Kawash et al., 2022) и хемотаксономии (Leisner et al., 2017), имели меньше противоречий с классическими представлениями. Секвенирование геномов весьма трудоемко и дорогостояще, поэтому проводилось только на экономически значимых видах. *rolB/C*-подобный ген в качестве филогенетического маркера позволил вовлечь в работу больше видов.

У анализируемых видов внутривидовая изменчивость трансгена (в отличие от видов камелий секции *Thea*) оказалась ниже, чем межвидовая, а мозаичность некоторых клад могла быть объяснена актами гибридизации между видами, входящими в эти клады. Данные филогенетического анализа демонстрировали объединение в отдельные клады представителей секций *Oxycoccus*, *Vaccinium*, *Myrtillus* и *Conchophyllum*, тогда как виды из секции *Cyanococcus* не образовывали монофилетической группы, что может быть связано с ее полифилетичностью или событиями гибридизации в ходе создания сортов североамериканских голубик. Остальные исследованные в работе виды являются единичными представителями секций *Bracteata*, *Hemimyrtillus*, *Vitis-idaea*, *Oxycoccoides* и *Praestantia*, поэтому для изучения этих секций требуется дополнительный материал. Учитывая простоту и дешевизну разработанного молекулярного маркера, новые виды могут быть легко включены в дополнительный анализ. Иными словами, филогения рода *Vaccinium*, определенная по последовательностям *rolB/C*-подобного гена, имеет большее сходство с классической системой рода, чем с построенной на основе маркеров ITS и *matK*. Результаты, схожие с полученными при использовании маркера на основе *rolB/C*-подобного гена, были получены и другими авторами на основе NGS-подходов (next generation sequencing – секвенирование следующего поколения) (Diaz-Garcia et al., 2019; Kawash et al., 2022).

Полиморфизм клТ-ДНК у видов рода *Arachis*

Род *Arachis* L. насчитывает 80 видов и подразделяется на девять таксономических секций: *Arachis* (с геномами А, В, К), *Erectoides* (геном Е), *Extranervosae* (геном EX), *Procumbentes* (геном PR), *Caulorrhizae* (геном С), *Heteranthae* (геном H), *Rhizomatosae* (геном R), *Trierectoides* (геном TE) и *Triseminatae* (геном Т) (Stalker, 2017). Современное понимание эволюционных связей представителей рода *Arachis* основано на морфологических, географических, молекулярно-генетических и цитогенетических данных, но до сих пор существует много дискуссионных вопросов относительно структуры рода (Krapovickas, Gregory, 2007; Koppolu et al., 2010; Stalker, 2017; Tian et al., 2021).

Изначально пГМО были обнаружены в пределах секции *Arachis* у тетраплоидного культурного арахиса (*Arachis hypogaea* L.) и его предков: *A. duranensis* Krapov. & W.C. Greg. и *A. ipaensis* Krapov. & W.C. Greg., близкого к культурному арахису тетраплоидного вида *A. monticola* Krapov. & Rigoni (Matveeva, Otten, 2019) и диплоида *A. stenosperma* Krapov. & W.C. Greg. (Matveeva, Otten, 2021).

Арахис культурный и *A. monticola* содержат предковые геномы, известные как А и В. Геном А присутствует также в *A. duranensis* и *A. stenosperma*, тогда как *A. ipaensis* содержит геном В (Matveeva, Otten, 2019, 2021). В дальнейшем список пГМО был расширен представителями секций *Erectoides*, *Extranervosae*, *Procumbentes*, *Caulorrhizae* и *Heteranthae* (Bogomaz et al., 2024).

Всего сейчас известно 23 природно-трансгенных вида из этого рода. Велика вероятность того, что в ближайшее время перечень будет пополнен, поскольку изученные пГМО образуют монофилетическую группу, общий предок которой был трансформирован до того, как изучаемые секции дивергировали. Во всех исследованных видах был обнаружен гомолог гена кукумопинсинтазы (*cus*). В дополнение к *cus*-подобному гену виды с геномом В содержали остатки гена *mas2'*, виды с геномом PR – остатки гена *mas1'*, *A. macedoi* содержал остатки гена *ags*, а *A. pusilla* – остатки как генов *mas2'*, так и *ags* (Bogomaz et al., 2024). Все эти гены кодируют ферменты, относящиеся к одному и тому же биосинтетическому пути. Они катализируют цепочку реакций, приводящих к синтезу агропина, и обнаруживаются сгруппированными вместе в одной и той же агробактериальной Т-ДНК (Ellis et al., 1984). Общий предок видов *Arachis*, вероятно, был трансформирован штаммом, содержащим все три гена, но в какой-то момент эти гены перестали обеспечивать селективные преимущества своим хозяевам, накопили мутации и были утрачены, оставшись в виде отдельных фрагментов у представителей разных клад.

Совершенно иная картина характерна для гомолога гена кукумопинсинтазы. Он сохранился интактным у большинства изученных видов. У культурного арахиса, являющегося тетраплоидом с геномами А и В, *cus*-подобный ген присутствует в обоих геномах, но в геноме А он интактный, а в геноме В – мутантный (Matveeva, Otten, 2019). Более детальные исследования показали, что среди 29 описанных аллелей гена из генома А только три содержали мутации, несовместимые с функцией (Bogomaz et al., 2024). Культурный арахис делится на два подвида: *hypogaea* и *fastigiata* (Krapovickas, Gregory, 2007; Bertioli et al., 2011). Наиболее распространенная аллель А *cus*-подобного гена обнаружена у представителей обоих подвигов, а также у *A. duranensis*, что подтверждает их близкое родство. Однако у вида *A. monticola* эта аллель пока не выявлена, а его прочие аллели были равномерно распределены по отдельным субкладам внутри клады, содержащей аллели А генома культурного арахиса и родственных ему видов (Bogomaz et al., 2024), подтверждая тесное родство *A. monticola* и *A. hypogaea*, описанное и другими авторами (Tian et al., 2021). На основе данных, полученных с использованием *cus*-подобного гена как филогенетического маркера, нельзя исключать, что генетический материал *A. paraguariensis* Chodat & Hassl. был вовлечен в формирование по крайней мере некоторых разновидностей *A. hypogaea*. Данные других ученых подтверждают тесную связь между *A. paraguariensis* и *A. duranensis* (предковый вид культурного арахиса) (Moretzsohn et al., 2013).

Вместе с тем филогенетические исследования арахиса на основе *cus*-подобного гена показали, что наиболее четкая клада на филогенетическом древе представлена

мутантными аллелями из генома В. Эти последовательности быстрее дивергируют и являются более удобным материалом для изучения филогении, чем те, в отношении которых действует стабилизирующий отбор.

Таким образом, на примере рода *Arachis* были продемонстрированы некоторые ограничения использования клТ-ДНК в филогенетических исследованиях.

Филогенетические связи генов Т-ДНК агробактерий и пГМО

В предыдущих разделах мы рассмотрели вопросы, связанные с изучением филогении растений с применением Т-ДНК как молекулярного маркера. Однако подобный маркер может быть использован и для прослеживания родства между клТ-ДНК пГМО и Т-ДНК ныне известных штаммов агробактерий (Suzuki et al., 2002; Matveeva, Otten, 2021). В одной из работ нашей группы (Matveeva, Otten, 2021) были построены филогенетические деревья на основе отдельных генов синтеза опинов из всех известных к 2021 г. пГМО, а также охарактеризованных на тот момент штаммов ризобий. В результате было показано, что у параспонии *Parasponia andersonii* Planch. клТ-ДНК, содержащие гомологи гена *susL*, получены от различных штаммов агробактерий на разных этапах эволюции. Аналогичная картина была характерна для хурмы *Diospyros lotus* L. В то же время клТ-ДНК, содержащие гомологи микмопинсинтазы, скорее всего, были получены разными видами табака *Nicotiana* L. и льнянками *Linaria* Mill. от одного или очень близкородственных штаммов (Matveeva, Otten, 2021). Эти наблюдения позволяют лучше понять биоразнообразие агробактерий и подойти ближе к пониманию механизмов их хозяйской специфичности (Anderson, Moore, 1979), которая часто обусловлена особенностями структуры и функционирования *vir*-генов, наследуемых сцепленно с Т-ДНК в составе Ti(Ri)-плазмид. Изучение вопросов хозяйской специфичности важно для оптимизации протоколов генетической трансформации растений.

Закключение

Таким образом, клТ-ДНК помогают уточнить многие спорные моменты в филогенетических исследованиях. Вставки различных Т-ДНК маркируют эволюционные события, указывая нам на группы видов, имеющих общего предка. Инвертированные повторы могут дать информацию о возрасте этого предка и выстроить порядок попадания независимых Т-ДНК в геномы растений. Изучение тонкого полиморфизма клТ-ДНК с учетом аллельных состояний маркера позволяет отслеживать микроэволюционные события, последствия актов гибридизации при незавершенном видообразовании. К настоящему времени молекулярные маркеры на основе клТ-ДНК были успешно использованы при изучении родов *Nicotiana*, *Camellia*, *Vaccinium* и *Arachis*. В случае рода *Nicotiana* с помощью клТ-ДНК удалось маркировать и датировать основные этапы эволюции в пределах секции *Tomentosae*. У *Camellia* главным итогом использования маркера была четкая демонстрация незавершенного видообразования в пределах секции *Thea*, у представителей рода *Vaccinium* маркер позволил подтвердить некоторые классические представления о системе рода, которые шли вразрез с данными

на основе ITS, но согласовывались с данными NGS на небольшой выборке видов. У арахиса изучение клТ-ДНК позволило четко продемонстрировать различные эволюционные судьбы трансгенов в случае стабилизирующего отбора и без него, а также наметить некоторые недостатки разрешающей способности маркера, находящегося под сильным давлением стабилизирующего отбора.

Список литературы / References

- Матвеева Т.В. Зачем растениям агробактериальные гены? *Экологическая генетика*. 2021;19(4):365-375. doi 10.17816/ecogen89905 [Matveeva T.V. Why do plants need agrobacterial genes? *Ecol Genet*. 2021;19(4):365-375. doi 10.17816/ecogen89905]
- Матвеева Т.В., Павлова О.А., Богомаз Д.И., Демкович А.Е., Лутова Л.А. Молекулярные маркеры для видоидентификации и филогенетики растений. *Экологическая генетика*. 2011;9(1):32-43. doi 10.17816/ecogen9132-43 [Matveeva T.V., Pavlova O.A., Bogomaz D.I., Demkovich A.E., Lutova L.A. Molecular markers for plant species identification and phylogenetics. *Ecol Genet*. 2011;9(1):32-43. doi 10.17816/ecogen9132-43 (in Russian)]
- Хафизова Г.В., Матвеева Т.В. Полиморфизм последовательностей ДНК агробактериального происхождения в сортах *Nicotiana tabacum*. *Генетика*. 2020;56(10):1215-1217. doi 10.31857/S0016675820100057 [Khafizova G.V., Matveeva T.V. Polymorphism in sequences of agrobacterial origin in *Nicotiana tabacum* cultivars. *Russ J Genet*. 2020;56(10):1215-1217. doi 10.1134/S1022795420100051]
- Anderson A., Moore L. Host specificity in the genus *Agrobacterium*. *Phytopathology*. 1979;69(4):320-323. doi 10.1094/Phyto-69-320
- Bahramnejad B., Naji M., Bose R., Jha S. A critical review on use of *Agrobacterium rhizogenes* and their associated binary vectors for plant transformation. *Biotechnol Adv*. 2019;37(7):107405. doi 10.1016/j.biotechadv.2019.06.004
- Becker A.L., Crowl A.A., Luteyn J.L., Chanderbali A.S., Judd W.S., Manos P.S., Soltis D.E., Smith S.A., Goncalves D.J.P., Dick C.W., Weaver W.N., Soltis P.S., Cellinese N., Fritsch P.W. A global blueberry phylogeny: evolution, diversification, and biogeography of Vacciniaceae (Ericaceae). *Mol Phylogenet Evol*. 2024;201:108202. doi 10.1016/j.ympev.2024.108202
- Bertioli D.J., Seijo G., Freitas F.O., Valls J.F.M., Leal-Bertioli S.C.M., Moretzsohn M.C. An overview of peanut and its wild relatives. *Plant Genet Resour*. 2011;9(1):134-149. doi 10.1017/S1479262110000444
- Bogomaz O.D., Bemova V.D., Mirgorodskii N.A., Matveeva T.V. Evolutionary fate of the opine synthesis genes in the *Arachis* L. genomes. *Biology (Basel)*. 2024;13(8):601. doi 10.3390/biology13080601
- Brandle J., Bai D. Biotechnology: uses and applications in tobacco improvement. In: Davis D.L., Nielsen M.T. (Eds) Tobacco: Production, Chemistry and Technology. Wiley, 1999;49-65
- Browning S.R., Browning B.L. Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 2011;12(10):703-714. doi 10.1038/nrg3054
- Camp W.H. The North American blueberries with notes on other groups of Vacciniaceae. *Brittonia*. 1945;5(3):203-275. doi 10.2307/2804880
- Camp W.H., Gilly C.L. The structure and origin of species. *Brittonia*. 1943;4(3):323-385. doi 10.2307/2804896
- Carr I.M., Robinson J.I., Dimitriou R., Markham A.F., Morgan A.W., Bonthron D.T. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics*. 2009;25(24):3244-3250. doi 10.1093/bioinformatics/btp583
- Chen K., Dorlhac de Borne F., Szegedi E., Otten L. Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. *Plant J*. 2014;80(4):669-682. doi 10.1111/tbj.12661
- Chen K., Dorlhac de Borne F., Sierro N., Ivanov N.V., Alouia M., Koechler S., Otten L. Organization of the TC and TE cellular T-DNA

- regions in *Nicotiana otophora* and functional analysis of three diverged TE-6b genes. *Plant J.* 2018;94(2):274-287. doi 10.1111/tpj.13853
- Chen K., Zhurbenko P., Danilov L., Matveeva T., Otten L. Conservation of an *Agrobacterium* cT-DNA insert in *Camellia* section *Thea* reveals the ancient origin of tea plants from a genetically modified ancestor. *Front Plant Sci.* 2022;13:997762. doi 10.3389/fpls.2022.997762
- Chen K., Liu H., Blevins T., Hao J., Otten L. Extensive natural *Agrobacterium*-induced transformation in the genus *Camellia*. *Planta.* 2023;258(4):81. doi 10.1007/s00425-023-04234-9
- Christey M.C. Use of Ri-mediated transformation for production of transgenic plants. *In Vitro Cell Dev Biol Plant.* 2001;37:687-700. doi 10.1007/s11627-001-0120-0
- Dehairs J., Talebi A., Cherifi Y., Swinnen J.V. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep.* 2016;6: 28973. doi 10.1038/srep28973
- Diaz-Garcia L., Rodriguez-Bonilla L., Rohde J., Smith T., Zalapa J. Pacbio sequencing reveals identical organelle genomes between American cranberry (*Vaccinium macrocarpon* Ait.) and a wild relative. *Genes (Basel).* 2019;10(4):291. doi 10.3390/genes10040291
- Doronina L., Feigin C.Y., Schmitz J. Reunion of Australasian possums by shared SINE insertions. *Syst Biol.* 2022;71(5):1045-1053. doi 10.1093/sysbio/syac025
- Ellis J.G., Ryder M.M., Tate M.E. *Agrobacterium tumefaciens* T_R-DNA encodes a pathway for agropine biosynthesis. *Mol Gen Genet.* 1984; 195:466-473. doi 10.1007/BF00341448
- Fricano A., Bakaher N., Del Corvo M., Piffanelli P., Donini P., Stella A., Ivanov N.V., Pozzi C. Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. *BMC Genet.* 2012;13:18. doi 10.1186/1471-2156-13-18
- Furner I.J., Huffman G.A., Amasino R.M. An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nature.* 1986;319: 422-427. doi 10.1038/319422a0
- Gaut B.S., Mortont B.R., McCaig B.C., Clegg M.T. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL* (orthology/nucleotide substitution rates/molecular clock). *Proc Natl Acad Sci USA.* 1996;93(19):10274-10279. doi 10.1073/pnas.93.19.10274
- Gelvin S.B. Plant DNA repair and *Agrobacterium* T-DNA integration. *Int J Mol Sci.* 2021;22(16):8458. doi 10.3390/ijms22168458
- Hancock J.F., Lyrene P., Finn C.E., Vorsa N., Lobos G.A. Blueberries and cranberries. In: Hancock J.F. (Ed.) *Temperate Fruit Crop Breeding*. Springer, 2008;115-150. doi 10.1007/978-1-4020-6907-9_4
- Hansen G., Larribe M., Vaubert D., Tempe J., Biermann B.J., Montoya A.L., Chilton M.-D., Brevet J. *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc Natl Acad Sci USA.* 1991;88(17):7763-7767. doi 10.1073/pnas.88.17.7763
- Haubold B., Wiehe T. Statistics of divergence times. *Mol Biol Evol.* 2001; 18(7):1157-1160. doi 10.1093/oxfordjournals.molbev.a003902
- Intrieri M.C., Buiatti M. The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*. *Mol Phylogenet Evol.* 2001;20(1):100-110. doi 10.1006/mpev.2001.0927
- Jouanin L. Restriction map of an agropine-type Ri plasmid and its homologies with Ti plasmids. *Plasmid.* 1984;12(2):91-102. doi 10.1016/0147-619x(84)90055-6
- Kawash J., Colt K., Hartwick N.T., Abramson B.W., Vorsa N., Polashock J.J., Michael T.P. Contrasting a reference cranberry genome to a crop wild relative provides insights into adaptation, domestication, and breeding. *PLoS One.* 2022;17(3):e0264966. doi 10.1371/journal.pone.0264966
- Khafizova G.V., Sierro N., Ivanov N.V., Sokornova S.V., Polev D.E., Matveeva T.V. *Nicotiana noctiflora* Hook. genome contains two cellular T-DNAs with functional genes. *Plants (Basel).* 2023;12(22): 3787. doi 10.3390/plants12223787
- Knapp S., Chase M.W., Clarkso J.J. Nomenclature changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon.* 2004; 53(1):73-82. doi 10.2307/4135490
- Koppolu R., Upadhyaya H.D., Dwivedi S.L., Hoisington D.A., Varshney R.K. Genetic relationships among seven sections of genus *Arachis* studied by using SSR markers. *BMC Plant Biol.* 2010; 10:15. doi 10.1186/1471-2229-10-15
- Krapovickas A., Gregory W.C. Taxonomy of the genus *Arachis* (Leguminosae). *Bonplandia.* 2007;16(Supl.):1-205. doi 10.30972/bon.160158
- Kron K., Powell E., Luteyn J. Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from MATK and nuclear ribosomal ITS regions, with comments on the placement of *Satyria*. *Am J Bot.* 2002;89(2):327-336. doi 10.3732/ajb.89.2.327
- Kuritzin A., Kischka T., Schmitz J., Churakov G. Incomplete lineage sorting and hybridization statistics for large-scale retroposon insertion data. *PLoS Comput Biol.* 2016;12(3):e1004812. doi 10.1371/journal.pcbi.1004812
- Leisner C.P., Kamileen M.O., Conway M.E., O'Connor S.E., Buell C.R. Differential iridoid production as revealed by a diversity panel of 84 cultivated and wild blueberry species. *PLoS One.* 2017; 12(6):e0179417. doi 10.1371/journal.pone.0179417
- Leitch I.J., Hanson L., Lim K.Y., Kovarik A., Chase M.W., Clarkson J.J., Leitch A.R. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann Bot.* 2008;101(6): 805-814. doi 10.1093/aob/mcm326
- Lewis R.S., Nicholson J.S. Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. *Genet Resour Crop Evol.* 2007;54:727-740. doi 10.1007/s10722-006-0024-2
- Lynch M., Conery J.S. The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290(5494):1151-1155. doi 10.1126/science.290.5494.1151
- Martin M., Patterson M., Garg S., Fischer S., Pisanti N., Klau G.W., Schöenhuth A., Marschall T. WhatsHap: fast and accurate read-based phasing. *bioRxiv.* 2016. doi 10.1101/085050
- Matveeva T. New naturally transgenic plants: 2020 update. *Bio Comm.* 2021;66(1):36-46. doi 10.21638/SPBU03.2021.105
- Matveeva T.V., Otten L. Widespread occurrence of natural genetic transformation of plants by *Agrobacterium*. *Plant Mol Biol.* 2019; 101(4-5):415-437. doi 10.1007/s11103-019-00913-y
- Matveeva T.V., Otten L. Opine biosynthesis in naturally transgenic plants: genes and products. *Phytochemistry.* 2021;189:112813. doi 10.1016/j.phytochem.2021.112813
- Meyer A.D., Ichikawa T., Meins F. Horizontal gene transfer: regulated expression of a tobacco homologue of the *Agrobacterium rhizogenes* *rolC* gene. *Mol Gen Genet.* 1995;249(3):265-273. doi 10.1007/BF00290526
- Min T.L., Bartholomew B. Theaceae. In: Wu Z.Y., Raven P.H. (Eds) *Flora of China*. Vol. 12. St Louis: Science Press, Beijing & Missouri Botanical Garden Press, 2007;323-450
- Mohajjel-Shoja H., Clément B., Perot J., Alioua M., Otten L. Biological activity of the *Agrobacterium rhizogenes*-derived *trnK* gene of *Nicotiana tabacum* and its functional relation to other *trnK* genes. *Mol Plant Microbe Interact.* 2011;24(1):44-53. doi 10.1094/MPMI-06-10-0139
- Moon H.S., Nifong J.M., Nicholson J.S., Heineman A., Lion K., Van der Hoeven R., Lewis R.S. Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources. *Crop Sci.* 2009;49(6): 2149-2159. doi 10.2135/cropsci2009.01.0024
- Moretzsohn M.C., Gouvea E.G., Inglis P.W., Leal-Bertioli S.C.M., Valls J.F.M., Bertioli D.J. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann Bot.* 2013; 111(1):113-126. doi 10.1093/aob/mcs237
- Nester E.W. *Agrobacterium*: nature's genetic engineer. *Front Plant Sci.* 2015;5:730. doi 10.3389/fpls.2014.00730

- Otten L. Natural *Agrobacterium*-mediated transformation in the genus *Nicotiana*. In: Ivanov N.V., Sierro N., Peitsch M.C. (Eds) The Tobacco Plant Genome. Compendium of Plant Genomes. Cham: Springer, 2020;195-209. doi 10.1007/978-3-030-29493-9_12
- Powell E., Kron K. Molecular systematics of the northern Andean blueberries (Vaccinieae, Vaccinioideae, Ericaceae). *Int J Plant Sci.* 2003;164(6):987-995. doi 10.1086/378653
- Ren N., Timko M.P. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. *Genome.* 2001;44(4):559-571. doi 10.1139/g01-060
- Rossi L., Bindler G., Pijnenburg H., Isaac P.G., Giraud-Henry I., Mahe M., Gadani F. Potential of molecular marker analysis for variety identification in processed tobacco. *Plant Varieties Seeds.* 2001;14(2):89-102
- Schell J., Van Montagu M. The Ti-plasmid of *Agrobacterium tumefaciens*, a natural vector for the introduction of *nif* genes in plants? In: Genetic Engineering for Nitrogen Fixation. Basic Life Sciences. Vol. 9. Springer, 1977;159-179. doi 10.1007/978-1-4684-0880-5_12
- Schlautman B., Covarrubias-Pazarán G., Fajardo D., Steffan S., Zalapa J. Discriminating power of microsatellites in cranberry organelles for taxonomic studies in *Vaccinium* and Ericaceae. *Genet Resour Crop Evol.* 2017;64(3):451-466. doi 10.1007/s10722-016-0371-6
- Schrinner S.D., Mari R.S., Ebler J., Rautiainen M., Seillier L., Reimer J.J., Usadel B., Marschall T., Klau G.W. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biol.* 2020; 21(1):252. doi 10.1186/s13059-020-02158-1
- Shedlock A.M., Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays.* 2000;22(2):148-160. doi 10.1002/(SICI)1521-1878(200002)22:2<148::AID-BIES6>3.0.CO;2-Z
- Sierro N., Battey J.N., Ouadi S., Bakaher N., Bovet L., Willig A., Goepfert S., Peitsch M.C., Ivanov N.V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun.* 2014;5(1):3833. doi 10.1038/ncomms4833
- Singer K. The mechanism of T-DNA integration: some major unresolved questions. *Curr Top Microbiol Immunol.* 2018;418:287-317. doi 10.1007/82_2018_98
- Snyder M., Adey A., Kitzman J.O., Shendure J. Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet.* 2015;16(6):344-358. doi 10.1038/nrg3903
- Soltis D.E., Mavrodiev E.V., Doyle J.J., Rauscher J., Soltis P.S. ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Syst Bot.* 2008;33(1):7-20. doi 10.1600/036364408783887401
- Stalker H.T. Utilizing wild species for peanut improvement. *Crop Sci.* 2017;57(3):1102-1120. doi 10.2135/cropsci2016.09.0824
- Sultana N., Menzel G., Heikam T., Kojima K., Bao W., Serçe S. Bioinformatic and molecular analysis of satellite repeat diversity in *Vaccinium* genomes. *Genes (Basel).* 2020;11(5):527. doi 10.3390/genes11050527
- Sun X., Jiao C., Schwaninger H., Chao C.T., Ma Y., Duan N., Khan A., Ban S., Xu K., Cheng L., Zhong G., Fei Z. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet.* 2020;52(12):1423-1432. doi 10.1038/s41588-020-00723-9
- Suzuki K., Yamashita I., Tanaka N. Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant J.* 2002;32(5):775-787. doi 10.1046/j.1365-3113X.2002.01468.x
- Tanaka H., Hori T., Yamamoto S., Toyoda A., Yano K., Yamane K., Itoh T. Haplotype-resolved chromosomal-level assembly of wasabi (*Eutrema japonicum*) genome. *Sci Data.* 2023;10(1):441. doi 10.1038/s41597-023-02356-z
- Tepfer D. Genetic transformation using *Agrobacterium rhizogenes*. *Physiol Plant.* 1990;79(1):140-146. doi 10.1111/j.1399-3054.1990.tb05876.x
- Tian X., Shi L., Guo J., Fu L., Du P., Huang B., Wu Y., Zhang X., Wang Z. Chloroplast phylogenomic analyses reveal a maternal hybridization event leading to the formation of cultivated peanuts. *Front Plant Sci.* 2021;12:804568. doi 10.3389/fpls.2021.804568
- Tiley G.P., Crowl A.A., Manos P.S., Sessa E.B., Solís-Lemus C., Yoder A.D., Burleigh J.G., Smith S. Benefits and limits of phasing alleles for network inference of allopolyploid complexes. *Syst Biol.* 2024;73(4):666-682. doi 10.1093/sysbio/syae024
- Tong Z., Fang D., Chen X., Jiao F., Zhang Y., Li Y., Xiao B. Genome-wide association study of leaf chemistry traits in tobacco. *Breed Sci.* 2020;70(3):253-264. doi 10.1270/jsbbs.19067
- Tzfira T., Li J., Lacroix B., Citovsky V. *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet.* 2004;20(8):375-383. doi 10.1016/j.tig.2004.06.004
- Vorsa N., Zalapa J. Domestication, genetics, and genomics of the American cranberry. In: Goldman I. (Ed.) Plant Breeding Reviews. Vol. 43. John Wiley & Sons, Inc., 2019;279-315. doi 10.1002/9781119616801.ch8
- Wang H., Guo X., Hu X., Li T., Fu X., Liu R. Comparison of phytochemical profiles, antioxidant and cellular antioxidant activities of different varieties of blueberry (*Vaccinium* spp.). *Food Chem.* 2017; 217:773-781. doi 10.1016/j.foodchem.2016.09.002
- Wang Y., Zhao Y., Bollas A., Wang Y., Au K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021; 39(11):1348-1365. doi 10.1038/s41587-021-01108-x
- Wenger A.M., Peluso P., Rowell W.J., Chang P., Hall R.J., Concepcion G.T., Ebler J., ... Li H., Koren S., Carroll A., Rank D.R., Hunkapiller M.W. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol.* 2019;37(10):1155-1162. doi 10.1038/s41587-019-0217-9
- White F.F., Ghidossi G., Gordon M.P., Nester E.W. Tumor induction by *Agrobacterium rhizogenes* involves the transfer of plasmid DNA to the plant genome. *Proc Natl Acad Sci USA.* 1982;79(10):3193-3197. doi 10.1073/pnas.79.10.3193
- White F.F., Garfinkel D.J., Huffman G.A., Gordon M.P., Nester E.W. Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature.* 1983;301:348-350. doi 10.1038/301348a0
- White F., Taylor B., Huffman G., Gordon M., Nester E. Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of *Agrobacterium rhizogenes*. *J Bacteriol.* 1985;164(1): 33-44. doi 10.1128/jb.164.1.33-44.1985
- Wu Q., Tong W., Zhao H., Ge R., Li R., Huang J., Li F., Wang Y., Mallano A.I., Deng W., Wang W., Wan X., Zhang Z., Xia E. Comparative transcriptomic analysis unveils the deep phylogeny and secondary metabolite evolution of 116 *Camellia* plants. *Plant J.* 2022;111(2): 406-421. doi 10.1111/tpj.15799
- Xie X., Ma X., Liu Y.G. Decoding Sanger sequencing chromatograms from CRISPR-induced mutations. *Methods Mol Biol.* 2019;1917: 33-43. doi 10.1007/978-1-4939-8991-1_3
- Yukawa M., Tsudzuki T., Sugiura M. The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol Genet Genom.* 2006; 275(4):367-373. doi 10.1007/s00438-005-0092-6
- Zalapa J.E., Bougie T.C., Bougie T.A., Schlautman B.J., Wiesman E., Guzman A., Fajardo D.A., Steffan S., Smith T. Clonal diversity and genetic differentiation revealed by SSR markers in wild *Vaccinium macrocarpon* and *Vaccinium oxycoccos*. *Ann Appl Biol.* 2015; 166(2):196-207. doi 10.1111/aab.12173
- Zhang X., Wu R., Wang Y., Yu J., Tang H. Unzipping haplotypes in diploid and polyploid genomes. *Comput Struct Biotechnol J.* 2020; 18:66-72. doi 10.1016/j.csbj.2019.11.011
- Zhidkin R., Zhurbenko P., Bogomaz O., Gorodilova E., Katsapov I., Antropov D., Matveeva T. Biodiversity of *rolB/C*-like natural transgene in the genus *Vaccinium* L. and its application for phylogenetic studies. *Int J Mol Sci.* 2023;24(8):6932. doi 10.3390/ijms24086932