

УДК 575.852.113:575.858:57.082.13

## ДНК-ШТРИХКОДЫ РАСТЕНИЙ

© 2018 г. В. С. Шнеер<sup>1,\*</sup>, А. В. Родионов<sup>1,2,\*\*</sup>

<sup>1</sup>Ботанический институт им. В.Л. Комарова РАН, Санкт-Петербург, Россия

<sup>2</sup>Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

E-mail: \*shneyer@rambler.ru, \*\*avrodionov@mail.ru

Поступила в редакцию 30.03.2018 г.

Поиск универсального ДНК-штрихкода растений оказался очень сложной задачей. Одного участка (или комбинации из нескольких), по которому можно было бы определить конкретный вид, не найдено. Хотя обычно комбинация, рекомендованная в 2009 г. как стандарт (*rbcL* + *matK*+ITS), позволяет отнести вид к соответствующему роду. Вариабельность отдельных маркеров не одинакова в разных таксономических группах, что, как правило, позволяет подобрать ДНК-штрихкод (иногда даже укороченный мини-штрихкод) для конкретной группы, особенно в случае задач прикладного характера. Методы секвенирования нового поколения (NGS) дают перспективу массового получения протяженных ДНК-штрихкодов (последовательностей полных хлоропластных геномов и рибосомных генов), позволяющих преодолеть ограничения стандартных ДНК-штрихкодов. Очень важно, что NGS существенно повышает возможность использования гербарных образцов. Поиски ДНК-штрихкодов для растений продолжаются.

**Ключевые слова:** ДНК-штрихкодирование растений, детерминация видов, хлоропластная ДНК, секвенирование нового поколения.

**DOI:**

Потребности в безошибочной идентификации с точностью до вида или рода образцов живых растений в природе или видовой принадлежности продуктов растительного происхождения в науке и на практике не удовлетворяются ввиду нехватки опытных экспертов. Поэтому области применения ДНК-штрихкодирования в ботанике и экспертной службе весьма разнообразны. В науке это идентификация трудно определяемых видов (например, коротко- и редкоцветущих, особенно древесных), более точное определение видового состава локальных флор и растительных сообществ, выявление новых, в том числе криптических, видов. Практическое использование ДНК-штрихкодирования многообразно и продолжает расширяться по мере достигаемых успехов: охрана биоразнообразия и редких видов, в том числе пресечение их сбора и незаконной торговли, проверка растительного сырья, травяных сборов, меда и других коммерческих продуктов, контроль за сорняками, инвазионными видами, растениями-источниками аллергенов и др.

Между тем с самого начала реализации международной программы CBOL (Consortium for the

Barcodes of Life) на пути ДНК-штрихкодирования растений встретилось много трудностей. Так как стандартный ДНК-штрихкод большинства животных – фрагмент митохондриального гена *COI* – оказался неприменим для растений ввиду низкой и неравномерной скорости мутирования у них митохондриальной ДНК, была поставлена задача выбрать штрихкод из хлоропластных участков генома. Была организована специальная группа по поиску ДНК-штрихкода растений (Plant Working Group CBOL). Поиск велся среди нескольких перспективных хлоропластных последовательностей-кандидатов – кодирующих: *matK* (матураза), *rpoB* ( $\beta$ -субъединица РНК полимеразы), *rpoC1* ( $\beta^1$ -субъединица РНК полимеразы), *accD* ( $\beta^1$ -субъединица ацетил-КоА-карбоксилазы), *ndhJ* (субъединица никотинамиддегидрогеназы), *ccsA* (кодирующий белок, участвующий в биосинтезе цитохрома *c*), *rbcL* (большая субъединица фермента 4.1.1.39 рибулозодифосфат-карбоксилаза) и некодирующих: спейсеры *trnH-psbA*, *atpF-atpH*, *psbK-psbI*, позже к ним были добавлены участки внутренних транскрибируемых спейсеров ITS ядерных рибосомных генов. Эти участки тестировались на

разных группах растений, проверялось, насколько хорошо они амплифицируются с универсальными праймерами и как хорошо дифференцируют виды. Довольно скоро стало ясно, что нет одного участка, который мог бы служить штрихкодом для всех растений, нужна была комбинация. Начальные этапы поисков и особенности участков, позже выбранных в качестве ДНК-штрихкодов, были достаточно подробно описаны нами ранее (Шнеер, 2009). Стоит добавить, что поиски заняли гораздо больше времени, чем поначалу планировалось, и вряд ли можно сказать, что они завершены.

В 2009 г. Исполнительный Комитет CVOЛ предложил в качестве ДНК-штрихкодов растений комбинацию двух хлоропластных участков: фрагмент гена *rbcL* длиной около 600 нуклеотидов и фрагмент гена *matK* длиной около 800 нуклеотидов (Hollingsworth et al., 2009). В некоторых работах сообщалось о том, что эта комбинация хорошо разрешает виды (Burgess et al., 2011), но во многих других разрешение было явно недостаточным, например, в родах *Crocus*, *Berberis*, *Primula* (Seberg, Petersen, 2009; Roy et al., 2010; Yan et al., 2011). Специальное сравнительное исследование провели китайские ученые, которые секвенировали 5 разных последовательностей — потенциальных ДНК-штрихкодов 6286 образцов растений, представляющих 1757 видов 141 рода. Комбинация маркеров *matK* + *rbcL* была видоспецифична в 49.7% случаев (Li et al., 2011). При этом в некоторых работах разрешение в родах значительно повышалось при добавлении или участка ITS, например, у пальмовых (Jeanson et al., 2011), или спейсера *trnH-psbA*, например, у губоцветных (De Mattia et al., 2011). Некоторые авторы еще раньше выступали за то, чтобы разносторонне исследованный участок ITS вошел в число ДНК-штрихкодов (Feliner, Rosselló, 2007), а ряд китайских авторов, по результатам анализа данных по ITS для тысяч видов из разных семейств, нашли, что ITS2 является весьма эффективным ДНК-штрихкодом (Ren et al., 2010; Gao et al., 2010). В результате в 2011 г. было предложено добавить к стандартным ДНК-штрихкодам хлоропластный спейсер *trnH-psbA* (его долго именовали как альтернативный ДНК-штрихкод) и ядерные ITS1 и ITS2 и использовать их в случае необходимости (Hollingsworth, 2011; Li et al., 2011). При этом было ясно, что у этих маркеров есть определенные ограничения, замечалось, что, возможно, хорошие результаты могли бы быть получены с высоковариабельными низкокопийными ядерными генами, но из-за их неодинаковой у разных таксонов вариабельности и ряда других проблем (Шнеер, 2009) выбрать какие-то из них в качестве ДНК-штрихкодов было невозможно.

В последующие годы были проведены многочисленные исследования, где использовался весь или почти весь рекомендованный набор, изредка к ним добавляли другие маркеры, по большей части хлоропластные. Объекты выбирались как по систематическому принципу (виды одного или близких родов, семейства), так и по флористическому (виды какой-либо территории, заповедника). Было показано, что ген *rbcL*, хотя и амплифицируется, как правило, лучше остальных, обладает низким разрешением и далеко не всегда увеличивает видовое разрешение в комбинации с другими маркерами, хотя обычно позволяет отнести вид к соответствующему роду. Ген *matK* для амплификации нуждается в большем количестве праймеров (особенные проблемы возникают с мхами и папоротниками), но, как правило, обладает более высоким разрешением. Спейсер *trnH-psbA* демонстрировал много инделей и инверсий, его длина могла существенно варьировать даже у близких видов, что весьма затрудняло выравнивание, так было, например, при исследовании рода *Gentiana* (Liu et al., 2016). Но в тех случаях, когда таких изменений не было или было мало, он (или комбинация *trnH-psbA* + ITS) иногда давал лучшее разрешение, чем *rbcL* и/или *matK* (Costion et al., 2011; Parmentier et al., 2013; Christina, Annamalai, 2014; Huan et al., 2018). Впрочем, в ряде других исследований *trnH-psbA* лишь незначительно увеличивал разрешение, полученное с локусами *rbcL* + *matK* (Burgess et al., 2011).

По мере того, как в GenBank появлялось все больше полностью отсеквенированных хлоропластных геномов, активизировались поиски участков, более вариабельных и лучше дифференцирующих виды, чем стандартные ДНК-штрихкоды. Так, появилось сообщение о том, что ген *yef1*, второй по величине в хлоропластном геноме, длиной почти 6 тыс. нуклеотидов, показал лучшее разрешение, чем комбинация *rbcL* + *matK* + *trnH-psbA*, в выборке из 420 видов из 67 семейств, включавшей мхи, голо- и покрытосеменные (Dong et al., 2015). Хорошие результаты были получены с этим геном и при анализе видов рода *Paris* (Song et al., 2017). Однако при анализе видов дубов из Китая (Yang et al., 2017) и маревых из рода *Kalidium* (Liang et al., 2017) этот ген не обнаружил высокой разрешающей способности.

Однако во многих работах констатировалось, что все примененные хлоропластные маркеры давали низкое разрешение видов (Roy et al., 2010; Arca et al., 2012; Federici et al., 2013; Alves et al., 2014; Wu et al., 2017). Наблюдалась их высокая внутривидовая и низкая межвидовая вариабельность, отсутствие штрихкодowego разрыва между видами. Полной

неудачей закончилась попытка ДНК-штрихкодирования 71 вида ив (*Salix*) при использовании до 7 хлоропластных маркеров: только для одного вида удалось получить уникальный штрихкод (Pegu et al., 2014). Наиболее вероятной гипотезой, объясняющей такую картину, эти авторы сочли «выметание отбором», когда при отборе полезной мутации вместе с ней отбираются сопутствующие аллели, а другие аллели удаляются, обедняя общий полиморфизм. Другое объяснение – низкий поток генов между популяциями (Petit, Excoffier, 2009; Federici et al., 2013), так как поток генов, обеспечиваемый лишь семенами (а органеллы передаются по большей части семенами), меньше, чем семенами и пылью. Рекомендовалось усилить поиск ядерных маркеров для достижения большего успеха в разрешении видов при ДНК-штрихкодировании (Naciri et al., 2012).

Действительно, участок ядерного ITS в ряде работ давал лучшее разрешение, чем хлоропластные маркеры (Vivas et al., 2014; Liu et al., 2016). При исследовании бразильских видов малоизученного семейства Sapotaceae, преимущественно тропических вечнозеленых деревьев и кустарников, для видовой идентификации которых нужны цветки, являющиеся короткоживущими и поэтому редко доступными, ITS обладал наибольшим разрешением, которое не увеличивалось при добавлении хлоропластных маркеров (Vivas et al., 2014). В то же время у аллополиплоидных видов иногда наблюдался внутригеномный полиморфизм ITS из-за непрошедшей концертной эволюции (Zarrei et al., 2015).

Во многих публикациях, особенно в работах прикладного характера, в частности, по анализу лекарственных растений, констатировалось, что ITS2 обладает лучшим разрешением, чем ITS1 и даже ITS1+2 (Chen et al., 2010; Han et al., 2012; Michel et al., 2016; Wang et al., 2016). Виды семейства зонтичных Apiaceae плохо различаются неспециалистами, но среди них есть лекарственные растения. Были разработаны видоспецифичные праймеры для ITS2-участка для тех массовых видов зонтичных, которые могут быть спутаны с лекарственными, и создана система для проверки соответствующего растительного материала (Kim et al., 2016). Тем не менее, это не является общим правилом. Участок ITS1+2, один и в комбинации с остальными использованными в работе ДНК-штрихкодами, превосходил по разрешению ITS2, одного и в комбинациях, у видов *Primula* (Yan et al., 2015) и у таксонов семейства Lauraceae (Liu et al., 2017). Участок ITS1 давал лучшее разрешение видов, чем ITS2 и ITS1+2, при исследовании мхов сложного комплекса *Racomitrium canescens* (Stech

et al., 2013) и группы бобовых подтрибы Cassiinae (Mishra et al., 2016). Таким образом, применимость участков даже с примерно равными скоростью и характером эволюции может варьировать в зависимости от задач.

Иногда в рамках одной работы разные группы растений давали наилучшее разрешение видов разными локусами. В работе, где исследовались виды двух семейств из одного региона, оказалось, что ITS лучше разрешает виды семейства злаков Poaceae, а *matK* – виды маревых Chenopodiaceae (Yao et al., 2017). Разные результаты были получены в исследованиях двух родов, *Ocimum* и *Thymus*, из одного подсемейства Nepetoideae семейства Lamiaceae. Виды обоих родов не разрешались локусами *rbcL* и *matK*, но участок *trnH-psbA* разрешал виды *Ocimum* и не разрешал – *Thymus*. В самом масштабном на настоящее время проекте по ДНК-штрихкодированию растений «ДНК-штрихкодирование сосудистых растений Канады», включившем 96% от 5108 видов (Braukmann et al., 2017; Kuzmina et al., 2017), использовали три участка – *rbcL*, *matK* и ITS2. Все ДНК-штрихкоды верно относили виды к родам (91–98% успеха), но разрешение видов у них было разным – 81% у *matK*, 72% у ITS2 и лишь 44% – у *rbcL* (Braukmann et al., 2017). При этом разрешение видов из канадской Арктики было хуже, чем видов из более южных, лесных районов, хотя последние богаче по видовому составу.

Были и работы, где ни один из стандартных ДНК-штрихкодов или их комбинации не привели к успешному разрешению видов. Так было у некоторых мхов (Hassel et al., 2013), жузгуна *Calligonum* из бобовых (Li et al., 2014), боярышника *Crataegus* (Zarrei et al., 2015). Ограничения стандартных штрихкодов не раз были обсуждены (Hollingsworth et al., 2016; Coissac et al., 2016). Для многих прикладных целей стандартные штрихкоды обеспечивают вполне удовлетворительное разрешение, но определять любые виды, как изначально задумывалось, даже не всех растений, но крупных таксономических групп или региональных флор, они не могут.

Особенно низким бывало разрешение всех стандартных маркеров в эволюционно молодых группах, таких как макаронезийские виды рода *Lotus* (Ojeda et al., 2014). Попытка использовать для такой группы (гавайские виды из двух родов – *Clermontia* из семейства Campanulaceae и *Cyrtandra* из семейства Gesneriaceae) два низкокопийных ядерных гена к успеху не привела (Pillon et al., 2013). Из-за высокой внутривидовой изменчивости, удержания предковых аллелей и других факторов они разрешали виды еще хуже хлоропластных

участков. Несмотря на неоднократные попытки, на использование стандартных ДНК-штрихкодов и дополнительно еще нескольких хлоропластных участков, не удалось добиться разрешения видов рода *Araucaria* с островов Новой Каледонии – области, известной как горячая точка биоразнообразия (Escapa, Catalano, 2013; Gaudeul et al., 2014; Kranitz et al., 2014). Возникла идея использовать для этого полные хлоропластные последовательности. Еще несколько лет назад об этом нельзя было мечтать. Но новые технологии – секвенирование нового поколения NGS (next-generation sequencing) – дали такую возможность. Для 11 видов *Araucaria* из 13 произрастающих на острове, представленных двумя–тремя образцами, методами NGS были секвенированы полные хлоропластные геномы (Ruhsam et al., 2015). Дополнительно к этому были секвенированы 11 одно- и низкокопийных переменных ядерных генов, и проведено сравнение разрешения видов, получаемого с помощью хлоропластных геномов (около 147 тыс. нуклеотидов) и 11 ядерных генов (суммарно более 6 тыс. нуклеотидов). В первом случае оно было выше, чем во втором, хотя не полностью удовлетворило авторов, так как более половины видов не были монофилетичны (Ruhsam et al., 2015). Был сделан вывод, что для разрешения молодых, активно радирующих видов нужно использовать больше переменных ядерных генов. Тем не менее, использование полных хлоропластных геномов представлялось очень привлекательным.

Вскоре после появления методов NGS возникли предложения использовать получаемые последовательности для ДНК-штрихкодирования (Nock et al., 2011). Однако анализы с применением NGS весьма дороги. Разрабатывались более дешевые приемы, позволяющие получать эти последовательности (Stull et al., 2013). Методы NGS характеризуются глубиной покрытия (глубина секвенирования) – это показатель того, сколько раз прочитан нуклеотид в данном положении; чем он больше, тем выше вероятность избежать ошибок при сборке участка генома или целого генома. Был предложен подход, названный геномным скиммированием (неглубокое покрытие секвенирования). При этом путем моделирования подбирается нужная небольшая глубина покрытия и качество прочтений тотальной ДНК, достаточные для удовлетворительной сборки повторяющихся участков – рибосомных генов и геномов органелл, и именно она задается при секвенировании, дающим в результате примерно 1 Гб последовательностей (Straub et al., 2012). Такой подход удешевляет анализ, давая возможность анализировать тысячи образцов. Кроме того, метод позволяет исключить

редкие варианты и ошибки разного происхождения. Геномное скиммирование в этом уподобляют секвенированию по Сэнгеру, где редкие варианты не улавливаются (Bakker et al., 2016). Как мы показывали выше, часто у разных таксономических групп растений самыми переменными и, соответственно, дающими наилучшее разрешение, оказывались разные маркеры. Последовательность хлоропластного генома будет включать участки стандартных ДНК-штрихкодов, а кроме них, еще все остальные гены и спейсеры, из которых можно будет выбрать наиболее переменные для данной группы.

Для секвенирования низкокопийных ядерных генов предполагается использовать другой прием – таргетное или целевое обогащение, когда с помощью олигонуклеотидных зондов получают нужные последовательности генов из образцов геномной ДНК и затем их секвенируют методами NGS (Nicholls et al., 2015; Hollingsworth et al., 2016).

Так, по мере быстрого прогресса в технологиях секвенирования, возникли идеи использования для ДНК-штрихкодирования геномных данных нового формата, сформулирован так называемый двуединый подход, включающий: 1) накопление стандартных ДНК-штрихкодов и их библиотек; 2) активную разработку и применение расширенных штрихкодов с использованием геномного скиммирования (Coissac et al., 2016). Расширенный или протяженный штрихкод – хлоропластный геном + рибосомные гены, последовательности, получаемые с использованием NGS.

ДНК-штрихкодирование нередко не может обойтись без привлечения гербарных образцов, которых в мире накоплено десятки миллионов. Между тем часто из них могут быть выделены только короткие (до 250–400 нуклеотидов) фрагменты ДНК. Для получения ДНК удовлетворительной чистоты и количества приходится подбирать реагенты и условия (Särkinen et al., 2012). Было показано, что успешность выделения зависит не только от возраста образца (заметно понижается у образцов старше 50 лет), но и от того, как он был высушен, какими химикатами обрабатывались помещения от вредителей, а также от систематической принадлежности. Например, особо плохо выделялась ДНК из гербарных образцов семейства *Boaginiaceae* (Kuzmina et al., 2017). Для анализа по Сэнгеру материалов, в которых ДНК может быть сильно повреждена (старые гербарные образцы, обработанное растительное сырье, консервы), разрабатываются мини-штрихкоды – последовательности, более короткие, чем стандартные, например, менее 200 нуклеотидов.

Поиск таких участков в гене *rbcL* проводился *in silico*, и они позволяли определить образец до семейства (Little, 2014). Когда известны последовательности хлоропластных геномов, для решения прикладных задач, касающихся определенных видов, находятся частные решения в результате поиска наиболее переменных у этих видов участков. Так, было показано, что *Panax ginseng* (женьшень) можно достоверно отличить от близких видов по участку в 60 нуклеотидов гена *ycf1a*, в 100 нуклеотидов гена *ycf1b* или в 280 нуклеотидов гена *rps16* (Dong et al., 2014).

Важное преимущество методов NGS – возможность прочитать последовательности ядерных и хлоропластных маркеров сразу нескольких видов, что особенно существенно, если стоит задача раскрыть видовой состав растительных смесей, что актуально при оценке качества растительного сырья в фармакогнозии (Ivanova et al., 2016).

Так как методы NGS используют фрагментированную ДНК, была исследована возможность использовать для этого гербарные образцы, и результат был положительный (Bakker et al., 2016), то есть можно получать расширенные штрихкоды из старых образцов, среди которых особый интерес представляют типовые. Массовый анализ ДНК музейных, в том числе гербарных, образцов обозначают теперь специальным термином музеомика (Chomicki, Renner, 2015). Деградированная ДНК присутствует не только в образцах гербарных коллекций, но и, например, в травяных лекарственных сборах и другом растительном сырье. Большая эффективность методов NGS по сравнению с секвенированием по Сэнгеру при проверке аптечных сборов лекарственных растений уже была продемонстрирована (Ivanova et al., 2016).

На заре ДНК-штрихкодирования осознавалась и обсуждалась вероятность того, что близкие виды нельзя будет дифференцировать с помощью ДНК-штрихкодов (Шнеер, 2009). Но теперь, когда появилась возможность использовать в качестве ДНК-штрихкодов полные хлоропластные последовательности и участки рибосомных генов, секвенируемые с применением NGS, было предложено распространить ДНК-штрихкодирование и на более низкие таксономические уровни – подвиды и разновидности – и назвать это ультра-штрихкодированием (Kane et al., 2012).

Поиск оптимальных ДНК-штрихкодов растений продолжается.

Работа выполнена в соответствии с Программой фундаментальных исследований РАН 41 «Биоразнообразие природных систем и биологические

ресурсы России». Тема «Генетический полиморфизм родов и видов злаков и других дикорастущих покрытосеменных, выявляемый при сравнительном исследовании высокоизменчивых районов ядерного и хлоропластного геномов» (АААА-А18-118020190240-8)

## СПИСОК ЛИТЕРАТУРЫ

- Шнеер В.С. ДНК-штрихкодирование – новое направление в сравнительной геномике растений // Генетика. 2009. Т. 45. № 11. С. 1436–1448.
- Alves T.L., Chauveau O., Eggers L., De Souza-Chies T.T. Species discrimination in *Sisyrinchium* (Iridaceae): assessment of DNA barcodes in a taxonomically challenging genus // Mol. Ecol. Res. 2014. V. 14. P. 324–333.
- Arca M., Hinsinger D.D., Cruaud C. Deciduous trees and the application of universal DNA barcodes: a case study on the circumpolar *Fraxinus* // PLoS One. 2012. V. 7. P. e34089.
- Bakker F.T., Lei D., Yu J. et al. Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline // Bot. J. Linn. Soc. 2016. V. 117. P. 33–43.
- Braukmann T.W.A., Kuzmina M.L., Sills J. et al. Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada // PLoS One. 2017. V. 12. P. e0169515.
- Burgess K.S., Fazekas A.J., Kesanakurti P.R. et al. Discriminating plant species in a local temperate flora using the *rbcL*+*matK* DNA barcode // Meth. Ecol. Evol. 2011. V. 2. P. 333–340.
- Chen S., Yao H., Han J. et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species // PLoS ONE. 2010. V. 5. № 1. P. e8613.
- Chomicki G., Renner S.S. Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics // New Phytol. 2015. V. 205. P. 526–532.
- Christina V.L.P., Annamalai A. Nucleotide based validation of *Ocimum* species by evaluating three candidate barcodes of the chloroplast region // Mol. Ecol. Resour. 2014. V. 14. № 1. P. 60–68.
- Coissac E., Hollingsworth P.M., Lavergne S., Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding // Mol. Ecol. 2016. V. 25. P. 1423–1428.
- Costion C., Ford A., Cross H. et al. Plant DNA barcodes can accurately estimate species richness in poorly known flora // PLoS One. 2011. V. 6. № 11. P. e26841.
- De Mattia F., Bruni I., Galimbert A. et al. A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae // Food Res. Intern. 2011. V. 44. № 3. P. 693–702.
- Dong W., Liu H., Xu C. et al. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs // BMC Genet. 2014. V. 15. P. 138.

- Dong W., Xu C., Li C. et al. *ycf1*, the most promising plastid DNA barcode of land plants // *Sci/ Rep.* 2015. V. 5. № 8348.
- Escapa I.H., Catalano S.A. Phylogenetic analysis of Araucariaceae: integrating molecules, morphology, and fossils // *Int. J. Plant Sci.* 2013. V. 174. P. 1153–1170.
- Federici S., Galimberti A., Bartolucci F. et al. DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (Lamiaceae) // *Bot. J. Linn. Soc.* 2013. V. 171. P. 687–699.
- Feliner G.N., Rosselló J.A. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants // *Mol. Phylogenet. Evol.* 2007. V. 44. № 2. P. 911–919.
- Gao T., Yao H., Song J. et al. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family // *BMC Evol. Biol.* 2010. V. 10. P. e324.
- Gaudeul M., Gardner M.F., Thomas P. et al. Evolutionary dynamics of emblematic *Araucaria* species (Araucariaceae) in New Caledonia: nuclear and chloroplast markers suggest recent diversification, introgression, and a tight link between genetics and geography within species // *BMC Evol. Biol.* 2014. V. 14. P. 171.
- Han J.-P., Shi L.-C., Chen X.-C., Lin Y.-L. Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae // *J. Syst. Evol.* 2012. V. 50. № 3. P. 227–234.
- Hassel K., Segreto R., Ekrem T. Restricted variation in plant barcoding markers limits identification in closely related bryophyte species // *Mol. Ecol. Resour.* 2013. V. 13. № 6. P. 1047–1057.
- Hollingsworth P.M. Refining the DNA barcode for land plants // *PNAS USA.* 2011. V.108. P. 19451–19452.
- Hollingsworth P.M., Forrest L.L., Spouge J.L. et al. A DNA barcode for land plants // *PNAS USA.* 2009. V. 106. P. 12794–12797.
- Hollingsworth P.M., Li D.-Z., van der Bank M., Twyford A.D. Telling plant species apart with DNA: From barcodes to genomes // *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2016. V. 371. P. e20150338.
- Huan H.V., Trang H.M., Toan N.V. Identification of DNA barcode sequence and genetic relationship among some species of magnolia family // *As.J. Plant Sci.* 2018. V. 17. № 1. P. 56–64.
- Ivanova N.V., Kuzmina M.L., Braukmann T.W.A. et al. Authentication of herbal supplements using next-generation sequencing // *PLoS One.* 2016. V.11. № 5. P. e0156426.
- Jeanson M.L., Labat J.-N., Little D.P. DNA barcoding: a new tool for palm taxonomists? // *Ann. Bot.* 2011. V. 108. № 8. P. 1445–1451.
- Kane N., Sveinsson S., Dempewolf H. et al. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA // *Amer. J. Bot.* 2012. V. 99. № 2. P. 320–329.
- Kim W.J., Moon B.C., Yang S. et al. Rapid authentication of the herbal medicine plant species *Aralia continentalis* Kitag. and *Angelica biserrata* C.Q. Yuan and R.H. Shan using ITS2 sequences and multiplex-SCAR markers // *Molecules.* 2016. V. 21. № 3. P. 270.
- Kranitz M., Biffin E., Clark A. et al. Evolutionary diversification of New Caledonian *Araucaria* // *PLoS One.* 2014. V.9. P. e110308.
- Kuzmina M.L., Braukman T.W.A., Fazecas A.J. et al. Using herbarium-derived DNAs to assemble a large-scale DNA barcode library for the vascular plants of Canada // *Appl. Plant Sci.* 2017. V. 5. № 12: pii: apps.1700079.
- Li D.Z., Gao L.M., Li H.T. et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants // *PNAS USA.* 2011. V. 108. P. 19641–19646.
- Li Y., Feng Y., Wang X.-Y. et al. Failure of DNA barcoding in discriminating *Calligonum* species // *Nordic J. Bot.* 2014. V. 32. № 4. P. 511–517.
- Liang X.H., Wu Y.X. Identification of *Kalidium* species (Chenopodiaceae) by DNA barcoding // *Sciences in cold and arid regions.* 2017. V. 9. № 1. P. 89–96.
- Little D.P. A DNA mini-barcode for land plants // *Mol. Ecol. Resour.* 2014. V. 14. № 3. P. 437–446.
- Liu J., Yan H.F. Ge X.J. The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China // *PLoS One.* 2016. V. 11. № 4. P. e0153008.
- Liu Z.F., Ci X.Q., Li L. et al. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China // *PLoS One.* 2017. V. 124. № 4. P. e0175788.
- Michel C.-I., Meyer R.S., Taveras Y., Molina J. The nuclear internal transcribed spacer (ITS2) as a practical plant DNA barcode for herbal medicines // *J. Appl. Res. Med. Aromat. Plants.* 2016. V. 3. № 3. P. 94–100.
- Mishra P., Kumar A., Rodrigues V. et al. Feasibility of nuclear ribosomal region ITS1 over ITS2 in barcoding taxonomically challenging genera of subtribe Cassiinae (Fabaceae) // *Peer J.* 2016. V. 4. P. e2638.
- Naciri Y., Caetano S., Salamin N. Plant DNA barcodes and the influence of gene flow // *Mol. Ecol. Resour.* 2012. V. 12. P. 575–580.
- Nicholls J.A., Pennington R.T., Koenen E.J. et al. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae) // *Front. Plant Sci.* 2015. V. 6. P. 710.
- Nock C.J., Waters D.L., Edwards M.A. et al. Chloroplast genome sequences from total DNA for plant identification // *Plant Biotechnol. J.* 2011. V. 9. № 3. P. 328–333.
- Ojeda D.I., Santos-Guerra A., Oliva-Tejers F. et al. DNA barcodes successfully identified Macaronesian *Lotus* (Leguminosae) species within early diverged lineages of Cape Verde and mainland Africa // *AoB Plants.* 2014. V.6. DOI: 10.1093/aobpla/plu050
- Parmentier I., Dumini J., Kuzmina M. et al. How effective are DNA barcodes in the identification of African rainforest trees? // *PLoS One.* 2013. V. 8. № 4. P. e54921.

- Percy D.M., Argus G.W., Cronk Q.C. et al. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? // Mol. Ecol. 2014. V. 19. P. 4737–4756.
- Petit R.J., Excoffier L. Gene flow and species delimitation // Trends Ecol. Evol. 2009. V. 24. P. 386–393.
- Pillon Y., Johansen J., Sakishima T. et al. Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations // BMC Evol. Biol. 2013. V. 13. № 1. P. 35.
- Ren B.-Q., Xiang X.-G., Chen Z.-D. Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers // Mol. Ecol. Resour. 2010. V. 10. P. 594–605.
- Roy S., Tyagi A., Shukla V. et al. Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species // PLoS One. 2010. V. 5. № 10. P. e13674.
- Ruhsam M., Rai H.S., Mathews S. et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? // Mol. Ecol. 2015. V. 15. № 5. P. 1067–1078.
- Särkinen T., Staats M., Richardson J.E. et al. How to open the treasure chest? Optimising DNA extraction from herbarium specimens // PLoS ONE. 2012. V. 7. № 8. P. e43808.
- Seberg O., Petersen G. How many loci does it take to DNA barcode a *Crocus*? // PLoS One. 2009. V. 4. № 2. P. e4598.
- Song Y., Wang S., Ding Y. et al. Chloroplast genomic resource of *Paris* for species discrimination // Sci. Rep. 2017. V. 7. № 3427.
- Stech M., Veldman S., Larrain J. et al. Molecular species delimitation in the *Racomitrium canescens* complex (Grimmiaceae) and implications for DNA barcoding of species complexes in mosses // PLoS One. 2013. V. 8. № 1. P. e53134.
- Straub S.C., Parks M., Weitemier K. et al. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics // Amer. J. Bot. 2012. V. 99. P. 349–364.
- Stull G.W., Moore M.J., Mandala V.S. et al. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes // Appl. Plant Sci. 2013. V. 1. № 2. P. e1200497.
- Vivas C.V., Moraes R.C., Alves-Araújo A. et al. DNA barcoding in Atlantic Forest plants: what is the best marker for Sapotaceae species identification? // Gen. Mol. Biol. 2014. V. 37. № 4. P. 662–670.
- Wang X.-Y., Zheng S.-H., Liu Y., Han J.-P. ITS2, a better DNA barcode than ITS in identification of species in *Artemisia* L. // Chin. Herb. Med. 2016. V. 8. № 4. P. 352–358.
- Wu F., Ma J., Meng Y. et al. Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations // PLoS One. 2017. V. 12. № 9. P. e0182693.
- Yan H.-F., Hao G., Hu C.-M., Ge X.-J. DNA barcoding in closely related species: a case study of *Primula* L. sect. *Proliferae* Pax (Primulaceae) in China // J. Syst. Evol. 2011. V. 49. № 3. P. 225–236.
- Yan H.-F., Liu Y.J. Xie X.F. DNA barcoding evaluation and its taxonomic implications in the species-rich genus *Primula* L. in China // PLoS ONE. 2015. V. 10. № 4. P. e0122903.
- Yang J., Vázquez L., Chen X. et al. Development of chloroplast and nuclear DNA markers for chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes // Front. Plant Sci. 2017. V. 8. P. 816.
- Yao P.-C., Gao H.-Y., Wei Y.-N. et al. Evaluating sampling strategy for DNA barcoding study of coastal and inland halotolerant *Poaceae* and *Chenopodiaceae*: a case study for increased sample size // PLoS One. 2017. V. 12. № 9. P. e0185311.
- Zarrei M., Talent N., Kuzmina M. et al. DNA barcodes from four loci provide poor resolution of taxonomic groups in the genus *Crataegus* // AoB Plants. 2015. V. 7. DOI: 10.1093/aobpla/plv045

## Plant DNA Barcodes

V. S. Shneyer<sup>1,\*</sup>, A. V. Rodionov<sup>1,2,\*\*</sup>

<sup>1</sup>Komarov Botanical Institute, Russian Academy of Sciences, St. Petersburg, Russia

<sup>2</sup>Saint-Petersburg State University, St. Petersburg, Russia

E-mail: \*shneyer@rambler.ru, \*\*avrodionov@mail.ru

Universal plant DNA barcodes proved to be a big challenge. A single site or several ones that could be used to determine a particular species was not found, although the combination (rbcL + matK + ITS) recommended in 2009 as the standard allowed to refer the species to the definite genus. The variability of some markers varies in different taxonomic groups, which, as a rule, permits to select a DNA barcode (sometimes a mini-barcode) for a specific group, especially in the cases of applied nature. New sequencing technologies (NGS) are prospective to gain extended DNA barcodes (sequences of complete chloroplast genomes and of ribosomal genes), which can overcome the limitations of the current DNA barcodes. It is important that with NGS technologies, the sequence data are recoverable when using herbarium specimens. The search for plant DNA barcodes is still ongoing.

**Keywords:** plant DNA barcoding, species discrimination, chloroplast DNA, next generation sequencing.

