

Раздел «Взгляд эксперта» / Section “Expert’s view”

БИОЛОГИЧЕСКИЕ НАУКИ BIOLOGICAL SCIENCES

УДК 577.212.3: 581.174.1: 582.632.1: 582.685.4

Можаровская Л.В., Пантелеев С.В., Кирьянов П.С., Баранов О.Ю., Падутов В.Е.

СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ХЛОРОПЛАСТНОГО ГЕНОМА КАРЕЛЬСКОЙ БЕРЕЗЫ

ГНУ «Институт леса НАН Беларуси», г. Гомель, Беларусь

Проведено высокопроизводительное секвенирование хлоропластной ДНК карельской березы с последующей аннотацией. Произведена оценка частоты встречаемости различных олигонуклеотидных комбинаций, а также выполнен поиск олигонуклеотидных последовательностей, представляющих собой повторяющиеся мотивы – микросателлитные локусы. На основании полученных данных созданы праймеры для SSR-анализа хлДНК карельской березы.

Ключевые слова: карельская береза, высокопроизводительное секвенирование, микросателлитные локусы

Введение. Редкая и хозяйственно-важная разновидность березы повислой – карельская береза (*Betula pendula* Roth. var. *carelica* Mercl.), является ценным сырьем для лесоперерабатывающей промышленности. Данная порода представляет собой дерево второй величины высотой до 25 м или сильно ветвящийся кустарник с приподнимающимися стволиками до 3 м высотой. Характерной особенностью карельской березы является высокодекоративная текстура древесины, напоминающая мрамор с переливами различных оттенков и темными включениями разнообразной формы. По внешнему виду березу карельскую можно разделить на четыре основные формы: высокоствольная узорчатая форма, короткоствольная узорчатая форма, кустовидная узорчатая форма и полнодревесная безузорчатая форма [1].

Традиционное размножение карельской березы путем семян не является эффективным, что связано со сложным характером детерминации признака узорчатости и расщеплением его в потомстве. Отбор селекционного материала лесных древесных пород, основанный на анализе ряда морфологических признаков, не позволяет за короткие сроки достигать намеченных целей, что связано с длительным жизненным циклом древесных пород и проявлением многих хозяйственно-важных признаков на поздних этапах развития растений. Как показывает многолетняя мировая практика, использование ДНК-маркеров является быстрым и точным инструментом в ходе реализации той или иной селекционной программы. ДНК-маркеры по-

зволяют производить: отбор ценного селекционного материала на ранних этапах онтогенеза, подбор родительских пар и анализ системы скрещивания, что является также весьма существенным. Кроме того, использование ДНК-маркеров позволяет охарактеризовать не только селекционируемый признак (ген), но и предоставить информацию о состоянии других генов и генома индивидуума в целом, т.к. для обеспечения жизненных процессов организма необходимо нормальное функционирование большого числа генов [2].

Наиболее эффективными генетическими маркерами, используемыми для оценки уровня генетического полиморфизма и проведения селекционных исследований, являются локусы, локализованные в цитоплазматических органеллах. Цитоплазматические маркеры, как правило, наследуются по одной из родительских линий, что в совокупности с гаплоидностью и низким уровнем рекомбинации делает их информативным инструментом в исследованиях в области молекулярной таксономии и филогенетики, оценки состояния генетических ресурсов, фило- и геногеографии, селекции и др. [3]. При этом наибольшая информативность, отмечена для локусов хлоропластной ДНК и, в частности, микросателлитных маркеров (cpSSR) [4].

В настоящее время выявление и разработка ПЦР-подходов к диагностике микросателлитных регионов производится с помощью различных биоинформационных методов, основанных на анализе данных высокопроизводительного (NGS) секвенирования, как *de novo*, так и представленных в ин-

формационных генетических базах.

Исходя из всего выше сказанного, целью данной работы является определение и анализ нуклеотидной последовательности хпДНК карельской березы, выявление регионов, содержащих микросателлитные повторы, и разработка набора праймеров для проведения сrSSR-анализа.

Материалы и методы исследования. Объектом исследований являлись микроклональные растения короткоствольной формы карельской березы из коллекции культур *in vitro* Института леса НАН Беларуси, что обусловлено необходимостью снижения доли секвенированных артефактных нуклеотидных последовательностей, относящихся к эндофитной и эпифитной микрофлоре в получаемых биоинформационных данных.

В качестве экспериментального материала были использованы апикальные части побегов исследуемых растений.

Выделение хлоропластов из образцов растений проводилось согласно методике Янсона с соавторами [5]. Последующее получение препаратов нуклеиновых кислот, насыщенных хпДНК, выполнялось с применением модифицированного СТАВ-метода [6].

Создание библиотек ДНК-фрагментов (≈ 200 п.н.) осуществлялось с использованием набора Ion Plus Fragment Library Kit (Thermo Scientific, США). Эмульсионная ПЦР производилась в амплификаторе Ion OneTouch 2 (Thermo Scientific, США) согласно протокола Ion PGM Template OT2 Kit (Thermo Scientific, США). Секвенирующая реакция выполнялась с применением системы Ion PGM System (Thermo Scientific, США) на основе использования наборов Ion 314 Chip v2 и Ion PGM Sequencing 200 Kit v2 (Thermo Scientific, США).

Первоначальную обработку данных, поступающих от геномного анализатора, осуществляли в автоматическом режиме при помощи программного обеспечения Ion Torrent Suite (Thermo Scientific, США). Окончательную обработку информации, включая сборку и аннотацию последовательностей, выполняли с помощью программного пакета Lasergene v.11 (DNASTAR, Израиль).

Результаты и их обсуждение. В результате проведенного высокопроизводительного секвенирования и сборки хпДНК-обогащенной библиотеки *in vitro* линии карельской березы (короткоствольная форма) получено 459 259 парноконцевых чтений. Общий объем полученных данных секвенирования равнялся 105,0 миллионов нуклеотидов (показатель качества $Q \geq 20$). Длина прочтения варьировала от 229 до 263 нуклеотидов (среднее 259). Общий размер хпДНК после сборки составил ≈ 160 тыс. п.н.

Проведенная аннотация пластома позволила

выявить 134 кодирующих локуса, из которых: 40 – представлены генами тРНК (ассоциированы с 19 типами аминокислот), 8 – гены рРНК (по 2 копии генов 4,5S-, 5S-, 16S-, 23S рРНК), 25 – гены рибосомальных белков (большой и малой субъединиц), 25 – гены белков фотосистемы 1 (11) и 2 (14), 6 – кодируют субъединицы хпАТФазы, 4 – гены РНК-полимеразы, 12 – кодируют субъединицы НАДН-дегидрогеназы, 14 – гены белков с различными функциями (элементы электронно-транспортной цепи, ферменты карбоксилирования и др.).

Изучение частоты встречаемости нуклеотидных оснований в хпДНК карельской березы показало, что доминирующими видами нуклеотидов являются А и Т, суммарно составляя примерно 2/3 от всех оснований (таблица 1). Превалирование оснований, формирующих менее энергетически выраженные двойные связи между цепями ДНК, в целом, определяют ее большую способность к диссоциации, тем самым увеличивая эффективность протекания процессов транскрипции и репликации.

Таблица 1 – Частота встречаемости нуклеотидных оснований в хпДНК карельской березы (плюсовая цепь)

Нуклеотидное основание	Суммарное количество	Частота встречаемости, %
А	50791	31,5
С	29555	18,3
Г	28482	17,7
Т	52294	32,5

Анализ распределения регионов, характеризующихся повышенной частотой встречаемости определенных типов нуклеотидных оснований, в структурах хпДНК карельской березы выявил ряд особенностей (рисунок 1). Так, участки генома, насыщенные А и Т нуклеотидными основаниями, в хпДНК были распределены практически дисперсно. В то же время С-насыщенные регионы, характеризовались дисперсно-кластерной структурой, а участки с превалированием Г-нуклеотидов, в своем большинстве были организованы отдельными кластерами. Еще более выраженная кластерная структура была отмечена для регионов с превалированием энергетически низких двойных (А–Т) и высоких тройных (Г–С) связей. Так, основные ГС-насыщенные регионы были выявлены между 20000-30000 и 150000-160000 нуклеотидными позициями в линейной первичной структуре хпДНК карельской березы. При этом, размер элементов кластеров был незначительным и не превышал 300-500 нуклеотидов. В то же время протяженность АТ-насыщенных

участков генома могла составлять 15000-20000 нуклеотидов. Наличие кластерной структуры распределения регионов с различными термодинамическими характеристиками является одним из факторов, определяющих образование пространственной структуры молекулы хпДНК и обуславливающих эффективность протекания процессов трансляции и

репликации. Выявленные особенности нуклеотидной организации хпДНК – отсутствие дисперсного распределения – подтверждают выводы, сделанные широким кругом авторов по результатам геномных исследований различных видов живых организмов, о неприменимости третьего правила Чаргаффа на субгеномном уровне.

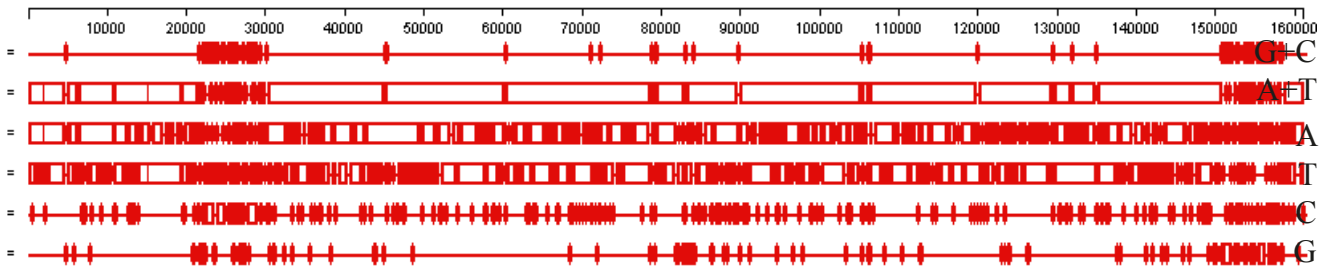


Рисунок 1 – Распределение регионов, характеризующихся повышенной частотой встречаемости определенных типов нуклеотидных оснований, в хпДНК карельской березы (плюсовая цепь)

Одним из аспектов оценки структуры хпДНК явилось выявление особенностей распределения и встречаемости различных видов олигонуклеотидных сочетаний в хпДНК карельской березы. Наиболее основными видами сочетаний, явились ТТ, АА, АТ, ТА, что соответствует статистической вероятности формирования данных типов комбинаций, исходя из частот встречаемости нуклеотидных оснований в хпДНК карельской березы. Аналогичная зависимость выявлена и для динуклеотидных комбинаций, включающих одновременно основания с высокой и низкой частотой встречаемости – представляющих группу сочетаний с умеренным уровнем встречаемости. Наименьшей распространенностью характеризовались динуклеотидные сочетания CG и GC. При этом для гомоолигомеров CC и GG встречаемость была выше от расчетной, что по всей видимости обусловлено положительным молекулярным отбором данных сочетаний, по сравнению с их альтернативными формами.

Сходные результаты были получены для три-, тетра-, пентануклеотидных последовательностей. Наибольшей частотой встречаемости характеризовались комбинации, включающие в свой состав нуклеотиды А и Т. При этом увеличение в составах комбинаций доли нуклеотидов С и G, постепенно вызывает уменьшение частоты встречаемости данных вариантов последовательностей. Так, например, последовательности CGGCCG и CGGCGC в первичной структуре встречаются в единичных случаях. Отсюда следует, что в большинстве случаев, частота встречаемости комбинаций олигону-

клеотидных фрагментов находится в прямой зависимости от долевого участия в геноме тех или иных нуклеотидных оснований.

Кроме оценки частоты встречаемости различных олигонуклеотидных комбинаций в хпДНК карельской березы, был выполнен и поиск олигонуклеотидных последовательностей, представляющих собой повторяющиеся мотивы – т.е. микросателлитные локусы. Проведенный анализ хпДНК карельской березы позволил выявить 384 микросателлитных локуса, мотивы которых представлены в таблице 2. Следует отметить, что критерии поиска мотивов были следующие: для мононуклеотидных – не менее 7 повторов, динуклеотидных – не менее 6 повторов, три (и более) нуклеотидных – не менее 5 повторов.

Как и в случае анализа нуклеотидных комбинаций, наибольшее число мотивов были представлены мононуклеотидными А- и Т-мотивы (94% от всех мононуклеотидных мотивов и 92% от всех типов мотивов). При этом, как и ожидалось, наблюдалась обратная зависимость количества диагностируемых локусов от длины повторяющегося региона. Так, например, количество локусов с семью повторами А мотива составило 94 шт., а с 12 повторами – 5 шт.

Среди динуклеотидных и тринуклеотидных мотивов были выявлены последовательности содержащие только А и Т нуклеотиды (8% от всех SSR-локусов). Несмотря на то, что общая частота встречаемости GC и CG сочетаний составила 5,8% в хпДНК карельской березы они не идентифицированы.

Таблица 2 – Представленность различных типов повторяющихся мотивов в хпДНК карельской березы (в абсолютном исчислении)

Мотив	Число повторов												Всего локусов
	5	6	7	8	9	10	11	12	13	14	15	16	
A	-	-	94	30	24	13	5	5	-	-	-	1	172
C	-	-	10	3	1	1	-	-	-	-	-	-	15
G	-	-	6	1	1	-	-	-	-	-	-	-	8
T	-	-	78	45	30	22	3	3	1	-	-	-	182
AT	-	1	2	1	-	-	-	-	-	-	-	-	4
TA	-	1	-	-	-	-	-	-	-	-	-	-	1
ATT	1	-	-	-	-	-	-	-	-	-	-	-	1
TAA	1	-	-	-	-	-	-	-	-	-	-	-	1
A/T	-	-	172	75	54	35	8	8	1	-	-	1	354
C/G	-	-	16	4	2	1	-	-	-	-	-	-	23
AT/AT	-	2	2	1	-	-	-	-	-	-	-	-	5
AAT/ATT	2	-	-	-	-	-	-	-	-	-	-	-	2

На основании анализа структуры микросателлитных повторов хпДНК и фланкирующих их регионов, для карельской березы нами была разработана нуклеотидная структура cpSSR-праймеров с целью последующего формирования набора хпДНК-маркеров для проведения популяционно-генетических исследований.

Заключение. В ходе секвенирования хпДНК ка-

рельской березы (короткоствольная форма) были описаны полные нуклеотидные последовательности хлоропластного генома, а также проведен их структурный анализ. Идентифицированы основные мотивы микросателлитных повторов хпДНК для молекулярно-генетического типирования индивидуальных деревьев и насаждений карельской березы, а также разработаны праймеры для их амплификации методом ПЦР.

Список источников:

1. Барсукова, Т.Л. Береза карельская в Белоруссии / Т.Л. Барсукова // Сельское хозяйство Белоруссии. – 1986. – № 8. – С. 38–45.
2. Хлесткина, Е.К. Молекулярные маркеры в генетических исследованиях и в селекции / Е.К. Хлесткина // Вавиловский журнал генетики и селекции. – 2013. – Т. 17. – №4-2. – С. 1044-1054
3. Kalia, R.K. Microsatellite markers: an overview of the recent progress in plants / R.K Kalia, [et al.] // Euphytica. – 2011. – V. 177. – №3. – P. 309-334.
4. Ruongis D. Aanalysis of the genetic diversity and population structure of latvian ash (*Fraxinus excelsior* L.) stands using nuclear and chloroplast SSR markers / D. Ruongis, A. Korica, A. Gailīte, I. Puspure, I. Veinberga // Proceedings of the Latvian academy of sciences. – 2016. – Vol. 70, № 3. – P. 101–108.
5. Jansen R.K. Methods for obtaining and analyzing whole chloroplast genome sequences / Methods in enzymology, 2005. – Vol. 395. – P. 348-384.
6. Падутов, В.Е. Методы молекулярно-генетического анализа / В.Е. Падутов, О.Ю. Баранов, Е.В. Воропаев. – Мн.: Юнипол, 2007. – 176 с.

Mozharovskaz L.V., Panteleev S.V., Kiryanov P.S., Baranov O.Yu., Padutov V.E.

STRUCTURAL-FUNCTIONAL ANALYSIS OF THE CHLOROPLAST GENOME OF THE KARELIAN BIRCH

Forest Research Institute, NAS of Belarus

Summary

Next-Generation Sequencing of chloroplast DNA of Karelian birch with annotation was carried out. The frequency of occurrence of different oligonucleotide combinations was estimated, and the search for oligonucleotide sequences, which are repetitive motifs - microsatellite loci, was performed. On the basis of the data obtained, primers for SSR analysis of the cpDNA of the Karelian birch were created.

Key words: Karelian birch, Next Generation Sequencing, microsatellite loci