

ИДЕНТИФИКАЦИЯ РАСТЕНИЙ СЕМЕЙСТВА *FABACEAE* ПРИ ИСПОЛЬЗОВАНИИ АНАЛИЗА МОЛЕКУЛЯРНОГО БАРКОДИРОВАНИЯҚали Б.Р.<sup>1</sup> , Сутула М.Ю.<sup>1</sup> , Ахметоллаева А.С.<sup>1</sup> , Манабаева Ш.А.<sup>1,2\*</sup> <sup>1</sup> Национальный центр биотехнологии, 010000, Казахстан, г. Астана, Кургальжинское шоссе, 13/5<sup>2</sup> Евразийский национальный университет им. Л.Н. Гумилева, 010000, Казахстан, г. Астана, ул. К. Сатпаева, 2

Автор для корреспонденции: \*manabayeva@biocenter.kz

## АБСТРАКТ

Семейство *Fabaceae* в Казахстане насчитывается 42 рода и около 650 видов. Бобовые культуры, такие как бобы, фасоль, соя, горох, а также кормовые растения – клевер, люцерна, эспарцет и донник – имеют высокое содержание белка и широко используются в пищевой и кормовой промышленности. Они также используются для производства биоразлагаемых материалов, масел, красителей и биодизельного топлива. ДНК-баркодирование эффективно для идентификации видов *Fabaceae*, несмотря на их морфологическую изменчивость, что подчеркивает важность сохранения их разнообразия. Для оценки генетического разнообразия 15 видов *Fabaceae* в Карагандинской области использовались три пары праймеров. Применялись хлоропластные маркеры *matK* и *rbcL*, а также ядерный маркер ITS. Полученные последовательности были проанализированы в Vector NTI и MEGA11, зарегистрированы в базе данных GenBank NCBI, ваучерные образцы хранятся в гербарии Национального центра биотехнологии в Астане. Исследование демонстрирует эффективность ДНК-баркодирования для идентификации видов *Fabaceae* на основе анализа маркеров ITS, *rbcL* и *matK* методом maximum likelihood. Результаты подчеркивают значимость выбора маркеров с учетом их изменчивости для точного филогенетического анализа и необходимость дальнейших популяционных исследований для сохранения биологического разнообразия в Казахстане.

**Ключевые слова:** *Fabaceae*, ДНК-баркодирование, *matK*, *rbcL*, ITS

## ВВЕДЕНИЕ

Семейство *Fabaceae*, широко известное как бобовые, имеет важное экологическое, экономическое и культурное значение во всем мире, в том числе в Казахстане, где оно представлено многочисленными местными и культивируемыми видами. Казахстан, обладая разнообразной флорой и обширными территориями, может похвастаться богатым ботаническим наследием, включающим различные экосистемы – от засушливых степей до альпийских лугов [1]. Семейство бобовых насчитывает около 650 родов и 18 000 видов, распространенных по всему земному шару, являясь одним из крупнейших семейств мировой флоры [2]. На евразийском континенте встречается более 110 родов и около 1700 видов, из которых в Казахстане произрастает 42 рода и около 650 видов [3].

Семейство *Fabaceae* является третьим по величине семейством цветковых растений в мире после *Orchidaceae* и *Asteraceae* [4, 5]. С точки зрения экономики семейство бобовых (*Fabaceae*) занимает второе место по значимости после семейства злаковых, (*Poaceae*) в сельском хозяйстве. Зерновые бобовые составляют 27% от общего мирового производства сельскохозяйственных культур и обеспечивают 33% потребляемого человеком пищевого белка [6]. Благодаря высокому содержанию белка бобовые культуры, такие как бобы, фасоль, соя, горох, а также кормовые растения, включая клевер, люцерну, эспарцет и донник, нашли широкое применение в пищевой и кормовой промышленности. Бобовые также активно используются в промышленности для производства биоразлагаемых материалов, масел, красителей и биодизельного топлива [7].

В современном мире проблема сохранения биоразнообразия становится всё более актуальной. Одним из важных аспектов этой проблемы является идентификация и изучение видов растений. Исследования, направленные на

изучение современного состояния и генетического разнообразия эндемичных, редких, исчезающих и хозяйственно ценных видов растений, до сих пор проводились, опираясь исключительно на морфологические признаки с применением ботанических методов. Однако эти подходы не всегда обеспечивают достаточную точность и требуют значительных физических усилий [8].

В последние десятилетия молекулярные методы стали неотъемлемой частью современной систематики растений, предоставляя исследователям инструменты для точной идентификации и классификации видов [9]. Баркодирование ДНК представляет собой новый метод идентификации, характеристики и классификации биологических образцов или видов организмов. Он основан на использовании короткой консервативной последовательности ДНК, охватывающей как ядерную, так и геномную ДНК органелл организма. Эти генетические коды могут быть сохранены в цифровой библиотеке и применены для идентификации неизвестных видов [10-12].

Баркодирование ДНК может преодолеть ограничения, связанные с морфологическими характеристиками, и ускорить процесс идентификации видов растений и животных, поскольку данным способом, возможно, определить организмы на любой стадии их роста. По данным Elansary Н.О. и др. [13] морфологическая идентификация не эффективна, особенно для сложных таксономических групп, таких как *Fabaceae* [14], *Convolvulaceae* [15, 16], *Ranunculaceae* [17]. Данная методика направлена на создание общего ресурса в виде последовательностей ДНК, который может быть использован для идентификации или таксономической классификации организмов в любом сообществе [18, 19]. Метод призван уменьшить возможные ошибки, которые могут возникнуть при традиционной идентификации [20]. Быстрое и точное определе-

ние видов играет ключевую роль в сохранении растений, исследовании биоразнообразия и оценке экологического состояния [21]. Баркодирование ДНК также может применяться для охраны редких эндемичных видов, находящихся под угрозой исчезновения, а также для изучения эволюции, экологии и сохранения растений. Это особенно важно, учитывая, что биоразнообразие подвергается угрозе из-за антропогенной деятельности, загрязнения окружающей среды, вырубки лесов и эксплуатации природных ресурсов [22].

Цель данного исследования заключается в использовании последовательностей ДНК для оценки эффективности ДНК баркодов *matK*, *rbcL* и ITS в идентификации растений семейства *Fabaceae*, а также в оценке их точности в реконструкции филогенетических отношений между wybranymi видами. Наши исследование представляют собой только небольшую часть литературы, посвященной идентификации растений семейства *Fabaceae* с использованием молекулярного баркодирования и стремится расширить и углубить наше понимание этой проблемы, внести свой вклад в развитие методов идентификации растений семейства *Fabaceae* на основе современных молекулярных технологий. В данной работе будут рассмотрены основные результаты таких исследований и их

вклад в сохранение биоразнообразия Казахстана.

## МАТЕРИАЛЫ И МЕТОДЫ

### Отбор проб и сбор данных

Растительный материал был собран в Карагандинской области учеными-ботаниками Карагандинского государственного университета имени академика Е.А. Букетова под руководством Ишмуратовой М.Ю. Растительный материал был идентифицирован ботаниками с применением специального идентификационного ключа ботанической базы данных [23]. Подробный список образцов приведен в Таблице 1.

Соответствующие ваучерные образцы хранятся в гербарном фонде Национального центра биотехнологий (Астана, Казахстан). Нуклеотидные последовательности контрольной внешней группы *Cladrastis sikokiana* (Makino), *Myrocarpus frondosus* (Allemao) взятые из GenBank, также представлены в электронных дополнительных материалах [24]. Образцы контрольной внешней группы были выбраны на основе их соответствующей длины и простоты выравнивания.

**Таблица 1.** Представители семейства *Fabaceae*, включая данные о трех использованных маркерах и соответствующих инвентарных номерах NCBI.

№	Виды растений	Инвентарные номера в NCBI		
		<i>matK</i>	<i>rbcL</i>	ITS
	<i>Astragalus sulcatus</i> (L.)*	OR683687	OR640487	OR676310
	<i>Astragalus cornutus</i> Pall.*	OR644024	OR640485	OR676320
	<i>Astragalus danicus</i> Retz.*	OR683689	OR640486	OR676321
	<i>Sophora alopecuroides</i> (L.)*	OR644015	OR640476	OR676311
	<i>Onobrychis arenaria</i> (Kit.) DC.*	OR683688	OR640477	OR676312
	<i>Trifolium lupinaster</i> (L.)*	OR644017	OR640478	OR676313
	<i>Trifolium repens</i> (L.)*	OR644018	OR640479	OR676314
	<i>Oxytropis pilosa</i> (L.) DC*	OR644019	OR640480	OR676315
	<i>Caragana halodendron</i> (Pall.) Dum.Cours.*	OR644021	OR640482	OR676317
	<i>Caragana frutex</i> (L.) K. Koch*	OR678648	OR678662	OR687291
	<i>Melilotus officinalis</i> (L.) Lam.*	OR644023	OR640484	OR676318
	<i>Melilotus officinalis</i> (L.) Lam.*	OR678653	OR678669	OR687297
	<i>Medicago platycarpus</i> (L.) Trautv.*	OR644022	OR640483	OR676319
	<i>Medicago sativa</i> subsp. <i>falcata</i> (L.) Arcang.*	OR644020	OR660064	OR676316
	<i>Glycyrrhiza uralensis</i> Fisch. ex DC.*	OQ722366	OQ722367	OR364050
	<i>Oxytropis coerulea</i> (Pall.) DC.	MH714120	MH713905	LR898460
	<i>Onobrychis pulchella</i> Schrenk ex Fisch. & C.A.Mey.	AB854599	MT931459	MT923631
	<i>Glycyrrhiza glabra</i> L.	PP798191	OR703725	PP158762
	<i>Sophora chrysophylla</i> Chock	ON996783	ON996708	ON228727
	<i>Melilotus albus</i> Medik.	MT918280	MT931452	MT923624
	<i>Medicago sativa</i> subsp. <i>falcata</i> (L.) Arcang.	MT918266	MT931443	MT923616
	<i>Medicago laciniata</i> (L.) Mill.	MK290560	KX282879	GQ246120
	<i>Trifolium pratense</i> L.	KX676980	MW497159	OR428717

	<i>Astragalus alopecurus</i> Pall.	MF543487	MF572163	MF543756
	<i>Cladrastis sikokiana</i>	MN497640	LC690939	JQ676968
	<i>Myrocarpus frondosus</i>	MG718721	MG718178	MH432197

Примечание: Звездочкой (\*) отмечены исследуемые в этой статье виды растений семейства бобовых.

### Выделение ДНК, амплификация и секвенирование

Геномную ДНК выделяли из образцов гербарных растений модифицированным методом СТАВ [25]. Извлеченную ДНК проверяли на целостность, гомогенность и чистоту с помощью электрофореза в 2% агарозном геле при 120 В в течение 30 мин. ДНК хранили при  $-20^{\circ}\text{C}$  до использования на следующем этапе эксперимента. Экстракцию геномной ДНК из образцов проводили с использованием трех и более независимых повторов. Все отобранные образцы имели ДНК высокого качества и достигали концентрации выше 50 нг/мкл.

Выбор универсальных праймеров для баркодирования основывался на соответствующей литературе, подробная информация представлена в Таблице 2.

Все праймеры были синтезированы в Лаборатории органического синтеза Национального центра биотехнологии (г. Астана, Казахстан). ПЦР проводили в общем объеме реакции 40 мкл, содержащем 2 мкл геномной ДНК (50 нг), 4 мкл 25 mM  $\text{MgCl}_2$  (Thermo Scientific), 4 мкл 10-кратного Taq-буфера (Thermo Scientific), 1 мкл 10 mM дНТФ (Thermo Scientific), 1 мкл прямого и обратного праймеров (10 мкмоль/л стокового раствора), 27,6 мкл дистиллированной воды и 0,4 мкл 10-кратной Taq-полимеразы (Gen Lab). ПЦР амплификацию проводили в термоциклере T100 (Bio-Rad) с использованием следующей программы ПЦР: 5 минут при  $95^{\circ}\text{C}$  для начальной денатурации (один цикл), 30 циклов по 1 минуте при  $95^{\circ}\text{C}$  для денатурации, 1 минуте при оптимальной температуре отжига для каждого праймера ( $50^{\circ}\text{C}$ – $58^{\circ}\text{C}$ ; Таблица 2) и 1 минуте при  $72^{\circ}\text{C}$  для элонгации; наконец, один цикл 10 минут при  $72^{\circ}\text{C}$  для увеличения количества копий целевого фрагмента последовательности ДНК, за которым следует хранение при  $4^{\circ}\text{C}$ . Полученные продукты ПЦР проверяли с помощью электрофореза в 2% агарозном геле, проводимого в течение 30 минут при 120 В, и очищали с использованием набора для быстрой экстракции ДНК из геля PureLink от Invitrogen. На основе экспериментальных данных температура отжига праймеров была установлена на уровне  $50^{\circ}\text{C}$  для *matK*,  $58^{\circ}\text{C}$  для *rbcL* и  $55^{\circ}\text{C}$  для ITS. Очищенные продукты ПЦР секвенировали методом Сэнгера на анализаторе ДНК 3730xl (Applied Biosystems). Полученные

последовательности прямых и обратных праймеров из каждого образца были анализированы с использованием программного обеспечения ContigExpress от Invitrogen (Vector NTI Advance 11.5) [30], а контиги собрали для минимизации возможных ошибок чтения. Собранные последовательности сравнивали с существующими последовательностями ДНК с помощью инструмента BLASTn [31] платформы Национального центра биотехнологической информации (NCBI) GenBank [32].

### Анализ данных

Последовательности были выравнены в программе MEGA 11 [33] с использованием автоматического выбора алгоритма и стандартных настроек. Выравненные последовательности проверяли в BioEdit v. 7.2 [34] и перепроверяли вручную. Выравнивания доступны в электронных дополнительных материалах (доп. Приложение S1–S3).

Анализ Maximum likelihood (ML) был выполнен в MEGA 11 с использованием стандартных настроек. Тестирование моделей проводилось автоматически в MEGA 11 с помощью Model Test. Наилучшие модели для данных nrDNA были: ITS K2+G, для последовательностей cpDNA: CDS *rbcL* T92+G, *matK* T92+G. Ветви с поддержкой бутстрепа менее 95% считались неподдержанными.

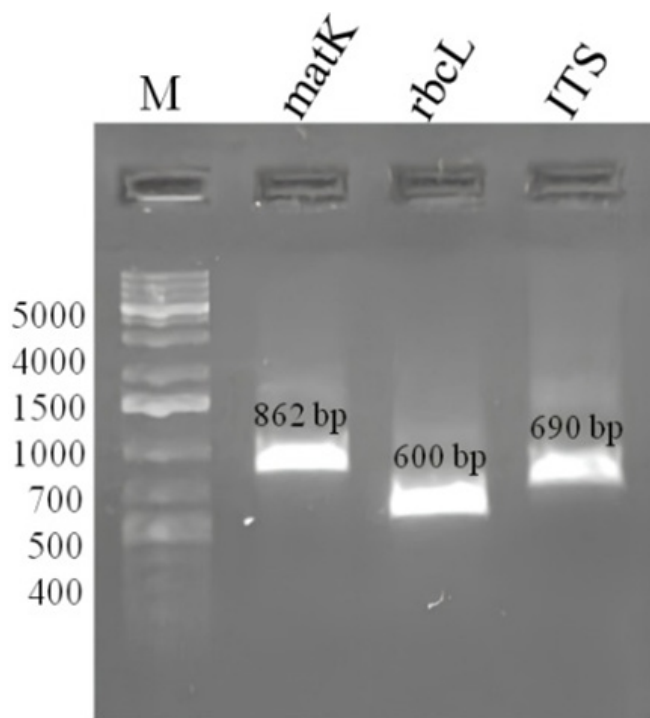
### РЕЗУЛЬТАТЫ

Электрофореграмма агарозного геля, представляющая ПЦР-продукты, показала, вариабельность длины ампликонов для различных ДНК маркеров: *rbcL* варьировался от 564 до 590 bp, *matK* – от 598 до 732 bp, ITS – от 607 до 744 bp (Рис. 1). Эффективность амплификации ДНК-баркодов составила 98%. Полученные последовательности ДНК представителей семейства *Fabaceae* из Казахстана были депонированы в базе данных NCBI, их номера доступа указаны в Таблице 1.

Основные характеристики генетического межродового разнообразия были проанализированы, включая содержание консервативных (C), полиморфных (V) и однонуклеотидных (S) регионов (Таблица 3). Регионы ITS (ITS1, полный ген 5.8S рДНК, ITS2 и небольшая часть гена 26S рДНК) демонстрировали наибольшее расхождение (0.16), при этом 59.2% составляли консервативные регионы, а

Таблица 2. Нуклеотидные последовательности праймеров ПЦР, используемых для ДНК-баркодирования

Название праймера	Нуклеотидная последовательность (5'-3')	Целевой locus	T °C отжига
3F_KIMf [26]	CGTACAGTACTTTTGTGTTTACGAG	<i>matK</i>	50
1R_KIMr [26]	ACCCCATTCATCTGGAAATCTTGGTTC	<i>matK</i>	50
<i>rbcLa</i> _F [27, 28]	ATGTCACCAACAACAGAGACTAAAGC	<i>rbcL</i>	58
<i>rbcLa</i> _R [27, 28]	GTAATAATCAAGTCCACCRG	<i>rbcL</i>	58
ITS4 [29]	TCCTCCGCTTATTGATATGC	ITS1 and ITS2	55
ITS5 [29]	GGAAGTAAAAGTCGTAACAAG	ITS1 and ITS2	55



**Рисунок 1** – Электрофореграмма ПЦР продуктов, амплифицированных с соответствующими праймерами (ITS, *matK*, *rbcL*)

доля полиморфных регионов варьировалась в пределах 34.5%. Регионы *rbcL* проявили более высокую степень консерватизма (0.07), с 85.3% консервативных и 14.7% вариативных регионов. Регионы *matK* характеризовались средними показателями (0.07), с 75.5% консервативных и 24.5% вариативных регионов. Содержание G+C после множественного выравнивания анализируемых маркеров варьировалось от 31.8 до 51.8 %.

Использование инструмента BLAST для поиска идентичных последовательностей в базе данных NCBI показало ограниченную эффективность на уровне вида для хлоропластных ДНК-маркеров *rbcL* и *matK*. Поиск в базе данных NCBI с использованием ДНК-маркеров (ITS, *rbcL* и *matK*) успешно идентифицировал только виды *Glycyrrhiza uralensis* (Fisch.), *Caragana frutex* (L.), *Melilotus officinalis* (L.), *Astragalus sulcatus* (L.), *Sophora alopecuroides* (L.), *Trifolium lupinaster* (L.), *Trifolium repens* (L.), *Oxytropis pilosa* (L.), *Medicago sativa* subsp. *falcata* (L.) на уровне рода с точностью 100%. Однако наблюда-

лись расхождения для некоторых видов, таких как *Onobrychis arenaria* (Kit.) DC. (100% сходства с *Trifolium lupinaster* (L.)), *Caragana halodendron* (Pall.) Dum.Cours. (100% сходства с *Halimodendron halodendron* (Pall.) Voss., 98.80% с *Caragana gerardiana* Benth.), *Melilotus officinalis* (L.) Lam. (99.83% сходства с *Melilotus polonicus* (L.) Pall.), *Medicago platycarpus* (L.) Trautv. (100% сходства с *Medicago sativa* (L.)), *Astragalus cornutus* Pall. (99.61% сходства с *Astragalus austriacus* Jacq.), что подтверждает необходимость дальнейших исследований для уточнения таксономических связей этих видов. Тем не менее, все исследованные ДНК-баркоды успешно идентифицировали виды на уровне рода с точностью 100%. Все маркерные гены были успешно секвенированы с точностью 100%.

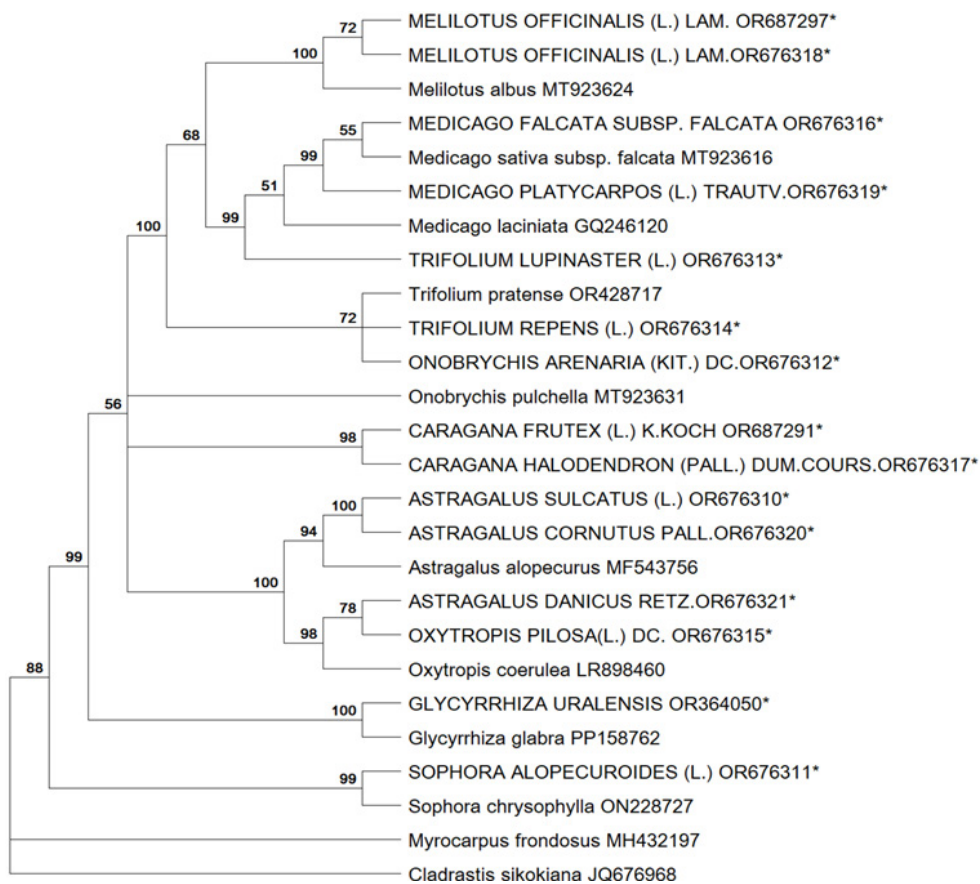
### Регион ITS

Топология, полученная в результате анализа методом максимального правдоподобия ML (Рис. 2) на основе 26 последовательностей ядерной ДНК региона ITS, в целом демонстрировала высокую статистическую поддержку, хотя некоторые ветви имели более низкую поддержку (<95% BS). Полученное дерево показывает, что таксоны разделены на несколько основных клад с различной степенью поддержки. Первая клада включает представителей родов *Trifolium*, *Melilotus* и *Medicago*, принадлежащих к трибе *Trifolieae*. Образец *Onobrychis arenaria* из трибы *Hedysareae* также попал в эту кладу, что, кроме того, согласно дереву ML, образец *Onobrychis arenaria* демонстрирует близкое родство к *Trifolium repens*, что подтверждается и анализом BLASTn в базе данных NCBI. Это объясняется тем, что роды *Trifolium*, *Melilotus* и *Medicago* являются близкородственными по морфологическим признакам (строение соцветий) и географическому происхождению. Род *Onobrychis* также достаточно таксономически близок к роду *Melilotus*, а также к родам *Oxytropis* и *Hedysarum*. Также известно, что *Onobrychis arenaria* имеет множество синонимов, включая *Hedysarum arenarium*, *Onobrychis sativa*, *Onobrychis vicifolia*, и *Onobrychis tanaitica*, что подчеркивает его сложную таксономическую историю. География распространения этого вида простирается от Европы и Кавказа до степных участков Сибири, Средней Азии и Казахстана. Широкий ареал обитания привел к высокой полиморфности вида.

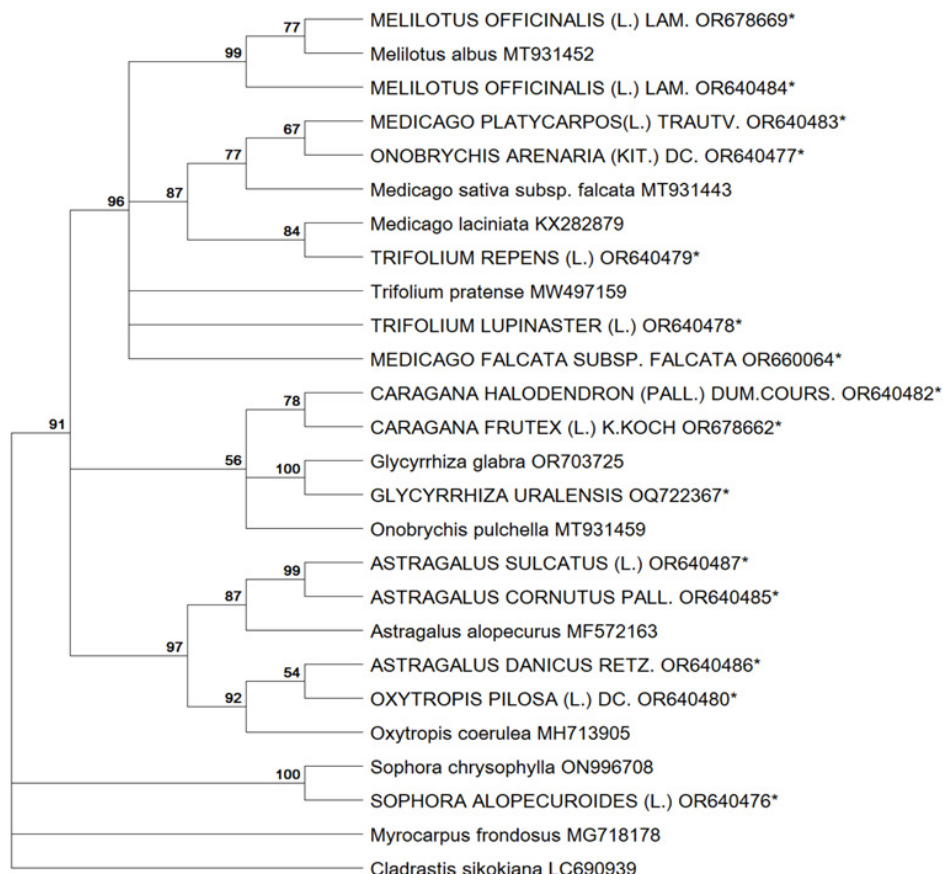
Вторая клада включает представители трибы *Galegeae* (род *Astragalus*, *Glycyrrhiza*, *Oxytropis*) и *Caraganeae*

**Таблица 3.** Характеристики выровненных последовательностей для анализов *rbcL*, *matK* и ITS

Параметры	Маркеры		
	<i>rbcL</i>	<i>matK</i>	ITS
Количество таксонов	15	15	15
Длина после выравнивания (bp)	476	612	495
Консервативные участки (C)	406	462	293
Вариативные сайты (V)	70	150	171
Потенциально информативные участки	42	56	114
Однонуклеотидные полиморфизмы (S)	28	94	57
Соотношение G + C (%)	43.9	31.8	51.8



**Рисунок 2** – Филогенетическое дерево, построенное методом maximum likelihood, на основе набора последовательностей ITS, с указанием значений бутстрепа (BS > 50%) над каждой ветвью. Звездочкой (\*) отмечены исследуемые казахстанские виды растений семейства бобовых.



**Рисунок 3** – Филогенетическое дерево, построенное методом maximum likelihood, на основе набора последовательностей *rbcL*, с указанием значений бутстрепа (BS > 50%) над каждой ветвью. Звездочкой (\*) отмечены исследуемые казахстанские виды растений семейства бобовых.

(*Caragana*). Следует отметить, что все исследуемые роды, за исключением рода *Sophora*, относятся к таксономической кладе, накапливающей не-белковые аминокислоты – NPAAA clade (Non-protein amino acid-accumulating clade). Род *Sophora* относится к трибе генистоид в широком смысле (*genistoids sensu lato*), что отчетливо видно на филогенетическом дереве (BS = 88%).

#### Регион *rbcL*

Для построения филогенетического дерева на основе более консервативного региона *rbcL* было использовано 26 последовательностей. Результаты показывают распределение последовательностей по родам со средней поддержкой (Рис. 3). Образцы рода *Sophora* были идентифицированы как представители отдельной трибы *Sophoreae* (BS = 100%). Далее дерево разделяется на две основные клады с высокой степенью поддержки (BS = 91%).

Первая клада включает представителей трибы *Galegeae* (род *Astragalus*, *Glycyrrhiza*, *Oxytropis*) и *Caraganeae* (*Caragana*). Образец *Oxytropis pilosa* демонстрирует близкое родство к представителям рода *Astragalus*.

Вторая клада с высокой поддержкой (BS = 96%), состоит из представителей трибы *Trifolieae*: роды *Trifolium*, *Melilotus* и *Medicago*. Однако, в отличие от дерева ITS, образец *Onobrychis arenaria* демонстрирует близкое родство к *Medicago platycarpus*. образец *Trifolium repens* показывает родство к *Medicago laciniata*.

#### Регион *matK*

Филогенетическое дерево, построенное на основе 26

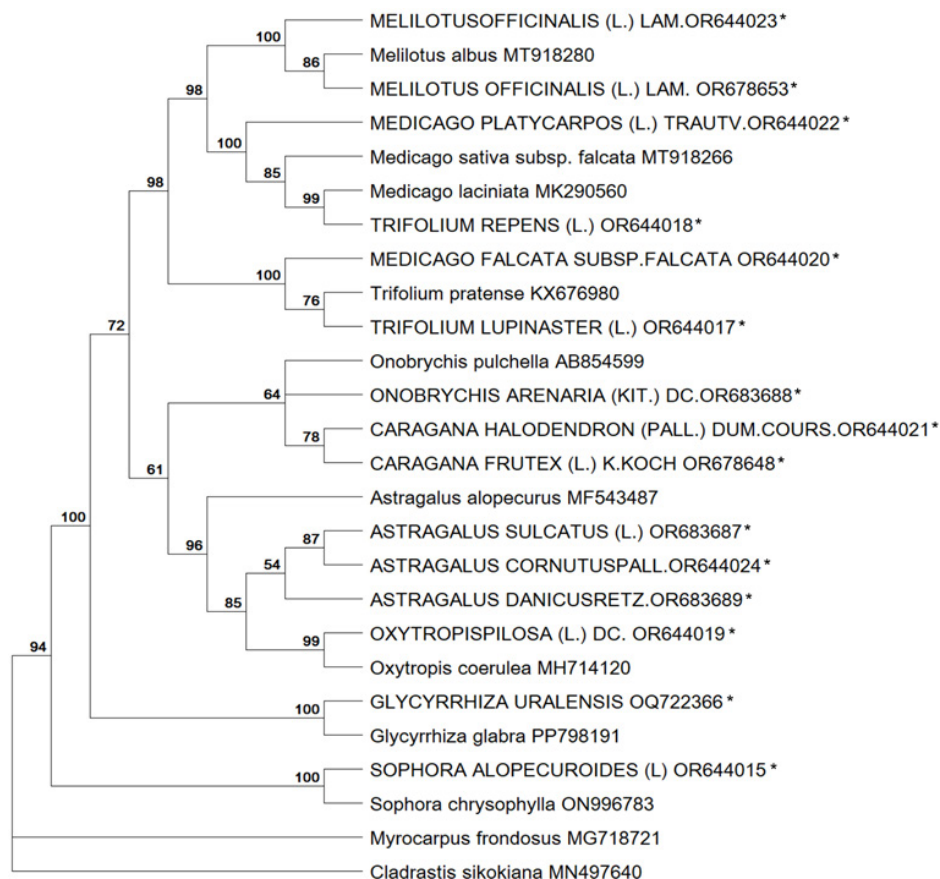
последовательностей *matK*, разделилось на две основные клады, представляющие трибы *Galegeae* и *Trifolieae*, а также отдельное ответвление образцов рода *Sophora*, представляющее трибу *Sophoreae* (BS = 100%) (Рис. 4). Образцы рода *Glycyrrhiza* также выделены в отдельную таксономическую кладу с высокой поддержкой (BS = 100%), формируя ответвление от трибы *Galegeae*.

В первой кладе четко сгруппированы представители трибы *Trifolieae*: роды *Trifolium*, *Melilotus* и *Medicago* (BS = 98%).

Структура второй клады включает экземпляры трибы *Galegeae* (роды *Astragalus*, *Glycyrrhiza*, *Oxytropis*) и *Caraganeae* (род *Caragana*) (BS = 61%). Образец *Oxytropis pilosa* демонстрирует близкое родство с представителями рода *Astragalus* (BS = 85%). Однако, в анализе региона *matK*, образец *Onobrychis arenaria* сгруппирован вместе с видами рода *Caragana* (BS = 64%).

### ОБСУЖДЕНИЕ

Метод ДНК-баркодирования полезен для точной идентификации видов и является важным первым шагом в планах по сохранению видов, находящихся под угрозой исчезновения. У растений для баркодирования чаще всего используют пластидную ДНК (районы *rbcL*, *matK*, *trnL* и *trnH-psbA*) и ядерную ДНК (регионы ITS1 и ITS2) [35]. В данном исследовании были проанализированы последовательности генов *matK*, *rbcL* и ITS с целью оценки их потенциала для точной идентификации представителей семейства *Fabaceae*.



**Рисунок 4** – Филогенетическое дерево, построенное методом maximum likelihood, на основе набора последовательностей *matK*, с указанием значений бутстрепа (BS > 50%) над каждой ветвью. Звездочкой (\*) отмечены исследуемые казахстанские виды растений семейства бобовых.

В ходе филогенетического анализа семейства бобовых (*Fabaceae*) на основе данных баркодирования по трем ДНК-маркерам мы обнаружили, что образец *Onobrychis arenaria* из трибы *Hedysareae* был размещен вместе с представителями родов *Trifolium*, *Melilotus* и *Medicago*, принадлежащих к трибе *Trifolieae*. Более того, сгенерированное дерево на основе данных анализа максимального правдоподобия (ML) подтверждает, что *Onobrychis arenaria* демонстрирует близкое родство с *Trifolium repens*. Эти результаты также подтверждаются анализом BLASTn в базе данных NCBI. Объясняется это тем, что роды *Trifolium*, *Melilotus* и *Medicago* являются близкородственными на основании строения соцветий и происхождения. В истории классификации были случаи, когда некоторые виды ботаники фенотипически относили как к роду *Melilotus* (донник), так и к роду *Medicago* (люцерна), что свидетельствует о таксономической близости этих родов. Род *Onobrychis* также достаточно таксономически близок к роду *Melilotus*, а также к родам *Oxytropis* и *Hedysarum* [36].

*Onobrychis arenaria* имеет множество синонимов, включая *Hedysarum arenarium*, *Onobrychis sativa*, *Onobrychis vicifolia*, и *Onobrychis tanaïtica*, что подчеркивает его сложную таксономическую историю. Этот вид отличается крайне широким распространением, встречаясь в Европе, на Кавказе, на степных участках Сибири, в Средней Азии и Казахстане. Первоначальный ареал группы видов, от которых произошел *Onobrychis arenaria*, приходится на территорию северной части Балканского полуострова. Широкий ареал обитания привел к высокой полиморфности вида [36].

Наши результаты показывают, что *matK* является наиболее подходящим локусом для идентификации видов внутри *Fabaceae*, и он особенно точен для некоторых родов, таких как *Melilotus*, *Caragana* и *Astragalus*. В исследовании [37] из 1355 образцов растений, собранных у 1079 видов, принадлежащих к 409 различным родам, *matK* точно идентифицировал примерно 80% на уровне вида и 96% на уровне рода. На основании характеристик отдельных баркодов и их комбинаций рекомендуется использовать ген *matK* как один из лучших баркодов для идентификации растений рода *Vicia* [38]. Таким образом, наши исследования подтверждают, что область *matK* является ценным маркером для идентификации видов растений внутри семейства *Fabaceae*.

Результаты наших исследований показывают, что область гена ITS у родов *Melilotus*, *Medicago* и *Caragana* является хорошим кандидатом для исследований по ДНК-баркодированию. Этот маркер проявляет высокую вариабельность и отлично подходит для определения родственных видов, что делает его популярным инструментом в филогенетических исследованиях и идентификации традиционных видов лекарственных растений [39]. Многие авторы подтверждают, что область ITS, являющаяся одним из основных регионов ДНК растений для баркодирования, позволяет различать большинство родов растений, даже сложных [40–43]. Например, маркер ITS рекомендуется для молекулярной идентификации лекарственных видов *Astragalus*, несмотря на сложности в морфологических различиях. Область ITS обычно полезна

для различения видов.

Наши данные показали, что область гена *rbcL* эффективно различает роды *Melilotus* и *Caragana*, однако обладает низким разрешением по сравнению с ITS и *matK*. Регион *rbcL* полезен для исследования филогении покрытосеменных, поскольку его легко амплифицировать и секвенировать в большинстве таксонов покрытосеменных, но его генетическая изменчивость относительно невелика [44]. Этот регион широко используют для филогенетического анализа, демонстрируя лучшие результаты на уровне семейства и класса, однако на уровне видов его способность распознавания умеренна [45–47]. В данном исследовании идентификация видов была наименее эффективной с использованием *rbcL* по сравнению с другими исследованными регионами. Молекулярная идентификация видов *Fabaceae* показала лучшие результаты при использовании гена *matK* по сравнению с *rbcL*, а использование мультилокусного подхода *matK+rbcL* улучшило эффективность идентификации [48].

## ЗАКЛЮЧЕНИЕ

Таким образом, наш анализ подтверждает, что *matK* и ITS являются наиболее эффективными маркерами для идентификации видов в семействе *Fabaceae*. Эти маркеры продемонстрировали высокую степень разрешающей способности, позволяя точно определять виды и роды внутри данного семейства. В частности, *matK* показал высокую эффективность в идентификации видов родов *Melilotus*, *Caragana* и *Astragalus*, а ITS проявил высокую вариабельность, что делает его полезным для определения родственных видов и проведения филогенетических исследований.

Ген *rbcL*, несмотря на его консервативность и легкость амплификации, продемонстрировал ограниченную эффективность при самостоятельном использовании для идентификации видов на уровне вида, но показал хорошие результаты при различении родов. Однако его низкая разрешающая способность по сравнению с *matK* и ITS указывает на необходимость его использования в комбинации с другими маркерами для повышения точности идентификации. Комбинация *matK* и *rbcL* улучшила эффективность молекулярной идентификации видов *Fabaceae*, что подтверждается нашими результатами.

Таким образом, для точной и надежной идентификации видов в семействе *Fabaceae* рекомендуется использовать мультилокусный подход, включающий маркеры *matK* и ITS, а также *rbcL* в комбинации с другими маркерами. Такой подход обеспечит высокую точность и надежность молекулярной идентификации, что является важным для сохранения биологического разнообразия и проведения дальнейших филогенетических исследований. Обнаружение близкого родства *Onobrychis arenaria* с представителями трибы *Trifolieae* в нашем филогенетическом анализе может быть объяснено исторической и морфологической близостью этих родов. Полиморфность и широкое распространение *Onobrychis arenaria* также могут способствовать его расположению в филогенетическом дереве рядом с видами, имеющими схожие эволюционные черты. Эти результаты подчеркивают необходимость комплексного подхода к таксономии, учитывающего как генетиче-

ские, так и морфологические данные.

## ФИНАНСИРОВАНИЕ

Исследование было профинансировано Министерством науки и высшего образования Республики Казахстан и выполнено в рамках научных программы ИРН: BR18574125.

## БЛАГОДАРНОСТИ

Мы выражаем благодарность коллективу ботаников Карагандинского государственного университета имени академика Е.А. Букетова под руководством Ишмуратовой М.Ю. за сбор и идентификацию образцов растений для ДНК-баркодирования.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов. Финансирующие организации не принимали участия в разработке исследования; в сборе, анализе или интерпретации данных; в написании рукописи; а также в принятии решения о публикации результатов.

## СПИСОК ЛИТЕРАТУРЫ

1. Аблаева А. Ф. Генетическое разнообразие семейства бобовых культур, собранных в национальных парках Северного Казахстана // *Wschodnioeuropejskie Czasopismo Naukowe*. – 2019. № 4-2. – С. 4-7.
2. Lewis G. P., Schrire B., Mackinder B., Lock M. *Legumes of the World* // Royal Botanic Gardens Kew. – 2005.
3. Сайт о грибах Казахстана // [[http://fungi.su/infusions/advanced\\_articles\\_sort/plant\\_family.php?sortby=Fabaceae](http://fungi.su/infusions/advanced_articles_sort/plant_family.php?sortby=Fabaceae)].
4. Zarre S., Azani N. Perspectives in taxonomy and phylogeny of the genus *Astragalus* (Fabaceae): a review // *Progress in biological sciences*. – 2013. – Vol. 3, № 1. – P. 1-6.
5. Khal L. H., Al-Anbari A. K., AL-Hadeethi M. A., Abdulrazaq R. T. Phylogenetic diversity of *Trifolium* L. species in Iraq // *Journal of Plant Production*. – 2020. – Vol. 11, № 11. – P. 1041-1044.
6. Muhammed D., Serkan A. Intercropping of legumes with Cereal crops in particular with the Perennials // *Perennial Crops for Food Security: Proceedings of the FAO Expert Workshop 28-30 August, 2013, Rome, Italy* – Food & Agriculture Org., 2018. – P. 221.
7. Graham P. H., Vance C. P. Legumes: importance and constraints to greater use // *Plant Physiol*. – 2003. – Vol. 131, № 3. – P. 872-7. <https://doi.org/10.1104/pp.017004>.
8. A.M. Danchenko, Kabanova S. A. Specially protected natural areas of the Republic of Kazakhstan and problems of conservation // *Conifers of the boreal zone*. – 2007. № 2-3. – P. 179-182.
9. Kress W. J., Erickson D. L. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region // *PLoS One*. – 2007. – Vol. 2, № 6. – P. 1-10. <https://doi.org/10.1371/journal.pone.0000508>.
10. Zahra N. B., Shinwari Z. K., Qaiser M. DNA barcoding: a tool for standardization of herbal medicinal products (HMPS) of Lamiaceae from Pakistan // *Pak. J. Bot*. – 2016. – Vol. 48, № 5. – P. 2167-2174.
11. Hebert P. D., Cywinska A., Ball S. L., deWaard J. R. Biological identifications through DNA barcodes // *Proc Biol Sci*. – 2003. – Vol. 270, № 1512. – P. 313-321. <https://doi.org/10.1098/rspb.2002.2218>.
12. Pettengill J. B., Neel M. C. An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae) // *American journal of botany*. – 2010. – Vol. 97, № 8. – P. 1391-1406.
13. Elansary H.O., Ashfaq M., Ali H. M., Yessoufou K. The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry // *PLoS One*. – 2017. – Vol. 12, № 2. – P. 1-15. <https://doi.org/10.1371/journal.pone.0172170>.
14. Han S., Sebastin R., Wang X., Lee K. J., Cho G. T., Hyun D. Y., Chung J. W. Identification of *Vicia* Species Native to South Korea Using Molecular and Morphological Characteristics // *Frontiers Plant Science*. – 2021. – Vol. 12. – P. 1-20. <https://doi.org/10.3389/fpls.2021.608559>.
15. Traiperm P., Chow J., Nopun P., Staples G., Swangpol S. C. Identification among morphologically similar *Argyrea* (Convolvulaceae) based on leaf anatomy and phenetic analyses // *Bot Stud*. – 2017. – Vol. 58, № 1. – P. 1-25. <https://doi.org/10.1186/s40529-017-0178-6>
16. Park I., Song J. H. *Cuscuta* Species Identification Based on the Morphology of Reproductive Organs and Complete Chloroplast Genome Sequences // . – 2019. – Vol. 20, № 11. – P. 1-15. <https://doi.org/10.3390/ijms20112726>.
17. Li Q. J., Wang X., Wang J. R., Su N., Zhang L., Ma Y. P., Chang Z. Y., Zhao L., Potter D. Efficient Identification of *Pulsatilla* (Ranunculaceae) Using DNA Barcodes and Micro-Morphological Characters // *Frontiers Plant Science*. – 2019. – Vol. 10. – P. 1-16. <https://doi.org/10.3389/fpls.2019.01196>.
18. Chakraborty C., Doss C. G. P., Patra B. C., Bandyopadhyay S. DNA barcoding to map the microbial communities: current advances and future directions // *Applied Microbiology and Biotechnology*. – 2014. – Vol. 98. – P. 3425-3436.
19. Park H. S., Jayakodi M., Lee S. H., Jeon J. H., Lee H. O., Park J. Y., Moon B. C., Kim C. K. Mitochondrial plastid DNA can cause DNA barcoding paradox in plants // . – 2020. – Vol. 10, № 1. – P. 1-12. <https://doi.org/10.1038/s41598-020-63233-y>.
20. Liu Z. F., Ci X. Q., Li L., Li H. W., Conran J. G., Li J. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China // *PLoS One*. – 2017. – Vol. 12, № 4. – P. 1-20. <https://doi.org/10.1371/journal.pone.0175788>.
21. Desalle R. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff // *Conserv Biol*. – 2006. – Vol. 20, № 5. – P. 1545-1557. <https://doi.org/10.1111/j.1523-1739.2006.00543.x>.
22. Jiang K.-W., Zhang R., Zhang Z.-F., Pan B., Tian B. DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage // *Plant*



Diversity. – 2020. – Vol. 42, № 5. – P. 376-385.

23. PlantNet. Plant identification and information website. – 2024. – URL: <https://plantnet.org/en/> (дата обращения: 26 фев. 2024).

24. Ma H.-l., Zhu Z.-b., Zhang X.-m., Miao Y.-y., Guo Q.-s. Species identification of the medicinal plant *Tulipa edulis* (Liliaceae) by DNA barcode marker // *Biochemical Systematics and Ecology*. – 2014. – Vol. 55. – P. 362-368. <https://doi.org/10.1016/j.bse.2014.03.038>.

25. Doyle J. J., Doyle J. L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue // *Phytochemical Bulletin*. – 1987. – Vol. 19, № 1. – P. 11-15.

26. Zhang C. Y., Wang F. Y., Yan H. F., Hao G., Hu C. M., Ge X. J. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae) // *Mol Ecol Resour*. – 2012. – Vol. 12, № 1. – P. 98-108. <https://doi.org/10.1111/j.1755-0998.2011.03076.x>.

27. Kuzmina M. L., Johnson K. L., Barron H. R., Hebert P. D. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library // *BMC Ecology*. – 2012. – Vol. 12, № 1. – P. 1-25. <https://doi.org/10.1186/1472-6785-12-25>.

28. Parmentier I., Dumnil J., Kuzmina M., Philippe M., Thomas D. W., Kenfack D., Chuyong G. B., Cruaud C., Hardy O. J. How effective are DNA barcodes in the identification of African rainforest trees? // *PLoS One*. – 2013. – Vol. 8, № 4. – P. 1-10. <https://doi.org/10.1371/journal.pone.0054921>.

29. Khan A. M., Bhadauria S. Molecular characterization of keratin degrading fungi isolated from semi-arid soil by PCR using ITS4 and ITS5 primers // *Journal of King Saud University - Science*. – 2019. – Vol. 31, № 4. – P. 1418-1423. <https://doi.org/10.1016/j.jksus.2018.04.014>.

30. Lu G., Moriyama E. Lu, G. & Moriyama, E.N. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief. Bioinform.* 5, 378-388 // *Briefings in bioinformatics*. – 2005. – Vol. 5. – P. 378-388. <https://doi.org/10.1093/bib/5.4.378>.

31. BLAST (Basic Local Alignment Search Tool) is a bioinformatics tool provided by the National Center for Biotechnology Information (NCBI) for comparing a query sequence against a database to find regions of similarity. – 2024. – URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (дата обращения: 26 2024).

32. GenBank. GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. – 2024. – URL: <https://www.ncbi.nlm.nih.gov/genbank/> (дата обращения: 26 фев.2024).

33. Tamura K., Stecher G., Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11 // *Molecular Biology and Evolution*. – 2021. – Vol. 38, № 7. – P. 3022-3027. <https://doi.org/10.1093/molbev/msab120>.

34. Alzohairy A. BioEdit: An important software for molecular biology // *GERF Bulletin of Biosciences*. – 2011. – Vol. 2. – P. 60-61.

35. Fazekas A. J., Kuzmina M. L., Newmaster S. G., Hollingsworth P. M. DNA barcoding methods for land plants // *Methods Mol Biol*. – 2012. – Vol. 858. – P. 223-252. [https://doi.org/10.1007/978-1-61779-591-6\\_11](https://doi.org/10.1007/978-1-61779-591-6_11).

36. Baker W.J., Bailey P., Barber V., Barker A., Bellot S., Bishop D., Botigue L.R., Brewer G., Carruthers T., Clarkson

J.J., Cook J., Cowan R.S., Dodsworth S., Epitawalage N., Francoso E., Gallego B., Johnson M., Kim J.T., Leempoel K., Maurin O., McGinnie C., Pokorny L., Roy S., Stone M., Toledo E., Wickett N.J., Zuntini A.R., Eiserhardt W.L., Kersey P.J., Leitch I.J. & Forest F. 2022. A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*. – Vol. 71. – P. 301–319. <https://doi.org/10.1093/sysbio/syab035>.

37. Gao T., Sun Z., Yao H., Song J., Zhu Y., Ma X., Chen S. Identification of Fabaceae plants using the DNA barcode matK // *Planta Med*. – 2011. – Vol. 77, № 1. – P. 92-94. <https://doi.org/10.1055/s-0030-1250050>.

38. Wu F. F., Gao Q., Liu F., Wang Z., Wang J. L., Wang X. G. DNA barcoding evaluation of *Vicia* (Fabaceae): Comparative efficacy of six universal barcode loci on abundant species // *Journal of systematics and evolution*. – 2020. – Vol. 58, № 1. – P. 77-88.

39. Yao H., Song J., Liu C., Luo K., Han J., Li Y., Pang X., Xu H., Zhu Y., Xiao P., Chen S. Use of ITS2 region as the universal DNA barcode for plants and animals // *PLoS One*. – 2010. – Vol. 5, № 10. – P. 1-9. <https://doi.org/10.1371/journal.pone.0013102>.

40. Qiang Z., Yu-long D., Chen X., Hui-yu Z., Min-ren H., Ming-xiu W. A preliminary analysis of phylogenetic relationships of Arundinaria and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trn L-F intergenic spacer) // *Journal of Forestry Research*. – 2005. – Vol. 16, № 1. – P. 5-8.

41. Doh E. J., Kim J.-H., Oh S. E., Lee G. Identification and monitoring of Korean medicines derived from *Cinnamomum* spp. by using ITS and DNA marker // *Genes & Genomics*. – 2017. – Vol. 39. – P. 101-109.

42. Wu F., Ma J., Meng Y., Zhang D., Pascal Muvunyi B., Luo K., Di H., Guo W., Wang Y., Feng B. Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations // *PLoS One*. – 2017. – Vol. 12, № 9. – P. 1-16.

43. Zhang D., Jiang B. Species identification in complex groups of medicinal plants based on DNA barcoding: a case study on *Astragalus* spp.(Fabaceae) from southwest China // *Conservation genetics resources*. – 2020. – Vol. 12, № 3. – P. 469-478.

44. Li X., Yang Y., Henry R. J., Rossetto M., Wang Y., Chen S. Plant DNA barcoding: from gene to genome // *Biol Rev Camb Philos Soc*. – 2015. – Vol. 90, № 1. – P. 157-166. <https://doi.org/10.1111/brv.12104>.

45. Li X., Yang Y., Henry R. J., Rossetto M., Wang Y., Chen S. Plant DNA barcoding: from gene to genome // *Biological Reviews*. – 2015. – Vol. 90, № 1. – P. 157-166.

46. Hollingsworth P. M., Forrest L. L., Spouge J. L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M. W., Cowan R. S., Erickson D. L. A DNA barcode for land plants // *Proceedings of the National Academy of Sciences*. – 2009. – Vol. 106, № 31. – P. 12794-12797.

47. Chen S., Yao H., Han J., Liu C., Song J., Shi L., Zhu Y., Ma X., Gao T., Pang X. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species // *PLoS one*. – 2010. – Vol. 5, № 1. – P. 1-8.

48. Wardani I. G. A. K., Amandita F. Y., de Melo Moura C. C., Gailing O., Siregar I. Z. Molecular taxonomy via DNA barcodes for species identification in selected genera of Fabaceae // *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*. – 2022. – Vol. 12, № 1. – P. 112-122.

## REFERENCES

1. Ablava A. G. Genetic diversity of the legume family collected in national parks of Northern Kazakhstan / Ablayeva A. G. Geneticheskoye raznoobrazie semeystva bobovykh kul'tur, sobrannykh v natsional'nykh parkakh Severnogo Kazakhstan // *Wschodnioeuropejskie Czasopismo Naukowe*. – 2019. № 4-2. – P. 4-7. [In Russian].
2. Lewis G. P., Schrire B., Mackinder B., Lock M. Legumes of the World // *Royal Botanic Gardens Kew*. – 2005.
3. Website about mushrooms of Kazakhstan // Sayt o gribakh Kazakhstana  
[[http://fungi.su/infusions/advanced\\_articles\\_sort/plant\\_family.php?sortby=Fabaceae](http://fungi.su/infusions/advanced_articles_sort/plant_family.php?sortby=Fabaceae)]. [In Russian].
4. Zarre S., Azani N. Perspectives in taxonomy and phylogeny of the genus *Astragalus* (Fabaceae): a review // *Progress in biological sciences*. – 2013. – Vol. 3, № 1. – P. 1-6.
5. Khal L. H., Al-Anbari A. K., AL-Hadeethi M. A., Abdulrazaq R. T. Phylogenetic diversity of *Trifolium* L. species in Iraq // *Journal of Plant Production*. – 2020. – Vol. 11, № 11. – P. 1041-1044.
6. Muhammed D., Serkan A. Intercropping of legumes with Cereal crops in particular with the Perennials // *Perennial Crops for Food Security: Proceedings of the FAO Expert Workshop 28-30 August, 2013, Rome, Italy – Food & Agriculture Org.*, 2018. – P. 221.
7. Graham P. H., Vance C. P. Legumes: importance and constraints to greater use // *Plant Physiol*. – 2003. – Vol. 131, № 3. – P. 872-7. <https://doi.org/10.1104/pp.017004>.
8. A.M. Danchenko, Kabanova S. A. Specially protected natural areas of the Republic of Kazakhstan and problems of conservation // *Conifers of the boreal zone*. – 2007. № 2-3. – P. 179-182.
9. Kress W. J., Erickson D. L. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region // *PLoS One*. – 2007. – Vol. 2, № 6. – P. 1-10. <https://doi.org/10.1371/journal.pone.0000508>.
10. Zahra N. B., Shinwari Z. K., Qaiser M. DNA barcoding: a tool for standardization of herbal medicinal products (HMPS) of Lamiaceae from Pakistan // *Pak. J. Bot*. – 2016. – Vol. 48, № 5. – P. 2167-2174.
11. Hebert P. D., Cywinska A., Ball S. L., deWaard J. R. Biological identifications through DNA barcodes // *Proc Biol Sci*. – 2003. – Vol. 270, № 1512. – P. 313-321. <https://doi.org/10.1098/rspb.2002.2218>.
12. Pettengill J. B., Neel M. C. An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae) // *American journal of botany*. – 2010. – Vol. 97, № 8. – P. 1391-1406.
13. Elansary H.O., Ashfaq M., Ali H. M., Yessoufou K. The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry // *PLoS One*. – 2017. – Vol. 12, № 2. – P. 1-15. <https://doi.org/10.1371/journal.pone.0172170>.
14. Han S., Sebastin R., Wang X., Lee K. J., Cho G. T., Hyun D. Y., Chung J. W. Identification of *Vicia* Species Native to South Korea Using Molecular and Morphological Characteristics // *Frontiers Plant Science*. – 2021. – Vol. 12. – P. 1-20. <https://doi.org/10.3389/fpls.2021.608559>.
15. Traiperm P., Chow J., Nopun P., Staples G., Swangpol S. C. Identification among morphologically similar *Argyrea* (Convolvulaceae) based on leaf anatomy and phenetic analyses // *Bot Stud*. – 2017. – Vol. 58, № 1. – P. 1-25. <https://doi.org/10.1186/s40529-017-0178-6>
16. Park I., Song J. H. *Cuscuta* Species Identification Based on the Morphology of Reproductive Organs and Complete Chloroplast Genome Sequences // . – 2019. – Vol. 20, № 11. – P. 1-15. <https://doi.org/10.3390/ijms20112726>.
17. Li Q. J., Wang X., Wang J. R., Su N., Zhang L., Ma Y. P., Chang Z. Y., Zhao L., Potter D. Efficient Identification of *Pulsatilla* (Ranunculaceae) Using DNA Barcodes and Micro-Morphological Characters // *Frontiers Plant Science*. – 2019. – Vol. 10. – P. 1-16. <https://doi.org/10.3389/fpls.2019.01196>.
18. Chakraborty C., Doss C. G. P., Patra B. C., Bandyopadhyay S. DNA barcoding to map the microbial communities: current advances and future directions // *Applied Microbiology and Biotechnology*. – 2014. – Vol. 98. – P. 3425-3436.
19. Park H. S., Jayakodi M., Lee S. H., Jeon J. H., Lee H. O., Park J. Y., Moon B. C., Kim C. K. Mitochondrial plastid DNA can cause DNA barcoding paradox in plants // . – 2020. – Vol. 10, № 1. – P. 1-12. <https://doi.org/10.1038/s41598-020-63233-y>.
20. Liu Z. F., Ci X. Q., Li L., Li H. W., Conran J. G., Li J. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China // *PLoS One*. – 2017. – Vol. 12, № 4. – P. 1-20. <https://doi.org/10.1371/journal.pone.0175788>.
21. Desalle R. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff // *Conserv Biol*. – 2006. – Vol. 20, № 5. – P. 1545-1557. <https://doi.org/10.1111/j.1523-1739.2006.00543.x>.
22. Jiang K.-W., Zhang R., Zhang Z.-F., Pan B., Tian B. DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage // *Plant Diversity*. – 2020. – Vol. 42, № 5. – P. 376-385.
23. PlantNet. Plant identification and information website. – 2024. – URL: <https://plantnet.org/en/> (date of the application: 26 Feb. 2024).
24. Ma H.-l., Zhu Z.-b., Zhang X.-m., Miao Y.-y., Guo Q.-s. Species identification of the medicinal plant *Tulipa edulis* (Liliaceae) by DNA barcode marker // *Biochemical Systematics and Ecology*. – 2014. – Vol. 55. – P. 362-368. <https://doi.org/10.1016/j.bse.2014.03.038>.
25. Doyle J. J., Doyle J. L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue // *Phytochemical Bulletin*. – 1987. – Vol. 19, № 1. – P. 11-15.

26. Zhang C. Y., Wang F. Y., Yan H. F., Hao G., Hu C. M., Ge X. J. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae) // *Mol Ecol Resour.* – 2012. – Vol. 12, № 1. – P. 98-108. <https://doi.org/10.1111/j.1755-0998.2011.03076.x>.
27. Kuzmina M. L., Johnson K. L., Barron H. R., Hebert P. D. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library // *BMC Ecology.* – 2012. – Vol. 12, № 1. – P. 1-25. <https://doi.org/10.1186/1472-6785-12-25>.
28. Parmentier I., Dumnil J., Kuzmina M., Philippe M., Thomas D. W., Kenfack D., Chuyong G. B., Cruaud C., Hardy O. J. How effective are DNA barcodes in the identification of African rainforest trees? // *PLoS One.* – 2013. – Vol. 8, № 4. – P. 1-10. <https://doi.org/10.1371/journal.pone.0054921>.
29. Khan A. M., Bhadauria S. Molecular characterization of keratin degrading fungi isolated from semi-arid soil by PCR using ITS4 and ITS5 primers // *Journal of King Saud University - Science.* – 2019. – Vol. 31, № 4. – P. 1418-1423. <https://doi.org/10.1016/j.jksus.2018.04.014>.
30. Lu G., Moriyama E. Lu, G. & Moriyama, E.N. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief. Bioinform.* 5, 378-388 // *Briefings in bioinformatics.* – 2005. – Vol. 5. – P. 378-388. <https://doi.org/10.1093/bib/5.4.378>.
31. BLAST (Basic Local Alignment Search Tool) is a bioinformatics tool provided by the National Center for Biotechnology Information (NCBI) for comparing a query sequence against a database to find regions of similarity. – 2024. – URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (дата обращения: 26 2024).
32. GenBank. GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. – 2024. – URL: <https://www.ncbi.nlm.nih.gov/genbank/> (date of the application: 26 Feb..2024).
33. Tamura K., Stecher G., Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11 // *Molecular Biology and Evolution.* – 2021. – Vol. 38, № 7. – P. 3022-3027. <https://doi.org/10.1093/molbev/msab120>.
34. Alzohairy A. BioEdit: An important software for molecular biology // *GERF Bulletin of Biosciences.* – 2011. – Vol. 2. – P. 60-61.
35. Fazekas A. J., Kuzmina M. L., Newmaster S. G., Hollingsworth P. M. DNA barcoding methods for land plants // *Methods Mol Biol.* – 2012. – Vol. 858. – P. 223-252. [https://doi.org/10.1007/978-1-61779-591-6\\_11](https://doi.org/10.1007/978-1-61779-591-6_11).
36. Baker W.J., Bailey P., Barber V., Barker A., Bellot S., Bishop D., Botigue L.R., Brewer G., Carruthers T., Clarkson J.J., Cook J., Cowan R.S., Dodsworth S., Epiawalage N., Francoso E., Gallego B., Johnson M., Kim J.T., Leempoel K., Maurin O., McGinnie C., Pokorny L., Roy S., Stone M., Toledo E., Wickett N.J., Zuntini A.R., Eiserhardt W.L., Kersey P.J., Leitch I.J. & Forest F. 2022. A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology.* – Vol. 71. – P. 301–319. <https://doi.org/10.1093/sysbio/syab035>.
37. Gao T., Sun Z., Yao H., Song J., Zhu Y., Ma X., Chen S. Identification of Fabaceae plants using the DNA barcode matK // *Planta Med.* – 2011. – Vol. 77, № 1. – P. 92-94. <https://doi.org/10.1055/s-0030-1250050>.
38. Wu F. F., Gao Q., Liu F., Wang Z., Wang J. L., Wang X. G. DNA barcoding evaluation of *Vicia* (Fabaceae): Comparative efficacy of six universal barcode loci on abundant species // *Journal of systematics and evolution.* – 2020. – Vol. 58, № 1. – P. 77-88.
39. Yao H., Song J., Liu C., Luo K., Han J., Li Y., Pang X., Xu H., Zhu Y., Xiao P., Chen S. Use of ITS2 region as the universal DNA barcode for plants and animals // *PLoS One.* – 2010. – Vol. 5, № 10. – P. 1-9. <https://doi.org/10.1371/journal.pone.0013102>.
40. Qiang Z., Yu-long D., Chen X., Hui-yu Z., Min-ren H., Ming-xiu W. A preliminary analysis of phylogenetic relationships of *Arundinaria* and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trn LF intergenic spacer) // *Journal of Forestry Research.* – 2005. – Vol. 16, № 1. – P. 5-8.
41. Doh E. J., Kim J.-H., Oh S. E., Lee G. Identification and monitoring of Korean medicines derived from *Cinnamomum* spp. by using ITS and DNA marker // *Genes & Genomics.* – 2017. – Vol. 39. –P. 101-109.
42. Wu F., Ma J., Meng Y., Zhang D., Pascal Muvunyi B., Luo K., Di H., Guo W., Wang Y., Feng B. Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations // *PLoS One.* – 2017. – Vol. 12, № 9. – P. 1-16.
43. Zhang D., Jiang B. Species identification in complex groups of medicinal plants based on DNA barcoding: a case study on *Astragalus* spp.(Fabaceae) from southwest China // *Conservation genetics resources.* – 2020. – Vol. 12, № 3. – P. 469-478.
44. Li X., Yang Y., Henry R. J., Rossetto M., Wang Y., Chen S. Plant DNA barcoding: from gene to genome // *Biol Rev Camb Philos Soc.* – 2015. – Vol. 90, № 1. – P. 157-166. <https://doi.org/10.1111/brv.12104>.
45. Li X., Yang Y., Henry R. J., Rossetto M., Wang Y., Chen S. Plant DNA barcoding: from gene to genome // *Biological Reviews.* – 2015. – Vol. 90, № 1. – P. 157-166.
46. Hollingsworth P. M., Forrest L. L., Spouge J. L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M. W., Cowan R. S., Erickson D. L. A DNA barcode for land plants // *Proceedings of the National Academy of Sciences.* – 2009. – Vol. 106, № 31. – P. 12794-12797.
47. Chen S., Yao H., Han J., Liu C., Song J., Shi L., Zhu Y., Ma X., Gao T., Pang X. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species // *PloS one.* – 2010. – Vol. 5, № 1. – P. 1-8.
48. Wardani I. G. A. K., Amandita F. Y., de Melo Moura C. C., Gailing O., Siregar I. Z. Molecular taxonomy via DNA barcodes for species identification in selected genera of Fabaceae // *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management).* – 2022. – Vol. 12, № 1. – P. 112-122.

UDC: 575.174.015.3; 575.86

IDENTIFICATION OF PLANTS OF THE FAMILY *FABACEAE* USING MOLECULAR BARCODING ANALYSISKali B.R.<sup>1</sup>, Sutula M.Y.<sup>1</sup>, Akhmetollayeva A.S.<sup>1</sup>, Manabayeva S.A.<sup>1,2\*</sup><sup>1</sup> National Center for Biotechnology, 13/5, Qorghalzhyn Hwy., Astana, 010000, Kazakhstan<sup>2</sup> L.N.Gumilyov Eurasian National University, Satpayev st. 2, Astana, 010000, Kazakhstan Corresponding author:

\*manabayeva@biocenter.kz

## ABSTRACT

The *Fabaceae* family in Kazakhstan comprises 42 genera and about 650 species. Leguminous crops, such as beans, peas, soybeans, and peas, as well as forage plants like clover, alfalfa, sainfoin, and sweet clover, are rich in protein and widely used in the food and feed industries. They are also utilized in the production of biodegradable materials, oils, dyes, and biodiesel fuel. DNA barcoding is effective for species identification within *Fabaceae* despite their morphological variability, highlighting the importance of preserving their diversity. To assess the genetic diversity of 15 *Fabaceae* species in the Karaganda region, three pairs of primers were used. The chloroplast markers *matK* and *rbcL*, as well as the nuclear marker ITS, were employed. The obtained sequences were analyzed using Vector NTI and MEGA11, registered in the GenBank NCBI database, and voucher specimens are stored in the herbarium of the National Center for Biotechnology in Astana. The study demonstrates the effectiveness of DNA barcoding for species identification of *Fabaceae* based on the analysis of ITS, *rbcL*, and *matK* markers using the maximum likelihood method. The results underscore the significance of selecting markers considering their variability for accurate phylogenetic analysis and the necessity of further population studies to preserve biological diversity in Kazakhstan.

**Keywords:** *Fabaceae*, DNA barcoding, *matK*, *rbcL*, ITS

ӘОК: 575.174.015.3; 575.86

МОЛЕКУЛАЛЫҚ БАРКОДТАУ ТАЛДАУЫН ПАЙДАЛАНУ АРҚЫЛЫ *FABACEAE* ТҰҚЫМДАСЫНЫҢ ӨСІМДІКТЕРІН АНЫҚТАУҚали Б.Р.<sup>1</sup>, Сутула М.Ю.<sup>1</sup>, Ахметоллаева А.С.<sup>1</sup>, Манабаева Ш.А.<sup>1,2\*</sup><sup>1</sup> Ұлттық биотехнология орталығы, Қорғалжын көшесі 13/5, Астана, 010000, Қазақстан<sup>2</sup> Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Сәтбаев көшесі 2, Астана, 010000, Қазақстан

Корреспондент автор: \*manabayeva@biocenter.kz

## АБСТРАКТ

*Fabaceae* тұқымдасы Қазақстанда 42 туыс және 650-ге жуық түрі бар. Бұршақ, фасоль соя, сияқты бұршақ дақылдары, сондай – ақ жемшөп өсімдіктері – беде, жоңышқа, эспарцет пен донник-ақуызға бай және тамақ және жем өнеркәсібінде кеңінен қолданылады. Олар сондай-ақ биологиялық ыдырайтын материалдарды, майларды, бояғыштарды және биодизельді өндіру үшін қолданылады. ДНҚ баркодтау морфологиялық өзгергіштігіне қарамастан *Fabaceae* түрлерін анықтауда тиімді, бұл олардың әртүрлілігін сақтаудың маңыздылығын көрсетеді. Қарағанды облысындағы *Fabaceae* тұқымдасының 15 түрінің генетикалық әртүрлілігін бағалау үшін үш жұп праймер қолданылды. *matK* және *rbcL* хлоропласт маркерлері, сондай-ақ ITS ядролық маркері қолданылды. Алынған тізбектер Vector NTI және MEGA11-де талданды, GenBank NCBI дерекқорында тіркелді, ал ваучерлік үлгілер Астанадағы Ұлттық биотехнология орталығының гербарийінде сақталады. Зерттеу maximum likelihood әдісімен ITS, *rbcL* және *matK* маркерлерін талдау негізінде *Fabaceae* түрлерін анықтау үшін ДНҚ баркодтау тиімділігін көрсетеді. Нәтижелер нақты филогенетикалық талдау үшін олардың өзгергіштігін ескере отырып, маркерлерді таңдаудың маңыздылығын және Қазақстанда биологиялық әртүрлілікті сақтау үшін одан әрі популяциялық зерттеулердің қажеттілігін көрсетеді.

**Түйін сөздер:** *Fabaceae*, ДНҚ баркодтау, *matK*, *rbcL*, ITS