

## КРАТКИЕ СООБЩЕНИЯ

Научная статья  
УДК 575:633.34:581.1  
DOI: 10.30901/2227-8834-2026-2-04



## Анализ *in silico* геномных районов сои, ассоциированных с содержанием белка в семенах

И. В. Розанова<sup>1</sup>, П. В. Климова<sup>1</sup>, А. Я. Евлаш<sup>1,2</sup>, И. В. Сеферова<sup>2</sup>

<sup>1</sup> Научно-технологический университет «Сириус», Центр генетики и наук о жизни, Краснодарский край, Россия

<sup>2</sup> Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, Санкт-Петербург, Россия

Автор, ответственный за переписку: Ирина Вениаминовна Розанова, rozanova.iv@talantiuspeh.ru

Соя (*Glycine max* (L.) Merr.) – один из наиболее важных объектов мирового сельского хозяйства. Благодаря высокому содержанию белка, соя активно возделывается во многих странах. На основании литературного анализа и анализа доступных онлайн-ресурсов, содержащих полногеномные данные разных сортов, были исследованы последовательности генов, которые кодируют белки аспарагинсинтетазу (AS) и аспарагиназу (ASPG), ассоциированные с содержанием белка в семенах сои. Проведено выравнивание исследуемых последовательностей в сортах 'Jack', 'Enrei', 'Williams 82' и др., выявлены SNP и InDel. Сконструированы праймеры для стандартной ПЦР и для ДНК-штрихкодирования. Также подобрана выборка из 23 образцов сои из коллекции ВИР, которая показала контрастное содержание белка и может быть использована в дальнейшем изучении и верификации.

**Ключевые слова:** соя, *Glycine max*, пангеном, аспарагиназа, аспарагинсинтетаза

**Благодарности:** работа выполнена в рамках государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории “Сириус”» (Соглашение №18-03 от 10.09.2024). Формирование выборки сои осуществлено в рамках тематического плана ВИР по проекту № FGEM-2022-0002.

**Для цитирования:** Розанова И.В., Климова П.В., Евлаш А.Я., Сеферова И.В. Анализ *in silico* геномных районов сои, ассоциированных с содержанием белка в семенах. *Труды по прикладной ботанике, генетике и селекции*. 2026;187(2):226-234. DOI: 10.30901/2227-8834-2026-2-04

Прозрачность финансовой деятельности: авторы не имеют финансовой заинтересованности в представленных материалах или методах. Авторы благодарят рецензентов за их вклад в экспертную оценку этой работы. Мнение журнала нейтрально к изложенным материалам, авторам и их месту работы.

## BRIEF REPORTS

Original article

DOI: 10.30901/2227-8834-2026-2-04

## ***In silico* analysis of soybean genomic regions associated with seed protein content**

**Irina V. Rozanova<sup>1</sup>, Polina V. Klimova<sup>1</sup>, Anastasia Ya. Evlash<sup>1,2</sup>, Irina V. Seferova<sup>2</sup>**<sup>1</sup> *Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, Krasnodar Territory, Russia*<sup>2</sup> *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia***Corresponding author:** Irina V. Rozanova, [rozanova.iv@talantiuspeh.ru](mailto:rozanova.iv@talantiuspeh.ru)

Soybean (*Glycine max* (L.) Merr.) is one of the most important crops in global agriculture. Due to its high protein content, soybean is actively cultivated in many countries. Upon reviewing published sources and analyzing available online resources containing whole-genome data for different cultivars, the sequences of genes encoding asparagine synthetase (AS) and asparaginase (ASPG) proteins associated with protein content in soybean seeds were analyzed. Multiple sequence alignments were performed for the target sequences in cvs. 'Jack', 'Enrei', 'Williams 82', and others, revealing SNPs and InDels. Primers were designed for standard PCR and DNA barcoding. Additionally, a panel of 23 soybean accessions from the VIR collection was selected, exhibiting contrasting protein content, which can be used for further studies and validation.

**Keywords:** soybean, *Glycine max*, pangenome, asparaginase, asparagine synthetase

**Acknowledgments:** this study was conducted with financial support from the state program for the Sirius Federal Territory "Scientific and technological development of the Sirius Federal Territory" (Agreement No. 18-03 dated Sept. 10, 2024). Soybean accessions were sampled within the framework of the thematic plan of VIR, Project No. FGEM-2022-0002.

**For citation:** Rozanova I.V., Klimova P.V., Evlash A.Ya., Seferova I.V. *In silico* analysis of soybean genomic regions associated with seed protein content. *Proceedings on Applied Botany, Genetics and Breeding*. 2026;187(2):226-234. (In Russ.). DOI: 10.30901/2227-8834-2026-2-04

Financial transparency: the authors have no financial interest in the presented materials or methods. The authors thank the reviewers for their contribution to the peer review of this work. The journal's opinion is neutral to the presented materials, the authors or their employers.

## Введение

Соя (*Glycine max* (L.) Merr.) – один из наиболее значимых объектов мирового сельского хозяйства, универсальная культура с высокой питательной ценностью и широким спектром применения. Благодаря высокому содержанию белка (30–52%), соя активно возделывается во многих странах как продукт питания, а также используется для кормления животных и в технических целях (Gaffield et al., 2024). По содержанию белка она превосходит все другие возделываемые культуры, включая горох (22–24%) (Emkani et al., 2023), подсолнечник (25%) (Petraru et al., 2021), пшеницу (9–26%) (Cao et al., 2021), рис (7%) (Fukai, Mitchell, 2024) и др. Она содержит все девять незаменимых аминокислот, что делает ее ключевым источником растительного белка и важным элементом продовольственной безопасности, особенно в регионах с недостатком животного белка (Guo et al., 2022). Пищевая ценность соевого белка определяется высоким содержанием незаменимых аминокислот, доля которых составляет около 20% от общего содержания белка, тогда как для пшеницы этот показатель не превышает 18% (Gorissen et al., 2018). По содержанию полиаминов, важных для клеточного деления и защиты от окислительного стресса, соевые бобы занимают второе место среди культур – после кукурузы (Egogov et al., 2021).

Исследования генетических механизмов, контролирующих содержание белка в сое, активно развиваются. Однако на сегодняшний момент нет четкого понимания того, как в семенах сои регулируется содержание белка. QTL, ассоциированные с признаком «содержание белка», распределены по всем хромосомам, однако их интервалы остаются слишком широкими для точного выделения генов-кандидатов. Например, для локуса на хромосоме 20, ассоциированного с высоким содержанием белка, область была последовательно уточнена с 8,4 до 1 Мб (Hwang et al., 2014; Vaughn et al., 2014; Bandillo et al., 2015; Zhang Q. et al., 2023), но все еще остается слишком обширной для выделения конкретных регуляторных генов. Известно, что содержание белка в семенах сои проявляет сильную зависимость от условий получения репродукции семян (Novikova et al., 2018). Некоторые сорта демонстрируют высокий уровень белка при выращивании в определенных географических регионах: так, генотипы PI407788A и 'Enrei' содержат аллель (CC+) гена *GmSWEET39* на хромосоме 15, ассоциированный с повышенным содержанием белка, в Южной Корее и Японии соответственно (Brzostowski, Diers, 2017; Zhang H. et al., 2020).

Сложность изучения механизма наследования у сои объясняется полиплоидизацией генома сои и перестройкой гомологичных хромосом, что ведет к дублированию локусов и усложняет картирование. Современные методы исследования, объединяющие данные QTL-картирования с анализами РНК-секвенирования, позволяют выявлять структурные вариации (SNP, InDel) между генотипами с контрастным содержанием белка, что способствует идентификации ряда генов-кандидатов.

Существуют работы, описывающие молекулярные регуляторы синтеза белков. Так, последовательность, кодирующая транспортер сахарозы *GmSWEET39*, содержит аллель (CC+) на хромосоме 15, наличие которой способствует снабжению развивающихся семян промежуточными продуктами метаболизма, влияя как на уровень белка, так и масла. Данная делеция приводит к укорочению белкового транспортера и изменению соотношения

этих компонентов (Zhang H. et al., 2020). Высокая экспрессия аспарагинсинтетазы и аспарагиназы положительно коррелирует с повышенным содержанием белка в семенах (Wan et al., 2006; Pandurangan et al., 2012). Регуляторные белки, такие как NF-YC4 и ген арабидопсиса *Qua-Quine Starch (QQS)*, также задействованы в контроле соотношения белка и масла: эктопическая экспрессия *QQS* повышает содержание белка на ~20% при одновременном снижении масла на ~10%, тогда как сверхэкспрессия его партнера *GmNF-YC4-2 (Glyma.04g196200)* повышает как белковость семян, так и урожайность.

Главным регулятором синтеза запасных белков выступает *GmLEC2*. Эктопическая экспрессия *LEC2a* в *Arabidopsis thaliana* приводит к увеличению доли белка, тогда как мутации с утратой функции *Atlec2* снижают этот показатель примерно на 40%. Похожие эффекты наблюдаются при экспрессии генов других аминокислотных путей. Например, устойчивый к отрицательной обратной связи вариант антранилатсинтазы риса (OASA1D) повышает уровень триптофана в семенах без подавления роста растений (Ishimoto et al., 2009; Kita et al., 2010). Аналогичная стратегия применена для метионина: экспрессия нечувствительной к регуляции формы цистатионин-γ-синтазы (CGS) из *Arabidopsis* увеличивает его уровень в семенах без снижения урожайности (Yu et al., 2018). Применение нечувствительных ферментов дигидродипиколинатсинтазы (DHDP) из *Escherichia coli* и аспартокиназы (AK) из *Corynebacterium* позволило увеличить содержание лизина более чем в пять раз.

Полногеномная последовательность сои впервые была получена в 2010 г. при секвенировании сорта 'Williams 82' с использованием технологии коротких прочтений (Schmutz et al., 2010). В том же году секвенировали геном дикорастущей сои (*G. soja* Siebold et Zucc.) и провели его детальное сравнение с культурной (*G. max*). Было показано, что генетическое разнообразие *G. soja* выше, чем у культурной сои, что подчеркнуло ценность диких форм как источника новых аллелей для селекции на устойчивость и продуктивность. Современные исследования геномной изменчивости выходят за рамки анализа одиночных геномов, переходя к формированию пангеномов – объединенных баз данных генетических вариаций внутри вида.

В рамках данного исследования проведен анализ *in silico* в пангеноме аллельных вариантов генов, кодирующих аспарагинсинтетазу (AS) и аспарагиназу (ASPG).

## Материалы и методы

### Анализ *in silico* потенциальных генов-мишеней

Поиск последовательностей генов белков аспарагинсинтетазы и аспарагиназы проводили с использованием алгоритма BLASTN в базе данных NCBI в референсном геноме. Были скачаны геномные последовательности сортов сои 'Jack', 'Cozusu', 'Tanba', 'Enrei', 'Pekin' и 'Fukujutaka'. С помощью веб-платформы Galaxy (Galaxy Community, 2024) по референсным кодирующим последовательностям из каждого генома выделяли соответствующие участки. Для множественного выравнивания нуклеотидных последовательностей применяли алгоритм MUSCLE, доступный в онлайн-ресурсе Multalin (версия 5.4.1) (Corpet, 1988). На основании результатов множественного выравнивания был проведен дизайн праймеров с помощью программного обеспечения Unipro UGENE v39.0 в модуле Primer Designer (Okonechnikov et al., 2012).

**Растительный материал**

Для исследования геномных локусов создали выборку из 23 образцов сои (*G. max*), контрастных по содержанию белка, из коллекции ВИР. Структуру выборки составили образцы, имеющие происхождение из 12 стран (Венгрия, Германия, Казахстан, Канада, Китай, Польша, Россия, США, Украина, Франция, Чехия, Япония). Содержание белка оценено на образцах, полученных из коллекции ВИР.

**Измерение белка**

Измерение содержания белка сои проводилось при помощи ИК-спектрофотометрии. Процесс ИК-анализа сводится к заполнению кюветы исследуемым материалом, установке ее в измерительную камеру прибора и получению результата в требуемых единицах измерения.

**Результаты**

В данной работе валидированы последовательности генов *asparagine synthetase (AS)* и *asparaginase (ASPG)*, изменчивость по которым, в соответствии с литературными данными, влияет на различия по содержанию белка (Wan et al., 2006; Pandurangan et al., 2012). Для дальнейшего изучения аллельных отличий этих генов у образцов сои из изучаемой выборки, контрастных по содержанию белка, в отчетный период проведен анализ *in silico*.

**Аспарагиназа (ASPG)**

Выравнивание нуклеотидных последовательностей гена, кодирующего ASPG, из шести сортов сои (*Glycine max*), геномы которых получены из NCBI (<https://www.ncbi.nlm.nih.gov>), представлено на рисунке 1. В консенсусной последовательности отражается самый частый вариант нуклеотида.

Множественное выравнивание нуклеотидных последовательностей гена, кодирующего ASPG, из разных сортов выявило следующие особенности. Сорт 'Fukujutaka' содержит инсерцию в одной позиции (показана пробелами в других последовательностях), что приводит к его частичному выпадению из общего консенсуса. У сорта 'Pekin' наблюдается InDel в виде пары нуклеотидов (AA/TT). Остальные сорта ('Jack', 'Tanba', 'Cozusu', 'Enrei') демонстрируют полное совпадение с консенсусной последовательностью без значимых вставок/делеций. Такая высокая степень гомологии подтверждает консервативность исследуемого генного локуса среди изученных генотипов. Праймеры на данную последовательность не разрабатывались.

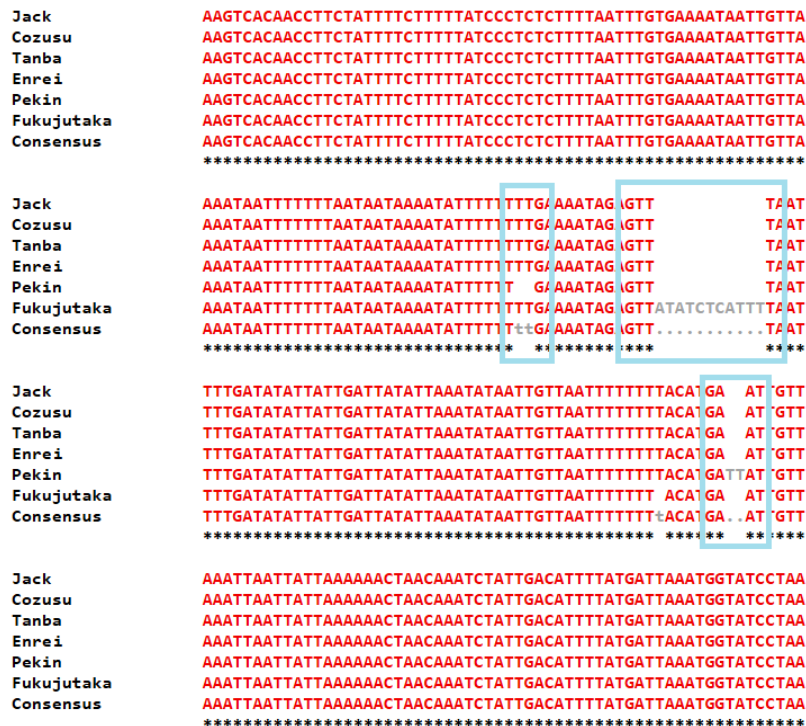
**Аспарагинсинтетаза (AS)**

Выравнивание нуклеотидных последовательностей гена, кодирующего аспарагинсинтетазу (AS), было выполнено для трех сортов сои (*G. max*), геномы которых получены из NCBI. Выравнивание представлено на рисунке 2. AS был выровнен в трех геномах: 'Enrei', 'Jack' и 'Williams 82'.

Было обнаружено два участка. Первый (размером в 200 пн), предположительно, может быть ассоциирован с содержанием белка в семенах сои. Второй участок показал выявленные инсерции/делеции – вставкой/делецией отличается сорт 'Enrei' (пробелы относительно референсного сорта 'Williams 82').

Проведенное множественное выравнивание идентифицированных нуклеотидных последовательностей позволило выявить наиболее консервативные участки, к которым был выполнен дизайн праймеров (табл. 1).

Для дальнейшего исследования была собрана выборка из 23 образцов ДНК, которая по результатам опубликованных каталогов коллекции ВИР включала в себя сорта с контрастным содержанием белка (табл. 2). Содерж



**Рис. 1.** Выравнивание последовательностей гена, кодирующего аспарагиназу (ASPG), из шести сортов сои (*Glycine max* (L.) Merr.): 'Jack', 'Cozusu', 'Tanba', 'Enrei', 'Pekin', 'Fukujutaka'

**Fig. 1.** Sequence alignment of the gene encoding asparaginase (ASPG) from six soybean (*Glycine max* (L.) Merr.) cultivars: 'Jack', 'Cozusu', 'Tanba', 'Enrei', 'Pekin', and 'Fukujutaka'



**Таблица 1. Последовательности праймеров, используемые для стандартной амплификации и амплификации ДНК-штрихкодов у образцов *Glycine max* (L.) Merr.**

**Table 1. Primer sequences used for standard and DNA barcode amplification in *Glycine max* (L.) Merr. accessions**

Название праймера	Последовательность (5' → 3')	Примечание
AS_Gmax_PCR_F	GATGGGCAGTTAGCTCCTGA	Праймеры для ПЦР
AS_Gmax_PCR_R	CTCAACATATACTTGCAGAACAGAC	
AS_Gmax_1_F	GTGACCAGAAGGAAGATAGGGAGca <b>CCAAGTCACATCAGATACAGGAC</b>	Праймеры для амплификации ДНК-штрихкодов
AS_Gmax_2_R	CATTGATAACACTGGCACAGACa <b>TGAAGTATATGTTTAATATGCCATTTCTCTC</b>	
AS_Gmax_3_F	GTGACCAGAAGGAAGATAGGGAGtca <b>CCAAGTCACATCAGATACAGGAC</b>	
AS_Gmax_4_R	CATTGATAACACTGGCACAGACat <b>TGAAGTATATGTTTAATATGCCATTTCTCTC</b>	

**Таблица 2. Содержание белка в сортах сои из коллекции ВИР, подобранных для дальнейшей работы**

**Table 2. Protein content in soybean cultivars from the VIR collection selected for further studies**

Номер по каталогу ВИР	Название	Происхождение	Содержание белка на ИК (%)	Содержание белка по классификатору*
4878	‘Урожайная’	Россия, Амурская обл.	49,64	Высокое
4909	Без названия	Канада	48,48	Высокое
5267	‘Wolfsthaler’	Германия	46,63	Высокое
5797	‘Balvanska’	Венгрия	43,00	Среднее
5884	‘Tokio vert’	Германия	51,79	Высокое
6020	‘Bulharska’	Чехословакия	47,70	Высокое
6197	‘Vilenska 18297’	Чехословакия	48,92	Высокое
6226	‘Grignon 48’	Франция	50,25	Высокое
6473	‘Амурская 402’	Россия, Амурская обл.	45,03	Высокое
6972	R-11/17-76	Польша	44,83	Среднее
9108	‘Weber’	США	41,14	Среднее
9233	‘Kador’	Франция	42,58	Среднее
9351	‘Kou Kei 95’	Япония	46,54	Среднее
10323	‘Heinong 4’	Китай	38,46	Среднее
10364	T 313	США	44,04	Среднее
10620	Л.263	Россия, Приморский край	44,52	Среднее
10696	‘Oni hadakasai’	Япония	41,93	Среднее
11129	‘МОК’	Россия, Хабаровский край	46,02	Высокое
11353	‘Murzynka’	Чехия	45,99	Высокое
11441	N 8	Китай	44,01	–
11607	‘Бриз’	Россия, Приморский край	46,77	Высокое
11615	‘Барс’	Россия, Краснодарский край	41,81	Среднее
11724	‘Грея’	Россия, Краснодарский край	47,07	Высокое

Примечание: \* – использованы градации, предлагаемые классификатором (Shchelko et al., 1990)

Note: \* – gradations proposed in the list of descriptors were used (Shchelko et al., 1990)

жание белка в семенах образцов сои, выявленное с помощью ИК-спектрофотометрии, и соотношение с классификатором показаны в таблице 2.

### Обсуждение

Соя (*G. max*) является одним из наиболее хорошо изученных объектов среди культурных растений. Такое же пристальное внимание уделяется ее дикому сородичу, *G. soja*. Первый пангеном сои, включавший семь образцов *G. soja*, был создан в 2014 г., и на данной работе было показано, что быстрее всего эволюционируют геномные локусы, обогащенные *R*-генами устойчивости (Li et al., 2014). Затем пангеном был расширен до 30 образцов, что позволило выявить географическую структуру популяций и зависимость некоторых локусов, ассоциированных с масличностью семян, от региона происхождения (Zhou et al., 2015). Последующие проекты дополнили пангеном новыми эталонными сборками: китайская дикая соя W05 (Shen et al., 2018), восемь сортов из различных широт (Chu et al., 2021), а также создан обширный пангеном PanSoy, включающий 204 филогенетически и географически разнообразных образцов (Torkamaneh et al., 2021). В 2024 г. были опубликованы результаты секвенирования геномов сорта 'Jack' и отличающегося высоким содержанием белка сорта 'HJ117' (Huang et al., 2024).

Исследования, выполненные с помощью GWAS и QTL-анализа, используют один референсный геном. При этом видовое разнообразие внутри культуры не учитывается. Также не учитываются делеции, инсерции и инверсии. Исследование пангеномных данных дает более полное представление о генетическом разнообразии вида и включает структурные варианты, которые отсутствуют в единичном референсном геноме.

Известно, что существует связь между уровнями аспарагинсинтетазы и аспарагиназы и уровнем белка в семенах сои (Pandurangan et al., 2012; Hooker et al., 2023). Но данных об исследовании в пангеноме последовательностей, кодирующих AS и ASPG, нет. Анализ переменных участков в составе пангенома открывает новые возможности для изучения полиморфных локусов, ассоциированных с желаемым признаком. В нашем исследовании проведен анализ *in silico* нуклеотидных последовательностей генов, которые кодируют аспарагинсинтетазу и аспарагиназу, на шести геномах сои ('Jack', 'Enrei', 'Williams 82' и др.). Выявлено пять SNP/InDel (см. рис. 1, 2), включая AATT-InDel в ASPG 'Fukujutaka' и делецию в 200 пн в AS 'Enrei' против 'Williams 82'.

Сформирована панель из 23 образцов из коллекции сои ВИР с контрастным содержанием белка (38–52%) (см. табл. 2), спроектированы праймеры, валидация которых позволит создать диагностическую панель для скрининга белка. Полученные данные – первый шаг к диагностической панели для скрининга коллекции сои на содержание белка.

### Заключение

Анализ *in silico* последовательностей генов, кодирующих белки аспарагинсинтетазу и аспарагиназу, выявил полиморфные локусы, ассоциированные с содержанием белка в сортах сои. Спроектированные праймеры создают основу диагностической модели для скрининга контрастных образцов.

### References / Литература

- Bandillo N., Jarquin D., Song Q., Nelson R., Cregan P., Specht J. et al. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *The Plant Genome*. 2015;8(3):plantgenome2015.04.0024. DOI: 10.3835/plantgenome2015.04.0024
- Brzostowski L.F., Diers B.W. Agronomic evaluation of a high protein allele from PI407788A on chromosome 15 across two soybean backgrounds. *Crop Science*. 2017;57(6):2972–2978. DOI: 10.2135/cropsci2017.02.0083
- Cao H., Duncan O., Islam S., Zhang J., Ma W., Millar A.H. Increased wheat protein content via introgression of an HMW glutenin selectively reshapes the grain proteome. *Molecular and Cellular Proteomics*. 2021;20:100097. DOI: 10.1016/j.mcpro.2021.100097
- Chu J.C.C., Peng B., Tang K., Yi X., Zhou H., Wang H. et al. Eight soybean reference genome resources from varying latitudes and agronomic traits. *Scientific Data*. 2021;8(1):164. DOI: 10.1038/s41597-021-00947-2
- Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*. 1988;16(22):10881–10890. DOI: 10.1093/nar/16.22.10881
- Egorov O.S., Borisova N.Yu., Borisova E.Ya., Rezhabbaev M.L., Afanas'eva E.Yu., Arzamastsev E.V. Structure and biological analysis of analogs and derivatives of biogenic polyamines. *Fine Chemical Technologies*. 2021;16(4):287–306. [in Russian] (Егоров О.С., Борисова Н.Ю., Борисова Е.Я., Режаббаев М.Л., Афанасьева Е.Ю., Арзамасцев Е.В. Структура и биологическое действие аналогов и производных биогенных полиаминов. *Тонкие химические технологии*. 2021;16(4):287–306). DOI: 10.32362/2410-6593-2021-16-4-287-306
- Emkani M., Moundanga S., Oliete B., Saurel R. Protein composition and nutritional aspects of pea protein fractions obtained by a modified isoelectric precipitation method using fermentation. *Frontiers in Nutrition*. 2023;10:1284413. DOI: 10.3389/fnut.2023.1284413
- Fukai S., Mitchell J. Grain yield and protein concentration relationships in rice. *Crop and Environment*. 2024;3(1):12–24. DOI: 10.1016/j.crope.2023.11.002
- Gaffield K.N., Goodband R.D., DeRouchey J.M., Tokach M.D., Woodworth J.C., Denny G. et al. A review of soybean processing byproducts and their use in swine and poultry diets. *Translational Animal Science*. 2024;8:txae063. DOI: 10.1093/tas/txae063
- Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*. 2024;52(W1):W83–W92. DOI: 10.1093/nar/gkae410
- Gorissen S.H.M., Crombag J.J.R., Senden J.M.G., Waterval W.A.H., Bierau J., Verdijk L.B. et al. Protein content and amino acid composition of commercially available plant-based protein isolates. *Amino Acids*. 2018;50(12):1685–1695. DOI: 10.1007/s00726-018-2640-5
- Guo B., Sun L., Jiang S., Ren H., Sun R., Wei Z. et al. Soybean genetic resources contributing to sustainable protein production. *Theoretical and Applied Genetics*. 2022;135(11):4095–4121. DOI: 10.1007/s00122-022-04222-9
- Hooker J.C., Smith M., Zapata G., Charette M., Luckert D., Mohr R.M. et al. Differential gene expression provides leads to environmentally regulated soybean seed protein content. *Frontiers in Plant Science*. 2023;14:1260393. DOI: 10.3389/fpls.2023.1260393
- Huang Y., Koo D.H., Mao Y., Herman E.M., Zhang J., Schmidt M.A. A complete reference genome for the soybean cv. Jack

- Plant Communications*. 2024;5(2):100765. DOI: 10.1016/j.xplc.2023.100765
- Hwang E.Y., Song Q., Jia G., Specht J.E., Hyten D.L., Costa J. et al. A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*. 2014;15:1. DOI: 10.1186/1471-2164-15-1
- Ishimoto M., Rahman S.M., Hanafy M.S., Khalafalla M.M., El-Shehmy H.A., Nakamoto Y. et al. Evaluation of amino acid content and nutritional quality of transgenic soybean seeds with high-level tryptophan accumulation. *Molecular Breeding*. 2009;25(2):313-326. DOI: 10.1007/s11032-009-9334-3
- Kita Y., Nakamoto Y., Takahashi M., Kitamura K., Wakasa K., Ishimoto M. Manipulation of amino acid composition in soybean seeds by the combination of deregulated tryptophan biosynthesis and storage protein deficiency. *Plant Cell Reports*. 2010;29(1):87-95. DOI: 10.1007/s00299-009-0800-5
- Li Y.H., Zhou G., Ma J., Jiang W., Jin L.G., Zhang Z. Et al. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*. 2014;32(10):1045-1052. DOI: 10.1038/nbt.2979
- NCBI. National Center for Biotechnology Information: [website]. Available from: <https://www.ncbi.nlm.nih.gov> [accessed Dec. 11, 2025].
- Novikova L.Yu., Seferova I.V., Nekrasov A.Yu., Perchuk I.N., Shelenga T.V., Samsonova M.G. et al. Impact of weather and climate on seed protein and oil content of soybean in the North Caucasus. *Vavilov Journal of Genetics and Breeding*. 2018;22(6):708-715. [in Russian] (Новикова Л.Ю., Сеферова И.В., Некрасов А.Ю., Перчук И.Н., Шеленга Т.В., Самсонова М.Г. и др. Влияние погодных-климатических условий на содержание белка и масла в семенах сои на Северном Кавказе. *Вавиловский журнал генетики и селекции*. 2018;22(6):708-715). DOI: 10.18699/VJ18.414
- Okonechnikov K., Golosova O., Fursov M.; the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-1167. DOI: 10.1093/bioinformatics/bts091
- Pandurangan S., Pajak A., Molnar S.J., Cober E.R., Dhaubhadel S., Hernández-Sebastià C. Kaiser W.M. et al. Relationship between asparagine metabolism and protein concentration in soybean seed. *Journal of Experimental Botany*. 2012;63(8):3173-3184. DOI: 10.1093/jxb/ers039
- Petraru A., Ursachi F., Amariei S. Nutritional characteristics assessment of sunflower seeds, oil and cake. Perspective of using sunflower oilcakes as a functional ingredient. *Plants (Basel)*. 2021;10(11):2487. DOI: 10.3390/plants10112487
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W. et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463(7278):178-183. DOI: 10.1038/nature08670
- Shchelko L., Sedova T., Korneychuk V., Pastucha L., Sinsky T., Hofirek P., Bares I., Sehmalova J. The international COMECON list of descriptors for the genus *Glycine* Willd. Leningrad: VIR; 1990. [in Russian] (Шелко Л., Седова Т., Корнейчук В., Пастуха Л., Синский Т., Гофирек П., Бареш И., Сегналова Я. Международный классификатор СЭВ рода *Glycine* Willd. Ленинград: ВИР; 1990).
- Shen Y., Liu J., Geng H., Zhang J., Liu Y., Zhang H. et al. *De novo* assembly of a Chinese soybean genome. *Science China. Life Sciences*. 2018;61(8):871-884. DOI: 10.1007/s11427-018-9360-0
- Torkamaneh D., Lemay M., Belzile F. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnology Journal*. 2021;19(9):1852-1862. DOI: 10.1111/pbi.13600
- Vaughn J.N., Nelson R.L., Song Q., Cregan P.B., Li Z. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3 (Bethesda)*. 2014;4(11):2283-2294. DOI: 10.1534/g3.114.013433
- Wan T.F., Shao G.H., Shan X.C., Zeng N.Y., Lam H.M. Correlation between *AS1* gene expression and seed protein contents in different soybean (*Glycine max* [L.] Merr.) cultivars. *Plant Biology (Stuttgart)*. 2006;8(2):271-276. DOI: 10.1055/s-2006-923876
- Yu Y., Hou W.S., Hacham Y., Sun S., Wu C.X., Matityahu I. et al. Constitutive expression of feedback-insensitive cystathionine  $\gamma$ -synthase increases methionine levels in soybean leaves and seeds. *Journal of Integrative Agriculture*. 2018;17(1):54-62. DOI: 10.1016/S2095-3119(16)61599-X
- Zhang H., Goettel W., Song Q., Jiang H., Hu Z., Wang M.L. et al. Selection of GmSWEET39 for oil and protein improvement in soybean. *PLoS Genetics*. 2020;16(11):e1009114. DOI: 10.1371/journal.pgen.1009114
- Zhang Q., Sun T., Wang J., Fei J., Liu Y., Liu L. et al. Genome-wide association study and high-quality gene mining related to soybean protein and fat. *BMC genomics*. 2023;24(1):596. DOI: 10.1186/s12864-023-09687-6
- Zhou Z., Jiang Y., Wang Z., Gou Z., Lyu J., Li W. Et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*. 2015;33(4):408-414. DOI: 10.1038/nbt.3096

### Информация об авторах

**Ирина Вениаминовна Розанова**, кандидат биологических наук, старший научный сотрудник, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, rožanova.iv@talantiuspeh.ru, <https://orcid.org/0000-0003-4341-0766>

**Полина Владимировна Климова**, магистрант, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, Klimova.polya04@gmail.com, <https://orcid.org/0009-0003-7808-3782>

**Анастасия Ярославовна Евлаш**, младший научный сотрудник, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, 190000 Россия, Санкт-Петербург, ул. Б. Морская, 42, 44, evlash.ae@talantiuspeh.ru, <https://orcid.org/0000-0003-0246-8929>

**Ирина Владимировна Сеферова**, кандидат биологических наук, ведущий научный сотрудник, Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, 190000 Россия, Санкт-Петербург, ул. Б. Морская, 42, 44, i.seferova@vir.nw.ru, <https://orcid.org/0000-0003-3308-9198>

**Information about the authors**

**Irina V. Rozanova**, Cand. Sci. (Biology), Senior Researcher, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiysky Ave., Sirius Settle., Sirius Federal Territory, Krasnodar Territory 354340, Russia, rozanova.iv@talantiuspeh.ru, <https://orcid.org/0000-0003-4341-0766>

**Polina V. Klimova**, MSc Student, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiysky Ave., Sirius Settle., Sirius Federal Territory, Krasnodar Territory 354340, Russia, Klimova.polya04@gmail.com, <https://orcid.org/0009-0003-7808-3782>

**Anastasia Ya. Evlash**, Associate Researcher, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiysky Ave., Sirius Settle., Sirius Federal Territory, Krasnodar Territory 354340, Russia, N.I. Vavilov All-Russian Institute of Plant Genetic Resources, 42, 44 Bolshaya Morskaya Street, St. Petersburg 190000, Russia, evlash.ae@talantiuspeh.ru, <https://orcid.org/0000-0003-0246-8929>

**Irina V. Seferova**, Cand. Sci. (Biology), Leading Researcher, N.I. Vavilov All-Russian Institute of Plant Genetic Resources, 42, 44 Bolshaya Morskaya Street, St. Petersburg 190000, Russia, i.seferova@vir.nw.ru, <https://orcid.org/0000-0003-3308-9198>

**Вклад авторов:** все авторы сделали эквивалентный вклад в подготовку публикации.

**Contribution of the authors:** the authors contributed equally to this article.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Conflict of interests:** the authors declare no conflicts of interests.

Статья поступила в редакцию 04.01.2026; одобрена после рецензирования 24.01.2026; принята к публикации 25.01.2026.  
The article was submitted on 04.01.2026; approved after reviewing on 24.01.2026; accepted for publication on 25.01.2026.