Вавиловский журнал генетики и селекции. 2022;26(8):798-805
DOI 10.18699/VJGB-22-96

Original Russian text https://sites.icgbio.ru/vogis/

Promoters of genes encoding β-amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: *in silico* analysis

O.V. Vishnevsky[1], I.V. Chadaeva[1], E.B. Sharypova[1], B.M. Khandaev[1], K.A. Zolotareva[1], A.V. Kazachek[1], P.M. Ponomarenko[1], N.L. Podkolodny[1, 2], D.A. Rasskazov[1], A.G. Bogomolov[1], O.A. Podkolodnaya[1], L.K. Savinkova[1], E.V. Zemlyanskaya[1], M.P. Ponomarenko[1] ✉

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ pon@bionet.nsc.ru

Abstract. It is generally accepted that during the domestication of food plants, selection was focused on their productivity, the ease of their technological processing into food, and resistance to pathogens and environmental stressors. Besides, the palatability of plant foods and their health benefits could also be subjected to selection by humans in the past. Nonetheless, it is unclear whether in antiquity, aside from positive selection for beneficial properties of plants, humans simultaneously selected against such detrimental properties as allergenicity. This topic is becoming increasingly relevant as the allergization of the population grows, being a major challenge for modern medicine. That is why intensive research by breeders is already underway for creating hypoallergenic forms of food plants. Accordingly, in this paper, albumin, globulin, and β-amylase of common wheat *Triticum aestivum* L. (1753) are analyzed, which have been identified earlier as targets for attacks by human class E immunoglobulins. At the genomic level, we wanted to find signs of past negative selection against the allergenicity of these three proteins (albumin, globulin, and β-amylase) during the domestication of ancestral forms of modern food plants. We focused the search on the TATA-binding protein (TBP)-binding site because it is located within a narrow region (between positions –70 and –20 relative to the corresponding transcription start sites), is the most conserved, necessary for primary transcription initiation, and is the best-studied regulatory genomic signal in eukaryotes. Our previous studies presented our publicly available Web service Plant_SNP_TATA_Z-tester, which makes it possible to estimate the equilibrium dissociation constant (K_D) of TBP complexes with plant proximal promoters (as output data) using 90 bp of their DNA sequences (as input data). In this work, by means of this bioinformatics tool, 363 gene promoter DNA sequences representing 43 plant species were analyzed. It was found that compared with non-food plants, food plants are characterized by significantly weaker affinity of TBP for proximal promoters of their genes homologous to the genes of common-wheat globulin, albumin, and β-amylase (food allergens) ($p < 0.01$, Fisher's Z-test). This evidence suggests that in the past humans carried out selective breeding to reduce the expression of food plant genes encoding these allergenic proteins.

Key words: food allergen; albumin; globulin; β-amylase; gene; promoter; common wheat *Triticum aestivum* L. (1753); plants; TATA-binding protein; TATA box; domestication; selection; *in silico* estimate.

For citation: Vishnevsky O.V., Chadaeva I.V., Sharypova E.B., Khandaev B.M., Zolotareva K.A., Kazachek A.V., Ponomarenko P.M., Podkolodny N.L., Rasskazov D.A., Bogomolov A.G., Podkolodnaya O.A., Savinkova L.K., Zemlyanskaya E.V., Ponomarenko M.P. Promoters of genes encoding β-amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: *in silico* analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):798-805. DOI 10.18699/VJGB-22-96

Промоторы генов, кодирующих β-амилазу, альбумин и глобулин пищевых растений в сравнении с непищевыми, характеризуются более низкой аффинностью к TATA-связывающему белку: *in silico* анализ

О.В. Вишневский[1], И.В. Чадаева[1], Е.Б. Шарыпова[1], Б.М. Хандаев[1], К.А. Золотарева[1], А.В. Казачек[1], П.М. Пономаренко[1], Н.Л. Подколодный[1, 2], Д.А. Рассказов[1], А.Г. Богомолов[1], О.А. Подколодная[1], Л.К. Савинкова[1], Е.В. Землянская[1], М.П. Пономаренко[1] ✉

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ pon@bionet.nsc.ru

Аннотация. Принято считать, что при доместикации пищевых растений отбор шел на урожайность, технологичность переработки в продукты питания, устойчивость к патогенам и стрессовым воздействиям окружающей среды. При этом также могли оцениваться вкусовые качества продуктов питания растительного происхождения и их ценность для здоровья. Однако неясно, проводил ли человек в прошлом наряду с положительным отбором на полезные свойства растений одновременно отбор против таких вредоносных свойств, как способность вызывать аллергические реакции. Этот вопрос становится все более актуальным по мере роста аллергизации населения как вызова современной медицине. В связи с этим селекционерами уже ведутся интенсивные исследования по созданию гипоаллергенных форм пищевых растений. В этой работе рассмотрены альбумин, глобулин и β-амилаза мягкой пшеницы *Triticum aestivum* L. (1753), идентифицированные ранее как мишени для атак иммуноглобулинов класса Е человека. Нашей целью было найти на геномном уровне следы отрицательного отбора в прошлом против гипераллергенности трех белков (альбумин, глобулин и β-амилаза) при одомашнивании предковых форм современных пищевых растений. Для этого мы сфокусировали поиск на сайте связывания TATA-связывающего белка (TBP) как локализованном в узком районе [–70; –20] относительно старта транскрипции, консервативном, необходимом для первичной инициации транскрипции и наиболее изученном регуляторном сигнале в геномах эукариот. Ранее нами был создан свободно доступный веб-сервис Plant_SNP_TATA_Z-tester для оценки величин равновесной константы диссоциации (K_D) комплексов TBP с проксимальными промоторами генов растений по их последовательностям ДНК длиной 90 п. о. В настоящей работе с его помощью проанализированы 363 последовательности ДНК промоторов генов 43 видов растений. Обнаружено, что пищевые растения, в сравнении с непищевыми, характеризуются достоверно более низкой аффинностью TBP к проксимальным промоторам их генов, гомологичных генам глобулина, альбумина и β-амилазы мягкой пшеницы как пищевых аллергенов ($p < 0.01$, Z-критерий Фишера). Это свидетельствует об отборе при доместикации пищевых растений в прошлом на снижение уровня данных аллергенных белков.

Ключевые слова: пищевые аллергены; альбумин; глобулин; β-амилаза; ген; промотор; мягкая пшеница *Triticum aestivum* L. (1753); растения; TATA-связывающий белок; TATA-бокс; доместикация; отбор; оценки *in silico*.

Introduction

Currently, the problem of food allergenicity is extremely relevant because the documented rapid growth of population allergization is becoming one of the key challenges for modern medicine (Prescott et al., 2022). In this regard, modern plant breeders are working in two directions: (1) creation of new hypoallergenic forms of agricultural food plants and (2) identification of new plant food allergens and of molecular mechanisms of their action (Hong et al., 2021; Cavazza et al., 2022).

The aim of our work was to search at the molecular genetic level for signs of negative selection against allergens during the domestication of ancestral forms of modern food plants. Three food allergens from common wheat *Triticum aestivum* L. (1753) were studied: β-amylase, albumin, and globulin, previously identified as targets of allergic reactions mediated by human class E immunoglobulins (Wang et al., 2021).

The current study was conducted using our previously developed freely available Web service Plant_SNP_TATA_Z-tester, which is designed to estimate the equilibrium dissociation constant (K_D) of a complex of *Arabidopsis thaliana* (L.) Heynh. (1842) TBP-1 (hereafter: "plant TBP") with a proximal promoter of various plant genes (Rasskazov et al., 2022). This tool was utilized to analyze 363 nucleotide sequences of proximal promoters of relevant genes from 43 plant species. As a result, compared to non-food plants, food plants were found to have significantly weaker affinity of plant TBP toward promoters of genes homologous to common-wheat genes of β-amylase, albumin, and globulin (food allergens). These data indicate that in the past, selection was carried out by humans for reducing the expression of food plant genes encoding allergenic proteins when such plants were domesticated.

Materials and methods

Nucleotide sequences of plant gene promoters analyzed in this work. Three allergenic proteins from common wheat *T. aestivum* were investigated: β-amylase, albumin, and globulin, which have previously been experimentally identified as targets for human class E immunoglobulins (Wang et al., 2021). From the GenBank database (Benson et al., 2015), nucleotide sequences of 90 bp proximal promoters were retrieved that are located immediately upstream of transcription start sites of plant genes homologous to the genes of β-amylase, albumin, and globulin from common wheat *T. aestivum*. After the exclusion of promoter DNA sequences with unknown nucleotides w, s, r, y, k, m, b, d, h, v, and n (according to the nomenclature of (IUPAC-IUB…, 1970)), we had 363 promoter sequences belonging to 43 plant species. Then, all 43 plant species were categorized into two nonoverlapping groups: group I, represented by 235 proximal promoters from 28 food plant species for which there was information about their centuries-old use by humans as foods (Table 1), and group II, represented by 128 proximal promoters from non-food plants (the other 15 species) (Table 2).

Nucleotide sequence analysis of proximal promoters of plants. Using Web service Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022), which we have created earlier, we calculated K_D (in moles per liter; M) for complexes of plant TBP with each promoter by means of the nucleotide sequence of each promoter (characterized in Tables 1 and 2).

The calculations were performed in accordance with our previously formulated model of three-step binding of TBP to a promoter (i) TBP slides along the double helix of promoter DNA (Coleman, Pugh, 1995) ↔ (ii) TBP stops at a potential site of TBP binding (Berg, von Hippel, 1987; Bucher,

Table 1. Characteristics of 235 nucleotide sequences of proximal promoters of food plant genes homologous to the studied globulin (*Glo*), albumin (*Alb*), and β-amylase (*Bmy*) genes from common wheat *T. aestivum*

Food plant species		Number of promoters		
No.	Name	*Glo*	*Alb*	*Bmy*
1	Buckwheat *Fagopyrum esculentum* Moench, 1794	1	–	–
2	Maidenhair tree *Ginkgo biloba* L., 1771	1	–	–
3	Yoshino cherry *Prunus yedoensis* var. *nudiflora* Koehne, 1912	1	–	2
4	Maize *Zea mays* L., 1753	1	–	–
5	Oat *Avena sativa* L., 1753	2	–	–
6	Waxberry *Morella rubra* Siebold & Zucc.	2	2	1
7	Quinoa *Chenopodium quinoa* Willd., 1798	2	–	–
8	Rice *Oryza sativa* L., 1753	3	–	–
9	Melon *Cucumis melo* L., 1753	4	2	6
10	Cardoon *Cynara cardunculus* L.	4	–	–
11	Cork oak *Quercus suber* L.	4	–	–
12	Wine grape *Vitis vinifera* L.	9	–	9
13	Congolese coffee *Coffea canephora* Pierre ex A. Froehner, 1897	1	–	–
14	Pepper *Capsicum annuum* L., 1753	26	8	27
15	Sesame *Sesamum indicum* L.	–	1	–
16	Kiwifruit nashi-kazura *Actinidia rufa* Franch. & Sav.	–	1	–
17	Brazil nut *Bertholletia excelsa* Humb. & Bonpl.	–	1	–
18	Soybean *Glycine max* (L.) Merr., 1917	–	2	–
19	Pea *Pisum sativum* L., 1753	–	4	–
20	Perilla *Perilla frutescens* var. *hirtella* (Nakai) Makino	–	5	–
21	Almond *Prunus dulcis* (Mill.) D.A. Webb, 1967	–	8	4
22	Mandarin unshiu *Citrus unshiu* (Tanaka ex Swingle) Marcow., 1921	–	15	–
23	Tea *Camellia sinensis* (L.) Kuntze, 1887	–	–	1
24	Barley *Hordeum vulgare* L. (1753)	–	–	2
25	Hibiscus *Hibiscus syriacus* L. (1753)	–	–	2
26	Pineapple *Ananas comosus* (L.) Merr., 1917	–	–	3
27	Olive *Olea europaea* L., 1753	–	–	4
28	Sweet wormwood *Artemisia annua* L.	13	35	16
Total number of food plant species		15	12	12

1990) ↔ (iii) the TBP/promoter complex is stabilized by bending of the DNA double-helix axis at a right angle (Flatters, Lavery, 1998), as subsequently demonstrated experimentally *in vitro* (Delgadillo et al., 2009).

Statistical analysis. In this work, using standard software package Statistica (Statsoft™, USA), we averaged the Plant_SNP_TATA_Z-tester-generated (Rasskazov et al., 2022) estimates of K_D – for complexes of plant TBP with promoters of β-amylase, albumin, and globulin genes – for food and non-food plants separately. On the basis of these data, statistical significance of differences between food and non-food plants was evaluated by Fisher's Z-test.

Results

Globulin

Table 3 presents the *in silico* estimates of K_D for complexes of plant TBP with 74 proximal promoters of globulin genes from 15 food plant species in comparison with 53 such promoters from 12 non-food plant species, as determined using Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022). One can see in this table that in the case of food plants, the estimates of K_D for complexes of plant TBP with promoters of these genes varied from 1.67 ± 0.12 (mean ± SEM) to 6.75 ± 5.23 nM, with an average of 2.97 ± 0.21 nM, whereas for non-food plants,

Table 2. Characteristics of 128 nucleotide sequences of proximal promoters from non-food plant genes homologous to the studied globulin (*Glo*), albumin (*Alb*), and β-amylase (*Bmy*) genes from common wheat *T. aestivum*

Non-food plant species		Number of promoters		
No.	Name	*Glo*	*Alb*	*Bmy*
1	Five-seeded plume-poppy *Macleaya cordata* (Willd.) R. Br.	1	–	–
2	Witchweed *Striga asiatica* (L.) Kuntze	1	1	1
3	Genlisea *Genlisea aurea* A.St. Hil. (1833)	1	–	2
4	Florida teosinte *Zea luxurians* (Durieu & Asch.) R.M. Bird, 1978	1	–	–
5	Noccidium *Microthlaspi erraticum* (Jord.) T. Ali & Thines, 2016	2	1	2
6	Gama grass *Tripsacum dactyloides* (L.) L., 1759	2	–	–
7	Balsas teosinte *Zea mays* subsp. *parviglumis* Iltis & Doebley, 1980	3	–	–
8	Thale cress *Arabidopsis thaliana* (L.) Heynh., 1842	4	5	3
9	Panic grass *Dichanthelium oligosanthes* (Schult.) Gould	6	–	6
10	Water lily *Nymphaea thermarum* Eb. Fisch., 1988	8	1	9
11	Rue-anemone *Thalictrum thalictroides* (L.) A.J. Eames & B. Boivin	9	15	6
12	Chile tomato *Solanum chilense* (Dunal) Reiche	15	13	6
13	Purple witchweed, *Striga hermonthica* (Delile) Benth.	–	1	–
14	Gerardia *Phtheirospermum japonicum* (Thunb.) Kanitz	–	–	1
15	Chinese rose *Rosa chinensis* Jacq., 1768	–	–	2
Total number of food plant species		12	7	10

these values varied from 1.25±0.06 to 3.33±0.23 nM, with an average of 2.15±0.08 nM.

In Fig. 1, arithmetic mean estimates of K_D for complexes of plant TBP with globulin-coding gene promoters are compared between two groups (food and non-food plants) by Fisher's Z-test. The difference between the groups was significant, with Z = 3.59 and $p < 0.001$.

Albumin

Table 4 shows data obtained by Web service Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022) regarding estimates of K_D for complexes of plant TBP with 84 albumin gene promoters from 12 food plant species and with 37 promoters from 7 non-food plant species. As readers can see in this table, in the case of food plants, the estimates of K_D of TBP-promoter complexes for these genes ranged between 1.65±0.12 and 4.49±1.39 nM (average: 3.10±0.22 nM), whereas for non-food plants, they ranged from 1.65±0.05 to 2.70±0.22 nM (average: 2.18±0.10 nM).

A comparison of the two groups (food and non-food plants) by Fisher's Z-test is displayed in Fig. 2. Here one can see a significant difference between food plants and non-food plants (Z = 3.85, $p < 0.001$).

β-Amylase

Table 5 lists estimated K_D values of complexes of plant TBP with 77 proximal promoters of β-amylase genes from 12 food plant species and with 38 promoters from 10 non-food plant species, as calculated by Web service Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022). For food plants, this table presents the range of K_D from 1.30±0.09 to 8.77±7.36 nM, with an arithmetic mean of 2.85±0.21 nM, whereas for non-

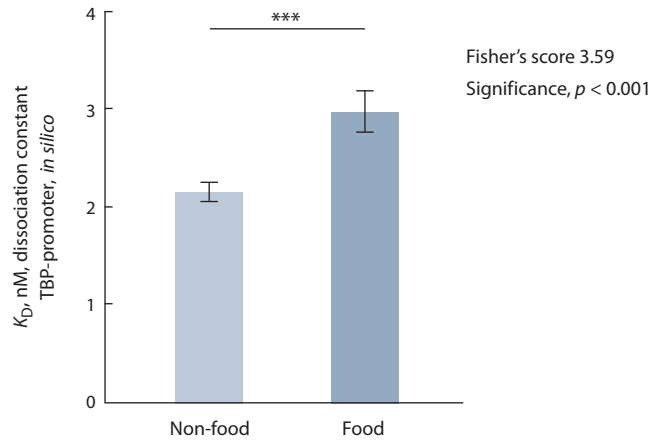

Fig. 1. The statistically significant difference between the studied food plants and non-food plants in the *in silico* estimates of K_D for complexes of plant TBP with 90 bp proximal promoters of their genes encoding globulins.

Here and in Fig. 2: *** statistical significance $p < 0.001$ according to Fisher's Z-test.

Table 3. Arithmetic mean estimates (M_0) of the equilibrium dissociation constant (K_D) of complexes between plant TBP and 90 bp proximal promoters of the plant globulin genes analyzed in this work

No.	Plant species	*N*	K_D, $M_0 \pm \Delta$, nM
	Food plants		
1	Congolese coffee	1	2.17 ± 0.13
2	Buckwheat	1	2.04 ± 0.14
3	Maidenhair tree	1	1.67 ± 0.12
4	Yoshino cherry	1	1.76 ± 0.12
5	Maize	1	2.30 ± 0.16
6	Oat	2	2.57 ± 0.15
7	Waxberry	2	6.75 ± 5.23
8	Quinoa	2	2.47 ± 0.12
9	Rice	3	2.84 ± 0.38
10	Melon	4	2.66 ± 0.30
11	Cardoon	4	3.22 ± 0.37
12	Cork oak	4	4.84 ± 1.41
13	Wine grape	9	2.67 ± 0.35
14	Sweet wormwood	13	2.51 ± 0.25
15	Pepper	26	3.01 ± 0.37
Total		74	2.97 ± 0.21
	Non-food plants		
1	Five-seeded plume-poppy	1	1.25 ± 0.06
2	Witchweed	1	3.33 ± 0.23
3	Genlisea	1	2.70 ± 0.19
4	Florida teosinte	1	1.96 ± 0.14
5	Noccidium	2	2.25 ± 0.83
6	Gama grass	2	2.19 ± 0.07
7	Balsas teosinte	3	2.12 ± 0.11
8	Thale cress	4	1.90 ± 0.13
9	Panic grass	6	2.27 ± 0.15
10	Water lily	8	2.60 ± 0.39
11	Rue-anemone	9	2.01 ± 0.16
12	Chile tomato	15	1.97 ± 0.11
Total		53	2.15 ± 0.08

Note. Here and in Tables 4 and 5: *N* – total number of the promoter studied; M_0 – arithmetic mean score; Δ – standard error of the mean (SEM).

food plants, the range of K_D was found to be 1.66±0.32 to 6.75±5.23 nM, with an average of 3.89±0.32 nM. Fig. 3 presents a comparison between the analyzed food and non-food plants by Fisher's Z-test, according to which these groups are statistically significantly different at Z = 2.74 and $p < 0.01$.

Table 4. Arithmetic mean estimates (M_0) of the equilibrium dissociation constant (K_D) for complexes between plant TBP and 90 bp proximal promoters of the plant albumin genes investigated in this work

No.	Plant species	*N*	K_D, $M_0 \pm \Delta$, nM
	Food plants		
1	Sesame	1	2.28 ± 0.16
2	Kiwifruit nashi-kazura	1	1.77 ± 0.12
3	Brazil nut	1	3.04 ± 0.15
4	Waxberry	2	1.96 ± 0.14
5	Melon	2	2.04 ± 0.14
6	Soybean	2	1.65 ± 0.12
7	Pea	4	4.49 ± 1.39
8	Perilla	5	1.98 ± 0.40
9	Almond	8	3.74 ± 0.66
10	Pepper	8	3.00 ± 0.59
11	Mandarin unshiu	15	3.51 ± 0.61
12	Sweet wormwood	35	3.07 ± 0.35
Total		84	3.10 ± 0.22
	Non-food plants		
1	Water lily	1	2.70 ± 0.22
2	Noccidium	1	2.00 ± 0.14
3	Purple witchweed	1	1.65 ± 0.12
4	Witchweed	1	2.19 ± 0.15
5	Thale cress	5	2.03 ± 0.21
6	Chile tomato	13	2.33 ± 0.20
7	Rue-anemone	15	2.11 ± 0.16
Total		37	2.18 ± 0.10

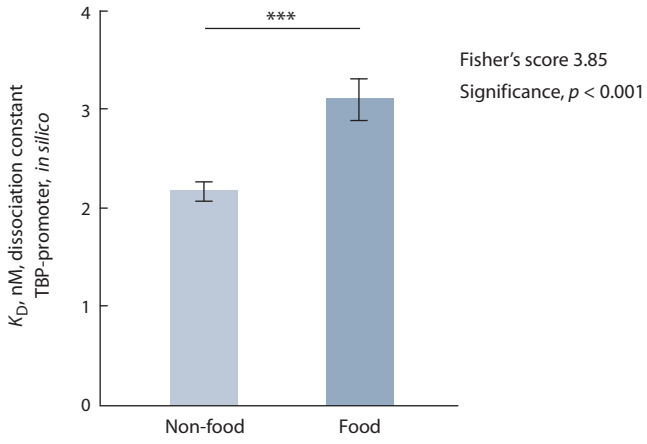

Fig. 2. The statistically significant difference between the studied food plants and non-food plants in the *in silico* estimates of K_D for the complexes of plant TBP with 90 bp proximal promoters of their genes encoding albumins.

Table 5. Arithmetic mean estimates (M_0) of the equilibrium dissociation constant (K_D) of complexes between plant TBP and 90 bp proximal promoters of the plant β-amylase genes examined in this work

No.	Plant species	*N*	K_D, $M_0 \pm \Delta$, nM
	Food plants		
1	Tea	1	4.59 ± 0.28
2	Waxberry	1	2.21 ± 0.13
3	Barley	2	1.30 ± 0.09
4	Hibiscus	2	3.58 ± 1.86
5	Yoshino cherry	2	3.19 ± 1.68
6	Pineapple	3	8.77 ± 7.36
7	Almond	4	6.56 ± 1.63
8	Olive	4	5.24 ± 0.93
9	Melon	6	4.79 ± 0.96
10	Wine grape	9	3.97 ± 0.73
11	Sweet wormwood	16	4.29 ± 0.77
12	Pepper	27	2.59 ± 0.23
Total		77	3.89 ± 0.32
	Non-food plants		
1	Witchweed	1	3.50 ± 0.25
2	Gerardia	1	3.43 ± 0.21
3	Genlisea	2	1.38 ± 0.82
4	Noccidium	2	1.93 ± 0.37
5	Chinese rose	2	1.79 ± 0.23
6	Thale cress	3	1.66 ± 0.32
7	Panic grass	6	3.26 ± 0.43
8	Rue-anemone	6	2.91 ± 0.43
9	Chile tomato	6	2.89 ± 0.47
10	Water lily	9	3.30 ± 0.62
Total		38	2.85 ± 0.21

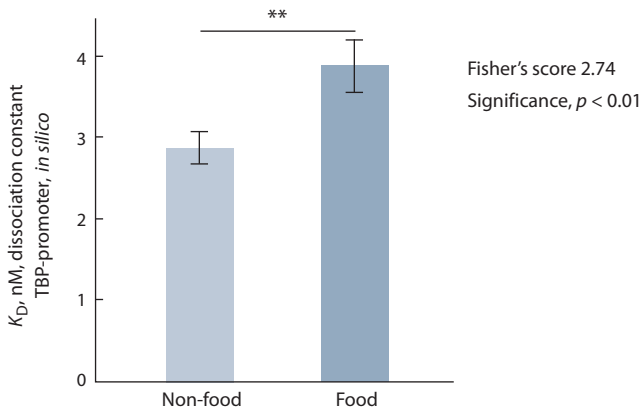

Fig. 3. The statistically significant difference between the studied food plants and non-food plants in the *in silico* estimates of K_D for the complexes of plant TBP with 90 bp proximal promoters of their genes encoding β-amylases.

** Statistical significance $p < 0.01$ according to Fisher's Z-test.

Discussion

It is well known that in the process of spontaneous domestication of ancestral forms of modern food plants, the selection was primarily based on their economically valuable traits, such as productivity, resistance to pathogens and to environmental stressors, and the ease of technological processing into final food products. Additionally, during the plant domestication, humans assessed the palatability of food products and their benefits for health.

It remains unclear whether in addition to the positive selection for the beneficial properties of agricultural plants, there was also simultaneous selection against their detrimental properties, which include allergenicity of the dishes prepared from these plants. To answer this question, we concentrated on the search for molecular genetic selection markers related to structural and functional organization of proximal promoters of plant genes.

Accordingly, plant genes were analyzed that are homologous to three *T. aestivum* genes encoding food allergens β-amylase, albumin, and globulin, earlier identified as targets for human IgE (Wang et al., 2021). Thus, 363 homologous genes were investigated belonging to 28 and 15 species of food plants and non-food plants, respectively. With the help of Web service Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022), for each homologous gene, K_D of the complex of plant TBP with this gene's proximal promoter was computed.

Interest in the TBP protein and its binding site in the proximal promoter (canonical form: the TATA box) is due to the fact that they play a key role in the initiation of eukaryotic gene transcription. It has been experimentally established (Coleman, Pugh, 1995) that TBP slides along the DNA double helix owing to nonspecific affinity between them: $K_D \sim 10^{-5}$ M (Hahn et al., 1989). TBP then stops at a site of TBP binding because of their mutual molecular recognition (Berg, von Hippel, 1987; Bucher, 1990) mediated by stronger (specific) affinity of TBP for this site: $K_D \sim 10^{-9}$ M (Hahn et al., 1989). Next, under the action of TBP, the DNA double helix melts at the site of TBP binding, and kinking of the DNA axis at a right angle takes place, which stabilizes the TBP-promoter complex (Flatters, Lavery, 1998). The resultant TBP-promoter complex is considered an obligatory DNA anchor, which is required for the binding of RNA polymerase II (Muller et al., 2001; Martianov et al., 2002; Choukrallah et al., 2012; Rhee, Pugh, 2012) as a key step in the assembly of the transcription preinitiation complex (Auble, 2009) responsible for basal transcription (Fire et al., 1984). Due to the key importance of TATA boxes, mutations located in proximal promoters have a well-pronounced effect on the magnitude of gene expression (Savinkova et al., 2009).

The molecular mechanism underlying the binding of TBP to a promoter of various eukaryotic genes via the three successive steps was first proposed by P. Ponomarenko et al. (2008) and later confirmed experimentally (Delgadillo et al., 2009). Based on this mechanism, a bioinformatic model was devised previously for calculating a change in K_D (of a complex between TBP and a proximal promoter of a eukaryotic gene) for a polymorphism of the TBP-binding site(s) in the promoter as compared to the wild type (Ponomarenko et al., 2009). Results of computations based on this model have been

confirmed by independent *ex vivo* experiments on cell cultures transfected with the pGL4.10 plasmid (Promega, USA) carrying a wild-type or mutant promoter inserted before a luciferase reporter gene (Ponomarenko et al., 2017) as well as *in vitro* in real time (Arkova et al., 2017) by means of stopped-flow spectrometer SX.20 (Applied Photophysics, UK) under equilibrium conditions (Savinkova et al., 2013) and under non-equilibrium conditions (Drachkova et al., 2014) with the help of an electrophoretic mobility shift assay. As a result of such comprehensive verification of this bioinformatic model, on its basis, the Web service Plant_SNP_TATA_Z-tester (Rasskazov et al., 2022) was created, which was employed in the current project for estimating K_D of complexes of plant TBP with proximal promoters of genes from food and non-food plants.

Our analysis revealed that in comparison with non-food plants, food plants are characterized by significantly weaker affinity of TBP for promoters of genes homologous to common-wheat β-amylase, albumin, and globulin (food allergens) ($p < 0.01$, as estimated by the above software and Fisher's Z-test). When interpreting the obtained results, let us take into account the experimentally proven fact that the level of expression of eukaryotic genes increases with enhancement of the affinity of TBP for the promoters of these genes (Mogno et al., 2010). This observation allows us to interpret the food plants' weaker TBP affinity – for promoters of genes homologous to genes of food allergens (common-wheat β-amylase, albumin, and globulin) in comparison with non-food plants – as evidence of selection by humans for low amounts of these allergenic proteins in food plants in the past, during the domestication of the plants.

Conclusion

In this work, DNA sequences of proximal promoters of genes homologous to genes of food allergens (Wang et al., 2021) were consistently analyzed *in silico* for the first time for food compared to non-food plants. As a result, weaker *in silico* affinity of TBP was observed for promoters of the investigated food plant genes as compared to genes of non-food plants. This finding is suggestive of artificial selection – in antiquity, for the purpose of reducing the expression of food plant genes encoding allergenic proteins – carried out by humans in the course of domestication of plants as food products.

References

Arkova O., Kuznetsov N., Fedorova O., Savinkova L. A real-time study of the interaction of TBP with a TATA box-containing duplex identical to an ancestral or minor allele of human gene *LEP* or *TPI*. *J. Biomol. Struct. Dyn*. 2017;35(14):3070-3081. DOI 10.1080/07391102.2016.1241190.

Auble D.T. The dynamic personality of TATA-binding protein. *Trends Biochem. Sci.* 2009;34(2):49-52. DOI 10.1016/j.tibs.2008.10.008.

Benson D.A., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. GenBank. *Nucleic Acids Res*. 2015;43(Database issue):D30-D35. DOI 10.1093/nar/gku1216.

Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol*. 1987;193(4):723-750. DOI 10.1016/0022-2836(87)90354-8.

Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol*. 1990;212(4):563-578. DOI 10.1016/0022-2836(90)90223-9.

Cavazza A., Mattarozzi M., Franzoni A., Careri M. A spotlight on analytical prospects in food allergens: From emerging allergens and novel foods to bioplastics and plant-based sustainable food contact materials. *Food Chem*. 2022;388:132951. DOI 10.1016/j.foodchem.2022.132951.

Choukrallah M.A., Kobi D., Martianov I., Pijnappel W.W., Mischerikow N., Ye T., Heck A.J., Timmers H.T., Davidson I. Interconversion between active and inactive TATA-binding protein transcription complexes in the mouse genome. *Nucleic Acids Res*. 2012;40(4):1446-1459. DOI 10.1093/nar/gkr802.

Coleman R.A., Pugh B.F. Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. *J. Biol. Chem*. 1995;270(23):13850-13859. DOI 10.1074/jbc.270.23.13850.

Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry*. 2009;48(8):1801-1809. DOI 10.1021/bi8018724.

Drachkova I., Savinkova L., Arshinova T., Ponomarenko M., Peltek S., Kolchanov N. The mechanism by which TATA-box polymorphisms associated with human hereditary diseases influence interactions with the TATA-binding protein. *Hum. Mutat*. 2014;35(5):601-608. DOI 10.1002/humu.22535.

Fire A., Samuels M., Sharp P.A. Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *J. Biol. Chem*. 1984;259(4):2509-2516. DOI 10.1016/S0021-9258(17)43382-5.

Flatters D., Lavery R. Sequence-dependent dynamics of TATA-Box binding sites. *Biophys. J*. 1998;75(1):372-381. DOI 10.1016/S0006-3495(98)77521-6.

Hahn S., Buratowski S., Sharp P.A., Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci. USA*. 1989;86(15):5718-5722. DOI 10.1073/pnas.86.15.5718.

Hong L., Pan M., Xie X., Liu K., Yang J., Wang S., Wang S. Aptamer-based fluorescent biosensor for the rapid and sensitive detection of allergens in food matrices. *Foods*. 2021;10(11):2598. DOI 10.3390/foods10112598.

IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. Recommendations 1970. *Biochem. J.* 1970;120(3):449-454. DOI 10.1042/bj1200449.

Martianov I., Viville S., Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science*. 2002;298(5595):1036-1039. DOI 10.1126/science.1076327.

Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters. *Genome Res*. 2010;20(10):1391-1397. DOI 10.1101/gr.106732.110.

Muller F., Lakatos L., Dantonel J., Strahle U., Tora L. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* 2001;11(4):282-287. DOI 10.1016/s0960-9822(01)00076-8.

Ponomarenko P., Chadaeva I., Rasskazov D.A., Sharypova E., Kashina E.V., Drachkova I., Zhechev D., Ponomarenko M.P., Savinkova L.K., Kolchanov N. Candidate SNP markers of familial and sporadic Alzheimer's diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front. Aging Neurosci*. 2017;9:231. DOI 10.3389/fnagi.2017.00231.

Ponomarenko P.M., Ponomarenko M.P., Drachkova I.A., Lysova M.V., Arshinova T.V., Savinkova L.K., Kolchanov N.A. Prediction of the affinity of the TATA-binding protein to TATA boxes with single nucleotide polymorphisms. *Mol. Biol. (Moscow)*. 2009; 43(3):472-479. DOI 10.1134/S0026893309030157.

Ponomarenko P., Savinkova L., Drachkova I., Lysova M., Arshinova T., Ponomarenko M., Kolchanov N. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl. Biochem. Biophys*. 2008;419:88-92. DOI 10.1134/S1607672908020117.

Prescott S.L., Logan A.C., Bristow J., Rozzi R., Moodie R., Redvers N., Haahtela T., Warber S., Poland B., Hancock T., Berman B. Exiting the anthropocene: achieving personal and planetary health in the 21st century. *Allergy*. 2022;77(12):3498-3512. DOI 10.1111/all.15419.

Rasskazov D., Chadaeva I., Sharypova E., Zolotareva K., Khandaev B., Ponomarenko P., Podkolodnyy N., Tverdokhleb N., Vishnevsky O., Bogomolov A., Podkolodnaya O., Savinkova L., Zemlyanskaya E., Golubyatnikov V., Kolchanov N., Ponomarenko M. Plant_SNP_TATA_Z-tester: a Web service that unequivocally estimates the impact of proximal promoter mutations on plant gene expression. *Int. J. Mol. Sci*. 2022;23(15):8684. DOI 10.3390/ijms23158684.

Rhee H., Pugh B. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*. 2012;483(7389):295-301. DOI 10.1038/nature10799.

Savinkova L., Drachkova I., Arshinova T., Ponomarenko P., Ponomarenko M., Kolchanov N. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. *PLoS One*. 2013;8(2):e54626. DOI 10.1371/journal.pone.0054626.

Savinkova L.K., Ponomarenko M.P., Ponomarenko P.M., Drachkova I.A., Lysova M.V., Arshinova T.V., Kolchanov N.A. TATA box polymorphisms in human gene promoters and associated hereditary pathologies. *Biochemistry (Moscow)*. 2009;74(2):117-129. DOI 10.1134/s0006297909020011.

Wang Y., Weng J., Zhu C., Ai R., Zhou J., Wang C., Chen Q., Fu L. Allergenicity assessment and allergen profile analysis of different Chinese wheat cultivars. *World Allergy Organ. J.* 2021;14(7):100559. DOI 10.1016/j.waojou.2021.100559.

ORCID ID

O.V. Vishnevsky orcid.org/0000-0002-0347-2252
I.V. Chadaeva orcid.org/0000-0002-2724-5441
E.B. Sharypova orcid.org/0000-0002-1467-9312
P.M. Ponomarenko orcid.org/0000-0003-2715-9612
N.L. Podkolodny orcid.org/0000-0001-9132-7997
D.A. Rasskazov orcid.org/0000-0003-4795-0954
A.G. Bogomolov orcid.org/0000-0003-4359-6089
O.A. Podkolodnaya orcid.org/0000-0003-3247-0114
L.K. Savinkova orcid.org/0000-0003-4543-4104
E.V. Zemlyanskaya orcid.org/0000-0002-4304-1190
M.P. Ponomarenko orcid.org/0000-0003-1663-318X

Acknowledgements. Russian Science Foundation grant No. 20-14-00140 supported this study. The authors are thankful to the multi-access Center "Bioinformatics" for the use of computational resources as supported by Russian government project FWNR-2022-0020 and the Russian Federal Science and Technology Program for the Development of Genetic Technologies.

Conflict of interest. The authors declare no conflict of interest.

Received September 16, 2022. Revised November 29, 2022. Accepted November 30, 2022.