COMPREHENSIVE PLANT BIOLOGY

http://doi.org/10.2298/CPB2502181D

(2025) 49 (2): 181–193
ISSN 3042-3201
formerly Botanica Serbica

Original Scientific Paper

What genes are involved in the adaptive intraspecies divergence of *Aegilops tauschii*?

Alexander Ju. Dudnikov* and Gennady V. Vasiliev

Institute of Cytology and Genetics, Novosibirsk 630090 Russia

Corresponding author: dudnikov@bionet.nsc.ru

ABSTRACT:
The coding sequences of 27,448 genes each of 72 accessions of *Aegilops tauschii*, representing the entire species range, were obtained using next generation sequencing of transcriptomes from leaf tissue. 9,660 of these genes exhibited high Ka/Ks values exceeding 0.8. 4,336 different GO terms were present in the annotations for this gene set, with 193 of them occurring statistically significantly more often among the genes with high Ka/Ks, indicating variation in what functions and processes were involved in the adaptive intraspecies divergence of *A. tauschii*. It was shown that *A. tauschii* genes with high Ka/Ks ratios are primarily those which help plants to adapt to different ecological conditions and respond to changing environments. Those genes whose allelic variation played an essential role in *A. tauschii* intraspecies divergence included R-Genes (marked with GO terms: GO:0043531, ADP binding, + GO:0051707, response to other organisms, + GO:0006952, defence response); as well as genes involved in photosynthesis, response to light stimulus, the respiratory electron transport chain, transcriptional regulation, microtubule motor activity, and the regulation of the jasmonic acid mediated signalling pathway, etc.

Keywords: *Aegilops tauschii*, intraspecies divergence, Ka/Ks, natural selection, next generation sequencing, resistance genes, transcriptome sequences.

Received:
25 November 2024

Revision accepted:
10 April 2025

UDC: 582.542.11:575.113

INTRODUCTION

Aegilops tauschii Coss. (syn. *Aegilops squarrosa* auct. non L.) is a wild diploid, mostly self-pollinating goat-grass (genome DD, 2n = 14) which donated its D genome to common wheat, *Triticum aestivum* L. (genome AABBDD, 2n = 42). As the most important wild progenitor of common wheat, *Aegilops tauschii* is a key source of agriculturally important genes for its improvement (Kimber & Feldman 1987; Kilian *et al.* 2011). It inhabits a vast area in Central Eurasia, expanding from Turkey and Georgia to Kyrghystan and the Yellow River region (China), with a high level of genetic variation (Dudnikov 1998, 2021; Pestsova *et al.* 2000; Mizuno *et al.* 2010; Matsuoka *et al.* 2013; Zhou *et al.* 2021). The species has occupied its range since the end of the Tertiary period (Zhukovsky 1928). It is distributed in hilly or mountainous regions where its primary habitats are dwarf-shrub steppe-like formations, usually just below the edge of a forest belt. Such habitats have not been disturbed by man since they are not suitable for any kind of agriculture. Therefore *A. tauschii* is a perfect model object for evolutionary studies.

The species is presented by its two subspecies: *A. tauschii* subsp. *tauschii* and *A. tauschii* subsp. *strangulata* (Eig) Tzvelev (Eig 1929; Hammer 1980;

Published by Institute of Botany and Botanical Garden Jevremovac, Belgrade
Faculty of Biology, University of Belgrade, Belgrade, Serbia
www. cpb.bio.bg.ac.rs

Jaaska 1981). Molecular-genetic methods display *Aegilops tauschii* subsp. *tauschii* as gene-pool TauL1 and *A. tauschii* subsp. *strangulata* is presented as gene-pools TauL2 + TauL3. TauL3 is the most ancient, relict gene-pool of *A. tauschii* (Dudnikov 1998, 2012, 2017; Pestsova *et al.* 2000; Matsuoka *et al.* 2013)). Studies have shown that the intraspecies divergence of *A. tauschii* was mostly adaptive, and different phylogenetic lineages of the species essentially differ ecologically (Dudnikov 2014, 2021). Therefore, it seems interesting to investigate which genes were involved in the adaptive evolution of *A. tauschii*. (It is of interest both from a theoretical point of view, to understand the evolution of the plant species, and also from an applied point of view, since genetic variation, essential for the adaptation of *A. tauschii* in the wild, could also be promising in *T. aestivum* breeding). The peculiarities of natural selection could be reflected by the ratio of non-synonymous (Ka) to synonymous (Ks) substitutions in the coding sequence of a gene. A low Ka/Ks value indicates that the gene variation is under purifying selection and the trait it encodes remains rather unchanged in the course of a species' evolutionary history. A high Ka/Ks reveals that a gene is under positive natural selection and its variation is involved in adaptive changes in the species' population (Hurst 2002). In this study we used transcriptome sequences of a set of *A. tauschii* accessions representing all entire species range, and the results are presented below.

MATERIALS AND METHODS

Plant materials. 72 *Aegilops tauschii* accessions were used in the study (Table S1). The sources of the plant material are as follows: (1) N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), (''k''); (2) Kyoto University (''KU''); (3) IPK Gatersleben (''AE''); Triticeae Research Institute, Sichuan Agricultural University ("As") and (4) the collection of Dudnikov (1998) (''t''). The accessions for the study were specially chosen using cpDNA and nuclear gene *Got2* DNA sequencing data (Dudnikov 2021; Dudnikov *et al.* 2023) to present as much *A. tauschii* genetic variability as possible in a sample of such a size.

Molecular genetic methods. *Aegilops tauschii* plants were grown in a greenhouse at a temperature of 22 degrees Celsius, 80% humidity and round the clock illumination. RNA was extracted from the leaves of two-week old plants using a Qiagen RNeasy Plant Mini Kit. The quality of the RNA samples was evaluated using an Agilent Bioanalyzer 2100. All the samples had RIN of 7.8 or higher. RNAseq library preparations were carried out with 1mkg of total RNA using a TruSeq Stranded mRNA Library Prep Kit (Illumina, USA) according to the manufacturer's instructions for barcoded libraries with minor modifications (4 min RNA fragmentation time and 12 PCR cycles were used). After normalisation the barcoded libraries were pooled and sequenced on a NextSeq550 sequencer 2 × 150 bp using NextSeq 500/550 High Output v2.5 Kit 300 cycles (Illumina).

Data analysis. The obtained raw sets of about 20 billion 150 b.p. paired-end reads were analysed as follows. The adapter sequences were removed using Cutadapt (Martin 2011), followed by trimming and subsequent filtering with PRINSEQ. De novo transcriptome assembly was carried out with the help of Trinity (Grabherr *et al.* 2011; Haas *et al.* 2013). The transcriptomes were then aligned to the reference genome using minimap2 (Li 2018). (Aegilops_tauschii.Aet_v4.0.dna.toplevel.fa.gz genome from Ensembl Plants was used as a reference.) Samtools were used to convert the alignments from SAM to BAM format and then to sort BAM for SNP calling. SNPs calling was carried out using bcftools, and bcftools was also used to filter SNPs, converting

Table 1. GO terms with the most distinct effects in *Aegilops tauschii* adaptive divergence: GO terms with "Ghl" higher than 1.7 and a p-value lower than 0.009 are presented. ("nh" - the number of times the GO term occurred among genes with Ka/Ks higher than 0.8 and some GO term(s) available (6,565 in total); "nl" - the number of times the GO term occurred among genes with Ka/Ks equal to or lower than 0.8 and having some GO term(s) (12,137 in total); Ghl = ((nh/(6565)/((nl + 1)/(12137)); "F" – molecular function; "C" – cellular component; "P" – biological process.)

GO term	nh	nl	Ghl	Barnard's test score	p-value	GO term type	GO term definition
GO:0043531	310	159	3.581	14.243	1.097e-43	F	ADP binding
GO:0051707	287	166	3.177	12.754	2.057e-35	P	response to other organisms
GO:0006952	337	229	2.708	12.369	1.903e-33	P	defence response
GO:0006353	46	28	2.932	4.886	1.122e-06	P	DNA-templated transcription termination
GO:0008194	98	97	1.848	4.456	5.955e-06	F	UDP-glycosyltransferase activity
GO:0007166	70	66	1.931	4.013	4.751e-05	P	cell surface receptor signalling pathway
GO:2000022	16	7	3.697	3.464	0.000446	P	regulation of jasmonic acid mediated signalling pathway
GO:0009882	6	0	11.09	3.331	0.000528	F	blue light photoreceptor activity
GO:0003690	49	47	1.887	3.280	0.000595	F	double-stranded DNA binding
GO:0003777	27	21	2.268	3.073	0.00118	F	microtubule motor activity
GO:0009538	5	0	9.243	3.040	0.00134	C	photosystem I reaction centre
GO:0008017	50	51	1.777	3.041	0.00134	F	microtubule binding
GO:0007018	26	21	2.184	2.907	0.00207	P	microtubule-based movement
GO:0043488	11	5	3.389	2.821	0.00269	P	regulation of mRNA stability
GO:0070940	6	1	5.546	2.806	0.00285	F	dephosphorylation of RNA polymerase II C-terminal domain
GO:0036297	6	1	5.546	2.806	0.00285	P	DNA interstrand cross-link repair
GO:0019344	4	0	7.394	2.719	0.00364	P	cysteine biosynthetic process
GO:0000226	18	13	2.377	2.681	0.00395	P	microtubule cytoskeleton organisation
GO:0022904	8	3	3.697	2.615	0.00476	P	respiratory electron transport chain
GO:0004402	8	3	3.697	2.615	0.00476	F	histone acetyltransferase activity
GO:1990247	9	4	3.328	2.579	0.00530	F	N6-methyladenosine-containing RNA reader activity
GO:0009522	21	18	2.043	2.455	0.00744	C	photosystem I
GO:0045892	27	26	1.849	2.420	0.00822	P	negative regulation of DNA-templated transcription

BCF into final VCF. VCF was annotated using SnpEff (Cingolani *et al.* 2012). The resulting snpEff_genes.txt file containing Ka and Ks values for each gene studied was used for further analysis using original programmes for Python (File S1). Since many genes were presented by several different transcripts, the Ka/Ks value for each gene was calculated as follows: (average Ka value per one transcript) / ((average Ks value per one transcript) + 1). The Go term(s) for each gene under study were taken from Ensembl Plants annotation for Aegilops_tauschii.Aet_v4.0 genome (https://ftp.ebi.ac.uk/ensemblgenomes/pub/release-57/plants/embl/aegilops_tauschii/). The value 0.8 was taken as an arbitrary dividing line between high and low Ka/Ks (Jia *et al.* 2013). For each GO term four values were used to test the hypothesis as to whether the GO term occurs more frequently among genes with high Ka/Ks: the total number of genes with low Ka/Ks and GO terms available ("nlow"); the total number of genes with high Ka/Ks and GO terms available ("nhigh"); the number of occurrences of the particular GO term among genes with low Ka/Ks ("nl"); the number of occurrences of the particular GO term among genes with high Ka/Ks ("nh")) The hypothesis was tested using Barnard's exact test, which is a more powerful alternative to Fisher's exact test for 2 × 2 contingency tables (Barnard 1947; Mehta & Pralay 2003), using SciPy v1.12.0 for python (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.barnard_exact.html#scipy.stats.barnard_exact).

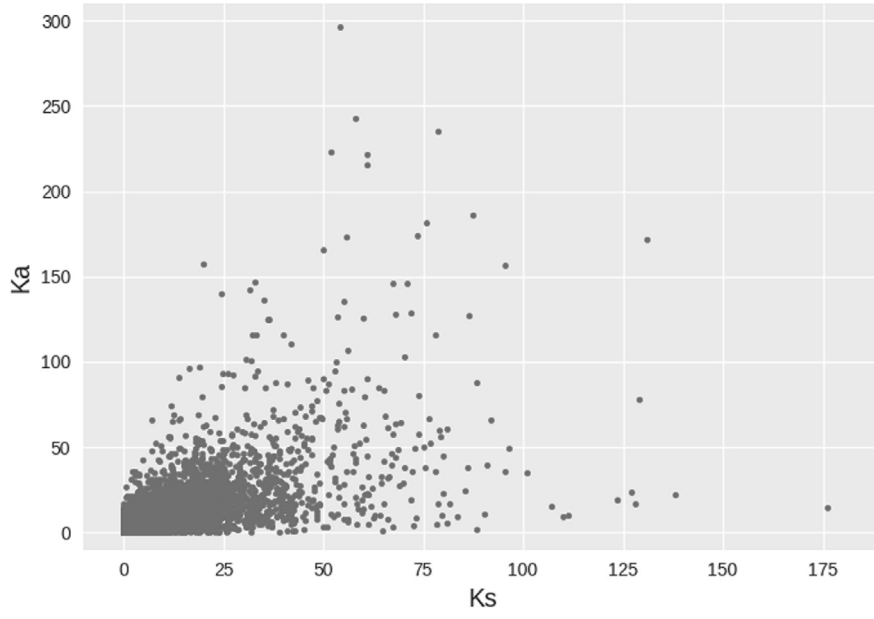


Fig. 1. 27,448 genes of *Aegilops tauschii* are placed on a scatter plot in accordance with their Ka and Ks values.

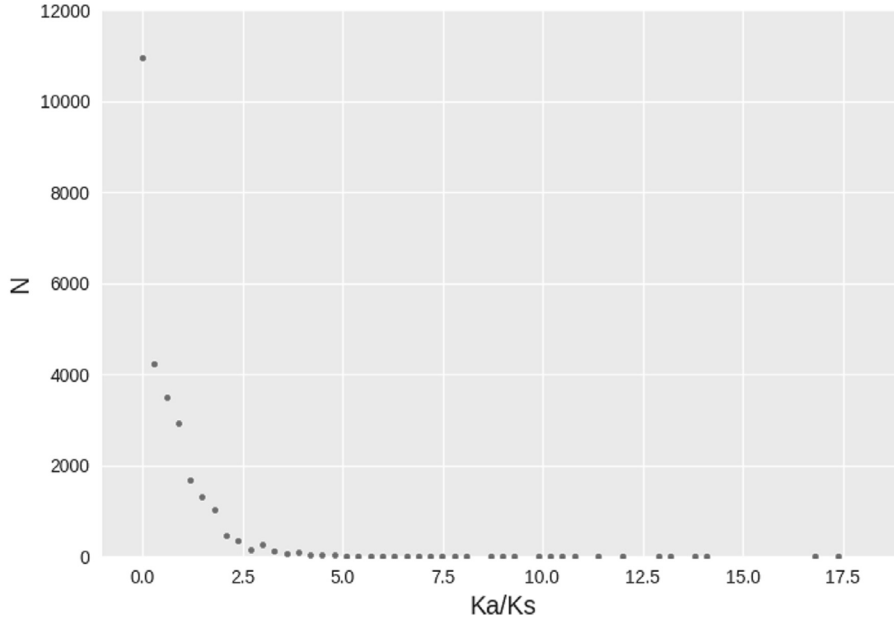


Fig. 2. The distribution of the occurrences of Ka/Ks values among 27,448 genes of *Aegilops tauschii*. The circles on the plot show the number of genes exhibiting Ka/Ks values within the intervals of 0.3.

RESULTS

The coding sequences of 27,448 genes were obtained for each of the 72 *A. tauschii* accessions. 9660 of these genes exhibited high Ka/Ks values exceeding 0.8. File S2 presents, for each gene: its ID; chromosomal location; Ka and Ks (their average values per transcript); the Ka/Ks ratio (defined as Ka/(Ks + 1)); and GO terms. The list ranged from highest to lowest Ka/Ks. A scatter plot of Ka vs. Ks values for all the genes is shown in Fig. 1. The frequency distribution of Ka/Ks is shown in Fig. 2. As expected, it is L-shaped.

The genes studied had 4,336 different GO terms in total. (Among the 9,660 genes with high Ka/Ks, – 6,565 (nhigh) had GO terms, while among the 17,788 genes with low Ka/Ks, 12,137 (nlow) had GO terms.) In File S3 each of these GO terms is presented together with the number of occurrences among genes with high Ka/Ks (i.e. higher than 0.8) (nh), the number of times it occurred among genes with low Ka/Ks (nl), and the ratio: Ghl = ((nh/(nhigh)/((nl + 1)/

Table 2. *Aegilops tauschii* genes exhibiting GO:0009882 or GO:0009538. (The average Ka and Ks values per transcript are presented.)

GO term	Gene ID	Location	Ka	Ks	Ka/Ks	Description
GO:0009882 blue light photoreceptor activity	AET5Gv20917800	5D:479446729-479723471	16.04	9.31	1.72	
	AET2Gv20723000	2D:405917414-405922243	4.7	4.7	1.00	
	AET2Gv20613400	2D:333812790-333832637	11.74	4.13	2.84	Phototropin-2
	AET3Gv20935900	3D:535845557-535862024	10.32	4.32	2.39	
	AET6Gv20603300	6D:322175163-322181028	2.8	0.7	4.00	Cryptochrome 2, blue light photoreceptor, Promotion of flowering time
	AET6Gv20511100	6D:243988399-243998488	5.11	4.88	1.05	
GO:0009538 photosystem I reaction centre	AET3Gv20112600	3D:21931135-21932637	49.5	18.0	2.75	
	AET5Gv21059900	5D:522276054-522278708	56.25	11.75	4.79	
	AET1Gv20942900	1D:473773388-473776240	21.66	8.66	2.50	Photosystem I reaction centre subunit VI, chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein)
	AET5Gv21114600	5D:538800495-538801666	68.75	12.5	5.50	
	AET5Gv20956200	5D:492422895-492427465	166.0	50.0	3.32	

(nlow)). The list was sorted in descending order according to this ratio, from high to low values. From this list, the top 1,180 GO terms with the highest Ghl are presented in File S4. Each entry contains: the GO term name; nh; nl; the Ghl ratio; the Barnard's test score statistic; and the p-value. (The list ranged from lower to higher p-values). It can be seen that 193 of these GO terms occurred significantly ($P < 0.05$) more frequently among genes with high Ka/Ks, indicating that these functions (processes and cellular components) were important in *A. tauschii* adaptive intraspecies divergence (Fig. 3). Table S2 presents all these 193 GO terms, ranged in accordance with their p-values, while Table 1 provides a shorter list of only those GO terms with a "Ghl" level higher than 1.7 and a p-value lower than 0.009. (These GO terms are located in the upper left part of the plot from Fig. 3a). Among the GO terms with the lowest p-values, GO:0006952, "defence response", took third place ($P < 1.903 \times 10^{-33}$), while GO:0051707,"response to other organism", ranked second place ($P < 2.057 \times 10^{-35}$ The lowest p-value, 1.097×10^{-43}, was recorded for GO:0043531, "ADP binding". A total of 469 of the genes studied had this GO term, with 310 of them having high Ka/Ks, indicating that this GO term occurred 3.58 times more frequently among genes with high Ka/Ks in comparison with those with low Ka/Ks (Table 1). (These three GO terms could be considered together since many genes possess all of them (Fig. 4, File S2)).

GO:0009882, "blue light photoreceptor activity", (11.092; 0.0005) (the first value in brackets is the Ghl ratio; the second is the p-value) had the highest nh/(nl + 1) value. There are six genes with this GO term in the Aegilops_tauschii.

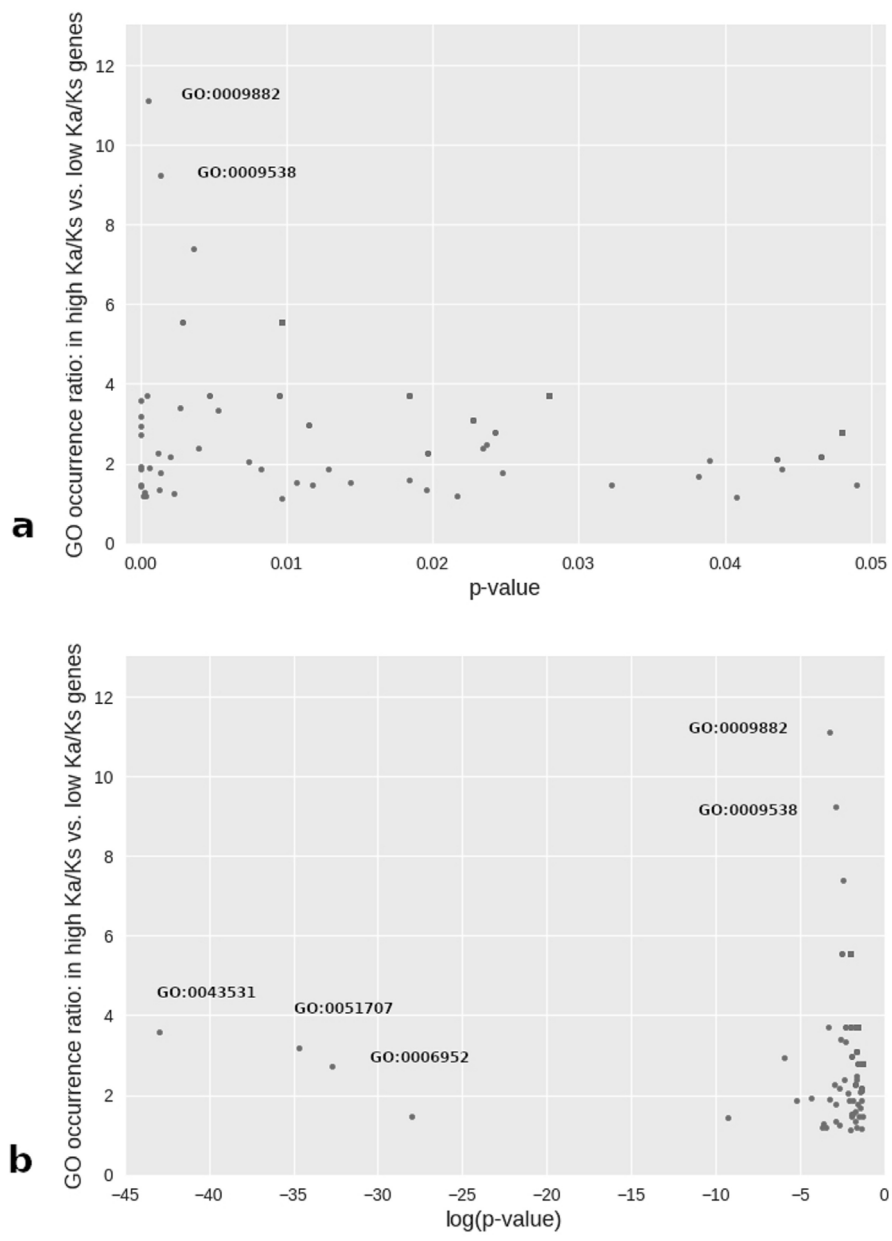


Fig. 3. The scatter plot presents the GO terms which occurred significantly ($P < 0.05$) more frequently among *Aegilops tauschii* genes with high Ka/Ks values (i.e. higher than 0.8) than among those with low Ka/Ks. The Y axis shows the ratio Ghl = (nh/nhigh)/((nl + 1)/nlow), where "nhigh" and "nlow" denote the number of genes with high and low Ka/Ks, respectively, and with GO term information available; "nh" and "nl" indicate the number of times the GO term occurred among genes with high and low Ka/Ks, respectively. The X axis represents either (a) the p-value or (b) the log(p-value).

Aet_v4.0 genome, all of them present in the list of our genes studied (File S2), and all exhibiting high Ka/Ks (Tables 1 & 2). GO:0009538, "photosystem I reaction centre", (9.2437; 0.0013) took second place. There are five genes with this GO term in the *A. tauschii* genome, all of which are present in the list of our genes studied (File S2) and all with high Ka/Ks (Tables 1 & 2). Table 1 also presents several other GO terms whose effects in *A. tauschii* adaptive divergence were statistically very distinct.

The data obtained indicated that the *A. tauschii* gene polymorphisms which played an important role in the intraspecies adaptive divergence, deal with the following processes, functions and cellular components.
(Some of the GO terms listed in Tables 1 and S2 are presented below. The numbers in brackets indicate: the number of genes with the given GO term among the set of 6,565 genes with high Ka/Ks and some GO term(s) available; the number of genes with this GO term among the set of 12,137 genes with

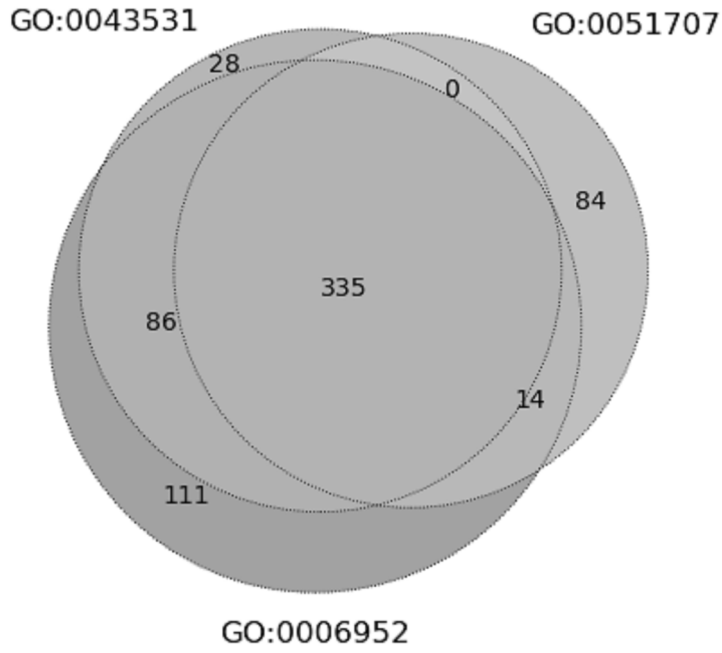


Fig. 4. The number of genes (among the 27,448 genes studied) exhibiting GO terms: GO:0043531, ADP binding; GO:0051707, response to other organisms; GO:0006952, defence response.

low Ka/Ks and some GO term(s) available; and the probability (p-value) of this happening by chance, respectively. The letters "P", "F", "C" denote "biological process", "molecular function" and "cellular component", respectively.)

Photosynthesis:
GO:0009507 (168; 229; 0.00132) chloroplast (C); GO:0009538 (5; 0; 0.00134) photosystem I reaction centre (C); GO:0009522 (21; 18; 0.00744) photosystem I (C); GO:0015979 (60; 75; 0.01175) photosynthesis (P); GO:0009535 (35; 40; 0.01839) the chloroplast thylakoid membrane (C); GO:0010196 (5; 2; 0.0227) nonphotochemical quenching (P); GO:0009570 (20; 20; 0.02477) chloroplast stroma (C); GO:1904821 (2; 0; 0.028) chloroplast disassembly (P); GO:1902326 (2; 0; 0.028) positive regulation of the chlorophyll biosynthetic process (P); GO:0047746 (2; 0; 0.028) chlorophyllase activity (F); GO:0009579 (38; 47; 0.03226) thylakoid (C); GO:0019253 (3; 1; 0.0480) reductive pentose-phosphate cycle (P); GO:0045036 (3; 1; 0.0480) protein targeting to chloroplast (P).

Cell energy flow:
GO:0022904 (8; 3; 0.00476) the respiratory electron transport chain (P); GO:0019646 (2; 0; 0.028) the aerobic electron transport chain (P); GO:0045275 (2; 0; 0.028) respiratory chain complex II (C).

Response to biotic and abiotic factors:
GO:0051707 (287; 166; 2.057×10^{-35}) response to other organisms (P); GO:0006952 (337; 229; 1.903×10^{-33}) defence response (P); GO:0050896 (49; 58; 0.01068) response to stimulus (P); GO:0048544 (46; 55; 0.01438) recognition of pollen (P); GO:0031347 (11; 8; 0.01963) regulation of defence response (P); GO:0010118 (5; 2; 0.02273) stomatal movement (P); GO:0031540 (2; 0; 0.028) regulation of the anthocyanin biosynthetic process (P); GO:0009624 (2; 0; 0.028) response to nematode (P).

ADP binding:
(It should be part of the previous section, as explained further in the text.)
GO:0043531 (310; 159; 1.097×10^{-43}) ADP binding (F).

Regulation of the jasmonic acid mediated signalling pathway:
(It should be part of the response to biotic and abiotic factors section since jasmonic acid and its precursors and derivatives, referred to as jasmonates are known to be important molecules in the regulation of numerous physiological processes in plant growth and development, and especially the mediation of plant responses to biotic and abiotic stresses (Ruan *at al.* 2019).)
GO:2000022 (16; 7; 0.000446) regulation of the jasmonic acid mediated signalling pathway (P);

Response to light stimulus:
GO:0009882 (6; 0; 0.000529) blue light photoreceptor activity (F); GO:0009785 (6; 2; 0.009475) blue light signalling pathway (P); GO:0048579 (4; 1; 0.0184) negative regulation of long-day photoperiodism, flowering (P); GO:0009416 (11; 8; 0.0196) response to light stimulus (P); GO:0009637 (5; 2; 0.0227) response to blue light (P);

Cysteine biosynthetic process:
(It could probably be included in the previous section according to the explanation presented further in the text.)
GO:0019344 (4; 0; 0.00364) the cysteine biosynthetic process (P);

Regulation of transcription:
GO:0006355 (494; 634; 5.0396×10^{-10}) the regulation of DNA-templated transcription (P); GO:0006353 (46; 28; 1.122×10^{-6}) DNA-templated transcription termination (P); GO:0003700 (236; 344; 0.00229) DNA-binding transcription factor activity (F); GO:0043488 (11; 5; 0.00269) the regulation of mRNA stability (P); GO:0070940 (6; 1; 0.00285) dephosphorylation of the RNA polymerase II C-terminal domain (F); GO:0036297 (6; 1; 0.00285) interstrand cross-link repair (P); GO:1990247 (9; 4; 0.005296) N6-methyladenosine-containing RNA reader activity (F); GO:0045892 (27; 26; 0.008216) the negative regulation of DNA-templated transcription (P); GO:0008420 (6; 2; 0.009475) RNA polymerase II CTD heptapeptide repeat phosphatase activity (F); GO:1900369 (3; 0; 0.00970) the negative regulation of post-transcriptional gene silencing by regulatory ncRNA (P); GO:0003714 (12; 11; 0.04388) transcription corepressor activity (F).

Cell surface receptor signalling pathway:
(This could be included in the previous section, since this pathway may culminate in the regulation of transcription.)
GO:0007166 (70; 66; 4.75×10^{-5}) cell surface receptor signalling pathway (P).

Histone acetyltransferase activity:
(This could be included in the regulation of transcription section, since histone acetylation influences the levels of gene transcription (Verdone *et al.* 2006))
GO:0004402 (8; 3; 0.00476) histone acetyltransferase activity (F); GO:0070461 (4; 1; 0.0184) SAGA-type complex (C);

Double-stranded DNA binding:
(This could be included in the regulation of transcription section, since a set of transcription termination factors belongs to GO:0003690)
GO:0003690 (49; 47; 0.000596) double-stranded DNA binding (F).

Microtubule processes and functions:
GO:0003777 (27; 21; 0.00118) microtubule motor activity (F); GO:0008017 (50; 51; 0.00134) microtubule binding (F); GO:0007018 (26; 21; 0.00207) microtubule-based movement (P); GO:0000226 (18; 13; 0.00395) microtubule cytoskeleton organisation (P); GO:0003774 (23; 22; 0.01289) cytoskeletal motor activity (F).

Plant organ morphogenesis:
GO:1905392 (2; 0; 0.028) plant organ morphogenesis (P); GO:0010051 (2; 0; 0.028) xylem and phloem pattern formation (P); GO:2000014 (2; 0; 0.028) the regulation of endosperm development (P); GO:0010222 (2; 0; 0.028) stem vascular tissue pattern formation (P); GO:0080006 (2; 0; 0.028) internode patterning (P); GO:2000280 (3; 1; 0.0480) the regulation of root development (P).

Synthesis of wax:
GO:1904278 (2; 0; 0.028) the positive regulation of the wax biosynthetic process (P).

DISCUSSION

It can be seen that among the genes whose polymorphisms played an essential role in *A. tauschii* intraspecies divergence are those involved in photosynthesis, cell energy flow, transcription regulation, response to biotic and abiotic factors, response to light stimulus, and microtubule motor activity, etc. The genes in *A. tauschii* with high Ka/Ks ratio are predominantly those which facilitate the plant's adaptation to different ecological conditions and effective responses to environmental changes.

Previous studies have shown that the TauL1, TauL2 and TauL3 gene-pools of *A. tauschii* essentially differ in their genetic expression patterns. More than two thousand genes exhibit statistically significantly differences in their genetic expression levels in the leaves of TauL1, TauL2 and TauL3 (Dudnikov *et al.* 2024). These findings are in line with those obtained in the present study, indicating that positive natural selection in genes regulating transcription in *A. tauschii* have made a considerable impact on the species' adaptive divergence.

Those genes which control response to light stimulus have been subject to positive natural selection in *A. tauschii*. (Among them is the gene AET-2Gv20613400 encoding Phototropin-2, and the gene AET6Gv20603300 encoding Cryptochrome-2 (Table 2)). They affect flowering time which is of particular importance for *A. tauschii*. The species inhabits environments characterised by moist winters and hot dry summers. Plants which flower too early, before the optimal period for tillering, will form fewer seeds than they potentially could. Conversely, plants which start flowering late do not have enough time to form seeds before the dry season begins, and consequently may form no seeds at all. Therefore, the precise regulation of the flowering time for each environment is crucial for *A. tauschii.*

Phototropins are also involved in chloroplast movements in response to changes in light intensity, light-induced stomatal opening, etc.

Genes encoding the cysteine biosynthetic process exhibit high Ka/Ks in *A. tauschii* (Table 1). The reason for this could be the fact that cysteine residue in the photosensoric part of phototropin plays an important role in its functioning. Phototropin is a light-activated autophosphorylating serine/threonine kinase which binds two flavin mononucleotide (FMN) molecules which function as blue light-absorbing chromophores. In darkness the chromophore of phototropin is noncovalently bound to the protein and absorbs light at 447 nm. Illumination activates FMN to form a covalent bond with the cysteine residue in the photosensoric part of phototropin. (Briggs *et al.* 2001).

As well as the blue light mediated regulatory pathway, the regulation of the jasmonic acid mediated signalling pathway has also been subject to positive natural selection during the course of the intraspecies divergence of *A. tauschii*. This is not unexpected, taking into account that jasmonic acid and its precursors and derivatives are important molecules in the regulation of many physiological processes in plant growth and development, and especially the mediation of plant responses to biotic and abiotic stresses. Jasmonic acid signalling plays a central role in plant defences against necrotrophic pathogens and herbivorous insects, which target both roots and shoots. This pathway is also activated during interactions with beneficial microbes which may lead to induced systemic resistance. (Ruan *et al.* 2019).

The data obtained in this study indicated that many resistance genes (R-Genes) (marked with GO terms: GO:0043531, ADP binding, + GO:0051707, response to other organisms, + GO:0006952, defence response) were under positive natural selection in *A. tauschii* (Table 1, Fig 3b).

In the *A. tauschii* genome, as annotated in Ensemble Plants, 639 genes with GO:0043531, ADP binding, are presented, one of which, AET3Gv21194500, had an available description: NB-ARC domain containing protein (https://plants.ensembl.org/Aegilops_tauschii/Gene/Summary?g=AET3Gv21194500;r=3D:611383074-611395000;t=AET3Gv21194500.1;db=core). The gene was also associated with GO terms GO:0051707, response to other organisms, and GO:0006952, defence response. The AET3Gv21194500 gene is also included in our dataset, exhibiting Ka, Ks, Ka/(Ks + 1) - 10.76, 2.53 and 3.05 respectively (File S2). The NB-ARC domain is a functional ATPase domain, and its nucleotide-binding state is thought to regulate R protein activity. Resistance (R) proteins in plants are involved in pathogen recognition and the subsequent activation of innate immune responses (van Ooijen *et al.* 2008). R-genes are of considerable interest in crop breeding, providing a large part of the immunity required by agricultural pathosystems (Arora *et al.* 2019).

Microtubule motor activity provides the intracellular transport of vesicles and organelles through the cytoplasm. It is required for maintaining homeostasis within the cell by responding to physiological signals (Barlan & Gelfand 2017*).* In this study microtubule motor activity was found to be under positive selection in *A. tauschii* (Table 1).

Previous research has shown that the trait of "waxiness" exhibits specific geographic patterns in *A. tauschii*. Waxy plants have been observed only in *A. tauschii* subsp. *strangulata*, and only in relatively dry habitats which were not optimal for this subspecies (Dudnikov 2011). Two genes in the *A. tauschii* genome Aegilops_tauschii.Aet_v4.0 are known to control the positive regulation of the wax biosynthetic process (GO:1904278): AET4Gv20312600 and AET6Gv20348000. Both genes are included in the set of genes investigated in this study, and both exhibited high Ka/Ks values (File S2).

It can be seen that the vast range of the *A. tauschii* set of accessions used in this study reveals the adaptive allelic polymorphisms of numerous genes involved in the functions and processes which are of key importance for adaptation to different environments, effective responses to environmental changes, and the energy efficiency of plants. The information obtained and the plant material of *A. tauschii* investigated in this study could be further used for both detailed studies of the molecular genetic basis of adaptation and for applied purposes. *A. tauschii* genetic lines with known geographical origin and distinct allelic variants of genes encoding different regulatory pathways may serve for the development of new prospective varieties of *T. aestivum*.

The requirements of plants in nature can differ greatly from those valued in agriculture, e.g. agriculturally important traits such as soft glumes and tough rachis are deleterious for wild *Triticeae* species. However, as demonstrated in the present study, in general, the genes whose variation was particu-

larly important for *A. tauschii* adaptive evolution in the wild are extremely also interesting for *T. aestivum* breeding. Notably, resistant genes which have a very important and distinct effect in *A. tauschii* evolution (Fig. 3b) are nowadays among the most popular and effective tools for creating new prospective varieties of common wheat. The set of 72 genetic lines of *A. tauschii*, specially chosen in order to present as much genetic variability of the species as possible for a sample of such a size (Table S1) together with information about 27,448 genes, their chromosomal location, function and Ka/Ks values (File S2), could serve as the perfect tool for the breeding of *T. aestivum.* When a breeder transfers a segment of an *A. tauschii* chromosome into *T. aestivum,* its gene content is already known. In addition, the high value of Ka/Ks of the gene (File S2) indicates that its different alleles obtained from ecologically different lineages of *A. tauschii,* e.g. from the TauL1 and TauL2 gene-pools, will have different effects on plant phenotype.

Acknowledgments – We would like to express our sincere gratitude to Dr. Olesya Y. Shoeva for her help in the course of the study. NGS sequencing was performed in the ICG Joint Centre for Genome Studies. The work was supported by ICG budget project FWNR-2022-0016.

REFERENCES

Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, *et al*. 2019. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology* **37**(2): 139–143. https://doi.org/10.1038/s41587-018-0007-9

Barlan K & Gelfand VI. 2017. Microtubule-Based Transport and the Distribution, Tethering, and Organization of Organelles. *Cold Spring Harbor Perspectives in Biology* **9**(5): a025817. https://doi.org/10.1101%2Fcshperspect.a025817

Barnard GA. 1947. Significance Tests for 2x2 Tables. *Biometrika* **34**(1/2): 123–138.

Briggs WR, Christie JM & Salomon M. 2001. Phototropins: a new family of flavin-binding blue light receptors in plants. *Antioxid Redox Signal* **3**(5):775–788. https://doi.org/10.1089/15230860152664975.

Cingolani P, Platt A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X & Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**(2):80–92. https://doi.org/10.4161/fly.19695.

Dudnikov AJ. 1998. Allozyme variation in Transcaucasian populations of *Aegilops squarrosa*. *Heredity* **80**: 248–258.

Dudnikov AJ. 2011. Waxiness in *Aegilops tauschii:* Its Occurrence in Natural Habitats of the Species. *Cereal Research Communications* **39**: 283–288. https://doi.org/10.1556/crc.39.2011.2.12

Dudnikov AJ. 2012. Chloroplast DNA non-coding sequences variation in *Aegilops tauschii* Coss.: evolutionary history of the species. *Genetic Resources and Crop Evolution* **59**: 683–699. https://doi.org/10.1007/s10722-011-9711-8

Dudnikov AJ. 2014. *Aegilops tauschii* Coss.: allelic variation of enzyme-encoding genes and ecological differentiation of the species. *Genetic Resources and Crop Evolution* **61**: 1329–1344. https://doi.org/10.1007/s10722-014-0115-4

Dudnikov AJ. 2017. Polymorphism of Got2 DNA sequences sheds light on *Aegilops tauschii* Coss. intraspecies divergence and origin of *Triticum aestivum* L. *Genetic Resources and Crop Evolution* **64**: 1623–1640. https://doi.org/10.1007/s10722-016-0461-5

Dudnikov AJ. 2021. *Aegilops tauschii* Coss. molecular phylogeny: nuclear gene *Got2* versus chloroplast DNA data. *Genetic Resources and Crop Evolution* **68**: 2469–2482. https://doi.org/10.1007/s10722-021-01143-2

Dudnikov AJ, Hao M & Liu D. 2023. Genetic variation in local populations of *Aegilops tauschii* Coss. helps to elucidate the evolutionary history of the species. *Genetic Resources and Crop Evolution* **70**: 2367–2373. https://doi.org/10.1007/s10722-023-01565-0

Dudnikov AJ, Vasiliev GV, Hao M, Liu D, Xing F, Mansoury M, Afonnikov DA & Shmakov NA. 2024. TauL1, TauL2 and TauL3 gene-pools of *Aegilops tauschii* Coss. essentially differ in their genetic expression patterns. *Botanica Serbica* **48**(2): 239–246.

https://doi.org/10.2298/BOTSERB2402239D

Eig A. 1929. Monographisch-kritische Übersicht der Gattung *Aegilops*. *Repertorium Speciorum Novarum Regni Vegetabilis. Beihefte* **55**: 1–228.

Grabherr MG, Haas BJ, Yassour M, *et al*. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**(7): 644–652. https://doi.org/10.1038/nbt.1883

Haas BJ, Papanicolaou A, Yassour M, *et al*. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**(8):1494–1512. https://doi.org/10.1038/nprot.2013.084

Hammer K. 1980. Vorarbeiten zur monographischen Darstellung von Wildpflanzensortimenten: *Aegilops* L. *Kulturpflanze* **28**: 33–180.

Hurst DL. 2002. The *Ka*/*Ks* ratio: diagnosing the form of sequence evolution. *Trends in Genetics* **18**(9): 486–487. https://doi.org/10.1016/S0168-9525(02)02722-1

Jaaska V. 1981. Aspartate aminotransferase and alcohol dehydrogenase enzymes: intraspecific differentiation in *Aegilops tauschii* and the origin of the D genome polyploids in the wheat group. *Plant Systematics and Evolution* **137**: 259–273. https://doi.org/10.1007/BF00982790

Jia J, Zhao S, Kong X. *et al*. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**: 91–95. https://doi.org/10.1038/nature12028

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K & Ozkan H. 2011. *Aegilops*. In: Kole C (ed.), *Wild crop relatives: genomic and breeding resources: Cereals,* pp. 1–76, Springer Verlag, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14228-4_1

Kimber G & Feldman M. 1987. *Wild wheat. An introduction*. College of Agricultural University of Missouri, Columbia, Special report 353.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**: 10–12. https://doi.org/10.14806/ej.17.1.200

Matsuoka Y, Nasuda S, Ashida Y, Nitta M, Tsujimoto H, Takumi S & Kawahara T. 2013. Genetic Basis for Spontaneous Hybrid Genome Doubling during Allopolyploid Speciation of Common Wheat Shown by Natural Variation Analyses of the Paternal Species. *PloS One* **8**: e68310. https://doi.org/10.1371/journal.pone.0068310

Mehta CR & Pralay S. 2003. Conditional versus unconditional exact tests for comparing two binomials. *Cytel Software Corporation* **675**: 1–5.

Mizuno N, Yamasaki M, Matsuoka Y, Kawahara T & Takumi S. 2010. Population structure of wild wheat D-genome progenitor *Aegilops tauschii* Coss.: implications for intraspecific lineage diversification and evolution of common wheat. *Molecular Ecology* **19**: 999–1013. https://doi.org/10.1111/j.1365-294X.2010.04537.x

Pestsova E, Korzun V, Goncharov NP, Hammer K, Ganal MW & Roder MS. 2000. Microsatellite analysis of *Aegilops tauschii* germplasm. *Theoretical and Applied Genetics* **101**: 100–106. https://doi.org/10.1007/s001220051456

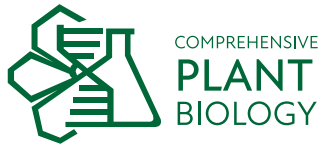
Ruan J, Zhou Y, Zhou M, Yan J, Khurshid M, Weng W, Cheng J & Zhang K. 2019. Jasmonic Acid Signaling Pathway in Plants. *International Journal of Molecular Sciences* **20**(10): 2479. https://doi.org/10.3390%2Fijms20102479

van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC & Takken FLW. 2008. Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of Experimental Botany* **59**(6): 1383–1397. https://doi.org/10.1093/jxb/ern045

Verdone L, Agricola E, Caserta M & Di Mauro E. 2006. Histone acetylation in gene regulation. *Briefings in Functional Genomics & Proteomics* **5**(3): 209–221. https://doi.org/10.1093%2Fbfgp%2Fell028

Zhou Y, Bai S, Li H, Sun G, Zhang D, Ma F, Zhao X, Nie F, Li J, Chen L, Lv L, Zhu L, Fan R, Ge Y, Shaheen A, Guo W, Huang J, Li S, Zou C & Song C. 2021. Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. *Nature Plants* **7**(6): 774–786. https://doi.org/10.1038/s41477-021-00934-w

Zhukovsky PM. 1928. A critical-systematical survey of the species of the genus *Aegilops* L. *Bulletin of Applied Botany,of Genetics and Plant Breeding* **18**: 417–609.



REZIME

Koji geni su uključeni u adaptivnu intraspecijsku divergenciju kod *Aegilops tauschii*?

Alexander Ju. Dudnikov i Gennady V. Vasiliev

Kodirajuće sekvence 27,448 gena u 72 uzorka *Aegilops tauschii*, koji predstavljaju čitav areal vrste, dobijene su korišćenjem sekvenciranja transkriptoma sledeće generacije iz tkiva lista. 9.660 ovih gena pokazalo je visoke vrednosti Ka/Ks koje prelaze 0,8. 4.336 različitih GO termina bilo je prisutno u anotacijama za ovaj skup gena, od kojih je 193 bilo statistički značajno i često se javljalo među genima sa visokim Ka/Ks, što ukazuje na varijacije u relevantnim funkcijama i procesima uključenim u adaptivnu intraspecijsku divergenciju *A. tauschii*. Pokazano je da su geni *A. tauschii* sa visokim odnosima Ka/Ks prvenstveno oni koji pomažu biljkama da se prilagode različitim ekološkim uslovima i reaguju na promenljiva okruženja. Oni geni čija je alelna varijacija igrala suštinsku ulogu u intraspecijskoj divergenciji A. tauschii uključivala je R-gene (označene GO terminima: GO:0043531, vezivanje za ADP, + GO:0051707, odgovor na druge organizme, + GO:0006952, odbrambeni odgovor); kao i gene uključene u fotosintezu, odgovor na svetlosni stimulus, respiratorni lanac transporta elektrona, transkripcionu regulaciju, motoričku aktivnost mikrotubula i regulaciju signalnog puta posredovanog jasmonskom kiselinom.

Ključne reči: *Aegilops tauschii*, intraspecijska divergencija, Ka/Ks, prirodna selekcija, sekvenciranje sledeće generacije, geni otpornosti, transkriptomske sekvence