

RESEARCH

Open Access



Proteome evaluation of homolog abundance patterns in *Arachis hypogaea* cv. Tifrunner

Zhenquan Duan^{1,2†}, Yongli Zhang^{1†}, Tian Zhang¹, Mingwei Chen¹ and Hui Song^{1*} 

Abstract

Background: Cultivated peanut (*Arachis hypogaea*, AABB genome), an allotetraploid from a cross between *A. duranensis* (AA genome) and *A. ipaensis* (BB genome), is an important oil and protein crop with released genome and RNA-seq sequence datasets. These datasets provide the molecular foundation for studying gene expression and evolutionary patterns. However, there are no reports on the proteomic data of *A. hypogaea* cv. Tifrunner, which limits understanding of its gene function and protein level evolution.

Results: This study sequenced the *A. hypogaea* cv. Tifrunner leaf and root proteome using the tandem mass tag technology. A total of 4803 abundant proteins were identified. The 364 differentially abundant proteins were estimated by comparing protein abundances between leaf and root proteomes. The differentially abundant proteins enriched the photosystem process. The number of biased abundant homeologs between the two sub-genomes A (87 homeologs in leaf and root) and B (69 and 68 homeologs in leaf and root, respectively) was not significantly different. However, homeologous proteins with biased abundances in different sub-genomes enriched different biological processes. In the leaf, homeologs biased to sub-genome A enriched biosynthetic and metabolic process, while homeologs biased to sub-genome B enriched iron ion homeostasis process. In the root, homeologs with biased abundance in sub-genome A enriched inorganic biosynthesis and metabolism process, while homeologs with biased abundance in sub-genome B enriched organic biosynthesis and metabolism process. Purifying selection mainly acted on paralogs and homeologs. The selective pressure values were negatively correlated with paralogous protein abundance. About 77.42% (24/31) homeologous and 80% (48/60) paralogous protein pairs had asymmetric abundance, and several protein pairs had conserved abundances in the leaf and root tissues.

Conclusions: This study sequenced the proteome of *A. hypogaea* cv. Tifrunner using the leaf and root tissues. Differentially abundant proteins were identified, and revealed functions. Paralog abundance divergence and homeolog bias abundance was elucidated. These results indicate that divergent abundance caused retention of homologs in *A. hypogaea* cv. Tifrunner.

Keywords: Abundance divergence, Homeolog, Paralog, Peanut

Introduction

Cultivated peanut (*Arachis hypogaea*), an important oil and protein crop, is an allotetraploid from a cross between *A. duranensis* and *A. ipaensis* [1–3]. *Arachis* spp. genomes, including *A. duranensis*, *A. ipaensis*, and *A. hypogaea* cv. Tifrunner have been sequenced [1, 2]. *Arachis hypogaea* cv. Tifrunner is highly resistant to tomato

*Correspondence: biosonghui@outlook.com

†Zhenquan Duan and Yongli Zhang contributed equally to this study

¹ Grassland Agri-Husbandry Research Center, College of Grassland Science, Qingdao Agricultural University, Qingdao, China
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

spotted wilt virus and moderately resistant to early and late leaf spot bacteria [4]. The *A. hypogaea* cv. Tifrunner genome has been used for quantitative trait locus analyses and RNA-seq assemblies [5–8]. In 2016, RNA-seq datasets from 22 different tissues of *A. hypogaea* cv. Tifrunner were sequenced under normal growth conditions, and the cleaned raw data was released on PeanutBase [9, 10]. However, there are no proteomic datasets for *A. hypogaea* cv. Tifrunner, which limits the understanding of its gene functions.

Homeolog expression bias is an important feature in polyploidy [11, 12]. *Arabidopsis suecica* (AATT genome) derives from crossing *Arabidopsis thaliana* (AA genome) and *Arabidopsis arenosa* (TT genome) [13]. An RNA-seq data showed that *A. suecica* preferentially expressed homeologous genes towards sub-genome T [14]. Nevertheless, different results in the *Gossypium hirsutum* (AADD genome), an allopolyploid from a cross of *Gossypium arboreum* (AA genome) and *Gossypium raimondii* (DD genome), is a crucial model for homeologous expression bias [11, 12, 15, 16]. The homeologs from sub-genome A had biased expression in diploid hybrids and natural allopolyploids, but the expression was reversed in synthetic allopolyploids [16]. However, homeologs from reproductive tissues of cultivated peanut, including *A. hypogaea* cv. Shitouqi, *A. hypogaea* cv. Tifrunner, and *A. hypogaea* cv. Fuhuasheng, had biased expression in sub-genome B [2, 17, 18]. To date, no study has analyzed homeolog bias using proteomic data. Therefore, this study sequenced the *A. hypogaea* cv. Tifrunner proteome from leaf and root tissues using the tandem mass tag technology. The results revealed proteome level of homolog (including paralogs and homeologs) abundance patterns, providing insights into homolog expression.

Methods

Plant materials

Sterile *A. hypogaea* cv. Tifrunner seeds were cultured on sterile, wet Petri dishes at 28 °C in the dark. When the first cotyledon expanded, the germinated plants were transferred into Hoagland solution and continually cultured at 28 °C under 16 h light/8 h dark cycles. Leaf (containing lateral stem leaf, mainstem leaf and seeding leaf) and root tissues were collected from three biological replicates, and samples were stored at – 80 °C for further use.

Proteome sequencing and assemble

The samples were mixed with 8 M Urea/100 mM Tris-Cl and treated with water bath sonication for 10 min in ice-water bath, and centrifuged for 15 min at 12,000g. After centrifugation, the supernatant was subjected to a reduction reaction (10 mM DTT, 37 °C for 1 h), followed

by an alkylation reaction (40 mM iodoacetamide, room temperature/dark place for 30 min). The protein concentration was measured by the Bradford method [19]. Total protein was released using a lysis buffer (2.5% SDS, 100 mM Tris-Cl, pH 8.0), and the protein was digested in trypsin at 37 °C. The digested protein was sequenced and quantified using TMT technology. Peptides were TMT-labeled following the manufacturer's instructions (Thermo Fisher Scientific, MA, USA). Briefly, peptides were reconstituted in TMT reagent buffer (Thermo Fisher Scientific, MA, USA), and the samples were separately labeled with different TMT-labeling reagents. The TMT-labeled samples were mixed and desalted using Sep-Pak C18 (Thermo Fisher Scientific, MA, USA). The complex mixture was fractionated using high pH reverse phase chromatography and combined into 15 fractions. Each fraction was vacuum-dried and stored at – 80 °C until mass spectrometry (MS) analysis.

LC–MS/MS datasets were produced by the Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, MA, USA) and Easy-nLC 1000 system (Thermo Fisher Scientific, MA, USA). The peptides were loaded through an auto-sampler and separated in a C18 analytical column (50 $\mu\text{m} \times 15 \text{ cm}$, C18, 2 μm , 100 Å). The mobile phases A (0.1% formic acid) and B (80% ACN, 0.1% formic acid) were used to establish the separation gradient.

The MaxQuant software (v1.66) estimated the MS raw data using the Andromeda search engine. The *A. hypogaea* cv. Tifrunner primary proteins were used to search the proteome database. Proteins denoted as decoy hits, contaminants, or only identified by site were removed. The remaining hits were used for further quantification analysis. Both proteins and peptides were filtered at 1% FDR. The raw datasets were deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the iProX partner repository [20] with the dataset identifier PXD027553.

Differentially abundant proteins

Differential abundance was calculated between leaf and root tissues using protein quantification (abundance). The leaf and root average protein abundance was estimated using three biological replicates. The differential abundance was estimated by $\log_2(\text{average abundance in leaf/average abundance in root})$. Proteins with $\log_2(\text{foldchange})$ was larger than two or less than negative two at $p \leq 0.05$ significance level using *t*-test, and they were considered differentially abundant.

GO and KEGG enrichment analyses were performed to reveal differential abundance in protein function. The GO enrichment was performed using the eggNOG-mapper tool on differentially expressed *A. hypogaea* cv. Tifrunner protein sequences. The same differentially expressed

protein sequences were uploaded to the KEGG database (<http://www.genome.jp/tools/kaas>) for KEGG annotation. The Fisher test was used for statistical analysis.

Homologous protein abundance

Paralogs are genes produced via gene duplication within a genome. Homeologs are genes formed by hybridization or polyploidization but not through gene duplication events. Paralogs and homeologs were identified in *A. hypogaea* cv. Tifrunner to reveal proteome-level homolog abundance. The *A. hypogaea* cv. Tifrunner homeologs were inferred from the *A. duranensis* and *A. ipaensis* genome sequences. The *A. duranensis*, *A. ipaensis*, and *A. hypogaea* cv. Tifrunner genome sequences were downloaded from the Peanutbase database (<https://www.peanutbase.org/>) [1, 2, 9]. Homologs were identified using the previously described BLAST-based method [10, 21, 22]. The following thresholds were used: (1) The alignment sequence length is >80% in each sequence. (2) the sequence similarity is >80%, and (3) the E-value is $1E-10$.

This study also analyzed the paralog divergence and homeolog abundance bias. A paralogous pair was considered abundantly divergent when the absolute value of $\log_2(\text{copy1 abundance}/\text{copy2 abundance})$ was larger than two in the paralogous pair at $p \leq 0.05$ significance level using the *t*-test. For $\log_2(\text{sub-genome A abundance}/\text{sub-genome B abundance})$ values was larger than two or less than negative two in a homeologous pair at $p \leq 0.05$ significance level on the *t*-test, the homeologous pair was considered biased for sub-genomes A or B.

To estimate the selective pressures of paralogs and homeologs, the homologous protein pairs were aligned using the MAFFT program [23] and converted to CDS using the PAL2NAL program [24]. The nonsynonymous substitutions per nonsynonymous site (K_a) and synonymous substitutions per synonymous site (K_s) were estimated. The sequences with low sequence divergence (K_a/K_s value < 0.001 or $K_a=0$) were excluded because low sequence divergence causes unknown results. Neutrality, positive selection, and purifying selection were indicated by $K_a/K_s = 1$, $K_a/K_s > 1$, and $K_a/K_s < 1$, respectively.

Results

Evaluation of proteomic sequencing

Contamination is a major setback during protein sampling, transportation, and sequencing. Therefore, this study examined all proteome sequences and confirmed purity (Additional file 1: Fig. S1a). Total protein was digested using trypsin (hydration C-terminal end of arginine and lysine), but several arginine and lysine residues were modified after translation. The modified amino acids were indigestible. Most proteins were digested

totally, but a few had missed cleavage (Additional file 1: Fig. S1b). The average peptide weight was 110 Da, and the peptide length was 7 to 27 amino acids using the MS analysis (Additional file 1: Fig. S1c). The distribution of mass deviation (effective peptide weight/expected peptide weight) was -4 to 4 (Additional file 1: Fig. S1d), consistent with the standard, -10 to 10 . Several proteins lacking during the proteome sequencing of leaf and root tissues could not be quantified (Additional file 1: Fig. S1e). The miss rate was 2.53% and 2.61% in the three biological repetitions between leaf and root tissues. The leaf and root tissues were not statistically different (Mann–Whitney test, $p > 0.05$), indicating that missed data did not significantly affect differential abundance. Consequently, the proteome sequencing data were used for further analyses.

Protein quantification and differentially abundant protein functional annotation

A total of 34,547 peptides (19,257 unique peptides) were identified in the leaf and root tissues, translating to 4803 proteins via the primary transcript CDS (Fig. 1A and Additional file 6: Table S1). A correlation analysis showed that the protein abundance patterns were similar across the three-leaf and root biological replicates (Additional file 1: Fig. S1f). There were 364 differentially abundant proteins between leaf and root tissues, including 300 up-regulated and 64 down-regulated proteins (Fig. 1B).

The GO enrichment results revealed that differentially abundant proteins were preferentially involved in the photosystem process (Fig. 1C). Meanwhile, the KEGG results showed that differentially abundant proteins were enriched for photosynthesis (Fig. 1D). These results indicate that differentially abundant proteins correlated with the environmental changes between leaf and root tissues. The leaf and root are distinct tissues, which leaf is above-ground, and the root is underground.

Paralog abundance divergence and homeolog abundance bias

The *A. hypogaea* cv. Tifrunner proteome had 336 paralogous and 197 homeologous pairs of abundant proteins (Fig. 2A and Additional file 7: Table S2). The selective pressure results showed that the average K_a/K_s value of paralogous and homeologous pairs was 0.24 and 0.31, respectively (Fig. 2B). The results indicate that major purifying selection acted on these proteins. Additionally, the K_a/K_s values of the three paralogous and six homeologous pairs were larger than one, indicating that these proteins underwent positive selection (Fig. 2B).

A total of 248 and 249 paralogous pairs were divergently abundant in the leaf and root tissues, respectively (Fig. 3a, B) There were 232 common paralogous pairs

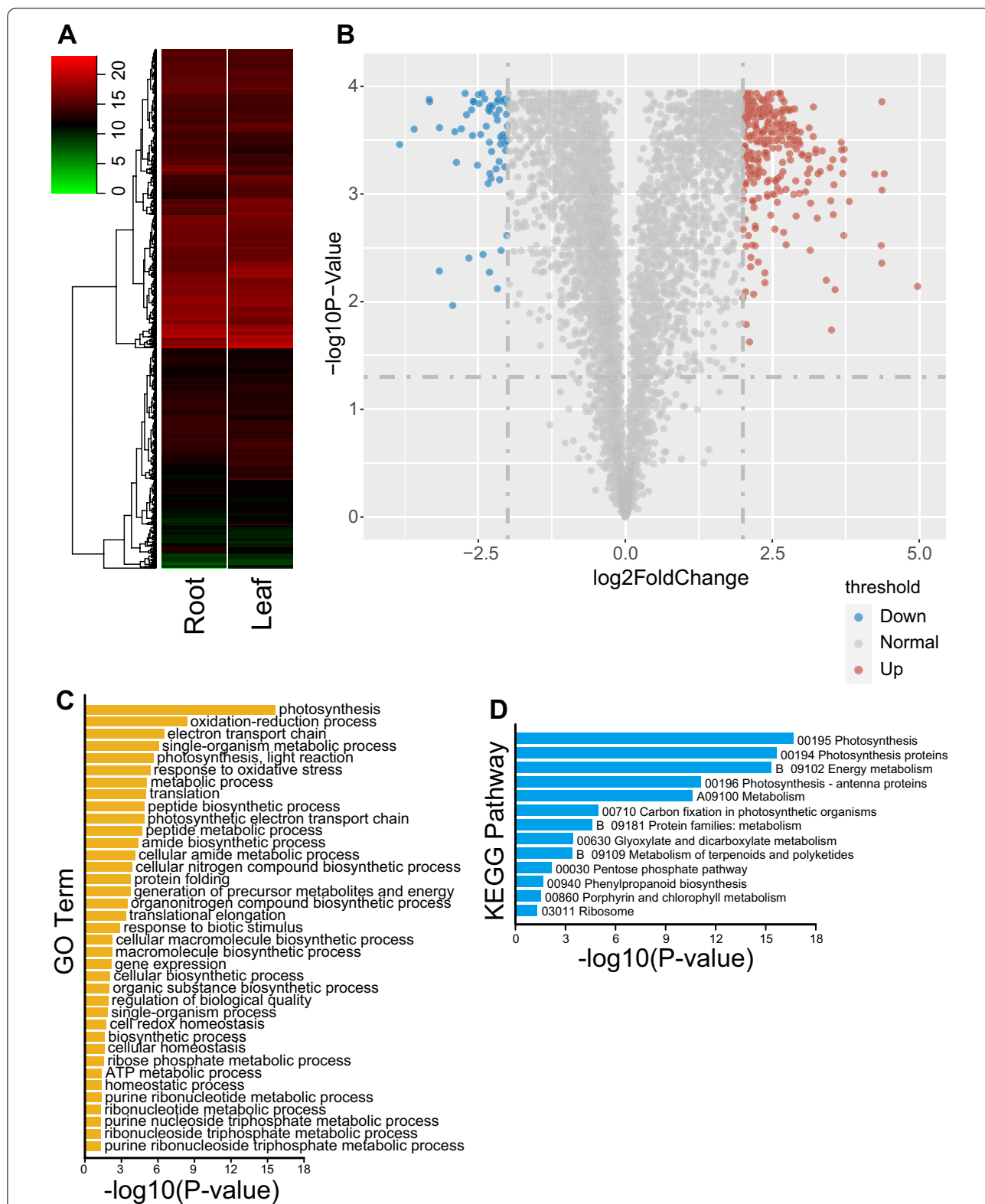
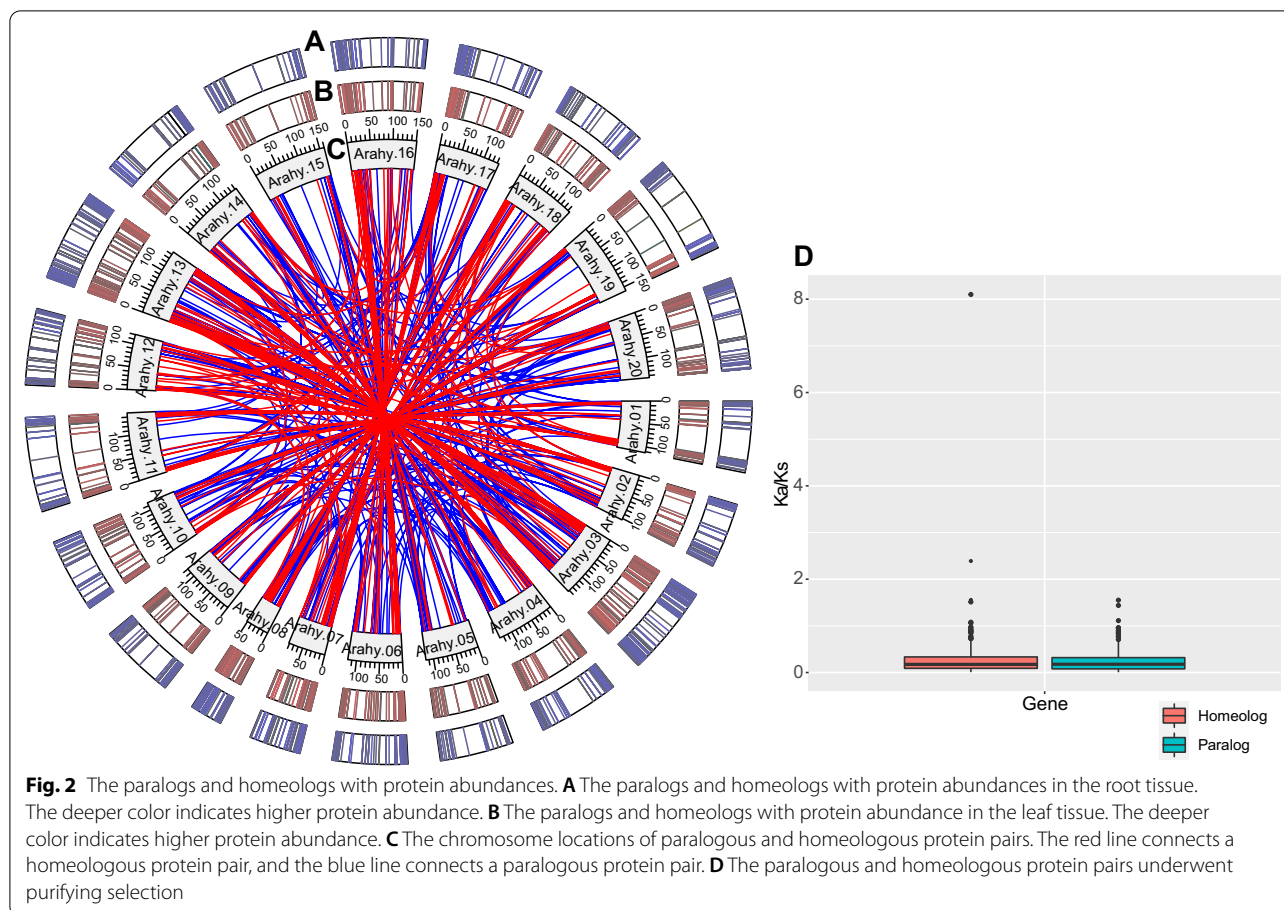


Fig. 1 The proteomic abundance of leaf and root tissues. **A** A total of 4803 abundant proteins in the leaf and root tissues. **B** A total of 364 differentially abundant proteins between leaf and root tissues, with 300 up-regulated and 64 down-regulated proteins, respectively. **C** Gene ontology (GO) enrichment analyses using differentially abundant proteins. The GO terms of the biological process were listed. **D** Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses using differentially abundant proteins

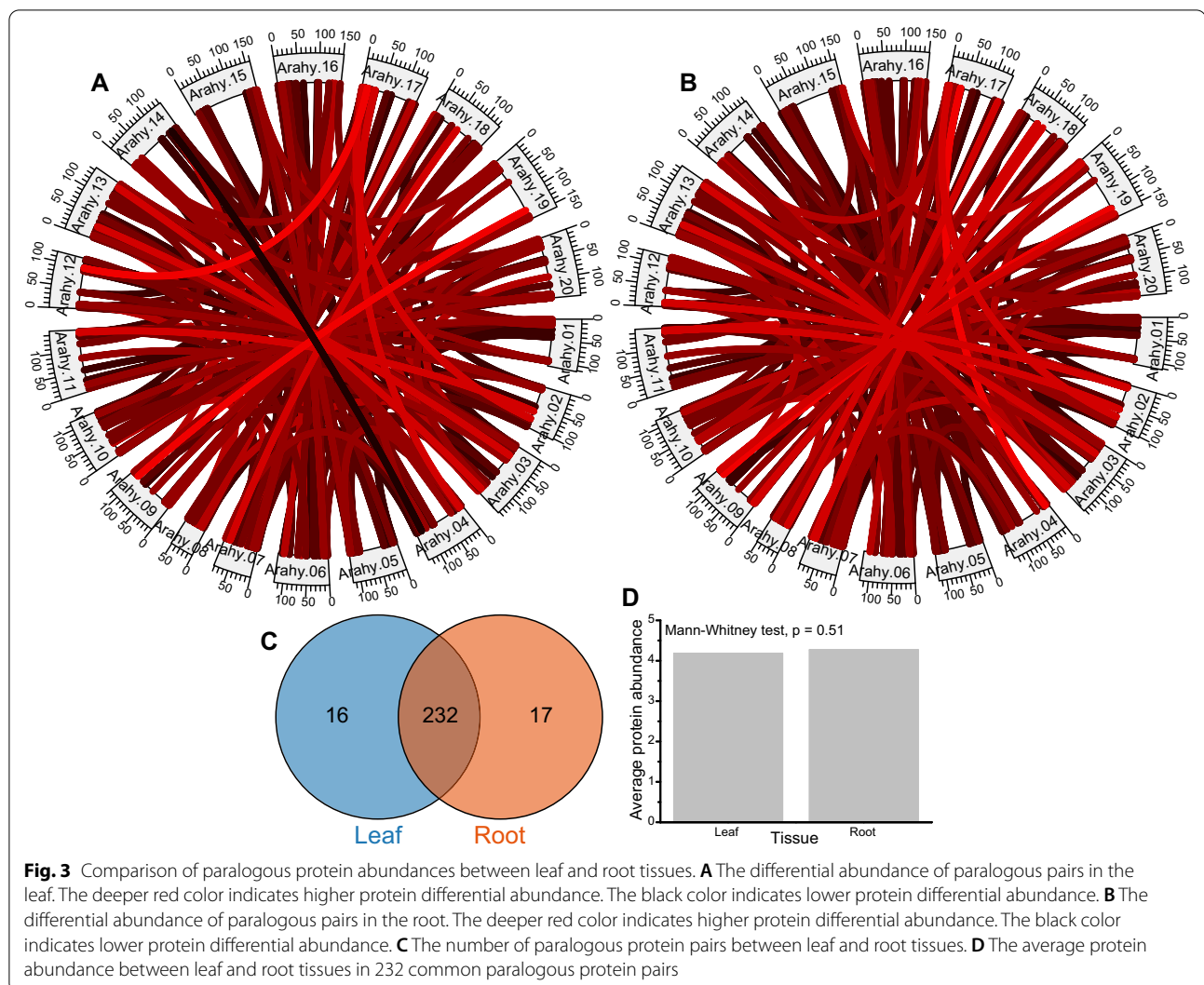


between leaf and root tissues (Fig. 3C). However, divergent abundance in the 232 paralogous pairs was not significantly different (Fig. 3D). The results indicate that the paralogous pairs from leaf and root tissues had similar divergent abundances. The same leaf and root tissues had 156 and 155 homeologous pairs with biased abundances (Fig. 4A, B), including 87 homeologous pairs biased to sub-genome A and 69 biased to sub-genome B leaf tissue. In the root, 87 homeologous pairs were biased to sub-genome A, while 68 were biased to sub-genome B. The leaf and root tissue were not significantly different, considering the number of homeologs with biased abundances (Fig. 4C).

GO enrichment analyses were executed in the leaf and root tissue to reveal homeologous protein function. In the leaf, homeologs biased to sub-genome A enriched biosynthetic and metabolic process, while homeologs biased to sub-genome B enriched iron ion homeostasis process (Additional file 2: Fig. S2 and Additional file 3: Fig. S3). In the root, homeologs with biased abundance enriched biosynthetic and metabolic processes (Additional file 4: Fig. S4 and Additional file 5: Fig. S5). However, sub-genomes A and B

had different homeologs in the root. Homeologs with biased abundance in sub-genome A enriched inorganic biosynthesis and metabolism process, while homeologs with biased abundance in sub-genome B enriched organic biosynthesis and metabolism process (Additional file 4: Fig. S4 and Additional file 5: Fig. S5). These results indicate that homeologs had a functional bias in the different sub-genomes and tissues.

The K_a/K_s values were not correlated with homeologous protein abundance in the leaf and root tissues but negatively correlated with paralogous protein abundance (Fig. 5A). The correlation between K_a/K_s value and protein abundance in paralogs is consistent with the expression rate of sequence evolution anticorrelation model (E-R anticorrelation) [25], where decreasing protein misfolding promotes fitness [26]. Moreover, the K_a/K_s values and differential protein abundances in paralogs and homeologs of leaf and root tissues were not correlated (Fig. 5B). These results indicate that selective pressure cannot affect differential protein abundance, consistent with a previous study where *A. duranensis* CDS architectures and differential gene expression were not correlated under drought and nematode stress [27].

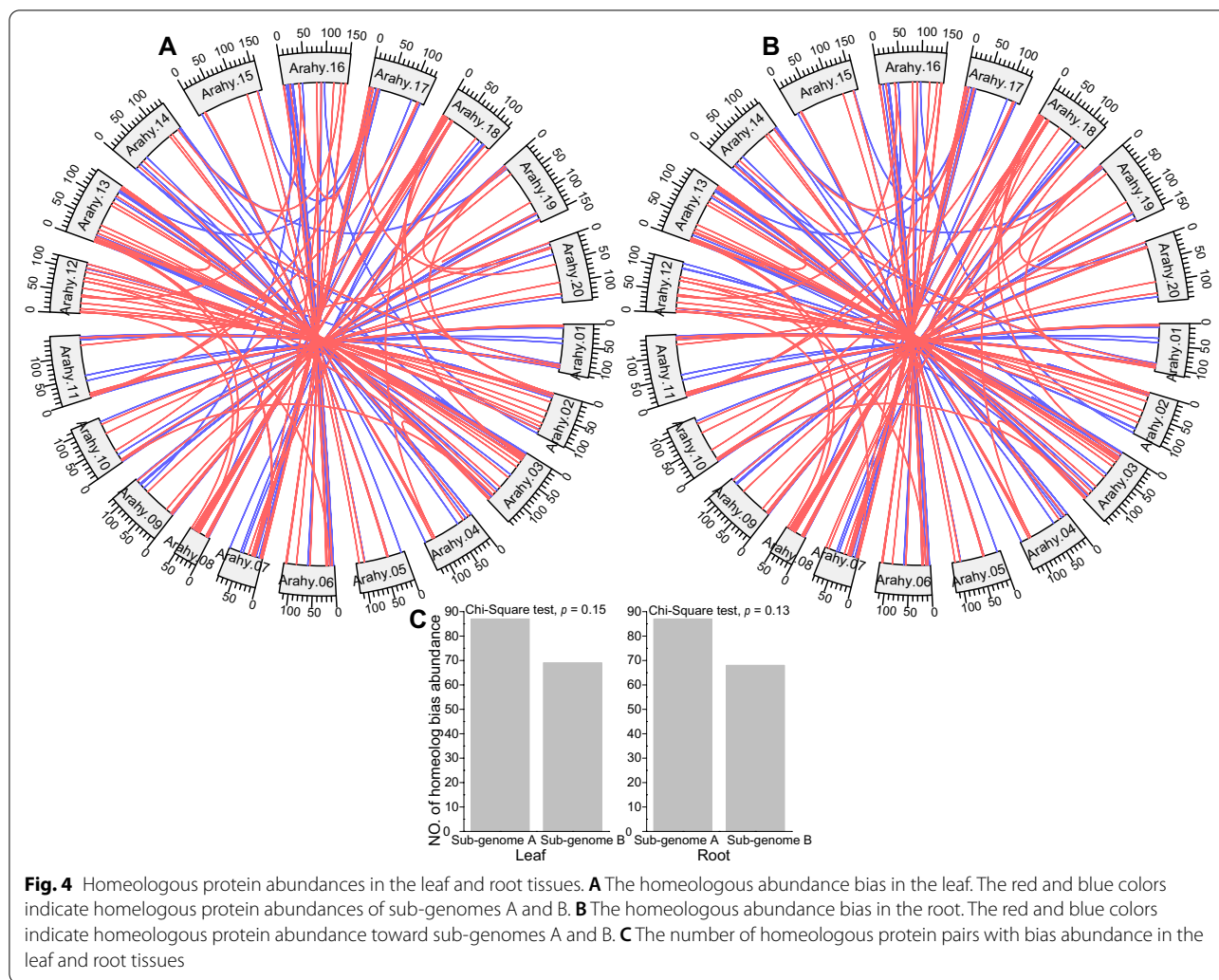


Paralogs and homeologs have differential abundance between leaf and root

This study identified differentially abundant proteins between leaf and root tissues. However, little is known about the abundance patterns of paralogs and homeologs in the differentially abundant proteins. One copy of the 48 paralogous protein pairs was differentially abundant in the leaf and root tissues (Additional file 8: Table S3). Twelve paralogous protein pairs were differentially abundant in the leaf and root tissues (Additional file 8: Table S3). In the leaf and root tissues, 24 homeologous protein pairs had one differentially abundant copy (Additional file 9: Table S4). Seven homeologous protein pairs were differentially abundant in the leaf and root (Additional file 9: Table S4). The two copies from paralogous and homeologous protein pairs had differential abundances, indicating that the two copies had similar gene functions. In contrast, only one copy from paralogous

and homeologous protein pairs had differential abundances, indicating that the two copies had potentially divergent functions. Furthermore, the difference value of differential abundance from protein pairs with two differentially abundant copies was less than that of differential abundance from protein pairs with one differentially abundant copy (Fig. 6A, B).

In paralogous proteins, 33 paralogous protein pairs with one differentially abundant copy were divergently abundant in the leaf and root. In addition, eight and nine paralogous protein pairs were divergently abundant in the root and leaf, respectively (Fig. 6C). These results indicate that paralogous protein pairs with differential abundance had similar abundance patterns in the leaf and root tissues. Ten homeologous protein pairs were biased to sub-genome A, and nine gene pairs were biased to sub-genome B in the leaf (Fig. 6D). In the root, five and ten homeologous protein pairs had biased abundance



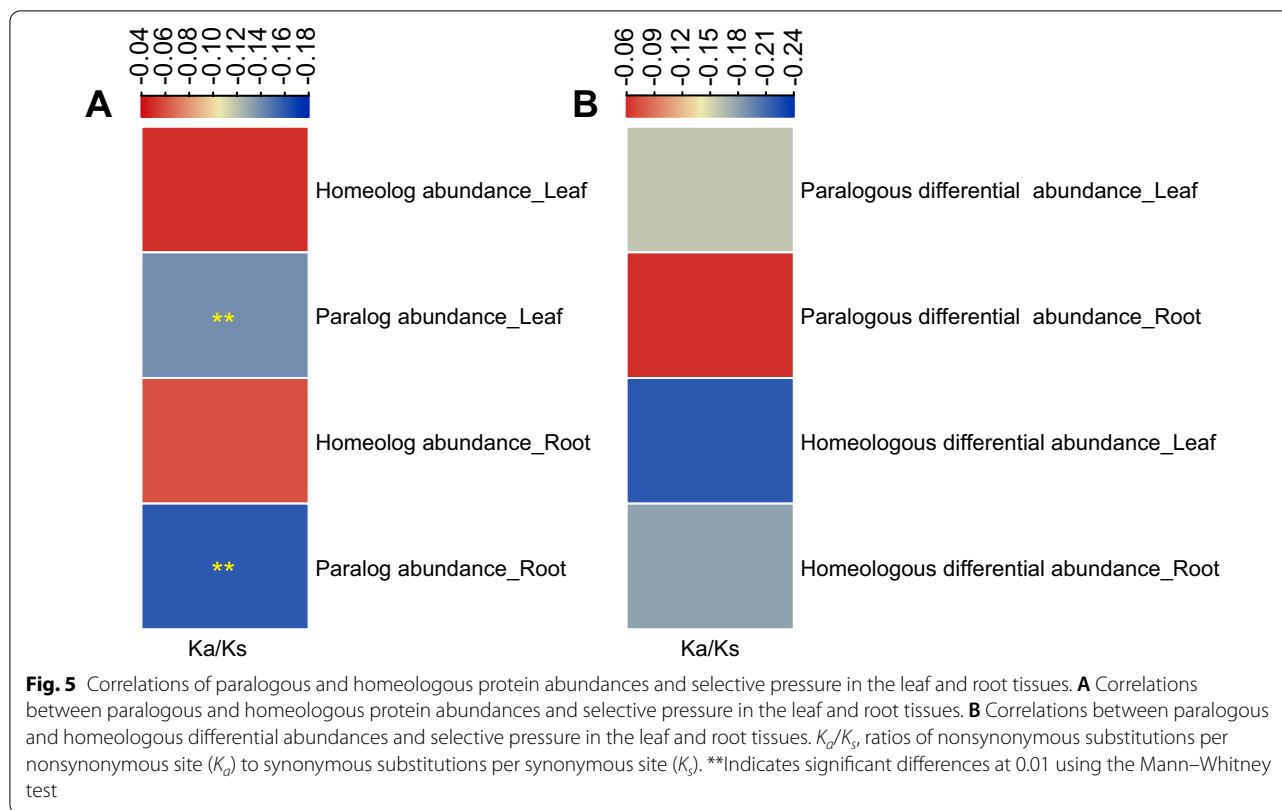
in the root sub-genomes B and A, respectively (Fig. 6E). These results indicate that homeologous protein pairs with differential abundance had similar abundance patterns in the leaf and root tissues.

Discussion

This work provides a foundation for revealing gene functions and improving molecular breeding of *A. hypogaea* cv. Tifrunner [2]. To date, homeolog expression bias was only elucidated using the RNA-seq analysis of cultivated peanut [2, 10, 17, 18, 28]. However, proteome level homeolog abundance bias in *A. hypogaea* cv. Tifrunner is scarcely known. This study obtained proteomic datasets from leaf and root tissues using the TMT technology and estimated differentially abundant proteins between leaf and root tissues. The paralogous protein abundance divergence and homeologous abundance bias were revealed. The study showed that 364 proteins were differentially abundant between leaf and root tissues. These

differentially divergent proteins enriched the photosystem process. Several homeologous proteins had biased abundances in sub-genomes A and B, but the number was not significantly different in the leaf and root tissues. Furthermore, the homeologous biased proteins between sub-genomes A and B had different biological processes. Additionally, paralogous and homeologous proteins were divergently abundant, with potentially functional divergence.

To our knowledge, there are no reported studies on proteomic homeologous abundance bias in *A. hypogaea* cv. Tifrunner. This study identified 156 homeologous protein pairs with biased abundance, yet previous study identified 7404 preferentially expressed homeologous gene pairs [2]. Although the number of homeologs between these two studies is different, both studies showed that the number of biased homeologous pairs between sub-genomes A and B was not significantly different in the leaf and root tissues. Previous studies



demonstrated non-significant differences in homeolog expression bias in somatic tissues but reproduction tissues in *A. hypogaea* cv. Tifrunner, *A. hypogaea* cv. Shitouqi, and *A. hypogaea* cv. Fuhuasheng were significantly different [2, 17, 18]. Homeolog expression bias is tissue-specific in cultivated peanuts compared to other allotetraploid plants. Although homeolog expression bias varies between tissues, other allotetraploid plants, such as cotton (*Gossypium hirsutum*), report common expression bias. In *G. hirsutum*, 20–40% of homeologs had biased expression towards sub-genome D, and a few gene pairs were biased toward sub-genome A in 35 different tissues [29]. Additionally, homeologous expression was biased in sub-genome A during the fiber development of cultivated cotton [29].

There are different functions of homeologs between two sub-genomes in *A. hypogaea* cv. Tifrunner [2]. For example, A sub-genome-biased homeologous gene pairs are preferentially involved in mannose metabolic processes, nitrate assimilation, and cell wall assembly, and B sub-genome-biased homeologous gene pairs are enriched in response to biotic stimulus, sucrose transport, and glucan metabolic processes in subterranean peg tip [2]. In maturing pericarp, the A sub-genome-biased homeologous gene pairs were biased toward phosphorylation signal transduction, carbohydrate

metabolism, and cell wall biogenesis. In contrast, B sub-genome-biased homeologous gene pairs were enriched for response to biotic stimulus and inorganic ion transport [2]. Additionally, comparisons of genome sequences between *A. monticola* and *A. hypogaea* cv. Tifrunner revealed that domestication mainly affects the sub-genome A of *A. hypogaea* cv. Tifrunner [28]. However, this study showed that homeologous proteins had different biological processes in the two sub-genomes from leaf and root tissues, consistent with Bertioli et al. [2]. These results indicate that homeologs are crucial for functional divergence between the two sub-genomes.

Homologs have functionally divergent gene expression [30–32] due to the WGD that preferentially occurred 55–75 million years ago around the K/Pg boundary [33]. This period involved numerous cataclysmic events causing biological extinctions [34]. The WGDs prevented extinctions of flowering plants [33]. This study showed that most homologous protein pairs had asymmetric abundance, and several homologous protein pairs had conserved abundance between leaf and root tissues. These results provide a clue about asymmetric abundance with potential functions in homologs retention at genome and proteome levels.

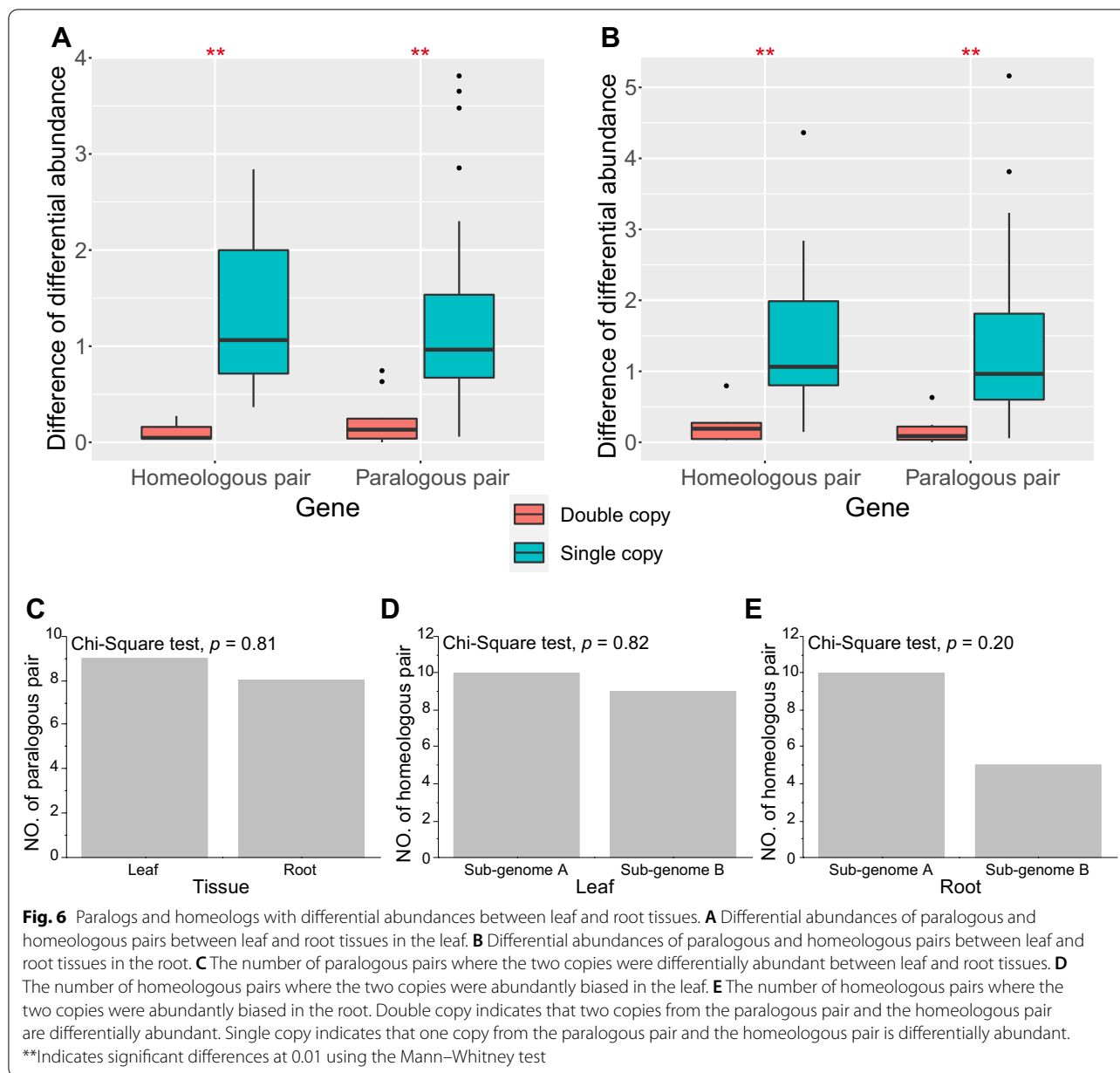


Fig. 6 Paralogs and homeologs with differential abundances between leaf and root tissues. **A** Differential abundances of paralogous and homeologous pairs between leaf and root tissues in the leaf. **B** Differential abundances of paralogous and homeologous pairs between leaf and root tissues in the root. **C** The number of paralogous pairs where the two copies were differentially abundant between leaf and root tissues. **D** The number of homeologous pairs where the two copies were abundantly biased in the leaf. **E** The number of homeologous pairs where the two copies were abundantly biased in the root. Double copy indicates that two copies from the paralogous pair and the homeologous pair are differentially abundant. Single copy indicates that one copy from the paralogous pair and the homeologous pair is differentially abundant. **Indicates significant differences at 0.01 using the Mann–Whitney test

Conclusions

In this study, proteome data were used to reveal homologous abundance patterns in *A. hypogaea* cv. Tifrunner. The results showed there is no difference in the number of biased abundant homeologs between the sub-genomes A and B. However, homeologous proteins enriched different biological processes between the two sub-genomes. Paralogous and homeologous biased abundant divergence in leaf and root tissues.

Abbreviations

FDR: False discovery rate; GO: Gene ontology; K_s : Nonsynonymous substitutions per nonsynonymous site; KEGG: Kyoto encyclopedia of genes and

genomes; K_a : Synonymous substitutions per synonymous site; MS: Mass spectrometry; TMT: Tandem mass tag.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-022-00840-y>.

Additional file 1: Fig. S1. Proteome sequencing results of leaf and root tissues by the tandem mass tag technology. A. The protein contaminant information during proteome sequencing. B. The number of missed cleavages. C. The peptide length. D. The distribution of mass deviation. E. The number of non-detected proteins. F. The protein abundance correlation across biological replications.

Additional file 2: Fig. S2. The gene ontology enrichment of biased homeologs in the leaf sub-genome A.

Additional file 3: Fig. S3. The gene ontology enrichment of biased homeologs in the leaf sub-genome B.

Additional file 4: Fig. S4. The gene ontology enrichment of biased homeologs in the root sub-genome A.

Additional file 5: Fig. S5. The gene ontology enrichment of biased homeologs in the root sub-genome B.

Additional file 6: Table S1. Proteins with abundances were detected using the tandem mass tag technology.

Additional file 7: Table S2. The paralogous and homeologous protein pairs with abundances in the leaf and root tissues.

Additional file 8: Table S3. The paralogous and homeologous protein pairs with differential abundances in the leaf. The differential abundance was estimated between leaf and root tissues.

Additional file 9: Table S4. The paralogous and homeologous protein pairs with differential abundances in the root. The differential abundance was estimated between leaf and root tissues.

Authors' contributions

HS conceived and designed this research, and wrote the manuscript. ZD and YZ analyzed data and wrote the partial manuscript. TZ, and MC provided help for plant cultivation. All authors read and approved the final manuscript.

Funding

This study was funded by the first Class Grassland Science Discipline Program of Shandong Province, China, Natural Science Foundation of Shandong Province, China (ZR2019QC017), Start-up Foundation for High Talents of Qingdao Agricultural University (No. 665/1120012), and Shandong Modern Agricultural Industrial and Technical System (SDAIT-23-01).

Availability of data and materials

All of the data and materials supporting our research findings are contained in the methods section of the manuscript. Details are provided in the attached additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Grassland Agri-Husbandry Research Center, College of Grassland Science, Qingdao Agricultural University, Qingdao, China. ²College of Animal Science and Technology, Qingdao Agricultural University, Qingdao, China.

Received: 9 October 2021 Accepted: 6 January 2022

Published online: 13 January 2022

References

- Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet.* 2016;48(4):438–46.
- Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet.* 2019;51:877–84.
- Bertioli DJ, Abernathy B, Seijo G, Clevenger J, Cannon SB. Evaluating two different models of peanut's origin. *Nat Genet.* 2020;52:557–9.
- Holbrook CC, Culbreath AK. Registration of "Tifrunner" peanut. *J Plant Registrations.* 2007;1:124.
- Zhao N, He M, Li L, Cui S, Hou M, Wang L, et al. Identification and expression analysis of WRKY gene family under drought stress in peanut (*Arachis hypogaea* L.). *PLoS ONE.* 2020;15(4):e0231396.
- Li L, Yang X, Cui S, Meng X, Mu G, Hou M, et al. Construction of high-density genetic map and mapping quantitative trait loci for growth habit-related traits of peanut (*Arachis hypogaea* L.). *Front Plant Sci.* 2019;10:745.
- Mondal S, Badigannavar AM. Identification of major consensus QTLs for seed size and monor QTLs for pod traits in cultivated groundnut (*Arachis hypogaea* L.). *3 Biotech.* 2019;9:347.
- Zhao Y, Ma J, Li M, Deng L, Li G, Xia H, et al. Whole-genome resequencing-based QTL-seq identified *AhTc1* gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol J.* 2020;18:96–105.
- Dash S, Cannon EKS, Kalberer SR, Farmer AD, Cannon SB. PeanutBase and other bioinformatic resources for peanut. In: Stalker HT, Wilson RF, editors. *Peanuts genetics, processing, and utilization.* Champaign: AOCS Press; 2016. p. 241–52.
- Clevenger J, Chu Y, Scheffler B, Ozias-Akins P. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front Plant Sci.* 2016;7:1446.
- Grover CE, Gallagher JP, Szadkowschi EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 2012;196(4):966–71.
- Yoo MJ, Liu X, Pires C, Soltis PS, Soltis DE. Nonadditive gene expression in polyploids. *Annu Rev Genet.* 2014;48:485–517.
- Novikova PY, Tschimatsu T, Simon S, Nizhynska V, Voronin V, Burns R, et al. Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol Biol Evol.* 2017;34(4):957–68.
- Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biol.* 2010;11(12):R125.
- Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, Chen JD, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fibre improvement. *Nat Biotechnol.* 2015;33(5):531–7.
- Yoo MJ, Szadkowschi E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity.* 2013;110(2):171–80.
- Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H, et al. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol Plant.* 2019;12(7):920–34.
- Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet.* 2019;51:865–76.
- Kruger NK. The Bradford method for protein quantitation. *Methods Mol Biol.* 1988;32:25–32.
- Ma J, Chen T, Wu S, Yang C, Bai M, Shu K, et al. iProX: an integrated proteome resource. *Nucleic Acids Res.* 2019;47(D1):D1211–7.
- Zhang Y, Yin D, Song H. Genome-wide identification and characterization of gene families in *Arachis*: methods and strategies. *Front Genet.* 2020;11:525.
- Song H, Sun J, Yang G. Comparative analysis of selection mode reveals different evolutionary rate and expression pattern in *Arachis duranensis* and *Arachis ipaensis* duplicated genes. *Plant Mol Biol.* 2018;98(4–5):349–61.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
- Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006;34(suppl 2):609–12.
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA.* 2005;102(40):14338–43.
- Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and

- trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci USA*. 2011;108(2):680–5.
27. Dong S, Zhang L, Pang W, Zhang Y, Wang C, Li Z, et al. Comprehensive analysis of coding sequence architecture features and gene expression in *Arachis duranensis*. *Physiol Mol Biol Plants*. 2021;27(2):213–22.
 28. Yin D, Ji C, Song Q, Zhang W, Zhang X, Zhao K, et al. Comparison of *Arachis monticola* with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. *Adv Sci*. 2019;7:1901672.
 29. Fang L, Guan X, Zhang T. Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.). *Crop J*. 2017;5(2):159–65.
 30. Ohno S. *Evolution by gene duplication*. New York: Springer; 1970.
 31. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. *New Phytol*. 2009;183:557–64.
 32. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. *Plant Physiol*. 2016;171(4):2294–316.
 33. Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017;18(7):411–24.
 34. Renne PR, Sprain CJ, Richards MA, Self S, Vanderkluyzen L, Pande K. State shift in Deccan volcanism at the Cretaceous-Paleogene boundary, possibly induced by impact. *Science*. 2015;350(6256):76–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

