ABS ACTA BIOLOGICA SIBIRICA
A journal on the biodiversity of Siberia and the adjacent lands

Acta Biologica Sibirica 10: 1689–1697 (2024)
doi: 10.5281/zenodo.14438667
http://journal.asu.ru

RESEARCH ARTICLE

Study of genetic diversity of globulin proteins in soybean (*Glycine max* (L.) Merr.) genotypes

Gunay H. Zeynalova[1], Saltanat A. Aghayeva[2]

1 *Ministry of Science and Education of the Republic of Azerbaijan, Institute of Bioresources, Nakhchivan, Azerbaijan*
2 *Western Caspian University, Baku, Azerbaijan*

Corresponding author: Saltanat A. Aghayeva (saltanat.genetic@wcu.edu.az)

Academic editor: R. Yakovlev | Received 17 October 2024 | Accepted 20 November 2024 | Published 15 December 2024

http://zoobank.org/2CA3B8F7-8688-4482-9C79-11592AE64732

Citation: Zeynalova GH, Agayeva SA (2024) Study of genetıc dıversıty of globulın proteıns ın soybean (*Glycıne max* (L.) Merr.) genotypes. Acta Biologica Sibirica 10: 1689–1697. https://doi.org/10.5281/zenodo.14438667

Abstract

In the study, electrophoretic analysis of globulin storage proteins was performed in 32 samples of local and introduced soybean plants. The goal was the identification, passportization, and study of the genetic diversity of soybean genotypes. In addition, the genetic diversity index (H-) was calculated for zones (ω-, ϒ- β- and α-) according to the frequency of occurrence of patterns in the electropherograms of globulin storage proteins in the seeds of soybean plant samples. 26 spectra and 56 patterns were detected in soybean samples, and polymorphism was observed in most of them. 9 spectra and 20 patterns were observed in the ω-zone, 5 spectra and 8 patterns in the ϒ-zone, 5 spectra and 7 patterns in the β-zone, and 7 spectra and 21 patterns in the α-zone. The genetic diversity index was calculated based on Nei's formula for each of the 4 zones - ω, ϒ, β, and α. According to the calculations, genetic diversity was observed to be higher in the α - zone (H=0.927), slightly lower in the ω- (H=0.796) and ϒ- (H=0.680) zones, and the lowest in the β- zone (H= 0.480). Based on cluster analysis, genotypes were divided into 6 groups and subgroups. Based on the obtained results, electrophoretic analysis of globulin storage proteins was performed for the first time in polyacrylamide gel (A-PAGE) and polymorphism was detected among soybean genotypes.

Keywords

Soybean, genotype, seed, globulin, storage protein, gene, pattern, electropherogram, cluster

Introduction

Studies conducted by the World Health Organization (WHO)/Food and Agriculture Organization (FAO) (1985) have shown that soybean protein can provide all the essential amino acids for the balanced nutrition for human. Soybean protein is considered to be the high biological value protein among the plant-based proteins (García et al. 1998). The quality of soybean protein is comparable to animal proteins from meat, milk, and eggs (Millward 2012; Kudełka et al. 2021).

The percent rates of nine amino acids contents are displayed in Fig. 1, including histidine, phenylalanine, methionine, serine, valine, isoleucine, leucine, tryptophan, and lysine. According to the Protein digestibility-corrected amino acid scores (PDCAAS), soybean protein ranks first among vegetable proteins and is comparable to that of milk and egg proteins (https://foodproteins.globalfoodforums.com/food-protein-articles/soy-protein-delivers-on-nutrition-quality-sustainability/).

The fact that the electrophoretic components of storage proteins in the seeds of cereals and legumes are passed down from generation to generation as genetically determined traits (Sadigov 2021), do not change and remain stable depending on the soil, climate and cultivation conditions, and their use as universal genetic markers in the study of genetic diversity and in the process of accelerating the selection process remains scientifically relevant.

Since storage proteins are the first products of gene expression, they play an important role as genetic markers in solving many scientific issues, such as the study of polymorphism and identification of plant samples, and the relationship between protein markers and bean quality traits (Kudryavtsev et al. 2014; Noveselskaya-Dragovich 2015).

Thus, the part of DNA (Khankishiyeva et al. 2016), associated with the location of any gene or genes in the genome are molecular markers. Markers are divided into three types. Morphological, biochemical and DNA-based markers. Molecular markers are classified into two types.

DNA markers have a wide range of applications, such as genetic identification of parents to improve plant genetic structure, evaluation and identification of variation at the genetic level, genetic confirmation, and development of high-resolution genetic linkage groups. A wide variety of molecular markers are available for genetic analysis of plants. Gene mapping and other genetic analysis approaches in molecular biology were developed by Mullis and Faloona (Mullis et al. 1987).

Legume seeds contain 20–35% albumins, 43–55% globulins, 0.73–2.70% prolamins, and 11.84–32.21% glutelins (Tchiagam et al. 2011).

Albumins are soluble in water; globulins in salt; prolamins in alcohol; and glutelins in alkali (Salem et all. 2019). Albumin and globulin together make up 63–90% of total seed proteins. The salt-soluble fraction (globulins) accounts for 45–50.3% of the total mass. Soluble proteins with a mean value of 47.7% are the main protein fraction. The studied soy protein is considered to be the water-soluble fraction. Al-

bumins make up 31.2–35.5% of total soluble proteins with their mean value. The third most abundant seed protein is glutelins, ranging from 15.1 to 20.5% .

The breeding of breakthrough varieties often depends on the utilization of rare and desired resources (Wehrmann et al. 1987). The availability of the soybean reference genome (Schmutz et al. 2010), wild soybean genome (Kim et al. 2010), pan-genome (Li et al. 2014), and graph-based pan-genome (Liu and Tian 2020) is conducive to the discovery of genes related to the protein content of soybean. The utilization of genes related to the protein content of wild soybean can improve the protein content of cultivated soybean. The discovery of QTLs and genes related to the protein content of cultivated soybean can facilitate the breeding of soybean varieties with high protein content by means of transformation and gene-editing technologies (Wu et al. 2021; Valliyodan et al. 2016). With the establishment and improvement of massive amounts of data, the comprehensive use of modern breeding technologies on the basis of bioinformatics and CRISPR/Cas9 has become an important method for plant improvement and germplasm creation (Gao et al. 2021). Li et al. (2019b) designed sgRNAs for nine different main storage protein genes and used CRISPR/Cas9 technology to edit the soybean seed storage protein gene family. The mutations in three storage protein genes were detected in soybean hairy roots, and the mutation frequency ranged between 3.8 and 43.7%. These studies laid a basis for the use of molecular design to boost the breeding of new soybean varieties with high protein content.

Materials and methods

In the research work, 32 varieties of soybean samples obtained from the Institute of Genetic Resources of the Ministry of Science and Education of the Republic of Azerbaijan and the Research Institute of Crop Husbandry were used. The research was performed in 2019-2021 in irrigated grey soils in the Botanical Garden of the Institute of Bioresources of the Ministry of Science and Education of the Republic of Azerbaijan.

Electrophoretic analysis of globulin protein was carried out in the "Biochemical genetics and technology" laboratory of the Institute of Genetic Resources of the Ministry of Science and Education of the Republic of Azerbaijan. Extraction and electrophoretic analysis of globulin storage proteins in beans of soybean varieties was carried out in polyacrylamide gel (A-PAGE), by a new method improved based on the modification of F.A. Poperelya's method. So, after the soybean sample was crushed, it was extracted twice with 500 µl of 70% alcohol, centrifuged at 3500 rpm each time, and then washed twice with 500 µl of 0.03% vinegar and acetone solution and after dissolving each time with a mechanical stirrer, rapidly centrifuged at 3500 rpm. After the fourth time, 500 µl of 9 molar acetic-urea solution was added to the extract and analyzed in a vertical electrophoresis apparatus in glycine-acetate buffer (pH-3.5).

Pattern numbering was performed by comparing them to each other in each zone and then numbering all patterns without considering repetitions. So, if any pattern is repeated in the samples, a new number is not assigned to that pattern, and all patterns are recognized by this rule. The occurrence frequency of each pattern of soybean samples was calculated by the following formula based on the Nei (Nei 1979), genetic diversity index for all zones:

$$H = 1 - \sum P_i^2 ,$$

where H – genetic diversity index; P_i – frequency of each pattern in zones.

Cluster analysis was used to determine the affinity between samples according to the UPGMA method through SPSS software (Rohlf 2000).

Results and discussion

Protein markers are one of the main markers used in the genetic identification of plants. The electropherograms of globulin proteins obtained during the vertical electrophoretic analysis of leguminous plants modified to the A-PAGE method for the first time in Azerbaijan and carried out by a new method were conditionally divided into 4 zones: these are ω-, ϒ-, β-, and α-globulins. High molecular weight proteins are localized in the ω-zone and low molecular weight proteins are localized in the α-zone (Figure 1).

A total of 26 spectra and 56 patterns were found among the examined soybean samples, and polymorphism was determined among them based on the frequency of occurrence of the patterns formed by the electrophoretic spectra. 9 spectra and 20 different patterns were studied in the ω-zone of electropherograms of globulin storage proteins. In this zone, ω-7 pattern was found with a frequency of 12.5%, ω-1 pattern with a frequency of 9.3% in 2 samples, ω-3 pattern with a frequency of 6.3% in 5 samples, and ω-2 pattern with a frequency of 3.1% in 12 samples. Among the spectra, ω_9s showed a high frequency of 100%, ω_8- medim frequency of 53.0% and ω4 -low frequency of 3.1%.

Five spectra and 8 patterns were observed in the ϒ- zone of the electropherograms of globulin storage proteins. ϒ-5 pattern was found in 2 samples with a frequency of 25%, ϒ-1 pattern 15.6 %, ϒ-3 pattern 12.5 %, ϒ-4 pattern 9.4 %, ϒ-2 pattern 6.3 %, ϒ-7 pattern was found in 2 samples with a frequency of 3.1 %. Among the spectra, Υ_4s showed a high frequency of 81.3 %, Υ_1s-medium frequency of 31.3 % and Υ_2s -low frequency of 15.6 %.

Five spectra and 7 patterns were determined in the β-zone of the electropherograms. β-2 pattern was found with a frequency of 37.5 %, β-3 pattern with a frequency of 25.0 %, β-5 pattern with a frequency of 12.5 % in two samples, β-4 pattern with a frequency of 6.3 % and β-1 pattern with a frequency of 3.1% in two samples. β_{3S} was observed with a high frequency of 93.8 %, β_{5S} with a medium frequency of 71.9 % and β_{1S} with a low frequency of 21.9 %.

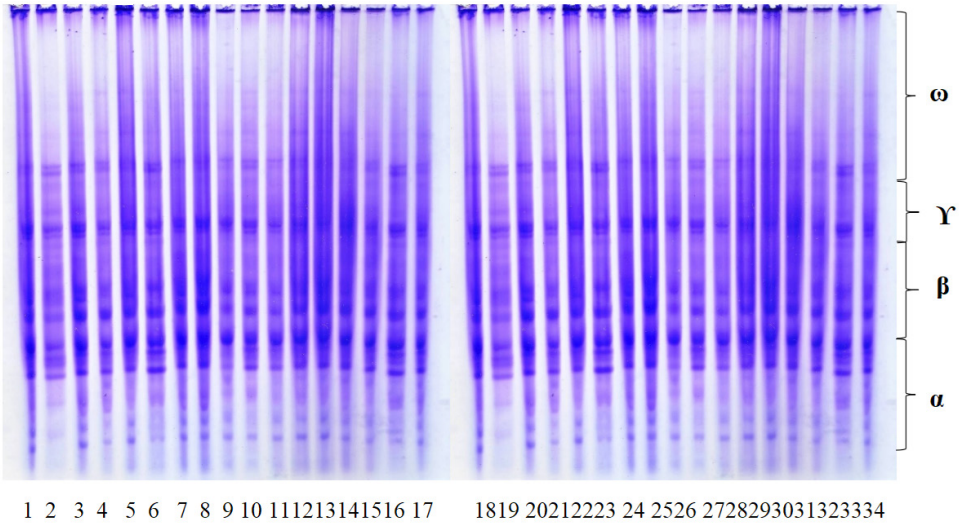

Figure 1. Electropherograms of globulin protein obtained from the beans of the soybean plant. 1. Opus st.; 2. Kofu st.; 3. Bravo; 4. Kanata; 5. Regale; 6. Bravo*; 7. Asuka; 8. Regale*; 9. Alexa; 10. Krasnodar-68; 11. Sinara; 12. Angelica; 13. Arisa; 14. Kofu; 15. Kyota; 16. Antonia; 17. Ukraniya; 18. Biyson; 19. Angelica t; 20. CU-7; 21. Kanata № 4; 22. Opus t; 23. CU-11; 24. Avstriya t; 25. CU-4; 26. CU-14; 27. Karisa; 28. Regaliya; 29. CU-1; 30. Kanata №7; 31. Kioto; 32. Antonia.

Seven spectra and 21 patterns were detected in the α-zone of the electropherograms. α- 3 pattern was found with a frequency of 18.8 %, α- 4 pattern with a frequency of 9.4 %, α- 1 pattern with a frequency of 6.3% in 4 samples, α- 2 pattern with a frequency of 3.1 % in 15 samples. α4s was found with a maximum frequency of 59.4 %, α_{2S} with a medium frequency of 37.5 % and α_{3S} with a low frequency of 12.5 %. The genetic diversity index was calculated for all 4 zones by applying Nei's formula among genotypes. As a result of the calculations, high genetic diversity was found in α-zone (H=0.927), relatively low in ω- (H=0.796) and ϒ-zones (H=0.680), and the lowest genetic diversity was observed in β-zone (H=0.480).

After the extraction of globulin storage proteins from the beans of soybean samples and their electrophoretic analysis, the dams (electrophoretic spectrum) were numbered according to the binary number system between genotypes. Dams located in the same place are numbered "1", and dams not in the corresponding area are numbered "0" according to binary nomenclature. The UPGMA computer program was used to determine the genetic affinity of the samples, and a dendrogram was constructed to study the genetic affinity of soybean genotypes through globulin protein markers. As can be seen from Figure 2, in the dendrogram, genotypes No. 21, 11, 22, 13, 10, 12, 20, 14 are classified in the 1st cluster; samples No. 30, 29, 31, 25, 24 are classified in the 2nd cluster; samples No. 28, 27, 26, 32,23 are classified in the 3rd cluster; samples No. 16, 15, 18, 6, 5, 4, 2, 1 are classified in the 4th cluster; sam-

ples No. 19, 3, 17 are classified in the 5th cluster and samples No. 9, 8, 7 are classified in the 6th cluster. According to the obtained results, by carrying out hybridization, it is appropriate to use the samples that are distant in terms of genetic distance in the selection of parental forms in marker-based selection and accelerate the selection process.

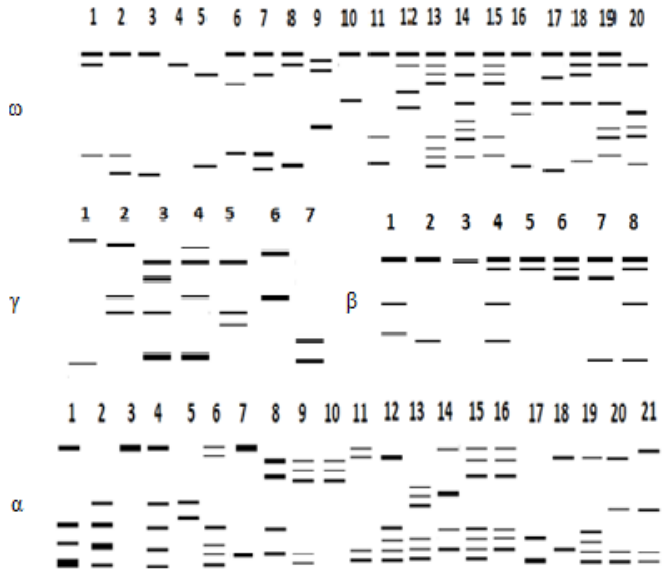

Figure 2. Idiogram of different patterns in ω-, ϒ-, β- and α-zones observed in soybean samples.

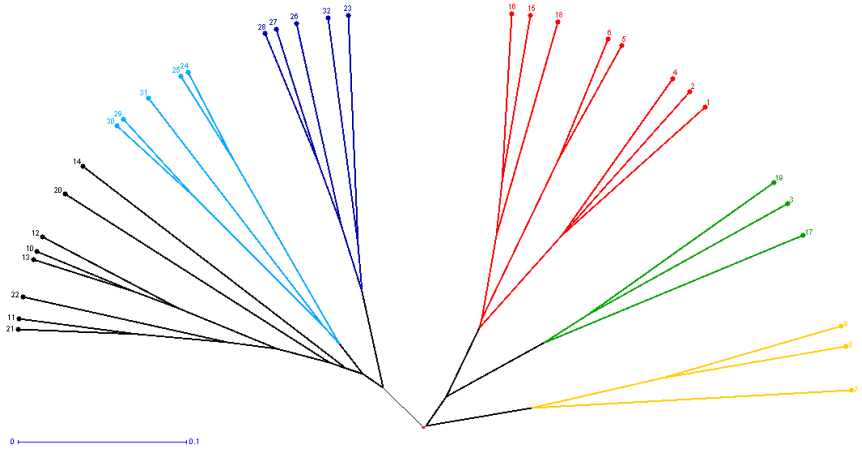

Figure 3. Dendrogram showing genetic distance between different soybean samples based on polymorphism of globulin protein electropherograms.

Conclusion

Based on the Nei formula, the genetic diversity index was calculated for each of the 4 zones ω-, ϒ- β- and α-. According to the calculations, the genetic diversity was observed to be higher in α- zone (H=0.927), relatively low in ω- (H=0.796) and ϒ- (H=0.680) zones, and the lowest in β- zone (H= 0.480). Based on cluster analysis, genotypes were divided into 6 groups and subgroups. Based on the obtained results, electrophoretic analysis of globulin storage proteins was performed for the first time in polyacrylamide gel (A-PAGE) and polymorphism was detected among soybean genotypes.

According to the results of our research, the coefficient of genetic diversity (H-), the highest α-zone (H=0.927), and the lowest β-zone (H= 0.480). According to the results of Orkhan B (2024) the highest was β-zone (0.947), and the lowest was ω-zone (0.731). There is no significant difference between the results we obtained. The cultivation method, the variety and the climate have made a relative difference. This proves the correctness of our analysis.

References

Alghamdi SS, Khan MA, Migdadi HM, El-Harty EH, Afzal M, Farooq M (2019) Biochemical and molecular characterization of cowpea landraces using seed storage proteins and SRAP marker patterns. Saudi Journal of Biological Sciences 26(1): 74–82. https://doi.org/10.1016/j.sjbs.2018.09.004

Gao Y, Ma S, Wang M, Feng XY (2017) Characterization of free, conjugated, and bound phenolic acids in seven commonly consumed vegetables. Molecules 22: 1878. https://doi.org/10.3390/molecules22111878

García MC, Marina ML, Laborda F, Torre M (1998) Chemical characterization of commercial soybean products. Food Chemistry 62: 325–331. https://doi.org/10.1016/S0308-8146(97)00231-8

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y (2010). A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): Implications from microsatellites and nucleotide sequences. Annals of Botany 106(3): 505–514. https://doi.org/10.1093/aob/mcq125

Khankishiyeva EM, Shikhlinski HM (2016) Use of molecular markers in resistance breeding. Research Institute of Crop Husbandry. Collection of scientific works XXVII, Baku, 160–167.

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee Y-S, Park D, Ma J, Kim W-Y, Kim B-Ch, Park S, Lee K-A, Kim DH, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee Y-H, Kim K-H, Moon J-K, Schmutz J, Jackson SA, Bhak J, Lee S-H (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proceedings of the National

Academy of Sciences of the United States of America 107(51): 22032–22037. https://doi.org/10.1073/pnas.1009526107

Kudryavtsev AM, Dedova LV, Melnik VA, Shishkina AA, Upelnik VP, Novoselskaya-Dragovich AYu (2014) Genetic diversity of modern Russian durum wheat cultivars at the gliadin-coding loci. Genetics 50(5): 554–559.

Liu Ch, Jin H, Yu Y, Sun J, Zheng H, Zhang Y, Xu J, Zhu X (2020) The improvement of nanoemulsion stability and antioxidation via protein-chlorogenic acid-dextran conjugates as emulsifiers. Nanomaterials 10(6): 1094. https://doi.org/10.3390/nano10061094

Liu Y, Yang J, Lei L, Wang L, Wang X, Ma KY, Yang X, Chen ZY (2019) Isoflavones enhance the plasma cholesterol-lowering activity of 7S protein in hypercholesterolemic hamsters. Food & Function 10: 7378–7386. https://doi.org/10.1039/C9FO01432B

Millward DJ (2012) Amino acid scoring patterns for protein quality assessment. British Journal of Nutrition 108: 31–34. https://doi.org/10.1017/S0007114512002462

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335–350. https://doi.org/10.1016/0076-6879(87)55023-6

Nei M (1979) Analysis of gene diversity in a subdivided population. Proceedings of the National Academy of Sciences of the United States of America 70: 3321–3323.

Noveselskaya-Dragovich AY (2015) Genetics and genomics of wheat: storage proteins, ecological plasticity, and immunity. Russian Journal of Genetics 51(5): 476–490.

Orkhan B (2024) Delineating Genetic Diversity in Soybean (*Glycine Max* L.) Genotypes: insights from A-page analysis of globulin reserve proteins. Advances in Biology & Earth Sciences 9(2): 259–266. https://doi.org/10.62476/abes9259

Poperelya FA (1989) Gliadin polymorphism and its relationship with grain quality, productivity, and adaptive properties of winter soft wheat varieties. Agropromizdat, Moscow, 138–149. [In Russian]

Rohlf F J (2000) NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System:Version 2.1. Exeter Publishing Setauket, New York.

Sadigov HB (2021) Protein polymorphism of tetraploid wheat genotypes and the relationship of qualitative characteristics with genetic markers. Abstract of the dissertation for the degree of Doctor of Biological Sciences. Baku, 57 pp.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu Sh, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Zh, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-Ch, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183. https://doi.org/10.1038/nature08670

Sugiyama A, Ueda Y, Takase H, Yazaki K (2015) Do soybeans select specific species of Bradyrhizobium during growth? Communications and Integrative Biology 8(4): e992734. https://doi.org/10.4161/19420889.2014.992734

Tchiagam JBN, Bell JM, Nassourou AM, Njintang NY, Youmbi E (2011) Genetic analysis of seed proteins contents in cowpea (*Vigna unguiculata* L. Walp.). African Journal of Biotechnology 10 (16): 3077–3086. https://doi.org/10.5897/AJB10.2469

United States Department of Agriculture FAS (2011) Oilseeds: World market and trade archives. Full report (10–11). http://www.fas.usda.gov/oilseeds/Current/default.asp

United States Department of Agriculture Foreign Agricultural Service (2011) Oilseeds: World market and trade archives. http://www.fas.usda.gov/oilseeds_arc.asp

United States Department of Agriculture (2013) Soybeans: Supply, disappearance and price, US, 1980/81–2012/13. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1290

Wehrmann V, Fehr W, Cianzio S, Cavins J (1987) Transfer of high seed protein to high-yielding soybean cultivars. Crop Science 27(5): 927–937. https://doi.org/10.2135/cropsci1987.0011183X002700050020x

World Health Organization/Food and Agriculture Organization/United Nations University (1985) Energy and protein requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. In: WHO technical report series 724.