Structural homology of metal-dependent proteins of woody plants used in agroforestry of arid areas

Pavel A. Krylov[*]*,* and *Anna* V. Tretyakova

Laboratory of Genomic and Postgenomic Technologies, Federal Scientific Center of Agroecology, Complex Melioration and Protective Afforestation of the Russian Academy of Sciences, Volgograd, Russian Federation

Abstract. Fighting against desertification is one of the priorities in the world. In areas subject to desertification, there is a deficiency of heavy metal ions, such as Fe, Mn, Zn, Co, Cu and Ni, which are involved in the implementation of metabolic processes in woody plants. In this study, an assessment metal-dependent proteins and an analysis of the structural homology between them in Quercus robur, Robinia pseudoacacia, Gleditsia triacanthos, Ulmus pumila and Fraxinus excelsior used in agroforestry of arid territories was carried out. Bioinformatic analysis included multiple alignment of amino acid sequences using the Ugene program by the ClustalW algorithm (BLOSUM62 matrix). According to the results of studying the plants metalloproteome, it was found that Quercus robur has 18 proteins, Robinia pseudoacacia – 24, Gleditsia triacanthos – 3, Ulmus pumila – 19, Fraxinus excelsior – 14 metal-dependent proteins. The main part of metal-dependent proteins is involved in photosynthesis and respiration, in particular Fe, Zn and Co-dependent proteins. Multiple alignment showed a high degree of protein homology between species of woody plants, where similarity varied from 65% to 100%. The results can be used to create new agricultural technologies for productivity management and the formation of adaptation to adverse environmental factors.

1 Introduction

Fighting against desertification is one of the main tasks of agroforestry in the world, the solution of which lies in increasing the area of forests. Currently, all protective afforestation works have a comprehensive approach, which includes agrotechnical measures, construction of species composition of protective forest plantations, as well as taking into account the properties of territories and zonal geomorphological processes [1]. One of the key factors affecting the productivity of protective forest belts is the soil and its composition [2].

Metal ions contained in the soil are able to have a sufficiently wide range of effects on the implementation of molecular genetic mechanisms in woody plants. Metal ions play an important role in plant life. They are involved in the basic processes of plant life:

---

* Corresponding author: krylov-p@vfanc.ru

photosynthesis, respiration, signal transmission, growth and development. However, if the content of metal ions in the soil exceeds the allowable norm, this can negatively affect plant crops [3-4].

In desert and semi-desert zones, there is a metal ions deficit in the soil necessary for the implementation of the physiological functions of plants [5]. It should be noted that the content of heavy metals in the soil of deserts and semi-deserts is poorly studied, as well as the impact on the molecular genetics adaptation mechanisms of woody plants growing in these territories and on the transitional boundaries between zones.

Insufficient knowledge of metal-dependent proteins, including interactions between metal ions and binding sites, does not allow for a complete understanding of plant physiology at the molecular level for individual species. This knowledge is important for determining what constitutes a healthy organism and provides a starting point for understanding not only the consequences of metal deficiency, but also their toxicity.

Metal ions perform three fundamental tasks in proteins: structural (domains), regulatory (activation or inhibition, signaling molecule), and enzymatic (cofactors). The main biologically significant metals belong to two groups of metals: non-transition metals such as Zn and Mg, and transition metals such as Mn, Fe, Co, and Cu [6].

Among the woody plants used in agroforestry, the proteome of Poplar and Walnut has been studied in the greatest detail, with over 1500 annotated metal-dependent proteins involved in various physiological and biochemical processes. The degree of study of the Poplar and Walnut proteome far exceeds that of other woody plant species, which may be due to their being sequenced among the first (compared, for example, to petiolate Oak) and continue to be actively studied [7-10]. The proteome of other woody plants has been practically unexplored and the number of annotated proteins is very small. Woody-shrub species used in fighting against desertification belong to different classes and families [11].

The structural homology of metal-dependent proteins involved in photosynthesis, respiration, and other processes has not been sufficiently studied. The presence or absence of structural homology may lay the basis for new agromeliorative technologies that enable the control of physiological and biochemical processes of adaptation to drought in woody plants.

Therefore, the aim of this study is to assess the knowledge of the proteome of woody plants used in agroforestry to detect annotated metal-dependent proteins and search for structural homology between them.

2 Material and methods

To assess the knowledge of the woody plants proteome for annotated metal-dependent proteins, the main species used in agroforestry in southern Russia were selected: Quercus robur, Robinia pseudoacacia, Gleditsia triacanthos, Ulmus pumila, and Fraxinus excelsior. The next step was to conduct bioinformatics analysis using virtual screening of the UniProtKB database (https://www.uniprot.org/). The main selected metals were iron, manganese, zinc, nickel, cobalt, and copper. Screening was conducted based on the presence of the target metals in the protein's composition (catalytic activity, cofactor, domain) and metal binding. Metal-dependent proteins were categorized based on their involvement in main physiological processes: photosynthesis (anabolic process), respiration (catabolic process), and others (regulation of molecular-genetic mechanisms).

Data on the structure of amino acid sequences of the studied metal-dependent proteins of woody plants were obtained from the UniProtKB free access database (https://www.uniprot.org/). Full amino acid sequences were taken to search for homology between proteins present in all species of the studied woody plants. Bioinformatic analysis included multiple alignment of the amino acid sequences of the studied proteins using the Ugene program (UNIPRO, Russia) by the ClustalW algorithm (BLOSUM62 matrix).

Sequences were chosen for alignment according to the following criteria: presence of a complete sequence (fragmented sequences were excluded), protein annotation and the newest sequence assemblies. The alignment results were evaluated by identity (%) and score.

3 Results and discussion

3.1 Analysis of the woody plants proteome for the presence of metal-dependent proteins

As a result of the proteome analysis of woody plants growing in the South of Russia and used in agroforestry, the presence of metal-dependent proteins was identified and is presented in Table 1.

Table 1. Evaluation of metal-dependent proteins in the proteome of woody plants, according to the UniProtKB database.

Forest species	Metal-dependent proteins					
	Total	Fe	Mn	Zn	Co	Cu
Quercus robur	18	8	3	5	2	-
Robinia pseudoacacia	26	9	8	6	2	1
Gleditsia triacanthos	3	2	-	-	-	1
Ulmus pumila	19	7	1	6	4	1
Fraxinus excelsior	14	7	-	6	1	-

The majority of the studied proteins have a complex structure, consisting of subunits that are individually encoded by different genes. Thus, when counting the number of metal-dependent proteins, we took into account the oligomeric proteins themselves. Analysis of the metalproteomes showed the absence of annotated Ni-dependent proteins in the studied woody plants.

Quercus robur was found to have 18 metal-dependent proteins, with no annotated Cu-dependent proteins found in databases. Analysis of the metalproteome of *Robinia pseudoacacia* and *Ulmus pumila* showed the presence of proteins related to all metal ions. *Gleditsia triacanthos* was found to have only two Fe-dependent and one Cu-dependent proteins. *Fraxinus excelsior* was described to have a total of 14 proteins related to Fe, Zn, and Co. It is worth noting that the number of annotated Fe-dependent proteins is 2-4 times higher than other metal-dependent proteins. Analysis of the UniProtKB database showed that Cu-dependent proteins are the least described and were only found in *Robinia pseudoacacia*, *Gleditsia triacanthos*, and *Ulmus pumila*.

In general, we can conclude that the proportion of annotated metal-dependent proteins in the studied woody plants is very small and requires further study. The low level of study of some woody plants metalproteomes may be due to the absence of sequenced chromosomal, plastid/mitochondrial genomes, as well as the fact that their practical application in agroforestry is at an early stage.

3.2 The participation of metal-dependent proteins in the physiological processes of woody plants

The influence of metal-dependent proteins on the functioning of various biochemical and physiological processes in plants is very diverse. The analysis of metal-dependent proteins made it possible to identify which of them are involved in certain processes in woody plants (Figure 1).

Analysis of the participation of Fe-dependent proteins in physiological processes in woody plants showed that they are involved in photosynthesis and respiration. Six Fe-dependent proteins were identified with one or more amino acid sequence variants in the studied trees, except for *Gleditsia triacanthos*. All of these proteins participate in the functioning of photosystems I and II. The cytochrome b6-f complex transfers electrons between these systems.
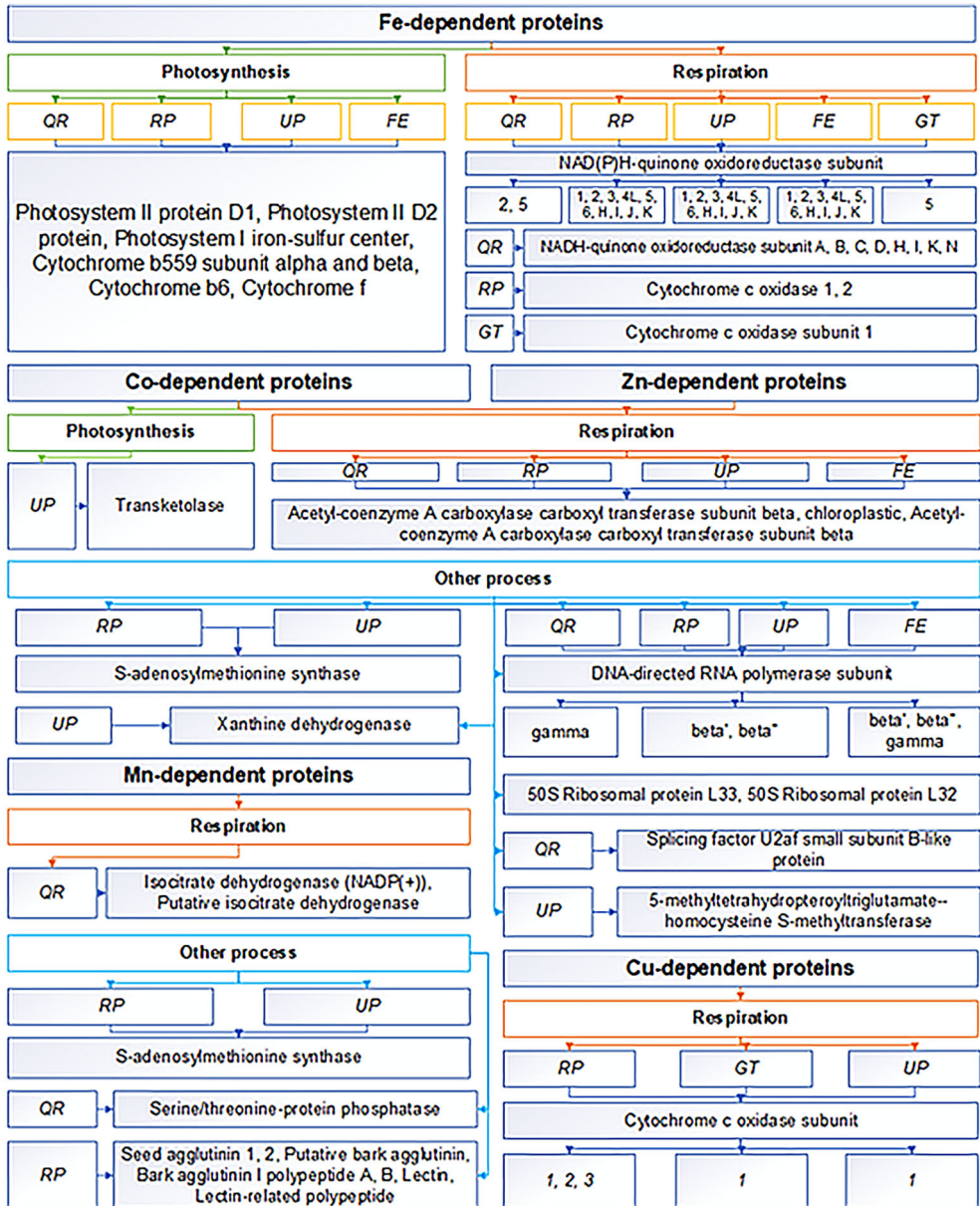
Fig. 1. Metal-dependent proteins involved in key processes in woody plants. QS – Quercus robur, RP – Robinia pseudoacacia, GT – Gleditsia triacanthos, UP – Ulmus pumila, FE – Fraxinus excelsior.

Fe-dependent subunits of NAD(P)H-quinone oxidoreductase were annotated in all studied woody plants. This enzyme is localized on the inner membrane of mitochondria and

participates in respiration. NADH-quinone oxidoreductase subunits were only described in *Quercus robur*. Their function is the translocation of protons across the inner membrane of mitochondria. Physicochemical properties of cytochrome c oxidase that catalyzes electron transfer from cytochrome c to oxygen with the formation of water were described in the database for two species: *Robinia pseudoacacia* and *Gleditsia triacanthos*. Mn-dependent proteins in plant proteomes were represented by 12 in *Quercus robur*, *Robinia pseudoacacia*, and *Ulmus pumila*.

Mn-dependent proteins did involve in photosynthesis were found in the studied plants, according to the UniProt database. Isocitrate dehydrogenase (NADP(+)) and Putative isocitrate dehydrogenase, involved in isocitrate metabolism in *Quercus robur*, which is linked to the Krebs cycle and amino acid synthesis, were identified. The Serine/threonine-protein phosphatase enzyme was present in *Quercus robur*, which provides for the phosphorylation of serine/threonine involved in cellular processes. In *Robinia pseudoacacia*, a number of proteins (lectins and agglutinins) were found to bind to manganese ions. S-adenosylmethionine synthase in *Robinia pseudoacacia* and *Ulmus pumila* participates in the expression of genes and their patterns.

Zn-dependent subunits of Acetyl-coenzyme A carboxylase carboxyl transferase participating in respiration were identified in *Quercus robur*, *Robinia pseudoacacia*, *Ulmus pumila*, and *Fraxinus excelsior*. In nature, two different types of enzymes with different structures are found. The heteromeric structure, consisting of four subunits, is more common in prokaryotes, while the homomeric structure, consisting of one large polypeptide, is found in eukaryotes. Both structures are found in plants, where the heteromeric form is located in plastids and the homomeric form is in chloroplasts. The enzyme is responsible for the synthesis of fatty acids [12].

Zn-dependent subunits of DNA-directed RNA polymerase have been identified in these tree species, providing transcription from matrix DNA. 50S ribosomal protein L33, chloroplastic, and 50S ribosomal protein L32 are responsible for the synthesis of proteins encoded by the chloroplast genome, including transcription and translation mechanisms and components of the photosynthetic apparatus. The enzyme Allantoinase participates in purine base metabolism. A Zn-dependent subunit of the Splicing factor U2af small subunit B-like protein was detected in *Quercus robur*. The enzyme 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase was annotated in *Ulmus pumila*, involved in L-methionine metabolism, where Zn is a cofactor.

Co-dependent proteins are involved in photosynthesis and respiration, as well as gene expression and patterns. Among the annotated Co-dependent proteins, the enzyme Transketolase was described in *Ulmus pumila*, involved in the photosynthesis process by participating in both the pentose phosphate pathway and the Calvin cycle [13]. For four tree species, excluding *Gleditsia triacanthos*, a subunit of Acetyl-coenzyme A carboxylase carboxyl transferase beta was annotated by analogy with Zn-dependent proteins. Another Co-dependent protein, Xanthine dehydrogenase, was annotated in *Ulmus pumila*, which provides active oxygen species for stress signaling and protection against pathogens [14]. *Robinia pseudoacacia* and *Ulmus pumila* share an annotated protein, S-adenosylmethionine synthase, which is responsible for the production of S-adenosylmethionine, a cofactor required for various methylation reactions, polyamines formation, and ethylene plant hormone [15].

Subunits of Cytochrome c oxidase were annotated in *Robinia pseudoacacia*, *Gleditsia triacanthos*, and *Ulmus pumila*, responsible for receiving an electron from each of the four molecules of cytochrome c and transferring them to one molecule of oxygen and four protons, forming two molecules of water [16].

Among the examined proteins that are part of the metalloproteome of woody plants, the most significant are Fe-, Zn-, and Co-dependent proteins, as they are more involved in photosynthesis, respiration, and the implementation of genetic information: transcription,

translation. The next step was to identify structural homology of proteins and their subunits among the studied woody plants.

3.3 The participation of metal-dependent proteins in the physiological processes of woody plants

The multiple alignment of metal-binding proteins allowed to reveal the different degree of homology of these proteins among woody plants. (Figures 2, 3, 4).

We determined all Fe-dependent proteins have high similarity values. The similarity of proteins involved in photosynthesis ranged from 90% to 100%. For example, amino acid sequences of Photosystem I iron-sulfur center protein have 100% similarity in different woody plants (Figure 2).
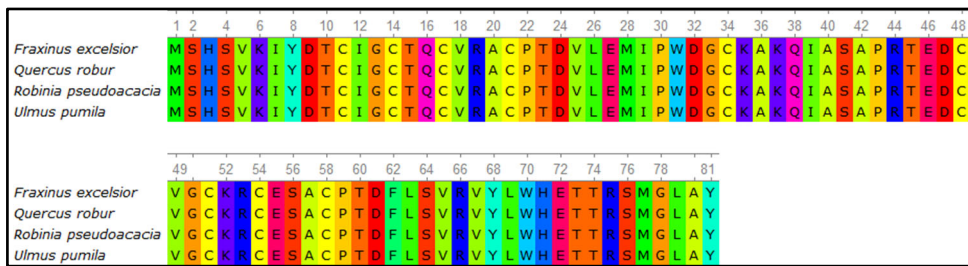
Fig. 2. Multiple alignment of Photosystem I iron-sulfur center. The sequences are 100% identical.

Fe-dependent proteins involved in respiration also have a high level of homology. Less than 80% similarity was found only in the proteins NAD(P)H-quinone oxidoreductase subunit 5 and NAD(P)H-quinone oxidoreductase subunit 6. Nevertheless, there is a minimum percentage - 74%. The obtained results of multiple alignment indicate the conservation of these proteins.
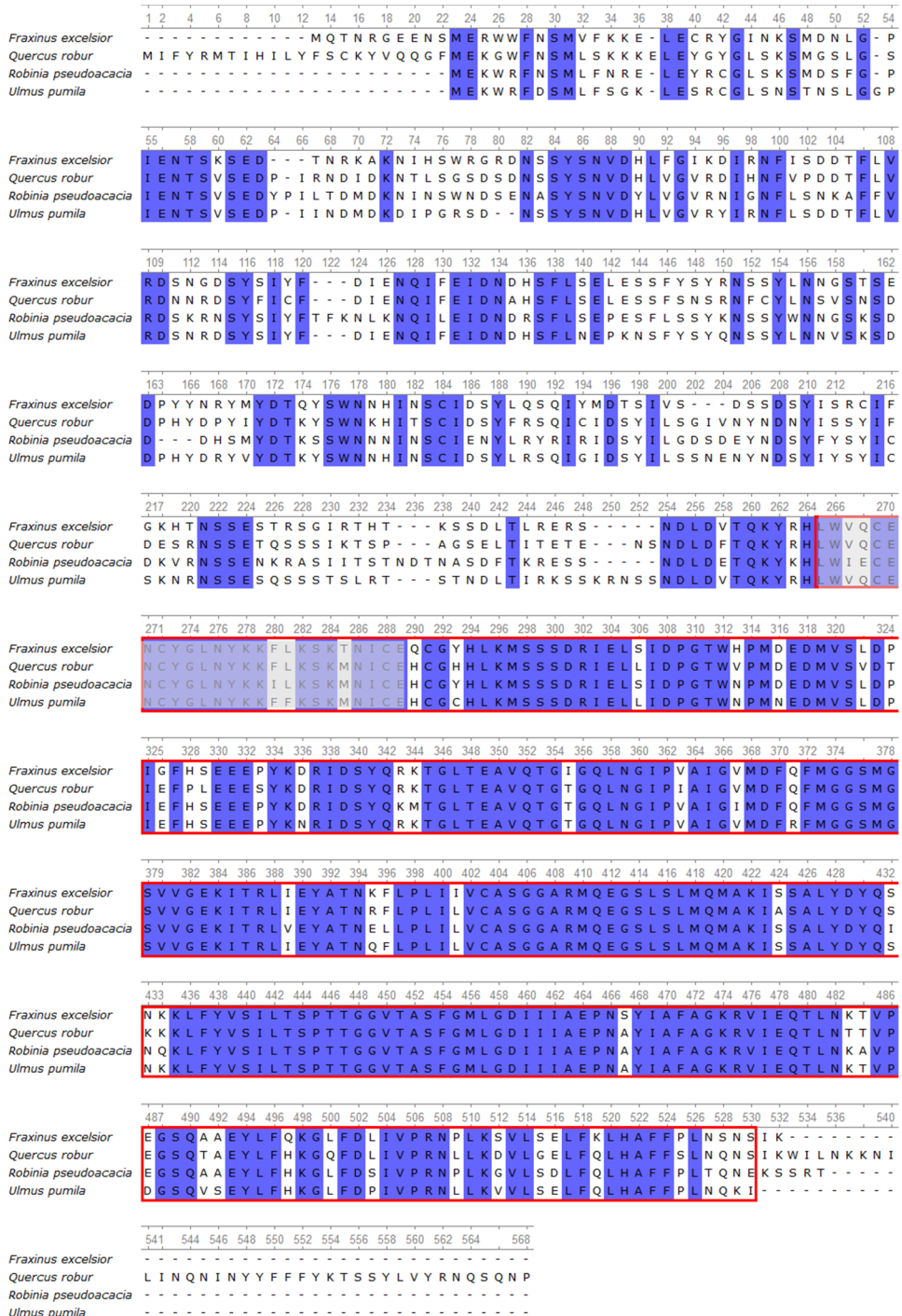
Zn-dependent proteins also have high similarity values among the plants. However, Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta, chloroplastic (65%) and 50S ribosomal protein L32, chloroplastic (65%) have the lowest values of similarity. Co-dependent proteins have Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta, chloroplastic too. Due to the low percentage of similarity compared to other proteins, conservative sites were identified in these proteins. Conservative sites were identified in these proteins due to the low similarity compared to other proteins (Figures 3, 4).
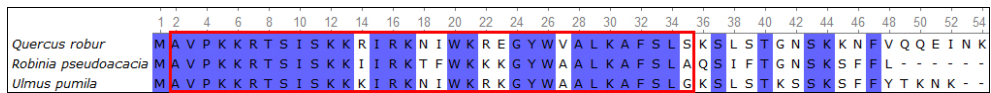
Fig. 3. Multiple alignment of 50S ribosomal protein L32. Areas with a 90 % conservstion level or more are colored in blue. A fragment characteristic of bacterial ribosomal protein bL32 family is highlighted in red.

We searched for domains in proteins with the lowest percentage similarity: 50S ribosomal protein L32, chloroplastic and Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta, chloroplastic. The protein 50S ribosomal protein L32 belongs to the bacterial ribosomal protein family bL32. This protein is a structural component of the ribosome and is involved in the translation process. No domains have been identified for this protein according to the UniProt database. The Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta, chloroplastic protein contains a CoA carboxyltransferase N-terminal domain and a zinc

finger domain. The domain is a component of the acetyl-CoA carboxylase complex. The zinc finger domain is involved in binding zinc ions.

Fig. 4. Multiple alignment of Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta, chloroplastic. Areas with a 90 % conservstion level or more are colored in blue. The CoA carboxyltransferase N-terminal domain is highlighted in red.

Quercus robur had proteins missing for amino acid sequence comparison more often than other trees considered. Sometimes proteins were also absent in *Ulmus pumila* (NAD(P)H-quinone oxidoreductase subunit 2) and *Fraxinus excelsior* (50S ribosomal protein L32, chloroplastic). This may be due to various factors, such as genetic differences (different tree species may have different genetic characteristics, which can lead to diversity in their protein composition), environmental factors (different environments and growth conditions can also affect the presence or absence of certain proteins in trees), different ages of trees (trees of different ages may have different protein compositions, as metabolic processes will differ), climatic conditions (changes in climatic conditions can also affect the protein composition of trees), nutritional peculiarities (trees growing in different locations or on different soils may receive different nutrients, which can also affect the protein content in them), as well as incomplete studies of the species.

Multiple alignment was not performed for Mn-dependent proteins and Cu-dependent proteins of woody plants due to the absence of identical proteins in organisms or due to the presence of fragmented sequences. *Gleditsia triacanthos* is currently poorly studied at the genetic level, and the sequences of metal-dependent proteins found are only present in fragmented form. Therefore, it is not possible to compare the degree of similarity with full sequences of other proteins.

Thus, metal-dependent proteins in these woody plants have not undergone significant evolutionary changes and, on the contrary, have maintained a high degree of similarity in their amino acid sequences. Conservation in amino acid sequences implies the preservation of certain amino acid substitutions during the evolution of the protein gene. This means that some substitutions may occur, but they will not change the fundamental functions of the protein. Such substitutions typically occur slowly and are unlikely to arise. Conservative regions of protein genes are crucial for preserving protein function, and their alteration may result in loss of protein functionality. In this case, the primary functions of proteins involved in photosynthesis, respiration, and other processes have been preserved. This could be due to the stability of the protein structure and its importance for the survival of the plant organism. Particularly, the sequences of Fe-dependent proteins involved in photosynthesis and respiration are not distinctive. Amino acid sequence conservatism can be caused by various reasons, with one of the main reasons being the preservation of protein functionality that has existed evolutionarily for a long time.

4 Conclusion

The study showed a low degree of knowledge of the metalloproteome of a number of forest plants used in agroforestry. Protein descriptions prevail for those involved in major physiological and biochemical processes. Among the studied forest plants, the metalloproteome of *Robinia pseudoacacia* is the most described, which includes 26 proteins including their subunits comparing to other plants.

Structural homology analysis showed high conservative values among the proteins involved in key processes such as photosynthesis and respiration. The degree of conservative amino acid sequences between tree species ranges from 65% to 100%, with higher values, for example, for Fe-dependent proteins. Multiple alignment of Mn-dependent and Cu-dependent proteins of forest plants was not performed due to the absence of identical proteins among the studied plants or due to the presence of a fragmented sequence.

The obtained data can serve as the basis for the development of new technologies for managing productivity and adapting to adverse environmental factors such as drought and soil salinization in forest plants based on the introduction or removal of ions of the studied metals.

References

1. K.N. Kulik, Scientific Agronomy Journal **3,** 08-13 (2022)

2. V.M. Kretinin, K.N. Kulik, A.V. Koshelev, Vestnik of the Russian agricultural science **1**, 23-26 (2020)

3. G. DalCorso, A. Manara, S. Piasentin, A. Furini, Metallomics **10**, 1770-1788 (2014)

4. M. Moustakas, Materials **14**, 549 (2021)

5. Yu.N. Vodyanitskii, Dokuchaev soil bulletin **68**, 56-82 (2011)

6. E.A. Permyakov, Encyclopedia **1**, 261-292 (2021)

7. A. Bernard, F. Lheureux, E. Dirlewanger, Tree Genet. Genomes **14**, 1 (2018)

8. J. Zhang, W. Zhang, F. Ji, J. Qiu, X. Song, D. Bu, G. Pan, Q. Ma, J. Chen, R. Huang, Y. Chang, D.Pei, Plant Biotechnol. J. **18**, 1848-1850 (2020)

9. F.Ji, Q. Ma, W. Zhang, J. Liu, Y. Feng, P. Zhao, X. Song, J. Chen, J. Zhang, X. Wei, Y. Zhou, Y. Chang, P. Zhang, X. Huang, J. Qiu, D. Pei, Genome Biol. **22**, 300 (2021)

10. X. Yang, K. Zhang, T. Nvsvrot, Y. Zhang, G. Cai, L. Huang, W. Ren, Y. Ding, J. Hammond, L. Shi, N. Wang, Plant J. **111**, 1753-1767 (2022)

11. J. Vendruscolo, A.M. Perez Marin, E. dos Santos Felix, K.R. Ferreira, W.C.S. Cavalheiro, I.M. Fernandes, Land Degrad. Dev. **32**, 684-698 (2021)

12. Y. Sasaki, Y. Nagano, Biosci. Biotechnol. Biochem **68**, 1175-1184 (2004)

13. S. Gerhardt, S. Echt, M. Busch, J. Freigang, G. Auerbach, G. Bader, W.F. Martin, A. Bacher, R. Huber, M. Fischer, Plant Physiol. **132**, 1941-1949 (2003)

14. X. Ma, W. Wang, F. Bittner, N. Schmidt, R. Berkey, L. Zhang, H. King, Y. Zhang, J. Feng, Y. Wen, L.Tan, Y. Li, Q. Zhang, Z. Deng, X. Xiong, S. Xiao, Plant Cell. **28**, 1108-1126 (2016)

15. B. Sekula, M. Ruszkowski, Z. Dauter, Int. J. Biol. Macromol. **151**, 554-565 (2020)

16. A.H. Millar, H. Eubel, L. Jänsch, V. Kruft, J.L. Heazlewood, H.P. Braun, Plant Mol. Biol. **56**, 77-90 (2004)