Search for motifs, associated with response to abscisic acid in promoters of genes of *Quercus robur* and *Populus trichocarpa* plants

Vsevolod Malov*, and *Ludmila* Derevshchikova

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

Abstract. Climate changing and increasing anthropogenic impact put more and more fertile lands under treat of degradation. To counteract this process, diverse methods, including agroforestry, should be applied, which, regarding changing environmental condition, requires development of new drought resistant tree forms and varieties. *Quercus robur* and *Populus trichocarpa* are actively used for this cause on the territories of southern Russia, their genomes are sequenced and annotated automatically, which facilitates search for potential genes of interest, which impact drought tolerance. Abscisic acid is a key participant of water deficiency and other abiotic stress response regulation, so genes, which expression is activated by it, are of interest as targets for further molecular selection and expression regulation researches. Promoter elements determining gene expression in response to this signal are already known. In the absence of experimental data about certain genes and their products, discovering such elements in their promoters allows to predict promising genes of interest with high probability. In this research, we conducted search for abscisic acid response elements ABRE and CE1 in promoters of genes of *Quercus robur* and *Populus trichocarpa*, and potential genes of interest were found.

1 Introduction

Every year we can observe increasing area of degraded lands and emerging of new deserts and semi-deserts. Desertification starts under anthropogenic pressure and increasing of mean temperatures. This leads to expansion of arid climate zones. This climate is characterized by hot dry summer and cold dry winter. Precipitation is low, or even absent for years. Expansion of degraded lands affects ecology and economy of the world. Ecological problem – disappearing of certain territory biocenosis. Economical problem – reducing area suitable for agriculture. [1]

Agroforestry is among the methods of degraded land reclamation and climate conditions improvement. [1-3] This technique is based on using plants, already adapted to the environment. Usage of this technique assumes planting previously selected trees forming

* Corresponding author : malov-v@vfanc.ru

forest belts. Main purpose of forest belts is formation of a new ecosystem or preserving the existing one, protection agricultural lands from soil erosion, reduction of the area of degraded land. Previously, trees for forest belts were selected by phenotypical traits such as root system, leaves survival at the dry period. Newly developed methods of drought adaptation mechanism studying at the molecular-genetic level, allows selection desired genotypes, required for reclamation of degraded lands.

Drought response pathways are well studied in model plant *Arabidopsis thaliana*, but genomes of many other, especially, non-agricultural plants, are sequenced, but annotated automatically, without isolation of gene products and their characterization. Nevertheless, there still is a task of selection of genes and regulatory elements in genomes of trees used for agroforestry of lands in southern Russia for further molecular selection and expression regulation researches.

Predunculate oak *(Quercus robur L.*) and Black cottonwood (*Populus trichocarpa Torr. et A.Gray*) are widely used for agroforestry because they are relatively drought resistant [4]. Genomes of these trees are sequenced and annotated automatically.

Abscisic acid (ABA) pathway is one of the main pathways of drought response, it is well studied at the model plant *Arabidopsis thaliana*, and there is data supporting its common functionality among most of flowering plants. ABA is produced at water deficiency and influences at all organism levels: transcription regulation, metabolism and transportation of substances. It spreads through different tissues of a plant and influences its life processes, such as stomata closing, seed germination and shoot growth control. [5, 6] As a transmitter of drought response signal is initiates expression of genes taking part in drought tolerance mechanism. Proteins AREB/ABF, are discovered to be ABA-dependent transcription factors binding ABRE motif. [7]

They are binding promoters – regulatory areas placed before genes to activate transcription, if promoter contains certain motifs – short, often partially variable sequences with a certain function.

If promoter sites determining expression in response to certain factors are known, their presence or absence makes possible to predict which factors activates gene expression, in which signal pathways it takes part and which function it may have.

Availability of sequenced genomes and knowledge of functional elements of promoters allows us to automatize motif search in promoters and, so, predict genes, taking part in certain processes in cases, where experimental studying of gene products was not conducted. This allows us to simplify gene selection for further, more detailed research, which will be useful for molecular selection and expression regulation experiments planning.

The aim of this research is to find motifs associated with ABA response in gene promoters of trees used in agroforestry of southern Russia, *Q. robur* and *P. trichocarpa*.

2 Materials and methods

Because of small length (sometimes less than 10 nucleotides), large number of occurrences of motifs appears in genome just by chance, not executing any biological function. Motifs, located in promoter areas before (5’ side) genes act as cis-regulatory elements.

In the case of ABRE (ABA-responsive element) motif, one element in the promoter does not cause gene expression activation in response to ABA, at least two of ABRE motifs required, or ABRE with one of coupling elements, among which are CE1-3 (Coupling element 1-3) and DRE (Drought responsive element), which also takes part in cold response. [8]

To search for motifs associated with ABA response, we used FIMO program [9], which is a part of a MEME-SUIT 5.5.5 package for nucleic acid motifs manipulation. A pipeline for search process automation was created. It combines FIMO program and selection of

fitting motifs frim its output. Program FIMO accepts input files with motifs in form of nucleotide frequency matrices in meme format, searches for these motifs in DNA sequences and counts p-value for each motif and rejects those with p-value higher than 1e-4. Changing these criteria would be inappropriate because further selection is planned and more strict criteria bears risk of not finding functional sites.

Pipeline takes genome files in fasta format and annotation in gff format (NCBI-pattern of columns required) as input, promoter area length before genes, files with motifs and number of occurrences of each motif. From the motifs found by FIMO, those located in promoter areas (user determined number of nucleotides before gene) are selected, and there, motifs with none less than specified number occurrences are selected from them.

The most used in agroforestry in the south of Russia tree species, which genome is sequenced and annotated are *Q. robur* and *P. trichocarpa*. In this research, we conducted search for cis-regulatory elements ABRE and CE1 in the gene promoters of these trees.
Tree genomes were taken from Refseq and Genbank NCBI databases, identifiers are: GCF_932294415.1 (*Q. robur*) and GCA_000002775.4 (*P. trichocarpa*).

3 Results

Motif sequences of ABRE are taken from research [10], for CE1 – from [11]. Nucleotide frequency matrices based on regulatory elements sequences from those researches were created in meme format for search for motifs, which fit those matrices. Nucleotide frequency matrix for ABRE motif (Figure 1) and for CE1 motif (Figure 2) were used as input for FIMO program.

```
MEME version 5.5.5 (Thu Sep 14 08:48:04 2023 +1000)

ALPHABET= ACGT

strands: + -

Background letter frequencies (from uniform background):
A 0.25000 C 0.25000 G 0.25000 T 0.25000

MOTIF ABRE from

letter-probability matrix: alength= 4 w= 9 nsites= 20 E= 0
  0.42    0.42    0.04    0.12
  0.125   0.625   0.125   0.125
  1.0     0.00    0.00    0.00
  0.00    1.0     0.00    0.00
  0.00    0.00    1.0     0.00
  0.00    0.00    0.00    1.0
  0.00    0.00    1.0     0.00
  0.04    0.04    0.46    0.46
  0.167   0.708   0.000   0.125
```

Fig. 1. Nucleotide probability matrix for ABRE motif in the meme format.

```
MEME version 5.5.5 (Thu Sep 14 08:48:04 2023 +1000)

ALPHABET= ACGT

strands: + -

Background letter frequencies (from uniform background):
A 0.25000 C 0.25000 G 0.25000 T 0.25000

MOTIF CE1 from

letter-probability matrix: alength= 4 w= 9 nsites= 20 E= 0
  0.231 0.231   0.384   0.154
  0.231 0.461   0.231   0.077
  0.000 0.308   0.384   0.308
  0.00  1.00    0.00    0.00
  1.00  0.00    0.00    0.00
  0.00  1.00    0.00    0.00
  0.00  1.00    0.00    0.00
  0.153 0.385   0.385   0.077
  0.154 0.3075  0.3075  0.231
```

Fig. 2. Nucleotide probability matrix for CE1 motif in the meme format.

In the *P. trichocarpa* genome 2 or more ABRE motifs were found in promoters of 247 genes, ABRE and CE1 motifs are found in promoters of 161 genes. No promoter contained both several ABRE motifs and CE1 motif.

Among these genes are homologues of *A. thaliana* genes, for with expression activation with ABA or ABF was experimentally proven:

POPTR_012G141300v4 (low-temperature-induced 65 kDa protein) — contains 2 ABRE in promoter, homologue of LTI65 (RD29B). [12]

POPTR_004G158500v4 — late embryogenesis abundant protein, homologue of dehydrins and Lea-proteins, gropes, which participation in ABA-response is known by Gene Ontology.

POPTR_001G097600v4 —cell division control protein 48 homolog C, closest homologue of AAA family ATPase *AIA1 (AT1G64110),* which is one of the genes most strongly induced by ABA in A. thaliana [13].

POPTR_014G094200v4 — common plant regulatory factor 1, has similarity with ABA-induced transcription factor GBF3 [13], has ABRE and CE1 in promoter.

POPTR_014G094200v4 — common plant regulatory factor 1, closest homologue of GBF2 [13]. Has ABRE and CE1 in promoter.

POPTR_006G065700v4 — probable galactinol--sucrose galactosyltransferase 6, has similarity at amino acid sequence level with ABA-inducible glycoside-hydrolase AT3G57520 from *A. thaliana*, has 2 ABRE in promoter.

Among genes, for which ABA-induction of expression was stated in Gene Ontology database, ABRE and CE1 elements were found in promoters of these genes:

- POPTR_004G158500, uncharacterized protein, 2 ABRE
- POPTR_010G157900, uncharacterized protein, 2 ABRE

In the *Q. robur* genome two or more ABRE motifs were found in promoters of 190 genes and ABRE and CE1 were found in promoters of 185 genes. No promoter contained both several ABRE motifs and CE1 motif.

Among them, closest homolog of transcription factor GBF3, LOC126721195, common plant regulatory factor 1-like, which is one of the most ABA-inducible *A. thaliana* genes. [13]. Has ABRE and CE1 in promoter.

LOC126689664 — dehydrin Xero 1-like, closest homologue of RAB18 dehydrin [13] in *Q. robur* genome, has 3 ABRE in promoter.

LOC126733101 — dehydrin Rab25-like, second closest homologue of RAB18 in *Q. robur* genome, has 4 ABRE-motifs in promoter.

LOC126721195 — common plant regulatory factor 1, closest homologue of GBF2 [13]. Has ABRE and CE1 in promoter.

4 Discussion

Regulatory elements in gene promoters are not always same sequences, nucleotides in different position can vary, that is why frequency matrices are used. Nucleotide frequencies can vary between species. The value of researches [10] and [11] is presence of regulatory element sequences from divers, very distant species, both monocots and dicots, which shows conservativeness of CE1 and ABRE.

In this research more than hundred potential genes – drought tolerance factors were found and may be targets of further researches aiming to produce more drought tolerant varieties and forms of trees. Obviously, such short sequences, like transcription factors recognizable motifs, can appear in random genome location, being non-functional. It is known, than single ABRE motif in gene promotor does not cause expression activation in response to ABA, and frequency of single ABRE occurrences in random data does not differ significantly from those in genomic data. But two such motifs are already a signal for transcription factors ABF/ABRE [14] and, so, should undergo selection pressure. Random occurrence of such motif pair is unlikely. Automatic annotation haven't shown homologs of AREB/ABF in *Q. robur* genome, but ABA-response elements typical for majority of plants are found in promoters of several hundred genes. Among genes, containing such elements in both *Q. robur* and *P. trichocarpa*, are *A. thaliana* proven ABA-inducible genes homologs, which speaks for existing of transcription activation pathway analogous to ABF/AREB in *Q. robur*.

5 Conclusion

The search for promoters, containing motifs of interest was done automatically, using a newly developed program, which automatically processes FIMO output and selects genes with specified number of motifs of different types in promoter. This way motifs associated with response to ABA were found in promoters of *Q. robur* and *P. trichocarpa*, and genes of interest for further researches were determined. In addition to this, convenient instrument for finding known motifs in gene promoters of any organisms which genome is sequenced and gene borders are known.

This research work supported by the framework of the state task Ministry of Science and Higher Education of the Russian Federation No. FNFE-2022-0022 "Search and management of patterns of expression of forest and cultural plant genes responsible for adaptation to environmental hazards and productivity".

References

1. K.N. Kulik, M.V. Vlasenko, CSCEE **9** (2024)
2. J. Dollinger, S. Jose, Agroforestry Systems **92**, 213-219 (2018)
3. V. Rolo, D. Rivest, E. Maillard, G. Moreno, Soil Use and Management **39**, 1006-1032 (2023)

4. J. K. Rauschendorfer, C*haracterizing drought adaptations, phenotypic plasticity, and fixed gene expression patterns within quercus* (Technological University, 2022)
5. A. Daszkowska-Golec, Drought Stress Tolerance in Plants **2**, 123-151 (2016)
6. A.M. Muhammad, M. Waseem, B.H. Jakada, E.J. Okal, Z. Lei, H.A.S Saqib, W. Xu, W. Yuan, Q. Zhang, Int J Mol Sci **23**, 1084 (2022)
7. H. Choi, J. Hong, J. Ha, J. Kang, S.Y. Kim, J Biol Chem. **21**, 1723-30 (2000)
8. Q.J Shen, J.A Casaretto, P. Zhang, TH Ho, Plant Mol Biol. **54** (2004)
9. C.E. Grant, T.L. Bailey, W.S. Noble, Bioinformatics **27**, 1017-1018 (2011)
10. W. Li, X. Cui, Z. Meng, X. Huang, Q. Xie, H. Wu, H. Jin, D. Zhang, W. Liang, Plant Physiol **158**, 1279-92 (2012)
11. W. Li, X. Cui, Z. Meng, X. Huang, Q. Xie, H. Wu, H. Jin, D. Zhang, W. Liang, Plant Cell **7**, 295-307 (1995)
12. K. Shinozaki, K. Yamaguchi-Shinozaki, Curr Opin Plant Biol. **3**, 217-23 (2003)
13. Y. Fujita, M. Fujita, R. Satoh, K. Maruyama, M.M. Parvez, M. Seki, K. Hiratsu, M. Ohme-Takagi, K. Shinozaki, K. Yamaguchi-Shinozaki, Plant Cell. **17**, 3470-88 (2005)
14. J.L. Gómez-Porras, D.M. Riaño-Pachón, I. Dreyer, J.E. Mayer, B. Mueller-Roeber **1(8)**, 260 **(**2007)