DOI: 10.1007/s10535-017-0765-0 BIOLOGIA PLANTARUM 62 (2): 299-306, 2018

Identification of genes associated with drought tolerance in barley

S.F. ABOU-ELWAFA

Agronomy department, College of Agriculture, Assiut University, 71526 Assiut, Egypt

Abstract

Mapping of quantitative trait genes (QTGs) associated with drought related traits is essential for improving drought tolerance in crop species. *In silico* identification of candidate genes relies on annotation of critical QTGs to a variety of web resource-based datasets. The barley reference sequence was employed to map QTGs significantly associated with the proline accumulation and osmotic potential. Annotation of the critical QTGs contigs to the NCBI protein database identified 72 gene orthologs located on chromosomes 1H, 2H, and 7H, from which seven genes were identified as candidates. Expression analysis of all seven candidate genes revealed differential expression pattern between plants grown under well-watered conditions and drought-stress. The results represent a successful and highly powerful implementation of genome-wide scanning approach based on *in silico* mapping of QTGs to identify gene clusters having a common transcript pattern with similar function.

Additional key words: genome-wide-scanning, *Hordeum vulgare*, *in silico* mapping, phylogenetic analysis.

Introduction

Barley represents the fourth most abundant cereal crop worldwide. Almost 75 % of the global barley production is used for animal feeding, 20 % for malting and beverages industry, and 5 % as an ingredient in a panel food product (http://faostat.fao.org). Cultivated barley is among the oldest domesticated crop species and is widely adapted to diverse environmental conditions (Purugganan and Fuller 2009). Barley has a complex genome with an estimated size of 5.1 billion bp with approximately 84 % of mobile elements and repeat structures (Ariyadasa *et al.* 2014).

Drought tolerance is the ability of a plant to survive, grow, and produce yield with limited water supply. Great progress has been achieved in the last two decades in understanding the genetic control of drought tolerance (Abou-Elwafa 2016a). Several studies have indicated that proline content increases during drought stress, and proline accumulation is associated with improvement in drought tolerance in tall fescue and other plants (Seki *et al.* 2007, Zhang *et al.* 2009, Man *et al.* 2011). Proline is involved in cell osmotic adjustment (OA) and protection of cell components during dehydration (Zhang *et al.* 2009). Osmotic adjustment helps to preserve pressure potential, which allows cell enlargement and plant growth as well as stomata to remain at least partially opened and CO_2 assimilation to continue during water stress (Alves and Setter 2004).

Association mapping is an approach applied to identify QTLs in natural populations with high mapping resolution and lesser research effort. Furthermore, association mapping implement linkage disequilibrium (LD) between markers and closely linked QTLs present in a population consists of large number of accessions (Abou-Elwafa 2016b,c, Ersoz *et al.* 2007). However, identification of genes underlying QTLs of complex traits necessitates a careful rethinking about combining both molecular breeding approaches and new genomic and bioinformatic platforms.

Recent advances of whole genome sequence analysis tools and technologies have clearly facilitated the exploring black box of polygenic complex traits with a more accurate description of the number, distribution, and interaction of loci affecting phenotypic variation. Major progress has been achieved in this field with the advent of genomics and its potential contribution to development of quantitative genetics. Recent advances of whole genome

Submitted 21 March 2017, *last revision* 24 August 2017, *accepted* 26 August 2017.
Abbreviations: DArT - diversity arrays technology; IBSC - International Barley Sequencing Consortium; LD - linkage disequilibrium; NCBI - National Center for Biotechnology Information; OP - osmotic potential; PA - proline accumulation; QTGs - quantitative trait genes; QTL - quantitative trait loci; RT-PCR - reverse transcription polymerase chain reaction.
Acknowledgements: The author gratefully acknowledge the City of Scientific Research & Technology Applications (SRTA-City), Egypt, for the excellent help in conducting expression analysis.
Fax: (+20) 882331483, e-mail: elwafa75@aun.edu.eg

sequence analysis tools and technologies have accelerated and facilitated *in silico* identification of candidate genes in model organisms with a relatively small genome size (Abe *et al.* 2012, James *et al.* 2013, Takagi *et al.* 2013). *In silico* identification as computation-based approach includes methods that describe computational framework to prioritize the most likely candidate genes through annotation of genomic sequences to gene functional information from biological ontology sources available through the internet. The typical example of this approach is the positional candidate gene by using gene ontology (Harhay and Keele 2003, Hristovski *et al.* 2005, Rossi *et al.* 2006).

The present study aimed to *1*) *in silico* mapping of molecular markers significantly associated with drought-stress related physiological traits, *i.e.*, proline accumulation and osmotic potential, to identify critical quantitative trait genes (QTGs), *2*) identification and phylogenetic analysis of candidate genes located to critical chromosomal regions, and *3*) analysis of the expression of identified candidate genes.

Materials and methods

In silico identification of drought tolerance associated genes: Based on a previous association mapping study for drought tolerance in barley (Abou-Elwafa 2016b), molecular markers revealing significant marker-trait associations with proline accumulation (PA) and osmotic potential (OP) were used to *in silico* identification of chromosomal regions harboring genes responsible for proline metabolism and osmotic adjustment. The publicly available barley genome sequences (a sequence-enriched barley physical map; International Barley Sequencing Consortium (IBSC) 2012, http://mips.helmholtz-muenchen.de/plant/barley/) were employed to generate a consensus reference sequence. Contigs were concatenated according to their order in the physical map using the online bioinformatics tool Galaxy Project (Giardine *et al.* 2005).

Molecular markers significantly associated with proline accumulation and osmotic potential were *in silico* mapped to the consensus sequence of the respective chromosomes. In that context, molecular markers significantly associated with proline accumulation were *in silico* mapped to chromosomes 1H and 7H. Marker significantly associated with osmotic potential was mapped to chromosome 2H. To ensure the physical positions of the significantly associated markers, tightly linked genetic markers as identified from the GrainGenes 2.0 database (http://wheat.pw.usda.gov/GG2/index.shtml) were *in silico* mapped to their corresponding chromosomes as well. The identified chromosomal regions were annotated to the NCBI Reference Sequence (RefSeq 8) protein database (http://www.ncbi.nlm.nih.gov/refseq) by *BLASTX* searches and the barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/).

Multiple sequence alignment and phylogenetic analysis: Orthologous genes underlying putative proline accumulation and osmotic potential in other plant species were identified by *BLASTP* searches of the NCBI RefSeq protein database. To help infer orthology by bidirectional best hit (BBH), analysis was performed using barley sequences retrieved previously as queries for *BLASTP* searches against other plant species. The sequences were aligned using *CLUSTAL W*, and unrooted phylogenetic trees were constructed implementing the neighbor-joining algorithm and the *Dayhoff PAM* matrix as implemented in the *MEGA7* software (Kumar *et al.* 2016a).

Expression analysis of candidate genes: Five barley (*Hordeum vulgare* L.) accessions (IG_119451, Kaskade, BR4645a, Sonja, and SW 16199) were provided by National Small Grains Germplasm Research Facility (USDA, ARS, Idaho, USA). They were previously identified as drought tolerant accessions (Abou-Elwafa 2016b). Plants were sown in July 2016 in a glasshouse at Assiut University Experimental Farm, Assiut, Egypt (27° 03' N, 31° 01' E, and 70 m asl). Seeds were sown in plastic pots filled with soil. The well-watered plants were irrigated with about 500 cm^3 of water per pot each day by drip irrigation. The drought-stressed plants were irrigated as well-watered plants until they were 8-week-old, then water was decreased to 125 cm^3 twice a week. The second upper leaf was harvested from the 10-week-old plants, frozen in liquid nitrogen, and stored. Total RNA was extracted by the guanidine hydrochloride RNA extraction method as described by Logemann *et al.* (1987). The purity and quality of RNA was verified by running an aliquot on a 1 % (m/v) agarose gel. An amount of *ca*. 0.5 µg DNase-treated RNA was reverse transcribed using the first strand cDNA synthesis kit (*Thermo Scientific Fermentas Molecular Biology Solutions*, St. Leon-Rot, Germany) and the cDNA was diluted ten times before semi-quantitative RT-PCR assay. Primers for semi-quantitative RT-PCR were designed and optimized for *HvARR12, HvPip, HvDI19-2, HvADC, HvHRGP, HvPRP*, and *HvC4H* in addition to the housekeeping gene *HvGAPDH* as a reference (Table 1 Suppl.). Expression was assayed in three biological replicates and three technical replicates by separating PCR-amplified fragments on 2 % agarose gels in 0.5× Tris-borate-EDTA (TBE) buffer, stained with ethidium bromide, and visualized with UV radiation. Semi-quantitative RT-PCR data were quantified and transformed into a graphical format using the gelanalyzer software (http://www.gelanalyzer.com/). Analysis of variance (*ANOVA*) of expression of candidate genes was performed using *PROC MIXED* in *SAS v 9.3*. (*SAS Institute Inc.*, Cary, NC, USA). Fisher's least significant

difference (LSD) test at a probability level of 5 % was employed to analyze samples exhibiting significantly different means.

Results

We employed the barley reference sequence (IBSC 2012) to identify quantitative trait genes (QTG) associated with proline accumulation and osmotic potential as drought tolerance related traits. Contigs from draft reference sequences (IBSC 2012) were concatenated to seven pseudo molecules according to their position on the physical map. Contigs assembly resulted in a total sequence of 2 136 billion bp assigned to the seven chromosomes in addition to unassigned sequences referred to as chromosomes 0H. The physical length of pseudo chromosomes ranged from 251.62 Mb (1H) to 340.52 Mb (2H) (Table 1). A molecular marker on chromosome 1H (*Bmag0872*) revealed a significant association with proline accumulation and five closely linked diversity arrays technology (DArT) markers at genetic distance of 1.3 - 5.7 cM (*bPb-9418*, *bPb-4813*, *bPb-3217*, *bPb-9333*, and *bPb-1231*) were *in silico* mapped to *ca*. 11.2 Mb contig of the reference sequence of barley chromosome 1H (Fig. 1 Suppl.). Similarly, marker *GBM1012* was significantly associated with osmotic potential and four closely linked DArT markers (*bPb-1611*, *bPb-1181*, *bPb-4601*, and *bPb-4232*) were *in silico* mapped to *ca.* 10.1 Mb contig of the reference sequence of barley chromosome 2H (Fig. 1 Suppl.). Moreover, marker *GBMS183* which exhibited a significant association with proline accumulation and four closely linked DArT markers (*bPb-0375*, *bPb-3020*, *bPb-0995*, and *bPb-5403*) were *in silico* mapped to *ca*. 14.3 Mb contig of the reference sequence of barley chromosome 7H (Fig. 1 Suppl.).

Table 1. Physical length of different barley chromosomes.

Chromosome	Length [bp]
1H	251 619 801
2H	340 517 144
3H	302 677 518
4H	285 183 503
5H	309 083 703
6H	266 363 473
7H	335 273 546
Non-anchored contigs	45 099 519
Total	2 135 818 207

We focused on the critical regions on chromosomes 1H, 2H, and 7H to select candidate sequences as drought tolerance QTGs. The genomic sequences spanning the marker significantly associated with drought tolerance related traits of each chromosome, *i.e.*, *ca*. 6 Mb of each of chromosomes 1H, 2H, and 7H, were annotated to the NCBI RefSeq protein database by *BLASTX* searches. Thirty-four sequences with high sequence similarities to functionally characterized genes from other plant species were identified on the critical genomic region of chromosome 1H (Table 2 Suppl.). These candidate genes including genes which are known to be involved in resistance to plant pathogens, *e.g*., *Disease resistance protein RPM1* (*RPM1*), *Receptor-like serine/threonine-protein kinase* (*LRR*), *Wall-associated receptor kinase 3* (*WAK3)*, and *Kinase-like protein* (*TMKL1*). Furthermore, three genes *Dehydration-induced 19 homolog 2* (*DI19-2*), *hydroxyproline-rich glycoprotein* (*HRGP*), *Proline iminopeptidase* (*Pip*), and two-component response regulator orthologs (*At*ARR12) implicated in plant defense and drought tolerance were identified. Annotation of 6 Mb of chromosome 2H to *RefSeq* identified 18 gene orthologs, three of which *Arginine decarboxylase* (*ADC*), *Hydroxyproline-rich glycoprotein* (*HRGP*), and *Proline-rich family protein* (*PR*) were involved in the response to drought stress (Table 3 Suppl.). Annotation of the respective critical QTGs on chromosome 7H revealed the identification of 20 gene orthologs, two of which *Cinnamate-4-hydroxylase* (*C4H*) and *Hydroxyproline-rich glycoprotein-like* (*HRGPL*) were proven to be implicated in the response to drought stress (Hayat *et al.* 2012, Le Gall *et al.* 2015). Most of the remaining identified gene orthologs are involved in the resistance to various biotic and abiotic stresses such as *Disease resistance protein RPM1* (*RPM1*) (Boyes *et al.* 1998), *Heat shock Protein 70* (*HSP70*) (Yu *et al.* 2015), and *Receptor-like serine/threonine-protein kinase* (*LRR*) (Afzal *et al.* 2008).

Having identified candidate genes proposed to be implicated in response to drought tolerance, a set of gene orthologs from multiple species were retrieved from the *NCBI RefSeq* protein database using *BLASTP* and bidirectional best hit analysis. A phylogenetic analysis was performed in order to understand the evolutionary ancestry underpinning candidate genes. Phylogenetic analysis revealed two major clades of ARR12 proteins (Fig. 1*A*). The dicot plant *Arabidopsis* was represented within an entire clade of its own. The other clade consists of monocot and dicot sequences with *H. vulgare* represented in the same subclade with *Brachypodium distachyon*, *Triticum urartu*, and *Aegilops tauschii*. Protein sequences of proline iminopeptidase (Pip) were assigned to two distinct clades, one of which comprised only monocot sequences while the other clade involved barley and other monocot and dicot sequences (Fig. 1*B*). Phylogenetic analysis of protein sequences of dehydration-induced 19 homolog 2 (DI19-2) exhibited two major clades and the sequence of *Oryza sativa* reveled the highest degree of homology to barley (Fig. 1*C*). Protein sequences of arginine decarboxylase (ADC), hydroxyproline-rich glycoprotein (HRGP),

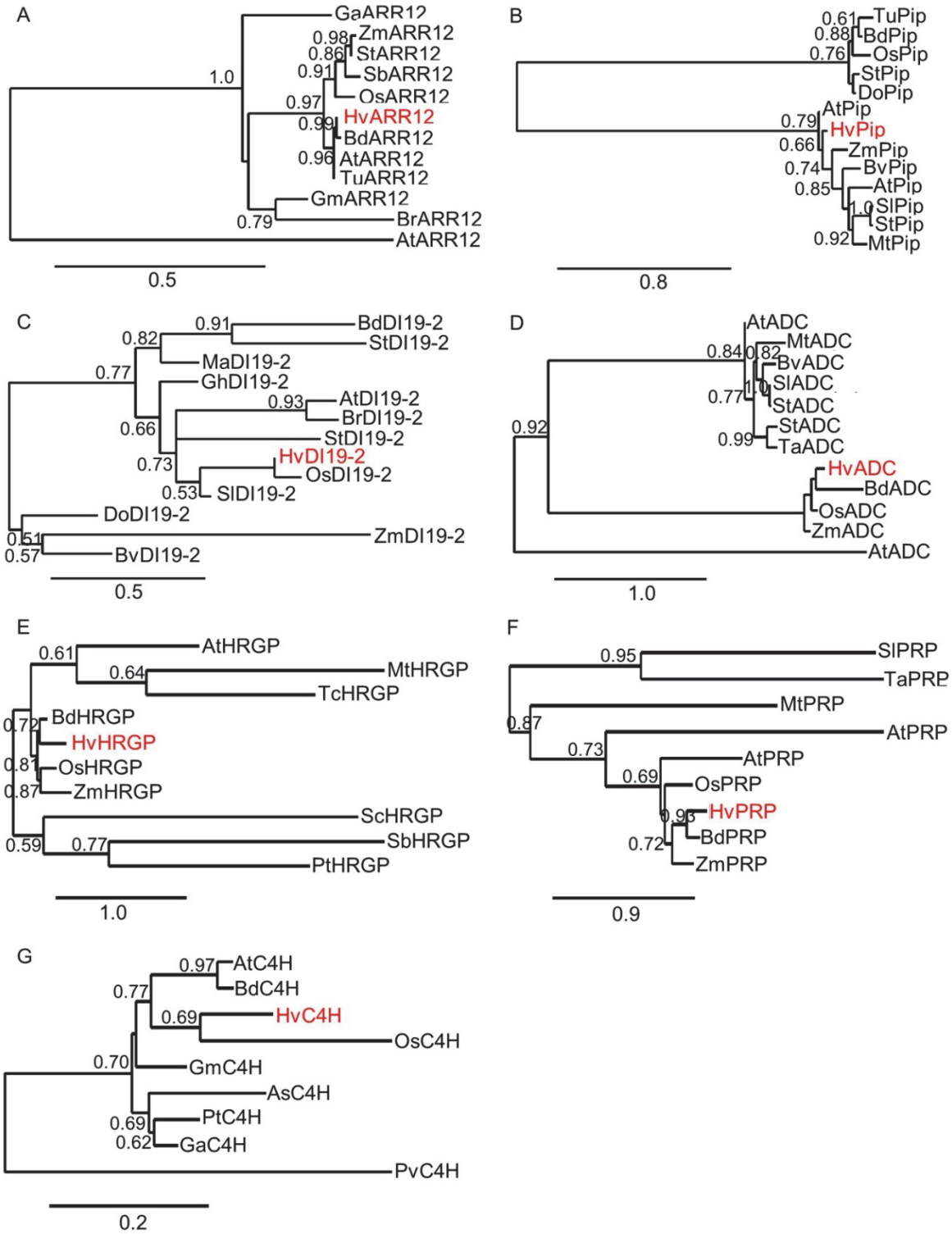

Fig. 1. Phylogenetic analysis of identified candidate genes included all putative plant orthologs retrieved from the *NCBI RefSeq* protein database using *BLASTP* and bidirectional best hit analysis as described in Materials and methods. Unrooted neighbor-joining tree of: *A* - two-component response regulator (ARR12), *B* - proline iminopeptidase (Pip), *C* - dehydration-induced 19 homolog 2 (DI19-2), *D* - arginine decarboxylase (ADC), *E* - hydroxyproline-rich glycoprotein (HRGP), *F* - proline-rich family protein (PRP), and *G* - cinnamate-4-hydroxylase (C4H). Numbers at branching nodes indicate the branching probabilities.

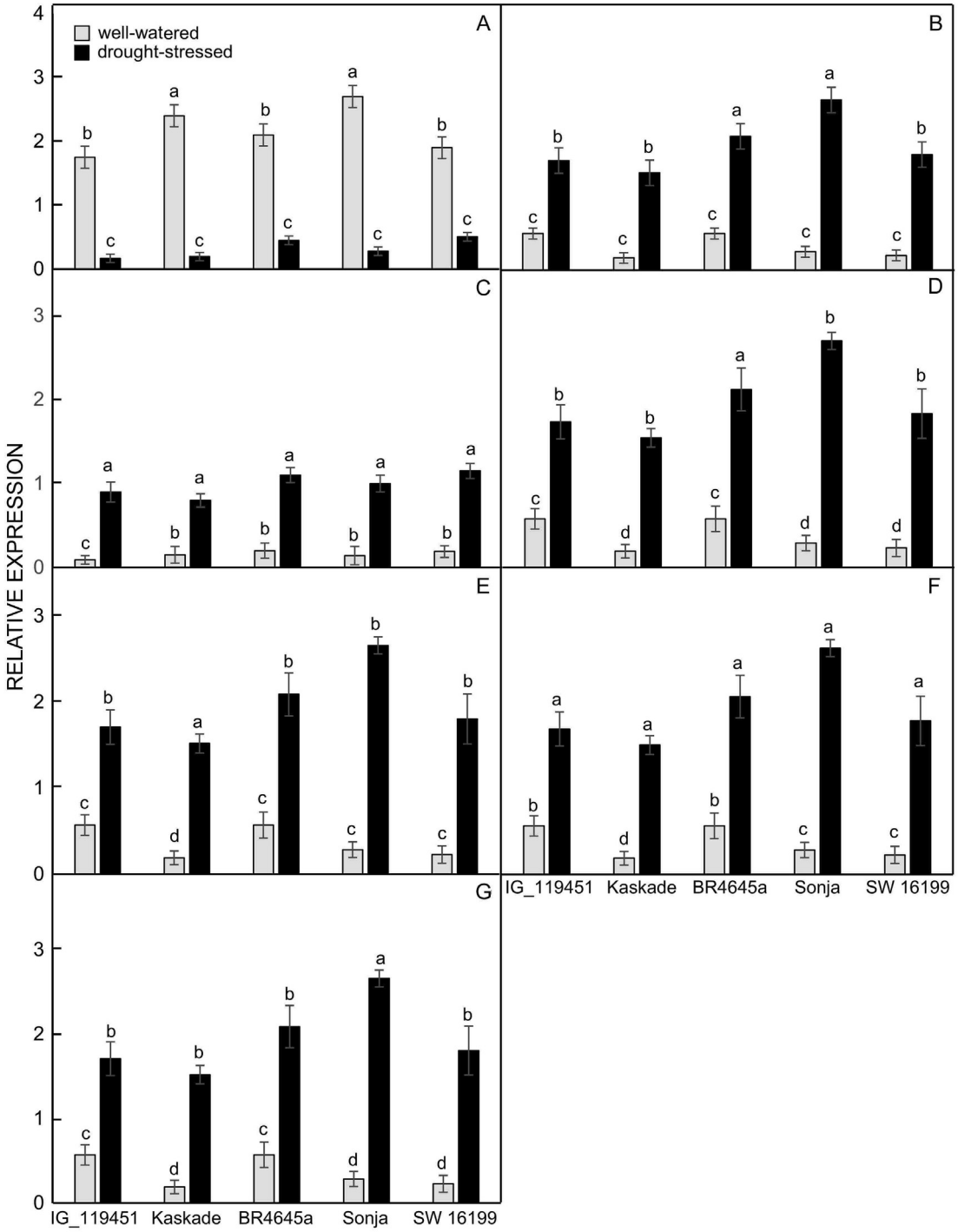

Fig. 2. Semi-quantitative PCR for drought tolerance candidate genes: *A* - *HvARR12*, *B* - *HvPip*, *C* - *HvDI19-2*, *D* - *HvADC*, *E* - *HvHRGP*, *F* - *HvPRP*, and *G* - *HvC4H* in five barley accessions (IG_119451, Kaskade, BR4645a, Sonja, and SW 16199) under drought-stress and well-watered conditions. Gel data were quantified and transformed into a graphical format using the gelanalyzer software (http://www.gelanalyzer.com/). Different letters indicate significantly different expression at a probability level of 5 % as indicated by analysis of variance and LSD.

and proline-rich protein family (PR) were phylogenetically clustered into two major clades. Barley was represented in the same subclade with *Brachypodium distachyon*, *Oryza sativa*, and *Zea mays* (Fig. 1*D*,*E*,*F*). Protein sequences of cinnamate-4-hydroxylase (C4H) were clustered to two major clades, with sequence of *Oryza sativa* closest to barley (Fig. 1*G*).

Based on sequence similarity and presupposed

biological function of the identified candidate gene orthologs, we reasoned that the transcriptional activity of the identified candidate genes could respond to drought stress. Therefore, the expressions of seven candidate genes (*HvARR12*, *HvPip*, *HvDI19*-2, *HvADC*, *HvHRGP*, *HvPRP*, and *HvC4H*) were analyzed in leaves of ten-weeks old plants grown under well-watered and drought-stress conditions. The seven candidate genes were expressed in all plants analyzed, but the expression patterns of all seven genes were different. The expressions of all genes except *ARR12* were highly up-regulated in response to drought stress (Fig. 2).

Discussion

Two approaches could be implemented to dissect genetics of quantitative traits: candidate gene approach and genome-wide scanning. Each approach has specific advantages and disadvantages. With the aid of DNA markers, genome-wide scanning could be applied to located glancing chromosomal regions, which usually embed a large number of candidate genes. Alternatively, candidate gene approach, which is more powerful and economical method for direct gene discovery, has been presented as an extremely effective approach to study the genetic architecture of quantitative traits. However, the successful application of traditional candidate gene approach depends on available knowledge about known or presumed biology of the phenotype under investigation at the molecular level, which is unfortunately still vague for most biological traits. Therefore, it is of great importance to develop alternative approaches to break the limitation of information bottleneck. Rapid development in next generation sequencing technologies has facilitated sequencing of targeted regions from large genomes including genes of interest and linkage regions allowing focus on a critical region of the genome (Teer and Mullikin 2010, Ekblom and Galindo 2011). We are interested in an approach for rapid identification of quantitative traits in crop genomes to empower marker assisted breeding aimed at overcoming global food insecurity by accelerating breeding programs.

In the present study, a bioinformatics pipeline targeting candidate genes for drought-stress response was successfully implemented and led to the identification of seven candidate genes. This pipeline implements both molecular breeding and genomic sequence to identify QTGs. It could be employed in any plant species for which molecular and genomic sequence data are available. The pipeline consists of four complementary steps, *i.e.*, *1*) *in silico* mapping of molecular markers located to the QTGs of interest to determine their physical positions, *2*) identification of candidate genes underlying these QTGs by annotation of critical genomic sequences to the *NCBI RefSeq* protein database behind, *3*) analysis of sequence similarity of candidate genes to other gene orthologs to predict conserved functional roles, and *4*) implication of candidate genes in the response to drought stress based on their expression profiles in plants grown under optimum and stressed conditions. The pipeline implemented here combining both genome-wide scanning and candidate gene approach reveals the genes underlying proline metabolism and osmotic potential as biochemical traits associated with drought tolerance in plants. With the aid of molecular markers, genome-wide scanning was applied to locate the chromosomal regions encompassing QTLs responsible for proline metabolism and osmotic potential accumulation (Abou-Elwafa 2016b). Based on existing knowledge about the known or presumed biology of proline metabolism and osmotic potential, candidate gene approach was implemented.

Annotation of more than 6 Mbp of sequences spanning QTGs on chromosomes 1H, 2H, and 7H to the protein database, revealed the presence of 34, 18, and 20 gene orthologs, respectively. The relatively high number of gene orthologs (34) identified on a comparable genomic region of chromosome 1H compared to both chromosomes 2H and 7H, could be attributed to the position of the QTG on each chromosome. The QTG on chromosome 1H is located close to the centromeric region which contains a large number of functional genes, whereas the QTGs on chromosomes 2H and 7H are located close to the telomeric regions where number of functional genes are drastically reduced. The vast majority of those genes are involved in resistance to biotic and abiotic stresses which may indicate an evolutionary trajectory following barley domestication and natural selection. *ARR12* gene in *Arabidopsis* acts as negative regulator of drought responses in both ABA-dependent and -independent pathways. The loss-of-function of *ARR12* gene enhances drought tolerance, suggesting that ARR12 protein mediates *Arabidopsis* responses to drought as a negative regulator (Nguyen *et al.* 2016). Pip protein is involved in the proline pathway and the expression of *Pip* gene is upregulated during drought treatment (Zhou *et al.* 2015). *DI19* is a drought-induced gene family which encodes seven hydrophilic proteins that contain two atypical Cys2/His2 zinc finger-like domains that are evolutionarily conserved. The expression of DI19-2 in *Arabidopsis* is elevated in response to salt stress (Milla *et al.* 2006). *DI19* gene family is functioning in ABA-independent, dehydration and salinity stress signaling pathways. However, they may also be regulated by other abiotic stresses (Milla *et al.* 2006). Arginine decarboxylase (ADC) is a key plant enzyme that converts arginine into putrescine, which is an important mediator of abiotic stress tolerance (Peremarti *et al.* 2010). Besides, overexpression of *ADC* gene in *Arabidopsis* caused higher content of putrescine and hence improved drought tolerance (Alcazar *et al.* 2006, 2010). Hydroxyproline-rich glycoproteins (HRGPs) are the dominant group of

non-enzymatic cell wall glycoproteins and are mostly categorized into three classes: extensins, proline-rich proteins (PRPs), and arabinogalactan-proteins (AGPs). HRGPs play a crucial role in biotic and abiotic stress responses, partly due to the oxidative cross-linking properties of extensins and PRPs (Deepak *et al.* 2007, 2010, Sujeeth *et al.* 2010). Cinnamate-4-hydroxylase (C4H) converts trans-cinnamic acid (CA) to *p*-coumaric acid (COA) in the phenylpropanoid/lignin biosynthesis pathway. The expression of *C4H* in *Artemisia annua* is significantly elevated during drought (Kumar *et al.* 2016b).

The transcriptional activities of candidate genes indicate a possible functional association among these genes and provide strong evidence for their implication in drought tolerance. Besides, the presence of gene clusters indicates that clustering affords a selective advantage and that some evolutionary mechanism exists to enhance the presence and maintenance of clusters which may belong to common metabolic pathways or directly interact with each other to produce a protein complex or serve as receptors in signaling cascades (Teichmann and Veitia 2004, Yi *et al.* 2007). It is believed that selective pressure is the most common force that promotes clustering and may arise through coordinated gene expression. Alternatively, coinheritance may provide the common power for operating such clustering of genes (Nei 2003).

In conclusion, the results represent a successful and highly efficient implementation of our approach of genome-wide scanning based on *in silico* mapping of molecular markers for the identification of QTGs. These results revealed the presence of gene clusters having a common transcript pattern with similar function points at an evolutionary mechanism to establish quantitative drought stress tolerance.

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S., Terauchi, R.: Genome sequencing reveals agronomically important loci in rice using MutMap. - Natur. Biotechnol. **30**: 174-178, 2012.

Abou-Elwafa, S.F.: Association mapping for drought tolerance in barley. - Gene Technol. **5**: 1-2, 2016a.

Abou-Elwafa, S.F.: Association mapping for drought tolerance in barley at the reproductive stage. - Compt. rend. Biol. **339**: 51-59, 2016b.

Abou-Elwafa, S.F.: Association mapping for yield and yield-contributing traits in barley under drought conditions with genome-based SSR markers. - Compt. rend. Biol. **339**: 153-162, 2016c.

Afzal, A.J., Wood, A.J., Lightfoot, D.A.: Plant receptor-like serine threonine kinases: roles in signaling and plant defense. - Mol. Plant Microbe Interact. **21**: 507-517, 2008.

Alcazar, R., Marco, F., Cuevas, J.C., Patron, M., Ferrando, A., Carrasco, P., Tiburcio, A.F., Altabella, T.: Involvement of polyamines in plant response to abiotic stress. - Biotechnol. Lett. **28**: 1867-1876, 2006.

Alcazar, R., Planas, J., Saxena, T., Zarza, X., Bortolotti, C., Cuevas, J., Bitrian, M., Tiburcio, A.F., Altabella, T.: Putrescine accumulation confers drought tolerance in transgenic *Arabidopsis* plants over-expressing the homologous arginine decarboxylase 2 gene. - Plant Physiol. Biochem. **48**: 547-552, 2010.

Alves, A.A.G., Setter, T.L.: Abscisic acid accumulation and osmotic adjustment in cassava under water deficit. - Environ. exp. Bot. **51**: 259-279, 2004.

Ariyadasa, R., Mascher, M., Nussbaumer, T., Schulte, D., Frenkel, Z., Poursarebani, N., Zhou, R., Steuernagel, B., Gundlach, H., Taudien, S., Felder, M., Platzer, M., Himmelbach, A., Schmutzer, T., Hedley, P.E., Muehlbauer, G.J., Scholz, U., Korol, A., Mayer, K.F., Waugh, R., Langridge, P., Graner, A., Stein, N.: A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. - Plant Physiol. **164**: 412-423, 2014.

Boyes, D.C., Nam, J., Dangl, J.L.: The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. - PNAS **95**: 15849-15854, 1998.

Deepak, S., Shailasree, S., Kini, R.K., Hause, B., Shetty, S.H., Mithöfer, A.: Role of hydroxyproline-rich glycoproteins in resistance of pearl millet against downy mildew pathogen *Sclerospora graminicola*. - Planta **226**: 323-333, 2007.

Deepak, S., Shailasree, S., Kini, R.K., Muck, A., Mithöfer, A., Shetty S.H.: Hydroxyproline-rich glycoproteins and plant defence. - Phytopathology **158**: 585-593, 2010.

Ekblom, R., Galindo, J.: Applications of next generation sequencing in molecular ecology of non-model organisms. - Heredity **107**: 1-15, 2011.

Ersoz, E.S., Yu, J., Buckler, E.S.: Applications of linkage disequilibrium and association mapping in crop plants. - In: Varshney, R.K., Tuberosa, R. (ed.): Genomics Assisted Crop improvement, Genomics Approaches and Platforms. Vol. 1. Pp. 97-119. Springer, Dordrecht 2007.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A.: Galaxy, a platform for interactive large-scale genome analysis. - Genome Res. **15**: 1451-1455, 2005.

Harhay, G.P., Keele, J.W.: Positional candidate gene selection from livestock EST databases using gene ontology. Bioinformatics **19**: 249-255, 2003.

Hayat, S., Hayat, Q., Alyemeni, M.N., Wani, A.S., Pichtel, J., Ahmad, A.: Role of proline under changing environments: a review. - Plant Signal. Behav. **7**: 1456-1466, 2012.

Hristovski, D., Peterlin, B., Mitchell, J.A., Humphrey, S.M.: Using literature based discovery to identify disease candidate genes. - Int. J. med. Inform. **74**: 289-298, 2005.

James, G.V., Patel, V., Nordström, K.J., Klasen, J.R., Salomé, P.A., Weigel, D., Schneeberger, K.: User guide for mapping-by-sequencing in *Arabidopsis*. - Genome Biol. **14**: R61, 2013.

Kumar, S., Stecher, G., Tamura, K.: MEGA7, molecular evolutionary genetics analysis version 7.0 for bigger datasets. - Mol. Biol. Evol. **33**: 1870-1874, 2016a.

Kumar, R., Vashisth, D., Misra, A., Akhtar, M.Q., Jalil, S.U., Shanker, K., Gupta, M.M., Rout, P.K., Gupta, A.K.,

Shasany, A.K.: RNAi down-regulation of cinnamate-4-hydroxylase increases artemisinin biosynthesis in *Artemisia annua*. - Natur. Rev. Sci. Rep. **25**: 26458, 2016b.

Le Gall, H., Philippe, F., Domon, J.M., Gillet, F., Pelloux, J., Rayon, C.: Cell wall metabolism in response to abiotic stress. - Plants **4**: 112-166, 2015.

Logemann, J., Schell, J., Willmitzer, L.: Improved method for the isolation of RNA from plant tissues. - Anal. Biochem. **163**: 16-20, 1987.

Man, D., Bao, Y.X., Han, L.B., Zhang, X.: Drought tolerance associated with proline and hormone metabolism in two tall fescue cultivars. - HortScience **46**: 1027-1032, 2011.

Milla, M.A., Townsend, J., Chang, I.F., Cushman, J.C.: The *Arabidopsis* AtDi19 gene family encodes a novel type of Cys2/His2 Zinc-finger protein implicated in ABA independent dehydration, high-salinity stress and light signaling pathways. - Plant mol. Biol. **61**: 13-30, 2006.

Nei, M.: Genome evolution: let's stick together. - Heredity **90**: 411-412, 2003.

Nguyen, K.H., Hab, C.V., Nishiyamaa, R., Watanabea, Y., Leyva-Gonzálezc, M.A., Fujitad, Y., Trana, U.T., Lia, W., Tanakae, M., Sekie, M., Schallerg, G.E., HerreraEstrellah, L., Trana, L.S.P.: *Arabidopsis* type B cytokinin response regulators ARR1, ARR10, and ARR12 negatively regulate plant responses to drought. - PNAS **13**: 3090-3095, 2016.

Peremarti, A., Bassie, L., Zhu, C.F., Christou, P., Capell, T.: Molecular characterization of the arginine decarboxylase gene family in rice. - Transgenic Res. **19**: 785-797, 2010.

Purugganan, D., Fuller, D.Q.: The nature of selection during plant domestication. - Nature **457**: 843-848, 2009.

Rossi, S., Masotti, D., Nardini, C., Bonora, E., Romeo, G., Macii, E., Benini, L., Volinia, S.: TOM: a web-based integrated approach for identification of candidate disease genes. - Nucl. Acids Res. **34**: W285-W292, 2006.

Seki, M., Umezawa, T., Urano, K., Shinozaki, K.: Regulatory metabolic networks in drought stress responses. - Curr. Opin. Plant Biol. **10**: 296-302, 2007.

Sujeeth, N., Deepak, S., Shailasree, S., Kini, R.K., Shetty, S.H., Hille, J.: Hydroxyproline-rich glycoproteins accumulate in pearl millet after seed treatment with elicitors of defense responses against *Sclerospora graminicola*. - Physiol. mol. Plant Pathol. **74**: 230-237, 2010.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., Innan, H., Cano, L.M., Kamoun, S., Terauchi, R.: QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. - Plant J. cell. mol. Biol. **74**: 174-183, 2013.

Teer, J.K., Mullikin, J.C.: Exome sequencing: the sweet spot before whole genomes. - Human mol. Genet. **19**: R145-R151, 2010.

Teichmann, S.A., Veitia, R.A.: Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. - Genetics **167**: 2121-2125, 2004.

Yi, G., Sze, S.H., Thon, M.R.: Identifying clusters of functionally related genes in genomes. - Bioinformatics **23**: 1053-1060, 2007.

Yu, A.M., Li, P., Tang, T., Wang, J.C., Chen, Y., Liu, L.: Roles of Hsp70s in stress responses of microorganisms, plants, and animals. - Biomed. Res. Int. **2015**: 1-8, 2015.

Zhang, X., Ervin, E.H., Evanylo, G.K., Haering, K.C.: Impact of biosolids on hormone metabolism in drought-stressed tall fescue. - Crop Sci. **49**: 1893-1901, 2009.

Zhou, Z., Ma, H., Lin, K., Zhao, Y., Chen, Y., Xiong, Z., Wang, L., Tian, B.: RNA-seq reveals complicated transcriptomic responses to drought stress in a nonmodel tropic plant, *Bombax ceiba* L. - Evol. Bioinformatics **11**: 27-37, 2015.