African Journal of Biotechnology Vol. 7 (3), pp. 224-233, 5 February, 2008
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB07.212
ISSN 1684–5315 © 2008 Academic Journals

*Full Length Research Paper*

# Nucleotide variation in *ATHK*1 region of *Arabidopsis thaliana* and its association study with drought tolerance

**Jian-Jun Zhang[1,2], Gang-Ping Hao[2,3], Zhong-Yi Wu[2], Xiu-Hai Zhang[2], Yong-Qin Wang[2], Qing Yang[4], Ming-Qing Cao[2], Shi-Gui Liu[1] and Cong-Lin Huang[2*]**

[1]School of Life Science, Sichuan University, Chengdu 610064, China.
[2]Beijing Agro-Biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097.
[3]Taishan medical University, Taian 271000, China.
[4]College of Life Science, Nanjing Agricultural University, Nanjing 210095, China.

Accepted 2 July, 2007

**The *ATHK*1 gene in *Arabidopsis* encodes a putative histidine kinase that is transcriptionally up-regulated in response to changes in external osmolarity. In this work, we investigated the nucleotide variability of the *ATHK*1 gene in a sample of 32 core *Arabidopsis* accessions originating from different ecoclimatic regions and their drought tolerance. The results showed that different accessions had quite difference in adaptation to drought stress. Thirty-two *Arabidopsis* accessions were clustered into four groups according to their drought tolerance capacity. Relative water content of the leaves (RWC) combined anyone of membrane permeability of leaves (MP) and water retain capacity of detached leaves (WRC) were selected as two representative physiological indexes for evaluation of comprehensive drought tolerance. Sequencing 5 515 bp encompassing *ATHK*1 coding region in 32 core accessions revealed 39 polymorphisms, which formed 24 haplotypes. The polymorphism (including single nucleotide polymorphism (SNP) and insertion/deletion (Indel)) frequency was 1 SNP per 131.2 bp. In coding region of *ATHK*1, the ratio of average number of nucleotide difference $\pi_n/\pi_s$ ratio was 0.727, suggesting that the *ATHK1* protein is not constrained against amino acid changes within the species and this gene belonged to the middle evolution rate gene. Using ANOVA analysis, it showed that the 1199 site amino acid (Ser↔stop) variation of the eleventh haplotype (257 and 266av) was associated with not only RWC but also WRC, indicating that the change of Ser↔stop is associated with comprehensive drought tolerance of 257 and 266av. This amino acid change may cause 257 and 266 av accessions originating from moist ecoclimatic region to be sensitive to dry climate, and likely be the evidence of adaptive evolution.**

**Key words:** *Arabidopsis*, *ATHK*1, nucleotide variation, drought tolerance, association study.

## INTRODUCTION

Drought is a major agronomic problem, resulting in yield reduction of crops (Boyer, 1982). Therefore, this type of abiotic stress has been the focus of considerable attention, revealing physiological mechanisms of adaptation (Bohnert and Sheveleva, 1998). With the occurrence of water deficit, many of the physiological processes associated with growth are affected, and under severe deficit plants may die. The effect of water stress may vary with the variety, degree and duration of water stress and the growth stage of plants.

Plant drought tolerance is not only related to biochemical and physiological processes, but also related to the variety, genotype, degree and duration of drought stress and the developmental stage (Lanceras et al., 2004). Drought tolerance is a very complex quantitative trait and affected by not only difference of plants but also

*Corresponding author. E-mail: conglinhuang@hotmail.com. Tel: 86-10-51503801. Fax: 86-10-51503980.

the phase and level of drought. Plant drought tolerance is a quantitative trait controlled by multiple genes, most of which are minor genes, and there is probably one major gene playing important role in this complex genetic background. So the polymorphism of any gene can play a role in plant drought tolerance (Hao et al., 2004). We can dissect complex quantitative trait as a set of single locus trait by analyzing QTLs marker. Historically, linkage analysis is used to measure the genetic proximity of loci to each other, to map quantitative trait. However, linkage analysis has some limitation. First, the limited number of recombination events results in poor resolution for quantitative trait. Second, only two alleles at any given locus can be studied simultaneously (Flint-Garcia, 2003).

Recently, association analysis emerges as a powerful tool to identify QTL in plants. It has been suggested that association analysis could be more powerful than linkage analysis for identification of genes that control quantitative trait (Risch and Merikangas, 1996). The principle of association study is that the joint relationship of genotypes and phenotypes in population samples are studied (Long and Langley, 1999). It is a population-based survey used to identify trait-marker relationships based on linkage disequilibrium (LD). Genetic diversity is evaluated across nature population to identify polymorphism that correlate with phenotypic variation. With the advent of efficient molecular marker technologies and specific statistical methods, association study is developed and extensively used in dissection of quantitative trait. Alternatively, association studies could take advantage of the recent progress made in genomics and use the available set of candidate genes (Long and Langley, 1999).

Drought tolerance is a complex quantitative trait being understood further at the molecular genetic level. Genetic analyses have identified several candidate loci demonstrated to affect drought tolerance of plants (Gutterson and Zhang, 2004). Most of these candidate genes have already been cloned. This provides the opportunity to test naturally occurring molecular variation at the candidate genes for an association with the phenotype of drought tolerance.

Our interest focuses on regulatory genes response to environmental stress. In this context, *ATHK1* gene is a good working model. *ATHK1* was cloned and studied by Urao et al. (1999). The predicted ATHK1 protein in *Arabidopsis* has two putative transmembrane regions in the N-terminal, and has similar structure to the yeast osmosensor synthetic lethal of N-end rule 1 (SLN1). Introduction of the *ATHK1* cDNA into the yeast double mutant sln1Δ sho1Δ, which lacks two osmosensors, suppresses lethality in high-salinity media and activate the high-osmolarity glycerol response 1 (HOG1) and nitogen-activated protein kinase (MAPK) (Urao et al., 1999). The *ATHK1* transcript is more abundant in roots than other tissues under normal growth conditions and accumulates under conditions of high or low osmolarity (Urao et al., 1999). Histochemical analysis of ß-glucuronidase activities driven by the *ATHK*1 promoter further indicates that the *ATHK*1 is transcriptionally upregulated in response to changes in external osmolarity (Urao et al., 1999). These results suggest that ATHK1 functions as an osmosensor and transmits the stress signal to a downstream MAPK cascade.

In this study, nucleotide variation at the *ATHK1* locus was analyzed in worldwide samples of 32 core accessions of *Arabidopsis thaliana.* Drought tolerance evaluation in the same set of *A. thaliana* accessions was used to search for an association with nucleotide variation at the *ATHK1* locus. Our objectives, firstly, to compare the level and pattern of nucleotide sequence variation in the *ATHK1* gene with that of other previously studied genes, secondly, to examine the relationship between the polymorphism in the *ATHK1* genes and drought tolerance.

## MATERIALS AND METHODS

### Plant materials

Thirty-two accessions of *A. thaliana* were used in this study (Table 1). They originated from different ecoclimatic regions, and all were obtained from Versailles stock center (http://dbsgap.versailles.inra.fr/vnat/). Table 1 shows list of 32 *Arabidopsis* core ecotypes

### Measurement of relative water content of the leaves (RWC)

Arabidopsis seeds were germinated on MS agar medium, and then 1-week-old seedlings were transferred to 7 cm pots filled with gravel and nutritive soil (1:1) and cultured at 23/19°C day/night temperature, a day/night cycle of 16/8 h. After growing under well-watered conditions for 4 weeks, the plants were subjected to drought treatment without further watering. Plant leaves were sampled after 7 days of drought treatment. Relative water content of the leaves (RWC) was assayed according to the method of Mata and Lamattina (2001). Three independent experiments were carried out and for each experiment two plants were used from each accession, 12 leaves were sampled per plant.

### Measurement of membrane permeability (MP)

Drought treatment was done according to method of RWC measurement. One gram of *Arabidopsis* leaves was used for analysis of the membrane permeability according to the method of Mata and Lamattina (2001). Mean values were calculated from three independent experiments.

### Measurement of water usage efficiency (WUE)

Drought treatment was done according to method of RWC measurement. WUE was determined with the equation: WUE = net photosynthetic rate / transpiration rate. Net photosynthetic rate and transpiration rate were measured with the LI-6400 Portable Photosynthesis System every 3 days for two times, and for each experiment two plants were selected from each ecotype.

**Table 1.** List of 32 *Arabidopsis* core accessions.

| Number | Versailles number* | Location/country | Latitude | Longitude | Mean monthly number of rain days |
|---|---|---|---|---|---|
| 1 | 163av | Canary Island/Spain | N 28°00' | W 15°30' | 0.67d |
| 2 | 94av | Martuba/cyrenaika(Libya) | N 31°34' | E 22°46' | 3.6d |
| 3 | 76av | Blanes/Spain | N 41°41' | E 2°48' | 4.1d |
| 4 | 178av | Madrid/Spain | N 40°29' | W 3°22' | 4.6d |
| 5 | 157av | Ibel Tazekka/Morocco | N 34°04' | W 4°12' | 5.6d |
| 6 | 162av | Catania/Italy | N 37°30' | E 15°06' | 6.0d |
| 7 | 50av | Palemo/Italy | N 38°07' | E 13°22' | 6.2d |
| 8 | 62av | Stockholm/Sweden | N 59°19' | E 18°03' | 6.2d |
| 9 | 25av | St Jean Cap Ferrat/France | N 43°41' | E 7°20' | 6.3d |
| 10 | 236av | Shakdara river/ Tadjikistan | N 37°29' | E 71°30' | 7.5d |
| 11 | 203av | Tadjikistan/ Tadjikistan | N 38°55' | E 68°47' | 7.6d |
| 12 | 91av | Tsu/Japan | N 34°19' | E 129°19' | 7.8d |
| 13 | 8av | Le Pyla/France | N 46°12' | E 6°10' | 7.9d |
| 14 | 101av | Geneve/Switzerland | N 46°12' | E 6°10' | 8.0d |
| 15 | 42av | Bologna/Italy | N 44°29' | E 11°20' | 8.1d |
| 16 | 186av | Warthe /Poland | N 52°44' | E 15°15' | 10.0d |
| 17 | 234av | Slapy/Cezchoslovakia | N 49°49' | E 14°24' | 11.5d |
| 18 | 200av | Greenville/USA | N 43°11' | E 85°15' | 11.7d |
| 19 | 209av | Kindalville/ USA | N 43°00' | E 85°00' | 11.8d |
| 20 | 180av | Bulhary/Cezchoslovakia | N48°49' | E 16°45' | 12.1d |
| 21 | 252av | Akita pref/Japan | N 39°43' | E 140°06' | 12.7d |
| 22 | 92av | Stobowa/Russia | N 52°57' | E 36°04' | 13.7d |
| 23 | 68av | Tenela/Finland | N 60°04' | E 23°18' | 13.8d |
| 24 | 244av | Wassilewskija/Russia | N 52°13' | E 30°38' | 14.0d |
| 25 | 83av | Edinburgh/UK | N 50°57' | E 3°13' | 14.1d |
| 26 | 70av | Kaunas/Lithuania | N 54°54' | E 23°54' | 14.4d |
| 27 | 257av | Sakata,Yamagata pref/Japan | N 33°55' | E 139°50' | 14.2d |
| 28 | 266av | Konchezero/ Russia | N 60°07' | E 34°01' | 15.8d |
| 29 | 224av | Oystese/Norway | N 60°23' | E 6°13' | 17.2d |
| 30 | 172av | Burren/Eire(Europe) | N 57°07' | E 9°04' | 18.6d |
| 31 | 166av | Cape Verdi Islands/ Cape Verde Islands | N 16°00' | W 24°00' | No data |
| 32 | 215av | Ostpr/Poland | N 53°31' | E 20°12' | No data |

*The number name is from Versailles stock centre (http://dbsgap.versailles.inra.fr/vnat/).

**Measurement of water retain capacity of detached leaves (WRC)**

Drought treatment was done according to method of RWC measurement. Two plants were selected from each ecotype and we used average water retain capacity of detached leaves in 12 h (accumulative water retain capacity/ 12 h). Detached leaves were weighed in the morning as fresh weight (FW). Instant weight (IW) was determined by subjecting detached leaves in the container with constant temperature (20°C) and constant humidity (30%) every one hour until 12 h. Dry weight (DW) was obtained after drying the samples at 80°C for at least 24 h. WRC measurements were calculated according to the formula: WRC (%) = [(FW-IW)/(FW-DW)] × 100. Mean values were calculated from three independent experiments.

**Measurement of ABA content**

Drought treatment was done according to method of RWC measurement. One gram of the tissue was suspended in 15 mL of extraction solution containing 80% methanol, 100 mg/L butylated hydroxytoluene, and 0.5 g/L citric acid monohydrate. The suspension was stirred overnight at 4°C and centrifuged at 1000 *g* for 20 min. The supernatant was transferred to a new tube and frozenly dried under vacuum. The dry residue was dissolved with 100 µL of methanol plus 900 µL of Tris-buffered saline (50 mM Tris, 0.1 mM $MgCl_2 \cdot 6H_2O$, and 0.15 M NaCl, pH 7.8). ABA concentration in the solution then was determined using the Phytodetek ABA immunoassay kit (Idetek, Inc., Sunnyvale, CA).

**Comprehensive drought tolerance analysis**

K-means cluster program of SPSS was used to analyze above five physiological index of drought tolerance (RWC, MP, WUE, WRC and ABA content) of 32 *Arabidopsis* accessions, then clustered with comprehensive drought tolerance.

**Table 2.** Primer pairs of *ATHK*1 gene for SNP analysis.

| Number of primer pairs | Upper primer | Lower primer | Product length (bp) |
|---|---|---|---|
| P1 | 5'-TCCTTCATGGGTTTCTGATTTGT-3' | 5'-CTTGTGAGATTCCTGCGATGTT-3' | 1269 |
| P2 | 5'-TGGGCTCTGTTTGCGAGT-3' | 5'-CTGTTGCGTATTGCTCATTTG-3' | 1274 |
| P3 | 5'-CAGAACTGATAAGGCAACTGGATG-3' | 5'-GCGTTCTTACAAGACTTCTGCAA-3' | 1097 |
| P4 | 5'-GGATGGTGCGAGAACATAAAT-3' | 5'-ACTTCCGTTTCCTCTGTTTCTTA-3' | 1298 |
| P5 | 5'-ACCTCGACAGATTCTCCAACA-3' | 5'-CTGCGAAACTTTGTCTTGGTC-3' | 1285 |

PCR amplification fragment of *ATHK*1 is located between nucleotide positions from 421 to 6001 of GenBank No.AC003952.

**Selection of representative index of comprehensive drought tolerance**

Representative index of comprehensive drought tolerance was selected with R cluster program of SPSS.

**DNA extraction**

The total DNA was isolated using a modified CTAB method (Doyle and Doyle, 1987). 2 g fresh young leaves were harvested from 4 weeks Arabidopsis plants, it was then grind with a mortar and pestle in liquid nitrogen, and later transfer to 50 ml centrifuge tube. Add actived 10 mL CTAB DNA extraction solution whose temperature is 65℃ (2% CTAB, 1.4 M NaCl, 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 0.2% β-mercaptoethanol), incubate at 60°C for 50 min, swirl one times per 5 - 10 min. Add 10 mL chloroform:isoamyl alcohol (24:1 v/v), mix gently but thoroughly. After RNase treatment and a second chloroform/isoamyl alcohol extraction, the DNA was pelleted and dissolved in 200 µL TE (10 mM Tris-HCl and 1 mM EDTA, pH 8.0) to a final concentration of 100 - 200 ng/µL DNA.

**Primer design and testing of *ATHK*1**

*ATHK*1 genome sequence was selected from *Arabidopsis* genome of Columbia (Genebank No. AC003952). Five primer pairs were designed from *ATHK*1 genomic sequence using Oligo 6.0 software. Two PCR products are overlapping (Table 2).

**PCR amplification**

PCR amplification fragment of *ATHK*1 is located between nucleotide position from 421 to 6001 of GenBank No.AC003952. PCR amplification of *ATHK*1 was performed in a 50 µL volume with PTC-100 thermocycler. Fifty microlitre reaction solution contained 1 µL genomic DNA (20ng/µL), 5 µL 10× PCR buffer, 4 µL 2.5 mM dNTP, 2 µL 10 mM upper primer, 2 µL 10 mM lower primer, 0.4 µL Taq enzyme (5U/µL), double distilled water 35.6 µL. The thermocycler was programmed as follows: one cycle at 94℃ for 4 min, followed by 30 cycles of 94℃ for 1 min, 53℃ for 1 min, and 72℃ for 2 min, then one cycle at 72℃ for 7 min.

**DNA sequencing**

PCR products were purified with gel extract kit of Omega and sequenced directly on CEQ2000 DNA analysis system (Beckman). Sequencing reaction was performed in 10 µL volume using the Beckman CEQ2000XL sequencing Kit with following PCR condition: 30 cycles of 96℃ for 20 s, 50℃ for 1min, and 60℃ for 4 min. Any ambiguous sequences were resolved by repeated sequencing of the PCR product.

**Sequencing analysis**

Sequence contigs and sequence alignment were analyzed with Vector NTI Suite 7.0 and DNAStar softwares. The analyzed region is located between nucleotide position from 724 to 5840 (Genebank No. AC003952). Each SNP was confirmed by repeated sequencing of the PCR products from both ends. Sequence polymorphisms, Tajima'S D and Fu-Li'S D* test were analyzed using the DnaSP 4.0 (Rozas and Rozas, 1999), and nucleotide variation was estimated as nucleotide diversity ($\pi$(pi), $\theta$), and 4N$\mu$($\theta$).

**Association analysis**

Phylogenetic reconstruction was carried out with the MEGA2 software. ANOVA analysis was carried out with the SPSS software package. To evaluate the statistical significance of differences in comprehensive drought tolerance among *A. thaliana* accessions, we used representative physiological index of comprehensive drought tolerance per accession for a one-way ANOVA analysis. To account for the correlation of molecular variation, we tested for a significant association between ATHK1 amino acid sequence variation and representative index variation of comprehensive drought tolerance by analyzing significant differentiation between haplotypes with one-way ANOVA analysis (Tables 7 - 8 and Figure 2). To determine the significance level for representative index change of comprehensive drought tolerance relative to reference accession Columbia, we evaluated the significance of the interaction between haplotypes ( haplotype including Columbia vs. the focal haplotype) using ANOVA analysis.

## RESULTS

### Comprehensive drought tolerance analysis

Drought tolerance is a complex response to soil water deficit, high temperature, low temperature and salt stress. Drought tolerance may include a range of morphological, physiological and biochemical adaptations that enable plant to survive period of unfavourable growth conditions. Plant drought tolerance is a quantitative trait affected by many factors, most of which are minor factors. So, it is difficult to evaluate plant comprehensive drought tolerance with only one index. There is no comprehensive standard system for evaluating drought tolerance. Many researchers have suggested appraising plant drought

**Table 3.** Primal data for assessment of comprehensive drought resistance.

| Versailles number | RWC (%) | MP (%) | WUE | ABA (ng/g FW) | WRC (%) |
|---|---|---|---|---|---|
| 8av | 86.90±0.03 | 4.700±0.132 | 1.997±0.776 | 153.996±1.810 | 4.887±0.517 |
| 25av | 91.00±0.08 | 4.380±0.321 | 3.610±0.564 | 132.377±5.660 | 4.798±1.169 |
| 42av | 81.20±0.02 | 5.300±0.381 | 1.973±1.082 | 102.290±6.210 | 5.441±0.435 |
| 50av | 83.12±0.02 | 8.150±0.231 | 2.664±0.476 | 137.988±6.660 | 4.348±0.493 |
| 62av | 78.60±0.02 | 7.230±0.424 | 2.171±0.726 | 68.865±4.780 | 5.018±0.675 |
| 68av | 84.05±0.02 | 3.900±0.465 | 2.565±0.283 | 227.000±20.88 | 4.674±0.911 |
| 70av | 86.80±0.03 | 6.200±0.196 | 3.543±0.881 | 474.322±16.84 | 3.434±0.765 |
| 76av | 83.60±0.08 | 3.150±1.016 | 2.404±0.669 | 119.552±11.58 | 5.490±0.432 |
| 83av | 84.20±0.02 | 10.70±0.123 | 3.531±0.215 | 148.882±2.070 | 5.368±0.276 |
| 91av | 80.39±0.05 | 6.900±4.132 | 2.078±0.171 | 128.759±5.340 | 5.649±0.245 |
| 92av | 84.97±0.03 | 3.750±0.324 | 2.345±0.267 | 103.002±5.150 | 5.708±0.632 |
| 94av | 84.09±0.07 | 5.430±2.336 | 2.510±0.521 | 139.065±6.010 | 4.570±0.321 |
| 101av | 85.34±0.02 | 3.780±0.445 | 2.340±0.509 | 156.255±11.29 | 2.859±0.616 |
| 157av | 95.34±0.01 | 7.000±0.432 | 3.670±0.737 | 454.890±19.94 | 3.424±0.178 |
| 162av | 82.18±0.02 | 7.900±0.298 | 0.567±0.317 | 327.401±4.020 | 5.889±0.212 |
| 163av | 86.53±0.03 | 3.650±0.728 | 2.671±0.522 | 649.657±33.19 | 3.958±0.342 |
| 166av | 85.25±0.04 | 4.890±0.141 | 2.774±0.399 | 167.016±1.940 | 5.373±0.278 |
| 172av | 82.63±0.03 | 18.78±0.876 | 2.982±1.344 | 147.396±23.12 | 5.889±0.341 |
| 178av | 78.24±0.03 | 2.350±0.289 | 3.326±0.194 | 438.490±10.67 | 3.565±0.287 |
| 180av | 90.00±0.03 | 9.500±0.212 | 2.500±0.445 | 63.599±1.520 | 4.661±0.376 |
| 186av | 82.80±0.02 | 5.550±0.201 | 0.675±0.272 | 109.597±2.660 | 3.008±0.276 |
| 200av | 75.44±0.01 | 4.560±0.090 | 2.530±0.487 | 59.730±6.640 | 5.405±0.024 |
| 203av | 91.21±0.04 | 3.000±0.112 | 2.767±0.090 | 277.734±13.67 | 3.127±0.267 |
| 209av | 88.46±0.07 | 3.890±0.278 | 1.845±0.987 | 257.630±4.990 | 3.150±0.222 |
| 215av | 87.16±0.08 | 6.500±0.265 | 1.832±1.160 | 117.721±1.410 | 3.152±0.510 |
| 224av | 81.53±0.03 | 2.150±0.325 | 1.007±0.348 | 42.240±2.030 | 5.714±0.380 |
| 234av | 78.43±0.03 | 4.900±0.178 | 2.191±0.312 | 80.789±1.050 | 4.283±0.643 |
| 236av | 79.42±0.02 | 5.000±0.367 | 3.335±0.336 | 40.476±1.330 | 4.601±0.189 |
| 244av | 87.85±0.02 | 4.200±0.078 | 2.780±0.366 | 103.561±3.360 | 4.340±0.453 |
| 252av | 85.43±0.03 | 12.310±0.116 | 1.757±0.906 | 153.996±1.810 | 6.076±0.564 |
| 257av | 76.24±0.10 | 11.250±0.378 | 0.488±0.556 | 89.138±7.750 | 6.073±0.648 |
| 266av | 82.71±0.87 | 15.430±0.110 | 2.132±0.171 | 103.561±3.360 | 5.717±1.166 |

RWC, Relative water content of leaves; MP, Membrane permeability of leaves; WUE, Water usage efficiency; WRC, Water retain capacity of detached leaves.

tolerance with comprehensive method for more physiological indexes. We scored the drought tolerance with five physiological indexes (RWC, MP, WUE, WRC and ABA content) of 32 *Arabidopsis* accessions (Table 3) with K-means cluster program of SPSS, and then clustered them into four groups: most tolerant groups, tolerant group, middle tolerant group and drought sensitive group:

Most tolerant group: 163av; Tolerant group: 203av, 209av, 70av, 162av, 178av, 157av.
Middle tolerant group: 8av, 166av, 215av, 172av, 68av, 50av, 94av, 91av, 83av, 25av, 101av.
Sensitive to drought group: 252av, 244av, 236av, 224av, 92av, 62av, 186av, 234av, 180av, 266av, 76av, 42av, 200av, 257av.

In order to carry out association analysis between *ATHK*1 sequence variation and representative index of comprehensive drought tolerance variation, representative physiological index were selected by R cluster program of SPSS (Figure 1). The result showed that MP and WRC were clustered into one group; other three indexes (RWC, WUE and ABA content) were clustered into another group.

We calculated the relative exponent $R^2$ among values (X1, X3 and X4) in group 2, and then selected maximum value ($R_{X1}^2$=0.132) RWC as one representative physiological index. Comprehensive drought tolerance could be evaluated by RWC combining anyone of MP and WRC.

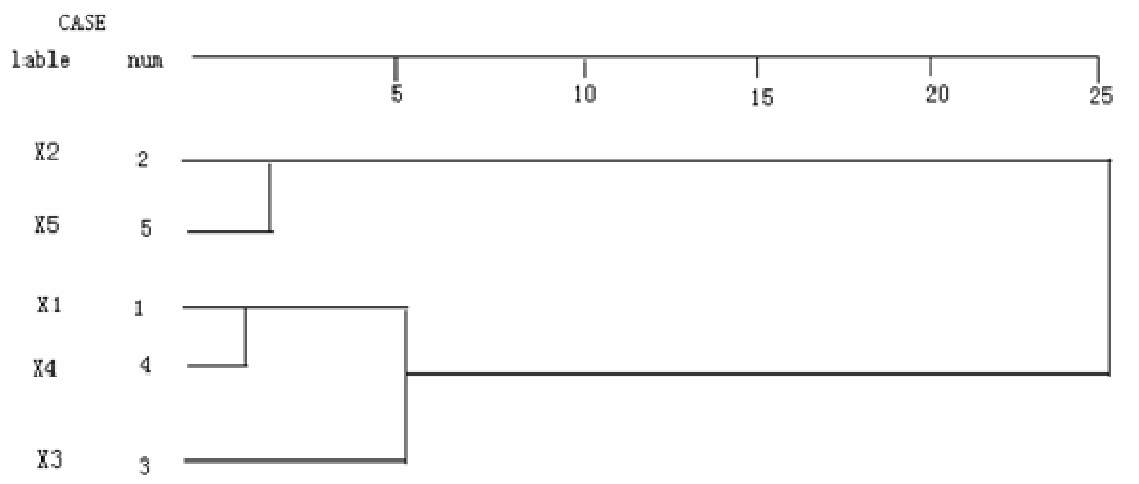


**Figure 1.** Cluster analysis of representative physiological index of comprehensive drought tolerance. X1, relative water content of leaves (RWC); X2, membrane permeability of leaves (MP); X3, water usage efficiency (WUE); X4, ABA content; and X5, water retain capacity of detached leaves (WRC).

**Table 4.** Patterns of nucleotide variation in the *ATHK*1 region of *Arabidopsis thaliana*.

| Region | Size (bp) | SNP | Indel | π | θ | Transition/Transversion |
|---|---|---|---|---|---|---|
| Whole region | 5117 | 36 | 3 | 0.00086 | 0.0016 | 1.57 |
| Coding region | 3624 | 19 | 0 | 0.00056 | 0.00137 | 1.71 |
| Non-coding region | 1493 | 17 | 3 | 0.00184 | 0.00221 | 1.42 |
| 5'- flanking | 131 | 2 | 0 | 0.0048 | 0.00379 | 1.00 |
| 3' - flanking | 105 | 0 | 0 | 0 | 0 | 0 |
| Introns | 1257 | 15 | 3 | 0.00155 | 0.00206 | 1.5 |

The analyzed region is located between nucleotide position from 724 to 5840 (Genebank No. AC003952). Sequence polymorphisms were analyzed using the DnaSP 4.0 (Rozas and Rozas, 1999) and nucleotide variation was estimated as nucleotide diversity (π, θ).

## DNA polymorphisms at the ATHK1 locus

At *ATHK1* locus, the nucleotide polymorphism (including single nucleotide polymorphism (SNP) and insertion /deletion (Indel)) frequency was 1 polymorphism per 131.2bp, on average 1 SNP per 142.1bp and 1 Indel per 1705.7bp (Table 4).

Nineteen SNPs and no Indel were found in the coding region, SNP frequency was 1 SNP per 190.7bp, and they formed 11 clades according to amino acid variation (Table 4 and Figure 2). Seventeen SNPs and 3 Indels were discovered in non-coding region, on average 1 SNP per 87.8bp and 1 Indel per 497.7 bp. Among them, two SNP were discovered in the 5'– flanking sequence, on average 1 SNP per 65.5 bp; fifteen SNPs and 3 Indels were discovered in introns, on average 1 SNP per 83.8 bp and 1 Indel per 419bp. Nucleotide polymorphism in non-coding region was 1.55 times higher than that in coding region. The nucleotide diversity (π)for the whole region was 0.00086 (Table 4), which was lower than that of previously studied genes ChiA, ChiB and FRI of *Arabidopsis* (Kawabe et al., 1997, 1999; Le Corre et al., 2002). For ChiA (Kawabe et al., 1997), ChiB (Kawabe and Miyashita, 1999) and FRI (Le Corre et al., 2002), nucleotide diversity (π) was 0.00214, 0.01040 and 0.0091 respectively. Similar to previously studied genes, nucleotide diversity was significantly lower in coding region than that in non-coding region. Table 4 shows patterns of nucleotide variation in the *ATHK*1 region of *Arabidopsis thaliana.*

Nineteen SNPs were discovered at the *ATHK*1 coding region. Five synonymous and fourteen nonsynonymous mutations were detected in this region. The ratio of average number of nucleotide difference $\pi_n / \pi_s$ was 0.727 (Table 5), higher than that of those previously studied genes *CAL, AP3* and *PI* (Le Corre et al., 2002).

This result indicates that the *ATHK1* protein is not constrained against amino acid changes within the species, and this gene belonged to the middle evolution rate gene. $\pi_n / \pi_s$ ratio in exon 1 is 5.36, this high value indicates that evolution rate in this region is very quick. Table 5 show analysis of nucleotide diversity in *ATHK*1 coding region of *A. thaliana.*

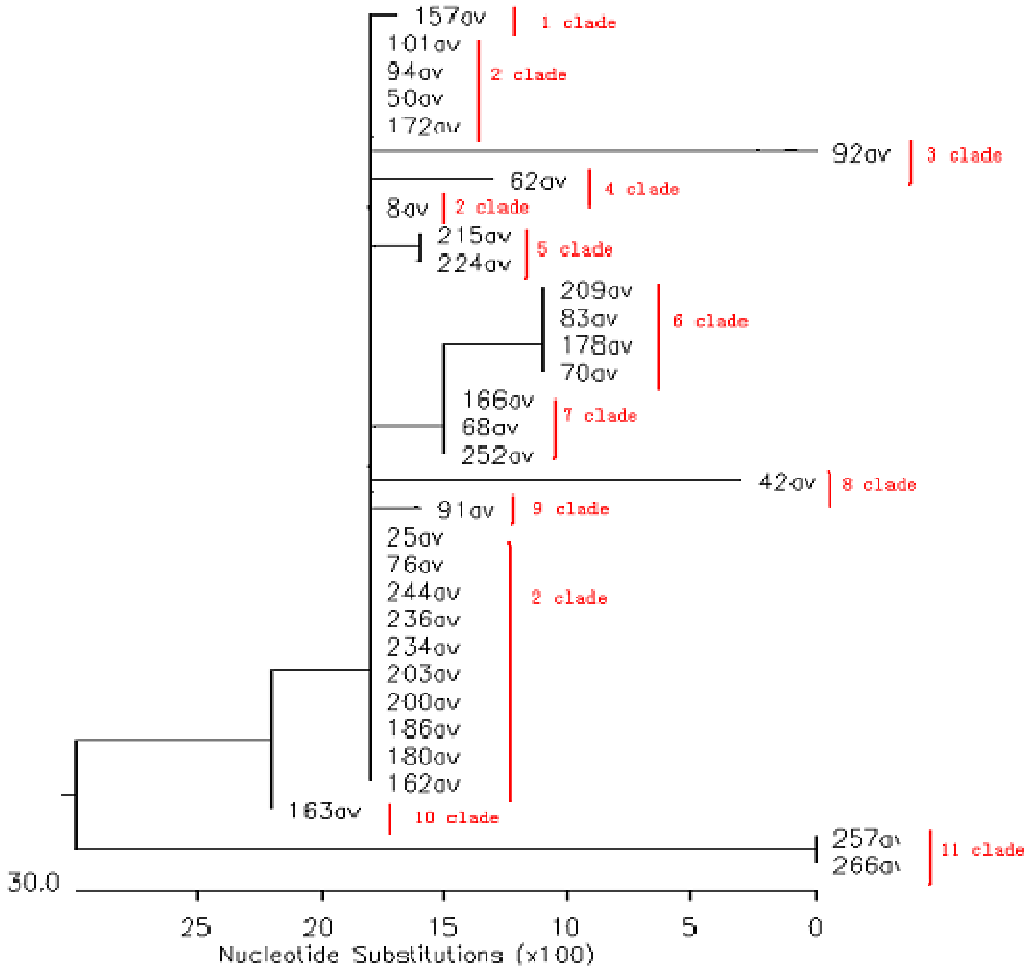


**Figure 2.** Neighbor-joining trees representing *ATHK*1 amino acid clade relationships.

**Table 5.** Analysis of nucleotide diversity in *ATHK*1 coding region of *Arabidopsis thaliana.*

| Region | Length | Synonymous | | | Nonsynonymous | | | $\pi_n / \pi_s$ |
|---|---|---|---|---|---|---|---|---|
| | | N | S | $\pi_s$ | N | S | $\pi_n$ | |
| All coding regions | 3624 | 823 | 5 | 0.00071 | 2797 | 14 | 0.00052 | 0.727 |
| Exon 1 | 259 | | 1 | 0.00125 | | 2 | 0.0067 | 5.36 |
| Exon 3 | 777 | | 0 | 0 | | 1 | 0.00125 | - |
| Exon 4 | 245 | | 1 | 0.0054 | | 0 | 0 | 0 |
| Exon 6 | 125 | | 2 | 0.00125 | | 0 | 0 | 0 |
| Exon 9 | 70 | | 1 | 0.00125 | | 1 | 0.00125 | 1 |
| Exon 12 | 1042 | | 0 | 0 | | 7 | 0.0016 | - |
| Exon 13 | 234 | | 0 | 0 | | 3 | 0.0020 | - |

N = Number of sites; S = Number of polymorphic sites.

## The relationship between polymorphisms in the ATHK1 coding region and comprehensive drought tolerance

To find the correlation of genotypic and phenotypic variation, we tested for a significant association between *ATHK*1 sequence variation and representative index of comprehensive drought tolerance variation by analyzing significant differentiation between haplotypes with ANOVA analysis (Table 6 and Figure 2). In the coding region of *ATHK*1, the sequencing result showed that eleven different clades were detected in 32 accessions (Figure 2).

**Table 6A.** ANOVA analysis on 5 physiological indexes based on 11 clades of *ATHK*1 protein.

| Parameter | Relative water content of leaves (RWC) | | | Membrane permeability of leaves (MP) | | | Water usage efficiency (WUE) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Between groups | Within groups | Total | Between groups | Within groups | Total | Between groups | Within groups | Total |
| Degree of freedom | 10 | 80 | 90 | 10 | 51 | 61 | 10 | 90 | 100 |
| Sum of squares | 5.606E-02 | 0.188 | 0.244 | 2.521E-02 | 6.747E-02 | 9.268E-02 | 33.252 | 51.818 | 85.070 |
| Mean squares | 5.606E-03 | 2.350E-03 | | 2.521E-03 | 1.323E-03 | | 3.325 | 0.576 | |
| *F* value | 2.386 | | | 1.906 | | | 5.775 | | |
| Significance | 0.016 | | | 0.066 | | | 0.000 | | |

**Table 6B.** ANOVA analysis on 5 physiological indexes based on 11 clades of *ATHK*1 protein.

| Parameter | ABA content | | | Water retain capacity of detached leaves (WRC) | | |
|---|---|---|---|---|---|---|
| | Between groups | Within groups | Total | Between groups | Within groups | Total |
| Degree of freedom | 10 | 85 | 95 | 10 | 76 | 86 |
| Sum of squares | 1425061.649 | 491373.212 | 1916434.861 | 31.632 | 74.494 | 106.127 |
| Mean squares | 142506.165 | 5780.861 | | 3.163 | 0.980 | |
| *F* value | 24.651 | | | 3.227 | | |
| Significance | 0.000 | | | 0.002 | | |

**Table 7.** Multiple comparison of ANOVA analysis on relative water content of leaves (RWC) based on 11 clades of ATHK1 protein. Multiple Comparisons, Dependable Variable: X, LSD

| (I)Y | (J)Y | Mean Difference (I-J) | Standard Error | Significance | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 2.00 | 1.00 | -.1033 | 2.895E-02 | .001* | -.1609 | -4.5657E-02 |
| | 3.00 | -5.1775E-03 | 2.895E-02 | .858 | -6.2782E-02 | 5.243E-02 |
| | 4.00 | 6.267E-02 | 2.895E-02 | .033* | 5.066E-03 | .1203 |
| | 5.00 | 1.040E-03 | 2.112E-02 | .961 | -4.1000E-02 | 4.308E-02 |
| | 6.00 | -1.0641E-03 | 1.683E-02 | .948 | -3.3658E-02 | 3.153E-02 |
| | 7.00 | -5.7670E-03 | 1.886E-02 | .758 | -4.2910E-02 | 3.138E-02 |
| | 8.00 | 3.266E-02 | 2.895E-02 | .263 | -2.4944E-02 | 9.026E-02 |
| | 9.00 | 4.051E-02 | 2.895E-02 | .165 | -1.7091E-02 | 9.812E-02 |
| | 10.00 | -1.6021E-02 | 2.895E-02 | .581 | -7.3625E-02 | 4.158E-02 |
| | 11.00 | 2.513E-02 | 2.290E-02 | .047* | -2.0448E-02 | 7.071E-02 |

*: The mean difference is significant at 0.05 level.

Except MP, the differentiation of other four physiological indexes between clades is highly significant (Table 6). Multiple comparisons indicated that only the 1199 amino acid site (Ser↔stop resulted from 4984th G↔C) variation of the eleventh clade (257av and 266av) is associated with not only RWC but also WRC (Table 7 and 8). This means that the change of Ser↔stop is associated with comprehensive drought tolerance of 257 and 266 av. Table 6 shows ANOVA analysis on 5 physiological indexes based on 11 clades of *ATHK*1 protein.

Table 7 shows multiple comparison of ANOVA analysis on relative water content of leaves (RWC) based on 11 clades of ATHK1 protein. Table 8 shows multiple comparison of ANOVA analysis on water retain capacity of detached leaves (WRC) based on 11 clades of ATHK1 protein. Figure 2 shows neighbor-joining trees representing *ATHK*1 amino acid clade relationships

## DISCUSSION

In the *ATHK*1 gene region, the polymorphism (SNP and

**Table 8.** Multiple comparison of ANOVA analysis on water retain capacity of detached leaves (WRC) based on 11 clades of ATHK1 protein. Multiple Comparisons, Dependable Variable: X, LSD.

| (I)Y | (J)Y | Mean Difference (I-J) | Standard Error | Significance | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 2.00 | 1.00 | 1.2410 | .7174 | .088 | -.1877 | 2.6697 |
| | 3.00 | -0.9073 | .5927 | .130 | -2.0877 | .2730 |
| | 4.00 | -.7707 | .5927 | .197 | -1.9510 | .4097 |
| | 5.00 | .1060 | .4334 | .807 | -.7573 | .9693 |
| | 6.00 | .5450 | .3500 | .124 | -.1522 | 1.2422 |
| | 7.00 | -.9215 | .3834 | .019* | -1.6852 | -.1578 |
| | 8.00 | -.6840 | .5927 | .252 | -1.8644 | .4964 |
| | 9.00 | -1.1040 | .5927 | .066 | -2.2844 | 7.637E-02 |
| | 10.00 | .9527 | .5927 | .112 | -.2277 | 2.1330 |
| | 11.00 | -1.2123 | .4334 | .007* | -2.0756 | -.3491 |

Indels) frequency was 1 SNP per 131.16 bp. Nucleotide polymorphism in non-coding region was 1.55 times higher than that in coding region. This distribution was similar to that observed in human (Wang et al., 1998; Bensen et al., 2003), maize (Ching et al., 2002; Tenaillon et al., 2001), rice (Nasu et al., 2002) and soybean (Zhu et al., 2003; Voryell et al., 1999). One explanation for this lay in the possible heterosis conferred by great selection pressure on coding region.

In *ATHK*1 coding region, high density of single nucleotide polymorphism (SNP) was discovered, SNP frequency was 1 SNP per 190.7 bp. Among them, five synonymous and fourteen nonsynonymous mutations were detected in this region. The ratio of average number of nucleotide difference $\pi_n / \pi_s$ was 0.727, indicating that the ATHK1 protein is not constrained against amino acid changes within species and positive selection affected these region intensively (Table 5), and this gene was undergoing adaptive evolution.

The major goal of this study is to link phenotypic and genotypic variation to learn more about the function of the *ATHK*1. *ATHK*1 exhibits marked sequence variability in association with the representative index of comprehensive drought tolerance. The strongest evidence for association analysis comes from the amino acid Ser↔stop variation of the eleventh clade (257 and 266av). This polymorphism was discovered near the reaction regulation region of histone protein kinase (ATHK1). The change of Ser↔stop may affect the osmosis signal output function of ATHK1. It is noteworthy that the branches of 257 and 266av, the two accessions grow in relative moist environment. The result of this amino acid change may be their sensitivity to dry climate. 257 and 266 av grow respectively in Sakata Yamagata pref (Japan) and Konchezero (Russia). The mean monthly number of rain days all year in these two places is respective 14.2 and 15.8 days (Table 1) (http://dbsgap.versailles.inra.fr/vnat/). It is highly significant than that of reference population 186av, which is 10 days. This case is likely to be the evidence of adaptive evolution.

## ACKNOWLEDGMENTS

This study is financially supported by grants from China National Key Basic Research Program (No.2003CB114302), National Natural Science Foundation of China (No.3040228), Foundation of Beijing Municipal Committee of Science and Technology (No.H012010240240113) and the Beijing Agricultural Innovative Platform - Beijing Natural Science Fund Program (YZPT02-06).

### REFERENCES

Bensen JT, Langefeld CD, Hawkins GA, Green LE, Mychaleckyj JC, Brewer CS, Kiger DS, Binford SM, Colicigno CJ, Allred DC, Freedman BI, Bowden DW (2003). Nucleotide variation, haplotype structure, and association with end-stage renal disease of the human interleukin-1 gene cluster. Genomics 82(2): 194-217.

Bohnert HJ, Sheveleva E (1998). Plant stress adaptations—making metabolism move. Curr. Opin. Plant Biol. 1: 267-274.

Boyer JS (1982). Plant productivity and environment, Science 218: 444-448.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3: 1-14.

Doyle JJ, Doyle JL (1987). Isolation of DNA from fresh plant tissue. Focus 12: 13-15.

Flint-Garcia SA (2003). Structure of linkage disequilibrium in plants, Annu. Rev. Plant. Biol. 54: 357-374.

Gutterson N, Zhang JZ (2004). Genomics applications to biotech traits: a revolution in progress? Curr. Opin. Plant Biol. 7(2): 226-230.

Hao GP, Wu ZY, Chen MS, Cao MQ, Dominique B, Georges P, Huang CL, Yang Q (2004). Single nucleotide polymorphisms of CBF4 locus region of Arabidopsis thaliana correspond to drought tolerance. Chin. J. Agric. Biotechnol. 1(3): 181-190.

Kawabe A, Innan H, Terauchi R, Miyashita NT (1997). Nucleotide polymorphism in the acidic chitinase locus (ChiA) region of the wild plant Arabidopsis thaliana. Mol. Biol. Evol. 14: 1303-1315.

Kawabe A, Miyashita TN (1999). DNA Variation in the Basic Chitinase Locus (ChiB) Region of the Wild Plant Arabidopsis thaliana. Genetics 153: 1445-1453.

Lanceras JC, Pantuwan G, Jongdee B, Toojinda T (2004). Quantitative trait loci associated with drought tolerance at reproductive stage in rice. Plant Physiol. 135(1): 384-399.

Le Corre V, Roux F, Reboud X (2002). DNA polymorphism at the FRI-

GIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol. Biol. Evol. 19: 1261-1271.
Long AD, Langley CH (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. 9: 720-731.
Mata CG, Lamattina L (2001). Nitric oxide induces stomatal closure and enhances the adaptive plant responses against drought stress. Plant Physiol. 126, 1196-1204.
Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002). Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (Oryza sativa, Oryza rufipogon) and establishment of SNP markers. DNA Res. 9: 163-171.
Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases, Science 273: 1516-1517.
Rozas J, Rozas R (1999). DnaSP version 3: an intergrated program for molecular population genetics and molecular population analysis. Bioinformatics 15: 174-175.
Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize ( *Zea mays* ssp. *Mays* L.). Proc Natl Acad Sci USA. 98: 9161-9166.
Urao T, Yakubov B, Satoh R, Yamaguchi-Shinozaki K, Seki M, Hirayama T, Shinozaki K (1999). A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. Plant Cell 11: 1743-1754.
Voryell VH, Jessen H, Schup JM, Webb D, Keim P (1999). Allele-specificx hybridization markers for soybean. Theor. Appl. Genet. 101: 1291-1298.
Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366): 1077-1082.
Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003). Single-nucleotide polymorphisms in soybean. Genetics 163: 1123-1134.