Heredity (2007) 99, 423–431

www.nature.com/hdy

ORIGINAL ARTICLE

Genetic diversity of the natural populations of *Arabidopsis thaliana* in China

F He[1], D Kang[1], Y Ren[1], L-J Qu[1,2], Y Zhen[1] and H Gu[1,2]
[1]*Peking-Yale Joint Center for Plant Molecular Genetics and AgroBiotechnology, The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing, People's Republic of China and* [2]*The National Plant Gene Research Center (Beijing), Beijing, People's Republic of China*

Although extensive studies have been conducted on the genetic structure of *Arabidopsis thaliana* (*A. thaliana*) populations worldwide, the populations from China have never been studied. In this study, we collected 560 individuals from 19 natural populations of *A. thaliana* distributed in East China along the lower reaches of the Yangtze River, and two populations from northwest China (Xinjiang Province). We adopted two kinds of molecular marker, inter-simple sequence repeats (ISSRs) and random amplified polymorphic DNA (RAPDs) to investigate the genetic diversity within and among populations, and the correlation between the genetic and geographic distances. Thirteen ISSR primers produced 165 polymorphic bands (PPB) (96%) and 11 RAPD primers produced 162 polymorphic bands (98%) in about 560 individuals. The two marker systems generated similar patterns of genetic diversity in these natural populations. The AMOVA analysis indicated about 42–45% of the total genetic variation existed within populations, and found possible geographic structure. The Mantel test revealed a significant correlation between the geographic distance and the genetic distance of these populations in general. A close genetic relationship was found among four populations in the Jiangxi Province, and these always appeared clustered together as a monophyletic group in unweighted pair-group method with arithmetic averages dendrograms based on both ISSR and RAPD data sets. Based on the observation of recolonization and extinction of naturally distributed populations of *A. thaliana*, and the pattern of their genetic differentiation, the distribution of this species in China might be a result of natural dispersal under the strong influence of human activity.
Heredity (2007) **99,** 423–431; doi:10.1038/sj.hdy.6801020; published online 27 June 2007

Keywords: *Arabidopsis thaliana*; natural populations; China; RAPD; ISSR; genetic diversity

Introduction

Arabidopsis thaliana (L.) Heynh is an annual, weedy and mostly autogamous species that is native to Europe and central Asia and now naturalized worldwide (Al-Shehbaz and O'Kane, 2002). *Arabidopsis thaliana* (*A. thaliana*) has been adopted as a model organism for establishing an in-depth understanding of plant biology. The use of this species as a model organism has become increasingly important since the elucidation of its genome sequence (The Arabidopsis Genome Initiative, 2000). In addition to a few well-known ecotypes such as Colombia and Landsberg, which have been cultivated in laboratory conditions for plant physiological, developmental, molecular and functional genomic studies, more and more naturally distributed populations (or ecotypes) have been collected and characterized for their genetic variation (King *et al*., 1993; Innan *et al*., 1997; Kuittinen *et al*., 1997; Ullrich *et al*., 1997; Bergelson *et al*., 1998; Loridon *et al*., 1998; Vos, 1998; Breyne *et al*., 1999; Miyashita *et al*., 1999; Erschadi *et al*., 2000; Sharbel *et al*., 2000; Barth *et al*., 2002; Hoffmann *et al*., 2003; Jorgensen and Mauricio, 2004; Nordborg *et al*., 2005; Stenoien *et al*., 2005; Schmid *et al*., 2006). Understanding the amount and distribution of genetic variation between and among populations is not only crucial for ecological and evolutionary studies, but also serves as a base for functional genomic studies. More than 300 accessions collected from different locations around the world are available through *Arabidopsis* stock centers, such as the Nottingham *Arabidopsis* Stock Center (NASC) and the *Arabidopsis* Biological Resource Center (ABRC), which have become an important resource for the Arabidopsis-research community. The analysis of variants found in nature has proven to be very successful in gaining insight into the control of important processes in plants (Koornneef *et al*., 2004; Mitchell-Olds and Schmitt, 2006).

Although *A. thaliana* is widely distributed in China, as recorded in the Chinese flora (Cheo *et al*., 1987, 2001), no accession from China is available in any of the international *Arabidopsis* stock centers. In China, the herbarium information on this species is fragmental, and the genetic background of the naturally distributed *A. thaliana* is largely unknown. Therefore, it is important to examine the genetic differentiation among the populations in natural habitats in China to construct a broader picture of the genetic diversity of this species around the world. In this study, we collected plant samples from 21 populations of *A. thaliana* from three provinces in East China and one province in Northwest China, and used

Correspondence: Professor H Gu, College of Life Sciences, Peking University, Beijing 100871, People's Republic of China.
E-mail: guhy@pku.edu.cn
Received 31 January 2007; revised 27 April 2007; accepted 28 May 2007; published online 27 June 2007

random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) markers to investigate the genetic diversity in these populations. The correlation of the geographic distance and genetic distance among these populations was also analyzed.

Materials and methods

Plant materials

All naturally distributed populations used in this study were collected by the authors. Nineteen populations were collected from three provinces, Anhui, Jiangxi and Zhejiang in East China, and two populations from Xinjiang Province in Northwest China (Figure 1). The Columbia (Col) ecotype was used for comparison. The locations and habitat descriptions for the naturally distributed populations are listed in Table 1, and each population is denoted by five letters: the first two uppercase letters representing the province followed by three lowercase letters representing the city or county where the population was collected. For example, AHngs represents the population collected in Ningguo City of Anhui Province. For most populations, at least 30 individual plants with matured seeds per population were collected. The whole plants were collected and stored individually in plastic bags with silica gel for immediate drying. However, for some small populations such as the two populations obtained from Xinjiang Province, which had less than 30 individuals per population, only 14–20 individuals were collected. One voucher specimen was collected for each population. Each individual plant was numbered and propagated in a growth room (16 h light, 8 h dark at $21 \pm 2^{\circ}C$).

DNA isolation

The total DNA was isolated from the rosette leaves of the silica-dried materials (for ISSR analysis) or from the plants propagated in a greenhouse (for RAPD analysis). For the fresh materials, the CTAB method was adopted (Qu *et al*., 2000); while for the silica-dried materials, the plant materials were pretreated with 1.5% PVP and proteinase K, and then treated by the CTAB method. For each population (including Col ecotype), the DNA was isolated from 14 to 30 individuals separately.

ISSR and RAPD analysis

Twelve ISSR primers were chosen from those primers previously used by Barth *et al*. (2002), while another five were designed based on the *Arabidopsis* genome sequence (The Arabidopsis Genome Initiative, 2000). In total, 17 primers were used for a two-step primer screening process. For the first step, the total DNA obtained from Col ecotype was used to identify the primers that produced clear and reproducible bands. For the second step, six individuals from each population were randomly selected for RAPD analysis with primers selected in the first step. Following the two-step primer screening, 13 primers with clear and reproducible bands were chosen for the amplification reactions for all individuals in the present study (Table 2). PCR amplification was carried out in 25 μl with 20 mM Tris–HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1% Triton X-100 (pH 8.8), 2 mM $MgCl_2$, 200 μM each of dATP, dGTP, dCTP and dTTP, 0.4 pmol of primer, 0.8 U *Taq* DNA polymerase (Dingguo Biotech, Beijing, China) and 20 ng template DNA. Thermal cycling (Thermocycler PTC-225, MJ Research, Watertown) started with 5 min at 94°C, and 45 cycles of 1 min at 94°C, 45 s at 50°C and 2 min at 72°C and ended by an extension for 7 min at 72°C. The amplification products were separated on 1.4% agarose gels in TAE buffer for about 4 h (5–7 V/cm), stained with ethidium bromide and documented using an AmpGene Gel 100 system. The 100 bp ladder marker (Promega (Beijing) Biotech Co. Ltd, China) and λDNA/*Eco*RI + *Hin*dIII marker were used as size standards.

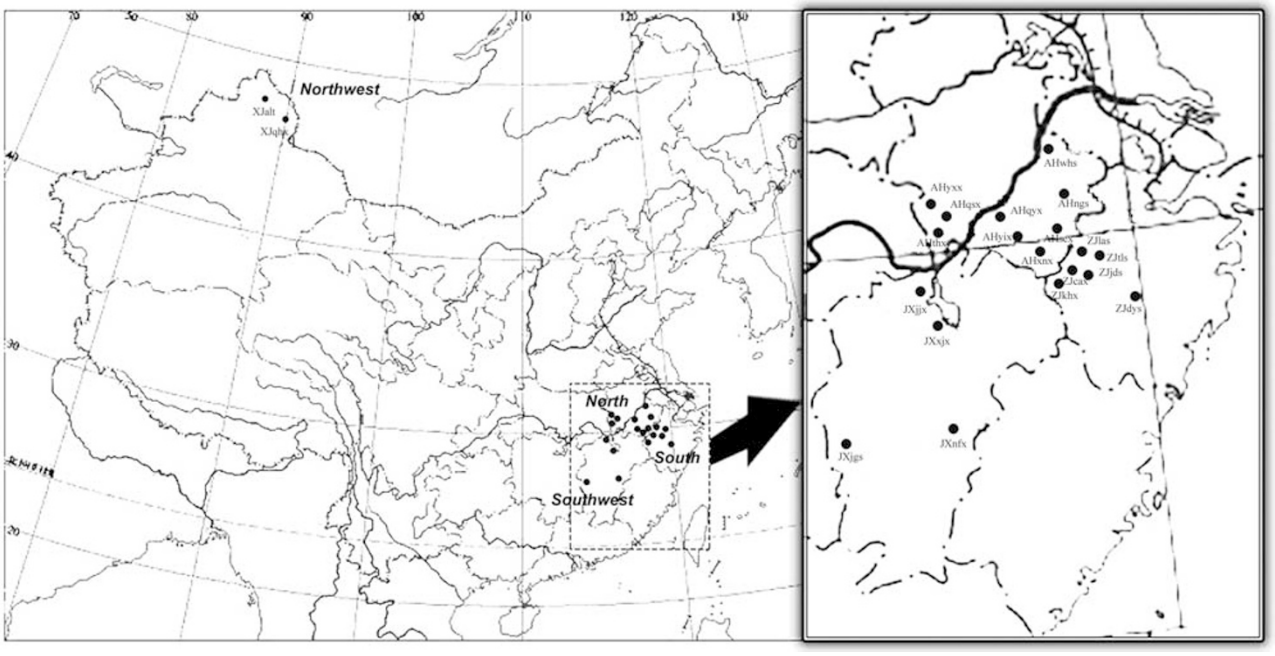

Figure 1 Distribution map of the populations of *Arabidopsis thaliana* used in this study.

Table 1 Naturally distributed populations collected in China

Group	*Location*	*Latitude and longitude*	*Altitude (m)*	*Habitat*
South	Anhui, ningguoshi (AHngs)	30°41.08 N 118°58.36E	160	River side, naked area
	Anhui, qingyangxian (AHqyx)	30°37.15 N 117°45.39 E	70	River side, field border
	Anhui, shexian (AHsex)	29°51.02 N 118°22.75 E	200	Railway bed
	Anhui, wuhushi (AHwhs)	31°23.13 N 118°22.74 E	20	Field border near railway
	Anhui, xiuningxian (AHxnx)	29°47.78 N 118°10.50 E	310	Railway slope
	Anhui, yixian (AHyix)	29°57.48 N 117°57.32 E	420	Farmland
	Zhejiang, chun'anxian (ZJcax)	29°48.28 N 118°50.97 E	100	Near the rivulet, road side
	Zhejiang, dongyangshi (ZJdys)	29°05.02 N 120°25.65 E	290	Field border
	Zhejiang, jiandeshi (ZJjds)	29°32.13 N 119°29.61 E	100	Bank near a river
	Zhejiang, kaihuaxian (ZJkhx)	29°06.50 N 118°23.03 E	165	Road side near the river
	Zhejiang, lin'anshi (ZJlas)	30°03.16 N 119°54.88 E	420	Near the rivulet
	Zhejiang tonglushi (ZJtls)	29°41.39 N 119°39.90 E	210	Road side near the river
Southwest	Jiangxi, jingganshan (JXjgs)	26°44.78 N 114°17.98 E	390	Near the rivulet
	Jiangxi, jiujiangxian (JXjjx)	29°35.68 N 115°54.74 E	80	Road side along the river
	Jiangxi, nanfengxian (JXnfx)	26°59.18 N 116°14.51 E	360	River side
	Jiangxi, xinjianxian (JXxjx)	28°44.33 N 115°44.11 E	100	Near the rivulet, naked area
North	Anhui, qianshanxian (AHqsx)	30°44.77 N 117°37.43 E	150	Road side
	Anhui, taihuxian (AHthx)	30°27.80 N 117°17.79 E	120	Railway bed
	Anhui, yuexixian (AHyxx)	30°42.89 N 116°15.33 E	600–800	Road side
Northwest	Xinjiang, aletaishi (XJalt)	47°46.72 N 88°20.64 E	830	Sandy grassplot near the river
	Xinjiang, qinghexian (XJqhx)	46°48.72 N 90°20.39 E	1400	Rivulet side in the valley

Table 2 The ISSR primers and the polymorphism of their PCR products

Primer sequence	*Total bands*	*Polymorphic bands*	*Shannon's information index*
(AG)8T	12	12	0.4207
(TC)8A	13	12	0.3403
(AC)8G	14	13	0.3277
(AC)8(CT)T	17	17	0.4163
(AC)8(CT)G	13	13	0.2895
(TG)8(AG)A	13	12	0.2297
(ATG)5	16	15	0.2182
(AGC)4G(AG)	16	15	0.3047
(AC)8T	11	11	0.3432
(GAA)6	12	12	0.5179
(GGAGA)3	8	8	0.4259
(GCT)4(CT)	18	17	0.4137
(CA)6(AG)	9	8	0.2723

For RAPD analysis, 153 primers (S2001∼S2100, S201∼220, S401∼S420, S480∼S500, Sangon Co. Ltd, Shanghai, China) were screened using the same procedure as in the ISSR analysis. Eleven primers with clear and reproducible bands were selected (Table 3). The procedure of PCR amplification and its product documentation for RAPD analysis were the same as that used in the ISSR analysis with only a slight modification in the thermal cycles: a hot start for 7 min at 94°C, and 45 cycles of 45 s at 94°C, 1 min at 36°C and 2 min at 72°C and ended by an extension of 7 min at 72°C.

Data analysis

The amplified RAPD and ISSR fragments were scored for the presence or absence of bands (1 = present, 0 = absent). Only clear and reproducible bands were scored. Each of the bands (DNA fragments) was considered a single, unique locus with Mendelian segregation. A locus was considered polymorphic if the relevant band was present in one or more, but not all, individuals of the population. A two-dimensional matrix was generated for each of the two molecular marker systems. The percentage of polymorphic bands (PPB) was calculated from the matrix by the number of PBs/total number of bands × 100%. The Shannon's Information Index (Lewontin, 1972), $I = 1 - \Sigma pi \ln pi$, was calculated for the genetic diversity in natural populations, where pi is the frequency of the ith band. Nei (1973), measurements of genetic diversity among natural populations were also calculated, including the total genetic diversity (that is, expected heterozygosity) (H_T), mean genetic diversity within populations (H_S), and the proportion of genetic diversity occurring among populations, $G_{ST} = (H_T - H_S)/H_T$ (Nei, 1973). All of these genetic diversity parameters were estimated using POPGENE version 1.32 (Yeh *et al*., 1999). The 21 natural populations were divided into four geographic groups: the northwestern group (two Xinjiang populations in the most northwestern part of China), which is geographically far from the other populations; the northern group (three populations north of the Yangtze River in the Anhui Province), southwestern group (four populations south of the Yangtze River in the Jiangxi Province), and southern group (the remaining populations south of the Yangtze River in the Anhui and Zhejiang Provinces). The southwestern group is separated from the southern group by Poyang Lake, the largest freshwater lake in China. The genetic variation among the natural populations was also estimated from the analysis of molecular variance (AMOVA) (Excoffier *et al*., 1992), in which the total genetic variance was partitioned into 'among population' and 'within population' components. The Hillis distance was calculated between each pair of individuals (Hillis, 1984). The genetic relationships among the natural populations were estimated using

the unweighted pair-group method with arithmetic averages (UPGMA). Cophenetic values (r_{cp}) based on the results of the UPGMA cluster analysis were calculated as a measure of the quality of clustering (Rohlf, 1982). The Mantel test (Mantel, 1967) was used to test the correspondence between RAPD and ISSR marker-based similarity matrices (Lapointe and Legendre, 1992), and to test the correlation between the geographical distance and genetic distance (Sharbel *et al*., 2000). The geographic distance was defined by GIS, and was calculated using the World Book Atlas module in IBM Book MAC Edition (version 2004). These calculations were carried out using NTSYS-pc version 2.11Q (Rohlf, 2002).

Table 3 The RAPD primers and the polymorphism of their PCR products

Primer order	*Primer sequence*	*Total bands*	*Polymorphic bands*	*Shannon's information index*
S97	ACGACCGACA	10	9	0.1843
S2002	TGCTCGGCTC	18	17	0.2696
S2006	GGACGACCGT	14	13	0.3307
S2055	CCAGACTCCA	15	15	0.3517
S2056	CTGGTGCTCA	15	15	0.4348
S2064	TCGGGTGTTG	11	11	0.5017
S2075	TGTCGTGGTC	15	15	0.3755
S2080	AGGCGGCACA	16	16	0.4820
S2083	TGGACTCGGT	17	17	0.3338
S2097	GGGAAAAGCC	18	18	0.3983
S2099	AGGCCAACAG	18	18	0.3832

Results

Thirteen ISSR primers produced 172 scorable bands in 566 individuals of 21 natural populations and Col ecotype, with 13.24 bands per primer in average ranging from 8 (by primer $(AC)_8$) to 18 bands (by primer $(GCT)_4(CT)$). Among the total number of bands, 95.93% (165) were polymorphic (Table 2), and 6.98% (12) were not significantly different among populations ($P>0.05$) while 85.47% (147) were significantly different ($P<0.001$). For the RAPD analysis, 11 primers yielded 165 reliable bands in the 560 individuals of 21 natural populations and Col ecotype, with 15 bands per primer in average ranging from 10 (by primer S2002) to 18 bands (by primers S97, S2006 and S2083). Of the bands, 98.18% (162) were polymorphic (Table 3) and 7.78% (13) were not significantly different among populations ($P>0.05$) while 83.63% (138) were significantly different ($P<0.001$). In both cases, more than 50% of the polymorphic fragments occurred at frequencies less than 0.10. For the ISSR markers, the mean percentage of PPB within natural populations was 35.94%, ranging from 18.60% (JXnfx) to 56.40% (AHyix) (Table 4). Geographically, the four southwestern populations were found to have low PPBs, while the populations in Anhui Province had relatively high PPBs in both the north and south groups. Two individuals from JXjgs and JXnfx respectively were determined to have identical banding profiles, but no identical profile was shared by individuals from different populations. For the RAPD markers, the mean PPB was 35.38%, ranging from 21.21% (ZJdys,

Table 4 Genetic diversity of the natural populations from China based on ISSR and RAPD data sets

Group	*Population*	*ISSR*				*RAPD*			
		Sample size	*Polymorphic bands*	*PPB (%)*	*I*[a]	*Sample size*	*Polymorphic bands*	*PPB (%)*	*I*[a]
South	AHngs	29	85	49.42	0.2529	29	82	49.70	0.2403
	AHqyx	29	72	41.86	0.2242	29	70	42.42	0.2003
	AHsex	31	71	41.28	0.2212	29	44	26.67	0.1321
	AHwhs	29	71	41.28	0.2014	29	69	41.82	0.1897
	AHxnx	29	83	48.26	0.2402	28	67	40.61	0.1974
	AHyix	30	97	56.40	0.2940	29	88	53.33	0.2669
	ZJcax	26	68	39.53	0.1831	26	60	36.36	0.1849
	ZJdys	16	49	28.49	0.1385	16	21	21.21	0.1109
	ZJjds	23	42	24.42	0.1430	23	63	38.18	0.1755
	ZJkhx	29	55	31.98	0.1562	29	40	24.24	0.1386
	ZJlas	14	55	31.98	0.1751	13	44	26.67	0.1393
	ZJtls	17	50	29.07	0.1546	17	52	31.52	0.1815
Southwest	JXjgs	28	43	25.00	0.1229	28	69	41.82	0.1857
	JXjjx	30	38	22.09	0.1057	29	55	33.33	0.1477
	JXnfx	30	32	18.60	0.0966	30	47	28.48	0.1228
	JXxjx	29	48	27.91	0.1468	30	61	36.97	0.1797
North	AHqsx	30	82	47.67	0.2350	30	63	38.18	0.1781
	AHthx	27	89	51.74	0.2580	28	78	47.27	0.2234
	AHyxx	30	59	34.30	0.1791	30	69	41.82	0.1968
Northwest	XJalt	16	55	31.98	0.1759	14	35	21.21	0.1248
	XJqhx	15	54	31.40	0.1840	15	21	21.21	0.1231
Col		29	13	7.56	0.0411	29	32	19.39	0.0999
Species		566	165	95.93	0.3462	560	162	98.18	0.3698

Abbreviations: ISSR, inter-simple sequence repeat; PPB, percentage of polymorphic bands; RAPD, random amplified polymorphic DNA.
[a]I, Shannon's information index.

XJalt and XJqhx) to 53.33% (AHyix) (Table 4). Two populations from the northwest group and one from the south group (ZJdys) have the lowest PPB among the natural populations, while the populations from Anhui province have relatively high PPBs in both the north and south groups. Two individuals from JXnfx, JXjjx and JXxjx, respectively, have identical banding profiles, but no identical profile was shared by individuals from different populations. Compared to the natural populations, the Col ecotype has the lowest PPB, as revealed by both the ISSR and RAPD markers. The mean Shannon's Information index within natural populations was 0.1852, ranging from 0.0966 (JXnfx) to 0.2940 (AHyix) for ISSR markers; and 0.1733, ranging from 0.1109 (ZJdys) to 0.2669 (AHyix) for RAPD markers (Table 4).

At population level, the genetic diversity of all populations is 0.4267 as detected by using ISSR markers, and 0.4946 by RAPD markers. At the group level, both ISSR and RAPD markers detected the lowest genetic diversity in the northwest group. However, ISSR markers detected the highest genetic diversity in the southwest group ($G_{st} = 0.3372$), while RAPD detected the highest diversity in the south group ($G_{st} = 0.4768$). The AMOVA analysis detected about 54 and 58% of the variation due to the genetic variation among populations based on the ISSR and RAPD data sets, respectively (Table 5). It is interesting to note that at group level, in contrast to population level, less genetic variation was found among populations, as revealed by both ISSR and RAPD markers, except for the southern group where more genetic variation was found among populations as detected by RAPD markers. For example, in the northern group, only about 41 and 39% of variation is found among populations revealed by ISSR and RAPD markers, respectively. Further analysis indicated that this resulted from a relatively high proportion of genetic variation among the groups (Table 5).

UPGMA dendrograms based on the Hillis distance matrixes generated from the raw data of the ISSR or RAPD markers did not group the populations exactly according to their geographic distribution. However, the four populations of the southwestern group were always clustered together in both ISSR and RAPD dendrograms as a monophyletic group (Figures 2 and 3). Although two of the populations in the northwestern group were always clustered together, it is interesting to note that these two were also clustered with some populations in the southern group. The Col ecotype was the 'basal group' in both the ISSR and RAPD dendrograms. The cophenetic values (r_{cp}) of the UPGMA clustering based on the ISSR and RAPD data sets was significantly correlated to the primary data matrixes of ISSR and RAPD, respectively ($r_{cp} = 0.80793$ for ISSR data set and $r_{cp} = 0.91622$ for RAPD data set).

The Mantel test on the ISSR data set indicated that the geographical distance was significantly correlated with the genetic distance ($r_{cp} = 0.58924$, $P < 0.001$) when using the complete data set (including Col). When Col or Col and the Xinjiang's populations were excluded from the data set respectively, the correlations were still significant ($r_{cp} = 0.37952$, $P < 0.005$ and $r_{cp} = 0.49100$, $P < 0.0001$ respectively). Similar correlations were also found for the RAPD data set when the Mantel test was applied ($r_{cp} = 0.71907$, $P < 0.00001$ for complete data set; $r_{cp} = 0.32985$, $P < 0.01$ excluding Col; $r_{cp} = 0.30710$, $P < 0.02$ excluding Col and Xinjiang's populations). The Mantel test also detected a significant correspondence between the genetic similarities based on the ISSR and RAPD data sets ($r_{cp} = 0.6792$, $P < 0.00001$).

Discussion

The distribution and natural habitats of *Arabidopsis thaliana* in China

A. thaliana is widely distributed in China and has been recorded in Eastern, Central, Northwestern, Western and Southwestern China (Cheo *et al*., 1987, 2001), covering both temperate and subtropical zones. Although *A. thaliana* has a wide distribution in China, relatively few herbarium specimens are found in this country. The earliest dated specimens we examined were collected from Yixian County, Anhui Province, in 1910 and deposited in the herbarium of the Institute of Botany, Beijing (PE). This lack of herbarium specimens may partially be attributed to the fact that this species has very short flowering–fruiting times, usually in early spring and/or that it is an inconspicuous weedy species.

The habitats of the natural populations of *A. thaliana* collected for this study are diverse, such as from along the roadside or from farmland, slopes or abandoned fields, with altitudes ranging from 20 m (AHthx) to 1400 m (XJqhx). The populations are usually distributed in relatively moist areas. The two populations from Xinjiang Province (XJqhx and XJalt) were obtained from the Altai Mountain range. Among the samples collected for this study, the XJqhx and XJalt populations grow at the highest altitude and lowest temperature during the growth season. When the authors collected samples from Qinghe (XJqhx) in early June, it was during a period of snowfall. The phenotype of the plants varies greatly in the fields. For example, the average height of individuals of the northwestern group is 16–18 cm, while that of the southern group is 26–46 cm. When the individuals from different populations were planted in the greenhouse at $22 \pm 2^{\circ}$C, those in the northwest group tended to flower earlier than the populations from East China.

These natural populations are usually found in disturbed habitats with strong influences of human activity. It seems that some newly found distributions could be attributed to human-aided dispersal. Since *A. thaliana* produces numerous tiny seeds, wind could also cause long-distance dispersal (Tackenberg *et al*., 2003; Jorgensen and Mauricio, 2004). The dynamic distribution pattern of *A. thaliana* in East China might reflect a history of combination of natural and human-aided dispersal.

Comparison of the results based on RAPD and ISSR data sets

Although there have been many debates on the reproducibility of RAPD markers, many studies have shown that RAPDs are useful molecular markers to detect genetic diversity at the population level in well-controlled experiments (Bartish *et al*., 2000; Diaz *et al*., 2001; Reisch *et al*., 2003; Fontaine *et al*., 2004; Nybom, 2004). Since the ISSR markers are generated by longer primers (15–24 bp), these were thought to be more stable than the RAPD markers (Yang *et al*., 1996; Nagaoka and Ogihara, 1997; Parsons *et al*., 1997; Esselman *et al*., 1999). In this study, the reproducibility of RAPD and ISSR was

Table 5 Genetic variations within and among populations revealed by AMOVA based on ISSR and RAPD data sets

Source of variation	*ISSR*					*RAPD*				
	d.f.	*Sum of squares*	*Variance components*	*Percentage of variation*	P-*value*	*d.f.*	*Sum of squares*	*Variance components*	*Percentage of variation*	P-*value*
South group										
Among populations	11	1868.169	6.38303	39.00	$P<0.001$	11	2683.812	9.56261	53.02	$P<0.001$
Within populations	92	2895.632	9.98494	61.00	$P<0.001$	285	2415.111	8.47407	46.98	$P<0.001$
Southwest group										
Among populations	3	363.572	3.99176	47.21	$P<0.001$	3	352.698	3.82326	39.87	$P<0.001$
Within populations	111	504.309	4.46291	52.79	$P<0.001$	113	651.508	5.76555	60.13	$P<0.001$
North group										
Among populations	2	427.481	7.02520	40.66	$P<0.001$	2	357.032	5.78348	39.24	$P<0.001$
Within populations	84	861.185	10.25220	59.34	$P<0.001$	5	761.150	8.95471	60.76	$P<0.001$
Northwest group										
Among populations	1	90.885	5.28773	36.98	$P<0.001$	1	43.801	2.64236	32.32	$P<0.001$
Within populations	29	261.308	9.01063	63.02	$P<0.001$	27	149.371	5.53228	67.68	$P<0.001$
East China										
Among groups	2	1586.248	4.32132	22.68	$P<0.001$	2	1231.218	2.70323	14.65	$P<0.001$
Within groups	16	2659.222	5.98210	31.40	$P<0.001$	16	3393.541	7.82881	42.42	$P<0.001$
Within populations	487	4261.125	8.74974	45.92	$P<0.001$	483	3827.768	7.92499	42.94	$P<0.001$
China										
Among groups	3	1972.475	4.63004	23.93	$P<0.001$	3	1627.269	3.24804	17.37	$P<0.001$
Within groups	17	2750.108	5.95750	30.79	$P<0.001$	17	3428.617	7.63584	40.83	$P<0.001$
Within populations	516	4522.434	8.76441	45.29	$P<0.001$	510	3985.864	7.81542	41.80	$P<0.001$

Abbreviations: ISSR, inter-simple sequence repeat; RAPD, random amplified polymorphic DNA.

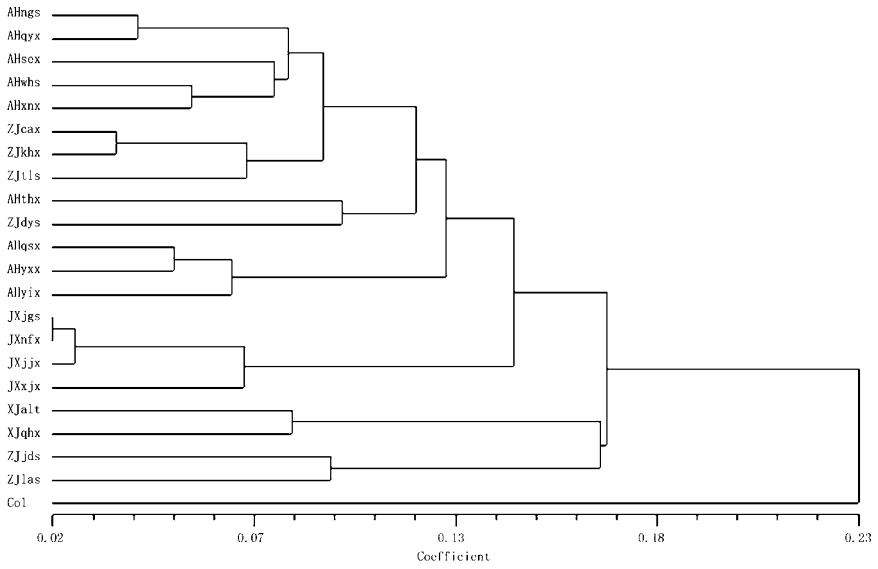

Figure 2 Unweighted pair-group method with arithmetic averages clustering of *Arabidopsis thaliana* based on Hillis distance calculated from inter-simple sequence repeats data set.

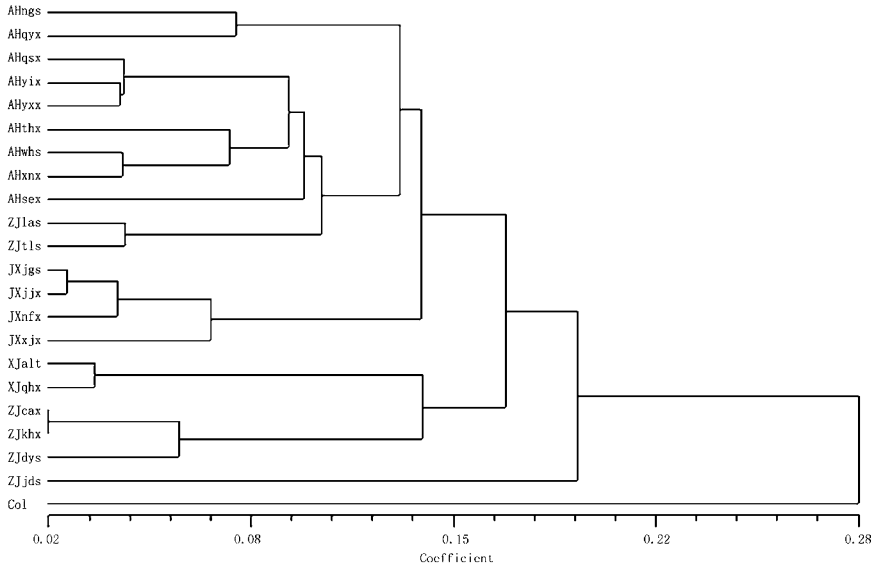

Figure 3 Unweighted pair-group method with arithmetic averages clustering of *Arabidopsis thaliana* based on Hillis distance calculated from random amplified polymorphic DNA data set.

assured by a repeated PCR amplification for at least six individuals in each population. Both ISSR and RAPD amplifications produced stable and repeatable fragments by the selected primers. In general, the ISSR and RAPD markers generated similar results on the genetic diversity within and among the natural populations, as revealed by Nei's measurement (Nei, 1973) of genetic diversity and AMOVA analysis. Moreover, the UPGMA clustering based on these two data sets were significantly correlated to their distance data matrixes respectively, although the r_{cp} value of ISSR ($r_{cp} = 0.80793$) is lower than that of RAPD ($r_{cp} = 0.91622$).

Genetic variation among and within populations

Although extensive studies have been conducted on the genetic diversity of *A. thaliana* around the world, none of the populations from China have ever been included. This study is focused on the genetic diversity of naturally distributed populations in China. The overall genetic diversity of the 21 natural populations, $G_{st} = 0.4946$ (RAPD), $G_{st} = 0.4267$ (ISSR), was approximately the same as that seen in other selfing species with gravity-dispersed seeds, whose average $G_{st} = 0.5$ (Hamrick and Godt, 1996), less than that of the native North European populations ($F_{st} = 0.88$; Stenoien *et al*, 2005) or native populations in France ($F_{st} = 0.59$; Le Corre, 2005), but significantly greater than that of the North American populations ($G_{st} = 0.28$) reported by Jorgensen and Mauricio (2004). When the northwestern populations were excluded, the genetic divergence of the 19 East China populations reduced only slightly (data not shown). Although *A. thaliana* was introduced into East China (Al-Shehbaz and O'Kane, 2002), its overall genetic

diversity is much greater than those recently introduced populations in North America, but still less than those native populations in Europe. This finding might indicate that the populations in China have a longer 'introduction' history than those in North America, which could also be supported by the fact that no identical banding profile was found between/among individuals of different populations. When the genetic variation is dissected into 'within' and 'among' population variations, typically, selfing species have low levels of genetic diversity within populations, but a substantial differentiation among populations (Hamrick and Godt, 1996). In previous studies, the hierarchical AMOVA on the genetic variation of the natural population of *A. thaliana* revealed different patterns for different populations. In most cases, the observed genetic variation was consistent with its lifestyle: less variation existed within populations and more genetic variation was found among populations (Hanfstingl *et al*., 1994; Bergelson *et al*., 1998; Breyne *et al*., 1999; Miyashita *et al*., 1999; Jorgensen and Mauricio, 2004; Stenoien *et al*., 2005). For example, Stenoien *et al*. (2005) found that, on average, only 12% of the genetic variation occurred among individuals within 10 northern European populations of *A. thaliana*, as screened by microsatellite markers. In other cases, higher genetic variation was found *within* populations than *among* populations. For example, Jorgensen and Mauricio (2004) detected that approximately 77% of the genetic variation occurred among individuals within six North American populations, as revealed by AFLP markers; and Bakker *et al*. (2006) found 56.7% of genetic variation within populations by the sequences of six genes and five microsatellite loci over the species range. In this study, in the hierarchical AMOVA for the 21 Chinese populations, it was found that slightly more genetic variation occurred *among* populations rather than *within* populations, as revealed by ISSR (54.7%) and RAPD (58.2%) makers, respectively. When two Xinjiang populations were excluded, the genetic variations within 19 Chinese populations were similar to that within 21 populations (Table 5). Although it is difficult to compare directly the results of the studies mentioned above due to different population size and methodology being adopted, the general trend can be referred. The relatively high amount of genetic variation found among groups of Chinese populations may indicate a geographic structure.

The correlation of genetic and geographic distance

The correlation between the genetic distance of populations and their geographic distance has been discussed extensively in previous studies using various molecular markers in *A. thaliana*. In most cases, no clear association between geographical origin and genetic similarity was detected in populations distributed in different regions of the world (King *et al*., 1993; Hanfstingl *et al*., 1994; Innan *et al*., 1997; Ullrich *et al*., 1997; Bergelson *et al*., 1998; Loridon *et al*., 1998; Breyne *et al*., 1999; Miyashita *et al*., 1999; Erschadi *et al*., 2000; Jorgensen and Mauricio, 2004; Stenoien *et al*., 2005; Bakker *et al*., 2006). However, Sharbel *et al*. (2000) detected significant, but weak, isolation by distance among populations sampled from the presumed native range of *A. thaliana* in Eurasia; and Barth *et al*. (2002) found some Asian accessions clustered separately from the central European plants. A recent study by Nordborg *et al*. (2005) on the fragments of 480 kb obtained from each of 96 individual genomes revealed a strong population structure, despite the fact that individual populations harbor much of the variation present species–wide. In this study, generally significant correlations between the geographical and genetic distance were detected using the Mantel test for Chinese populations, based on both ISSR and RAPD markers ($P<0.005$ for ISSR and $P<0.01$ for RAPD) and for the eastern Chinese populations ($P<0.01$ for ISSR and $P<0.02$ for RAPD).

Although *A. thaliana* is present within China and usually distributed in disturbed habitats strongly influenced by human activities, the correlation between its genetic and geographic distance at population level suggests that some natural dispersal mechanism may also be involved in the distribution of this species. This distribution in China is unlike the population distributions in North America, where the populations were possibly originated very recent from a mixed origin (Jorgensen and Mauricio, 2004). However, it is impossible to trace the origin of the Chinese populations only based on the ISSR or RAPD data of these limited populations. Data on precise sequences from more populations worldwide will be needed to explore the origin of Chinese populations and their phylogenetic relationships with those from other parts of the world.

Acknowledgements

We thank Professors Bingyang Ding and Jianzhang Shao for providing information on the distribution of *A. thaliana* in China, and thank Ms Mengxin Xie for her technique assistance. We are very grateful to Professor Xiaoru Wang at the Swedish University of Agricultural Sciences and Dr Xiaoquan Wang at the Institute of Botany, the Chinese Academy of Sciences for their valuable comments and suggestions on this work. This work is supported by the State Key Basic Research and Development Plan (2006CB100105).

References

Al-Shehbaz IA, O'Kane SL (2002). Taxonomy and phylogeny of Arabidopsis (Brassicaceae). In: Somerville CR, Meyerowitz EM (eds). *The Arabidopsis Book*. Rockville, MD. *Am Soc Plant Biol* http://www.aspb.org/publications/arabidopsis149a, p 22.

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006). A genome-wide survey of R gene polymorphisms in Arabidopsis. *Plant Cell* **18**: 1803–1818.

Barth S, Melchinger AE, Lubberstedt T (2002). Genetic diversity in Arabidopsis thaliana (L.) Heynh. investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. *Mol Ecol* **11**: 495–505.

Bartish IV, Rumpunen K, Nybom H (2000). Combined analyses of RAPDs, cpDNA and morphology demonstrate spontaneous hybridization in the plant genus Chaenomeles. *Heredity* **85**: 383–392.

Bergelson J, Stahl E, Dudek S, Kreitman M (1998). Genetic variation within and among populations of Arabidopsis thaliana. *Genetics* **148**: 1311–1323.

Breyne P, Rombaut D, Van Gysel A, Van Montagu M, Gerats T (1999). AFLP analysis of genetic diversity within and between Arabidopsis thaliana ecotypes. *Mol Gen Genet* **261**: 627–634.

Cheo T, Gue R, Lan Y, Lou L, Kuan K, An Z (1987). Arabidopsis. In: Cheo T (ed). *Flora Reipublicae Popularis Sincae*, Vol. 33. Science Press: Beijing, China. pp 281–287 (in Chinese).

Cheo T, Lu L, Yang G, Al-Shehbaz IA, Dorofeyev VI (2001). Brassicaceae. In: Wu Z, Raven PH (eds). *Flora of China*, Vol. 8. Science Press: Beijing, China; and the Missouri Botanical Garden Press: St Louis, USA. pp 120–121.

Diaz V, Muniz LM, Ferrer E (2001). Random amplified polymorphic DNA and amplified fragment length polymorphism assessment of genetic variation in Nicaraguan populations of Pinus oocarpa. *Mol Ecol* **10**: 2593–2603.

Erschadi S, Haberer G, Schöniger M, Torres-Ruiz RA (2000). Estimating genetic diversity of Arabidopsis thaliana ecotypes with amplified fragment length polymorphisms (AFLP). *Theor App Genet* **100**: 633–640.

Esselman EJ, Jianqiang L, Crawford DJ, Windus JL, Wolfe AD (1999). Clonal diversity in the rare Calamagrostis porteri ssp. insperata (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) markers. *Mol Ecol* **8**: 443–451.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM (2004). Genetic diversity of the shea tree (Vitellaria paradoxa C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* **93**: 639–648.

Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plant species. *Philos T Roy Soc B* **351**: 1291–1298.

Hanfstingl U, Berry A, Kellogg EA, Costa III JT, Rudiger W, Ausubel FM (1994). Haplotypic divergence coupled with lack of diversity at the Arabidopsis thaliana alcohol dehydrogenase locus: roles for both balancing and directional selection? *Genetics* **138**: 811–828.

Hillis DM (1984). Misuse and modification of Nei's genetic distance. *Syst Zool* **33**: 238–240.

Hoffmann MH, Glass AS, Tomiuk J, Schmuths H, Fritsch RM, Bachmann K (2003). Analysis of molecular data of Arabidopsis thaliana (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). *Mol Ecol* **12**: 1007–1019.

Innan H, Terauchi R, Miyashita NT (1997). Microsatellite polymorphism in natural populations of the wild plant Arabidopsis thaliana. *Genetics* **146**: 1441–1452.

Jorgensen S, Mauricio R (2004). Neutral genetic variation among wild North American populations of the weedy plant Arabidopsis thaliana is not geographically structured. *Mol Ecol* **13**: 3403–3413.

King G, Nienhuis J, Hussey C (1993). Genetic similarity among ecotypes of Arabidopsis thaliana estimated by analysis of restriction fragment length polymorphisms. *Theor Appl Genet* **86**: 1028–1032.

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004). Naturally occurring genetic variation in Arabidopsis thaliana. *Annu Rev Plant Biol* **55**: 141–172.

Kuittinen H, Mattila A, Savolainen O (1997). Genetic variation at marker loci and in quantitative traits in natural populations of Arabidopsis thaliana. *Heredity* **79**: 144–152.

Lapointe F-J, Legendre P (1992). Statistical significance of the matrix correlation coefficient for comparing independent phylogenetic trees. *Syst Biol* **41**: 378–384.

Le Corre V (2005). Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol Ecol* **14**: 4181–4192.

Lewontin R (1972). Apportionment of human diversity. *Evol Biol* **6**: 381–398.

Loridon K, Cournoyer B, Goubely C, Depeiges A, Picard G (1998). Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of Arabidopsis thaliana. *Theor Appl Genet* **97**: 591–604.

Mantel NA (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

Mitchell-Olds T, Schmitt J (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**: 947–952.

Miyashita NT, Kawabe A, Innan H (1999). DNA variation in the wild plant Arabidopsis thaliana revealed by amplified fragment length polymorphism analysis. *Genetics* **152**: 1723–1731.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.

Nagaoka T, Ogihara Y (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor Appl Genet* **94**: 597–602.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H *et al.* (2005). The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol* **3**: e196.

Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* **13**: 1143–1155.

Parsons BJ, Newbury HJ, Jackson MT, Ford-Lloyd BV (1997). Contrasting genetic diversity relationships are revealed in rice (Oryza sativa L.) using different marker types. *Mol Breeding* **3**: 115–125.

Qu L-J, Li L, Zhang Y, Liu M, Gu H, Chen Z (2000). Cloning and expression of a cDNA encoding ribosomal protein S4 from rice (Oryza sativa). *Chin Sci Bull* **45**: 168–173.

Reisch C, Poschlod P, Wingender R (2003). Genetic differentiation among populations of Sesleria albicans Kit. ex Schultes (Poaceae) from ecologically different habitats in central Europe. *Heredity* **91**: 519–527.

Rohlf FJ (1982). Consensus indices for comparing classifications. *Math Biosci* **59**: 131–144.

Rohlf FJ (2002). *NTSYSpc: Numerical Taxonomy System, Ver. 2.1*. Exeter Publishing, Ltd. Setauket, New York.

Schmid KJ, Torjek O, Meyer R, Schmuths H, Hoffmann MH, Altmann T (2006). Evidence for a large-scale population structure of Arabidopsis thaliana from genome-wide single nucleotide polymorphism markers. *Theor Appl Genet* **112**: 1104–1114.

Sharbel TF, Haubold B, Mitchell-Olds T (2000). Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. *Mol Ecol* **9**: 2109–2118.

Stenoien HK, Fenster CB, Tonteri A, Savolainen O (2005). Genetic variability in natural populations of Arabidopsis thaliana in northern Europe. *Mol Ecol* **14**: 137–148.

Tackenberg O, Poschlod P, Kahmen S (2003). Dandelion seed dispersal: the horizontal wind speed does not matter for long distance dispersal – it is updraft!. *Plant Biol* **5**: 451–454.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796–815.

Ullrich H, Lattig K, Brennicke A, Knoop V (1997). Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 Arabidopsis thaliana ecotypes. *Plant Mol Biol* **33**: 37–45.

Vos P (1998). AFLP fingerprinting of Arabidopsis. *Methods Mol Biol* **82**: 147–155.

Yang W, de Oliveira AC, Godwin I, Scheritz K, Bennetzen JL (1996). Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Sci* **36**: 1669–1676.

Yeh FC, Yang R, Boyle T (1999). *POPGENE, Version 1.32. Microsoft Window-Based Freeware for Population Genetic Analysis*. University of Alberta: Edmonton. Available at: http://www.ualberta.ca/~fyeh/index.htm.