Journal of Medicinal Plants Research Vol. 6(37), pp. 5086-5093, 26 September, 2012
Available online at http://www.academicjournals.org/JMPR
DOI: 10.5897/JMPR12.222
ISSN 1996-0875 ©2012 Academic Journals

Full Length Research Paper

Genetic diversity of the chloroplast trnL-trnF intergenic spacer and nuclear internal transcribed spacer of great burdock (*Arctium lappa* L.) in Taiwan

Mei-Chen Tseng[1], Saou-Lien Wong[1], Dai-Shion Hsiung[1], Jieh-Horng Hwang[2], Shih-Chiang Lee[3] and Fu-An Chen[2]*

[1]Department of Aquaculture, National Pingtung University of Science and Technology, Pingtung 912, Taiwan.
[2]Department of Pharmacy, Tajen University, Pingtung 907, Taiwan.
[3]Tong Yuan Biopharmaceutical Co. Ltd., Kaohsiang 813, Taiwan.

Accepted 10 April, 2012

Great burdock (*Arctium lappa* L.) of the Compositae has economic importance in East Asian countries. In Taiwan, major cultivated areas are located in Gueilai Township, Pingtung County and Jiali Township, Tainan County. The present work was undertaken with two objectives of determining intraspecific genetic variations and studying the population dynamics of *A. lappa* L. In total, 23 individuals were collected from Gueilai (n = 15) and Jiali (n = 8). Chloroplast trnL-trnF intergenic spacer (IGS) and nuclear ribosomal DNA internal transcribed spacer (ITS) sequences were subcloned and sequenced. Length variations among IGS sequences ranged from 805 to 808 bp and among ITS sequences ranged from 360 to 362 bp. Intraspecific genetic distances within IGS sequences ranged from 0 to 0.01 and within ITS sequences ranged from 0 to 0.023, which were lower than interspecific distances. All IGS and ITS sequences being separately scattered on each of the neighbor-joining trees with insignificant genetic differentiation indices (p > 0.05) suggested that both samples shared a recent common ancestor. Minimum spanning trees, Tajima's and Fu's parameters, and mismatch distributions implied that *A. lappa* L. had experienced a historical population expansion.

Key words: Compositae, Gueilai, Jiali, genetic distance, population expansion.

INTRODUCTION

Great burdock (*Arctium lappa* L.), an herb belonging to the Compositae, is mainly cultivated in Orient including Taiwan, China, Japan, South Korea, and a few other Southeast Asian countries. The original cultivated area of great burdock in southern Taiwan was in Gueilai Township, Pingtung County. The sown area exceeded 10 ha. Nowadays, the highest production area in Taiwan has shifted to Jiali Township, Tainan County, where over 70 ha are cultivated. Most of these products are exported to the Japanese market.

Great burdock is very nutritious, and many of its constituents have important physical activities, including plentiful proteins, vitamins, carbohydrates (inulin and dietary fiber), minerals (zinc, iron, magnesium, and iodine), unsaturated fatty acids, and polyphenols (chlorogenic acid, caffeic acid, isochlorogenic acid, caffeic acid derivatives, etc) (Murata et al., 1995). Its root has antioxidative (free radical scavenging activity) (Lin et al., 1996; Duh, 1998; Chen et al., 2004), antibacterial (Chow et al., 1997), antitumor, and anti-mutant activities (Morita et al., 1984, 1985; Kasai et al., 2000; Matsumoto et al., 2006; Lee and Kim, 2010; Yao et al., 2011). It also has some advantages in promoting human health, such as reducing cough, protecting against allergies (Wei and Wang, 2006; Knipping et al., 2008; Li et al., 2008), eliminating constipation (Dos Santos et al., 2008), decreasing blood lipids (Wei, 2006), lowering blood sugar and pressure (Cicero et al., 2004; Wei, 2006), resisting weariness and senility (Wei, 2006), protecting liver activities and the mucous membrane of the stomach, promoting inflammatory activities (Lin et al., 1996, 2000, 2002; Dos Santos et al., 2008; Chen et al., 2009), and

*Corresponding author. E-mail: fachen@mail.tajen.edu.tw.

inducing AMP-activated protein kinase activity (Miele and Beguinot, 2012). Cultivation of great burdock strains on Taiwanese farms depends on introducing seeds from Japan every year. In general, these seed strains have several beneficial characters achieved by artificial selection including long roots, white stalks and root bark, and a short sowing time. Most consumers prefer the great burdock with longer roots (more than 70 cm long with a diameter of about 2 to 3 cm) and white root bark. However, these strains are only distinguished by their morphological characters.

In the past decade, molecular genetics rapidly developed. Some technologies are commonly used to identify strains. Szczepaniak et al. (2002) explored morphological polymorphisms and genetic variations of *Elymus repens* (L.) by amplified fragment length polymorphism (AFLP). Wen and Hsiao (1999) examined genetic differentiation of *Lilium longiflorum* Thunb. var. *scabrum* Masam (Liliaceae), in Taiwan, by random amplification of polymorphic DNA (RAPD) and morphological characters. Sequence analyses of specific genes from nuclear (nr) and chloroplast (cp) DNA were also extensively utilized in studies of systematics (Ouborg et al., 1999). The trnL-trnF intergenic spacer (IGS) of the chloroplast genome has high variability and is suitable for analyzing different strains (Guzmán and Vargas, 2005). Moreover, nrDNA can provide more information for barcoding than organellar DNA (Chase and Fay, 2009). The ribosomal internal transcribed spacer (ITS)-1 is considered a good marker due to its ease of amplification. Nowadays, the ITS sequence is the most common tool in determining Chinese medicine germplasm, including discrimination among *Rhizoma curcumae* strains with different volatile oil ingredients (Xia et al., 2005) and systematic studies at the generic and specific levels of *Bupleurum*, *Fritillaria*, and *Dendrobium* species (Neves and Watson, 2004). The ITS region has the property of undergoing rapid concerted evolution. Its small size and ease of amplification makes the ribosomal (rDNA) ITS a very important tool in examining Chinese medicine germplasm.

Great burdock has been cultivated in Taiwan since Japanese colonial rule (1895 to 1945). At present, Gueilai and Jiali are major cultivation areas in southern Taiwan. The objectives of this study were to examine sequence diversities of the trnL-trnF IGS and ribosomal ITS-1 between 2 sets of samples from Gueilai and Jiali and to explore the population dynamics of the species.

MATERIALS AND METHODS

Sampling

In total, 23 *A. lappa* L. specimens were collected from Gueilai (n = 15) and Jiali (n = 8) in June 2009. Fresh leaves and roots were carried back to the laboratory and washed using distilled water, after being dried; leaves of these specimens were maintained at -70°C until DNA was extracted.

DNA extraction, amplification, and sequencing

Genomic DNA was extracted from 200 mg of leaf using the method of Sambrook et al. (1989). The trnL-trnF IGS was amplified using the forward primer, 5'-CGAAATCGGTAGACGCTACG-3', and reverse primer, 5'-ATTTGAACTGGTGACACGAG-3', which were developed for this study by referring to NBCI GenBank sequence AF129824. Primer pair of rDNA, forward primer, 5'-GGAGAAGTCGTAACAAGGTTTCCG-3' and reverse primer, 5'-GCTACGTTCTTCATCGATGCGTG-3' were used to amplify the ribosomal ITS-1 region. A polymerase chain reaction (PCR) consist of approximately 5 ng genomic DNA, 40 pmol each of the forward and reverse primers, 25 mM dNTP, 0.05 to 0.1 mM $MgCl_2$, 10× buffer, and 4 U *Taq* polymerase (Takara Shuzo, Shiga, Japan) brought to 100 µl with Milli-Q water. The PCR program used for IGS amplification included 1 cycle of 5 min at 95°C; 35 cycles of 1 min at 94°C, 30 s at 54°C, and 1 min at 72°C; followed by a single further extension of 10 min at 72°C. The program used for ITS-1 amplification included 1 cycle of 3 min at 95°C; 5 cycles of 30 s at 95°C, 30 s at 50°C, and 30 s at 72°C; and 38 cycles of 30 s at 90°C, 30 s at 56°C, and 30 s at 72°C; followed by a single further extension of 10 min at 72°C. We evaluated 10 µl of each product on a 1% agarose gel to check the PCR success and confirm the product sizes. Under these conditions, a single strong fragment was amplified in all specimens used. Amplified fragments of the ITS were purified using a DNA Clean/Extraction kit (GeneMark, Taichung, Taiwan), cloned with the pGEM-T Easy Vector System (Promega, Madison, WI, USA) and transformed into the *Escherichia coli* JM109 strain. Plasmid DNA was isolated using a mini plasmid kit (Geneaid, Taichung, Taiwan). Each colony from 1 individual was randomly chosen for cloning followed by sequencing in an Applied Biosystems (ABI, Foster City, CA, USA) automated DNA sequencer 377 (vers. 3.3) with a Bigdye sequencing kit (Perkin-Elmer, Wellesley, MA, USA). The T7 or SP6 primers were used in the sequencing reaction, and PCR cycle parameters for sequencing were 35 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. Sequences were determined on an Applied Biosystems (ABI) automated DNA sequencer 377 (vers. 3.3).

Genetic data analysis

Nucleotide sequences were aligned with the program, CLUSTAL W (Thompson et al., 1994), and alignments were verified by eye. The number of polymorphic sites was estimated using Arlequin vers. 3.1 (Excoffier et al., 2010). Levels of inter- and intra-sample genetic diversity were quantified by indices of haplotype diversity (h_d) and pairwise estimates of nucleotide divergence (d_{ij}) both among and within samples. The average number of nucleotide substitutions per site and pairwise differences (π) were determined using DnaSP vers. 4.5 (Rozas et al., 2003). Genetic distances between samples were analyzed by the Kimura 2-parameter (K2P) method (Kimura, 1980). Two homologous sequences of *Arctium minus* (AY772274) and *Saussurea tomentosa* (EF420981.1) from NCBI GenBank were used as outgroups in the IGS tree. *Saussurea glacialis* (AB118121.1) was selected as outgroup in the ITS tree. Phylogenetic trees were constructed using the Neighbor-joining (NJ) method in the MEGA 3 program (Saitou and Nei, 1987; Kumar and Nei, 2004). The confidence of each node of the tree was tested by bootstrapping (Felsenstein, 1985) with 1000 replicates. Nodes with bootstrap values of >70 were significantly supported by a ≥95% probability (Hillis and Bull, 1993). The minimum spanning tree (MST) was computed from the matrix of pairwise distances between all pairs of haplotypes in each sample using a modification of the algorithm described in Rohlf (1973). An exact test of genetic differentiation between the 2 samples was also estimated with F statistics (F_{St}) (Raymond and Rousset, 1995).

```
[              11122 2222233333 3333444444 4455557777 7777788]
[          4667904700 0156800113 3457122335 7833490046 6779900]
[          8230909036 8607257471 4670179035 6604483881 2145647]
G-1        TTTAAATAAG GTTAATATTA GACATTAGTA TATCTTATAT TAAATGT
G-2        .......... .....C.... .......... .G.T...... .......
G-3        .......... .......... .......... ...T..G... .......
G-5        .......... .........C. .....A.... ...T...... ...G...
G-7        .......... .........G .......... ...T...... .......
G-8        ..C....... .......... .......... ...T...... .......
G-9        .......... .......... ......G... ...T...... .......
G-10       .......... .......... .......... ...T...... ...-C..
G-11       .......... .......... .........G ...TC...G. C......
G-12       .......... ........C. .......... ...T...... .......
G-13       .......... .......... .......... ...T.C...C .......
G-15       .......... ..C...GC.. .......... ...T...... .......
J-1        .......... .C........ A......... ...T...... .......
J-2        .......GT. .......... .......... ...T...... .......
J-3        A........A ..A....... .G........ ...T...... .......
J-4        .C........ .......... ...G...... ..CT...C.. .......
J-5        ...G...... .......... C...C..... ...T...... .......
J-6        .......... ...G...... ..T....A.. ...T...... .......
J-7        ......C... .......... .......... ...T...... .......
J-8        .......... ....G..... ........C. A..T...... .......
AY914854   .......... .......... .......... ...T...... .......
AY772274   ......... .......... .......... ....T..... ........
EF420981.1 ....CC... .T........ .......... ....T..... ..CC..AC
```

Figure 1. Variable sites at aligned nucleotide sequences of the IGS region of 23 specimens of *Arctium lappa* L. from Gueilai (G) and Jiali (J). Three sequences, AY914854, AY772274, and EF420981.1, were obtained from the NCBI GenBank and were described in "Materials and methods". Gueilai-4, -6, -14, and AY914854 shared the identical haplotype.

Population dynamics

Tajima's (1989) and Fu's (1997) tests are based on an infinite-site model without recombinations, which is appropriate for short DNA sequences. The significance of the statistics from the 2 tests is calculated by generating random samples under a hypothesis of selective neutrality and population equilibrium, using a coalescent simulation algorithm adapted from Hudson (1990). A significant *D* value can be due to factors other than selective effects, like population expansion, bottlenecks, or heterogeneity of mutation rates (Tajima, 1996). The Fu's (*F*s) statistic is very sensitive to population demographic expansion, which generally leads to large negative values. The mismatch distribution is the distribution of the observed number of differences between pairs of haplotypes. The distribution is usually multimodal in samples drawn from populations at demographic equilibrium, but it is usually unimodal in populations that have passed through a recent demographic expansion (Rogers and Harpending, 1992). A minimum spanning tree (MST) was computed from the matrix of pairwise distances calculated between all pairs of haplotypes using a modification of the algorithm described by Rohlf (1973).

RESULTS AND DISCUSSION

The trnL-trnF IGS region of *A. lappa* L. was sequenced and ranged from 805 to 808 bp long among 23 specimens. The sequence contained the partial trnL gene, the complete trnL-trnF IGS, and the partial trnF gene. *Bidens pilosa* of the Compositae exhibited two copies of a 21 bp repeating unit in the IGS (Tsai et al., 2008), but there were no repeating units observed in the IGS region of *A. lappa* L.

Overall, 21 haplotypes were determined from 23 specimens with a high haplotype diversity (0.91). The consensus sequence was 813 bp in total length. Forty variable (polymorphic) sites included 35 single-variation sites and 4 parsimoniously informative site. In total, one indel was found within these sequences (Figure 1). The average number of differences (*k*) and mean nucleotide diversity (P_i) among the 21 haplotypes were 3.23 and

Table 1. Intergenic spacer (IGS) genetic distance (above the diagonal) and different nucleotide numbers (below the diagonal) between haplotypes. Sequence AY914854 is an IGS homological sequence of *Arctium lappa* from NCBI. Two sequences, AY772274 and EF420981.1, were described in "Materials and methods". G-4, -6, -14, and AY914854 shared the identical haplotype.

	G-1	G-2	G-3	G-5	G-7	G-8	G-9	G-10	G-11	G-12	G-13	G-15	J-1	J-2	J-3	J-4	J-5	J-6	J-7	J-8	J-9	AY914854	AY772274	EF 420981.1
G-1	***	0.004	0.002	0.005	0.002	0.002	0.002	0.002	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-2	3	***	0.004	0.006	0.004	0.004	0.004	0.004	0.008	0.004	0.005	0.006	0.005	0.005	0.007	0.008	0.006	0.006	0.004	0.006	0.004	0.002	0.002	0.11
G-3	2	3	***	0.005	0.002	0.002	0.002	0.002	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-5	4	5	4	***	0.005	0.005	0.005	0.004	0.009	0.002	0.006	0.007	0.006	0.006	0.009	0.009	0.007	0.007	0.005	0.007	0.005	0.004	0.004	0.13
G-7	2	3	2	4	***	0.002	0.002	0.002	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-8	2	3	2	4	2	***	0.002	0.002	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-9	2	3	2	4	2	2	***	0.002	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-10	2	3	2	3	2	2	2	***	0.006	0.002	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-11	5	6	5	7	5	5	5	5	***	0.006	0.008	0.009	0.008	0.008	0.01	0.01	0.009	0.009	0.006	0.009	0.006	0.005	0.005	0.14
G-12	2	3	2	2	2	2	2	2	5	***	0.004	0.005	0.004	0.004	0.006	0.006	0.005	0.005	0.002	0.005	0.002	0.001	0.001	0.10
G-13	3	4	3	5	3	3	3	3	6	3	***	0.006	0.005	0.005	0.007	0.008	0.006	0.006	0.004	0.006	0.004	0.002	0.002	0.11
G-15	4	5	4	6	4	4	4	4	7	4	5	***	0.006	0.006	0.007	0.009	0.007	0.008	0.005	0.007	0.005	0.004	0.004	0.13
J-1	3	4	3	5	3	3	3	3	6	3	4	5	***	0.005	0.007	0.008	0.005	0.006	0.004	0.006	0.004	0.002	0.002	0.11
J-2	3	4	3	5	3	3	3	3	6	3	4	5	4	***	0.007	0.008	0.006	0.006	0.004	0.006	0.004	0.002	0.002	0.11
J-3	5	6	5	7	5	5	5	5	8	5	6	6	6	6	***	0.01	0.009	0.009	0.006	0.009	0.006	0.005	0.005	0.14
J-4	5	6	5	7	5	5	5	5	8	5	6	7	6	6	8	***	0.009	0.009	0.006	0.009	0.006	0.005	0.005	0.14
J-5	4	5	4	6	4	4	4	4	7	4	5	6	4	5	7	7	***	0.007	0.005	0.007	0.005	0.004	0.004	0.13
J-6	4	5	4	6	4	4	4	4	7	4	5	6	5	5	7	7	6	***	0.005	0.007	0.005	0.004	0.004	0.13
J-7	2	3	2	4	2	2	2	2	5	2	3	4	3	3	5	5	4	4	***	0.005	0.002	0.001	0.001	0.10
J-8	4	5	4	6	4	4	4	4	7	4	5	6	5	5	7	7	6	6	4	***	0.005	0.004	0.004	0.13
J-9	2	3	2	4	2	2	2	2	5	2	3	4	3	3	5	5	4	4	2	4	***	0.001	0.001	0.10
AY914854	1	2	1	3	1	1	1	1	4	1	2	3	2	2	4	4	3	3	1	3	1	***	0	0.009
AY772274	1	2	1	3	1	1	1	1	4	1	2	3	2	2	4	4	3	3	1	3	1	0	***	0.009
EF 420981.1	8	9	8	10	8	8	8	8	11	8	9	10	9	9	11	11	10	10	8	10	8	7	7	***

0.004, respectively. Intraspecific and interspecific genetic distances of IGS sequences ranged from 0 to 0.01 and 0 (*A. minus* vs. G-4, G-6, and G-14) to 0.014 (*S. tomentosa* vs. G-11, J-3, and J-4). Different nucleotide numbers within species ranged from 0 to 8 and between species ranged from 0 to 11 (Table 1). All 17 polymorphic sites were found in 13 haplotypes of 15 specimens from Gueilai. An identical haplotype was shared by 3 specimens from Gueilai, k and P_i values in the sample from Gueilai were 2.74 and 0.003, and in the sample from Jiali were 4 and 0.005, respectively. All 16 polymorphic sites were found in the 8 specimens from Jiali. No fixed difference was found between the 2 sets of samples.

Average numbers of nucleotide differences and nucleotide substitution per site (D_{xy}) between the 2 sets of samples were estimated to be 3.38 and 0.004, respectively.

After alignment, the length of the consensus sequence was 822 bp. Percentages of 4 nucleotides differed (A, 30.95 to 31.20%; T, 34.77 to 34.90%; G, 17.02 to 17.14%; and C, 16.89 to 17.02%) with a high AT-rich region in all sequences. All specimens from the 2 sets of samples were scattered over the NJ tree (Figure 2a). The interspecific genetic distance of IGS sequences between *A. lappa* L. and *A. minus* (0 to 0.005) overlapped with intraspecific genetic distances (0 to 0.01). The most common haplotype was identical with some individuals and the congeneric *A. minus* which suggests interspecific parallel evolution of the IGS. Compared to *B. pilosa* of the Compositae, *A. lappa* L. had a higher h_d value. Intraspecific genetic distances of the IGS in *A. lappa* L. were

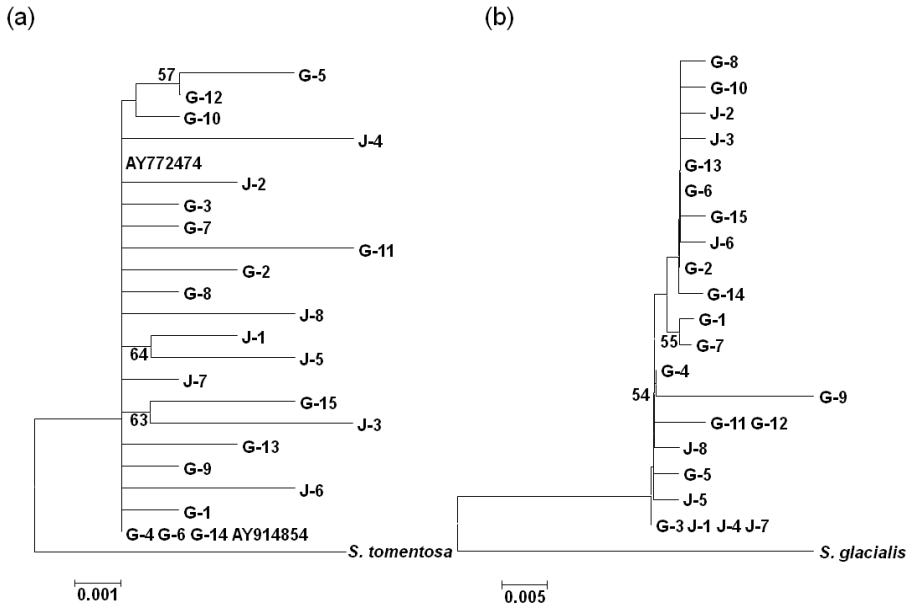

Figure 2. Neighbor-joining trees of (a) intergenic spacer (IGS) and (b) internal transcribed spacer (ITS) were constructed using the genetic distance method. Bootstrap values of > 50% (of 1000 replicates) are shown at the nodes.

```
[                    1111 1111111111 1111222222 2223333333 ]
[                 666880012 3334455777 8899015577 9990001234 ]
[                5058584968 3571958158 4703373945 2361264667
G-1              GTAATGAGTT CTCGCATCCA TGTTCCGCGC CGCCGCCGAC
G-2              .....A.... .......... .......... T.........
G-4              .....A.... .......... .......... ..........
G-5              .....A.... .......... .....T.... ..........
G-6              .....A.... .......... .......... T.........
G-7              .......... .......... .......... T.........
G-8              .....A.... .........G .......... T.........
G-9              ...G.A.... .......... .A....A... .......AGT
G-10             .....A.... .......... ........A. T.........
G-11_G-12        .....A.... ...A...... .......... ...T......
G-13             .....A.... .......... .......... T.........
G-14             .....A.... ....T..... .......... T.........
G-15              ..G..A.... .......... .......... T.........
J-2               .C...A.... .......... .......... T.........
J-3               .....A.... ......C... .......... T.........
J-5               .....A.... .......... .......... ....A.....
J-6               .....A.... .......... .........T T.........
G-3_J-1_J-4_J-7   .....A.... .......... .......... ..........
J-8               .....A.... T......... .......... ..........
AB118121.1        A...CAGACC .CT.-G.TT. C.CCT..T.. .AT..TT...
```

Figure 3. Variable sites at aligned nucleotide sequences of the internal transcribed spacer (ITS) region of 23 specimens of *Arctium lappa* L. from Gueilai (G) and Jiali (J). The sequence, AB118121.1 (*Saussurea glacialis*), was obtained from NCBI GenBank. G11 and G12 shared an identical sequence; G3, J1, J4, and J7 shared an identical sequence.

Table 2. Intraspecific and interspecific Kimura 2-parameter (K2P) genetic distances above the diagonal and different nucleotide numbers below the diagonal.

	G-1	G-2	G-4	G-5	G-6	G-7	G-8	G-9	G-10	G-11	G-13	G-14	G-15	J-2	J-3	J-5	J-6	J-7	J-8	AB118121.1
G-1	***	0.006	0.003	0.006	0.006	0.003	0.008	0.02	0.008	0.008	0.006	0.008	0.008	0.008	0.008	0.006	0.008	0.003	0.006	0.063
G-2	2	***	0.003	0.006	0	0.003	0.003	0.02	0.003	0.008	0	0.003	0.003	0.003	0.003	0.006	0.003	0.003	0.006	0.063
G-4	1	1	***	0.003	0.003	0.006	0.006	0.017	0.006	0.006	0.003	0.006	0.006	0.006	0.006	0.003	0.006	0	0.003	0.06
G-5	2	2	1	***	0.006	0.008	0.008	0.02	0.008	0.008	0.006	0.008	0.008	0.008	0.008	0.006	0.008	0.003	0.006	0.063
G-6	2	0	1	2	***	0.003	0.003	0.02	0.003	0.008	0	0.003	0.003	0.003	0.003	0.006	0.003	0.003	0.006	0.063
G-7	1	1	2	3	1	***	0.006	0.023	0.006	0.011	0.003	0.006	0.006	0.006	0.006	0.008	0.006	0.006	0.008	0.066
G-8	3	1	2	3	1	2	***	0.023	0.006	0.011	0.003	0.006	0.006	0.006	0.006	0.008	0.006	0.006	0.008	0.067
G-9	7	7	6	7	7	8	8	***	0.023	0.023	0.02	0.023	0.023	0.023	0.023	0.02	0.023	0.017	0.02	0.081
G-10	3	1	2	3	1	2	2	8	***	0.011	0.003	0.006	0.006	0.006	0.006	0.008	0.006	0.006	0.008	0.067
G-11	3	3	2	3	3	4	4	8	4	***	0.008	0.011	0.011	0.011	0.011	0.008	0.011	0.006	0.008	0.066
G-13	2	0	1	2	0	1	1	7	1	3	***	0.003	0.003	0.003	0.003	0.006	0.003	0.003	0.006	0.063
G-14	3	1	2	3	1	2	2	8	2	4	1	***	0.006	0.006	0.006	0.008	0.006	0.006	0.008	0.063
G-15	3	1	2	3	1	2	2	8	2	4	1	2	***	0.006	0.006	0.008	0.006	0.006	0.008	0.067
J-2	3	1	2	3	1	2	2	8	2	4	1	2	2	***	0.006	0.008	0.006	0.006	0.008	0.066
J-3	3	1	2	3	1	2	2	8	2	4	1	2	2	2	***	0.008	0.006	0.006	0.008	0.066
J-5	2	2	1	2	2	3	3	7	3	3	2	3	3	3	3	***	0.008	0.003	0.006	0.063
J-6	3	1	2	3	1	2	2	8	2	4	1	2	2	2	2	3	***	0.006	0.008	0.066
J-7	1	1	0	1	1	2	2	6	2	2	1	2	2	2	2	1	2	***	0.003	0.06
J-8	2	2	1	2	2	3	3	7	3	3	2	3	3	3	3	2	3	1	***	0.063
AB118121.1	21	21	20	21	21	22	22	26	22	22	21	21	22	22	22	21	22	20	21	***

G11 and G12 shared an identical sequence; G3, J1, J4, and J7 shared an identical sequence.

similar to those of other species, e.g., *Jatropha* spp. (Sudheer-Pamidimarri et al., 2009).

The total length of the ITS ranged from 361 to 62 bp. The mean percentages of 4 nucleotides were 24.36% for A, 26.01% for C, 30.06% for G, and 19.58% for T with a high GC-rich region. In total, 19 haplotypes (Figure 3) were determined from 23 ITS sequences of *A. lappa* L. with a haplotype diversity of 0.83. Twenty-seven variable sites contained 22 singleton variable sites and 5 parsimoniously informative sites. An average of 3.37 differences in nucleotides was present between sequences. The nucleotide diversity (P_i) was 0.009. Thirteen haplotypes with 16 polymorphic sites were determined among the 15 ITS-1 sequences from the Gueilai sample. Intraspecific and interspecific ITS genetic distances respectively ranged from 0 to 0.023 and 0.06 to 0.081. Different nucleotide numbers within and between species respectively ranged from 0 to 8 and 20 to 26 (Table 2). Values of k and P_i were 4.019 and 0.011 respectively. Seven haplotypes were discovered in 8 ITS-1 sequences from the Jiali sample with a haplotype diversity of 0.875. All 7 polymorphic sites were located among these haplotypes. Values of k and P_i were 2.21 and 0.006, respectively. Only 2 polymorphic mutations were shared by both samples. Average numbers of nucleotide differences (k) and substitutions per site (D_{xy}) between the 2 sets of samples were 3.038 and 0.009, respectively. The K2P distance between the 2 samples was 0.01 ± 0.002. All specimens from the 2 sets of samples were scattered over the NJ tree (Figure 2b). Pairwise *F*st between the 2 samples was estimated to be -0.027 and was not significant ($p = 0.773$). The insignificant genetic differentiation between the Gueilai and Jiali samples indicated that these samples shared a recent common

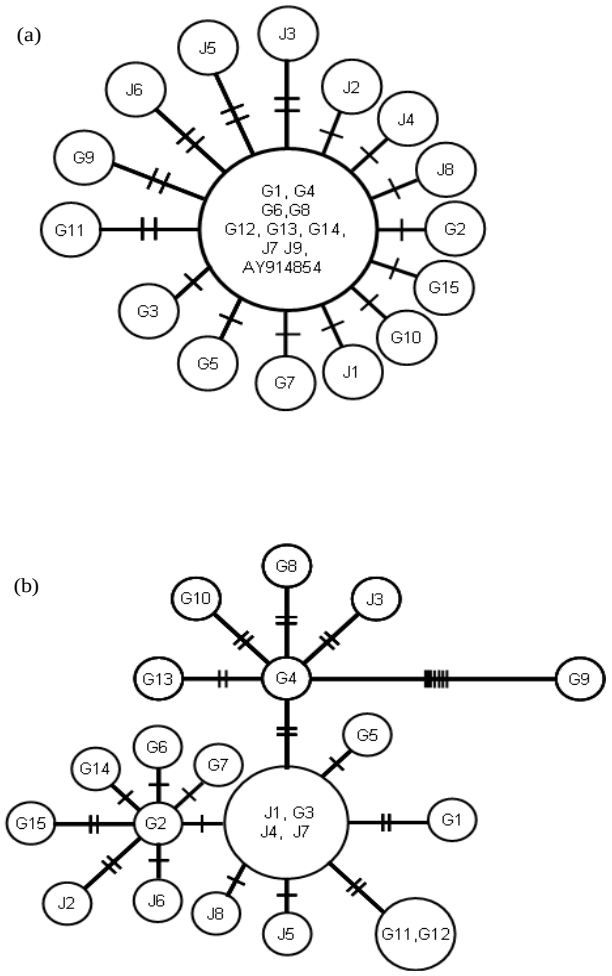

Figure 4. Minimum spanning trees of *Arctium lappa* L. reconstructed by (a) intergenic spacer (IGS) and (b) internal transcribed spacer (ITS) nucleotide sequences. The size of the circles indicates the number of repetitive units belonging to the same haplotype. The length of the branches indicates the numbers of different nucleotides.

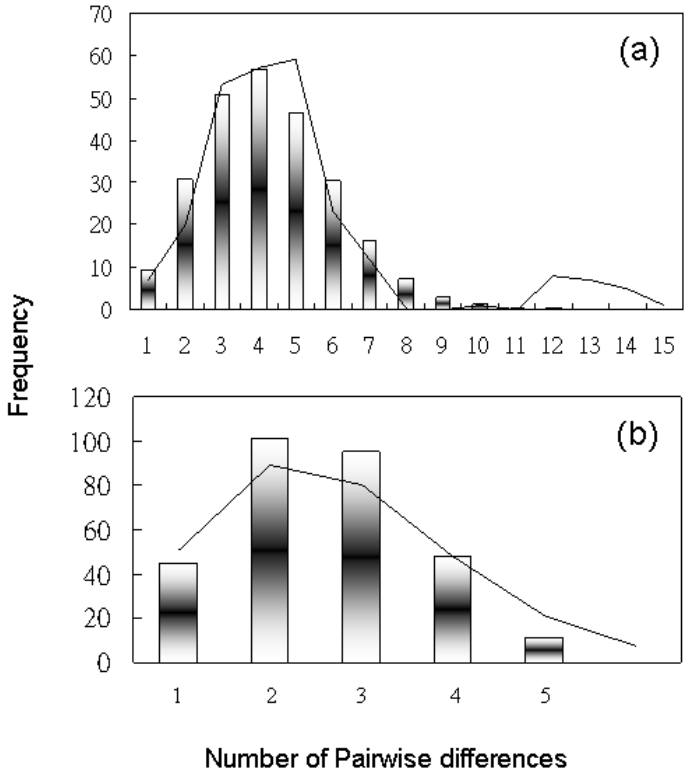

Figure 5. Mismatch distributions obtained from (a) intergenic spacer (IGS) and (b) internal transcribed spacer (ITS) data. Bars of the histogram represent the observed pairwise differences under a sudden expansion model. The curve is the expected distribution.

ancestor. The NJ tree also presented the same result, that all individuals were scattered on the tree with low bootstrapping values.

The most common haplotype was located at the centers of the 2 MSTs, and most haplotypes were located at the tips (Figure 4), which suggests that *A. lappa* L. had experienced a population expansion. Tajima's *D* values for the IGS and ITS were - 2.42 (p = 0) and - 2.033 (p = 0.01), while Fu's *F*s values for the IGS and ITS were - 27.58 (p = 0) and - 25.86 (p = 0), respectively. Both neutrality tests presented a statistically significant negative value, which implies that the populations experienced a historical expansion. Mismatch distributions of pairwise nucleotide differences of both the IGS and ITS appeared to be unimodai (Figure 5). Small values for Harpending's raggedness index of 0.076 and 0.051 from the IGS and ITS data were associated with a lack of significance (p = 0.282 and p = 0.208) which further supports a population expansion.

In the past, interspecific and intraspecific genetic diversities were ignored in the economic species *A. lappa* L. In this study, we observed that many variable sites were monomorphic in these specimens, which suggests that the population has accumulated different point mutations for a long time after a bottleneck. Both MSTs of the IGS and ITS demonstrated a common haplotype located at the center and all terminal haplotypes around the center. This reticulate phylogeny indicates that *A. lappa* L. underwent a historical population expansion effect. Results from both Tajima's and Fu's neutrality tests and mismatch distribution also agreed with a hypothesis of a historical population expansion. The present study proves the usefulness of the nrDNA ITS sequence in germplasm identification of *A. lappa* L. and paves the way for further studies on intraspecific populations, their phylogenetic origins, and evolutionary studies.

ACKNOWLEDGMENTS

We are extremely grateful to YW Shih and NC Chen for their help with laboratory work.

REFERENCES

Chase MW, Fay MF (2009). Barcoding of plants and fungi. Science, 325: 682-683.

Chen FA, Lee SC, Chao HR, Fu WC, Hsu MC, Horng CT, Wang CC, Matsui HA (2009). Effects of a burdock extract preparation on gastric mucosal protection. Asian J. Chem. 21:3015-3022.

Chen FA, Wu AB, Chen CY (2004). The influence of different treatments on free radical scavenging activity of burdock and the variation of its active components. Food Chem. 86:479-484.

Chow LW, Wang SJ, Duh PD (1997). Antibacterial activity of burdock. Food Sci. 24:195-202.

Cicero AFG, Derosa G, Gaddi A (2004). What do herbalists suggest to diabetic patients in order to improve glycemic control? Evaluation of scientific evidence and potential risks. Acta Diabetol. 41:91-98.

Dos Santos AC, Baggio CH, Freitas CS, Lepieszynski J, Mayer B, Twardowschy A, Missau FC, Dos Santos EP, Pizzolatti MG, Marques MC (2008). Gastroprotective activity of the chloroform extract of the roots from *Arctium lappa* L. J. Pharm. Pharmacol. 60:795-801.

Duh PD (1998). Antioxidant activity of burdock (*Arctium lappa* Linne): its scavenging effect on free-radical and active oxygen. J. Am. Oil Chem. Soc. 75:455-461.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564-567.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evol. 39:783-791.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 147:915-925.

Guzmán B, Vargas P (2005). Historical biogeography and character evolution of Cistaceae (Malvales) based on analysis of plastid *rbc*L and *trn*L-*trn*F sequences. Mol. Phylogenet. Evol. 37:644-660.

Hillis DM, Bull JJ (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182-192.

Hudson RR (1990). Gene genealogies and the coalescent process. In Futuyama and Antonovics (eds) Oxford Surveys in Evolutionary Biology, Oxford University, New York. pp. 1-44.

Kasai H, Fukada S, Yamaizumi Z, Sugie S, Mori H (2000). Action of chlorogenic acid in vegetable and fruits as an inhibitor of 8-hydroxydeoxyguanosine formation *in vitro* and in a rat carcinogenesis model. Food Chem. Toxicol. 38:467-471.

Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120.

Knipping K, van Esch EC, Wijering SC, van der Heide S, Dubois AE, Garssen J (2008). *In Vitro* and *in Vivo* anti-allergic effects of *Arctium lappa* L. Exp. Biol. Med. (Maywood), 233:1469-1477.

Kumar SK, Nei T (2004). MEGA 3: Integrated Software for Molecular Evolution Genetics Analysis and Sequence Alignment. Briefings Bioinf. 5:150-163.

Lee JY, Kim CJ (2010). Arctigenin, a phenylpropanoid dibenzyl-butyrolactone lignan, inhibits type I-IV allergic inflammation and pro-inflammatory enzymes. Arch. Pharm. Res. 33:947-957.

Li D, Kim JM, Jin Z, Zhou J (2008). Prebiotic effectiveness of inulin extracted from edible burdock. Anaerobe 14:29-34.

Lin CC, Lin JM, Yang JJ, Chuang SC, Ujiie T (1996). Anti-inflammatory and radical scavenging effect of *Arctium lappa*. Am. J. Chin. Med. 24:127-137.

Lin SC, Chung TC, Lin CC, Ueng TH, Lin YH, Lin SY, Wang LY (2000). Hepatoprotective effects of *Arctium lappa* on carbon tetrachloride- and acetaminophen-induced liver damage. Am. J. Chin. Med. 28:163-173.

Lin SC, Lin CH, Lin CC, Lin YH, Chen CF, Chen IC, Wang LY (2002). Hepatoprotective effects of *Arctium lappa* Linne on liver injuries induced by chronic ethanol consumption and potentiated by carbon tetrachloride. J. Biomed. Sci. 9:401-409.

Matsumoto T, Hosono-Nishiyama K, Yamada H (2006). Antiproliferative and apoptotic effects of butyrolactone lignans from *Arctium lappa* on leukemic cells. Planta Med. 72:276-278.

Miele C, Beguinot F (2012). New expectations from the well-known medicinal properties of *Arctium lappa*. Diabetologia DOI 10.1007/s00125-012-2503-7.

Morita K, Kada T, Namiki M (1984). A desmutagenic factor isolated from burdock. Mutation Res. 129:25-31.

Morita K, Nishijima Y, Kada T (1985). Chemical nature of a desmutagenic factor from burdock (*Arctium lappa* Linne). Agric. Biol. Chem. 49:925-932.

Murata Y, Kawabara J, Niki R (1995). Antioxidative caffeoylquinic acid derivatives in the roots of burdock. J. Agric. Food. Chem. 43:2592-2595.

Neves SS, Watson MF (2004). Phylogenetic relationships in *Bupleurum* (Apiaceae) based on nuclear ribosomal DNA ITS sequence data. Ann. Bot. 93:379-398.

Ouborg NJ, Piquot Y, Goenendael MV (1999). Population genetics, molecular markers and the study of dispersal in plant. J. Ecol. 87: 551-568.

Raymond M, Rousset F (1995). An exact test for population differentiation. Evolution 49:1280-1283.

Rogers AR, Harpending H (1992). Population growth makes waxes in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9:552-569.

Rohlf FJ (1973). Algorithm 76 Hierarchical clustering using the minimum spanning tree. Comput. J. 16:93-95.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas, R (2003). DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496-2497.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular Cloning, A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York.

Sudheer-Pamidimarri DV, Chattopadhyay B, Reddy MP (2009). Genetic divergence and phylogenetic analysis of genus *Jatropha* based on nuclear ribosomal DNA ITS sequence. Mol. Biol. Rep. 36:1929-1935.

Szczepaniak M, Cieślak E, Bednarek PT (2002). Morphological and AFLP variation of *Elymus repens* (L.) Gould (Poaceae). Cell Mol. Biol. Lett. 7:547-558.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tajima F (1996). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. Genetics 143:1457-1465.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucl. Acids Res. 22:4673-4680.

Tsai LC, Wang JC, Hsieh HM, Liu KL, Linacre AMT, Lee JCI (2008). Bidens identification using the noncoding regions of chloroplast genome and nuclear ribosomal DNA. Foren. Sci. Int. Gen. 2:35-40.

Wei D (2006). A study on anti-weary function of the burdock extract. J. Anhui Agric. Sci. 34:3171-3172.

Wei D, Wang LC (2006). Research development of burdock root. J. Anhui Agric. Sci. 34:3716-3717.

Wen CS, Hsiao JY (1999). Genetic differentiation of *Lilium longiflorum Thunb. var. scabrum* Masam. (Liliaceae) in Taiwan using random amplified polymorphic DNA and morphological characters. Bot. Bull. Acad. Sin. 40:65-71.

Xia Q, Zhao KJ, Huang ZG, Zhang P, Dong TTX, Li SP, Tsim KWK (2005). Molecular genetic and chemical assessment of *Rhizoma curcumae* in China. J. Agric. Food Chem. 53:6019-6026.

Yao X, Zhu F, Zhao Z, Liu C, Luo L, Yin Z (2011). Arctigenin enhances chemosensitivity of cancer cells to cisplatin through inhibition of the STAT3 signaling pathway. J. Cell Biochem. 112:2837-2849.