# Application of Molecular Markers in Predicting Production Quality of Cultivated *Cistanche deserticola*

Yan Wu,[a] Hai Ming Shi,[a] Zhong Bao,[b] Meng Yue Wang,[a] Peng Fei Tu,[b] and Xiao Bo Li*,[a]

[a] *School of Pharmacy, Shanghai Jiao Tong University; Shanghai 200240, P. R. China: and* [b] *Department of Natural Medicines, School of Pharmaceutical Sciences, Peking University Health Science Center; Beijing 100083, P. R. China.*
Received September 2, 2009; accepted November 10, 2009; published online November 20, 2009

**We developed a set of molecular markers in *Cistanche deserticola* Y. C. Ma to evaluate the production quality of cultivated *C. deserticola* individuals. This application utilizes the inter-simple-sequence repeat (ISSR) polymerase chain reaction (PCR) and random amplified polymorphic DNA (RAPD) PCR as molecular markers to determine the echinacoside content in cultivated *C. deserticola* individuals. The unweighted pair-group method using arithmetic average clustering (UPGMA) confirmed that the combined ISSR and RAPD data could categorize all *C. deserticola* individuals into three groups according to their respective echinacoside content. The stepwise multiple regression analysis (SMRA) revealed six potential markers associated with echinacoside accumulation in *C. deserticola* and produced 18 echinacoside-marker prediction models, four of which were successfully used to predict the quality of *C. deserticola* from Neimenggu. Both clustering and SMRA showed a correlation between the echinacoside content and molecular markers in cultivated *C. deserticola*. The relative average deviation of prediction (RADP) of the prediction models could be improved by simplifying and adjusting the model. It was found that the RADP value could reach 2.6% after adjustment and the simplified prediction models could successfully predict the quality of cultivated *C. deserticola* individuals.**

**Key words** prediction model; echinacoside; inter-simple-sequence repeat; random amplified polymorphic DNA; *Cistanche deserticola*

*Cistanche deserticola* Y. C. Ma, a member of Orobanchaceae family, is a holoparasite mainly distributed in Neimenggu and Xinjiang, northwest of China. The dried fleshy stems of the parasitic plant are known as Herba Cistanches. In China, Herba Cistanches is a commonly prescribed traditional Chinese medicine to treat various diseases including impotence, female infertility, morbid leukorrhea, profuse metrorrhagia, cold sensation in the loins and knees, and chronic constipation in the aged.[1)] Nowadays, it is reported to enhance sex and learning behavior,[2)] and to possess the sedative and vasorelaxant effects.[3,4)] In addition, Herba Cistanches is a healthy food supplement for men's care in Japan and Southeast Asia. The ingredient associated with the pharmacological activities of Herba Cistanches is echinacoside, which has significant vasorelaxant, neuroprotective, hepatoprotective effects.[4—6)] The echinacoside content is commonly used as a quality control marker of Herba Cistanches, therefore.

On the other hand, due to its wide medical use and consequent over exploitation, the natural source of *C. deserticola* has become rare and has been listed as a class II endangered species in China. Nowadays, the plants cultivated in Neimenggu and Xinjiang have become the main source of *C. deserticola*. However, the production quality of cultivated plant varies resulting from lack of standards for selecting high quality cultivars.

Harnessing genetic variability by adopting conventional breeding strategies entails a huge investment of time and resources. To speed up progress in classical breeding programs and detecting of high quality plants at a very early stage of cultivation, it is important to identify certain DNA markers related to genomic regions for traits such as quality and productivity. This would enable the breeders to make selections among seedlings grown in a non-target environment.[7)]

As the greater adoption of marker-assistant selection (MAS) is inevitable,[8)] molecular markers associated with quantity traits is widely studied in recent years. Random amplified polymorphic DNA (RAPD)-linked markers associated with production yield and its components in *Miscanthus sinensis*,[9)] fruit sugar concentration in tomato,[10)] *S*-glycoprotein in *Brassica campestris*,[11)] and inter-simple-sequence repeat (ISSR) markers linked to yield traits in mulberry,[12)] horticultural traits for watermelon,[13)] seasonal flowering locus in *Fragaria* have been successfully identified. Also the prediction models for quantity traits have been effectively applied in the crop productions, such as prediction models for mite resistance with molecular markers in coconut,[14)] genetic and environmental effects on chemical composition related to sensory traits in common beans,[15)] and prediction of yield component performances in hybrid rice using molecular markers technology.[16,17)] However, there are no reports on prediction models or specific molecular markers associated with quantity traits in *C. deserticola*.

Our previous chemical and genetic analyses of *Cistanche* species have shown a close relation between the bioactive compounds and ISSR markers.[18)] The objectives of this paper were to: (1) develop prediction models of echinacoside content with RAPD and ISSR markers in naturally occurring genotypes of *C. deserticola*; (2) find out the potential molecular markers associated with echinacoside accumulation in *C. deserticola* and (3) apply them into the breeding program.

## MATERIALS AND METHODS

**Plant Material and Pre-processing** A total of 44 *C. deserticola* individuals were sampled from Xinjiang and Neimenggu, authenticated by one of the authors (Prof. Pengfei Tu), and deposited at the School of Pharmacy, Shanghai Jiao Tong University. Among the 44 samples, 21

* To whom correspondence should be addressed. e-mail: xbli@sjtu.edu.cn

individuals (group A) from Xinjiang were collected in 2006, and 23 individuals (group B) from Neimenggu were collected in 2 different years, 9 individuals (group $B_1$) and 7 individuals (group $B_2$) collected in 2005, and 7 individuals (group $B_3$) in 2007. Each plant sample was pre-processed according to our previous method.[18] Genomic DNA of individuals was isolated from their fresh tissues using an extraction protocol by DNeasy Plant mini Kit (QIAGEN Inc.).

**Content Determination** The HPLC method for determination of echinacoside content in *C. deserticola* was developed and validated in our previous study.[18]

**Genetic Analysis** The polymerase chain reaction (PCR) was performed in a total volume of 25 $\mu$l reaction solution using MJ-Research Thermocycler (Model PTC-100). The amplified products were resolved on 2% agarose gel containing ethidium bromide (0.5 mg/ml) in 1×TAE (Tris–acetate-ethylene diamine tetraacetic acid (EDTA)) buffer, and bands were captured using a BIODoc-It$^{TM}$ System (UVP Inc.) The amplification products were scored in terms of a binary code as present (1) or absent (0), regardless of their intensity.

**RAPD** Each reaction was composed of 50 ng of template DNA, 1×PCR buffer, 200 $\mu$M $MgCl_2$, 100 $\mu$M each of deoxyribonucleotide triphosphate (dNTPs), 0.2 $\mu$M random decamer primers and 2 units *Taq* DNA polymerase (Sangon Inc., Shanghai, China). The RAPD PCR amplification was carried out with an initial denaturation at 94 °C for 5 min followed by 40 cycles, each step consisting of denaturing at 94 °C for 45 s, annealing at 34 °C for 1 min, and extension at 72 °C for 1 min with a final extension step at 72 °C for 5 min.

**ISSR** The ISSR PCR took place in a total volume of 25 $\mu$l containing 50 ng of template DNA, 1×PCR buffer, 200 $\mu$M $MgCl_2$, 100 $\mu$M each of dNTPs, 0.4 $\mu$M ISSR primers, and 2 units *Taq* DNA polymerase (Sangon Inc., Shanghai, China). The PCR amplification was carried out with an initial denaturation at 94 °C for 5 min followed by 40 cycles, each step consisting of denaturing at 94 °C for 50 s, annealing at 54 °C for 1 min, and extension at 72 °C for 1 min with a final extension step at 72 °C for 8 min.

**Clustering** To study the genetic diversity of all individuals, a genetic identity matrix was constructed using the Nei's genetic distance, and a dendrogram, based on the constructed genetic identity matrix, were established using the unweighted pair group method of arithmetic averages (PopGen v.1.31).

**Stepwise Multiple Regression Analysis (SMRA)** Stepwise regression of molecular markers (ISSR and RAPD) against the echinacoside content was performed to identify suitable markers that would account for echinacoside accumulation. A multiple regression approach was adopted using the echinacoside content as a dependent variable and genotypic data as the independent variable. The analysis model was

$$Y=a+b_1[m_1]+b_2[m_2]+b_3[m_3]+\ldots+e$$

Where $Y$ represents the echinacoside content, $m_j$ the RAPD and ISSR markers, $b_j$ the partial regression coefficients that specify the empirical relationship between $Y$ and $m_j$, and $e$ is the random error of $Y$, including environmental variation. This allowed us to foresee the unknown echinacoside content of individuals (SPSS package v13.0).

In the development of prediction models, a full cross-validation was used to evaluate the quality of the models. The samples of two of groups $B_1$, $B_2$ and $B_3$, were used as calibration sets, and the remaining samples were used as the validation set to study the prediction power in the same area. The samples of one of groups A and B, were used as the calibration set, and the other as the validation set to study the prediction power in different areas. Groups A, $B_1$ and $B_2$ were used as the calibration sets, and group $B_3$ was used as the validation set to study the prediction power in partially same areas. Correlative relationship between RAPD and ISSR markers and echinacoside content was performed indi-

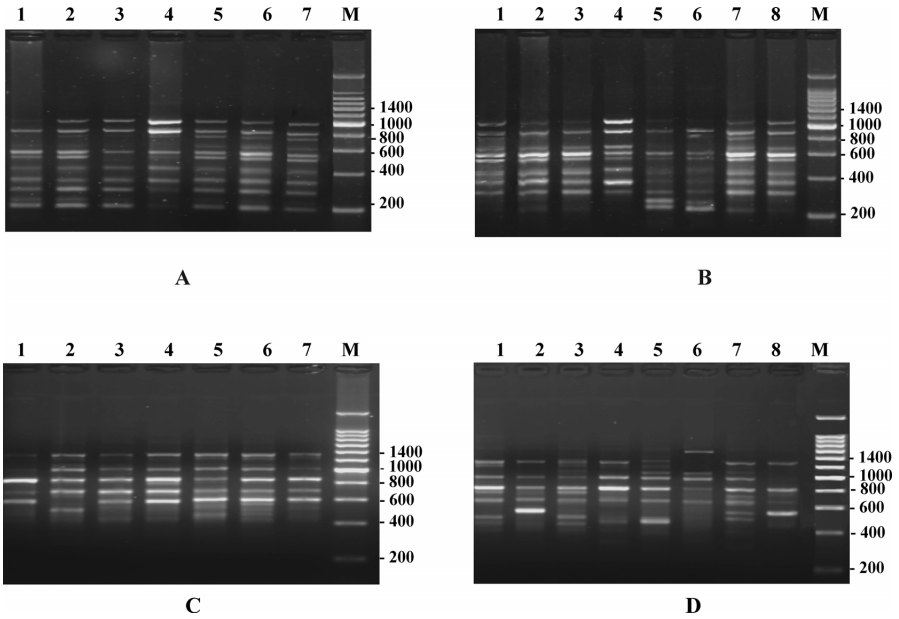


Fig. 1. RAPD and ISSR Fingerprint Patterns of *C. deserticola* Individuals from Neimenggu (A, C) and Xinjiang (B, D) Using the Primer S28 (A, B) and Primer I16 (C, D)

Lanes 1—7 of A and C represent CdeN-10 to CdeN-16. Lanes 1—8 of B and D represent CdeX-1 to CdeX-8. Lane M: 200 bp DNA Ladder.

vidually and collectively. The regression power of the model and its prediction performance were evaluated by the determination coefficient ($R^2$) and relative average deviation of prediction (RADP), respectively.

$$\mathrm{RADP} = \frac{\sum \frac{\left|\hat{C}_i - C_i\right|}{C_i} \times 100\%}{m}$$

Where $C_i$ represents the measured value of echinacoside, $\hat{C}_i$ the predictive value of echinacoside, and $m$ is the number of samples in the set. $R^2$ values assumed positive or negative values (parameter estimate, PE) indicating the association of the marker (bands of appropriate size) with increased or decreased content of echinacoside. The ideal model should have high $R^2$ value and low RADP value.

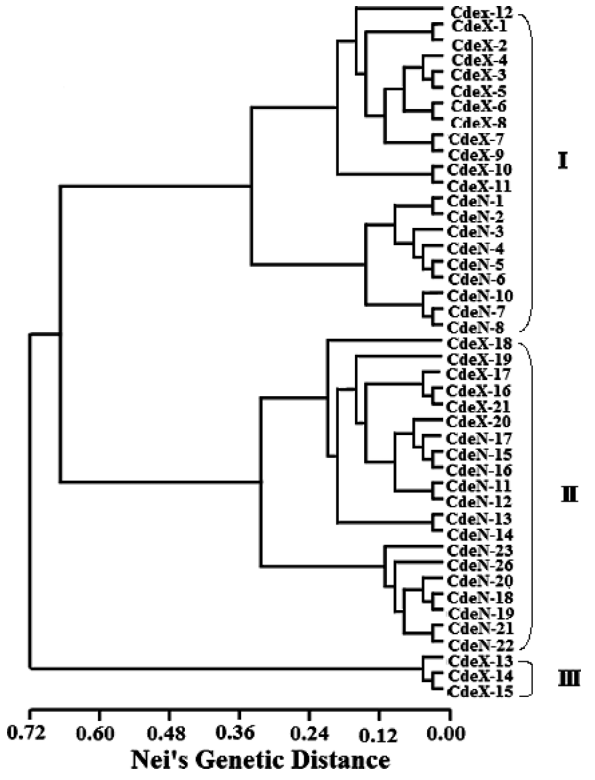


Fig. 2. UPGMA Dendrogram of 44 *C. deserticola* Individuals on the Basis of Combined ISSR and RAPD Fingerprints

## RESULTS AND DISCUSSION

**Content Determination** The analytical results of echinacoside content in each individual by HPLC were shown in Table 1.

**RAPD** Of the 52 random decamer primers used, 13

Table 1. Content of Echinacoside from *Cistanche deserticola*

| Sample | Geographical origins | Harvesting time | Content (mg/g) |
|---|---|---|---|
| CdeN-1 | Yabulai, Neimenggu | 1 May, 2005 | 21.0 |
| CdeN-2 | Yabulai, Neimenggu | 1 May, 2005 | 6.1 |
| CdeN-3 | Yabulai, Neimenggu | 1 May, 2005 | 19.9 |
| CdeN-4 | Yabulai, Neimenggu | 1 May, 2005 | 1.7 |
| CdeN-5 | Yabulai, Neimenggu | 1 May, 2005 | 12.9 |
| CdeN-6 | Yabulai, Neimenggu | 1 May, 2005 | 1.5 |
| CdeN-7 | Yabulai, Neimenggu | 1 May, 2005 | 6.7 |
| CdeN-8 | Yabulai, Neimenggu | 1 May, 2005 | 4.7 |
| CdeN-9 | Yabulai, Neimenggu | 1 May, 2005 | 4.2 |
| CdeN-10 | Ejinaqi, Neimenggu | 2 May, 2005 | 21.0 |
| CdeN-11 | Ejinaqi, Neimenggu | 2 May, 2005 | 90.4 |
| CdeN-12 | Ejinaqi, Neimenggu | 2 May, 2005 | 21.6 |
| CdeN-13 | Ejinaqi, Neimenggu | 2 May, 2005 | 44.0 |
| CdeN-14 | Ejinaqi, Neimenggu | 2 May, 2005 | 64.8 |
| CdeN-15 | Ejinaqi, Neimenggu | 2 May, 2005 | 31.5 |
| CdeN-16 | Ejinaqi, Neimenggu | 2 May, 2005 | 145.2 |
| CdeN-17 | Jilantai, Neimenggu | 9 May, 2007 | 77.2 |
| CdeN-18 | Jilantai, Neimenggu | 9 May, 2007 | 65.0 |
| CdeN-19 | Jilantai, Neimenggu | 9 May, 2007 | 256.2 |
| CdeN-20 | Jilantai, Neimenggu | 9 May, 2007 | 20.8 |
| CdeN-21 | Jilantai, Neimenggu | 9 May, 2007 | 27.1 |
| CdeN-22 | Jilantai, Neimenggu | 9 May, 2007 | 12.3 |
| CdeN-23 | Jilantai, Neimenggu | 9 May, 2007 | 46.8 |
| CdeX-1 | Jimusaer Xinjiang | 29 April, 2006 | 11.0 |
| CdeX-2 | Jimusaer Xinjiang | 29 April, 2006 | 16.1 |
| CdeX-3 | Jimusaer Xinjiang | 29 April, 2006 | 21.7 |
| CdeX-4 | Jimusaer Xinjiang | 29 April, 2006 | 44.0 |
| CdeX-5 | Jimusaer Xinjiang | 29 April, 2006 | 81.5 |
| CdeX-6 | Jimusaer Xinjiang | 29 April, 2006 | 25.9 |
| CdeX-7 | Jimusaer Xinjiang | 29 April, 2006 | 33.0 |
| CdeX-8 | Jimusaer Xinjiang | 29 April, 2006 | 71.2 |
| CdeX-9 | Jimusaer Xinjiang | 29 April, 2006 | 9.7 |
| CdeX-10 | Jimusaer Xinjiang | 29 April, 2006 | 8.8 |
| CdeX-11 | Jimusaer Xinjiang | 29 April, 2006 | 47.8 |
| CdeX-12 | Jimusaer Xinjiang | 29 April, 2006 | 27.5 |
| CdeX-13 | Jimusaer Xinjiang | 29 April, 2006 | 50.6 |
| CdeX-14 | Jimusaer Xinjiang | 29 April, 2006 | 130.7 |
| CdeX-15 | Jimusaer Xinjiang | 29 April, 2006 | 165.4 |
| CdeX-16 | Jimusaer Xinjiang | 29 April, 2006 | 12.6 |
| CdeX-17 | Jimusaer Xinjiang | 29 April, 2006 | 12.1 |
| CdeX-18 | Jimusaer Xinjiang | 29 April, 2006 | 55.7 |
| CdeX-19 | Jimusaer Xinjiang | 29 April, 2006 | 47.9 |
| CdeX-20 | Jimusaer Xinjiang | 29 April, 2006 | 31.2 |
| CdeX-21 | Jimusaer Xinjiang | 29 April, 2006 | 30.5 |

Table 2. Characteristics of Random Primers and ISSR Primers

| Primer | Sequence (5′–3′) | $T_m$, °C | Primer | Sequence (5′–3′) | $T_m$, °C |
|---|---|---|---|---|---|
| S17 | AGGGAACGAG | 36.9 | S1218 | CTACCGGCAC | 41.0 |
| S24 | AATCGGGCTG | 36.9 | S2015 | GAAGACCTGG | 36.9 |
| S28 | GTGACGTAGG | 36.9 | S2140 | TGGTACCTGG | 36.9 |
| S32 | TCGGCGATAG | 36.9 | ISSR16 | $(AG)_8CT$ | 54.0 |
| S40 | GTTGCGATCC | 36.9 | ISSR17 | $(AG)_8CA$ | 54.0 |
| S490 | TGTGCCCGAA | 36.9 | ISSR18 | $(GA)_8CT$ | 54.0 |
| S1101 | TCACGTACGG | 36.9 | ISSR20 | $(GA)_8CC$ | 56.0 |
| S1131 | GTCCATGCAG | 36.9 | ISSR54 | $(GA)_8C$ | 54.6 |
| S1137 | TCAGCACAGG | 36.9 | ISSR64 | $(AG)_8TC$ | 55.0 |
| S1165 | GACTTCAGGG | 36.9 | ISSR71 | $(GA)_8T$ | 52.2 |

primers, namely S17, S24, S28, S32, S40, S490, S1101, S1131, S1137, S1165, S1218, S2015 and S2140 (Table 2), were selected for further analysis. 213 bands ranging in size from 200 to 3600 bp, which were consistent, unambiguous and reproducible, were used in the statistical analysis (Fig. 1). The number of bands per primer varied from 13 to 17 with an average of 16 bands. Of these, 210 bands were polymorphic (98.6%).

**ISSR** Of the 8 ISSR primers, 7 primers, namely I16, I17, I18, I20, I54, I64 and I71 (Table 2), produced a total of 121 consistent, unambiguous and reproducible bands with sizes ranging from 200 to 2300 bp (Fig. 1), and were selected for further analysis. The number of bands per primer varied from 15 to 21 with an average of 17 bands. Of these, 118 bands were polymorphic (97.5%).

**Cluster Analysis** To investigate the correlation between the echinacoside content and genetic markers, cluster analysis was carried out before SMRA. The dendrogram based on combined RAPD and ISSR profiles formed 3 major groups, which were related to the echinacoside content. Genotypes from the first major group were lower echinacoside content, with an average of 22.7 mg/g, those from the second major group were moderate echinacoside content, with an average of 55.7 mg/g, and those from the third major group were higher echinacoside content, with an average of 115.6 mg/g (Fig. 2). The result further verified our previous chemical and genetic analysis of *C. deserticola*.[18)]

**SMRA** To identify putative markers associated with the echinacoside content, 18 prediction models were developed with RAPD and ISSR markers individually and collectively. Specifically, 9 models were used to predict the validation set from the same area as the calibration set, 6 models to predict two sets from different areas, and 3 models to predict two sets from partially the same areas.

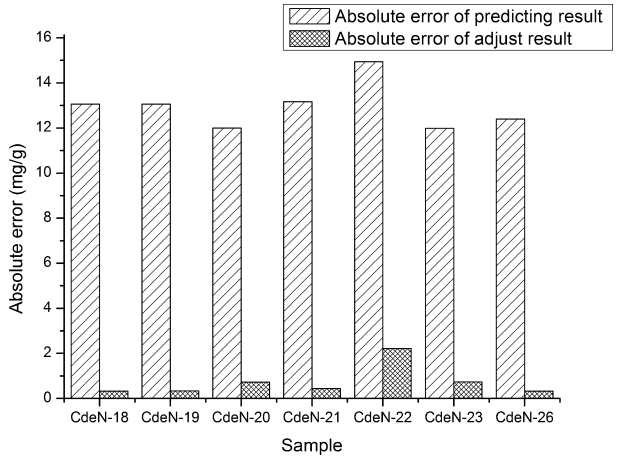


Fig. 3. The Absolute Error of Prediction Results and Adjusted Results of Model E-18 for 7 *C. deserticola* Individuals

As shown in Table 3, four models with $R^2$ values higher than 0.900 and RADP values lower than 100% were selected to predict individuals from certain areas. Specifically, two models were based on ISSR markers individually and two were based on ISSR and RAPD markers collectively. All four selected models used calibration and validation sets from the same or partially the same areas, therefore environmental effect on the accumulation of echinacoside content becomes an important factor. The phenotypes of individuals, especially the quantity traits, are the result of combined effects of genome and environment. When individuals are from the same area, the environmental effect is at a relatively same level and errors caused by the environmental differences can be minimized, and genetic difference may be detected. Models E-2 and E-14 were employed to predict echinacoside content of individuals from Ejinaqi, and models E-1 and E-18 to predict echinacoside content of individuals from Jilantai. In these four prediction models, E-18 showed the best prediction performance, with RADP value of 40.5%. However, the RADP values from other models, 72.1%, 95.8% and 75.3%, respectively, were not so satisfied for prediction. As herbal samples collected from different areas showed different genetic fingerprints, utilization of suitable primers to reveal specific genetic markers associated with echinacoside accumulation in *C. deserticola* might enhance the prediction power of these models.

Model E-18 contains 33 markers from 13 primers, which increase the difficulty to obtain the genetic data. Therefore, a simplification step was carried out to reduce the number of markers used in the model without significantly lowering the predicting power. By eliminating 14 markers, without $R^2$ change, a new model with 19 markers was generated (Table 4). Although eliminated 14 markers, the predicting power only reduced by 2.9%. To further increase the predicting power, a set of individuals with known content of echinacoside was added to the validation set to adjust the predicted value. The adjust model was as follows:

$$X_{adj}=X_{pre}+1/n\sum(Y_{mes}-Y_{pre})$$

$X_{adj}$ represents the adjusted value of echinacoside of experimental individuals, $X_{pre}$ the predicted value of experimental individuals, $Y_{mes}$ the measured value of individuals with known echinacoside content, $Y_{pre}$ the predicted value of individuals with known echinacoside content, and $n$ is the number of individuals with known echinacoside content. By adding individuals with known content, the prediction performance of model E-18 increased significantly (Fig. 3), with RADP value improved from 40.5 to 2.6%.

It is understood that the developed models were only suitable for the case in this study. The generalization and stabilization of the model should be improved, and more efforts

Table 3. Four Selected Prediction Models for Echinacoside

| Model | Calibration set | Validation set | Genetic markers | $R^{2\,a)}$ | RADP[b)] |
|---|---|---|---|---|---|
| E-1 | $B_1$ plus $B_2$ | $B_3$ | ISSR | 0.960 | 72.1% |
| E-2 | $B_1$ plus $B_3$ | $B_2$ | ISSR | 1.000 | 95.8% |
| E-14 | $B_1$ plus $B_3$ | $B_2$ | ISSR plus RAPD | 1.000 | 75.3% |
| E-18 | A and $B_1$ plus $B_2$ | $B_3$ | ISSR plus RAPD | 1.000 | 40.5% |

*a*) Determination coefficients. *b*) Relative average deviation of prediction.

Table 4. Details of Model E-18

| Factor | $Marker_{size}$ | Partial $R^{2a)}$ | Total $R^{2a)}$ | Prob.>$F$ | Parameter estimate |
|---|---|---|---|---|---|
| X1 | $S28_{554}$ | 0.676 | 0.676 | 0.000 | −178.967 |
| X2 | $S490_{1667}$ | 0.117 | 0.793 | 0.016 | 134.784 |
| X3 | $S24_{572}$ | 0.047 | 0.840 | 0.005 | 103.224 |
| X4 | $I54_{1267}$ | 0.045 | 0.885 | 0.002 | −18.133 |
| X5 | $S24_{1092}$ | 0.021 | 0.906 | 0.127 | 33.607 |
| X6 | $S1131_{1142}$ | 0.022 | 0.928 | 0.006 | −26.927 |
| X7 | $S2015_{620}$ | 0.015 | 0.943 | 0.010 | 19.850 |
| X8 | $S1131_{415}$ | 0.013 | 0.957 | 0.008 | −23.079 |
| X9 | $I20_{848}$ | 0.013 | 0.969 | 0.003 | 27.054 |
| X10 | $S2015_{370}$ | 0.008 | 0.978 | 0.005 | −21.062 |
| X11 | $S21_{1100}$ | 0.009 | 0.987 | 0.001 | 33.118 |
| X12 | $I16_{863}$ | 0.005 | 0.991 | 0.002 | 12.866 |
| X13 | $I54_{338}$ | 0.002 | 0.993 | 0.020 | −14.864 |
| X14 | $S32_{553}$ | 0.002 | 0.995 | 0.010 | −8.102 |
| X15 | $I20_{956}$ | 0.002 | 0.997 | 0.002 | 8.575 |
| X16 | $I20_{693}$ | 0.001 | 0.998 | 0.009 | 9.536 |
| X17 | $I18_{333}$ | 0.001 | 0.999 | 0.012 | −5.586 |
| X18 | $I16_{2227}$ | 0.000 | 0.999 | 0.013 | 8.057 |
| X19 | $S2015_{680}$ | 0.000 | 0.999 | 0.005 | 8.958 |
| Constant | | | | | 78.318 |
| Error | | | | | −0.954 |

*a*) $R^2$, determination coefficients.

should be made to avoid or reduce the overfitting problem for which there was no mathematically exact final solution. In order to achieve practical applications, a large number of *C. deserticola* samples from orthodox areas, different collecting years, and different batches of one same variety should be taken into consideration for a more general and stable model. Furthermore, cross-validation or other effective validation methods should be used in the calibration stage.

An attempt was made to assemble a suite of markers associated with echinacoside content for their possible use in MAS. SMRA using echinacoside content as the dependent variable and molecular markers as independent variables revealed 6 markers that might be associated with echinacoside accumulation, 3 RAPD markers and 3 ISSR markers, with all of the partial $R^2$ over 0.900. The PE indicated the strength of their influence on the trait. A negative value of PE indicated decrease of echinacoside accumulation, the numerical value indicating the strength of the association. Combined analysis suggests that $S490_{1667}$ and $I64_{841}$ were potential markers for enhancing echinacoside accumulation, and $S28_{554}$, $S1131_{700}$, $I18_{975}$ and $I18_{297}$ for decrease of echinacoside accumulation (Table 5). To identify these markers associated with echinacoside accumulation, bulked segregant analysis (BSA) should be carried out and more individuals should be included.[19)]

## CONCLUSION

The high quality of crude herbal drugs is a prerequisite for their clinical applications. The ability to predict potential contents of the bioactive compounds at an early stage of the plant growth is a possible way to control their production quality, to reduce the blindness of cultivation, and to help to select the superior varieties. Both clustering analysis and SMRA showed a correlation between the echinacoside content and molecular markers in cultivated *C. deserticola*. By SMRA analysis, four prediction models have been obtained

Table 5. Stepwise Multiple Regression Analysis for Content of Echinacoside Using the Combined Data of RAPD and ISSR Analysis

| SI No. | $Marker_{size}$ | Partial $R^{2a)}$ | Total $R^{2a)}$ | Prob.>$F$ | Parameter estimate |
|---|---|---|---|---|---|
| RAPD | | | | | |
| 1 | $S28_{554}$ | 0.394 | 0.394 | 0.000 | −128.374 |
| 2 | $S490_{1667}$ | 0.154 | 0.548 | 0.001 | 84.120 |
| 3 | $S1131_{700}$ | 0.137 | 0.685 | 0.000 | −48.280 |
| ISSR | | | | | |
| 1 | $I18_{975}$ | 0.382 | 0.382 | 0.000 | −83.368 |
| 2 | $I18_{297}$ | 0.109 | 0.491 | 0.005 | −35.539 |
| 3 | $I64_{841}$ | 0.098 | 0.589 | 0.004 | 37.692 |

*a*) $R^2$, determination coefficients.

for echinacoside content prediction with $R^2$ values higher than 0.900 and RADP values between 40.5% and 95.8%. By simplifying the prediction models, the complexity of obtaining the genetic data could be reduced without significantly altering the prediction performance. After adjustment, the prediction performance could be significantly enhanced with RADP value reaching 2.6%. The present study found a set of molecular markers, $S490_{1667}$, $S28_{554}$, $S1131_{700}$, $I18_{975}$, $I18_{297}$ and $I64_{841}$, that might link to the echinacoside accumulation and had the potential use for evaluating *C. deserticola* germ plasma at early stages as well as mother plants for breeding purposes.

**Acknowledgements** We thank the Special Program for Key Basic Research of the Ministry of Science and Technology of China (Grant No. 2004CCA01200) and the National Natural Science Foundation of China (Grant No. 30371742) for financial support of this work.

## REFERENCES

1) Jiangsu New Medical College, "Dictionary of Chinese Materia Medica," Shanghai Science and Technology Press, Shanghai, 1977, pp. 895—897.
2) Sato T., Kozima S., Kobayashi K., Kobayashi H., *Yakugaku Zasshi*, **105**, 1131—1144 (1985).
3) Lu M. C., *J. Ethnopharmacol.*, **59**, 161—165(1998).
4) Yoshikawa M., Matsuda H., Morikawa T., Xie H. H., Nakamura S., Muraoka O., *Bioorg. Med. Chem.*, **14**, 7468—7475 (2006).
5) Geng X. C., Song L. W., Pu X. P., Tu P. F., *Biol. Pharm. Bull.*, **27**, 797—801 (2004).
6) Tian X. F., Pu X. P., *J. Ethnopharmacol.*, **97**, 59—63 (2005).
7) Tanksley S. D., Nelson J. C., *Theor. Appl. Genet.*, **92**, 191—203 (1996).
8) Collard B. C. Y., Mackill D. J., *Philos. Trans. R. Soc. B, Biol. Sci.*, **363**, 557—572 (2008).
9) Atienza S. G., Satovic Z., Petersen K. K., Dolstra O., Martin A., *Euphytica*, **132**, 353—361 (2003).
10) Georgelis N., Scott J. W., Baldwin E. A., *J. Am. Soc. Hortic. Sci.*, **129**, 839—845 (2004).
11) Nozaki T., Kumazaki A., Koba T., Ishikawa K., Ikehashi H., *Euphytica*, **95**, 115—123 (1997).
12) Vijayan K., Srivatsava P. P., Nair C. V., Awasthi A. K., Tikader A., Sreenivasa B., Urs S. R., *Plant Breed.*, **125**, 298—301 (2006).
13) Hashizume T., Shimamoto I., Hirai M., *Theor. Appl. Genet.*, **106**, 779—785 (2003).
14) Shalini K. V., Manjunatha S., Lebrun P., Berger A., Baudouin L., Pirany N., Ranganath R. M., Prasad D. T., *Genome*, **50**, 35—42 (2007).
15) Florez A., Pujola M., Valero J., Centelles E., Almirall A., Casanas F., *Food Chem.*, **113**, 950—956 (2009).
16) He G. H., Hou L., Li D. M., Luo X. Y., Niu G. Q. , Tang M., Pei Y., *Acta Genet. Sin.*, **29**, 438—444 (2002).

17) Zha R. M., "Prediction of Yield Component Performances in Hybrid Rice (Oryza sativa indica L.) Using Molecular Marker Technology," Ph.D. dissertation, Center of Biotechnology, Southwest University, Chongqing, 2007.
18) Shi H. M., Wang J., Wang M. Y., Tu P. F., Li X. B., *Biol. Pharm. Bull.*, **32**, 142—146 (2009).
19) Michelmore R. W., Paran I., Kesseli R. V., *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9828—9832 (1991).