

Principal Component and Cluster Analyses for Quantitative Traits in GT Sugarcane Germplasm (*Saccharum* Spp. Hybrids)

Hui Zhou, Rong-Zhong Yang, Yang-Rui Li*

Guangxi Key Laboratory of Sugarcane Genetic Improvement, Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture, Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences, Sugarcane Research Center, Chinese Academy of Agricultural Sciences Nanning, 530007, China

*Corresponding Authors E-mail: liyr@gxaas.net

Abstract – This evaluated the quantitative traits of Guitang (GT) sugarcane (*Saccharum officinarum* L. hybrid) germplasm. Nine quantitative traits of GT sugarcane germplasm were investigated with principal component and cluster analyses. There were four principal components, including sugar factor, stalk diameter-leaf factor, stalk number-leaf factor and stalk height factor, that were extracted from nine quantitative traits. These principal components accounted for 74.42 % of the variance proportion. Sugar factor was the first principal component. There was negative effect between stalk diameter and stalk number. Stalk height was relatively independent from other traits. Cluster analysis divided the 111 accessions of sugarcane germplasm into high sugar group (including 3 subsets) and low sugar group (including 2 subsets); most of the germplasm accessions collected in recent years were clustered into high sugar group, and most of germplasm collected in earlier years were clustered into low sugar group. Sugar trait is important in sugarcane breeding. Brix was the most important trait in seedling selection. In commercial sugarcane breeding, it was suggested Brix should be a key trait, followed by stalk height and millable stalk number, and focusing on medium to medium-large stalk diameter in clone selection. Sugar trait improvement exhibited the greatest progress in the GT sugarcane breeding program in recent decades.

Keywords – Guitang (GT) Variety, Sugarcane, Quantitative Trait, Principal Component, Cluster Analysis.

I. INTRODUCTION

Seedling selection is an important step in sugarcane breeding. Breeders commonly selected the superior individual clones by vigor, plant type, Brix and resistance to disease. Slight difference of selection criteria may be present among breeders [1]. Guitang (GT) series sugarcane varieties were bred by Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/ Sugarcane Research Center, Chinese Academy of Agricultural Sciences. The Institute has bred more than 40 sugarcane varieties and a number of excellent parent clones and has made a great contribution to the sugar industry of Guangxi and China in the past 50 years. Analysis of the previous varieties and parent clones may provide useful information to the selection criteria of the breeders.

Principal component analysis could transform the original interrelated traits into independent composite indicators. Breeders could evaluate the composite

indicator in breeding according to the variance contribution rate. Principal component analysis and cluster analysis have been widely applied to evaluate quantitative traits of crop germplasm resources, including cotton, soybean, rice, corn, onions, etc. [2] [3] [4] [5] [6]. In China, there were a few reports of the quantitative trait evaluation of sugarcane and related genera germplasm resources. Zhang *et al.* [7] evaluated the major traits of the *S. spontaneum* germplasm and provided a theoretical basis for further development and utilization of the germplasm. Gao *et al.* [8] evaluated 14 sugarcane varieties for economic traits by factor analysis and cluster analysis, and provided a reference for the screening of germplasm resources. Huang *et al.* [9] [10] suggested that stalk height, millable canes, and stalk diameter were the most important traits because of their contribution to cane yield. Zhou *et al.* [11] established a mathematical model for the seedling selection criteria of quantitative traits based on stalk diameter, millable canes, stalk height, and Brix. However, there was no large-scale sugarcane germplasm evaluation of quantitative traits and seedlings selection indicators. In this study, 111 accessions of sugarcane clones grown at the Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/ Sugarcane Research Center, Chinese Academy of Agricultural Sciences were used to evaluate the major quantitative traits by principal component analysis and cluster analysis, as well as analyses the indicators of clone selection, and evaluate the achievement of genetic improvement in Guangxi.

II. MATERIALS AND METHODS

A total of 111 accessions of GT series sugarcane varieties, which were maintained in sugarcane germplasm garden of Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/ Sugarcane Research Center, Chinese Academy of Agricultural Sciences, Nanning, China (Latitude 22°50' N; Longitude 108°14' E), were used in this study. These clones have been collected from 1956 to 2006. The clones were planted in single rows of 2.5 m length, 1.2 m row space, three replications during March, 2006, and uniform cultural practices were performed. There are nine traits that were investigated during the growing season. The number of millable canes was recorded on a per plot basis. Stalk height, stalk diameter, leaf length, leaf width, leaf sheath length were recorded as the average of 15, 15, 5, 5, 5 stalks, respectively. Brix (total soluble solids % juice w/w) was

recorded on 10 stalks at the seven and 10 months of age after planting. Sucrose % juice was estimated using one rotatory analysis at the 10th months after planting. The detailed data set was based on the *Descriptors and Data*

Standard for Sugarcane (Saccharum officinarum L.) [12] and *Sugarcane Breeding* [1]. DPS software [13] was used for principal component analysis and cluster analysis.

Table 1: Sugarcane germplasm accessions used in the present study

Time	No.	Sugarcane germplasm
Before 70s	13	GT57-625, GT58-266, GT58-75, GT60-104, GT60-149, GT60-289, GT62-154, GT64-73, GT65-469, GT65-63, GT68-78, GT69-156, GT69-213
70s	23	GT71-419, GT71-5, GT72-268, GT72-287, GT72-484, GT72-653, GT73-1, GT73-167, GT74-268, GT74-445, GT74-553, GT75-208, GT75-209, GT75-57, GT76-149, GT76-154, GT76-32, GT76-62, GT76-82, GT77-204, GT77-220, GT77-32, GT77-388
80s	13	GT81-516, GT83-492, GT84-332, GT85-415, GT85-418, GT85-8, GT86-267, GT86-74, GT87-189, GT87-7, GT88-73, GT89-204, GT89-5
90s	45	GT90-420, GT90-502, GT90-95, GT91-114, GT91-116, GT91-61, GT91-9, GT91-90, GT92-263, GT92-27, GT93-101, GT93-102, GT93-103, GT94-10, GT94-11, GT94-38, GT94-63, GT94-77, GT95-315, GT95-53, GT95-53①, GT96-118, GT96-143, GT96-154, GT96-164, GT96-167, GT96-211, GT96-236, GT96-287, GT96-37, GT96-44, GT97-161, GT97-217, GT97-228, GT97-27, GT97-27②, GT97-40, GT97-69, GT98-43, GT98-55, GT98-86, GT99-107, GT99-178, GT99-181, GT99-192
After 2000	17	GT00-122, GT00-245, GT01-163, GT02-208, GT02-210, GT02-237, GT02-342, GT02-351, GT02-399, GT02-413, GT02-899, GT02-901, GT02-99, GT04-2423, GT04-2429, GT04-2453, GT04-2455

III. RESULTS AND ANALYSIS

A. Principal Component Analysis

Based on the principal component analysis of nine quantitative traits, four latent roots ($\lambda \geq 0.9$) were selected as the principal component of the sugarcane (Table 2), and the cumulative contribution of four selected principal components was 74.42%, which contained the majority of the variance information.

The eigenvalue of the first principal component was 2.7996, the variance contribution rate was 31.11%, and its corresponding eigenvectors including 270 d Brix, 300 d Brix and 300 d sucrose content, named as the sugar factor. Meanwhile, the eigenvalue of stalk diameter, leaf length, leaf width, and leaf sheath length were negative, suggesting that the high sugar sugarcane genotypes commonly had thin stalks and narrow leaves.

The eigenvalue of the second principal component was 1.6070 and had a variance contribution rate of 17.86%. Its corresponding eigenvectors including stalk diameter, leaf length, leaf width and leaf sheath length, named as stalk diameter - leaf factor. Meanwhile, the eigenvalue of stalk height and millable canes were negative, suggesting that the large stalk diameter sugarcane genotypes were often accompanied by large leaf, while the millable canes decreased.

The eigenvalue of the third principal component was 1.3653, the variance contribution rate was 15.17%, and its corresponding eigenvectors including millable canes, leaf length and leaf sheath length, named as millable canes - leaf factor. Meanwhile, the eigenvalues of stalk height, leaf width, and stalk diameter were negative, suggesting that the sugarcane genotypes with high millable canes often were thin, not tall, and with narrow-long leaf.

The eigenvalue of the fourth principal component was 0.9259, the variance contribution rate was 10.29%, and the corresponding eigenvectors included only stalk height, named as stalk height factor. The eigenvalue of stalk height in the fourth principal component was greater than the others, suggesting that it was controlled by independent genetic factors.

B. Cluster analysis

Cluster analysis of 111 accessions of sugarcane germplasm was computed by Euclidean distance and sum of squares deviations based on the results of principal component analysis (Fig. 1). When the genetic distance located within 4.80-7.59, 111 accessions of sugarcane germplasm were divided into two groups, one group was a high sugar group, and the other was a low sugar group. The high sugar group could be further divided into three sub-groups, namely high sugar I, high sugar II, and high sugar III, within the genetic distance of 2.53-2.78. Meanwhile, the low sugar group also could be divided into low sugar I and low sugar II. Mean values of nine traits in the five groups were listed in Table 3. Brix at 270 d and 300 d showed significant differences among groups, and sucrose content was significantly different among groups except for high sugar I, high sugar II, and high sugar III. In the high sugar group, the order of sucrose content from high to low was high sugar I (15.57%) > high sugar II (15.03%) > high sugar III (14.47%). Brix performed in a similar manner. The performance of stalk height, stalk diameter and millable canes was similar in the three high sugar sub-groups. In the low sugar group, the sugar content of low sugar I was higher (12.96%) than low sugar II (11.16%).

Table 2: Principal component analysis results of 9 sugarcane quantitative traits

Items	λ_1	λ_2	λ_3	λ_4	Components
Eigenvalue	2.7996	1.6070	1.3653	0.9259	
Contribution (%)	31.11	17.86	15.17	10.29	
Cumulative (%)	31.11	48.96	64.13	74.42	
Eigenvectors	0.5268	0.1272	0.0198	-0.0569	Brix at 210 d
	0.5489	0.1852	-0.0526	0.0044	Brix at 300 d
	0.4974	0.2298	-0.1243	-0.1425	Sucrose content at 300 d
	0.0828	-0.0229	-0.4060	0.8958	Stalk height
	-0.3315	0.3155	-0.2908	-0.0139	Stalk diameter
	0.0867	-0.3563	0.5119	0.1963	Millable canes
	-0.0114	0.5234	0.4478	0.2844	Leaf length
	-0.1194	0.3310	-0.3769	-0.2097	Leaf width
	-0.1871	0.5350	0.3569	0.1021	Leaf sheath length
Principal component	First	Second	Third	Forth	

Table 3: Group mean of nine quantitative traits

Group	Brix at 210 d (°BX)	Brix at 300 d (°BX)	Sucrose content (%)	Stalk height (cm)	Stalk diameter (cm)	Millable canes(stal ks/plot)	Leaf length (cm)	Leaf width (cm)	Leaf sheath length (cm)
High sugar I	19.95aA ±1.17	21.59aA ±1.10	15.57aA ±1.11	292.46aA ±19.71	2.24bcB ±0.15	22.78abA ±2.94	163.60abAB ±11.55	5.36aA ±0.30	34.55bB ±1.93
High sugar II	18.83bB ±1.31	20.81bB ±0.95	15.03abA ±1.07	289.35aA ±20.44	2.23cB ±0.14	24.50aA ±4.05	156.66bcAB ±14.41	4.54cB ±0.30	33.34bcB ±3.27
High sugar III	17.45cC ±1.09	19.99cC ±0.57	14.47bB ±1.08	291.54aA ±15.13	2.34bB ±0.19	22.58abA ±4.61	165.63aA ±11.28	5.33abA ±0.33	37.17aA ±2.64
Low sugar I	16.14dD ±1.21	18.55dD ±0.73	12.96cC ±0.91	293.97aA ±19.03	2.26bcB ±0.16	25.08aA ±5.98	153.18cB ±14.73	5.10bA ±0.37	32.78cB ±3.11
Low sugar II	15.06eD ±0.58	16.76eE ±0.69	11.16dD ±1.47	281.70aA ±16.23	2.65aA ±0.25	20.67bA ±4.76	161.06abcA ±7.47	5.34abA ±0.49	39.33aA ±2.30

Note: Data are means \pm SD for each. Means followed by different small and capital letters within a column are significantly different at P = 0.05 and 0.01 level using *t* test, respectively.

To understand the proportion of different time sources of sugarcane germplasm within each group, 111 accessions of GT sugarcane germplasm were divided into five periods, namely, before 1970s, 1980s, 1990s, and after 2000. The data in Fig. 2 showed that in the 1970s and before, the groups composed mostly by low sugar germplasm, but in the time of 1980s and after, the groups were composed mostly of high sugar germplasm. The proportion of high sugar germplasm increased over time. In each time period, the high sugar germplasm proportion ratio was 46.15% (before 70s) > 39.13% (70s) < 76.92% (80s) < 88.89% (90s) < 94.12% (after 2000).

C. Comparison of sugarcane germplasm from different time periods

The performance of sugarcane germplasm from different time periods (Table 4) showed smaller changes in yield and leaf related traits of sugarcane germplasm over time, but larger changes with certain regularity in sugar-related traits. Sugarcane germplasm from 1970s was lowest in sucrose content, and Brix at 210th day also was low. The performance of sugarcane germplasm during the 1980s was significantly better than those at 1970s, with increases of sucrose content, Brix at 210th day, and Brix at 300 d being significant. The performance of clones during the 1990s showed a steady increase, but increases for sucrose content, Brix at 210 d, and Brix at 300 d were not significant. The performance of clones collected after 2000 showed improved performance, with a significant increase for sucrose content and Brix at 210th day.

Table 4: Means of 9 traits in sugarcane germplasm accessions from different time

Time	Brix at 210 d (°BX)	Brix at 300 d (°BX)	Sucrose content (%)	Stalk height (cm)	Stalk diameter (cm)	Millable canes (stalks/plot)	Leaf length (cm)	Leaf width (cm)	Leaf sheath length (cm)
After 2000	19.87** ±1.39	21.31 ±1.06	15.77** ±0.98	295.54 ±20.90	2.22 ±0.17	22.29 ±3.53	155.98 ±16.51	4.96 ±0.51	32.99* ±3.16
90s	18.35 ±1.58	20.65 ±1.25	14.64 ±1.26	290.92 ±18.36	2.27 ±0.15	23.53 ±4.06	163.79 ±12.80	5.15* ±0.49	35.20 ±2.99

80s	17.83*	19.90*	14.30*	282.50	2.31	22.62 ±3.50	157.90	4.84	34.56
	±1.43	±1.26	±1.34	±21.88	±0.14		±12.67	±0.46	±3.70
70s	16.45	18.87	13.18	292.99	2.26*	25.09 ±5.97	156.70	4.98	34.14
	±1.57	±1.28	±1.57	±16.79	±0.19		±13.80	±0.43	±3.94
Before	17.05	18.89	13.71	290.80	2.44	24.31 ±5.60	155.98	5.13	35.63
70s	±2.05	±1.77	±1.88	±15.78	±0.29		±10.90	±0.46	±3.31

Note: *, ** means 0.05 and 0.01 significance level between two close ages (marks on late time)

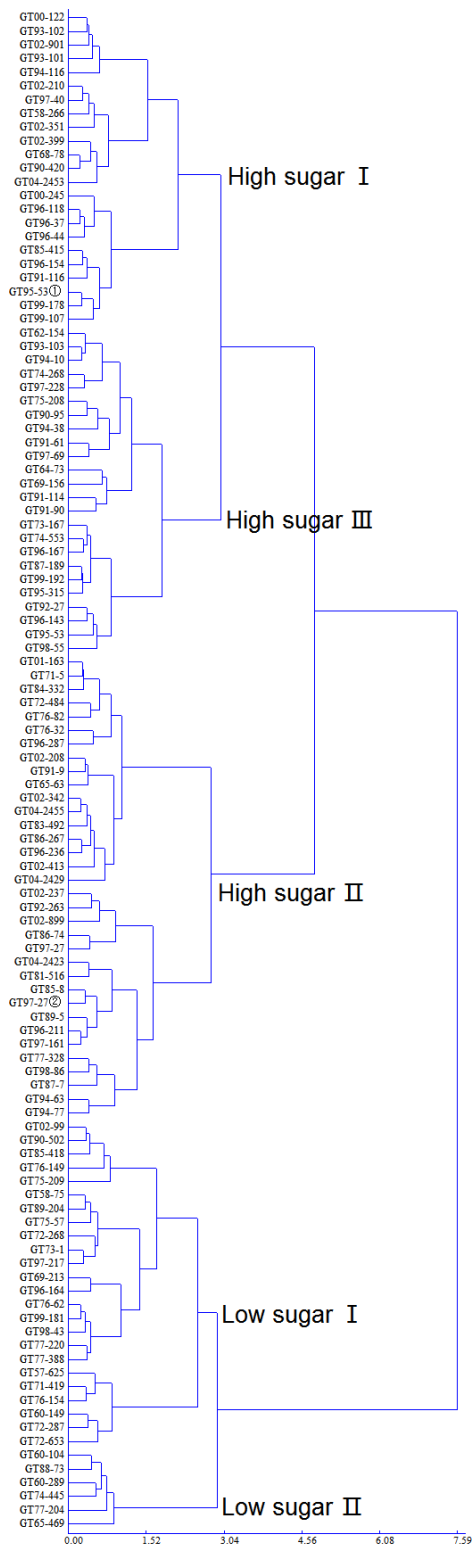


Fig.1. Cluster analysis map of 111 accessions of GT sugarcane germplasm

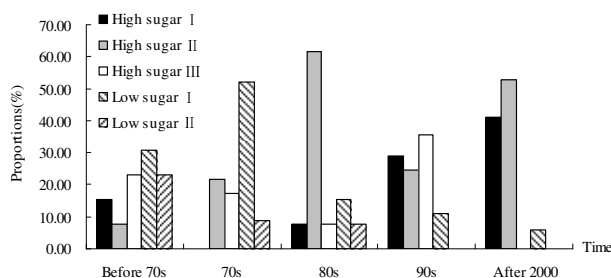


Fig.2. Accession proportions of different groups in different time

IV. DISCUSSIONS AND CONCLUSIONS

Compared to other crops, sugarcane has a special feature on the target product and the composition of yield. The production of other crops, mostly based on seeds or grains, while in the sugarcane, sucrose is the main target product. Production depends on the sucrose content and cane yield, and the cane yield was related to the traits of stalk height, stalk diameter and millable canes [1]. In the nine traits of this study, Brix at 210 d, Brix at 300 d and sucrose content were sugar-related traits; stalk height, stalk diameter, millable canes were cane yield related traits; stalk height, stalk diameter, leaf length, leaf width and leaf sheath length were sugarcane plant type related traits. These nine traits include cane yield, sucrose content and plant type factors which are important in sugarcane breeding programs. By principal component analysis, nine traits were simplified to four principal components including sugar, stalk diameter - leaf, millable canes - leaf, stalk height factors. Sugar factor was the first principal component. There was a large negative interaction between stalk diameter and millable canes. Stalk height factor was relatively independent of other measured traits. Principal component analysis showed that clonal selection should consider sugar factor first in sugarcane breeding selection, followed by yield and plant type related traits. Based on the eigenvalues of stalk height, stalk diameter and millable canes, stalk height should be considered first in sugarcane breeding, followed by stalk diameter and millable canes. In sugarcane breeding practice, large scale juice analysis was impracticable in the seedling stage selection, so breeders have to select superior clones with high yield potential to reduce population size, then perform further selection using Brix values. Based on the above analysis, we recommend the following for the seedling selection practice prioritt in Guangxi sugarcane breeding practice: Brix, stalk height, and millable canes; medium to medium-large stalk diameter would be good. Other traits, such as pipe, pith and disease resistance were

selected by traditional standards. This suggestion was similar to Huang *et al.* [9] [10].

Principal component analysis showed that the sugar factor was the first principal component, and cluster analysis divided 111 accessions of sugarcane germplasm into a high sugar group (including 3 sub-groups) and a low sugar groups (including 2 sub-groups) based on the sugar performance. Clones collected from different time periods showed regularity in the cluster analysis, later time periods had a higher proportion of high sugar clones within the groups. Comparison among different time periods showed that the sugar performance of sugarcane germplasm increased steadily after 1970s. The cluster analysis and the comparison among different time periods indicated that the improvement of sugar content was considered key in sugarcane breeding of the Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/ Sugarcane Research Center, Chinese Academy of Agricultural Sciences. After 2000, a major breakthrough has been made in the improvement of sucrose content in cane and early maturity. We have bred a number of new varieties with high sucrose content and early maturity including GTCP28 (GT00-122) [14], GT29 [15] and GT35 (GT02-901) [16], and these varieties have begun to be gradually increased in sugarcane production or used as a parent in sugarcane crossing.

ACKNOWLEDGMENTS

The present study was supported by Guangxi Natural Science Foundation (2011GXNSFF018002, 0991183), Special Fund for Bagui Scholars, Modern Agro-industry Technology Research System (CARS-20-3), and Science and Technology Development Fund of Guangxi Academy of Agricultural Sciences (2014YZ02, 2014YZ03).

REFERENCES

- [1] Peng, S.G. (1990). Sugarcane breeding. *Beijing: China Agriculture Press*. pp.7-37.
- [2] Zhu, Q.H. (1992). A preliminary study on genetics of quantitative characters in glandless cotton. *Acta Gossypii Sinica*, 4(2): 23-28.
- [3] Li, X.H., and Chang, R.Z. (1998). Cluster and Principal Component Analysis of the Spring Soybean Varieties in China. *Acta Agronomica Sinica*, 24(3): 325-332.
- [4] Liu, B.L., Shi, Y.M., Wang, W.H., Wei, S.F., Feng, R., Song, Z.P., and Mao, C.X. (2011). Principal component and cluster analysis of indica lines cultivars introduced from IRRI. *Journal of Southern Agriculture*, 42(12): 1449-1453.
- [5] Liu, Y.L., Zhang, S.K., Su, Q., Yang, Y.J., Teng, H.S., and Chen, K. (2011). Combining ability effects and cluster analysis of 16 maize inbred lines. *Journal of Southern Agriculture*, 42(6): 578-581.
- [6] Chen, Q.B., Hou, X.L., Zhang, B., Yang, J.M., Leng, Y.Q., and Jiang, F.L. (2007). Clustering and principal component analysis of germplasm resources of onion (*Allium cepa* L.). *Jiangsu Journal of Agricultural Sciences*, 23(4): 376-378.
- [7] Zhang, G.M., Yang, Y.Z., Liu, H.B., and Fang, W.K. (2006). Principal component analysis for 7 quantitative traits and cluster analysis based on 7 quantitative traits of *Saccharum spontaneum* L.. *Southwest China Journal of Agricultural Sciences*, 19(6): 1127-1131.
- [8] Gao, S.J., Chen, R.K., Zhang, H., Xu, L.N., Deng, Z.H., and Luo, J. (2006). Factor and clustering analysis of economic characters in sugarcane (*Saccharum* spp.). *Journal of Fujian Agriculture and Forestry University* (Natural Science Edition), 35(2):113-116.
- [9] Huang, J.Y., Zhu-Ge, Y., Liu, H.B., Tan, F., and Bi, S.L. (1999). Genetic correlation and path analysis of major traits in sugarcane. *Guangxi Agricultural Sciences*, 3: 6-9.
- [10] Huang, J.Y., Su, S.Q., and Wang, L.W. (2009). Genetic improvement of major characters of sugarcane varieties in Guangxi. *Guangxi Agricultural Sciences*, 40(1): 8-10.
- [11] Zhou, H.K., Ye, C.H., and Liu, G.F. (2009). Correlations analysis among a few quantitative traits of seedlings and the mathematical model of selected standard values in sugarcane. *Journal of Biomathematics*, 24(1): 183-191.
- [12] Cai, Q. and Fan, Y.H. (2006). Descriptors and data standard for sugarcane (*Saccharum officinarum* L.). China Agriculture Press, Beijing, China. pp.7-20.
- [13] Tang, Q.Y., and Feng, M.G. (2006). DPS Data processing system. *Beijing: Science Press*.
- [14] Liu, H.B., He, H., Deng, Z.Y., Zhang, G.M., Li, Y.R., Fang F.X., Xian, W., Liu, X.J., Fang, W.K., Bi, S.L., Rong, F.Y., and Tang, H.Q. (2010). Breeding of new sugarcane variety Guitang 28. *Sugar Crops of China*, 1: 13-15.
- [15] Zhang, R.H., He, H., Zhang, G.M., Liu, H.B., Li, Y.R., Fang, F.X., Song, H.Z., Fang, W.K., Bi, S.L. (2011). Breeding of new sugarcane variety Guitang 29 with high ratoon ability. *Sugar Crops of China*, (1):1-4.
- [16] Li, H.G., and Tan, Y.M. (2009). New high sugar content and early mature sugarcane variety - GT 02/901. *Guangxi Sugarcane & Canesugar*, (1): 48, 13.

AUTHOR'S PROFILE



Hui Zhou

was born in Funing City, Jiangsu, China in 1976. He graduated and earned B. A. degree in biology from Shandong University in July 1999, M.S. and Ph.D. degrees in agronomy from Fujian Agriculture and Forestry University in 2003 and 2006, respectively. He worked from 2006 until now in Sugarcane Research Institute, GXAAAS, China. From 2010-2011, he studied as a visiting scholar in U.S. Sugarcane Field Station, USDA-ARS, Canal Point, Florida.

His interests include sugarcane breeding, physiology, biochemistry and molecular biology. He has published more than 40 papers in these areas.



Yang-Rui Li

was born in Beiliu City, Guangxi, China in 1957. He graduated and earned B.A. degree in agronomy from Guangxi Agricultural University in January 1982, M.S. and Ph.D. degrees in agronomy from Fujian Agricultural University in 1985 and 1988, respectively.

He was appointed as a lecturer in Department of Agronomy, Guangxi Agricultural University from January 1988 to December 1989, and visiting scientist in Department of Agricultural Biochemistry, University of Hawaii, USA from December 1989 to February 1991. He joined Associate Professor (from March 1991 to November 1992) and Professor (since December 1992) in Department of Agronomy, Guangxi Agricultural University, and was a visiting scientist at University of Nebraska-Lincoln, USA from November 1994 to August 1996. He was appointed vice president of Guangxi University from April 1997 to April 1998, and president of Guangxi Academy of Agricultural Sciences from May 1998 to November 2012. He is now Director of Sugarcane Research Center, Chinese Academy of Agricultural Sciences. His research interests include crop sugarcane agronomy, breeding, physiology, biochemistry and molecular biology. He has more than 740 papers and 11 books in these areas. He received 29 scientific research achievements awards from Chinese Government till to date.

He was given Life Time Achievement Award of Society for Sugar Research & Promotion 2011, and Life Time Achievement Award of International Association of Professionals for Sugar and Integrated Technologies (IAPSIT) 2014. He is the founder President of IAPSIT.