Analysis of Genetic Diversity And DNA Fingerprinting In Early-Maturing Upland Cotton Using SSR Markers

Zhengcheng Kuang  
Hunan Institute of Cotton Sciences Research

Caisheng Xiao  
Hunan Institute of Cotton Sciences Research

LiShuang Guo  
Hunan Institute of Cotton Sciences Research

Wei Wang  
Agricultural Sciences Institute of Coastal Area of Jiangsu

Baohua Wang  
Nantong University

Hui Huang  
Hunan Institute of Cotton Sciences Research

Yujun Li  
Hunan Institute of Cotton Sciences Research

Yuqiang Li  
Hunan Institute of Cotton Sciences Research

Juyun Zheng  
Xinjiang Academy of Agricultural Sciences

Haodong Chen (✉ chdmks@163.com)  
Hunan Institute of Cotton Sciences Research

Research Article

Keywords: Early-maturing upland cotton, DNA fingerprinting, genetic diversity, molecular markers

Posted Date: January 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1227229/v1

License: © ️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

In this study, DNA fingerprinting and genetic diversity analysis of 79 early-maturing upland cotton (Gossypium hirsutum L.) cultivars were performed using Simple Sequence Repeat (SSR) molecular markers. From 126 pairs of SSR primers, we selected 71 pairs of primers that gave good polymorphisms and clear bands, had good stability, and showed even distribution on the cotton chromosomes, and 142 polymorphic genotypes were amplified. The average number of alleles amplified with the SSR primers was 2.01. The polymorphism information content (PIC) of the markers ranged from 0.1841 to 0.9043 with an average of 0.6494. The results of fingerprint analysis showed that nine varieties had characteristic bands, and at least six primer pairs could be used to completely distinguish all 79 cotton accessions. Using NTSYS-pc 2.11 cluster analysis, the genetic similarity coefficients between the cotton genotypes ranged from 0.3310-0.8705, with an average of 0.5861. All cotton accessions were grouped into five categories at a similarity coefficient of 0.57, which was consistent with the pedigree sources. At the same time, the average genetic similarity coefficients of early-maturing upland cotton varieties in China showed a low-high-low pattern of variation over time, revealing the development history of early-maturing upland cotton varieties from the 1980s to the present. This also indirectly reflects that in recent years, China's cotton breeders have focused on innovation and have continuously broadened the genetic resources for early-maturing upland cotton.

Introduction

Cotton is an important economic crop in China and the leading raw material used in the textile industry. In recent years, cotton cultivation in China has shown a trend of moving eastward, westward, and northward. In order to ensure safety of cotton production and utilization, China will continue to maintain the three existing cotton-producing regions; the Northwest inland (Xinjiang), the Yellow River Basin (YRB), and the Yangtze River Basin (YTRB) for the foreseeable future. However, expansion of cotton production in Xinjiang is limited by water shortages. The cotton plantation area in YTRB will be maintained at 660,000 hm² well into the future because of its suitable geographical climate, developed cotton spinning industry, and stable market demand. Cotton is the dominant crop in YTRB, although its competitiveness is weak due to its long growth period, high labor requirements, and high production cost. Therefore, it is urgent to select and breed new cotton varieties suitable for mechanized production in order to reduce labor costs and increase cotton planting efficiency.

Early-maturing upland cotton has the typical characteristics of a relatively short growth period and concentrated flowering and boll opening. Early-maturing upland cotton is one of the main targets of cotton breeding in YTRB for the future; this will not only allow for two crops per year by rotation with winter crops such as wheat and rape, but is also suitable for mechanized harvesting to achieve simple and efficient cotton production. A total of 79 early-maturing cotton accessions were collected and introduced from northern China to improve the local core germplasm resources that have long growth periods. To fully realize the genetic variation present in the introduced germplasm resources, it is necessary to study the genetic diversity of the 79 accessions.

Molecular marker technology is one of the main tools used for studying the genetic diversity of cotton varieties both in China and overseas, and the marker types include but are not limited to restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs). Of these marker
types, SSRs have the advantages of high polymorphism, good reproducibility, co-dominance, and simple operation\textsuperscript{12-13}. SSRs have been used for cotton DNA fingerprinting, genetic diversity analysis, and QTL mapping\textsuperscript{8,14-18}, and this has enhanced the protection of cotton germplasm resources and enabled the genetic improvement of cotton varieties in China. In the current research, DNA fingerprinting, genetic diversity, and the genetic relationships among 79 early-maturing upland cotton accessions were analyzed with SSR markers. The results of our study will provide genetic resources for the breeding of new varieties of early-maturing upland cotton in Hunan and YTRB through a systematic understanding of the genetic backgrounds of 79 early-maturing upland cotton germplasm accessions.

1. Materials And Methods

1.1 Experiment material

A total of 79 early-maturing upland cotton germplasm obtained from the National Cotton Mid-term Gene Bank (Anyang, Henan, China) were used for this study. All materials were used with the National Cotton Mid-term Gene Bank's permission and national guidelines. The maturity information of these materials is also provided by the Gene Bank. These materials were collected by the National Cotton Mid-term Gene Bank from six regions, including the Institute of Cotton Research of Chinese Academy of Agricultural Sciences (ICR-CAAS) which we consider to be a separate branch because it is a national institute and its cotton varieties are \textit{sui generis} (of their own kind), YRB (Henan, Shanxi, Shandong, Jiangsu), the Northwest Inland Region (Xinjiang, Gansu), the Liaohe River Basin, the United States, and the former Soviet Union (Table 1). All materials were planted in the Deshan Experimental Field of the Hunan Institute of Cotton Sciences Research in 2015. We investigated the maturity of these materials and all growth period of the materials were less than 110 days in Hunan province. Polymorphic SSR markers were selected from testing in other materials for many years, and the primers were synthesized by Shanghai Yingjun Biotechnology Co., Ltd. PCR reagents (Taq DNA polymerase, dNTPs, DNA marker size standard) were purchased from Beijing Quanjin Biotechnology Co., Ltd. Our studies did not involve any endangered or protected species.

Table 1. The names, certification years, and pedigrees of the 79 early-maturing cotton accessions
| Number | Cultivars name | Years | Region | Pedigrees |
|--------|----------------|-------|--------|-----------|
| 1      | ICR-CAAS10     | 1980  | Anyang | Heishanmian1-derived Lines |
| 2      | ICR-CAAS14     | 1989  | Anyang | 211×Liao6913 |
| 3      | ICR-CAAS16     | 1990  | Anyang | Zhong 211×ICR-CAAS10-derived Lines×Liao4086×Liaomian7-derived Lines |
| 4      | ICR-CAAS18     | 1990  | Anyang | (6908×Lambretta GL-5)F₄×Heishanmian1×Lambretta GL-5)F₃ |
| 5      | ICR-CAAS24     | 1995  | Anyang | Zhong 343× (ICR-CAAS10×USA early maturity cotton B) |
| 6      | ICR-CAAS20     | 1996  | Anyang | Shan 2942×Zhong 642×F₁×Zhong 607×Liao4086×PR80]F₁ |
| 7      | ICR-CAAS26     | 1996  | Anyang | 006×ICR-CAAS10 |
| 8      | ICR-CAAS27     | 1997  | Anyang | Zhong 0710×Zhong 5427 |
| 9      | ICR-CAAS35     | 1999  | Anyang | Zhong 23021×ICR-CAAS12×Chuan1704 |
| 10     | ICR-CAAS36     | 1999  | Anyang | H109×Zhong 662 |
| 11     | ICR-CAAS37     | 1999  | Anyang | Zhong R114-derived Lines |
| 12     | ICR-CAAS42     | 2002  | Anyang | Zhong 061723×Zhong 916448 |
| 13     | ICR-CAAS50     | 2005  | Anyang | H109×C662 |
| 14     | ICR-CAAS58     | 2006  | Anyang | SGK 27×92-047 |
| 15     | ICR-CAAS64     | 2007  | Anyang | SGK 27×Zhong 394 |
| 16     | Xinluzao1      | 1978  | the Northwest Inland Region Xinjiang | Xin 722-derived Lines69-1 |
| 17     | Xinluzao3      | 1988  | the Northwest Inland Region Xinjiang | 66-341×USA Glandless Cotton×Jingzhou 4588 |
| 18     | Xinluzao4      | 1994  | the Northwest Inland Region Xinjiang | 66-241×74-47W×DP 70 |
| 19     | Xinluzao6      | 1997  | the Northwest Inland Region Xinjiang | 85-174×Bailschlo |
| 20     | Xinluzao7      | 1997  | the Northwest Inland Region Xinjiang | 347-2×Tashkent 2 |
| 21     | Xinluzao9      | 1999  | the Northwest Inland Region Xinjiang | (Xinluzao6×Bailschlo)×ICR-CAAS17 |
| No. | Code/Culture | Year | Region | Parental Combination |
|-----|--------------|------|--------|----------------------|
| 22  | Xinluzao10   | 1999 | Xinjiang | (Heishanmian × 02) × Zhong 381 |
| 23  | Xinluzao11   | 1999 | Xinjiang | Yuzao 202-derived Lines |
| 24  | Xinluzao12   | 2000 | Xinjiang | Liao 95-25-derived Lines |
| 25  | Jiuxian2     | 2001 | Xinjiang | 77-118 × ICR-CAAS16 |
| 26  | Xinluzao13   | 2002 | Xinjiang | 83-14 × Zhong 5601 + 1639 |
| 27  | Xinluzao20   | 2005 | Xinjiang | Xinluzao16-derived Lines |
| 28  | Xinluzao22   | 2005 | Xinjiang | 45-1 × Xinluzao6 |
| 29  | Xinluzao23   | 2005 | Xinjiang | ICR-CAAS27-derived Lines |
| 30  | Xinluzao25   | 2006 | Xinjiang | [Xi 5 × Bailschlo) × Jin14] F1 × ICR-CAAS17 |
| 31  | Xinluzao26   | 2006 | Xinjiang | Xinluzao 8 Variant Plant |
| 32  | Xinluzao27   | 2006 | Xinjiang | 7147 × Bailschlo |
| 33  | Xinluzao30   | 2006 | Xinjiang | 97-185-derived Lines |
| 34  | Xinluzao31   | 2006 | Xinjiang | Xinluzao6 × Bailschlo × Daizimian |
| 35  | Xinluzao32   | 2006 | Xinjiang | Takla Makan 77-derived Lines |
| 36  | Xinluzao34   | 2007 | Xinjiang | Shi 87 Variant Plant |
| 37  | Xinluzao35   | 2007 | Xinjiang | (Xinluzao3 × Zhong 2621) × Kan 35 × 97-185 |
|   | Plant Name | Year | Region | Description |
|---|------------|------|--------|-------------|
|38 | Xinluzao36 | 2007 | the Northwest Inland Region Xinjiang | 1304×BD103 |
|39 | Xinluzao37 | 2007 | the Northwest Inland Region Xinjiang | Xi 8×(Liao 9001,Xi 5 and 90-2) |
|40 | Jiumian3   | 2007 | the Northwest Inland Region Gansu | Liaomian16 Variant Plant |
|41 | Xinluzao39 | 2008 | the Northwest Inland Region Xinjiang | Xinluzao4×Bailschlo |
|42 | Jiumian8   | 2008 | the Northwest Inland Region Gansu | Zhong 404×Xinluzao7 |
|43 | Jiumian9   | 2009 | the Northwest Inland Region Gansu | ICR-CAAS 27×Zhong 65110×Xinluzao7 |
|44 | Heishanmian1 | 1974 | the Liaohe River Basin | Jinmian1-derived Lines |
|45 | Liaomian5  | 1977 | the Liaohe River Basin | 614007(Liaomian1×Zhong 1470)×ICR-CAAS3 |
|46 | Liaomian6  | 1981 | the Liaohe River Basin | [Liaomian3×5998(F5)]×[Qiu623×Shuangqing(F4)] |
|47 | Liaomian7  | 1983 | the Liaohe River Basin | 2034+Xinlu209+Keke4104+Maste 111A+Daizimian16+64-15 |
|48 | Liaomian9  | 1985 | the Liaohe River Basin | Liao661×Hei68-35 |
|49 | Liaomian10 | 1990 | the Liaohe River Basin | Shan 3215-derived Lines |
|50 | Jinmian3   | 1991 | the Liaohe River Basin | Jin496×Sumian1 |
|51 | Liaomian12 | 1994 | the Liaohe River Basin | Liao4228×Liao4082 |
|52 | Liaomian15 | 1996 | the Liaohe River Basin | Liao 1038 Variant Plant |
|53 | Liaomian16 | 1999 | the Liaohe River Basin | 4086×Liaomian6,4696 and 4249 |
|54 | Liaomian17 | 2000 | the Liaohe River Basin | 78-4084×82-4258 |
|55 | Liaomian18 | 2002 | the Liaohe River Basin | 87-7109×Liao1038-6 |
|56 | Liaomian19 | 2003 | the Liaohe River Basin | Liao205×33B |
| No. | Variety       | Year  | Origin           | Description                                                                 |
|-----|---------------|-------|------------------|-----------------------------------------------------------------------------|
| 57  | Jinzhong200   | 1961  | YRB·Shanxi       | О-д1-derived Lines                                                          |
| 58  | Lumian1       | 1976  | YRB·Shandong     | ICR-CAAS2×1195                                                              |
| 59  | Sumian1       | 1988  | YRB·Jiangsu      | 86-1×[1087-2×Daihongdai×DP16×Heishanmian1]                                  |
| 60  | Yumian3       | 1988  | YRB·Henan        | Shangqiu17×Ke310                                                             |
| 61  | Yumian5       | 1988  | YRB·Henan        | ICR-CAAS10×Heishanmian1×Mianxiang1                                           |
| 62  | Lumian10      | 1990  | YRB·Shandong     | M-100×ICR-CAAS10                                                             |
| 63  | Yumian7       | 1992  | YRB·Henan        | ICR-CAAS10×Aiganzao×B743×F₃                                                |
| 64  | Yumian9       | 1993  | YRB·Henan        | Kan 5×ICR-CAAS10×F₁×ICR-CAAS14                                             |
| 65  | Yumian10      | 1993  | YRB·Henan        | Shang40×86-1×Ji366                                                          |
| 66  | Sumian10      | 1995  | YRB·Jiangsu      | Yanmian 48×20-9                                                             |
| 67  | Sumian11      | 1997  | YRB·Jiangsu      | Xiangkan 159×USA CAU’CS-2-81                                               |
| 68  | Jinmian23      | 1997  | YRB·Shanxi       | Lu 331-derived Lines                                                        |
| 69  | Jinmian26      | 1998  | YRB·Shanxi       | Bt gene into Jinmian 7                                                      |
| 70  | Jinmian28      | 1999  | YRB·Shanxi       | 91-57-derived Lines                                                         |
| 71  | Jinmian34      | 2002  | YRB·Shanxi       | VSG×95-1                                                                   |
| 72  | Jinmian36      | 2003  | YRB·Shanxi       | Jinmian 19×Yun 148                                                          |
| 73  | Jinmian44      | 2005  | YRB·Shanxi       | Cryl/Ac3 gene into Ji 492                                                  |
| 74  | Lumianyan27    | 2006  | YRB·Shandong     | Lu 613×GK12                                                                 |
| 75  | Lumianyan28    | 2006  | YRB·Shandong     | Lumian 14×Shiyuan321×F₁×5186,Yumian 19,Zhong 12,Zhong 19,Qinyuan 142,Lu8784 |
| 76  | Kings improved1| 1890  | the United States | Sugar block cotton                                                           |
| 77  | Foster cotton6243 | -     | the United States | Foster cotton                                                                |
| 78  | Soviet Union91-357 | -     | the former Soviet Union | Former USSR 61                                                               |
| 79  | Soviet Union10633 | -     | the former Soviet Union | Former USSR                                                                  |

### 1.2 Experimental methods

#### 1.2.1 DNA extraction

Mature leaves collected from the field were flash frozen in liquid nitrogen and ground to a powder. Genomic DNA was extracted from young leaf tissue, by CTAB DNA extraction procedure, as described by Zhang with some
modifications, and the final concentrations were adjusted to 50 ng•μL⁻¹ and stored at -20°C.

1.2.2 PCR amplification and electrophoresis detection

The PCR system consisted of 10 µL reactions containing 1 µL 10× Reaction buffer (including 10 mmol·L⁻¹ MgCl₂), 0.5 µL dNTPs (10 mmol·L⁻¹ of each), 0.4 µL of the forward and reverse primers (10 µmol·L⁻¹), 0.1 µL Taq DNA polymerase (5U·µL⁻¹), 0.5 µL cotton DNA (50 ng·µL⁻¹), and 7.1 µL ddH₂O. The PCR amplification protocol was as follows: pre-denaturation at 95°C for 30 min, followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 59°C for 45 s, and extension at 72°C for 1 min, with a final extension at 72°C for 3 min. Amplification reactions were stored at 4°C. The PCR products were separated by polyacrylamide gel electrophoresis (PAGE) on 8% gels. Electrophoresis was performed at 200V for 45 min, and the bands were observed by silver staining and photographed.

1.2.3 Band recording and data analysis

DNA fragments amplified with a primer pair that had the same migration position in the PAGE gel were recorded as 1, absence of a band was recorded as 0, bands that were blurred or had a deletion were recorded as 999, and the [0, 1] binary data matrix was constructed. The polymorphism information content of the SSR primers was calculated as ; genotypic diversity was calculated as ; the number of effective alleles per locus was , where Pi represents the gene frequency of the ith allelic variation at a certain locus. Genetic analysis of the 79 cotton accessions was performed using NTSYS-pc2.1 software. The Jaccard similarity coefficient was found using the Qualitative program in Similarity for the original [0,1] binary data matrix obtained from the EST-SSR markers. Based on the genetic similarity coefficient, the UPGMA (unweighted pair group method with arithmetic mean) algorithm in the SAHN program was used for cluster analysis, and the phenogram was generated using the Treeplot module under Graphics.

2. Results And Analysis

2.1 Selection of SSR primers and polymorphisms in the amplified products

A total of 126 pairs of candidate SSR primers were selected by using 8 samples of DNA from ICR-CAAS14 and ‘Xinluza06’ as templates (Fig. 1), and 71 pairs of SSR primers that gave good polymorphisms, clear amplified bands, and excellent stability were chosen. Except for the unknown chromosome information for 17 pairs of primers, the other primers covered all cotton chromosomes except for Chr. 04, Chr. 08, Chr. 16, and Chr. 26.

A total of 142 effective allelic variants were detected by amplifying DNA from the 79 accessions with the selected 71 pairs of SSR primers. The average number of alleles detected for each SSR primer pair was 2.01, with a range of 1-6. The effective allele numbers ranged from 1.2256 to 10.4502 with an average of 3.4379. The polymorphism information content (PIC) ranged from 0.1841 to 0.9043, with an average of 0.6494. The PIC of the primer pair MON_CGR5565 had the highest value of 0.9043, and the primer pair NAU4044 had the second highest PIC value of 0.8822, whereas the primer pair NAU3181 had the lowest PIC of 0.1841. Genetic diversity (H') ranged from 1.2256 to 10.4502 (Table 2).

Table 2. SSR marker loci, chromosomal locations, and SSR-PCR polymorphism data
| SN | SRR locus | Chromosome | PIC  | Genetic diversity (H') | Effective Number of Alleles (Ne) |
|----|-----------|------------|------|------------------------|-------------------------------|
| 1  | NAU4073   | Chr.01     | 0.8168 | 5.7030                 | 5.4582                        |
| 2  | NAU2457   | Chr.01     | 0.2577 | 1.5311                 | 1.3471                        |
| 3  | NAU2083   | Chr.01     | 0.4356 | 1.8725                 | 1.7717                        |
| 4  | NAU4044   | Chr.01     | 0.8822 | 9.1008                 | 8.4884                        |
| 5  | MON_COT064| Chr.02     | 0.4980 | 1.9960                 | 1.9920                        |
| 6  | NAU1190   | Chr.03     | 0.7956 | 5.3171                 | 4.8933                        |
| 7  | MON_CGR6528| Chr.03   | 0.7423 | 3.9381                 | 3.8800                        |
| 8  | MON_CGR6683| Chr.03   | 0.4576 | 1.9158                 | 1.8437                        |
| 9  | NAU1269   | Chr.05     | 0.7314 | 3.8504                 | 3.7234                        |
| 10 | MON_CGR5732| Chr.05  | 0.7307 | 3.8447                 | 3.7138                        |
| 11 | NAU1225   | Chr.05     | 0.7307 | 3.8447                 | 3.7138                        |
| 12 | MON_DC40122| Chr.05 | 0.6884 | 3.5157                 | 3.2096                        |
| 13 | NAU1221   | Chr.05     | 0.7307 | 3.8447                 | 3.7138                        |
| 14 | MON_CGR5651| Chr.06 | 0.6811 | 3.4515                 | 3.1354                        |
| 15 | MON_DPL0702| Chr.06 | 0.7349 | 3.8787                 | 3.7721                        |
| 16 | MON_COT002| Chr.06     | 0.7266 | 3.8116                 | 3.6572                        |
| 17 | BNL1694   | Chr.07     | 0.7442 | 3.9534                 | 3.9092                        |
| 18 | MUSS095   | Chr.07     | 0.7431 | 3.9447                 | 3.8921                        |
| 19 | MON_DC30218| Chr.07 | 0.5000 | 2.0000                 | 2.0000                        |
| 20 | NAU3859   | Chr.09     | 0.4980 | 1.9960                 | 1.9920                        |
| 21 | DPL0431   | Chr.10     | 0.6873 | 3.4966                 | 3.1981                        |
| 22 | NAU3784   | Chr.11     | 0.7071 | 3.6561                 | 3.4141                        |
| 23 | MON_CER0098| Chr.11  | 0.7250 | 3.8002                 | 3.6363                        |
| 24 | NAU3563   | Chr.11     | 0.8579 | 7.4448                 | 7.0361                        |
| 25 | NAU2671   | Chr.12     | 0.6978 | 3.5853                 | 3.3095                        |
| 26 | MON_DPL0491| Chr.12 | 0.2397 | 1.4972                 | 1.3153                        |
| 27 | NAU3991   | Chr.13     | 0.6801 | 3.4455                 | 3.1264                        |
| 28 | MON_COT009| Chr.13     | 0.6886 | 3.5149                 | 3.2116                        |
| 29 | BNL1421   | Chr.13     | 0.7107 | 3.6867                 | 3.4567                        |
|   |          | Chr. | 1      | 2      | 3      |
|---|----------|------|--------|--------|--------|
| 30| NAU3308  | 14   | 0.6721 | 3.3728 | 3.0500 |
| 31| GH304    | 15   | 0.7085 | 3.6704 | 3.4305 |
| 32| MUSS440  | 15   | 0.8784 | 8.9176 | 8.2218 |
| 33| NAU2343  | 15   | 0.7354 | 3.8824 | 3.7790 |
| 34| BNL2646  | 15   | 0.6360 | 3.1112 | 2.7475 |
| 35| NAU2742  | 17   | 0.6635 | 3.3170 | 2.9714 |
| 36| MON_DPL0308 | 18 | 0.5896 | 2.7536 | 2.4365 |
| 37| NAU3011  | 18   | 0.4980 | 1.9960 | 1.9920 |
| 38| NAU5262  | 18   | 0.7407 | 3.9252 | 3.8560 |
| 39| MON_CGR6151 | 19 | 0.4768 | 1.9539 | 1.9115 |
| 40| NAU1187  | 19   | 0.7307 | 3.8447 | 3.7138 |
| 41| NAU1042  | 19   | 0.7307 | 3.8447 | 3.7138 |
| 42| MON_CGR5590 | 19 | 0.7193 | 3.7542 | 3.5631 |
| 43| TMB1791  | 19   | 0.7387 | 3.9101 | 3.8277 |
| 44| MON_CGR6439 | 20 | 0.2604 | 1.5362 | 1.3520 |
| 45| DPL0442  | 20   | 0.7317 | 3.8542 | 3.7271 |
| 46| MON_SHIN1421 | 20 | 0.7418 | 3.9344 | 3.8728 |
| 47| MON_CGR5565 | 20 | 0.9043 | 11.0904 | 10.4502 |
| 48| BNL1551  | 21   | 0.7006 | 3.6046 | 3.3401 |
| 49| GH222    | 22   | 0.8094 | 5.5713 | 5.2460 |
| 50| MON_CGR6410 | 22 | 0.7378 | 3.9022 | 3.8136 |
| 51| CIR253   | 22   | 0.6158 | 2.9431 | 2.6031 |
| 52| MUSS139  | 23   | 0.7426 | 3.9408 | 3.8848 |
| 53| MON_CGR5202 | 24 | 0.3230 | 1.6552 | 1.4772 |
| 54| MON_CGR6932 | 25 | 0.7423 | 3.9381 | 3.8800 |
| 55| CRI151   | 25   | 0.7312 | 3.8485 | 3.7200 |
| 56| MON_CGR6389 |  | 0.7430 | 3.9442 | 3.8913 |
| 57| NAU3181  | 25   | 0.1841 | 1.3919 | 1.2256 |
| 58| MON_C2-0118 |  | 0.7397 | 3.9179 | 3.8418 |
| 59| CRI002   | 27   | 0.6296 | 3.0356 | 2.7000 |
| 60| MUCS375  | 28   | 0.3519 | 1.7103 | 1.5429 |
2.2 DNA fingerprinting analysis of the 79 cotton accessions

Fingerprinting analysis of the 79 early-maturing upland cotton accessions was performed using 71 pairs of SSR primers. We found that nine accessions had characteristic bands for which only one primer pair was needed to distinguish each accession from the others. Among them, ICR-CAAS64 had two characteristic primers, ‘Xinluzao20’, ‘Xinluzao25’, ‘Jiumian9’, ‘Liaomian5’, ‘Liaomian17’, ‘Liaomian19’, ‘Lumianyan28’, and ‘Jinmian23’ each had one characteristic primer (Table 3). The primer pair NAU4044 was able to uniquely identify four varieties including ‘Xinluzao25’, ‘Jiumian9’, ‘Lumianyan28’, and ‘Liaomian 5’. Primer pair NAU3254 could distinguish three varieties, ‘Xinluzao20’, ‘Liaomian17’, and ‘Liaomian19’. These results indicated that these two primer pairs had abundant polymorphism, strong discrimination power, and numerous characteristic bands, and could be used as preferred markers in the identification of fingerprints.

Table 3. Cotton accessions identified with specific SSR primer pairs

| Cultivar       | Specific primer       | Cultivar       | Specific primer       |
|----------------|-----------------------|----------------|-----------------------|
| ICR-CAAS64     | NAU1190,MUSS440       | Xinluzao20    | NAU3254               |
| Liaomian17     | NAU3254               | Liaomian19    | NAU3254               |
| Xinluzao25     | NAU4044               | Jiumian9      | NAU4044               |
| Liaomian5      | NAU4044               | Lumianyan28   | NAU4044               |
| Jinmian23      | NAU4073               |                |                       |

A total of 55 of the 79 cotton accessions could be identified by three pairs of primers, NAU4044, MUSS440, and MON_CGR5565, which were selected from the 71 pairs of core primers with high PIC values, strong discriminative power, clear bands on the gels, and high reproducibility. Seventy-two varieties could be identified by adding another primer pair, GH222. By adding primer pairs NAU1190 and BNL1694, all 79 of the early-maturing upland cotton varieties could be completely distinguished from one another (Table 4).
Table 4. Fingerprinting data for the 79 early-maturing upland cotton accessions
| SN | Cultivar                | NAU | MUSS | MON_CGR | GH | NAU | BNL |
|----|-------------------------|-----|------|---------|----|-----|-----|
| 1  | ICR-CAAS10              | 10001 | 00101 | 001100 | 001 | 101 | 10  |
| 2  | ICR-CAAS14              | 10001 | 11101 | 001100 | 010 | 101 | 01  |
| 3  | ICR-CAAS16              | 10001 | 11011 | 001100 | 001 | 101 | 01  |
| 4  | ICR-CAAS18              | 10010 | 00101 | 001100 | 001 | 011 | 10  |
| 5  | ICR-CAAS24              | 00101 | 00101 | 101101 | 101 | 101 | 11  |
| 6  | ICR-CAAS20              | 10010 | 00101 | 001100 | 100 | 011 | 10  |
| 7  | ICR-CAAS26              | 10101 | 00101 | 110001 | 001 | 011 | 01  |
| 8  | ICR-CAAS27              | 10001 | 11101 | 001100 | 110 | 111 | 10  |
| 9  | ICR-CAAS35              | 11000 | 11000 | 010001 | 001 | 101 | 10  |
| 10 | ICR-CAAS36              | 10001 | 11101 | 001100 | 101 | 101 | 10  |
| 11 | ICR-CAAS37              | 00100 | 11000 | 001010 | 010 | 101 | 01  |
| 12 | ICR-CAAS42              | 00100 | 00101 | 100001 | 100 | 011 | 10  |
| 13 | ICR-CAAS50              | 10001 | 11000 | 100001 | 100 | 101 | 10  |
| 14 | ICR-CAAS58              | 10100 | 00101 | 100001 | 100 | 011 | 10  |
| 15 | ICR-CAAS64              | 00100 | 00000 | 100001 | 100 | 000 | 10  |
| 16 | Xinluzao1               | 11000 | 00101 | 100001 | 100 | 101 | 01  |
| 17 | Xinluzao3               | 10101 | 00101 | 101101 | 010 | 101 | 01  |
| 18 | Xinluzao4               | 10001 | 00011 | 100001 | 100 | 010 | 01  |
| 19 | Xinluzao6               | 10010 | 11000 | 001100 | 001 | 101 | 01  |
| 20 | Xinluzao7               | 10001 | 11000 | 100001 | 010 | 111 | 01  |
| 21 | Xinluzao9               | 10010 | 00101 | 100001 | 001 | 101 | 01  |
| 22 | Xinluzao10              | 01010 | 11101 | 100001 | 001 | 011 | 01  |
| 23 | Xinluzao11              | 10101 | 11101 | 001110 | 001 | 111 | 11  |
| 24 | Xinluzao12              | 11101 | 00101 | 001110 | 010 | 101 | 11  |
| 25 | Jiumian2                | 10100 | 00101 | 001100 | 101 | 101 | 01  |
| 26 | Xinluzao13              | 10001 | 11000 | 010001 | 001 | 101 | 10  |
| 27 | Xinluzao20              | 11101 | 01111 | 001100 | 110 | 101 | 11  |
| 28 | Xinluzao22              | 11000 | 00101 | 001010 | 100 | 101 | 10  |
| 29 | Xinluzao23              | 10100 | 00101 | 111111 | 001 | 101 | 01  |
|   | Field         | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | Value 6 | Value 7 | Value 8 | Value 9 | Value 10 | Value 11 | Value 12 |
|---|--------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 30| Xinluzao25   | 11011   | 11011   | 001110  | 101     | 101     | 01      |
| 31| Xinluzao26   | 00011   | 11011   | 001110  | 001     | 110     | 11      |
| 32| Xinluzao27   | 10001   | 11000   | 001100  | 001     | 011     | 10      |
| 33| Xinluzao29   | 10100   | 00101   | 001100  | 100     | 101     | 01      |
| 34| Xinluzao30   | 10010   | 00101   | 001100  | 001     | 101     | 01      |
| 35| Xinluzao31   | 01010   | 11000   | 10001   | 001     | 011     | 10      |
| 36| Xinluzao32   | 01000   | 00011   | 10001   | 001     | 101     | 10      |
| 37| Xinluzao35   | 00101   | 00101   | 111111  | 001     | 111     | 11      |
| 38| Xinluzao36   | 10010   | 11000   | 10001   | 010     | 011     | 01      |
| 39| Xinluzao37   | 00101   | 01111   | 001110  | 001     | 101     | 01      |
| 40| Jiumian3     | 00100   | 00101   | 001100  | 101     | 101     | 01      |
| 41| Xinluzao39   | 01000   | 00101   | 01001   | 001     | 011     | 10      |
| 42| Jiumian8     | 00100   | 11000   | 001100  | 001     | 101     | 01      |
| 43| Jiumian9     | 10011   | 00101   | 101101  | 100     | 101     | 01      |
| 44| Heishanmian1 | 10001   | 11000   | 001010  | 001     | 011     | 10      |
| 45| Liaomian5    | 00001   | 00101   | 10001   | 010     | 011     | 01      |
| 46| Liaomian6    | 00100   | 00101   | 001010  | 010     | 101     | 01      |
| 47| Liaomian7    | 00100   | 00101   | 001100  | 100     | 101     | 01      |
| 48| Liaomian9    | 10101   | 00101   | 001100  | 011     | 101     | 11      |
| 49| Liaomian10   | 11101   | 00101   | 001100  | 010     | 101     | 10      |
| 50| Liaojinmian3 | 10100   | 11000   | 001010  | 001     | 101     | 01      |
| 51| Liaomian12   | 00100   | 00101   | 001100  | 011     | 101     | 01      |
| 52| Liaomian15   | 10100   | 00011   | 001100  | 001     | -       | 10      |
| 53| Liaomian16   | 10100   | 00101   | -       | 010     | 101     | 10      |
| 54| Liaomian17   | 00100   | 01111   | 001100  | 001     | 101     | 01      |
| 55| Liaomian18   | 10100   | 00011   | 001100  | 010     | 100     | 10      |
| 56| Liaomian19   | 00100   | 01111   | 101101  | 010     | 010     | 10      |
| 57| Jinzhong200  | 10001   | 00101   | 10001   | 001     | 101     | 10      |
| 58| Lumian1      | 10001   | 01111   | 111111  | 001     | 110     | 10      |
| 59| Sumian1      | 00100   | 00011   | 001010  | 010     | 101     | 10      |
| 60| Yumian3      | 10100   | 00101   | 111111  | 011     | 011     | 11      |
### 2.3 Genetic diversity analysis

Similarity coefficients between varieties were calculated using NTSYS-pc 2.11 and Microsoft Excel software. The results showed that the genetic similarity coefficients among the 79 early-maturing upland cotton accessions ranged from 0.3310 ('Jinmian36' and 'Xinluzao25') to 0.8705 ('Liaomian15' and 'Liaomian18'), with an average of 0.5861, indicating that 'Jinmian36' and 'Xinluzao25' have the highest genetic diversity, while 'Liaomian15' and 'Liaomian18' have the lowest. The similarity coefficients between varieties that were <0.4 (large genetic difference) accounted for 1.1%, 11.9% at 0.4-0.5, 44.1% at 0.5-0.6, 35.2% at 0.6-0.7, and 7.1% at 0.7-0.8. The similarity coefficients >0.8 accounted for 0.5% (Fig. 2). This demonstrates that the genetic relationships are relatively close among the 79 early-maturing upland cotton accessions, but that some genetic diversity is still present.

We performed genetic diversity analysis of the 79 early-maturing upland cotton accessions from six regions that included the China Cotton Institute, YRB (Henan, Shanxi, Shandong, Jiangsu), the Northwest Inland Cotton Region (Xinjiang, Gansu), the Liaohe River Basin, the United States, and the former Soviet Union. By comparison, the accessions from the former Soviet Union had the smallest average genetic similarity coefficients of the six regions, which increased in the order of the YRB cotton area, the United States, the China Cotton Institute, the

|   |            |       |     |        |        |       |        |       |       |
|---|------------|-------|-----|--------|--------|-------|--------|-------|-------|
| 61| Yumian5    | 11000 | 00101 | 001100 | 001   | 101   | 10     |
| 62| Lumian10   | 10010 | 00101 | 110001 | 001   | 101   | 01     |
| 63| Yumian7    | 10100 | 00101 | 001010 | 100   | 101   | 10     |
| 64| Yumian9    | 10001 | 00101 | 001100 | 001   | 101   | 01     |
| 65| Yumian10   | 11100 | 00101 | 010001 | 001   | 011   | 01     |
| 66| Sumian10   | 10101 | 00011 | 001100 | 011   | 010   | 01     |
| 67| Sumian11   | 11100 | 00101 | 110001 | 001   | 101   | 11     |
| 68| Jinmian23  | 11111 | 11101 | 111111 | 101   | 101   | 10     |
| 69| Jinmian26  | 11111 | 11101 | 101101 | 101   | 101   | 01     |
| 70| Jinmian28  | 10100 | 11101 | 101101 | 011   | 101   | 01     |
| 71| Jinmian34  | 10100 | 00101 | 100001 | 001   | 101   | 01     |
| 72| Jinmian36  | 11000 | 00011 | 010001 | 010   | 010   | 01     |
| 73| Jinmian44  | 00100 | 00011 | 100001 | 001   | 010   | 01     |
| 74| Lumianyan27| 10100 | 00011 | -      | 001   | 010   | 01     |
| 75| Lumianyan28| 01100 | 01111 | 110001 | 010   | 010   | 01     |
| 76| Kings improved1 | 10001 | 00101 | 001100 | 001   | 011   | 10     |
| 77| Foster cotton6243 | 10100 | 00101 | 001100 | 001   | 101   | 10     |
| 78| Soviet Union91-357 | 00011 | 00101 | 001100 | 011   | 110   | 10     |
| 79| Soviet Union10633 | 10100 | 00011 | 001100 | 001   | 100   | 01     |
Liaohe River Basin Early Maturing Cotton Area, and the Northwest Inland Cotton Area, indicating that there is ample genetic diversity in cotton resources imported from abroad. At the same time, the genetic diversity of accessions from the YRB cotton region is relatively high, which may be related to the complex geographical diversity of the YRB cotton region and the dispersion of breeders in Henan, Shanxi, Shandong, Jiangsu, and other provinces.

We found that the genetic similarity coefficients of accessions from the six regions were between 0.5575 and 0.6143, and the highest similarity coefficients were found between accessions from China and the USA. This indicates that early-maturing upland cotton varieties selected by ICR-CAAS have close genetic relationships to selections from the USA. In general Chinese cotton germplasm is more frequently exchanged for resources from the USA compared with other regions. The lowest genetic similarity coefficient in the early-maturing cotton areas is between YRB and the Liaohe River Basin, with a value of 0.5575, and the second lowest is between YRB and the Northwest Inland Cotton Area, with a value of 0.5636 (Table 5). The underlying reason for this may be that the YRB cotton-growing area has a better climate with warmer conditions, resulting in more varieties of early-maturing upland cotton and larger differences between varieties than the early-maturing cotton areas of the Liaohe River Basin and the Northwest Inland Cotton Area.

Comparisons of the genetic similarity coefficients between domestic and foreign early-maturing upland cotton varieties showed that except for the Northwestern Inland Cotton Area, the genetic similarity coefficients between cotton varieties from the ICR-CAAS, YRB, and the Liaohe River Basin are higher. This suggests that early-maturing upland cotton grown in the ICR-CAAS, YRB, and Liaohe River Basin in the early-maturing cotton area contains more American germplasm. Because of the introduction and utilization of early-maturing upland cotton in China, the majority of early maturity genetic resources came from American gold-colored cotton. The genetic similarity coefficients between accessions from the Northwestern Inland Cotton Region and the former Soviet Union is relatively high. This may be because the Northwest Inland Cotton Region is adjacent to the former Soviet Union, so it is easier to introduce germplasm resources into China from there.

Table 5. Genetic diversity of cotton cultivars from the six cotton-growing regions in China
The results of our study show that the average genetic similarity coefficient of bred and certified varieties of early-maturing upland cotton varieties from different areas in China showed a low-high-low pattern of variation as determined by genetic diversity analysis. ‘Jinzhong200’, ‘Xinluzao1’, ‘Lumian1’, ‘Heishanmian1’, and ‘Liaomia5’, which were selected prior to the 1980s, had the lowest genetic similarity coefficient of 0.5704. The nine varieties ‘ICR-CAAS10’, ‘ICR-CAAS14’, ‘Xinluzao3’, ‘Liaomian6’, ‘Liaomian7’, ‘Liaomian9’, ‘Sumian1’, ‘Yumian3’, and ‘Yumian5’, which were selected in the 1980s, have the highest average genetic similarity coefficient of 0.6306. The 29 varieties selected in the 1990s, which include ‘ICR-CAAS16’, ‘ICR-CAAS18’, ‘Xinluzao4’, ‘Liaomian10’, ‘Lumian10’, ‘Yumian7’, ‘Sumian10’, and ‘Jinmian23’, have an average genetic similarity coefficient of 0.5993. The average similarity coefficient of 32 varieties including ‘ICR-CAAS42’, ‘ICR-CAAS50’, ‘Xinluzao13’, ‘Jiumian2’, ‘Liaomian17’, ‘Jinmian34’, and ‘Lumianyan27’ selected after 2000 is 0.5791.

The average genetic similarity coefficients of early-maturing upland cotton varieties in China have shown a low-high-low pattern over time (Fig. 3). This may be because before the 1980s, domestic early-maturing upland cotton breeding was mainly carried out by introducing different early-maturing varieties from abroad and systematically using them in breeding. Since the early 1980s, cotton production and the cotton spinning industry

| Region                          | CAAS         | YRB         | the Northwest Inland Region | the Liaohe River Basin | the United States | the former Soviet Union |
|---------------------------------|--------------|-------------|-----------------------------|------------------------|-------------------|------------------------|
| CAAS                            | Max          | 0.8058      | 0.7817                      | 0.8169                 | 0.8451            | 0.7447                 | 0.6691                 |
|                                 | Min          | 0.3630      | 0.3643                      | 0.3704                 | 0.3582            | 0.4789                 | 0.4296                 |
|                                 | Mean         | 0.6032      | 0.5796                      | 0.5943                 | 0.5824            | 0.6143                 | 0.5672                 |
| YRB                             | Max          | 0.7899      | 0.8028                      | 0.7606                 | 0.7042            | 0.7042                 |                       |
|                                 | Min          | 0.4085      | 0.3310                      | 0.3475                 | 0.3521            | 0.3582                 | 0.4366                 |
|                                 | Mean         | 0.5802      | 0.5636                      | 0.5575                 | 0.5911            | 0.5661                 |                       |
| the Northwest Inland Region     | Max          |             | 0.8239                      | 0.8310                 | 0.7324            | 0.7254                 |                       |
|                                 | Min          |             | 0.4225                      | 0.3582                 | 0.4296            | 0.4296                 |                       |
|                                 | Mean         |             | 0.6221                      | 0.5936                 | 0.5799            | 0.5894                 |                       |
| the Liaohe River Basin          | Max          |             | 0.8705                      | 0.7465                 | 0.6812            |                       |                       |
|                                 | Min          |             | 0.3944                      | 0.4718                 | 0.4014            |                       |                       |
|                                 | Mean         |             | 0.6134                      | 0.5783                 | 0.5684            |                       |                       |
| the United States               | Max          |             |                             | 0.5845                 | 0.6549            |                       |                       |
|                                 | Min          |             |                             | 0.5845                 | 0.5357            |                       |                       |
|                                 | Mean         |             |                             | 0.5845                 | 0.6027            |                       |                       |
| the former Soviet Union         | Max          |             |                             |                       |                   | 0.5286                 |                       |
|                                 | Min          |             |                             |                       |                   | 0.5286                 |                       |
|                                 | Mean         |             |                             |                       |                   | 0.5286                 |                       |
have developed rapidly. Due to economic reform and the opening up of the country, transportation is more convenient, and the exchange of germplasm resources between breeding units has become frequent. In particular, a number of outstanding varieties (lines) such as ‘Heishanmian1’ and ‘ICR-CAAS10’ stand out from the competition and are used by other breeders as donor parents. This has resulted in closer genetic relationships between the varieties selected at this time, with higher genetic similarity coefficients and less genetic difference. In the 1990s, the difficulties of domestic distant hybridization were continuously overcome, and breeders consciously chose parental materials with complex genetic backgrounds for cross-breeding, which resulted in a significant reduction in the genetic similarity coefficients of cotton varieties and increased the genetic difference. After 2000, the use of modern breeding technologies (transgenics and molecular marker-assisted breeding) not only accelerated the cotton breeding process, but also broadened the source of available cotton genes, resulting in further reductions in the genetic similarity coefficients among new varieties of early-maturing upland cotton in China.

Based on the Jaccard similarity coefficient, 79 early-maturing upland cotton varieties were grouped using a hierarchical clustering method (UPGMA) (Fig. 4). The results showed that at a genetic similarity coefficient of 0.87, all 79 early-maturing upland cotton varieties were completely separated. At a similarity coefficient of 0.57, the 79 main varieties could be divided into five categories or classes. Class I contains 42 varieties, including seven varieties from the China Cotton Institute, 20 varieties from the Northwest Inland Cotton Area, nine varieties from the special early-maturing cotton area of the Liaohe River Basin, four varieties from the YRB cotton area, one variety from the US, and one variety from the former Soviet Union. Class II contains 27 varieties, including six varieties from the China Cotton Institute, six varieties from the Northwest Inland Cotton Area, two varieties from the special early-maturing cotton area of the Liaohe River Basin, 11 varieties from the YRB cotton area, and one variety each from the United States and the former Soviet Union. Class III has only varieties, both of which are from the Northwestern Inland Cotton Area; Class IV includes a single variety from YRB; and Class V contains seven varieties, including two from Zhongmian, two varieties from the Liaohe Basin early-maturing cotton area, and three varieties from the YRB cotton area. Among these five classes, most of the selected varieties of cotton grown in China clustered in Class I and Class II, accounting for 46.7% and 40.0% respectively. Most of the varieties from the Northwestern Inland cotton area are in Class I, accounting for 71.4%; most of varieties from the Liaohe Basin are concentrated in Class I, accounting for 69.2%; and most of the cotton varieties from YRB are concentrated in Class II, accounting for 57.9%. This shows that the clustering results reflect certain geographical distribution characteristics, and the genetic differences of the cultivars from the same area are relatively small, which is why they cluster together.

3. Discussion

In recent years, SSR molecular marker technology has been widely used in the genetic diversity and fingerprinting of cotton germplasm resources. For example, Han et al.\textsuperscript{21} used SSR markers to construct fingerprints and analyze the genetic diversity of 27 cotton accessions from 2009 to 2010. Li et al.\textsuperscript{22} used 20 pairs of SSR primers to construct fingerprints of the BaiMian cotton series. Kuang et al.\textsuperscript{23} found a correlation between cultivars and geographical origins by analyzing the genetic diversity in the main cotton varieties in China using SSR markers. These studies were important for the identification and purity detection of cotton varieties in China, as well as understanding the genetic relationships and origins of cotton varieties at the
molecular level, which provided a theoretical basis for rational selection of hybrid parents and the breeding of new varieties of cotton.

However, there are few reports on the use of SSR molecular markers to study early-maturing upland cotton in China. In our study, we constructed DNA fingerprints of 79 early-maturing upland cotton accessions using 73 SSR markers and analyzed the molecular data to determine genetic similarities. We found that 72 main varieties can be divided into five categories (Classes I, II, III, IV, and V) with a genetic similarity coefficient of 0.57, and the clustering results for the 79 early-maturing upland cotton accessions were consistent with the pedigree (Fig. 5). For example, ‘Liaomian15’ and ‘Liaomian18’ were first clustered together, traced back to their pedigree sources, and were both found to contain ‘Liao 1038’ in their ancestry. ‘Xinluzao6’, ‘Xinluzao9’, ‘Xinluzao27’, ‘Xinluzao31’, and ‘Xinluzao39’ are grouped together in Class I, and their pedigree sources show that they all are descended from ‘Bell Snow’. ‘Xinluzao6’, ‘Xinluzao9’, and ‘Xinluzao27’ clustered into a group with a genetic similarity coefficient of 0.87. We also identified several cases of incomplete matching. For example, ‘Jiu Mian 2’ is in the ‘ICR-CAAS16’ lineage; ‘Jinmian3’ is in the ‘Sumian1’ lineage; seven varieties including ‘ICR-CAAS16’, ‘Zhongmian24’, ‘ICR-CAAS26’, ‘Yumian5’, ‘Yumian7’, ‘Yumian9’, and ‘Lumian10’ are in the ‘Zhongmian10’ lineage; three varieties including ‘ICR-CAAS16’, ‘ICR-CAAS28’, and ‘Sumian 1’ are in the ‘Heishan Mian1’ lineage; and three varieties including ‘Heishan1’, ‘Jinzong200’, and ‘Jinzi 1’ descend from the gold-line cotton pedigree. The clustering results indicate that seven varieties including ‘ICR-CAAS10’, ‘Jiu Mian2’, ‘ICR-CAAS16’, ‘Yumian5’, ‘Yumian9’, ‘Jinmian3’, and ‘Lumian10’ group together in Class I. Eight varieties including ‘Sumian1’, ‘Zhongmian24’, ‘Zhongmian26’, ‘Yumian7’, ‘ICR-CAAS18’, ‘Heishan Mian1’, ‘Jinzong200’, and ‘Kings improved1’ are in Class II.

In addition, the genetic relationship of ‘ICR-CAAS37’ and ‘Liaomian6’ seem to be relatively distant but can be grouped together. However, the genetic relationship of ‘Liaomian6’ and ‘Liaomian16’ seem to be relatively close but do not cluster together. This suggests that the classification is not entirely dependent on pedigree, but is also influenced by the selection method, breeding process, and target traits. Pedigree analysis can only reflect the relative genetic information between the varieties. Molecular marker analysis using marker loci distributed over the whole genome can more accurately reflect the genetic differences between varieties24-25.

Declarations

Author Contributions: H.D.C designed the experiments. Z.C.K, H.D.C, L.S.G, C.S.X, and J.Y.Z conceived the experiments and analyzed the results. Z.C.K, H.D.C, C.S.X, L.S.G, J.Y.Z carried out all computational analyses. Z.C.K, H.D.C, C.S.X, L.S.G, J.Y.Z, W.W, B.H.W, H.H, Y.J.L and Y.Q.L participated in part of experiments. H.D.C, Z.C.K, and H.H drafted the manuscript. H.H proofread and H.D.C, B.H.W revised the manuscript. All authors read and approved the final manuscript.

Funding: This program was financially sponsored by The National Key Research and Development Program of China (2021YFE0101200), “Huxiang Young Talents Plan” Support Project of Hunan Province (2019RS2048), Natural Science Foundation of Hunan Province (2020JJ5291), “Tianshan” Innovation team program of the Xinjiang Uygur Autonomous Region (2021D14007) and State Key Laboratory of Cotton Biology Open Fund (CB2020A16)

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Dai JL, Kong XQ, Zhang DM, et al. Technologies and theoretical basis of light and simplified cotton cultivation in China. Field Crops Research, 2017, https://doi.org/10.1016/j.fcr.2017.09.005.

2. Maik W, Abid M A, Cheema H M N, et al. From Qutn to Bt Cotton: Development, adoption and prospects. A review. Tsitologiya i Genetika, 2015, 49(6):73-85.

3. Ma ZY, He SP, Wang XF, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nature Genetics, 2018, https://doi.org/10.1038/s41588-018-0119-7.

4. Liu BJ, Liang ML, Huang ZQ, et al. Duration - severity - area characteristics of drought events in eastern China determined using a three-dimensional clustering method. International Journal of Climatology, 2021, https://doi.org/10.1002/joc.6904.

5. Feng L, Dai JL, Tian LW, et al. Review of the technology for high-yielding and efficient cotton cultivation in the northwest inland cotton-growing region of China. Field Crops Research, 2017, https://doi.org/10.1016/j.fcr.2017.03.008.

6. Yu SX, Wang HT, Wei HL, et al. Research progress and application of early maturity in upland cotton. Cotton Science, 2017, 29(s): 1-10.

7. Li C, Wang X, Dong N, et al. QTL analysis for early-maturing traits in cotton using two upland cotton (Gossypium hirsutum L.) crosses. Breeding Science, 2013, 63(2):154-163.

8. Chen HD, Xiao CS, Li X, et al. DNA fingerprints construction and genetic diversity analysis of short-season cotton by SSR markers. Acta Agriculturae Boreali-Sinica, 2014, 29(4):98-104.

9. Iqbal MJ, Aziz N, Saeed NA. et al. Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. Theoretical and Applied Genetics, 1997, https://doi.org/10.1007/s001220050392.

10. Kumar P, Nimbal S, Budhlakoti N. et al. Genetic diversity and population structure analysis for morphological traits in upland cotton (Gossypium hirsutum L.). Journal of Applied Genetics, 2021, https://doi.org/10.1007/s13353-021-00667-8.

11. Noormohammadi Z, Rahnama A, Sheidai M. EST-SSR and SSR analyses of genetic diversity in diploid cotton genotypes from Iran. Nucleus, 2013, 56(3): 171-178.

12. Qian ZW, Chen HL, Cui YL. Analysis of the SSR loci and development of molecular markers in Spinacia oleracea transcriptome. Journal of Agricultural Biotechnology, 2016, 24 (11): 1688-1697.

13. Xiong Y, Lei X, Bai S. et al. Genomic survey sequencing, development and characterization of single- and multi-locus genomic SSR markers of Elymus sibiricus L. BMC Plant Biol, 2021, https://doi.org/10.1186/s12870-020-02770-0.

14. Santosh HB, Meshram M, Santhy V. et al. Microsatellite marker based diversity analysis and DNA fingerprinting of Asiatic cotton (Gossypium arboreum) varieties of India. Journal of Plant Biochemistry and...
15. Kumar P, Nimbal S, Sangwan R S. et al. Identification of novel marker-trait associations for lint yield contributing traits in upland cotton (*Gossypium hirsutum* L.) using SSRs. Frontiers in Plant Science, 2021, https://doi.org/10.3389/fpls.2021.653270.

16. Nie XH, You CY, Li XF, et al. Construction and genetic diversity analysis of DNA fingerprint of Xinlu Early Cotton varieties. Acta Agronomica Sinica, 2014, 40(12): 2104-2117.

17. Li X, Shahzad K, Guo LP, et al. Using yield quantitative trait locus targeted SSR markers to study the relationship between genetic distance and yield heterosis in upland cotton (*Gossypium hirsutum*). Plant Breeding, 2019, 138(1):105-113.

18. Wang XY, Li XY, Gong ZL, et al. DNA fingerprinting construction and genetic diversity analysis based on SSR markers for upland cotton in Xinjiang. Cotton Science, 2018, 30(4): 308-315.

19. Zhang J F, Stewart J M. Economical and rapid method for extracting cotton genomic DNA. J. Cotton Sci, 2000, 4, 193–201.

20. Yu SX, Fan SL, Wang HT, et al. Progresses in research on cotton high yield breeding in China. Scientia Agricultura Sinica, 2016, 49(18): 3465-3476.

21. Han ZF, Wang JH, Shen GF, et al. Construction of DNA fingerprinting and analysis of genetic diversity with SSR markers for major varieties from the Yellow River Valley. Cotton Science, 2011, 23(6): 545-551.

22. Li CQ, Wang XT, Zhang XF, et al. SSR Fingerprinting Establishment of Baimian series cotton inbred varieties (lines). Cotton Science, 2011, 23(3): 228-234.

23. Kuang M, Yang WH, Xu HX, et al. Construction of DNA fingerprinting and analysis of genetic diversity with SSR markers for cotton major cultivars in China. Scientia Agricultura Sinica, 2011, 44(1): 20-27.

24. Qin J, Li YH, Liu ZX, et al. Genetic relationship among parents of elite soybean (*Glycine max*) cultivars Suinong14 pedigree revealed by SSR markers. Scientia Agricultura Sinica, 2008, 41(12): 3999-4007.

25. Zhang L, Huang Y, Shen XJ, et al. Analysis of genetic relationship of local varieties of *Morus alba* var. *multicaulis* from the lower area of Yellow River based on ISSR marker. Journal of Plant Resources and Environment, 2010, 19(2): 21-27.

**Figures**
Figure 1

A silver stained gel showing primer screening for polymorphic SSRs

Figure 2

Similarity coefficients for the 79 early-maturing upland cotton accessions
Figure 3

The genetic similarity coefficients of early-maturing Upland cotton varieties developed over the past four decades
Figure 4

The genetic lineages of the 72 early-maturing upland cotton varieties based on UPGMA clustering
Figure 5

The pedigrees of 15 upland cotton varieties