Genetic diversity in Brazilian soybean germplasm

Raoni Gwinner¹*, Tesfahun Alemu Setotaw¹, Moacir Pasqual¹, Joao Bosco dos Santos², Alan Mario Zuffo¹, Everton Vinicius Zambiazzi¹ and Adriano Teodoro Bruzi¹

Abstract: Genetic diversity is an essential factor for the success of any plant breeding program and should be considered to ensure genetic gain through breeding. In Brazil, research on the genetic diversity and population structure of soybean is required since the species is an important commodity of the country. The study addressed the genetic diversity and population structure of 77 soybean genotypes using 35 SSR markers. The estimate of the diversity index showed that the level of genetic diversity in the soybean collection is low. Similarly, the Jaccard coefficient and Bayesian model based on clustering analysis confirmed the low diversity among soybean genotypes, providing evidence for the assumption of a genetic bottleneck effect on Brazilian soybean genotypes. The results also reinforced the importance of finding and incorporating new genetic resources of soybean in the genetic pool of Brazilian soybean to warrant genetic gain in soybean breeding in the future.

Key words: Glycine max, molecular markers, SSR, plant breeding.

INTRODUCTION

The expansion of soybean (Glycine max [L.] Merr.) around the world is remarkable. Over the past few decades, the soybean acreage increased drastically, making this crop one of the most widely cultivated worldwide (Phalan et al. 2013), mainly due to its outstanding role as protein and oil source (Clemente and Cahoon 2009). In Brazil, the increase in cultivated area was closely associated with management innovation. As a result, yield rates increased in most cultivated areas (Ray et al. 2012), and the country became the second largest producer in the world (USDA 2017).

With regard to innovation, the development of new crop management technologies and the release of better-adapted genotypes can be considered key factors for the establishment of soybean as one of the major crops in Brazil (Sediyama et al. 2012). Generally, plant breeding has contributed markedly to gains in crop yields, including of soybean. In Iowa for example, plant breeding obtained a genetic gain of 79% between 1930 and 2011 (Smith et al. 2014). However, the success of breeding programs in any crop species depends on the availability of a genetic pool with adequate diversity and its management.

A gradual reduction of genetic diversity would lead inevitably to a reduction in the potential of genetic gains, increasing the susceptibility to biotic stress.
and reducing the adaptability to environmental changes (Smith et al. 2015). Therefore, breeders should focus on the enrichment of genetic diversity in the breeding population, to ensure continuous increases in genetic gain in the long run.

The final goal of genetic diversity studies is the understanding of variation among genotypes or groups of genotypes (Mohammadi and Prasanna 2003). To this end, all types of data can be used, e.g., of genealogy, morphologic traits, the biochemical profile, genotypic data, and many others. Commonly, SSR markers are used in genetic studies, due to their multiallelic nature and repeatability in laboratories (Kaga et al. 2012, Rodrigues et al. 2008). The characterization of plant genotypes by PCR-based DNA markers provides reliable data, for being free of the influence of environmental factors, unlike other methods.

In view of the importance of deepening the understanding on the genetic diversity and population structure of soybean for breeding programs, this study assessed the genetic diversity and population structure of 77 soybean cultivars representative of the genetic resource in Brazil and routinely used in agriculture and research.

MATERIAL AND METHODS

Seventy-seven soybean genotypes, which represent different maturity groups released in Brazil and are part of the breeding program of the Federal University of Lavras (UFLA) were included in this study. The list of the 77 soybean genotypes and their description is shown in Table 1.

The study was carried out in the Laboratory of Molecular Genetics of the Department of Biology (DBI), UFLA. The DNA extraction was performed as described by Pereira et al. (2007). After DNA samples of each genotype were extracted, the genomic DNA was quantified using a NanoVue GE spectrophotometer and standardized to 30 ng µL$^{-1}$ for SSR genotyping. A total of 35 SSR primers were used in this study. The SSR genotyping was realized using the PCR reaction mixture composed of 5.7 µL Milli-Q water, 3 µL 5X buffer (Green GoTaq® Flexi Buffer - Promega), 2 µL MgCl$_2$ (Promega), 100 µM of each deoxyribonucleotide (dATP, dGTP, dCTP, and dTTP), 0.3 µL Taq (5 U µL$^{-1}$), and 3 µL of each primer pair.

Amplifications were performed in Eppendorf Mastercycler® Thermal Cyclers, using the following program: an initial denaturation step at 95 °C for 2 min, followed by 32 cycles at 94 °C for 30 sec each for DNA denaturation, 50 sec for primer annealing at 55 ºC, 40 sec for extension at 72 °C, followed by a final extension at 72 °C for 4 min.

The reaction products were maintained at 4°C and electrophoresed on polyacrylamide gel 6% in TBE buffer (0.045 M Tris-borate and 0.001 M EDTA), at 240V for 1 h 15 min. The gel was immersed in a silver nitrate solution (0.2% AgNO$_3$) for 10 min, washed with water and slowly stirred in a NaOH solution (3% NaOH, 0.5% formaldehyde) until clear visualization, and photographed with a digital camera.

The DNA fragments were codified as presence/absence to form the allelic matrix for each SSR primer.

Data analyses

To study the genetic diversity and population structure, the soybean genotypes were clustered according to their maturity group and variety development method (transgenic or conventional). The relative maturity group (RMG) was determined according to the macroregional classification in the state of Minas Gerais, Brazil (Macro Region 3) where: RMG I (RMG < 7.6); RMG II (7.6 < RMG < 8.2) and RMG III (RMG > 8.2).

The genetic distance of Slatkin (Slatkin 1995) among the soybean genotypes was estimated using software PowerMarker V3.25 (Liu and Muse 2005). Furthermore, the Nei diversity index (Nei 1973), Shannon-Weaver index and percentage of polymorphic loci (P%) were estimated with software Popgene V1.32 (Yeh and Boyle 1997). The Jaccard similarity coefficient was estimated using R statistical environment (R Development Core Team 2014), and used to plot a UPGMA dendrogram with software Mega (Tamura et al. 2007). Analysis of molecular variance (AMOVA) was performed using software GenAlEx 6.501 (Peakall and Smouse 2006).

The Bayesian model based population structure was analyzed using software Structure 2.3.4 (Pritchard et al. 2000) with 10000 burn-ins and 100000 MCMC iterations. To determine the appropriate number of clusters, the process was performed from K = 1 to 12, with 21 interactions for each K. Finally, the ideal cluster number (k), or true k, was determined by the DK statistics K (Evanno et al. 2005), using online software Structure Harvester (Earl and von Holdt 2012).
Table 1. Name, source, maturity group, type and year of release of the evaluated soybean genotypes in this study. "RMG – Relative Maturity Group"

<table>
<thead>
<tr>
<th>#</th>
<th>Genotypes</th>
<th>Source</th>
<th>Type</th>
<th>Year of release</th>
<th>RMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SUPREMA</td>
<td>Nidera</td>
<td>Conventional</td>
<td>1998</td>
<td>III</td>
</tr>
<tr>
<td>2</td>
<td>TMG 801</td>
<td>FMT / TMG / UNISOJA</td>
<td>Conventional</td>
<td>2008</td>
<td>II</td>
</tr>
<tr>
<td>3</td>
<td>BRSMG 820 RR</td>
<td>Embrapa / Epamig</td>
<td>Transgenic</td>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>BRSMG 250 (NOBREZA)</td>
<td>Embrapa</td>
<td>Conventional</td>
<td>2005</td>
<td>II</td>
</tr>
<tr>
<td>5</td>
<td>MG/BR 46 (CONQUISTA)</td>
<td>Embrapa / Epamig</td>
<td>Conventional</td>
<td>1998</td>
<td>II</td>
</tr>
<tr>
<td>6</td>
<td>NA 7337 RR</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2008</td>
<td>I</td>
</tr>
<tr>
<td>7</td>
<td>BRSGO 204 (GOIANIA)</td>
<td>Embrapa / AGDRF</td>
<td>Conventional</td>
<td>2009</td>
<td>II</td>
</tr>
<tr>
<td>8</td>
<td>BRSGO JATAI</td>
<td>Embrapa / EmaterGO</td>
<td>Conventional</td>
<td>1998</td>
<td>III</td>
</tr>
<tr>
<td>9</td>
<td>BRSMG 68 (Vencedora)</td>
<td>Embrapa / Epamig</td>
<td>Conventional</td>
<td>1998</td>
<td>II</td>
</tr>
<tr>
<td>10</td>
<td>SYN 1059 (V-TOP)</td>
<td>Syngenta</td>
<td>Transgenic</td>
<td>2010</td>
<td>I</td>
</tr>
<tr>
<td>11</td>
<td>TMG 123 RR</td>
<td>FMT / TMG / UNISOJA</td>
<td>Transgenic</td>
<td>2007</td>
<td>I</td>
</tr>
<tr>
<td>12</td>
<td>L8307 RR</td>
<td>Monsoy</td>
<td>Transgenic</td>
<td>2006</td>
<td>III</td>
</tr>
<tr>
<td>13</td>
<td>BRSMG GARANTIA</td>
<td>Embrapa / Epamig / APSEMG / COPAMIL</td>
<td>Conventional</td>
<td>2000</td>
<td>III</td>
</tr>
<tr>
<td>14</td>
<td>P 98Y11</td>
<td>Pioneer</td>
<td>Transgenic</td>
<td>2006</td>
<td>II</td>
</tr>
<tr>
<td>15</td>
<td>NK 7074 RR</td>
<td>Syngenta</td>
<td>Transgenic</td>
<td>2007</td>
<td>I</td>
</tr>
<tr>
<td>16</td>
<td>BRS SILVANIA RR</td>
<td>Embrapa</td>
<td>Transgenic</td>
<td>2003</td>
<td>III</td>
</tr>
<tr>
<td>17</td>
<td>M-SOY 6101</td>
<td>Monsoy / CTPA</td>
<td>Conventional</td>
<td>1998</td>
<td>I</td>
</tr>
<tr>
<td>18</td>
<td>ELITE</td>
<td>COOPADAP</td>
<td>Conventional</td>
<td>2001</td>
<td>III</td>
</tr>
<tr>
<td>19</td>
<td>M7211 RR</td>
<td>Monsoy</td>
<td>Transgenic</td>
<td>2007</td>
<td>I</td>
</tr>
<tr>
<td>20</td>
<td>BRSMG 850 GRR</td>
<td>Embrapa</td>
<td>Transgenic</td>
<td>2007</td>
<td>II</td>
</tr>
<tr>
<td>21</td>
<td>MONARCA</td>
<td>COOPADAP</td>
<td>Conventional</td>
<td>1999</td>
<td>III</td>
</tr>
<tr>
<td>22</td>
<td>BRSGO LUZIANIA RR</td>
<td>Embrapa</td>
<td>Transgenic</td>
<td>2007</td>
<td>III</td>
</tr>
<tr>
<td>23</td>
<td>TMG 1181 RR</td>
<td>FMT / TMG / UNISOJA</td>
<td>Transgenic</td>
<td>2009</td>
<td>II</td>
</tr>
<tr>
<td>24</td>
<td>P 98Y30</td>
<td>Pioneer</td>
<td>Transgenic</td>
<td>2009</td>
<td>I</td>
</tr>
<tr>
<td>25</td>
<td>M-SOY 8001</td>
<td>Monsoy</td>
<td>Conventional</td>
<td>1998</td>
<td>II</td>
</tr>
<tr>
<td>26</td>
<td>BRSMG 772</td>
<td>Embrapa / Epamig</td>
<td>Conventional</td>
<td>2013</td>
<td>II</td>
</tr>
<tr>
<td>27</td>
<td>BRS VALIOSA RR</td>
<td>Embrapa</td>
<td>Transgenic</td>
<td>2003</td>
<td>II</td>
</tr>
<tr>
<td>28</td>
<td>NS 5106</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2012</td>
<td>I</td>
</tr>
<tr>
<td>29</td>
<td>NS 5151 IPRO</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2012</td>
<td>I</td>
</tr>
<tr>
<td>30</td>
<td>NA 5909 RG</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2008</td>
<td>I</td>
</tr>
<tr>
<td>31</td>
<td>5D 615 RR</td>
<td>TMG</td>
<td>Transgenic</td>
<td>2011</td>
<td>I</td>
</tr>
<tr>
<td>32</td>
<td>CD 215 RR</td>
<td>COODETEC</td>
<td>Transgenic</td>
<td>2011</td>
<td>I</td>
</tr>
<tr>
<td>33</td>
<td>BMX POTENCIA</td>
<td>Brasmax</td>
<td>Transgenic</td>
<td>2007</td>
<td>I</td>
</tr>
<tr>
<td>34</td>
<td>M-SOY 7908</td>
<td>Monsoy</td>
<td>Transgenic</td>
<td>2004</td>
<td>II</td>
</tr>
<tr>
<td>35</td>
<td>TMG 7161</td>
<td>TMG</td>
<td>Transgenic</td>
<td>2010</td>
<td>I</td>
</tr>
<tr>
<td>36</td>
<td>TMG 1176</td>
<td>FMT / TMG / UNISOJA</td>
<td>Transgenic</td>
<td>2009</td>
<td>II</td>
</tr>
<tr>
<td>37</td>
<td>BRS MG 760 SRR</td>
<td>Embrapa / Epamig</td>
<td>Transgenic</td>
<td>2010</td>
<td>II</td>
</tr>
<tr>
<td>38</td>
<td>TMG 132 RR</td>
<td>TMG</td>
<td>Transgenic</td>
<td>2008</td>
<td>III</td>
</tr>
<tr>
<td>39</td>
<td>5D 688 RR</td>
<td>COODETEC</td>
<td>Transgenic</td>
<td>2010</td>
<td>I</td>
</tr>
<tr>
<td>40</td>
<td>BRS 137</td>
<td>Embrapa</td>
<td>Conventional</td>
<td>2001</td>
<td>I</td>
</tr>
<tr>
<td>41</td>
<td>BRSMG 752 S</td>
<td>Embrapa</td>
<td>Conventional</td>
<td>2008</td>
<td>I</td>
</tr>
<tr>
<td>42</td>
<td>BRSMG 800 A</td>
<td>Embrapa / Epamig</td>
<td>Conventional</td>
<td>2010</td>
<td>II</td>
</tr>
<tr>
<td>43</td>
<td>CD 238 RR</td>
<td>COODETEC</td>
<td>Transgenic</td>
<td>2010</td>
<td>I</td>
</tr>
<tr>
<td>44</td>
<td>CD 250 RR</td>
<td>COODETEC</td>
<td>Transgenic</td>
<td>2010</td>
<td>I</td>
</tr>
<tr>
<td>45</td>
<td>CD 237</td>
<td>COODETEC</td>
<td>Transgenic</td>
<td>2009</td>
<td>II</td>
</tr>
<tr>
<td>46</td>
<td>BRS 284</td>
<td>Embrapa</td>
<td>Conventional</td>
<td>2007</td>
<td>I</td>
</tr>
<tr>
<td>47</td>
<td>BRSMG 790A</td>
<td>Embrapa / Epamig</td>
<td>Conventional</td>
<td>2009</td>
<td>II</td>
</tr>
<tr>
<td>48</td>
<td>BRSMG 780 RR</td>
<td>Embrapa / Epamig</td>
<td>Transgenic</td>
<td>2011</td>
<td>II</td>
</tr>
<tr>
<td>49</td>
<td>NA 7255 RR</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2008</td>
<td>I</td>
</tr>
<tr>
<td>50</td>
<td>NS 7100</td>
<td>Nidera</td>
<td>Transgenic</td>
<td>2009</td>
<td>I</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

Of the analyzed primers, the lowest value for both genetic diversity indices was found for primer Satt222, and the highest Nei's genetic diversity index for primer Satt251 (Table 2). Primer Satt338 reached the highest Shannon index. The mean values for the genetic diversity indices were 0.3863 for Nei's index and 0.6294 for the Shannon index. The highest number of polymorphic alleles was observed at locus Satt270. However, 21 loci had only 2 polymorphic alleles (Table 2).

The polymorphic information content (PIC) was determined for each marker (Table 2), and the mean value was 0.3259. The highest PIC (0.6137) was observed for primer Satt338 and the lowest for Satt222 (0.0253).

A comparison of the soybean varieties obtained by transgenic and conventional breeding indicated slightly higher genetic diversity indices for the varieties derived from conventional breeding (Table 2). The transgenic cultivars had 100% and the conventional group 88.57% polymorphic loci.

Except for RMG III, the number of alleles was similar for all subsets. Subset RMG II had the highest and the RMG I the lowest index for both genetic diversity indices. Moreover, these two contained the highest PIC. For subset RMG III, the smallest number of polymorphic alleles and lowest Shannon diversity index were found (Table 3).

For the transgenic and conventional subsets, AMOVA showed that 99% of the total variation was partitioned within the subsets and the remaining 1% among subsets. Similarly, for the maturity groups, 94% of the total variation was partitioned within and 1% among the groups (Table 3).

The Bayesian model-based clustering analysis clustered the 77 genotypes in two principal groups (K=2). The first group (red) consisted of mostly short-season or early cultivars, while the second group (green) of mostly long-season or late cultivars. There is certain degree of admixture represented mostly by genotypes with medium cycle, but also by RMG I and III (Figure 1). The formation of only two groups indicated low genetic diversity among the assessed soybean genotypes.
Low genetic diversity was observed among the soybean genotypes and in all subsets analyzed in this study. The genetic diversity indices estimated in this study can be considered low when compared to results previously described for this species. In 100 vegetable soybean cultivars, Dong et al. (2014) analyzed the genetic diversity and reported a mean Nei index of 0.6286. In an analysis of the genetic diversity of 10 landrace populations in China, Wang et al. (2014) found a Shannon-Weaver index of 2.038. The high Shannon-Weaver indices observed by the authors in these two studies may be explained by the kind of genotypes used (vegetable and landrace). The selection pressure on vegetable soybean was lower than on the grain-type. Moreover, landrace genotypes are closer relatives of the wild genotypes (G. soja), which naturally have a high level of genetic diversity. Most genotypes included in this study are improved soybean genotypes subjected to a rather high selection pressure.

Low diversity in a germplasm is particularly alarming when we consider the vulnerability of agriculture under the impact of changes in climate patterns. The development of genotypes with better performance at high temperatures, high CO$_2$ concentration, and under water and salt stress has become a challenge that must be met with a certain urgency due to the ongoing climate changes (Ceccarelli et al. 2010). However, facing these challenges and developing soybean cultivars with the desirable traits requires a diversification of the genetic background of the current breeding population by incorporating new genetic backgrounds from other countries.

The genetic base of Brazilian soybean germplasm is known to be narrow (Wysmierski and Vello 2013, Miranda et al. 2001). The frequent use of a small set of genotypes in the breeding process could be considered a key factor for the loss of genetic diversity. Most of the Brazilian soybean germplasm is derived from four genotypes (CNS, S-100, Roanoke and Tokyo), which contribute with more than half of the genetic base to the cultivars released in Brazil.

<p>| Table 2. | Microsatellite loci, number of alleles (n), Nei’s genetic diversity index (H), Shannon-Weaver index (I) and Polymorphic Information Content (PIC) |</p>
<table>
<thead>
<tr>
<th>Primer name</th>
<th>n</th>
<th>H</th>
<th>I</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satt200</td>
<td>2.0</td>
<td>0.4979</td>
<td>0.6910</td>
<td>0.3739</td>
</tr>
<tr>
<td>Satt239</td>
<td>3.0</td>
<td>0.4876</td>
<td>0.8249</td>
<td>0.4241</td>
</tr>
<tr>
<td>Satt270</td>
<td>5.0</td>
<td>0.5943</td>
<td>1.1463</td>
<td>0.5508</td>
</tr>
<tr>
<td>Satt196</td>
<td>2.0</td>
<td>0.1214</td>
<td>0.2403</td>
<td>0.1141</td>
</tr>
<tr>
<td>Satt492</td>
<td>2.0</td>
<td>0.1847</td>
<td>0.3315</td>
<td>0.1676</td>
</tr>
<tr>
<td>Satt358</td>
<td>3.0</td>
<td>0.3771</td>
<td>0.6541</td>
<td>0.3298</td>
</tr>
<tr>
<td>Satt012</td>
<td>4.0</td>
<td>0.5158</td>
<td>0.9784</td>
<td>0.4749</td>
</tr>
<tr>
<td>Satt076</td>
<td>2.0</td>
<td>0.4252</td>
<td>0.6164</td>
<td>0.3348</td>
</tr>
<tr>
<td>Satt225</td>
<td>2.0</td>
<td>0.4898</td>
<td>0.6829</td>
<td>0.3698</td>
</tr>
<tr>
<td>Satt137</td>
<td>3.0</td>
<td>0.4192</td>
<td>0.7532</td>
<td>0.3810</td>
</tr>
<tr>
<td>Satt475</td>
<td>3.0</td>
<td>0.3287</td>
<td>0.6089</td>
<td>0.2995</td>
</tr>
<tr>
<td>Satt241</td>
<td>3.0</td>
<td>0.4557</td>
<td>0.6945</td>
<td>0.3631</td>
</tr>
<tr>
<td>Satt345</td>
<td>3.0</td>
<td>0.6044</td>
<td>0.9908</td>
<td>0.5213</td>
</tr>
<tr>
<td>Satt426</td>
<td>2.0</td>
<td>0.2088</td>
<td>0.3638</td>
<td>0.1870</td>
</tr>
<tr>
<td>Satt225</td>
<td>3.0</td>
<td>0.6610</td>
<td>1.0900</td>
<td>0.5869</td>
</tr>
<tr>
<td>Satt509</td>
<td>2.0</td>
<td>0.2922</td>
<td>0.4678</td>
<td>0.2495</td>
</tr>
<tr>
<td>Satt338</td>
<td>4.0</td>
<td>0.6606</td>
<td>1.2303</td>
<td>0.6137</td>
</tr>
<tr>
<td>Satt369</td>
<td>3.0</td>
<td>0.4459</td>
<td>0.7856</td>
<td>0.4005</td>
</tr>
<tr>
<td>Satt342</td>
<td>2.0</td>
<td>0.4581</td>
<td>0.6506</td>
<td>0.3532</td>
</tr>
<tr>
<td>Satt274</td>
<td>2.0</td>
<td>0.2449</td>
<td>0.4101</td>
<td>0.2149</td>
</tr>
<tr>
<td>Satt520</td>
<td>2.0</td>
<td>0.4554</td>
<td>0.6478</td>
<td>0.3517</td>
</tr>
<tr>
<td>Satt052</td>
<td>2.0</td>
<td>0.4932</td>
<td>0.6863</td>
<td>0.3716</td>
</tr>
<tr>
<td>Satt222</td>
<td>2.0</td>
<td>0.0256</td>
<td>0.0693</td>
<td>0.0253</td>
</tr>
<tr>
<td>Satt302</td>
<td>2.0</td>
<td>0.4992</td>
<td>0.6924</td>
<td>0.3746</td>
</tr>
<tr>
<td>Satt146</td>
<td>2.0</td>
<td>0.4369</td>
<td>0.6286</td>
<td>0.3415</td>
</tr>
<tr>
<td>Sct_034</td>
<td>2.0</td>
<td>0.2476</td>
<td>0.4135</td>
<td>0.2169</td>
</tr>
<tr>
<td>Satt553</td>
<td>3.0</td>
<td>0.3542</td>
<td>0.6589</td>
<td>0.3259</td>
</tr>
<tr>
<td>Satt486</td>
<td>2.0</td>
<td>0.0263</td>
<td>0.0708</td>
<td>0.0260</td>
</tr>
<tr>
<td>Satt386</td>
<td>2.0</td>
<td>0.2975</td>
<td>0.4741</td>
<td>0.2533</td>
</tr>
<tr>
<td>Satt513</td>
<td>4.0</td>
<td>0.4932</td>
<td>0.8421</td>
<td>0.4163</td>
</tr>
<tr>
<td>Satt476</td>
<td>3.0</td>
<td>0.4879</td>
<td>0.7285</td>
<td>0.3810</td>
</tr>
<tr>
<td>Satt471</td>
<td>2.0</td>
<td>0.1244</td>
<td>0.2449</td>
<td>0.1167</td>
</tr>
<tr>
<td>Satt150</td>
<td>2.0</td>
<td>0.4473</td>
<td>0.6395</td>
<td>0.3473</td>
</tr>
<tr>
<td>Satt567</td>
<td>2.0</td>
<td>0.3682</td>
<td>0.5548</td>
<td>0.3004</td>
</tr>
<tr>
<td>Satt384</td>
<td>2.0</td>
<td>0.2896</td>
<td>0.4648</td>
<td>0.2477</td>
</tr>
<tr>
<td>Mean</td>
<td>2.5429</td>
<td>0.3863</td>
<td>0.6294</td>
<td>0.3259</td>
</tr>
</tbody>
</table>

Table 3. Evaluated soybean subsets (Subsets), number of genotypes per subset (#), mean of observed alleles (n), Shannon-Weaver diversity index (I), Nei diversity index (H), polymorphism percentage (PP)

<table>
<thead>
<tr>
<th>Subsets</th>
<th>#</th>
<th>n</th>
<th>I</th>
<th>H</th>
<th>PP(%)</th>
<th>AMOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Among (%)</td>
</tr>
<tr>
<td>Conventional</td>
<td>23</td>
<td>2.4</td>
<td>0.6198</td>
<td>0.3817</td>
<td>88.57</td>
<td>1</td>
</tr>
<tr>
<td>Transgenic</td>
<td>54</td>
<td>2.4857</td>
<td>0.6102</td>
<td>0.3761</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>RMG I</td>
<td>38</td>
<td>2.4286</td>
<td>0.5611</td>
<td>0.3426</td>
<td>97.14</td>
<td>6</td>
</tr>
<tr>
<td>RMG II</td>
<td>28</td>
<td>2.4571</td>
<td>0.6230</td>
<td>0.3823</td>
<td>97.14</td>
<td></td>
</tr>
<tr>
<td>RMG III</td>
<td>11</td>
<td>2.0571</td>
<td>0.5532</td>
<td>0.3583</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td>2.5429</td>
<td>0.6294</td>
<td>0.3863</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
If we consider the 14 soybean genotypes used in the breeding program, their contribution to the genetic base can reach up to 92.4% (Wysmierski and Vello 2013). This conclusion is supported by the result of this study, where 77 soybean cultivars clustered into only two groups, with a high fixation index in each cluster. In summary, the soybean genetic base used in the breeding program of Brazil was constructed from a small number of genotypes. Our data confirmed this pattern and reinforced the results described by Wysmierski and Vello (2013). In addition, the methods of soybean breeding contribute to aggravate the genetic bottleneck, since backcrossing is a routinely applied method to introduce qualitative traits into an elite cultivar, particularly in the case of transgenic lines. The genetic diversity indices of the subset of transgenic cultivars in this study were lower than those of traditional genotypes, despite the higher number of genotypes included in this group (Table 3).

The development of transgenic cultivars by backcrossing can lead to a reduction in the genetic richness since a low number of recurrent parents and a few donor parents carrying the transgenic segment are used in the process. Furthermore, the small number of transformed lines reduces the range of genetic variation available for selection by the plant breeder (Sneller 2003).

Considering the rising demand for transgenic and early cultivars in the seed market, we can assume that the man-made genetic bottleneck is currently still an ongoing process that should be mitigated. The concern about genetic diversity resulting in efforts to enrich the soybean genetic pool can be helpful to preserve long-term performance increments by means of plant breeding.

The result of AMOVA showed that the absence of a clear differentiation among the groups of soybean genotypes indicates a narrow genetic base of the soybean cultivars grown in Brazil (Table 3). Therefore, soybean breeders should follow a strategy of broadening the genetic base of different groups by the incorporation of new genetic bases, particularly from landraces.

Most of the soybean cultivars analyzed in this study are early genotypes (RMG I). Nowadays, early cultivars have become rather popular among Brazilian farmers who use a production system based on crop succession, with an early-maturing cultivar in the summer followed by second-season crop after a short time gap. This widely applied system is called “safrinha” or late growing season in Brazil, and aims at a maximized land use in relation to time. Thus, this production model drives the seed companies to focus on an accelerated development of early cultivars, which may be another factor squeezing the genetic bottleneck.

The lowest Nei diversity index (H=0.3426) was observed in the early group (RMG I), suggesting low genetic diversity within this group. The subset RMG III obtained the lowest Shannon-Weaver index, suggesting low diversity as well. Nevertheless, RMG III is composed of a smaller number of genotypes, which can result in an underestimation of its diversity range (Table 3).

In our results, the absence of a clear clustering pattern in the dendrogram regarding relative maturity groups, and the short distance between clusters (Figure 2) suggested that the early and late cultivars may share common alleles obtained from the same parents used in the breeding process.

A pattern was observed in the population structure, where two groups were formed. The first group consisted of

---

**Figure 1.** Two subgroups inferred from STRUCTURE analysis. The vertical coordinate indicates the membership coefficients for each plant, and the digits on the horizontal coordinate represent the reference number of the accessions listed in Table 1.
Figure 2. Dendrogram based on Jaccard’s similarity coefficient, showing soybean genotypes of different maturity groups: RMG I (red), RMG II (yellow), RMG III (green).
73.9% plants from subset RMG I and the second group of 61.3% plants from subset RMG II and 25.8% from RMG III, indicating a higher allele frequency of alleles associated to early genotypes in the first group and in the second group a higher frequency of alleles associated to late genotypes (Figure 1).

In general, the result evidenced low genetic variability among soybean cultivars, reinforcing previous results. The popularization in the development and use of early transgenic cultivars should be followed by mitigating measures to alleviate the human-made bottleneck. In this context, the soybean breeding programs need to consider the genetic diversity issue as an important factor instead of focusing exclusively on yield and other agronomic traits.

In this regard, wild soybean species such as Glycine soja Seib. and Zucc. can be considered to increase genetic variation and used as an important source of alleles, despite their undomesticated characteristics (Akperetey et al. 2014). Interspecific crosses are commonly used in several crop species with the objective of introducing new alleles to overcome biotic and abiotic stress and increase the genetic diversity (Haussmann et al. 2004). Unfortunately, there is a lack of availability of this genetic material in the Brazilian germplasm (Carter et al. 2004).

Efforts for an international cooperation of sharing and exchanging soybean germplasm should be considered to avoid the surge of a genetic bottleneck in many countries of the world and to enrich the Brazilian soybean germplasm pool. In Brazil, attention must be paid to the threat of this bottleneck effect looming over the development of soybean breeding programs in the future.

An expansion of the soybean genetic base is fundamental to ensure the future progress of genetic gains by plant breeding. A more diverse genetic pool and germplasm set can facilitate the development of new varieties capable to overcome a hostile environment.

The analyzed soybean genotypes contain low genetic diversity, indicating the need to increase the variability in the genetic pool of the soybean breeding program of the Federal University of Lavras.

**ACKNOWLEDGEMENTS**

We would like to thank CNPq (Brazilian National Council for Scientific and Technological Development), CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education), FAPEMIG (Minas Gerais State Foundation for Research Development), and the Federal University of Lavras for financial support.

**REFERENCES**


Genetic diversity in Brazilian soybean germplasm


