Domestication to Crop Improvement: Genetic Resources for Sorghum and Saccharum (Andropogoneae)

SALLY L. DILLON, FRANCES M. SHAPTER, ROBERT J. HENRY, GIOVANNI CORDEIRO, LIZ IZQUIERDO and L. SLADE LEE

1Centre for Plant Conservation Genetics, Southern Cross University, PO Box 157, Lismore, NSW 2480, Australia and 2Australian Tropical Crops and Forages Centre, Queensland Department of Primary Industries and Fisheries, Biloela, QLD 4715, Australia

Received: 28 May 2007 Returned for revision: 2 July 2007 Accepted: 24 July 2007 Published electronically: 1 September 2007

Key words: Genomics, sorghum, Sorghum bicolor, sugarcane, Saccharum officinarum, crop improvement, domestication.

INTRODUCTION

Crop plants were first cultivated around 10,000 years ago. However, crop domestication and development began much more recently (Doggett, 1970). Innumerable varieties, races and cultivars of agricultural plants have been developed to support human and animal demand for food, fibre and building materials. The Poaceae are an important global source of dietary protein, carbohydrates and other nutrients. Sorghum [Sorghum bicolor (L.) Moench] is the fourth most important cereal crop behind wheat, rice and maize, and is grown throughout the arid and semi-arid tropics (Smith and Cordeiro, 2004). Sugarcane is the fourth most important sugar-producing crop globally and is grown throughout the arid and semi-arid tropics (Smith and Cordeiro, 2000). Sugarcane is a complex polyploid with a large and variable number of copies of each gene. The wild relatives of sugarcane represent a reservoir of genetic diversity for use in sugarcane improvement. Techniques for quantitative molecular analysis of gene or allele copy number in this genetically complex crop have been developed. SNP discovery and mapping in sugarcane has been advanced by the development of high-throughput techniques for ecoTILLING in sugarcane. Genetic linkage maps of the sugarcane genome are being improved for use in breeding selection. The improvement of both sorghum and sugarcane will be accelerated by the incorporation of more diverse germplasm into the domesticated gene pools using molecular tools and the improved knowledge of these genomes.

Sorghum and its wild relatives

Sorghum domestication

Archaeological evidence suggests that hunter-gatherers consumed sorghum as early as 8000 BC (Smith and Frederiksen, 2000). The domestication of sorghum has its origins in Ethiopia and surrounding countries, commencing around 4000–3000 BC. Numerous varieties of sorghum were created through the practice of disruptive selection, whereby selection for more than one level of a particular character within a population occurs (Doggett, 1970). This results from a balance of farmer selection for cultivated traits and natural selection for wild characteristics, generating both improved sorghum types, wild types and intermediate types (Doggett, 1970). These improved sorghum types were spread via the movement of people and trade routes into other regions of Africa, India (approx. 1500–1000 BC), the Middle East (approx. 900–700 BC) and eventually into the Far East (approx. AD 400). By the time sorghum was...
transported to America during the late 1800s to early 1900s, the diversity of new sorghum types, varieties and races created through the movement of people, disruptive selection, geographic isolation and recombination of these types in different environments would have been large (Wright, 1931; Doggett, 1970).

Initial domestication of sorghum would have focused primarily on converting wild types with small, shattering (dehiscent) seed to improved types with larger, non-shattering seed. Disruptive selection resulted in sorghum sections: *Eu-Sorghum*, *Chaetosorghum*, *Heterosorghum*, *Para-Sorghum* and *Stiposorghum* (Garber, 1950). Section *Eu-Sorghum* contains all domesticated/cultivated sorghum races and varieties as *Sorghum bicolor* subsp. *bicolor*, as well as the wild and weed species *S. halepense* (L.) Pers. (Johnsons grass), *S. propinquum* (Kunth) Hitchc, *S. × almum* Parodi, *S. × drummondii* (Steed.) Millsp. & Chase, and *S. arundinaceum* (Desv.) Stapf. (the known progenitor of *S. bicolor*) (Harlan and de Wet, 1971; Doggett, 1988). All *S. bicolor* subsp. *bicolor* have $2n = 2x = 20$ chromosomes, and are described as annual, with thick culms up to 5 m in height, often branched with many tillers. They have been classified into five basic races: *bicolor*, *guinea*, *caudatum*, *kafr* and *durra*, with ten intermediate races of these also recognized (Harlan and de Wet, 1972). These 15 races of cultivated sorghum are recognizable on spikelet/panicle morphology alone, and can be linked back to their specific environments and the nomadic peoples that first cultivated them (Smith and Frederiksen, 2000).

A comprehensive analysis of genetic diversity in sorghum landraces and core collections based on race, latitude of origin, photoperiod, seed quality, agronomic traits and DNA markers has demonstrated sorghum has considerable polymorphism that has been poorly exploited in terms of crop improvement (Wu et al., 2004; Abu Assat et al., 2005; Deu et al., 2006; Kayode et al., 2006). At the DNA level, two high-density maps have been completed, one intraspecific and another from an interspecific cross (between *S. bicolor* and *S. propinquum*). These maps showed a high colinearity from which the divergence between *Sorghum* species and the diversity within cultivated *S. bicolor* has been indicated (Feltus et al., 2006).

Changing characteristics/trait of domesticated sorghum and effects on yield

Early domestication of sorghum was associated with changing the small-seeded, shattering open panicles towards larger, non-shattering seeds and more compact panicles. This involved several factors: significantly increasing the number of branches within the inflorescence; decreasing the internode length of the rachis; and an increase in seed size so it protruded out of the glumes (House, 1985). These changes contributed to an increase in yield over the original sorghum landrace varieties.

Stable, high-yielding sorghum varieties have been recently developed through breeding/improvement programmes utilizing sorghum landrace varieties from Africa, India and China. This has involved selecting traits such as photoperiod insensitivity, reduced height (to reduce lodging), drought tolerance, and pest and disease resistance (Reddy et al., 2006).

Plant height and photoperiod insensitivity were the focus of conversion programmes that developed sorghum lines with desirable plant height and maturity that were usable in breeding programmes in both tropical, short-day environments and in long-day, temperate and subtropical environments. As sorghum originated in north-eastern Africa, the many landraces and early varieties were photoperiod sensitive, with a critical photoperiod of 12 h: once the day length is shorter than 12 h, the sorghum plant changes from vegetative to reproductive growth (Reddy et al., 2006). Growing these photoperiod-sensitive landraces/lines as a summer crop in temperate zones of America and Australia where the day length is longer than 13 h was difficult, especially as many growth-related characteristics are poorly expressed under these long-day conditions (Reddy et al., 2006). This made breeding improved varieties in semi-arid temperate and subtropical climates difficult. Cultivars and landraces were identified in India that had higher critical photoperiods, with no delay in flowering observed when grown in day lengths up to 17 h. These photoperiod-insensitive sorghum cultivars have since been widely adopted in breeding programmes throughout the world (Rai et al., 1999; Reddy et al., 2006).

Plant height and grain yield are highly correlated in some populations of sorghum, with maximum productivity achieved at heights of around 1.75–1.80 m and flowering at 68–70 d (Miller, 1982; Rao and Rana, 1982). However, plants of these heights easily lodge, and are not easily cultivated under modern farming practices. A selection of high-yielding, tall sorghum landraces/lines were crossed to shorter, photoperiod-insensitive sorghum lines to develop improved high-yielding cultivars with a shorter stature (Miller, 1980; Rosenow and Dahlberg, 2000).

Sorghum is grown predominantly in low-rainfall, arid to semi-arid environments. The occurrence of drought stress is a major constraint to sorghum production globally. Two forms of drought stress have been identified in sorghum: ‘pre-anthesis’ where plants are stressed during panicle differentiation prior to flowering; and ‘post-anthesis’ when moisture stress occurs during the grain fill stage (Rosenow and Clark, 1995). The identification of varieties and lines with naturally high levels of pre-anthesis drought tolerance and the selection of these for higher yields has developed sorghum varieties with stable, high yields (Ellis et al., 1997). Post-anthesis drought stress can result in significant yield loss due to small grain size, premature plant death and susceptibility to diseases.
Post-flowering drought tolerance is referred to as stay-green, with plants maintaining green leaf area and photosynthetic capability under severe moisture stress, which results in higher grain yields compared with senescent varieties (Borrell and Douglas, 1997; Borrell et al., 1999). The physiological components of stay-green (green leaf area at flowering; time of onset of senescence; rate of senescence) are independently inherited and easily combined through breeding, resulting in new sorghum varieties exhibiting high levels of stay-green with stable high yields and good levels of insect resistance (Borrell et al., 2000).

Sorghum production is affected by many pests and diseases globally. Some of the major pests include midge (*Stenodiplosis sorghicola* Coquillett), green bug (*Schizaphis*...
graminum Rondani), various aphids, shootfly (Atherigona soccata Rondani) and stem borer (Chilo partellus Swinhoe) (Sharma, 1993). Major diseases include downy mildew, anthracnose, sorghum rust, leaf blight, ergot and head and kernel smut (House, 1985). Success in breeding for insect resistance in sorghum varieties has been varied. Resistance to some pests is quantitatively inherited and therefore difficult to transfer into high-yielding cultivars (Tao et al., 2003). The exception to this is midge resistance, where high levels of midge immunity have been incorporated from Indian, American and Australian breeding lines into elite, high-yielding sorghum varieties in Australia, with greater than 80 % of the planted area utilizing these resistant varieties (Jordan et al., 1998; Tao et al., 2003).

Development of disease-resistant sorghum varieties has relied on identifying sorghum varieties/landraces with natural genetic resistance to the particular disease. To date, commercial sorghum varieties have been developed with resistance to grain moulds and anthracnose (Reddy et al., 2006).

The development of photoperiod-insensitive, dwarfed sorghum varieties with some levels of pest/disease resistance has improved the yields of cultivated sorghum varieties. However, the development of a hybrid cropping system is responsible for increases in yields of more than 300 % since the 1950s (Rooney and Smith, 2000). Hybrid cultivars make use of male sterility to enhance the combining abilities of the parental lines, resulting in heterosis and significant increases in phenotypic traits such as yield, plant height and days to flowering (Reddy et al., 2006).

Although the domestication and resulting super-domestication of sorghum has relied on principally S. bicolor subsp. bicolor varieties/landraces/lines for significant gains in agricultural production, the undomesticated Sorghum species offer an untapped wealth of novel traits for both biotic and abiotic stress resistance and yield.

Undomesticated Sorghum species as genetic resources for sorghum improvement

All cultivated sorghum varieties and landraces are S. bicolor subsp. bicolor of the Eu-Sorghum subgeneric section of the Sorghum genus. The other four sections, Chaetosorghum, Heterosorghum, Para-Sorghum and Stiposorghum contain 19, wild species native to Africa, Asia and Australia (Garber, 1950; Lazarides et al., 1991). These species are briefly outlined below, and contain new sources of genetic diversity for agronomic traits affecting yield, survivability and novel traits that may create new markets for sorghum products.


The adaptability of these undomesticated Sorghum species to colonize a wide range of soil and moisture conditions across a wide range of microenvironments is shown through their ability to survive very hot, dry, nutrient-limited environments. Due to their adaptability, many of the undomesticated Sorghum species have developed resistances to the many pests and diseases that affect sorghum grain production globally. Interestingly, many Australian undomesticated species contain resistances to the major pest/diseases of Africa and America, which are not yet present within Australia (Bapat and Mote, 1982; Karunakar et al., 1994; Franzmann and Hardy, 1996; Sharma and Franzmann, 2001; Kamala et al., 2002; Komolong et al., 2002).

Recent controlled-environment glasshouse trials have shown that the undomesticated Sorghum species, though adapted to specific abiotic conditions in the wild, showed prolific growth under moderate temperature in a standard potting mix and watered regularly (Table 1). These data show useful variations to germination times and time to flowering. Representatives of the undomesticated Heterosorghum, Para-Sorghum, Stiposorghum and a Eu-Sorghum were grown concurrently to compare their development under controlled conditions (Fig. 2). Cultivated S. bicolor takes 3–10 d to germinate depending on soil temperatures, with the first 30–35 d post-germination undergoing lower leaf growth followed by a rapid elongation in non-dwarf varieties. Flowering in S. bicolor occurs 55–70 d post-germination and seeds reach physiological maturity 30–40 d post-anthesis. It then takes 20–25 d to reduce the moisture content to the 12 % required for post-harvest storage (House et al., 1995). There appears to be limited differences between undomesticated species and S. bicolor for these traits (Table 1).

Undomesticated Sorghum species: grain attributes

The morphology of seed size and shape within the Sorghum genus varies greatly. Figure 3 shows the morphology of the domesticated S. bicolor subsp. bicolor and undomesticated Eu-Sorghum, Chaetosorghum, Heterosorghum, Para-Sorghum and Stiposorghum species. Variation in the grain morphology of representatives of the undomesticated Heterosorghum, Para-Sorghum and Stiposorghum species have also been evaluated at the microscopic level. Mature caryopses of 13 species were critically point dried, snap fractured and examined using a Leostereoscan 440 scanning electron
microscope to determine if novel variations existed in the undomesticated species (Shapter et al., 2007).

The endosperm of cultivated *S. bicolor* is described as having two distinct regions or layers. The floury central endosperm (Fig. 4A) contains simple round or lenticellar starch granules in a discontinuous protein matrix with few if any protein bodies present. The vitreous or conewy outer endosperm (Fig. 4B) is characterized by polygonal starch granules, 4–25 μm in diameter, the surface of which is typically indented from the protein bodies that are part of the continuous protein matrix surrounding the granules. Variations to the distribution and configuration of these two regions have been shown to alter the functional and putatively the nutritional value of sorghum flours and other foods (Serna-Saldivar and Rooney, 1995; Lindeboom et al., 2004; Tesso et al., 2006). One species appeared to have sections of the endosperm with small rice-like starch granules, usually only seen in the sub-aleurone layer in *S. bicolor* (Shapter et al., 2007). Several wild species also maintained a single morphology across the entire endosperm, rather than the two layers seen in *S. bicolor*. Amongst these differences some species retained the characteristic morphology of the *S. bicolor* vitreous layer (Fig. 4F).

The sub-aleurone of *S. bicolor* is described as being 15–30 μm wide and is an area of very small starch granules and denser protein matrix, the endosperm proper (Fig. 5A and B). In the *Para-Sorghum* and *Stiposorghum* species examined, areas of the sub-aleurone have a striated appearance (Fig. 5C) not previously reported in microscopy studies (F. M. Shapter et al., unpubl. res.). Investigation of these areas under high magnification showed what appeared to be a much denser protein matrix, embedded with spherical-shaped bodies reminiscent of protein bodies. Within this layer, small starch granules typical of sub-aleurone starch granules are interspersed (Fig. 5D and E). More investigation is needed to confirm if this layer is proteinaceous. From an adaptive point of view, the development of a highly proteinaceous layer directly below the aleurone would provide a rich nitrogen source for the germinating seedling when establishing itself in low nitrogen soils, typical of northern Australia where many of these undomesticated species are endemic. Protein/starch interactions in sorghum have been shown to decrease starch digestibility, especially after cooking (Duodo et al., 2003). The occurrence of increased protein content in the endosperm may therefore result in a further decrease in starch digestibility.

### Table 1. Germination, growth rates and flowering times for undomesticated Sorghum species

<table>
<thead>
<tr>
<th>Species</th>
<th>DTT*</th>
<th>30 d</th>
<th>60 d</th>
<th>90 d</th>
<th>130 d</th>
<th>Days till flower</th>
<th>10 dpa</th>
<th>20 dpa</th>
<th>30 dpa</th>
<th>40 dpa</th>
<th>Final max count</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. amplum</em></td>
<td>4-5</td>
<td>25</td>
<td>40</td>
<td>40</td>
<td>125 fl</td>
<td>136</td>
<td>md</td>
<td>16</td>
<td>29</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. angustatum</em></td>
<td>5-5</td>
<td>18</td>
<td>45</td>
<td>62 fl</td>
<td>95 fl</td>
<td>86</td>
<td>9</td>
<td>20-5</td>
<td>58</td>
<td>146</td>
<td>&gt;300</td>
</tr>
<tr>
<td><em>S. brachypodum</em></td>
<td>3-5</td>
<td>35</td>
<td>70</td>
<td>150 fl</td>
<td>210 fl</td>
<td>86</td>
<td>4</td>
<td>26-5</td>
<td>45</td>
<td>62-5</td>
<td>&gt;300</td>
</tr>
<tr>
<td><em>S. bulbosum</em></td>
<td>3-0</td>
<td>30</td>
<td>55</td>
<td>120 fl</td>
<td>170 fl</td>
<td>95</td>
<td>22</td>
<td>67</td>
<td>89</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. extans</em> (p)</td>
<td>3-5</td>
<td>12</td>
<td>45 fl</td>
<td>70 fl</td>
<td>95 fl</td>
<td>65</td>
<td>10</td>
<td>17</td>
<td>47</td>
<td>113</td>
<td>235</td>
</tr>
<tr>
<td><em>S. intrans</em> (p)</td>
<td>3-5</td>
<td>5</td>
<td>20 p</td>
<td>20 p</td>
<td>85-0</td>
<td>&gt;160</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. laxiflorum</em></td>
<td>5-5</td>
<td>60</td>
<td>60 fl</td>
<td>90 fl</td>
<td>95 fl</td>
<td>61</td>
<td>md</td>
<td>16-5</td>
<td>44</td>
<td>94</td>
<td>214</td>
</tr>
<tr>
<td><em>S. leiocladam</em></td>
<td>5-5</td>
<td>27</td>
<td>30</td>
<td>35</td>
<td>85-0</td>
<td>141</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. matorankeense</em></td>
<td>4-5</td>
<td>58</td>
<td>88 fl</td>
<td>130 fl</td>
<td>150 fl</td>
<td>81</td>
<td>2</td>
<td>5</td>
<td>23</td>
<td>53</td>
<td>69</td>
</tr>
<tr>
<td><em>S. nitidum</em></td>
<td>4-0</td>
<td>25</td>
<td>48</td>
<td>60</td>
<td>90</td>
<td>&gt;160</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. plumosum</em></td>
<td>8-0</td>
<td>8</td>
<td>15</td>
<td>15</td>
<td>60</td>
<td>&gt;160</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. propinquum</em></td>
<td>3-0</td>
<td>45</td>
<td>50</td>
<td>100</td>
<td>240 fl</td>
<td>106</td>
<td>19</td>
<td>35</td>
<td>–</td>
<td>–</td>
<td>&gt;300</td>
</tr>
<tr>
<td><em>S. stipoides</em></td>
<td>4-0</td>
<td>45</td>
<td>70</td>
<td>60 fl</td>
<td>130 fl</td>
<td>91</td>
<td>18</td>
<td>60</td>
<td>112</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>S. timorense</em></td>
<td>3-0</td>
<td>45</td>
<td>100 fl</td>
<td>100 fl</td>
<td>145 fl</td>
<td>69</td>
<td>md</td>
<td>3</td>
<td>14</td>
<td>18</td>
<td>63</td>
</tr>
</tbody>
</table>

* DTT. Days till transplant: seeds were germinated on damp filter paper; once a strong radicle and the first coleoptile had emerged they were transplanted to potting mix.

md, Missing data, (p), prostrate growth habit; fl, height was measured to the top of flag leaf or seed head.

A dash indicates that delayed onset of flowering caused the trial to be terminated before counts could be made.

FIG. 2. Growth trial of *Sorghum* species at seedling stage. Note the broader leaf (far left) of the *Eu-sorghum*, *S. propinquum* compared with the *Para-Sorghum*, *Stiposorghum* and *Heterosorghum* species.
which has utility for raising the glycaemic index of foods for Western diets.

**Hybridizing potential of undomesticated sorghum species**

Modern sorghum breeding programmes have not used species outside of section *Eu-Sorghum* as sources of genetic diversity due to a lack of information regarding the genetic relationships between the species. Recent phylogenetic analysis of all 25 *Sorghum* species based on the three gene sequences ITS1, *ndhF* and *Adh1* has identified *S. macrospermum* and *S. laxiflorum* as the undomesticated species outside of *Eu-Sorghum* most closely related to cultivated sorghum varieties (Dillon *et al.*, 2007). The relationships identified can now act as a guide for plant breeders.

Most of the undomesticated *Sorghum* species fall within the tertiary genepool, making gene transfer to domesticated species very difficult due to strong sterility barriers (Harlan and de Wet, 1971). The nature of the sterility barriers in *Sorghum* have recently been identified as pollen–pistil incompatibilities whereby the pollen of undomesticated species behaves abnormally in the pistils of *S. bicolor*, resulting in no hybrid embryo formation (Hodnett *et al.*, 2005). As a result, pollen rarely grew beyond the stigma of *S. bicolor*; however, a single embryo was formed using *S. macrospermum* pollen. The embryo of this *S. bicolor* × *S. macrospermum* cross was rescued and raised through tissue culture, with the seedling verified as a hybrid based upon cytological and morphological characteristics (Price *et al.*, 2005b).

Although a hybrid embryo was formed and able to be rescued via tissue culture, pollen–pistil incompatibilities make this an extremely rare occurrence. Methods of increasing the frequency of hybridization are required to successfully utilize the undomesticated *Sorghum* species. An *S. bicolor* accession was discovered containing a recessive gene (inhibition of alien pollen = *iap*) that allowed maize (*Zea mays* L.) pollen tubes to grow through *S. bicolor* pistils (Laurie and Bennett, 1989). This *S. bicolor* accession can successfully override the pollen–pistil incompatibilities between *S. bicolor* and undomesticated *Sorghum* species and lead to the production of hybrid embryos and plants (Price *et al.*, 2006). Hybrids between *S. bicolor* × *S. macrospermum* were obtained from germinated seeds, while the hybrids between *S. bicolor* × *S. angustum* and *S. bicolor* × *S. nitidum* were recovered through embryo rescue and tissue culture. The hybrid nature of these seedlings was again confirmed by the presence of genomes from both parental species that could be readily identified based upon chromosome size and number (Price *et al.*, 2006).

Introgression of the undomesticated *S. macrospermum* genome with cultivated *S. bicolor* has been tracked using FISH (fluorescent in situ hybridization) (Kuhlman *et al.*, 2006). FISH discriminated between the chromosomes of the two parent species, and confirmed through bivalent formation and allosyndetic pairing that recombination was occurring. Progeny of this novel hybrid when backcrossed to *S. bicolor* expressed altered fertility, again confirming that introgression from the undomesticated parent has
occurred (Kuhlman et al., 2006). The analysis of the amount of DNA introgressed from the undomesticated *S. macrospermum* is currently being undertaken using AFLPs (L. C. Kuhlman et al., unpubl. res.). The identification and use of the *iap S. bicolor* accession has enabled the successful introgression of genes from undomesticated *Sorghum* species into cultivated sorghum, and is the first step towards accessing these unique unexploited genes for both biotic and abiotic stresses and agronomic traits. The potential for improving the yield productivity through these traits in commercial sorghum varieties is now a reality.

The role of genomics in improving domesticated *S. bicolor*

*Sorghum bicolor*, a diploid, has a relatively small genome (735 Mbp), which although larger than rice (389 Mbp) is smaller than the other important cereals (wheat 16 900 Mbp, maize 2600 Mbp). The last genome duplication event for the *S. bicolor* genome seems to have occurred much earlier than the divergence of the major cereal crops from a common ancestor (Paterson et al., 2004). Completion of the whole genome sequencing project in 2007 will exponentially increase the sequence data available for *Sorghum* and will provide valuable information on cereal domestication in the African continent, an event that appears to have occurred independently of other continents though by similar reinforced selective pressures (Paterson et al., 2004). In a way, the sorghum genome sequencing will close a biographic triangle into the knowledge of the polymorphism shared before the divergence of these important grasses and ultimately in the understanding of the evolution in cereals crops between Africa, America and Asia (Kresovich et al., 2005). The tenets of colinearity and microlinearity of grass genomes mean that our knowledge of other cereals and their evolutionary ties will also greatly improve. Due to their economic and scientific value, cereal genomes have been studied over the last 15 years using highly advanced technologies. The similarity at the DNA level makes it possible to use comparative genetics to look for particular genes of unknown sequence.
between the genomes with the aim of using that information to develop new varieties or discovering new genes that could have a potential impact on traits that are of global importance (e.g. food quality, drought resistance).

The genetic diversity existing within and between Australian *Sorghum* species was recently evaluated using simple sequence repeats (SSRs) (Dillon et al., 2005). SSRs were sourced from the cultivated *S. bicolor* (Brown et al., 1996; Taramino et al., 1997; Kong et al., 2000) to determine diversity in these closely related taxa. This method has successfully evaluated diversity in the related species of many crop groups (e.g. Peacock et al., 1998; Hernández et al., 2001; Chen et al., 2002; Scott et al., 2003; González-Martínez et al., 2004; Sudupak, 2004). This evaluation of the Australian species has shown significantly higher levels of genetic diversity both between (inter-) and within (intra-) species compared with the intra-specific diversity of *S. bicolor* varieties. The relatively high transfer rate of *S. bicolor*-derived SSRs to the wild species and their high level of diversity suggests that these SSRs are an efficient, highly informative source of molecular markers for the undomesticated *Sorghum* species.

**Screening for novel genetic variation in *S. bicolor***

Mutations, both natural and artificially induced, provide an alternate source of genetic diversity. Mutants have long been a valuable resource in plant breeding (van Harten, 1998) and, in recent times, in plant genomics research (Henikoff and Comai, 2003; Till et al., 2003; Henikoff et al., 2004). However, the method employed (irradiation or chemical) to induce a mutated population will affect its usefulness and application for genomics research. A review of the comprehensive International Atomic Energy Agency’s Mutant Varieties Database (http://www-mvd.iaea.org/MVD/default.htm) shows only 15 induced sorghum mutant accessions amongst more than 2500 registered mutants.

As a result of the random nature of mutation induction, by physical and chemical means, each individual in a population will contain a unique range of gene mutations. This provides a powerful resource for genome analysis employing recent molecular technologies. It is well established that the ultimate goal in DNA research is to ascertain the DNA sequence of a gene. However, the existing technology for genotyping has become a powerful way to avoid the sequencing step or at least for reducing dramatically the sequencing cost and time through multiplexing (Kan et al., 2004).

**THE GENOME OF SUGARCANE AND ITS WILD RELATIVES***

Sugarcane is an important vegetatively propagated crop which is cultivated for its sugar-rich stalks. It contributes an estimated 75% of the world’s sucrose supply with its mature stem capable of accumulating 12–16% of its fresh weight and approx. 50% of its dry weight as sucrose (Bull and Glasziou, 1963). Sugarcane originated in South-east Asia and New Guinea (Lebot, 1999). Modern cultivated sugarcane (Saccharum spp.) is a hybrid complex originating from crosses between *S. officinarum* L. and *S. spontaneum* L., and in some lineages *S. sinense* Roxb., or *S. barberi* Jesw (Edme et al., 2005). Limited numbers of clones, and hence genetic variation, of the two major progenitors have been captured by commercial breeding programmes. Sugarcane, like sorghum, is a relatively recently domesticated species with little of the available genetic diversity having been incorporated or actively analysed for introgression into domesticated varieties. Breeding programmes in the early 1900s focused on hybridization of *S. officinarum* clones but soon progressed to interspecific crosses incorporating *S. spontaneum*. This resulted in improved agronomic traits, such as ratooning and disease resistance, but required a backcrossing programme to *S. officinarum*, called ‘nobilization’, to elevate the sucrose content (Roach, 1989; Edme et al., 2005). Since then the majority of breeding programmes have focused on intercrossing between the hybrids, though in recent decades the larger increases in genetic gain have been made by incorporating more diverse germplasm into the cultivated backgrounds (Edme et al., 2005).
Taxonomy of sugarcane

Sugarcane belongs to the genus *Saccharum*, first established by Linnaeus in *Species Plantarum* in 1753 with two species: *S. officinarum* and *S. spicatum* L. The original classification of Linnaeus’ has since been revised to contain six species: *S. officinarum*, known as the noble cane; *S. spontaneum* L., *S. robustum* E.W. Brandes & Jeswiet ex Grassl, and *S. edule* Hassk., classified as wild species; and *S. sinense* Roxb. and *S. barberi* Jeswiet, classified as ancient hybrids (Buzacott, 1965; Daniels and Roach, 1987; D’Hont and Layssac, 1998). The genus falls in the tribe Andropogoneae in the grass family, Poaceae, that includes other tropical grasses such as *Sorghum* and *Zea* (maize). Closely related to *Saccharum* are another four genera (*Erianthus* section *Ripidum*, *Miscanthus* section *Diandra*, *Narenga* and *Sclerostachya*) that purportedly readily interbreed, forming the ‘Saccharum complex’ (Daniels and Roach, 1987). They have in common a high level of polyploidy and aneuploidy (unbalanced number of chromosomes) that creates a challenge for both the taxonomist and molecular biologist (Daniels and Roach, 1987; Sreenivasan et al., 1987).

The sugarcane genome

The complexity and size of the sugarcane genome is a major limitation in genetic improvement. Whilst continued selective breeding for enhanced sucrose accumulation has been able to achieve over half of the yield increase in the past 50 years, it has been reported as having reached a plateau due to limits to the gene pool exploited in traditional breeding programmes (Mariotti, 2002). Individual research programmes, however, have been shown to still be making significant annual genetic gains by maintaining a diverse gene pool (Edme et al., 2005). The employment of new technologies to assist in the association of traits with genetic markers and genetic maps can aid in achieving further yield increases in breeding programmes.

Most sugarcane cultivars contain more than 100 chromosomes which can be assigned to eight homology groups (Rossi et al., 2003; Aitken et al., 2005). Over the past two decades, studies utilizing various molecular techniques to unravel the complexity of this important crop species have provided a greater understanding of its complex genetic make-up (Bonierbale et al., 1988; Wu et al., 1992; D’Hont, 1994; Sills et al., 1995; Grivet et al., 1996; Ming et al., 2001; Rossi et al., 2003). Significant achievements include milestones that demonstrate the use of single (markers present on one chromosome only) and double dose (marker present on two chromosomes) markers for mapping and QTL analysis (Ming et al., 2001, 2002; Hoarau et al., 2002; Aitken et al., 2004), and large-scale EST sequencing projects by SUCEST-Sugar Cane EST Genome Project (Vettore et al., 2001), SASRI-South African Sugar Research Institute (Carson and Botha, 2000), UGA-University of Georgia, USA (Ma et al., 2004), and CSIRO- Australia’s Commonwealth Scientific and Industrial Research Organization (Casu et al., 2004). Unfortunately, despite these achievements, the pace of progress with sugarcane genomics has lagged behind that achieved with other agricultural crops (Ramsay et al., 2000; Delseny et al., 2001; Mullet et al., 2002).

Analysis of variation in the sugarcane genome

In 1997, an effort was made by the International Consortium for Sugarcane Biotechnology to develop and evaluate simple sequence repeats (SSRs) or microsatellite sequences as a marker system for sugarcane. Markers were developed from an enriched microsatellite library and were shown to have the capacity to distinguish between sugarcane genotypes due to their ability to detect large numbers of alleles (Cordeiro et al., 2000). To date, this marker system has delivered a number of applications that have advanced both sugarcane research and breeding. Published applications include the mapping of alleles generated from 72 SSR primer pairs onto a genetic map constructed on the Australian hybrid cultivar, Q165A (Aitken et al., 2005); validation of the introgression of genes into F1 hybrids of crosses made between *S. spontaneum* and elite commercial clones (Pan et al., 2004); the confirmation of fertile intergeneric F1 hybrids of *S. officinarum* and *E. arundinaceus* as well as backcross (BC1) progeny from the F1 to hybrid sugarcane (Cai et al., 2005); and the use of the markers to register and confirm sugarcane varieties by the United States Department of Agriculture (USDA) (Tew et al., 2003). SSR markers have also been used to draw useful information on the relationships between various members of the ‘Saccharum complex’ (Cordeiro et al., 2003; Cai et al., 2005) as well as relationships between clonal cultivars of hybrid canes (Pan et al., 2003a). A fingerprint database of major Australian sugarcane cultivars has been developed using these markers (Piperidis et al., 2001) as has molecular genotyping of elite clones produced by the USDA (Pan et al., 2003a, b).

High-throughput SNP genotyping

High-throughput genotyping technologies based on single nucleotide polymorphisms (SNPs) or small-scale insertion/deletions (indel) could become efficient alternative tools for traditional markers because of their greater abundance in the genome and ease of measurement. SNPs are being identified and rapidly mapped to provide a rich source of genetic information with the potential for allowing a greater insight into understanding the genetic complexity of many organisms. SNPs are present in high frequency in any genome, amenable to high throughput analysis and have the ability to reveal hidden polymorphisms where other methods fail (Bhattiramakki and Rafalski, 2001). In plants, a number of studies have been able to link SNPs with phenotypic traits of agronomic interest, such as the putative betaine aldehyde dehydrogenase gene responsible for the fragrance trait in rice (Bradbury et al., 2005) and SNPs found in the starch synthase IIa gene associated with starch gelatinization temperature in rice (Waters et al., 2005). These studies highlight the usefulness...
of SNP markers, demonstrating both the abundance of this marker type and the potential causal association between a single nucleotide alteration and organism phenotype. A further major advantage of SNP markers is that they allow easy and unambiguous identification of alleles or haplotypes.

Whilst numerous technical methods have been developed for their detection (Gut, 2001), the majority are applicable mainly to diploid genomes where a simple presence/absence of either one or both of the alternative bases would indicate homozygosity or heterozygosity. Sugarcane, with its complex genome comprising an estimated 8–14 copies of every chromosome (Rossi et al., 2003; Aitken et al., 2004), can have up to 14 different alleles present, with individual alleles in varying numbers. Thus, the frequency of an SNP base at a gene locus will be determined by both the number of chromosomes carrying the gene, and the number of different alleles (or haplotypes) and frequency of each allele possessing each SNP base. Hence, any method used to detect SNPs at a particular locus in sugarcane must be able to determine the frequency of each SNP base in different genotypes, rather than simply detecting the presence and absence of SNPs. Such detection systems are generally more complex and expensive than simpler and more common methods used for detecting less complex genomes (Rossi et al., 1998; Ahmadian et al., 2000; Alderborn et al., 2000; Nurmi et al., 2001; Storm et al., 2003).

**Use of SNPs in sugarcane**

Currently, whilst there are only a limited number of papers describing the use of SNPs to understand the sugarcane genome, they point to this marker system as a valuable means of mapping candidate genes and for identifying the genetic basis of QTLs of agronomically important traits. These studies include a discussion on the ability of SNPs to: delineate a set of 64 ESTs into two groups that are likely to represent two gene family members of 6-phosphogluconate dehydrogenase (Grivet et al., 2001); delineation of 178 ESTs into three paralogous genes to reveal the expression of an Adh2 and two Adh1 genes in sugarcane (Grivet et al., 2003); the development of co-dominant cleaved amplified polymorphic sequence (CAPS) markers (Quint et al., 2002); and to map several candidate genes and ESTs (McIntyre et al., 2005).

In sugarcane, the proportional frequencies of each SNP base will vary depending on the number of alleles of the gene containing the SNP locus. The ability to capture this information accurately across several SNPs within a set of homo(eo)logous alleles can give an indication of the number of allele haplotypes present for a gene and potentially provide the haplotype sequences. This information could have implications for sugarcane breeding. High yield potential may be due to the presence of, or different number of copies of a specific allele(s) present at a gene locus, or possibly a combination of both. Knowledge of the sequence underlying each allele haplotype has the potential to allow allele-specific markers to be designed.

Quantitative methods to detect allele dosage in sugarcane are now possible with such techniques as pyrophosphate sequencing using the Pyrosequencer™ platform (Cordeiro et al., 2006b) and mass-spectrometry using the Sequenom™ platform (Cordeiro et al., 2006b). These methods have allowed the quantitative detection of frequencies of consensus to alternate SNP bases at any particular SNP locus. Utilizing a group of SNP markers developed to the same EST or gene, it becomes possible to infer the likely copy number of the EST or gene. This information then allows for possible haplotypes of a gene present in hybrid cane to be determined through statistical approaches (Cordeiro et al., 2006b).

In theory, the association of SNP variations with either the presence or absence of different phenotypes among individuals or among individuals from different populations appears straightforward. This simplistic view does not account for the majority of base polymorphisms that do not result in any amino acid change. Determining the haplotypes is more important for predicting individual phenotypes than are the underlying SNPs. Determining haplotypes also allows the ability to infer the evolutionary history of a DNA region (Templeton et al., 1988; Tishkoff et al., 1998). However, difficulties are encountered in determining SNP haplotypes when inbred or homozygous individuals are not available (Rafalski, 2002) as is usually the case with sugarcane.

The ability to determine SNP base frequencies provides the means to determine the likely copy number of homo(eo)logous loci in sugarcane. Where chromosome counts have been performed for a genotype, this information can be used to support the inference of the most likely copy number of homo(eo)logous loci. Knowledge of the number of homo(eo)logous loci will assist in the deduction of the allelic composition of the locus in any particular sugarcane genotype. The ability to determine haplotypes also opens possibilities in unraveling the complexities of the sugarcane genome. By defining haplotypes in parents of crosses, it may be possible to deduce their segregation in progeny, or to determine allele dosage and composition in any particular genotype in relation to phenotypic performance. A further level of analysis is required to determine the level of expression of each of the haplotypes in this complex genome.

**Genetic mapping**

Genetic maps are widely used in plant breeding to identify genomic regions controlling traits of interest. Such information assists in understanding the genetic basis of the target trait, as well as providing DNA markers for use in marker-assisted breeding. In sugarcane, only markers that are present as a single copy in one parent and absent in the second [i.e. single-dose (SD) marker] can be incorporated into maps using populations of conventional size (approx. 250 progeny) (Wu et al., 1992). In these populations, SD markers segregate in a 1 : 1 ratio.

The first maps of a cultivar were initiated on the selfed progeny of SP70-1006 (D’Hont, 1994). This map was later transferred and further developed on the cultivar R570.
(Givret et al., 1996) using RFLP probes from maize and sugarcane. By 2001, the R570 map, as it had become commonly known, contained some 600 RFLP markers derived from a number of grass (Poaceae) species (D’Hont and Glazmann, 2001). The markers on this map distribute over 98 cosegregation groups covering a total length of 2008 cM. A parallel mapping effort was also carried out to place 939 single-dose AFLP markers on R570, of which 887 were distributed into 120 cosegregation or linkage groups (Hoarau et al., 2001). A more recent map has been developed on a cross between the Australian commercial variety Q165A (2n = 115) with the S. officinarum clone IJ76-514 (2n = 80) using a combination of AFLP and SSR markers. A total of 967 single dose markers were generated from the two marker systems, and 910 were distributed across 116 linkage groups covering a total map length of 9058.3 cM (Aitken et al., 2005). Markers on these maps have all been generated through anonymous marker systems. However, the use of SNP markers are resulting in ESTs mapped onto the Q165A map.

EcoTILLING for mapping ESTs

Parallel to the development of quantitative SNP frequency scoring methods has been the adaptation of the EcoTILLING method for detecting and mapping sugarcane ESTs. TILLING utilizes the Cell mismatch-cleavage enzyme on heteroduplexed DNA strands with detection of end-labelled cleavage product (McCallum et al., 2000). A variant of this method utilizes natural populations for the discovery of polymorphisms (SNPs, SSRs and indels), and is referred to as EcoTILLING (Comai et al., 2004). Both methods as published rely largely on electrophoretic gels to separate and visualize the products. In sugarcane, this does not allow SNPs that occur on a single allele to be clearly detected. Modifying the protocol and moving the detection system to capillary electrophoresis has allowed the detection of single-dose SNPs in sugarcane to be identified (Cordeiro et al., 2006b) and mapped (McIntyre et al., 2006). Our early experience with family members of the sucrose phosphate synthase gene indicate straightforward detection of the presence of 5–11 SNPs in fragment lengths of genomic DNA between 300 bp and 400 bp in length. Neither prior knowledge of any SNP in the fragment nor the alignment of multiple ESTs are required to identify putative SNPs and their location. Whilst the method is as yet unable to indicate the frequency at which an individual SNP base is present, it has been demonstrated that the detected variation in base composition segregates as expected in progeny of mapping populations. Using the SPS gene family members as an example, the mapping of the gene family members through the EcoTILLING approach supports sequence information that three of the five gene family members may contain more than one gene, with each gene possessing from one to five alleles (McIntyre et al., 2006). This observation will in time allow further unravelling of the complexities of the sugarcane genome.

Sorghum genome information as a resource for sugarcane

Sorghum is the closest cultivated relative of sugarcane. Sugarcane has a large genome that has duplicated at least twice since it diverged from sorghum, around 5 million years ago (Al-Janbi et al., 1997). The extensive similarity in the gene order between these two genomes, where intercroses are still possible (Ming et al., 1998), makes sorghum the best model crop for the Androponea tribe (Price et al., 2005a) with the aim of understanding the extensive gene rearrangements and assisting the development of genetic maps in sugarcane.

Sequencing of Sorghum provides another model genome within the grasses, which particularly when utilized in conjunction with rice, will stimulate evolutionary understanding of the entire Poaceae. Sequencing will stimulate gene and allele discovery and crop improvement in Sorghum as it did in rice. Sugarcane genomics will be supported by the Sorghum sequence data. The sequences of Sorghum genes and to a lesser extent the location of genes in the genome should be useful in sugarcane.

Genetic resources for sorghum and sugarcane improvement have been enhanced by the application of genomic tools to analysis of wild relatives in the Sorghum and Saccharum genera. Mutant populations (including TILLING populations) of Sorghum expand the options for gene discovery and genetic manipulation. Protocols for EcoTILLING (Cordeiro et al., 2006a) and quantitative SNP analysis in the complex sugarcane genome should be valuable tools for gene mapping, gene discovery and association genetics in sugarcane. The availability of a Sorghum genome sequence will further accelerate the potential to apply these techniques in both Sorghum and sugarcane. Gene discovery in this germplasm will also be supported by application of advances in expression profiling tools as has been applied to other crop species in the Poaceae (McIntosh et al., 2007).

ACKNOWLEDGEMENTS

We thank the Grain Foods CRC, CRC for Sugarcane Industry Innovation through Biotechnology and the Australian Research Council for their support of this research. Funding to pay the Open Access publication charges for this article was provided by the OECD.

LITERATURE CITED


Felus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, et al. 2006. Alignment of genetic maps and QTLs between inter- and


