Isolation and Sequencing of a Single Copy of an Introgressed Chromosome From a Complex Genome for Gene and SNP Identification

Cushla J Metcalfe

CSIRO Agriculture and Food, Queensland Biosciences Precinct

Jingchuan Li

CSIRO Agriculture and Food, Queensland Biosciences Precinct

Bangyou Zheng

CSIRO Agriculture and Food, Queensland Biosciences Precinct

Jiri Stiller

CSIRO Agriculture and Food, Queensland Biosciences Precinct

Adam Healey

HudsonAlpha Institute for Biotechnology

Nathalie Piperidis

SRA Limited, Mackay

Karen S Aitken (✉ karen.aitken@csiro.au)

CSIRO Agriculture and Food, Queensland Biosciences Precinct <https://orcid.org/0000-0003-4372-0721>

Research Article

Keywords: Flow cytometry, karyotype, Single nucleotide polymorphism (SNP), single chromosome sequencing

Posted Date: July 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-629403/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on March 11th, 2022. See the published version at <https://doi.org/10.1007/s00122-022-04030-1>.

Abstract

The large complex genomes of many crops constrain the use of new technologies for genome-assisted selection and genetic improvement. One method to simplify a genome is to break it into individual chromosomes by flow cytometry, however, in many crop species most chromosomes cannot be isolated individually. Flow sorting of a single copy of a chromosome has been developed in wheat and here we demonstrate its use to identify markers of interest in an *Erianthus/Saccharum* hybrid. *Erianthus/Saccharum* hybrids are of interest because *Erianthus* is known to be highly resistant to soil borne diseases which cause extensive sugarcane yield losses in Australia. Sugarcane (*Saccharum*) cultivars are autopolyploids with a highly complex genome and over 100 chromosomes. Flow cytometry for sugarcane, as in most crops, does not resolve individual chromosomes to a karyotype peak for sorting. To isolate a single chromosome, we used genomic *in situ* hybridisation (GISH) to identify the flow karyotype region containing the *Erianthus* chromosomes, flow sorted single chromosomes from this region, PCR screened for the *Erianthus* chromosomes and sequenced them. One *Erianthus* chromosome amplified and sequenced well, and from this data we could identify 57 resistant type genes and SNPs in nearly half of these genes. We developed KASP SNP assays and demonstrated that the identified SNP markers segregated as expected in a small introgression population. The pipeline we developed here to flow sort and sequence single chromosomes could be used in any crop with a large complex genome to rapidly discover and develop markers to important loci.

Key Message

This manuscript describes the identification, isolation and sequencing of a single chromosome containing high value resistance genes from a complex polyploid where sequencing the whole genome is too costly.

Introduction

Many of our grass crop species of major economic significance for example, wheat, barley, oat, rye, and sugarcane have large genomes (Kellogg 1998). This presents challenges in the implementation of new technologies for genome-assisted selection and genetic improvement of crops. One approach that has been successfully used in many plants is to break down the complexity of the genome by using flow cytometric sorting to isolate chromosomes or groups of chromosomes according to their size. For most grass crops usually only one or few chromosomes can be identified and separated by monovariate flow sorting, while others cluster together forming composite peaks (Giorgi et al. 2013). We have recently developed the use of flow cytometry in sugarcane (Metcalf et al. 2019). Sugarcane cultivars are derived from crosses between the high sugar producing *S. officinarum* ($2n=10x, =80$) and the wild species *S. spontaneum* ($2n=8x, = 40-128$) followed by backcrossing to *S. officinarum* (Roach 1972). Further crosses have resulted in the modern hybrids with ploidy levels of 10 or more and large genomes (R570 cultivar, 10,000 Mb and $2n = 112$) (D'Hont 2005; D'Hont and Glaszmann 2001; Piperidis et al. 2010). In the flow karyotype of the two *S. officinarum* accessions and three hybrid cultivars, five main peaks, representing groups of chromosomes of similar DNA content or size, were identified (Metcalf et al. 2019). Using microsatellite markers to screen chromosomes from each of the peaks it was determined that at least one homo(eo)log of each of the 10 chromosomes was present in each peak. This indicates that like most other grass crops, a single peak does not contain a single chromosome. The composite peaks could represent ancestral genomes from previous whole genome duplications in this autopolyploid genome (Metcalf et al. 2019). Individual chromosomes cannot therefore be isolated by standard flow sorting in sugarcane from a single peak.

Fluorescence in situ hybridization in suspension (FISHIS) can be used to isolate chromosomes using a fluorescent label however, this has only been successfully demonstrated in wheat and barley with two microsatellite probes and

isolates only some chromosomes with > 90% purity (Giorgi et al. 2013; Sánchez-Martín et al. 2016). Attempts to use FISH in other species has been unsuccessful (unpublished data and pers. comm Jarsolav Doležel). As an alternative, isolation of a single copy of a chromosome by flow sorting has been developed in wheat (Cápal et al. 2015). Flow cytometry of a single chromosome results in very low quantities of DNA which need to be amplified before sequencing. This amplification process results in uneven coverage and some missing genes because of the random nature of the amplification (Cápal et al. 2015). However, it has been successfully used in wheat to identify chromosomes carrying a transgene (Cápal et al. 2016) and in pea to identify chromosome translocations (Kreplak et al. 2019). Here we demonstrate how flow sorting of one single copy of a chromosome can also be used for gene identification and SNP discovery.

We chose to isolate an *Erianthus* chromosome from an *Erianthus-Saccharum* hybrid because of the potential value to sugarcane breeders. The genetic base of sugarcane cultivars is limited due to the relatively few *S. spontaneum* and *S. officinarum* clones being used in their initial development (Aitken and McNeil 2010; Arceneaux 1965). To incorporate new genes or alleles for resistance into sugarcane cultivars, sugarcane breeders have established introgression breeding programs with sugarcane-related genera of the 'Saccharum complex' (Aitken et al. 2018; Santchurn et al. 2014; Waclawovsky et al. 2010). Of these, *Erianthus arundinaceus* is of most interest because some *E. arundinaceus* clones are highly resistant to root knot and root lesion nematodes and to Pachymetra root rot (Magarey and Croft 1996). Yield loss from nematodes are estimated at 12 to 15%, costing the Australian sugar industry about AU\$82 million per year (Blair and Stirling 2007). Similar yield losses, 10 to 15% are estimated for Pachymetra root rot, with up to nearly 40% losses in susceptible varieties (Magarey and Bull 2003). *E. arundinaceus* is a hexaploid and contains 60 chromosomes making whole genome sequencing and assembly very costly but introgression can be used to reduce the number of chromosomes. The first fertile genuine hybrids between *S. officinarum* and *E. arundinaceus* was reported at the Hainan Sugarcane Breeding Station of the Guangzhou Sugarcane Industry Research Institute in China in 2002 (Cai et al. 2005; Deng et al. 2002). The initial hybrid between *S. officinarum* and *E. arundinaceus* has 40 *Saccharum* chromosomes and 30 *Erianthus* derived chromosomes. This has been backcrossed to *Saccharum* to recover the high-sucrose phenotype, resulting in reduced numbers of *Erianthus* chromosomes, BC4 plants have 0-3 or 4 *Erianthus* chromosomes (Piperidis et al. 2010). Recombination does occur between the *Erianthus* and *Saccharum* chromosomes resulting in stable inheritance (unpublished data), so with further crossing recombination will reduce the linkage block and allow the incorporation of resistance genes into sugarcane.

The process of introgression in *Saccharum* using conventional breeding procedures is relatively long and risky. Following initial crosses with wild clones generally two or more backcrosses to elite commercial type parents are made and then lengthy (up to 10 years) field evaluation and selection programs are conducted before the commercial type progeny are developed (Hotta et al. 2010; Zhang et al. 2014). Without clear identification of specific traits or genes (markers) derived from the exotic germplasm, there is no guide for the selection for these traits or genes and valuable components from the exotic source may be easily lost after several generations of backcrossing. There is also the probability that undesirable traits from the exotic genome may be inherited making some progeny inferior agronomically. For these reasons markers specific to chromosomes carrying resistance genes of interest are desirable and can be used for selection of progeny that contain the resistant gene from the exotic germplasm.

To demonstrate that sequencing of a single flow-sorted chromosome can be used to identify markers to resistance genes, we isolated an *Erianthus* chromosome from the BC4 *Erianthus-Saccharum* hybrid, QC12-20006. The BC3 parent of QC12-20006 is known to have some resistance to Pachymetra root rot and root knot nematodes (Shamsul Bhuiyan pers.com.). We first identified the region of the flow karyotype representing the *Erianthus* chromosomes by genomic *in situ* hybridization (GISH) to chromosomes flow sorted onto slides, single chromosomes from this region were then flow sorted into plates, amplified using the non-PCR based DNA amplification method, multiple displacement

amplification (MDA) (Dean et al. 2002), and screened by PCR with an *Erianthus* specific sequence to isolate *Erianthus* chromosomes. Single *Erianthus* chromosomes were sequenced and used to identify genes and SNP markers for use in introgression breeding.

Materials And Methods

Plant Material

The *Saccharum-Erianthus* BC4 hybrid QC12-20006 was developed in a collaborative project conducted by the Sugar Research and Development Corporation (SRDC), the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sugar Research Australia (SRA), Yunnan Sugarcane Research Institute and Guangzhou Sugar Industry Research Institute. The pedigree for this clone is shown in Figure 1. F1 and BC1 plants were generated at the Hainan Sugarcane Breeding Station of Guangzhou Sugarcane Industry Research Institute (Wu et al. 2014). YCE96CYC96-40 and YCE01-102 were developed in Hainan Sugarcane Breeding Station of Guangzhou Sugarcane Industry Research Institute, China (Cai et al. 2005; Wu et al. 2014), QC12-20006 was developed by Sugar Research Australia (SRA) (Phillip Jackson, pers. comm.) The BC3 parent (KQ08-6013) has resistance to *Pachymetra* root rot and Root knot nematode.

QC12-20006 for the GISH analysis were grown in a glasshouse at Sugar Research Australia in Mackay, Australia (21°09'52"S; 149°07'11"E). Collection of root tips, metaphase preparation and GISH were performed as previously described (D'Hont et al. 1996). Metaphases were probed with whole *Erianthus* DNA labelled with Rhodamine 5-d-UTP and *S. officinarum* DNA labelled with Fluorescein 12-d-UTP according to (Piperidis et. 2010; Piperidis et al. 2019). Slides were counter-stained with 4'-6-Diamidino-2-phenylindole (DAPI). Images were captured with a CCD camera attached to a BX53 Olympus microscope and imaged with the Olympus Cellsens software.

Flow Cytometry and identification of composite peaks with Erianthus chromosomes

QC12-20006 for flow cytometry was grown in a glasshouse at CSIRO (St Lucia, Queensland) with a natural photoperiod, 12 hours 30 °C, 12 hours 24 °C and humidity > 55% / 70% (day/night relative humidity). Chromosome suspensions were prepared and sorted according to (Metcalf et al. 2019). To identify flow karyotype peaks containing *Erianthus* chromosomes, 2,000 to 3,000 chromosomes from each composite peak were flow sorted onto slides and probed with whole *Erianthus* genomic DNA by GISH. DNA was extracted from the *Erianthus arundinaceus* clone IJ76-346 using a standard CTAB method (CIMMYT Laboratory Protocols 2005), replacing the octanol with isoamyl alcohol. 1 µg of DNA was used in a 25 µL labelling reaction with the Digoxigenin (DIG)-11-dUTP-Nick Translation Mix (Roche). Hybridisation and detection was performed according to Vrána et al. 2016) except that in the hybridization mix salmon sperm DNA was used instead of calf thymus DNA and the probe was detected using Anti-DIG Rhodamine (Roche) (1 µg/mL).

Slides were examined with a Carl Zeiss AxioImager M2 Fluorescence microscope equipped with filter set appropriate for DAPI. Images were captured and digitized using an AxioCam 506 Mono camera and Zen 2 Pro (blue edition) version 2.0.0.0 imaging software (<https://www.zeiss.com/microscopy/int/products/microscope-software/zen.html>).

Chromosome amplification, evaluation and sequencing

Proteinase K (2 µg/mL) in 1 x TE buffer was filter sterilised through a Millex 0.22 µm syringe-driven filter unit (Merck Millipore), 7.5 µL aliquoted into a 96 well plate and then UV-irradiated for 30 seconds in a BioRad XR+ Gel Documentation System. Chromosomes were flow sorted into the 96 well plate at the rate of 100 events per second.

One hundred chromosomes were flow sorted into the first row as a positive control, single chromosomes into rows B to G, and no chromosomes into the last row as an amplification and flow cytometry negative control. Chromosomes were purified and amplified according to Vrána et al. (2016) and Cápál et al. (2015), with the REPLI-g Single Cell Kit (Qiagen) and the protocol 'Whole genome amplification from genomic DNA using the REPLI-g Single Cell Kit with increased sample volumes', with half the volumes described. We tested 8 hours incubation time vs 16 hours for increased amplification.

The multiple displacement amplification (MDA) products were evaluated by agarose gel electrophoresis, diluted 1:10 and quantified with the Qubit dsDNA Broad Range Assay kit (ThermoFisher). Diluted amplified DNA was PCR screened for sugarcane and *Erianthus* genomic DNA, and for bacterial contamination. For all PCR reactions 5 ng of purified DNA was used as a template in 15 µl PCR reaction with 10 µM of each primer and the MyTaq HS DNA Polymerase (Bioline) kit with dNTPs, MgCl₂ and enhancers included in the buffer provided. The following cycling conditions were used, initial denaturation 94°C 1 min, 35 cycles of (94°C 30 sec, 60°C 1 min, 72°C 30 sec), followed by a final extension of 72°C for 5 min.

Three sets of primers were used to identify an *Erianthus* chromosome (Table S1). The first set of primers are universal bacterial 16S rDNA primers (Wang et al. 2003). The second set of primers were designed against a conserved domain of a highly abundant transposable element in the sugarcane clone R570 (Domingues et al. 2012). These primers were used to verify that *Saccharum* or *Erianthus* DNA had been amplified and that there was low or no bacterial contamination. The third set of primers are specific to *Erianthus* (Besse and McIntyre 1998) and were used to identify the *Erianthus* or *Saccharum*/*Erianthus* hybrid chromosome.

Two MDA products identified as *Erianthus* DNA were prepared with the Nexera Illumina library kit and sequenced on an Illumina (<https://www.illumina.com/>) HiSeq 2500 platform at the Australian Genome Research Facility (AGRF)

Mapping to R570 monoploid genome and genes

Illumina sequencing reads from the MDA products were quality trimmed and the Nextera adapters removed using Trimmomatic v0.36 (Bolger et al. 2014). The following parameters were used: ILLUMINACLIP:NextraPE-PE.fa:2:30:10:8:TRUE LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50. After trimmomatic, the reads retained poor quality bases in the first 18 bases. SED (<https://www.gnu.org/software/sed/>) was used to hard trim the 18 bases. The quality of the trimmed reads was assessed with FastQC, a quality control tool for high throughput sequence data (Andrews 2015).

To check for bacterial contamination 5,000 reads were randomly selected using a custom perl script and compared with the NCBI nucleotide database using blastn version 2.7.1 (Altschul et al. 1990; Sayers et al. 2011). A custom R script was used to summarise the results.

The R570 STP chromosomes and BAC sequences were retrieved from the Sugarcane Genome hub (<https://sugarcane-genome.cirad.fr/content/download>) (Garsmeur et al. 2018). Simulated R570 sequencing reads were generated from BAC sequences using wgsim (<https://github.com/lh3/wgsim>). Gene coordinates were extracted from the GeneStructureAndFunction.gff3 file and used to extract gene only regions from the monoploid sugarcane genome using bedtools version 2.28.0 getfasta (Quinlan and Hall 2010).

Reads from the two MDA products and the simulated R570 reads were mapped to the R570 STP (Garsmeur et al. 2018), while the two MDA products were also mapped to the gene regions. Mapping was done using the Burrows-Wheeler mem aligner using default settings (Li 2013) and quality filtered (q=30) to keep only uniquely mapped reads

and filtered so that both reads mapped. Mapping statistics were obtained using samtools version 1.9.0 flagstat and view from the SAMtools package (Li et al. 2009). Bedtools version 2.28.0 genomecov (Quinlan and Hall 2010) with the -bga option was used to generate a coverage reports. Custom R scripts were used to generate graphs and information for the tables.

SNP identification and verification

Samtools version 1.9.0 mpileup and bcftools version 1.9.0 from the SAMtools package (Li et al. 2009) were used to identify SNPs from mapping to the sugarcane monoploid genome. A R570 resistance genes bed file was created by extracting annotations that included the terms LRR, NBS, RAG, NBC-ARC, resistance, TIR and WAK from the R570 monoploid reference GeneStructureAndFunction.gff3 file (<https://sugarcane-genome.cirad.fr/content/download>) (Garsmeur et al. 2018). Bedtools version 2.28.0 slop (Quinlan and Hall 2010) was used to expand the gene coordinates 100 bp for primer design. Bedtools version 2.28.0 intersect was then used to extract the resistance genes from the vcf file (-header -wa options) and to annotate the resulting file with resistance gene descriptions (-wb option). A custom R script was used to filter out SNPs within 20 bp.

KASP primers were designed to the selected SNPs using Primer3Plus (Untergasser et al. 2007). A pre-amplification step was used for the SNP genotyping to improve the allelic discrimination of the genotypes. The pre-amplification was carried out with 10x buffer, 25 mM MgCl₂, 10 mM dNTPs, Taq 1u/ul, Forward and Reverse primers 10uM in a final volume of 6 ul. PCR cycle was as follows, 94°C 3 min, (94 °C 20s, 55°C 30s, 72 °C 30s) for 30 cycles then 72 °C for 5 min. The product was diluted 1 in 100 then run on the ABI Viia7 instrument using the KASP (Biosearch Technologies) genotyping program using 2.5 ul of pre-amplification in a 5 ul reaction containing 2.5 ul KASP master mix, 0.07 ul KASP assay mix and 0.1 ul of primer mix. Amplification conditions were 95°C for 15 min, 95°C 20s, 65°C-57°C 60s Touchdown over 10 cycles then, (95°C 10s and 57°C 60s) for 20 cycles. Once the run was complete the data was analysed and the allelic discrimination plot used to call the genotypes.

To determine the segregation ratio the SNP markers were screened across 29 clones from the *Saccharum-Erianthus* BC3 population, from the QN80-3425 x QBYC06-30415 cross. The SNP markers were tested against the expected ratio for a single dose marker (1:1) using a χ^2 test. The clones were screened for pachymetra root rot response using the method in [19]. To determine association with the Pachymetra root rot rating the data was analysed using a t-test: Two-Sample assuming unequal variances to test for association.

Results

Chromosomal composition of the BC4 hybrid, QC12-20006, by GISH

Genomic *in situ* hybridisation analysis of QC12-20006 with whole *Erianthus* DNA identified three *Erianthus* chromosomes in the sugarcane background. The red chromosomes are inherited from *Erianthus* and the green chromosomes are inherited from *Saccharum* (Figure 2). One of these chromosomes, indicated with a yellow arrow which has a majority of red signal also contains a small introgression of green indicating a recombinant chromosome (Figure 2).

Flow Cytometry and identification of peaks with Erianthus chromosomes

In all sugarcane genotypes examined to date, five composite peaks (I, II, III, IV, V) were identified in the flow karyotype except for the genotype R570 which only presents the first four (Metcalf et al. 2019). Each genotype also had a characteristic flow karyotype profile. Five composite peaks were also identified in the QC12-20006 flow karyotype,

even though the genotype has three *Erianthus* chromosomes (Figure 3). The QC12-20006 flow karyotype is similar to other sugarcane cultivars but distinct from *S. officinarum* (Metcalf et al. 2019). The highest composite peak in the *S. officinarum* flow karyotype is II, whereas composite peaks I and II in QC12-20006 are the highest and equally abundant, with composite peaks III and IV also of equal height (Figure 3a).

To ensure enrichment of *Erianthus* chromosomes when sequencing, GISH with whole *Erianthus* genomic DNA was used to narrow down the screening of single flow sorted chromosomes to one or two composite peaks (Figure 3b). No signals were observed on chromosomes flow sorted onto slides from composite peaks I, II or V. Signals from the distinctive *Saccharum-Erianthus* recombinant chromosome was observed in QC12-20006 composite peaks III and IV (Table 1).

Amplification, screening and sequencing of chromosomes

Chromosomes from composite peaks III and IV were flow sorted into six 96 well plates and amplified. Using the standard amplification reaction time suggested with the REPLI-g Single Cell Kit (Qiagen), the amplification success rate was very poor. Only one out of the five samples processed resulted in a product at the expected yields and was positive for the *Saccharum/Erianthus* transposable element (TE). After increasing the amplification incubation time to 16 hrs, as described in the supplementary protocol for maximum yield, the success rate improved to 4 successfully amplified products out of 5 samples.

Thirty single chromosomes from the two QC12-20006 flow karyotype composite peaks were purified, amplified and PCR screened. The DNA from the single chromosome amplifications were evaluated on a 1.5% agarose gel. An example of the amplifications is shown in Figure S1. In Figure S1, all amplifications show high molecular weight DNA, except for Lane 2 which represents an amplification failure. The negative control produces a product of the same size as the samples because DNA is generated during the REPLI-g Single Cell reaction by random extension of primer dimers. Total yields ranged from 6 to 31 ug, with an average of 19 ug.

Purified and amplified samples were PCR screened with the three sets of primers (Table S1). Wells with no chromosomes flow sorted into them were PCR negative, indicating no contamination from the flow sorter or amplification process. Of the thirty single chromosome amplification products screened, twenty-four, 80%, were PCR positive for the *Saccharum/Erianthus* repeat, 17 of these 24 chromosomes, 71%, were negative for the universal bacterial 16S rDNA sequence. After screening with the *Erianthus* specific repeat, there were two MDA products that were positive for the *Saccharum/Erianthus* TE and the *Erianthus* specific repeat, and negative for the universal bacterial 16S rDNA. As we would expect 3 out of 105 chromosomes to be *Erianthus* (2.8%) and we identified 2 *Erianthus* chromosomes out of 17 tested (11.7%) by flow sorting from a selected peak we obtained a 4-fold enrichment for *Erianthus* chromosomes. Amplification yields for the MDA products that were *Saccharum/Erianthus* PCR positive and bacterial 16S rDNA negative were between 15 and 31 ug in a total volume of 25 uL.

Read trimming, alignment and mapping

Around 9 Gb of sequence was obtained for both single chromosomes sequenced (Table S2), resulting in approximately 110 x coverage, based on the estimated chromosome size of 80 Mb for sugarcane (cultivar R570: 10 Gb/112 chromosomes (Piperidis et al. 2010)). After trimming and quality checking the coverage dropped to approximately 62x, with a mean read length of 107 bp and approximately 80% of the reads retained (Table S2). The top hit for about ½ of a random sample of 5,000 reads with a blastn against the NCBI nucleotide database was *Saccharum*, following by *Sorghum* (Table 2). The top hit was *Saccharum* rather than *Erianthus* because there is very little sequence information available for *Erianthus*. There was no indication of bacterial or human contamination.

Approximately 70% of the reads were properly paired and mapped to the R570 single tilling path (STP) (Garsmeur et al. 2018), which dropped to 43% for the first MDA product and 36% for the second MDA product after quality filtering (Table S2). For mapping to gene regions only, 24% and 18% of reads from MDA products one and two were properly paired and mapped, respectively. This dropped to 17% (7,602,141 reads) and 12% (10,056,900) after BAM mapping quality filtering (Table S2).

Filtered MDA product reads were separately mapped to the R570 STP chromosomes and genes (Garsmeur et al. 2018). Mapping to genes was used to identify which chromosome had been amplified and to examine gene coverage. Mapping to STP chromosomes was used to identify resistance genes and SNPs and to examine overall coverage. The highest proportion of reads from the first MDA product mapped to genes from chromosome 7, consistent with chromosome mapping results (Table 3, S3 and S4). Results for the second MDA product were not concordant, where reads mapped with higher frequency to gene sequences on chromosome 4, but overall mapped best across the entirety of chromosome 5. However, the highest mean depth and second highest mean breadth of read mapping was to chromosome 5 genes (Table 3, S3 and S4). Results for the second MDA product are therefore shown for chromosome 5.

Coverage of STP chromosome seven (MDA one) and chromosome five (MDA two) was uneven, showing very high coverage in some regions and very low coverage in other regions (Figure 4a). Figure 4 (b) shows that for the first MDA product 80% of chromosome 7 had no reads mapped, 16% had less than 200 reads, while 2.4% had over 200 reads, while for the second MDA product 98% of chromosome 5 had no reads mapped, 1.39% had less than 200 reads, while only 0.27% had over 200 reads.

The average gene read depth and percentage mapping breadth was much higher for the first MDA product (74.05 and 44.15% respectively), than for the second MDA product (30.68 and 12.59% respectively) (Table 3 and S4). This is reflected in Figure 5 (a), a higher proportion of reads from MDA one map with higher coverage to chromosome 7 than to any chromosome. Reads from the first MDA product covered 5% of the R570 STP chromosome 7 genes 80-100%, 16% of genes 40-80%, 42% of genes >0-40%, while 38% of genes were missing (Table S5). One hundred resistance genes were identified on chromosome 7 and 68 on chromosome 5. For the first MDA product, over half had some read coverage, 16% had over 40% coverage. Only 16% of the chromosome 5 resistance genes had some coverage with reads from the second MDA product. (Figure 5 and Table S5).

SNP identification and verification

As the test case was to see if we could identify SNP markers inherited from the *Erianthus* chromosomes, genes that were annotated as resistance genes were targeted. SNPs were identified in 24 classic resistance genes in MDA 1 on chromosome 7, i.e. nearly half of the resistance genes with some read coverage and in 7 of the resistance genes in MDA 2 on chromosome 5, i.e. over half of the resistance genes with some read coverage. A total of 136 SNPs within 31 annotated resistant genes were identified as potential candidate markers (Table S6). Genes were identified that had SNPs at a coverage of 3 to 19 reads or above. Four KASP primers were designed against four genes identified from the first MDA product and two KASP primers against one gene identified from the second MDA product (Table 4).

The KASP primers were optimised for annealing temperature and run across the parents of the BC3 *Erianthus/Saccharum* population. Five of the SNP markers did not segregate as single dose markers. One SNP, in gene ID Sh07_g011420, segregated as single dose in a small BC3 population (Figure 3f). This SNP segregated as expected for a single dose marker (1:1) and verified the high quality of the sequence information. A preliminary association test between this SNP data from the BC3 clones and their pachymetra root rot rating revealed a significant association ($p \leq 0.01$).

Discussion

We aimed to demonstrate the use of single chromosome flow sorting as part of a genome-assisted selection and genetic improvement strategy in any crop with a large complex genome. Routine isolation of single chromosomes, either by flow cytometry or other methods, results in quantities of DNA too low to sequence without amplification (Vrána et al. 2016; Arumuganathan et al. 1994). Whole genome amplification (WGA) was developed for analysis of small pools of cells or single cells, for example, for forensic analysis, prenatal or preimplantation diagnosis, oncogenetics and paleo-archeology (Dean et al. 2002; Spits et al. 2006). The first methods were PCR-based which were hampered by non-specific amplification artefacts (Cheung and Nelson 1996), incomplete coverage of loci (Paunio et al. 1996) and the small size of the DNA products (Zhang et al. 1992; Telenius et al. 1992). Multiple displacement amplification (MDA) is a newer WGA method with limited sequence representation bias (Dean et al. 2002; Lizardi et al. 1998). Amplification and sequencing of single or small numbers of chromosomes (Hotta et al. 2010) has been developed in humans for haplotype phasing (Fan et al. 2011; Ma et al. 2010; Yang et al. 2011; Kirkness et al. 2013), in the Japanese eel to identify genetic linkage group sequences (Matsubara et al. 2018) and in plants (Cápal et al. 2015; Cápal et al. 2016; Kreplak et al. 2019) for wheat and pea.

Sugarcane (*Saccharum*) cultivars present particular challenges, they are polyploid aneuploid hybrids with eight or more homeologous chromosome copies and in total between 100 to 120 chromosomes (Piperidis et al. 2010; D'Hont et al. 1998). We have recently developed the use of flow cytometry with *Saccharum* (Metcalf et al. 2019), which provided the ability to carry out chromosome flow sorting to isolate a single targeted chromosome. The method we used was based on that developed by (Cápal et al. 2015), combining flow sorting of a single chromosome into a 96 well plate, amplification and sequencing. In Cápal et al. (2015) the amplified wheat chromosome 3B corresponds to only 0.002 ng template DNA. The total sugarcane genome is about 10 Gb (10.22 pg) (D'Hont and Glaszmann 2001; Garsmeur et al. 2018) and has small chromosomes all of similar size (Meng et al. 2019). A single sugarcane chromosome is therefore about 0.0001 ng, an order of magnitude smaller than a wheat chromosome.

To determine if the SNPs identified from the sorted chromosome were high quality and segregated as single dose, we tested several SNP markers identified in two genes that have been annotated as resistant genes across a small *Erianthus* introgression population. Although this population was small as we had *Pachymetra* root rot resistance ratings we tested for association to resistance to *Pachymetra* root rot. This pathogen was chosen as loci (or QTL) located on *Erianthus* chromosomes in QC12-20006 have been shown to contain resistance to this pathogen (Magarey and Croft 1996). GISH was used successfully to narrow down the two karyotype composite peaks that putatively contained *Erianthus* chromosomes by flow sorting chromosomes from individual peaks onto slides for detection. This allowed easier isolation, enrichment and sequencing. Two individual *Erianthus* chromosomes were isolated from flow-sorted chromosomes, verified by PCR and an *Erianthus* specific repeat. The two chromosomes were sequenced to high depth with Illumina reads and used to identify SNPs against the R570 monoploid sugarcane genome and genotyped using KASP in the BC3 population to identify regions associated with resistance to *Pachymetra* infection.

Two out of the three *Erianthus* chromosomes from an introgression line were isolated, amplified and sequenced. The second amplification product (MDA) had very low sequence coverage and depth which could be due to a lower amount of amplified product. The first MDA coverage and depth was comparable to, but lower, than reported by Cápal et al. (2015) for a single wheat chromosome and Kirkness et al. (2013) for a single sperm. Uneven coverage (Figure 4, 5) was also observed by Cápal et al. (2015) and Kirkness et al. (2013). The total region examined with no coverage (80%) is higher slightly than that reported by Cápal et al. (2015) (70%) and Kirkness et al. (2013) (57-72%) (Figure 4). The percent of genes 80-100% covered, 5%, is slightly lower than reported for the three wheat chromosomes (7-15%) (Figure 5) (Cápal et al. 2015). The depth is similar (10% of chromosome 7 covered by at least 10 reads) to that

reported by Cápál et al. (2015). There are probably several contributing factors to low coverage observed for the first MDA. The input (0.002 ng) is many fold lower than that recommended by the MDA amplification kit is designed for (10 ng) and 10 fold lower than a wheat chromosome (Cápál et al. 2015). The sequencing reads were mapped onto the evolutionary closest available genome, the sugarcane monoploid genome because an *Erianthus* genome is not available, whereas Cápál et al. (2015) were able to map directly back onto published sequence of the same chromosome that was isolated. The sugarcane monoploid genome sequence is generated from selecting gene rich *Saccharum* BAC clones aligned to the sorghum genome sequence so centromeric and other repeat regions are under-represented.

Despite the low coverage of the amplified chromosomes, we were able to identify SNP markers in nearly a quarter of the resistance genes identified on chromosome 7 (24 out of 100) with the first MDA product. Segregation of these SNP markers were verified using a KASP genotype assay on 29 clones from the BC3 population which includes KQ08-6013, the intergeneric parent carrying the *Erianthus* chromosomes. In most cases the KASP primers identified SNPs that were not segregating in the BC3 population. This was probably due to them being either multi-dose in the BC3 population, which contain from 3 to 7 *Erianthus* chromosomes, or the SNPs were not in fact *Erianthus* specific SNPs, but were also present in *Saccharum*, just not in the R570 monoploid sequence, so did not segregate. The SNP located in an IAA-alanine resistance gene was segregating in the BC3 population and was tested for association with the Pachymetra root rot ratings for these clones. This SNP was significantly associated ($p \leq 0.01$) with resistance to Pachymetra root rot and a larger population should be screened to verify this result. Interestingly IAA-alanine resistance genes are members of the ZIP transporter family which are probably involved in zinc uptake (Grotz et al. 1998). IAA is the major form of auxin in plants, and in rice, auxin signaling is closely associated with zinc efficiency (Begum et al. 2016). The homolog of the IAA-alanine resistance gene in Arabidopsis, IAR1, has been also been shown to participate in auxin metabolism or response (Lasswell et al. 2000). Auxin has been known to be a regulator of plant growth and development ever since its discovery, however, recent studies on plant–pathogen interactions have also identified auxin as a key character in pathogenesis and plant defence (Fu and Wang 2011).

We have successfully demonstrated that there is potential to couple GISH, single chromosome sorting, PCR screening and sequencing to rapidly identify genes/markers to a particular introgressed chromosome. This method allowed us to identify single dose segregating SNP markers that when tested showed potential association with an important trait in sugarcane. The genes/markers identified in this research will be tested on a larger population to verify their association with resistance. There are a number of strategies that could be used to improve the low and uneven coverage, and low depth and improve the chances of identifying resistance genes/markers. The increased yield of the kit we used would allow long read sequencing combined with Illumina sequencing. Amplification bias could be improved by amplification in a nanoliter reactor or using a microfluidic digital droplet MDA technique (Marcy et al. 2007; Rhee et al. 2016), which could be used in combination with single chromosome flow sorting. In Cápál et al. (2015) and Kreplak et al. (2019) their strategy was to merge reads from three amplification products and evaluate the amplification products before sequencing by PCR using a set of single loci specific to the 3B chromosome, which we were unable to do because of the lack of sequencing information for *Erianthus*, but this is a strategy which could be used to isolate sugarcane chromosomes.

Abbreviations

WGA	Whole genome amplification
FISHIS	Fluorescence <i>in situ</i> hybridization in suspension
GISH	Genomic <i>in situ</i> hybridization
MDA	Multiple displacement amplification
SNP	Single nucleotide polymorphism
STP	Single tiling path

Declarations

Funding: Work was funded by CSIRO, Sugar Research Australia

Conflicts of interest/Competing interest: no conflicts of interest or competing interests

Availability of data and material: All data is available on request to corresponding author

Code Availability: Not applicable

Authors' contributions: Conceptualization, KSA and NP; Funding acquisition, KSA and NP; Investigation, CJM and JL; Formal analysis, CJM, AH, BZ and JS; Original draft preparation, CJM; Review and editing, KSA, CJM and NP; Supervision, KSA.

Ethics approval: Not applicable

Consent for publication: All authors have read and agreed to the published version of the manuscript.

Acknowledgments: We would like to acknowledge the support of Sugar Research Australia who provided the genotype for flow sorting. We would also like to thank Dalia Khalil from the TRI Flow Cytometry Facility at the Translational Research Institute who was instrumental in getting the flow sorting parameters correct for sugarcane. This work was conducted in collaboration by the Joint Genome Institute and is supported by the Office of Science in the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

References

1. Aitken, K.; Li, J.; Piperidis, G.; Qing, C.; Yuanhong, F.; Jackson, P. Worldwide Genetic Diversity of the Wild Species and Level of Diversity Captured within Sugarcane Breeding Programs. *Crop Sci.* **2018**, *58*, 218, doi:10.2135/cropsci2017.06.0339.
2. Aitken, K.S.; McNeil, M. Diversity Analysis. In *Genetics Genomics and Breeding of Sugarcane*, Henry, R., Kole, C., Eds.; Science Publishers: Enfield, USA, 2010; pp. 19–42 ISBN 978-57808-684-9.
3. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–10, doi:10.1016/S0022-2836(05)80360-2.
4. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online] Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2015).

5. Arceneaux, G. Cultivated Sugarcane of the World and their Botanical Derivation. *Proc. Int. Soc. Sugarcane Technol.***1965**, *12*, 844–854.
6. Arumuganathan, K.; Martin, G.B.; Telenius, H.; Tanksley, S.D.; Earle, E.D. Chromosome 2-specific DNA clones from flow-sorted chromosomes of tomato. *Mol. Gen. Genet.***1994**, *242*, 551–558, doi:10.1007/BF00285278.
7. Begum, M.C.; Islam, M.; Sarkar, M.R.; Azad, M.A.S.; Huda, A.K.M.N.; Kabir, A.H. Auxin signaling is closely associated with Zn-efficiency in rice (*Oryza sativa* L.). *J. Plant Interact.***2016**, *11*, 124–129, doi:10.1080/17429145.2016.1220026.
8. Besse, P.; McIntyre, C.L. Isolation and characterisation of repeated DNA sequences from *Erianthus* spp. (Saccharinae: Andropogoneae). *Genome***1998**, *41*, 408–416, doi:10.1139/g98-034.
9. Blair, B.L.; Stirling, G.R. The role of plant-parasitic nematodes in reducing yield of sugarcane in fine-textured soils in Queensland, Australia. *Aust. J. Exp. Agric.***2007**, *47*, 620–634, doi:10.1071/EA05287.
10. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics***2014**, *30*, 2114–20, doi:10.1093/bioinformatics/btu170
11. Cai, Q.; Aitken, K.; Deng, H.H.; Chen, X.W.; Fu, C.; Jackson, P.A.; McIntyre, C.L. Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers. *Plant Breed.***2005**, *124*, 322–328, doi:10.1111/j.1439-0523.2005.01099.x.
12. Cápal, P.; Blavet, N.; Vrána, J.; Kubaláková, M.; Doležel, J. Multiple displacement amplification of the DNA from single flow-sorted plant chromosome. *Plant J.***2015**, *84*, 838–844, doi:10.1111/tpj.13035.
13. Cápal, P.; Endo, T.R.; Vrána, J.; Kubaláková, M.; Karafiátová, M.; Komínková, E.; Mora-Ramírez, I.; Weschke, W.; Doležel, J. The utility of flow sorting to identify chromosomes carrying a single copy transgene in wheat. *Plant Methods***2016**, *12*, 24, doi:10.1186/s13007-016-0124-8.
14. Cheung, V.G.; Nelson, S.F. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc. Natl. Acad. Sci. U. S. A.***1996**, *93*, 14676–14679, doi:10.1073/pnas.93.25.14676.
15. CIMMYT *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory*, Third Edit.; CIMMYT: Mexico, D.F., 2005; ISBN 9686923306.
16. Dean, F.B.; Hosono, S.; Fang, L.; Wu, X.; Faruqi, A.F.; Bray-Ward, P.; Sun, Z.; Zong, Q.; Du, Y.; Du, J.; et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U. S. A.***2002**, *99*, 5261–6, doi:10.1073/pnas.082089499.
17. Deng, H.H.; Liao, Z.Z.; Li, Q.W.; Lao, F.Y.; Fu, C.; Chen, X.W.; Zhang, C.M.; Liu, S.M.; Yang, Y.H. Breeding and isozyme marker assisted selection of F2 hybrids from *Saccharum* spp. x *Erianthus arundinaceus*. *Sugarcane and Canesugar***2002**, *1*, 1–5.
18. D'Hont, A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet. Genome Res.***2005**, *109*, 27–33, doi:10.1159/000082378.
19. D'Hont, A.; Glaszmann, J.C. Sugarcane genome analysis with molecular markers: a first decade of research. In *International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress*; 2001; pp. 556–559.
20. D'Hont, A.; Grivet, L.; Feldmann, P.; Rao, S.; Berding, N.; Glaszmann, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet.***1996**, *250*, 405–413.
21. D'Hont, A.; Ison, D.; Alix, K.; Roux, C.; Glaszmann, J.C. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome***1998**, *41*, 221–225, doi:10.1139/gen-41-2-221.

22. Domingues, D.S.; Cruz, G.M.Q.; Metcalfe, C.J.; Nogueira, F.T.S.; Vicentini, R.; de S Alves, C.; Van Sluys, M.-A. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics***2012**, *13*, 137, doi:10.1186/1471-2164-13-137.
23. Fan, H.C.; Wang, J.; Potanina, A.; Quake, S.R. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.***2011**, *29*, 51–7, doi:10.1038/nbt.1739.
24. Fu, J.; Wang, S. Insights into Auxin Signaling in Plant-Pathogen Interactions. *Front. Plant Sci.***2011**, *2*, 1–7, doi:10.3389/fpls.2011.00074.
25. Garsmeur, O.; Droc, G.; Antonise, R.; Grimwood, J.; Potier, B.; Aitken, K.; Jenkins, J.; Martin, G.; Charron, C.; Hervouet, C.; et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.***2018**, *9*, 2638, doi:10.1038/s41467-018-05051-5.
26. Giorgi, D.; Farina, A.; Grosso, V.; Gennaro, A.; Ceoloni, C.; Lucretti, S. FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy. *PLoS One***2013**, *8*, e57994, doi:10.1371/journal.pone.0057994.
27. Grotz, N.; Fox, T.; Connolly, E.; Park, W.; Guerinot, M. Lou; Eide, D. Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proc. Natl. Acad. Sci. U. S. A.***1998**, *95*, 7220–4, doi:10.1073/pnas.95.12.7220.
28. Hotta, C.T.; Lembke, C.G.; Domingues, D.S.; Ochoa, E.A.; Cruz, G.M.Q.; Melotto-Passarini, D.M.; Marconi, T.G.; Santos, M.O.; Mollinari, M.; Margarido, G.R.A.; et al. The Biotechnology Roadmap for Sugarcane Improvement. *Trop. Plant Biol.***2010**, *3*, 75–87, doi:10.1007/s12042-010-9050-5.
29. Kellogg, E.A. Relationships of cereal crops and other grasses. *Proc. Natl. Acad. Sci. U. S. A.***1998**, *95*, 2005–2010, doi:10.1073/pnas.95.5.2005.
30. Kirkness, E.F.; Grindberg, R. V.; Yee-Greenbaum, J.; Marshall, C.R.; Scherer, S.W.; Lasken, R.S.; Venter, J.C. Sequencing of isolated sperm cells for direct haplotyping of a human genome. *Genome Res.***2013**, *23*, 826–32, doi:10.1101/gr.144600.112
31. Kreplak, J.; Madoui, M.A.; Cápál, P.; Novák, P.; Labadie, K.; Aubert, G.; Bayer, P.; Gali, K.K.; Syme, R.A.; Main, D.; Areference genome for pea provides insight into legume genome evolution. *Nature genetics.* **2019**, *51*, 1411-1422
32. Lasswell, J.; Rogg, L.E.; Nelson, D.C.; Rongey, C.; Bartel, B. Cloning and characterization of *IAR1*, a gene required for auxin conjugate sensitivity in arabidopsis. *Plant Cell***2000**, *12*, 2395–2408, doi:10.1105/tpc.12.12.2395.
33. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM Available online: arxiv:1303.3997v1 [q-bio.GN] 2013.
34. Li Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics***2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
35. zardi, P.M.; Huang, X.; Zhu, Z.; Bray-Ward, P.; Thomas, D.C.; Ward, D.C. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.***1998**, *19*, 225–232, doi:10.1038/898.
36. Ma, L.; Xiao, Y.; Huang, H.; Wang, Q.; Rao, W.; Feng, Y.; Zhang, K.; Song, Q. Direct determination of molecular haplotypes by chromosome microdissection. *Nat. Methods***2010**, *7*, 299–301, doi:10.1038/nmeth.1443.
37. Magarey, R.C.; Bull, J.I. Relating cultivar *Pachymetra* root rot resistance to sugarcane yield using breeding selection trial analyses. *Aust. J. Exp. Agric.***2003**, *43*, 617–622, doi:10.1071/EA02144.
38. Magarey, R.; Croft, B. *Pachymetra* root rot: Incidence and potential solutions to its influence on yield decline in Queensland. In *Sugarcane: Research towards efficient and sustainable production*; Wilson, J., Hogarth, D.,

- Campbell, J., Garside, A., Eds.; CSIRO Division of Tropical Crops and Pastures: Brisbane, Australia, 1996; pp. 151–152.
39. Marcy, Y.; Ishoey, T.; Lasken, R.S.; Stockwell, T.B.; Walenz, B.P.; Halpern, A.L.; Beeson, K.Y.; Goldberg, S.M.D.; Quake, S.R. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet.***2007**, *3*, 1702–1708, doi:10.1371/journal.pgen.0030155.
 40. Matsubara, K.; Iwasaki, Y.; Nishiki, I.; Nomura, K.; Fujiwara, A. Identification of genetic linkage group 1-linked sequences in Japanese eel (*Anguilla japonica*) by single chromosome sorting and sequencing. *PLoS One***2018**, *13*, e0197040, doi:10.1371/journal.pone.0197040.
 41. Meng, Z.; Han, J.; Lin, Y.; Zhao, Y.; Lin, Q.; Wang, J.; Zhang, M.; Zhang, L.; Yang, Q.; Wang, K. Characterisation of a *Saccharum spontaneum* with a basic chromosomes number of $x=10$ provides new insights on genome evolution in genus *Saccharum*. *Theoretical and Applied Genetics***2019** doi.org/10.1007/s00122-019-03450-w
 42. Metcalfe, C.J.; Li, J.; Giorgi, D.; Doležel, J.; Piperidis, N.; Aitken, K.S. Flow cytometric characterisation of the complex polyploid genome of *Saccharum officinarum* and modern sugarcane cultivars. *Sci. Rep.***2019**, *9*, 19362, doi:10.1038/s41598-019-55652-3.
 43. Paunio, T.; Reima, I.; Syvänen, A.C. Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA. *Clin. Chem.***1996**, *42*, 1382–1390.
 44. Piperidis, N. GISH: Resolving Interspecific and Intergeneric Hybrids. In *Molecular Plant Taxonomy: Methods and Protocols, Methods in Molecular Biology*, Besse, P., Ed.; Springer Science+Business Media: New York, 2014; Vol. 1115, pp. 325–336 ISBN 978-1-62703-766-2.
 45. Piperidis, N.; Chen, J.; Deng, H.; Wang, L.-P.; Jackson, P.; Piperidis, G. GISH characterization of *Erianthus arundinaceus* chromosomes in three generations of sugarcane intergeneric hybrids. *Genome***2010**, *53*, 331–336, doi:10.1139/G10-010.
 46. Piperidis, G.; Piperidis, N.; D'Hont, A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. Genomics***2010**, *284*, 65–73, doi:10.1007/s00438-010-0546-3.
 47. Piperidis, N.; Tom, C.; Aitken, K.S.; Atkin, F.C.; Piperidis, G. Exploiting Erianthus diversity to enhance sugarcane cultivars. *41st Annu. Conf. - Aust. Soc. Sugar Cane Technol. 2019***2019**, *41*, 108–116.
 48. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics***2010**, *26*, 841–842, doi:10.1093/bioinformatics/btq033.
 49. Rhee, M.; Light, Y.K.; Meagher, R.J.; Singh, A.K. Digital Droplet Multiple Displacement Amplification (ddMDA) for Whole Genome Sequencing of Limited DNA Samples. *PLoS One***2016**, *11*, e0153699, doi:10.1371/journal.pone.0153699.
 50. Roach, B. Nobilisation of sugarcane. *Proc. Int. Soc. Sugar Cane Technol.***1972**, *14*, 206–216.
 51. Sánchez-Martín, J.; Steuernagel, B.; Ghosh, S.; Herren, G.; Hurni, S.; Adamski, N.; Vrána, J.; Kubaláková, M.; Krattinger, S.G.; Wicker, T.; et al. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.***2016**, *17*, 1–7, doi:10.1186/s13059-016-1082-1.
 52. Santchurn, D.; Ramdoyal, K.; Badaloo, M.G.H.; Labuschagne, M.T. From sugar industry to cane industry: Evaluation and simultaneous selection of different types of high biomass canes. *Biomass and Bioenergy***2014**, *61*, 82–92, doi:10.1016/j.biombioe.2013.11.023.
 53. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bolton, E.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.***2011**, *39*, D38–D51, doi:10.1093/nar/gkq1172.

54. Spits, C.; Le Caignec, C.; De Rycke, M.; Van Haute, L.; Van Steirteghem, A.; Liebaers, I.; Sermon, K. Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.***2006**, *1*, 1965–70, doi:10.1038/nprot.2006.326.
55. Telenius, H.; Carter, N.P.; Bebb, C.E.; Nordenskjöld, M.; Ponder, B.A.J.; Tunnacliffe, A. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics***1992**, *13*, 718–725, doi:10.1016/0888-7543(92)90147-K.
56. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A.M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 2007, *35*, W71–W74, doi:10.1093/nar/gkm306.
57. Vrána, J.; Cápal, P.; Šimková, H.; Karafiátová, M.; Čížková, J.; Doležel, J. Flow Analysis and Sorting of Plant Chromosomes. *Curr. Protoc. Cytom.***2016**, *78*, 5.3.1-5.3.43, doi:10.1002/cpcy.9.
58. Waclawovsky, A.J.; Sato, P.M.; Lembke, C.G.; Moore, P.H.; Souza, G.M. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.***2010**, *8*, 263–276, doi:10.1111/j.1467-7652.2009.00491.x.
59. Wang, X.; Heazlewood, S.P.; Krause, D.O.; Florin, T.H.J. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J. Appl. Microbiol.***2003**, *95*, 508–520, doi:10.1046/j.1365-2672.2003.02005.x.
60. Wu, J.; Huang, Y.; Lin, Y.; Fu, C.; Liu, S.; Deng, Z.; Li, Q.; Huang, Z.; Chen, R.; Zhang, M. Unexpected inheritance pattern of *Erianthus arundinaceus* chromosomes in the intergeneric progeny between *Saccharum* spp. and *Erianthus arundinaceus*. *PLoS One***2014**, *9*, doi:10.1371/journal.pone.0110390.
61. Yang, H.; Chen, X.; Wong, W.H. Completely phased genome sequencing through chromosome sorting. *Proc. Natl. Acad. Sci. U. S. A.***2011**, *108*, 12–17, doi:10.1073/pnas.1016725108.
62. Zhang, L.; Cui, X.; Schmitt, K.; Hubert, R.; Navidi, W.; Arnheim, N. Whole genome amplification from a single cell: Implications for genetic analysis. *Proc. Natl. Acad. Sci. U. S. A.***1992**, *89*, 5847–5851, doi:10.1073/pnas.89.13.5847.
63. Zhang, J.; Zhou, M.; Walsh, J.; Zhu, L.; Chen, Y.; Ming, R. Sugarcane Genetics and Genomics. *Sugarcane Physiol. Biochem. Funct. Biol.***2014**, 623–643, doi:10.1002/9781118771280.ch23.

Tables

Table 1. Fraction of chromosomes flow sorted from composite peaks onto slides labelled by *Erianthus* whole genomic DNA. The composite peak regions are shown in Figure 3. Composite peak IV was split into two equal regions, IVa and IVb.

Composite Peak/s	# unlabelled chromosomes	# labelled chromosomes	Total	Fraction labelled (%)	Fraction recombinant chromosome (%)
III and IV	966	143	1,109	12%	6%
III	344	16	360	4%	4%
IV	499	50	549	9%	8%
IVa	480	66	546	12%	7%
IVb	800	53	853	6%	4%

Table 2. Summary of the top blastn hits against the NCBI nucleotide database for a random sample of 5,000 reads from each of the Illumina sequenced MDA products.

	MDA one		MDA two	
	percentage	mean	percentage	mean
	of hits	e-value	of hits	e-value
<i>Saccharum</i>	40.48	1.25 e ⁻¹²	47.41	1.64 e ⁻¹²
<i>Sorghum</i>	7.06	1.39 e ⁻¹²	6.24	3.53 e ⁻¹²
<i>Zea</i>	2.97	2.59 e ⁻¹²	1.24	2.60 e ⁻¹²
<i>Oryza</i>	1.23	4.69 e ⁻¹²	0.54	5.43 e ⁻¹²
Other Poaceae	1.46	4.30 e ⁻¹²	1.11	2.88 e ⁻¹²
Other Plants	0.04	2.89 e ⁻¹²	0.15	8.24 e ⁻¹²
Other	0.08	5.08 e ⁻¹²	0.08	1.45 e ⁻¹²
No hit	46.68		43.23	

Table 3. Summary of sequencing, mapping and gene identification results for the two *Erianthus* single-chromosome MDA products.

Sequencing data				Mapping to R570 STP genome			
	Raw reads total data (Gb)	Trimmed reads total data (Gb)	Median read length	Chromosome with highest fraction of reads	% reads		
MDA One	9.13	5.62	107	Sh07	65		
MDA Two	9.88	5.98	107	Sh05	33		
Mapping to R570 STP genes (q20)							
	Chromosome with highest fraction of reads	% reads	Chromosome with highest mean depth	Mean depth	Mean % breath	% genes with reads	% resistance genes with reads
MDA One	Sh07	39	Sh07	74.05	44.15	62	57
MDA Two	Sh04	32	Sh05	30.68	12.59	20	16
SNPs							
	SNPs identified in resistance genes	Number Primers designed					
MDA One	120	4					
MDA Two	15	2					

Table 4. SNPs identified in classic resistance genes. For each resistance gene, the number of SNPs identified, primers designed, position of the SNP, the reference and allele nucleotide, quality at the SNP position are shown. One primer was designed for each gene identified, except for the MDA two gene identified, where two primers were designed.

MDA	Gene ID	Annotation	Number of SNPs identified	Chromosome	Position	Reference	Allele	Quality
One	Sh07_g006930	weakly similar to NB-ARC domain containing protein	4	Sh07	10,134,638	T	G	200
	Sh07_g008080	Putative disease resistance protein At3g14460	2	Sh07	12,986,768	A	G	33
	Sh07_g011420	IAA-alanine resistance protein 1	18	Sh07	20,170,249	C	T	103
	Sh07_g017720	F-box/LRR-repeat protein At3g26922	4	Sh07	29,262,158	T	C	180
Two	Sh05_g003680	Putative Disease resistance protein RPM1	7	Sh05	6,273,137	A	C	71
				Sh05	6,273,830	T	G	134

Figures

Figure 1

Pedigree for the *Saccharum-Erianthus* BC4 clone QC12-20006. Boxes shaded in blue = *Saccharum-Erianthus* hybrids, box shaded in orange = *Erianthus*.

Figure 2

GISH analysis of the *Saccharum*-*Erianthus* BC4 hybrid, QC12-20006. QC12-20006 metaphases were probed with whole *Erianthus* DNA labelled with biotin-14-dUTP and visualised amplified/detected with avidin-conjugated Texas red and genomic DNA from *S. officinarum* labelled with DIG and detected with FITC according to (Piperidis et al. 2010; Piperidis 2014). The recombinant chromosome is indicated with the yellow arrow.

Figure 3

Summary of results (a) Histogram of relative DNA fluorescence of chromosomes (flow karyotypes) obtained by flow cytometric analysis of DAPI-stained chromosome suspensions of QC12-20006. The flow karyotype was divided by peaks into five regions (dotted lines) and 2,000 to 3,000 chromosomes from each region were flow sorted onto microscope slides.; (b) Flow sorted chromosomes (grey) were probed with whole *Erianthus* genomic DNA (red) to identify regions with *Erianthus* chromosome. (c) Single chromosomes from regions most enriched for the *Erianthus* chromosomes (dark grey region in (a)) were flow sorted into a 96 well plate, purified and MD amplified. (d) amplified DNA from 30 single-flow sorted chromosome were PCR-screened using 3 sets of primers, a primer specific to a *Saccharum*/*Erianthus* transposable element, an *Erianthus* specific repeat primer, and a universal bacterial 16S rDNA primer. Two single chromosomes positive for the *Saccharum*/*Erianthus* transposable element and the *Erianthus* specific repeat, but negative for the universal bacterial primer, were Illumina sequenced and analysed for contamination and coverage. (e) 136 SNPs within 31 annotated resistance genes were identified from mapped

illumina reads from the two chromosomes. (f) Six KASP primers were designed to SNPs within 5 classic resistance genes and used to determine the segregation ratio of the SNP markers across 29 clones from a *Saccharum-Erianthus* BC3 population using the KASP system. One SNP segregated as a single dose marker separating the progeny into two groups one containing the *Erianthus* heterozygous SNP (red points) and the other the homozygous group inherited from *Saccharum* (green points).

Figure 4

Coverage for the R570 STP chromosome by Illumina reads obtained from the two *Erianthus* single-chromosome MDA products. Results are shown for chromosome seven (MDA one) and chromosome five (MDA two) (a). Coverage along chromosome (b). Depth of coverage. The read depth is indicated by the color-coded key. MDA one has a higher fraction of the chromosome with increased depth compare with MDA two.

Figure 5

Fraction of R570 STP genes covered by Illumina reads obtained from the two *Erianthus* single-chromosome MDA products. (a) Fraction of genes covered from all ten chromosomes. (b) Fraction of genes from chromosome seven (MDA one) or chromosome five (MDA two) that are missing (0%), covered up to 40%, 40-80%, and 80-100% of the breadth of the gene, (i) = all genes and (ii) = classic resistance genes. MDA one clearly maps to chromosome 7, while MDA two does not map clearly to a single chromosome. MDA one also has a higher fraction of genes with > 80% coverage breadth, suggesting that MDA one is the product of a more successful amplification.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.Primers.xlsx](#)
- [TableS2.sequencingandmappingdata.xlsx](#)
- [TableS3chromosomefrequencygenomemapping.xlsx](#)
- [TableS4.Statsreadmappingallchromosomes.xlsx](#)
- [TableS5.readmappingchr5andchr7.xlsx](#)
- [TableS6.AllSNPsidentified.xlsx](#)