Mol Breeding (2016) 36:60
DOI 10.1007/s11032-016-0482-y

Targeted sequencing of a complex locus within a polyploid genome using reduced representation libraries

Joann Conner · Sirjan Sapkota · Stéphane Deschamps · Kevin Fengler · Mark Cigan · Peggy Ozias-Akins

Received: 4 September 2015 / Accepted: 26 April 2016 / Published online: 4 May 2016

Abstract Apospory is a form of gametophytic apomixis in which embryos develop from unreduced embryo sacs derived from aposporous initials formed from nucellar cells of ovules to produce offspring genetically identical to the female plant. Apospory in *Pennisetum squamulatum* (8X) and *Cenchrus ciliaris* (4X) is a dominant trait controlled by a physically large, hemizygous, heterochromatic chromosomal block called the apospory-specific genomic region (ASGR). Both apomictic species are polyploid, with genome sizes estimated at 2600 to 3000 Mbp for *C. ciliaris* and 9400 to 10,300 Mbp for *P. squamulatum*. A study was conducted to determine whether duplex-specific nuclease (DSN) normalization of DNA from apomictic and sexual genotypes would reduce repetitive sequences and allow bioinformatic analysis to predict sequence contigs derived from the ASGR. DSN libraries from four genotypes were sequenced using Illumina® HiSeq 2000 technology. 39 out of 44 tested sequence characterized amplified region (SCAR) markers from in silico predicted ASGR-specific contigs were mapped to the ASGR in a *Pennisetum* F_1 mapping population. Eighteen SCARs showed apomict-specific amplification in *C. ciliaris*. The successful mapping of ~90 % of the SCAR markers to the ASGR in the *Pennisetum* F_1 mapping population shows that DSN normalization and Illumina sequencing can be used as an effective strategy for targeted mapping of a physically large locus rich in repetitive sequences, like that of the ASGR.

Keywords Apomixis · Duplex-specific nuclease · Apospory-specific genomic region · Polyploidy · Mapping

Electronic supplementary material The online version of this article (doi:10.1007/s11032-016-0482-y) contains supplementary material, which is available to authorized users.

J. Conner (✉) · S. Sapkota · P. Ozias-Akins
Department of Horticulture, University of Georgia – Tifton Campus, Tifton, GA 31793, USA
e-mail: jconner@uga.edu

S. Deschamps
DuPont Pioneer, DuPont Experimental Station, Wilmington, DE 19803, USA

K. Fengler · M. Cigan
DuPont Pioneer, Johnston, IA 50131, USA

Introduction

Apomixis is an asexual mode of reproduction in which viable embryos are produced from unreduced egg cells of the ovule without fertilization of gametes (Nogler 1984). Endosperm of apomictic plants can be derived either autonomously or through fertilization of the central cell (Ozias-Akins 2006). Because apomictic plants produce seeds that are genetically identical to the maternal plant, apomixis has been identified as a potential mechanism to fix heterosis for major crops

(Savidan 2000; Ozias-Akins 2006; Hand and Koltunow 2014). Although it has been documented in more than 120 angiosperm genera, apomixis is rarely found in major crop species (Carman 1997; Hand and Koltunow 2014). Efforts have been made to transfer the apomictic trait to maize (Leblanc et al. 1995) and pearl millet (Dujardin and Hanna 1989b) by interspecific hybridization. These efforts have not yet resulted in commercially viable germplasm (Savidan 2000; Ozias-Akins and Van Dijk 2007; Hand and Koltunow 2014). Alternatively, the transfer of key genes characterized to govern components of apomixis to sexual plants would also confer the trait of apomixis and enable application of apomixis to crop breeding (Koltunow et al. 1995; Grimanelli et al. 2001; Ozias-Akins 2006).

Apomixis is often associated with polyploidy (Nogler 1984; Voigt-Zielinski et al. 2012). Apomixis can be classified as sporophytic or gametophytic. Gametophytic apomixis is further categorized as diplospory or apospory. In diplospory, the unreduced embryo sac forms from the megaspore mother cell which does not complete the meiotic pathway but instead undergoes mitotic divisions to form an unreduced embryo sac. In apospory, the unreduced embryo sac develops from mitotic divisions of aposporous initial cells which are formed from one or more nucellar cells of the ovule (Ozias-Akins 2006; Hand and Koltunow 2014). The pathway of gametophytic apomixis requires two components, apomeiosis, the formation of unreduced embryo sacs, and parthenogenesis, the development of the embryo without fertilization, in addition to successful endosperm formation.

In *P. squamulatum*, both apomeiosis and parthenogenesis are controlled by a dominant genetic locus transmitted as a physically large (>50 Mb), hemizygous chromosomal block named the apospory-specific genomic region (ASGR) (Ozias-Akins et al. 1998; Goel et al. 2003; Akiyama et al. 2004). In other gametophytic species such as *Taraxacum officinale* (Van Dijk et al. 1999), *Poa pratensis* (Albertini et al. 2001), *Erigeron annuus* (Noyes and Rieseberg 2000) and *Hieracium caespitosum* (Catanach et al. 2006), apomeiosis and parthenogenesis are confirmed to be controlled by independent loci, which can be genetically linked. Suppressed recombination of at least one of the two components of gametophytic apomixis has been identified in apomicts characterized to date.

Sequence analysis of ASGR-linked and ASGR-recombining BAC clones, covering about 2.7 Mb of the DNA, showed that the ASGR has both gene-rich and gene-poor regions with several identified genes that could be postulated to play a role in the apomictic pathway (Conner et al. 2008). The ASGR also contains abundant repetitive elements, such as an Opie-2-like retrotransposon (Akiyama et al. 2004; Conner et al. 2008). Sequence analysis of the *ASGR-BABY BOOM-Like* (*ASGR-BBML*) gene in *Pennisetum*/*Cenchrus* species, along with FISH analysis of the low copy BAC containing this gene, suggests that this region of the ASGR is of relatively recent origin (Akiyama et al. 2011).

Next-generation sequencing has opened a platform for investigating the genomes of crop species. Polyploidy and repetitive sequences, which can constitute a large proportion of a plant's genome, have hindered the assembly of large genomes. Limiting the sequencing of repetitive elements can facilitate sequence analyses of large and complex genomes. Several approaches have been used for the elimination of repetitive sequences and the enrichment of low copy sequences. In maize, approaches like Methylation filtration (MF) (Palmer et al. 2003), development of hypomethylated partial restriction (HMPR) libraries (Emberton et al. 2005) and methylation-spanning linker libraries (MSLL) (Yuan et al. 2002) utilize the tendency of repetitive sequences to be hypermethylated in higher plants and have been used to eliminate repetitive sequences and enrich for low copy sequences. High-C_0t analysis is based on the renaturation kinetics of sheared DNA with the rate of reassociation of a particular sequence being directly proportional to its repetitiveness in the genome. The re-annealed double-stranded repetitive DNA can then be separated from the lower copy DNA by hydroxyapatite chromatography. This approach has been used in the study of the maize and sorghum genomes (Peterson et al. 2002; Yuan et al. 2003).

A more recent approach to the removal of repetitive DNA is duplex-specific nuclease (DSN) normalization technology. This technology, also based on hybridization kinetics, does not involve physical separation of single-stranded (ss) from double-stranded (ds) DNA. Instead, the DSN enzyme, which displays a strong preference for cleaving dsDNA compared to ssDNA, can be used to eliminate more rapidly annealing repetitive DNA (Shagin et al. 2002). DSN

normalization technology was initially used to enrich full-length cDNA sequences and to discover rare transcripts by removing intermediate and highly abundant cDNAs (Zhulidov et al. 2004, 2005). The application of DSN normalization in human whole-genome shotgun sequencing indicated that DSN normalization was effective for the elimination of evolutionarily young repetitive sequences from genomic DNA prior to sequencing (Shagina et al. 2010). The genomes of many higher plants contain a large number of highly conserved repetitive elements that are evolutionarily young and share high sequence identity (Smith and Flavell 1975; SanMiguel et al. 1998; Ramakrishna et al. 2002). Successful application of DSN normalization of plant genomic libraries has been reported for the lettuce and Arabidopsis genomes (Matvienko et al. 2013). This study identified that DSN normalization could enable efficient access to the low copy fraction of the genome for identifying sequence and copy number variants and also for mapping purposes. Furthermore, libraries for Illumina sequencing with short 300- to 500-bp fragments are suited for reducing highly repeated sequences without the concomitant loss of neighboring low copy sequences.

In *P. squamulatum,* 12 sequence characterized amplified region (SCAR) markers derived from RAPD analysis, 7 amplified fragment length polymorphism (AFLP) markers and 45 sequence-specific amplified polymorphism (SSAP) markers, based on the long terminal repeat region of an ASGR-abundant retrotransposon, have been mapped to the ASGR (Ozias-Akins et al. 1998; Goel et al. 2006; Huo et al. 2009). In the present study, we show that the use of DSN normalization technology, Illumina sequencing, and in silico analysis is an effective strategy for the identification of sequences unique to the ASGR.

Materials and methods

Plant material

Seven apomictic (R5-04, R5-19, R5-41, R5-46, R5-65, R5-99 and R5-102) and five sexual (R5-33, R5-44, R5-59, R5-77 and R5-87) plants derived from an F_1 population (R5-F_1) obtained from a cross between *P. squamulatum* (Ps26; $2n = 8x = 56$) and *P. glaucum* ($2n = 4x = 28$) (Goel et al. 2006) were used for nuclear DNA isolation. Four apomictic and five sexual plants derived from line BC_8 (06-A-58), which only contains the ASGR-carrier chromosome from Ps26 (Singh et al. 2010), were also used for nuclear DNA isolation.

Nuclear DNA extraction

Two to five grams of leaf tissue from 4 BC_8 apomictic plants (S1, BC_8-apo), 5 BC_8 sexual plants (S2, BC_8-sex), 5 F_1 apomictic plants (S3, F_1-apo) and 7 F_1 sexual plants (S4, F_1-sex) were used in separate nuclei DNA preparations as described in Roche et al. (2002). DNA was purified from the isolated nuclei using QIAprep® Spin Miniprep Kit (QIAGEN Inc., Valencia, CA, USA) or NucleoSpin® Plant II Kit (Clontech Laboratories, Mountain View, CA, USA). ASGR-linked SCAR p787/p788 was used to confirm nuclear DNA genotypes (Singh et al. 2010). DNA from each purified sample was quantitated with PicoGreen® (Life Technologies, Grand Island, NY, USA). DNA isolated for each sample type was equally pooled based on the picogreen reading to contain 11–12.5 μg of DNA.

DSN normalization

The following protocol is based on the protocol of Shagina et al. (2010). DNA of S1 and S2 was digested with restriction enzyme *Bts*CI (New England Biolabs) and purified with MinElute columns (Qiagen). *Bts*CI universal adapters (upper strand, 5′ACACTCTT TCCCTACACGACGCTCTTCCGATCTGGATGNN; lower strand, 5′CATCCAGATCGGAAGAGCGTCG TGTAGGGAAAGAGTGT) were ligated to the digested DNA fragments and purified with Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). An aliquot of the DNA was removed as the untreated control sample (S1a, S2a). The remaining DNA was denatured for 2 min at 98 °C and allowed to renature at 68 °C for 4 (S1b, S2b) or 8 h (S1c, S2c), followed by 25-min incubation at 68 °C with 2 Units of DSN. The DSN treatment was inactivated, and samples were purified with Agencourt AMPure XP beads (Beckman Coulter). The normalized single-stranded DNA fraction and untreated control were amplified by 20 cycles of PCR, using the upper strand adapter as primer, purified with Agencourt AMPure XP (Beckman Coulter) beads, digested with *Bts*CI and purified again with Agencourt

AMPure XP beads (Beckman Coulter). S3 and S4 were treated as above but only with an 8-h renaturation (S3c and S4c, respectively).

Illumina library construction and sequencing of DSN libraries

DNA from S1a, S2a, S1b, S2b, S1c and S2c was randomly sheared to 300–500 bp (Covaris Inc., Woburn, MA), purified with Agencourt AMPure XP beads (Beckman Coulter), and Illumina paired-end libraries were constructed according to the manufacturer's protocol. S1a, S2a, S1b, S2b, S1c and S2c libraries were sequenced at a 50-nt read length on Illumina® HiSeq 2000. S3c and S4c libraries were prepared as above. S1c, S2c, S3c and S4c libraries were sequenced at a 100-nt read length on Illumina® HiSeq 2000. Adapter and poor-quality sequences were removed from all sequence datasets prior to analysis.

Assembly of reference DSN library S1c

Velvet de novo (Zerbino 2010) and SPAdes (Bankevich et al. 2012) genome assembly algorithms were used to assemble contiguous sequences from quality-trimmed 100-bp paired-end dataset of Illumina sequencing for S1c. Five Velvet assemblies with K-mer values of 31, 39, 47, 55 and 63 and minimum contig length of 400 bp were created. Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990) analysis was used to identify unique contigs from each Velvet *de novo* assembly to build a more complete Velvet reference sequence assembly. The 55 k-mer assembly was the database for BlastN analysis against the 63 K-mer assembly with a cutoff of e^{-10}. Any contig from the 63 K-mer assembly with no BlastN hit to the 55 K-mer database was considered unique and added to the 55 K-mer assembly. This process was repeated with the 47, 39 and 31 K-mer assemblies, resulting in the BC8ApoDSN8_V assembly containing 135,324 contigs.

The SPAdes assembler (v3.5.0), with default parameters, generated 503,824 assembled contigs ranging from 8178 to 78 base pairs in length. After removal of all contigs under 400 bp, 221,389 contigs remained and were labeled the BC8ApoDSN8_S assembly.

Bowtie sequence alignment

The BC8ApoDSN8_V and BC8ApoDSN8_S assemblies were used as the reference genome to which reads from the S1c, S2c, S3c and S4c libraries were aligned using Bowtie version 1.1.0 (Langmead et al. 2009) at the Georgia Advanced Computing Resource Center (GACRC) at the University of Georgia (http://www.gacrc.uga.edu). Quality-trimmed reads from the 100-bp sequencing runs were aligned to both the BC8ApoDSN8_V and BC8ApoDSN8_S assemblies using Bowtie parameters–v0–a–best. Bowtie output was converted using SAMtools view, sort, index and idxstats (Li et al. 2009).

PCR conditions and primer screening

PCRs consisted of 0.375 U of *Jumpstart*^{TM} *Taq* DNA polymerase (Sigma-Aldrich Co., St. Louis, MO, USA) with 0.25-mM dNTP and 0.25-μM primers, or 0.5 U of *TaKaRa*^{TM} *Ex Taq* DNA polymerase Hot Start Version (Clontech Laboratories) with 0.2-mM dNTP and 0.5-μM primers, in a total reaction volume of 20 μl with 1X PCR buffer with 25- to 50-ng template DNA and were amplified for 30–35 cycles (Online Resource Table 1). The amplified PCR products were analyzed with a 1.5 % (w/v) agarose gel using electrophoresis.

Primer screening was done in two steps. The SCAR primers were first screened with DNA from 4 F_1 apomicts, 4 F_1 sexual individuals, 3 BC_8 (06-A-58) apomicts, 3 BC_8 (06-A-58) sexual individuals, Ps26 and *P. glaucum* (IA4X, an induced sexual tetraploid millet). Primers which only amplified DNA from apomicts were screened on a larger F_1 population (Online Resource Table 1) which consisted of 17 F_1 apomicts, 22 F_1 sexual lines, a BC_8 (06-A-58) apomict, a BC_8 (06-A-58) sexual, Ps26, apomictic *C. ciliaris* genotype B-12-9, sexual *C. ciliaris* genotype B-2S and recombinant *C. ciliaris* A8 (Conner et al. 2013). DNA extraction, if needed, was done from approximately 100 mg of young leaf tissue using DNeasy® Plant Mini Kit (Qiagen). All DNA used in the experiment had genotypes confirmed using ASGR-SCAR primers 787/788 and 792/793 (Conner et al. 2013).

Primer design

Primers were designed using Geneious (http://www.geneious.com) or primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3). Primer information is listed in Online Resource Table 2.

Results

DSN treatment and sequencing

Our first goal was to determine whether DSN normalization could be used to reduce the repetitiveness/complexity of the apomictic BC_8 (06-A-58) genome. This genome was chosen over the naturally apomictic, *P. squamulatum,* for its smaller genome size (7800 Mbp vs. ~9400 to 10,300 Mbp) and reduced ploidy level (4X vs. 8X). DNAs from both apomictic and sexual BC_8 (06-A-58) plants were tested at 4- and 8-h renaturation prior to DSN treatment and compared to untreated samples using Illumina 50-bp paired-end reads and 31-bp K-mer analysis. As shown in Fig. 1, the DNA samples undergoing DSN normalization treatment showed enrichment in abundance of the low copy number 31 K-mer sequences, with the 8 h DSN treatment slightly more enriched than the 4-h DSN treatment. The remaining samples (S3 and S4) were prepared as S1c/S2c, and all four sample libraries were sequenced as 100-bp paired-end reads. The number of reads generated from sequencing of samples S1c, S2c, S3c and S4c) are listed in Table 1.

Assembly and reference sequences/contigs

Sequence reads from S1c (BC_8-apo_DSN8) were assembled using Velvet (Zerbino 2010) and SPAdes assembly algorithms (Bankevich et al. 2012). The Velvet assembler was able to use ~30 % of the total library reads and assemble contigs that were a minimum 400 bp and on average 750 bp in length. To create the most comprehensive database of the Velvet assembly, unique contigs within each K-mer assembly were identified through BlastN analysis with cutoff value of e^{-10}. A total of 12,464; 9938; 1191; and 667 unique contigs were identified from the 63, 31, 39 and 47 K-mer assemblies, respectively, and added step by step to form the 135,324 BC8ApoDSN8_V assembly.

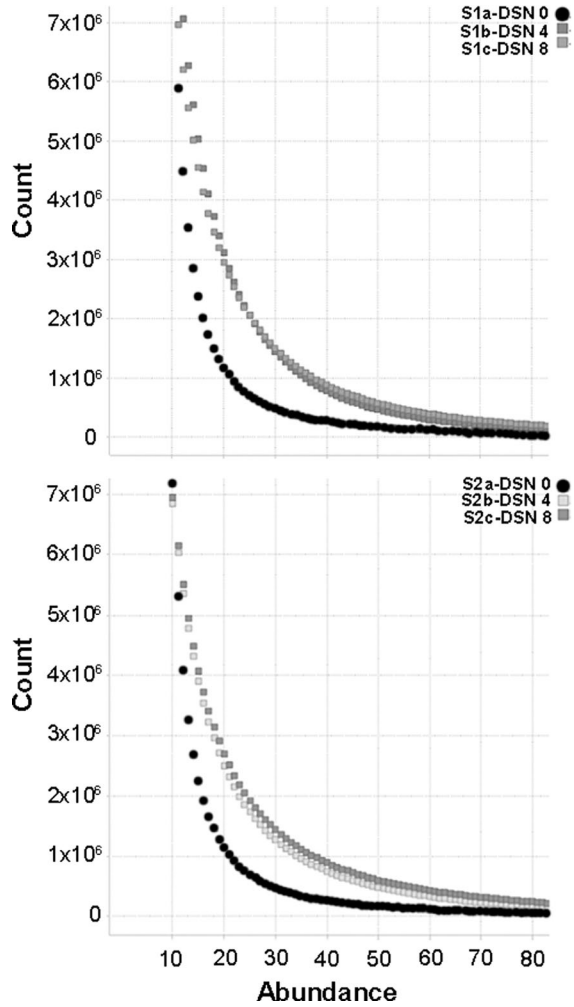

Fig. 1 Count versus abundance of 31K-mers for S1 and S2 control and DSN-treated sequences. Graphs shown compare count versus abundance of 31K-mers identified in S1 (*upper*) and S2 (*lower*) DSN untreated and treated libraries derived from BC_8 apomict (S1) and sexual (S2) DNA

The SPAdes assembler generated 503,824 contigs with 221,389 contigs 400 bp or longer. The BC8ApoDSN8_S assembly contained ~50 % of the S1c library reads and averaged 891 bp in length. Both the BC8ApoDSN8_V and BC8ApoDSN8_S assemblies were used for analysis.

Both assemblies were searched against the *Zea* and *Sorghum bicolor* repeat database (RM database version 20140131) using RepeatMasker (version open-4.0.5). The percent masked bases for repeat elements were 9.5 and 13.8 % for *Zea* and *Sorghum*, respectively, with retroelements representing 7.8 and 10.0 %

Table 1 Number of 100-bp Illumina sequence reads generated from DSN-treated libraries and percentage of read hits back to the BC8ApomicticDSN8 assemblies using Bowtie

Sample	Sample description	Single read count	Paired read count	Percent reads aligned to Velvet assembly (%)	Percent reads aligned to SPAdes assembly (%)
S1c	BC_8-apo_DSN8	22,723,940	268,942,912	28	51
S2c	BC_8-sex_DSN8	109,447,045	161,638,656	27	49
S3c	F_1-apo_DSN8	43,253,545	96,779,228	9	17
S4c	F_1-sex_DSN8	47,377,584	108,724,480	8	15

and DNA transposons represented 1.5 and 3.5 % for the BC8ApoDSN8_V assembly. The percent masked bases for repeat elements were 11.1 and 15.6 % for *Zea* and *Sorghum*, respectively, with retroelements representing 9.1 and 11.2 % and DNA transposons represented 1.8 and 4.1 % for the BC8ApoDSN8_S assembly.

Identifying potential ASGR-linked contigs using an in silico approach

As the apomictic BC_8 (06-A-58) plants contain the ASGR-carrier chromosome which is lacking in the sexual BC_8 (06-A-58) plants (Singh et al. 2010), contigs from BC8ApoDSN8_V or BC8ApoDSN8_S assembly with reads aligning from BC_8 apo_DSN8 but not from BC_8 sex_DSN8 should be highly enriched for sequences derived from the ASGR-carrier chromosome. A total of 2,811 BC8ApoDSN8_V contigs had zero BC_8 sex_DSN8 reads and a range of 11–5710 aligned BC8ApoDSN8 reads. In contrast, 5777 BC8ApoDSN8_S contigs had zero BC_8 sex_DSN8 reads and a range of 11 to 9758 aligned BC8ApoDSN8 reads.

We next determined whether the potential ASGR-carrier chromosome contigs identified above contained portions of *PsASGR-BBML*, an ASGR-linked gene extensively characterized in the laboratory (Gualtieri et al. 2006; Conner et al. 2015). Two contigs were identified in the subset of BC8ApoDSN8_V sequences to contain portions of *PsASGR-BBML*. One contig containing a portion of *PsASGR-BBML* had 37 F_1-apo_DSN8 and zero F_1-sex_DSN8 hits. The other BC8ApoDSN8_V contig with portions of the *PsASGR-BBML* gene did not have any F_1-apo_DSN8 or F_1-sex_DSN8 hits. In contrast, no contigs were identified in the subset of BC8ApoDSN8_S sequences to contain portions of *PsASGR-BBML*. Therefore, a search of the entire BC8ApoDSN8_S assembly was done. Two contigs with portions of *PsASGR-BBML* were identified. Both SPAdes contigs were larger than those of the Velvet assembly by 140 and 27 bp. This variation in size was enough to create bowtie hits with the BC_8 sex_DSN8 sequences thus eliminating them from the ASGR-carrier chromosome subset.

Given the *PsASGR-BBML* results, the ASGR-carrier chromosome subset of the BC8ApoDSN8_V contigs that had at least 20 F_1-apo_DSN8 hits and no F_1-sex_DSN8 hits was used for primer design to validate our bioinformatics approach. A total of 44 BC8ApoDSN8_V contigs were selected for primer design and mapping (Online Resource Table 1) out of 300 possible contigs. The 44 contigs chosen did not show similarity to repetitive elements/proteins, organelle proteins or non-plant hits based on a BlastX analysis against the NCBI (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov) non-redundant protein sequence (nr) database with a cutoff value of e^{-5}, did not share significant similarity to additional contigs within the BC8ApoDSN8_V assembly or have multiple hits to the sorghum genome using a cutoff value of e^{-20} (Online Resource Table 3).

PCR amplification and mapping

Results of preliminary screening identified 39 of the 44 SCARs showing specificity to the ASGR (Fig. 2 upper panel), 4 SCARs that showed specificity to the ASGR-carrier chromosome, but not to the ASGR (Fig. 2 lower panel), and one SCAR that was amplified from all DNA screened (Online Resource Table 1). The 39 SCAR primers continued to show ASGR-linked amplification when analyzed using a larger F_1 population (Fig. 3 and Online Resource Table 1). Two SCAR primer pairs 2897/2898 and 3036/3037 did not amplify any *C. ciliaris* genotypes. 18 SCARs were

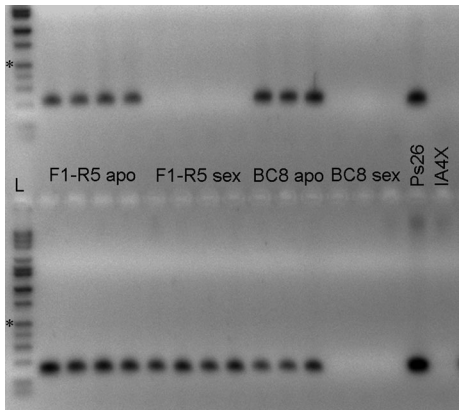

Fig. 2 Example of initial screening of DSN SCAR primers. Preliminary DSN SCAR primer screen with p3032/p3033 (*upper panel*) and p3068/p3069 (*lower panel*). L is HI-LO DNA marker (Minnesota Molecular, Minneapolis, MN) with *asterisk* denoting 500 bps

amplified from all *C. ciliaris* DNAs regardless of genotype. 15 SCARs were amplified from DNA of apomictic genotype B-12-9 but not from DNA of sexual B-2 s or recombinant A8. 4 SCARs amplified from *C. ciliaris* apomictic genotype B-12-9 and A8 but not from sexual *C. ciliaris* genotype B-2 s. The results of mapping of the SCAR primers with PCR conditions are summarized in Online Resource Table 1.

Comparison of BC8ApoDSN8_V and BC8ApoDSN8_S ASGR-linked contigs

As stated above, variation between the bioinformatics outputs between the two library assemblies was identified with *PsASGR-BBML*. We identified contigs of all 44 BC8ApoDSN8_V used for SCAR primers in the BC8ApoDSN8_S assembly, with the SPAdes contig length increased for all contigs in comparison with the Velvet assembly (Online Resource Table 4). However, only 18 of the corresponding SPAdes contigs were mapped via bioinformatic analysis with the stringent approach of zero BC_8-sex_DSN8, zero F_1-sex_DSN8 and 20 F_1-apo_DSN8 bowtie hits. As with the *PsASGR-BBML* analysis, the longer SPAdes contigs covered new areas conserved across apomicts and sexuals leading to bowtie hits in either the BC_8-sex_DSN8 and/or F_1-sex_DSN8 analysis.

Discussion

The ASGR and rationale for DSN treatment

The ASGR is characterized as a complex, hemizygous, dominant genetic locus that controls apospory in *Pennisetum* and *Cenchrus* species (Ozias-Akins et al. 1998; Roche et al. 1999; Goel et al. 2006). The ASGR is rich in several repetitive elements such as Opie-2-like retrotransposons (Akiyama et al. 2004; Conner et al. 2008). The ability to genetically map a trait at high resolution is crucial for map-based cloning. However, as recombination at the ASGR (as with many other apomictic loci) is heavily suppressed, map-based cloning is not a feasible strategy for identification and isolation of apomixis genes from the locus (Ozias-Akins et al. 1998; Kotani et al. 2014). Saturation of the ASGR with molecular markers and physical isolation of the BACs linked to these markers could be a potential strategy for the isolation of apomixis genes from the ASGR (Huo et al. 2009). Bioinformatic identification of contigs within the ASGR is another potential option both for the generation of markers and identification of gene content from the region.

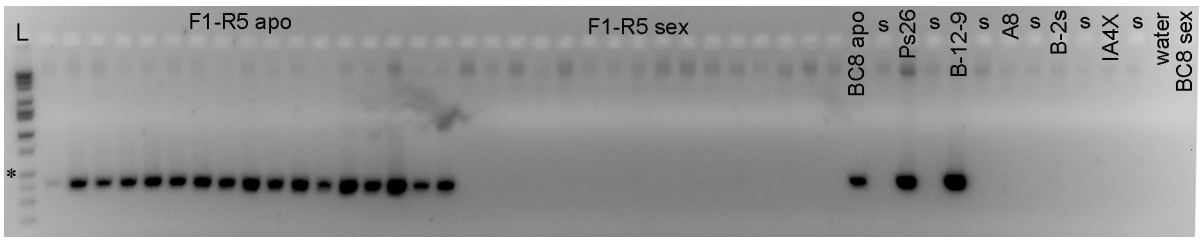

Fig. 3 Mapping of ASGR-linked DSN SCAR primers to a larger R5-F_1 population. DSN SCAR primer mapping using p2913/p2914. DNA samples from apomicts are F1-R5 apo, BC8 apo, Ps26 and B-12-9. DNA samples from sexuals are F1-R5 sex or s, B-2 s, IA4X and BC8 sex. Water is a PCR negative control. L is HI-LO DNA marker (Minnesota Molecular, Minneapolis, MN) with *asterisk* denoting 500 bps

DSN has been used as a successful strategy for normalization of complex genomes by eliminating evolutionarily young repetitive elements that contain low-complexity repeats (Shagina et al. 2010; Matvienko et al. 2013). Using reassociation kinetics of DNA during C_0t fractionation, Wimpee and Rawson (1979) estimated that the diploid pearl millet (*Pennisetum americanum*) genome consists of ~69 % repeated DNA, with much of the repeats present in long tandem arrays consisting of shorter elements varying slightly from one another in their nucleotide sequences. The high degree of sequence homogeneity of repetitive DNA of millet indicates that these sequences are either highly conserved or that they are relatively recent and have undergone little evolutionary divergence (Wimpee and Rawson 1979). Phylogenetic analysis of the *ASGR-BBML* gene in *Pennisetum*/*Cenchrus* along with FISH analysis using the low copy BAC containing this gene suggests a relatively recent origin of the ASGR in these species (Akiyama et al. 2011). These observations suggest that repeat DNA in the pearl millet genome along with repeats in the genomic region of the ASGR should be efficient targets for DSN normalization and that the use of bioinformatics to compare sequences generated from DSN-normalized Illumina libraries of apomictic and sexual plants should effectively target contigs mapping to the ASGR.

DSN normalization and enrichment of low copy sequences

We attempted to use DSN as a normalization strategy to eliminate the repetitive sequences from the genomes of BC_8 (06-A-58) apomictic and sexual plants (Singh et al. 2010) and F_1 progeny from a cross of *P. squamulatum* and *P. glaucum* (Goel et al. 2006). Analysis of count number and abundance of 31 K-mer sequences of the 4- and 8-h DSN-treated BC_8 (06-A-58) DNA from pools of sexual and apomictic plants showed enrichment of low abundance 31 K-mers, compared to the untreated control. This result is in agreement with DSN treatment of the lettuce genome where sequences with up to 43 copies were enriched rather than depleted in the normalized libraries (Matvienko et al. 2013).

Masking repeats in the BC8ApoDSN8_V and BC8ApoDSN8_S assembly using RepeatMasker maize and sorghum databases showed a similar level of repeats in both assemblies for both maize and sorghum with the highest masking at ~16 %. While the sequence of the pearl millet genome is not yet available for repeat analysis, approximately 61 % of the sorghum genome and 46 % of the foxtail millet genome are composed of repeats (Paterson et al. 2009; Zhang et al. 2012). These data suggest that DSN treatment of the BC_8 (06-A-58) apomict DNA was successful at removing highly repetitive DNA.

Reference assembly, screening and mapping of SCARs to the ASGR

In the absence of a reference genome assembly, we used the Velvet and SPAdes generated assembly of BC_8 (06-A-58) apomictic DSN8-treated DNA as the reference for Bowtie alignment of the sequenced reads. BlastN of assemblies created using different K-mers in Velvet proved to be a good approach to identify contigs unique to each K-mer assembly. A total of 24,260 contigs from the 63, 31, 39 and 47 K-mer assemblies were added to the 111,064 contigs from the 55 K-mer assembly. We verified that contigs generated from three of the different K-mer assemblies were linked to the ASGR. While the Velvet assembler was able to use ~30 % of the generated sequence for assembly, SPAdes was able to assemble ~50 % of the generated sequence. The properties of the SPAdes assembler were designed for single-cell sequencing and therefore are particularly well suited for DSN applications which need accommodations for low and variable coverage.

As the apomictic BC_8 (06-A-58) plants have one alien chromosome, the ASGR-carrier chromosome, from the apomictic donor *P. squamulatum* (Singh et al. 2010), the reference assembly can be used to screen for ASGR-carrier chromosome and ASGR sequences through Bowtie alignment of the reads from sexual BC_8 (06-A-58) and apomictic and sexual F_1 plants. With the objective of finding and mapping low copy sequences to the ASGR, we carefully screened against potential high copy sequences in the contig assembly by avoiding sequences with similarity to protein families, with similarity to multiple chromosomal regions in the sorghum reference genome or with contigs showing multiple hits to nucleotide Blast against the whole assembly.

The targeted mapping of the ASGR using the DSN-normalized Illumina sequences was successful since

89 % (39/44) of the tested SCARs showed linkage to the ASGR when mapped to apomicts in a F_1 mapping population consisting of 39 individuals (22 sexuals and 17 apomicts) from the cross between *P. squamulatum* ($2n = 56$) and *P. glaucum* (induced tetraploid, $2n = 4x = 28$). Testing of contigs displaying F_1-sex_DSN8 hits was not pursued. From previous studies, the ASGR is highly conserved between *P. squamulatum* and *C. ciliaris* (Goel et al. 2006) with ten of 12 original ASGR-linked SCAR markers from *P. squamulatum* present in *C. ciliaris*, of which six could be scored as SCAR markers present in aposporous but not in sexual *C. ciliaris* (Roche et al. 1999). Similarly, 18 DSN SCAR markers were scored as dominant markers present in aposporous but not in sexual *C. ciliaris*. Four of the 18 DSN SCAR markers were present in the *C. ciliaris* recombinant A8 plant. The A8 plant retains the ability to make aposporous embryo sacs but has lost the ability to undergo parthenogenesis (Conner et al. 2013). These newly identified A8-retained SCAR markers can be used to identify and isolate more A8-retained BACs and aid in delineation of the potential A8 recombination point and apomeiosis candidate genes in the ASGR.

Of 1372 BC8Apo_DSN8_V contigs identified as having zero BC_8 sex_DSN8 and zero F_1-sex_DSN8, 140 showed 95 % sequence identity and at least 100 bp overlap with ASGR and ASGR-recombinant partially sequenced BACs. Forty of these contigs were identified as ASGR-linked using the criterion of at least 20 F_1-apo_DSN8 hits. The remaining 100 had less than 20 F_1-apo_DSN8 hits used as a threshold with the majority showing no F_1-apo_DSN8 hits. This suggests that deeper F_1 sequencing may be necessary to identify more contigs derived from the ASGR within the assembly. Our analysis showed that the restriction digest/DSN treatment generated low copy sequences even within the "high copy" region of the

Table 2 Comparison of number of BC8Apo_DSN8_V contigs identified as ASGR linked versus ASGR-carrier chromosome linked based on bioinformatics

BAC identification	FISH signal location within the ASGR	ASGR linked	ASGR-carrier chromosome linked
c004/c014	Low copy	3	11
c018	n/a	4	4
c100/c111	Low copy	3	5
c1000	Low copy	7	13
c108	Low copy	2	4
c201	n/a	0	1
c205	n/a	0	2
c501/c522	n/a	0	0
c801	n/a	1	3
p002/p003/p004	Low copy	3	12
p1000	Low copy	4	12
p102	Low copy	1	1
p104	Low copy	2	7
p1200	Low copy	1	2
p1300[a]	Outside ASGR	1	2
p201/p207/p208	Low copy	4	9
p303	Low copy	0	5
P602	High copy	2	0
p708	n/a	1	3
p800/p801	High copy	1	3
p900	High copy	0	1

BC8Apo_DSN8_V contigs shared ≥95 % sequence identity and at least 100 bp overlap with ASGR-BAC and ASGR-recombinant BAC GSS sequences. The BAC identification and FISH signal location of the BACs within the ASGR is also shown

[a] ASGR-recombinant BAC

ASGR as indicated by the identification of BC8ApomicticDSN8 contigs with high similarity to sequences derived from the ASGR and ASGR-recombinant shotgun-cloned BACs physically mapped to the high copy regions of the ASGR (Table 2).

This research demonstrates the effectiveness of DSN treatment for enrichment of low copy sequences and the use of DSN-normalized Illumina sequences generated from apomictic and sexual BC_8 and F_1 plants to effectively assign sequences to the ASGR. Almost 89 % of the DSN SCAR primers designed for mapping showed amplification specific to the ASGR. This technique should also be applicable to work in plant genomes carrying introgressed chromosomal regions from wide crosses. However, as our analysis shows that different assemblies and assembly programs will identify both analogous and unique contigs for analysis, this technique may be more useful in generating markers for mapping versus being used to identify every gene of interest at a locus, as information may be lost due to assembly or strictly defined bioinformatics parameters.

Acknowledgments This study was supported by Pioneer Hi-Bred. We thank Gunawati Gunawan for providing technical support.

References

Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in *Pennisetum squamulatum* reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. Plant Physiol 134:1733–1741. doi:10.1104/pp.103.033969

Akiyama Y, Goel S, Conner JA, Hanna WW, Yamada-Akiyama H, Ozias-Akins P (2011) Evolution of the apomixis transmitting chromosome in *Pennisetum*. BMC Evol Biol 11:289–289. doi:10.1186/1471-2148-11-289

Albertini E, Porceddu A, Ferranti F, Reale L, Barcaccia G, Romano B, Falcinelli M (2001) Apospory and parthenogenesis may be uncoupled in *Poa pratensis*: a cytological investigation. Sex Plant Reprod 14:213–217

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc 61:51–94. doi:10.1006/bijl.1996.0118

Catanach AS, Erasmuson SK, Podivinsky E, Jordan BR, Bicknell R (2006) Deletion mapping of genetic regions associated with apomixis in Hieracium. Proc Natl Acad Sci USA 103:18650–18655

Conner JA, Goel S, Gunawan G, Cordonnier-Pratt MM, Johnson VE, Liang C, Wang H, Pratt LH, Mullet JE, DeBarry J, Yang L, Bennetzen JL, Klein PE, Ozias-Akins P (2008) Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of *Pennisetum* and *Cenchrus*. Plant Physiol 147:1396–1411. doi:10.1104/pp.108.119081

Conner JA, Gunawan G, Ozias-Akins P (2013) Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in *Cenchrus ciliaris*. Planta 238:51–63. doi:10.1007/s00425-013-1873-5

Conner JA, Mookkan M, Huo H, Chae K, Ozias-Akins P (2015) A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. PNAS 2015. doi:10.1073/pnas.1505856112 (Published ahead of print, 24 Aug 2015)

Dujardin M, Hanna WW (1989) Developing apomictic pearl millet—characterization of a BC3 plant. J Genet Breed 43:145–150

Emberton J, Ma J, Yuan Y, SanMiguel P, Bennetzen JL (2005) Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. Genome Res 15:1441–1446

Goel S, Chen Z, Conner JA, Akiyama Y, Hanna WW, Ozias-Akins P (2003) Delineation by fluorescence in situ hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in *Pennisetum squamulatum* and *Cenchrus ciliaris*. Genetics 163:1069–1082

Goel S, Chen Z, Akiyama Y, Conner JA, Basu M, Gualtieri G, Hanna WW, Ozias-Akins P (2006) Comparative physical mapping of the apospory-specific genomic region in two apomictic grasses: *Pennisetum squamulatum* and *Cenchrus ciliaris*. Genetics 173:389–400

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet 17:597–604

Gualtieri G, Conner JA, Morishige DT, Moore LD, Mullet JE, Ozias-Akins P (2006) A segment of the apospory-specific genomic region is highly microsyntenic not only between the apomicts *Pennisetum squamulatum* and Buffelgrass, but also with a rice chromosome 11 centromeric-proximal genomic region. Plant Physiol 140:963–971

Hand ML, Koltunow AM (2014) The genetic control of apomixis: asexual seed formation. Genetics 197:441–450

Huo H, Conner JA, Ozias-Akins P (2009) Genetic mapping of the apospory-specific genomic region in *Pennisetum squamulatum* using retrotransposon-based molecular markers. Theor Appl Genet 119:199–212. doi:10.1007/s00122-009-1029-y

Koltunow AM, Bicknell RA, Chaudhury AM (1995) Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiol 108(4):1345–1352

Kotani Y, Henderson ST, Suzuki G, Johnson SD, Okada T, Siddons H, Mukai Y, Koltunow AM (2014) The LOSS OF

APOMEIOSIS (LOA) locus in *Hieracium praealtum* can function independently of the associated large-scale repetitive chromosomal structure. New Phytol 201:973–981. doi:10.1111/nph.12574

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. doi:10.1186/gb-2009-10-3-r25

Leblanc O, Grimanelli D, González-de-León D, Savidan Y (1995) Detection of the apomictic mode of reproduction in maize-*Tripsacum* hybrids using maize RFLP markers. Theor Appl Genet 90:1198–1203. doi:10.1007/BF00222943

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. doi:10.1093/bioinformatics/btp352

Matvienko M, Kozik A, Froenicke L, Lavelle D, Martineau B, Perroud B, Michelmore R (2013) Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. PLoS ONE 8:1–17. doi:10.1371/journal.pone.0055913

Nogler GA (1984) Gametophytic apomixis. In: Johri BM (ed) Embryology of angiosperms. Springer, Berlin, pp 475–518

Noyes RD, Rieseberg LH (2000) Two independent loci control agamospermy (apomixis) in the triploid flowering plant *Erigeron annuus*. Genetics 155:379–390

Ozias-Akins P (2006) Apomixis: developmental characteristics and genetics. Crit Rev Plant Sci 25:199–214

Ozias-Akins P, van Dijk PJ (2007) Mendelian genetics of apomixis in plants. Ann Rev Genet 41:509–537

Ozias-Akins P, Roche D, Hanna WW (1998) Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes. Proc Natl Acad Sci USA 95:5127–5132

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, Dike S, de la Bastide M, Martienssen RA, McCombie WR (2003) Maize genome sequencing by methylation filtration. Science 302:2115–2117

Paterson A, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12:795–807

Ramakrishna W, Emberton J, Ogden M, SanMiguel P, Bennetzen JL (2002) Structural analysis of the maize Rp1 complex reveals numerous sites and unexpected mechanisms of local rearrangement. Plant Cell 14:3213–3223

Roche D, Cong P, Chen Z, Hanna WW, Gustine DL, Sherwood RT, Ozias-Akins P (1999) An apospory-specific genomic region is conserved between Buffelgrass (*Cenchrus ciliaris* L.) and Pennisetum squamulatum Fresen. Plant J 19:203–208

Roche D, Conner J, Budiman M, Frisch D, Wing R, Hanna W, Ozias-Akins P (2002) Construction of BAC libraries from two apomictic grasses to study the microcolinearity of their apospory-specific genomic regions. Theor Appl Genet 104:804–812

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

Savidan Y (2000) Apomixis: genetics and breeding. In: Janick J, Savidan Y (eds) Plant breeding reviews, vol 18. Wiley, New York, pp 13–86

Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S (2002) A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res 12(12):1935–1942

Shagina I, Bogdanova E, Mamedov IZ, Lebedev Y, Lukyanov S, Shagin D (2010) Normalization of genomic DNA using duplex-specific nuclease. Biotechniques 48:455–459. doi:10.2144/000113422

Singh M, Conner JA, Zeng YJ, Hanna WW, Johnson VE, Ozias-Akins P (2010) Characterization of apomictic BC_7 and BC_8 pearl millet: meiotic chromosome behavior and construction of an ASGR-carrier chromosome-specific library. Crop Sci 50:892–902

Smith DB, Flavell RB (1975) Characterisation of the wheat genome by renaturation kinetics. Chromosoma 50:223–242

Van Dijk PJ, Tas IC, Falque M, Bakx-Schotman T (1999) Crosses between sexual and apomictic dandelions (Taraxacum). II. The breakdown of apomixis. Heredity 83:715–721

Voigt-Zielinski ML, Piwczyński M, Sharbel TF (2012) Differential effects of polyploidy and diploidy on fitness of apomictic *Boechera*. Sex Plant Reprod 25:97–109. doi:10.1007/s00497-012-0181-8

Wimpee CF, Rawson JR (1979) Characterization of the nuclear genome of pearl millet. Biochim Biophys Acta 562:192–206

Yuan Y, SanMiguel PJ, Bennetzen JL (2002) Methylation-spanning linker libraries link gene-rich regions and identify epigenetic boundaries in *Zea mays*. Genome Res 12:1345–1349

Yuan Y, SanMiguel PJ, Bennetzen JL (2003) High-Cot sequence analysis of the maize genome. Plant J 34:249–255

Zerbino DR (2010) Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinform 11(11):5. doi:10.1002/0471250953.bi1105s31

Zhang G, Liu X, Quan Z, Cheng S, Xu X et al (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotech 30:549–554. doi:10.1038/nbt.2195

Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res 32:e37

Zhulidov PA, Bogdanova EA, Shcheglov AS, Shagina IA, Vagner LL, Khazpekov GL, Kozhemiako VV, Luk'ianov SA, Shagin DA (2005) A method for the preparation of normalized cDNA libraries enriched with full-length sequences. Bioorg Khim 31:186–194