2011

Isolation and characterisation of the first microsatellite markers for *Cyperus rotundus*

R. S. Arias
USDA-ARS, renee.arias@ars.usda.gov

W. T. Molin
USDA-ARS

J. D. Ray
USDA-ARS

M. D. Peel
USDA-ARS

B. E. Scheffler
USDA-ARS, brian.scheffler@ars.usda.gov

WEED RESEARCH An International Journal of Weed Biology, Ecology and Vegetation Management

DOI: 10.1111/j.1365-3180.2011.00861.x

Isolation and characterisation of the first microsatellite markers for *Cyperus rotundus*

R S ARIAS*, W T MOLIN†, J D RAY‡, M D PEEL§ & B E SCHEFFLER*

**Genomics and Bioinformatics Research Unit, USDA-ARS, Stoneville, MS, USA, †Southern Weed Science Research Unit, USDA-ARS, Stoneville, MS, USA, ‡Crop Genetics Research Unit, USDA-ARS, Stoneville, MS, USA, and §Forage and Range Research Unit, USDA-ARS, Logan, UT, USA*

Received 8 September 2010
Revised version accepted 8 March 2011
Subject Editor: Gavin Ash, CSU, Australia

Summary

This is the first report of microsatellite markers for *Cyperus rotundus*. A total of 191 sequence-specific microsatellite markers were isolated and used to screen 12 accessions of *C. rotundus* and one accession of *Cyperus esculentus* collected from 10 different countries. Polymorphisms were observed in 49% of the markers tested, 22% of the markers were monomorphic and 29% had weak or no amplification. The best 57 markers are reported, and cluster analysis was used to analyse their resolving power. BLASTx screening of the contig sequences was also performed. Multiallelic loci over all samples ranged from 24% to 60%. The maximum number of alleles detected by the markers suggests a polyploidy nature of all *C. rotundus* accessions tested, except for the sample N25-Brazil. Chromosome number was determined for N12-Taiwan and used as an internal flow cytometry standard to estimate the amount of DNA within haploid nuclei of the remaining material. Chromosome numbers estimated for *C. rotundus* were 16 and 24. The markers identified in this study can be used for the identification of biotypes and detection of potential crosses of *C. rotundus*, to implement management practices for the effective control of this weed.

Keywords: purple nutsedge, STR, simple sequence repeats, simple sequence repeat, molecular markers, *Cyperus esculentus*, yellow nutsedge, minisatellite, nutgrass.

ARIAS RS, MOLIN WT, RAY JD, PEEL MD & SCHEFFLER BE (2011). Isolation and characterisation of the first microsatellite markers for *Cyperus rotundus*. *Weed Research*.

Introduction

Adapted to warm environments, *Cyperus rotundus* L. (purple nutsedge) is a weed found throughout the world (Wills, 1998). In the USA, it is located mostly in the southern states, where it affects crops such as cotton, soyabean (Edenfield *et al*., 2005; Reddy & Bryson, 2009) and vegetables in general (Wang *et al*., 2008). Large morphological variations have been reported within the species of *C. rotundus* among accessions found in the USA and among other accessions from around the world (Wills, 1998). However, studies using randomly amplified polymorphic DNA (RAPD) analysis within the species have shown genetic variation in samples from around the world, but not necessarily in accessions from the continental USA (Okoli *et al*., 1997; Tayyar *et al*., 2003; Molin *et al*., 2009). A low level of genetic diversity was also observed in *Cyperus esculentus* L. analysed by amplified fragment length polymorphism, and the authors suggested that microsatellites could provide a

Correspondence: Renée S. Arias, USDA-ARS, Genomics and Bioinformatics Research Unit, 141 Experiment Station Rd., Stoneville, MS 38776, USA. Tel: (+1) 662 686 5355; Fax: (+1) 662 686 5372; E-mail: renee.arias@ars.usda.gov

more effective method to determine genetic variation (Dodet *et al*., 2008). Defining the genome in *C. rotundus* is challenging because its chromosomes, as is common in Cyperaceae, are holocentric and exhibit agmatoploidy. This is particularly displayed in *C. rotundus* for which counts of 2n = 42, 48, 50, 52, 54–56, 66, 69, 80 and 108 have been reported by the same groups (Heiser & Whitaker, 1948; Bir *et al*., 1992; Heenan & de Lange, 2005). Even so, the knowledge of genetic diversity of weeds is important for the development of chemical and biological control practices (Goolsby *et al*., 2006; Bodo Slotta, 2008).

Knowing the genetic diversity of weeds can help explain the differential resistance of biotypes to herbicides (Anderson *et al*., 2004), as well as the susceptibility of biotypes to biocontrol agents, such as fungal pathogens (Okoli *et al*., 1997) or phytophagous mites (Goolsby *et al*., 2006). The most efficient tools to make inferences about population structure and genetic relatedness in weed species are molecular markers (Anderson, 2008; Bodo Slotta, 2008). Molecular markers can help determine the origin of invasive weed species, monitor their transport to new areas (Baker *et al*., 2007; Lu *et al*., 2007), and they can facilitate reliable classification of weeds for the planning of cost-effective and sustainable control practices (Tabacchi *et al*., 2006). In general, understanding the inherent genetic traits, which govern weed plasticity and make them so competitive, can facilitate manipulating the effective range and productivity of important cultivated crop plants (Anderson, 2008).

Among molecular markers, microsatellites or simple sequence repeats (SSR) have become one of the most powerful genetic tools in biology. Microsatellites are short tandem DNA repeats, with 1- to 8-bp motifs [i.e., the microsatellite $(AGT)_5$ is composed of a 3-bp motif and has a 15-bp length]. Microsatellites are widely spread throughout eukaryotic genomes (Richard *et al*., 2008) and their use as molecular markers is highly desirable because of their abundance, high polymorphism and co-dominant inheritance (Weber, 1990; Varshney *et al*., 2005). Given the advantages of sensitivity, transferability, reproducibility and low cost of microsatellites over other molecular markers (Varshney *et al*., 2005; Sharma *et al*., 2008), we isolated and characterised microsatellite-enriched libraries for *C. rotundus* and tested their efficacy on a subset of accessions that had previously been evaluated with RAPDs (Molin *et al*., 2009). We did additional cytogenetic characterisation on these accessions to explain the allele distribution found by the markers. The objective of this work was to develop robust molecular tools that will allow the study of *C. rotundus* at multiple loci and analysis of the genetic diversity of this weed in accessions from different geographical areas.

Materials and methods

Construction of C. rotundus *SSR-enriched libraries and primer design*

Fresh leaves of *C. rotundus*, accession number N37-Mississippi-2, from the nutsedge collection maintained at the ARS facility in Stoneville, MS (Dr William Molin, curator), were processed with DNeasy Plant Maxi kit (Qiagen, Valencia, CA, USA) for DNA extraction. Twenty micrograms of DNA was used to generate SSR-enriched libraries following the protocol of Techen *et al*. (2010) and briefly described here. DNA from *C. rotundus* was digested with restriction enzymes *Alu*I, *Hae*III, *Dra*I and *Rsa*I (New England Biolabs, Ipswich, MA, USA) individually and in pairs of these enzymes (*Alu*I + *Rsa*I, *Dra*I + *Hae*III). The restriction-digested DNA was pooled and then separated by agarose gel electrophoresis; fragments between 300 and 2000 bp were purified as indicated by Techen *et al*. (2010).

The blunt-end DNA fragments were A-tailed with Taq DNA Polymerase (Promega, Madison, WI, USA) in the presence of dATP for 2 h, then ligated for 3 h at 16°C to the adapter SSRLIB3 (Techen *et al*., 2010), made from oligos SSRLIBF3: 5′- CGGGAGAGCAAG-GAAGGAGT-3′ and SSRLIBR3 5′Phos-TCCTTCCT-TGCTCTCTCCCGAAAA-3′. The ligated fragments were purified with MinElute (Qiagen) and amplified by 20 cycles of PCR using primer SSRLIBF3 and high-fidelity DNA Polymerase (Invitrogen, Carlsbad, CA, USA) at 94°C for 2 min and cycles of 94°C for 30 s, 60°C for 30 s and 68°C for 90 s. The amplified products were hybridised to four groups of biotinylated oligo (Invitrogen) repeats as indicated by Glenn and Schable (2005): group 1 [(AC)13, (AACC)5, (AACG)5, (AAGC)5, (AAGG)5, (ATCC)5], group 2 [(AG)12, (AAC)6, (AAG)8, (ACT)12, (ATC)8], group 3 [(AAAC)6, (AAAG)6, (AATC)6, (AATG)6, (ACAG)6, (ACCT)6, (ACTC)6, (ACTG)6] and group 4 [(AAAT)8, (AACT)8, (AAGT)8, (ACAT)8, (AGAT)8]. The final concentration of each oligo in the mix was 1 μM, and 2 μL of each oligo mix was used in 50-μL hybridisation reactions.

Hybridisations were performed in a gradient thermocycler at 95°C for 10 min, followed by 3 h at the annealing temperature using a gradient block (Group 1: 56°C, Groups 2 and 4: 50°C and Group 3: 53°C) and an extension step of 10 min at 68°C in the presence of high-fidelity Taq Polymerase (Invitrogen) as indicated in Hayden *et al*. (2002). Sequences containing repeats were captured using streptavidin-coated magnetic beads M-270 (Invitrogen) in a Labquake tube shaker/rotator (Barnstead/Thermoline, Dubuque, IA, USA) at 22°C for 1 h (Kijas *et al*., 1994). After binding, the beads were

washed with 2xSSC (sodium citrate buffer), 1xSSC at ambient temperature and 0.5xSSC at 50°C for 5 min each. Elution of the DNA from the biotinylated oligos was carried out with 100 μL MilliQ water at 96°C for 10 min twice. Please see Techen *et al*. (2010) for a detailed protocol. The eluate was PCR amplified for 20 cycles as indicated for the ligation step; the PCR products were cloned in vector TOPO4 (Invitrogen) transformed into *E. coli*, plated on x-gal and white colonies selected for sequencing. Sequencing was performed using Big dye terminator and analysed in an ABI 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Sequences were assembled in contigs using DNAStar Lasergene7 (DNASTAR, Madison, WI, USA) and visually checked. Singletons were not used for primer designing. Contigs are a length of contiguous sequences assembled from partial overlapping sequences generated from the PCR products. Repeats were searched using SSRFinder (Sharopova *et al*., 2002) and Sputnik (C. Abajian, http://espressosoftware.com/sputnik/index.html, accessed April 2010) using default conditions. Primers were designed using Primer3 (Rozen & Skaletsky, 2000) with stringent parameter conditions: Tm 63°C optimum (60/65) min/max, length 24 optimum (20/28) min/max and 3′ GC clamp.

Fingerprinting and cluster analysis

Using stringent conditions in Primer3 software, we designed 191 primers on the flanking regions of the repeats and tested 13 accessions of *Cyperus* (12 *C. rotundus* and one *C. esculentus*). The 12 accessions of *C. rotundus* were N02-Mississippi-1; N03-California; N12-Taiwan; N13-Sudan; N20-Indonesia; N21-Greece; N22-El Salvador; N25-Brazil; N26-Australia; N36-Arkansas; N37-Mississippi-2; and N41-Thailand. These accessions are clonally propagated. Leaves of several whorls per accession were freeze-dried and pulverised for DNA extractions and screening of microsatellites. These accessions are maintained by Dr. William Molin at SWSRU-USDA, Stoneville. Forward SSR primers were 5′ tailed with the sequence 5′-CAGTTTTCCCAGTCACGAC-3′ to permit product labelling (Waldbieser *et al*., 2003), and reverse primers were tailed at the 5′end with the sequence 5′-GTTT-3′ to promote non-template adenylation (Brownstein *et al*., 1996). Primer 5′-CAGTTTTCCCAGTCACGAC-3′ labelled with 6-carboxy-fluorescein (IDT-Technologies, Coralville, IA, USA) (FAM) was used for the amplification of 10-ng DNA using Titanium Taq DNA Polymerase (Clontech, Mountain View, CA, USA) according to manufacturer instructions in 5-μL reactions on an MJ Research Thermocycler (BioRad, Hercules, CA, USA) at 95°C for 1 min, 60°C for 1 min (2 cycles), 95°C for 30 s, 60°C for 30 s, 68°C for 30 s (27 cycles) and a final extension at 68°C for 4 min. Fluorescently labelled PCR fragments were analysed on an ABI 3730XL DNA Analyzer using ROX for allele size standard and data processed using GeneMapper v. 3.7 (both from Applied Biosystems). Presence of alleles was converted to a binary matrix. *Cyperus* accessions were clustered based on SSR markers using genetic distances according to Nei (1972) and the unweighted paired group method and arithmetic averages algorithm implemented in the SAHN program of NTSYSpc v. 2.2 (Exeter Software, Setauket, NY, USA). The confidence levels for the dendrograms were assessed by bootstrap resampling (5000 replicates) (Felsenstein, 1985; Efron *et al*., 1996) using Winboot (downloaded from http://www.irri.org/science/software/winboot.asp accessed August 2010).

BLASTx of the DNA sequences

Contig sequences containing microsatellites were screened against the National Center for Biotechnology Information (NCBI) protein databases (BLASTx) (Altschul *et al*., 1990).

Multiallelic loci and UPIC scores

The percentage of multiallelic loci was calculated for each accession across all SSR markers tested. We also calculated Unique Pattern Informative Combinations (UPIC) to determine the discriminatory power of the SSR markers and the minimum set of markers that could discriminate each and every one of the samples. Both percentage of multiallelic loci and UPIC scores were calculated using UPIC Perl scripts (Arias *et al*., 2009).

Cytogenetic analysis

Somatic cell chromosome numbers from N12-Taiwan were determined from root tips of clonally propagated plants grown in a glasshouse in Logan, Utah. Prior to harvesting root tips, the plants were subject to mild water stress to promote root growth and allowed to grow for 2–6 h. Twenty root tip preparations were made, and hundreds of cells were scanned to find few with discernable chromosomes. The harvested root tips were treated in an aqueous solution containing 0.05% colchicine plus 0.025% 8-hydroxyquinoline and 2% dimethylsulphoxide (DMSO) for 3 h. They were then fixed and stained in a 2% aceto-orcein at 5°C for 3 days. The meristematic portion of the root was then squashed in 45% acetic acid for examination with a microscope.

A Partec PA II flow cytometer (Partec, Münster, Germany) with DAPI technology was used to determine the ploidy level of the samples using Partec's CyStain UV Precise P kit and protocol. The Partec was equipped with a mercury lamp with UV light 360-nm excitation and 420-nm admittance filter. The average number of cells counted was 544 per accession, with a range from 392 to 953 cells per accession. In the initial testing phase to calibrate/set the gain on the ploidy analyser, we chose sample N25-Brazil, which had the lowest heterozygosity and the lowest number of alleles per locus in the microsatellite fingerprinting. To match the material being tested, N25-Brazil was found to reliably and consistently produce peaks at 2/3 the level of N12-Taiwan, and N12-Taiwan consistently produced peaks at 1.5, the level of N25-Brazil. Sample N20-Indonesia was not included in cytogenetic analysis. Leaf blade tissue from plants growing in the glasshouse was utilised for the flow cytometry with 1000–1200 nuclei reads from each plant sample for the final comparison. N12-Taiwan was used as an internal control for all other material tested. If one peak was detected, then the chromosome number of that accession was taken to be equal to N12-Taiwan. Conversely, if two peaks appeared, one for N12-Taiwan and another at two-thirds the gain of N12-Taiwan, its chromosome number was recorded as equal to two-thirds of N12-Taiwan.

Results

We constructed SSR-enriched libraries from *C. rotundus*, sequenced 2592 clones of those libraries and assembled the sequences into 780 contigs. The number of repeats detected in the contigs by SSRFinder and Sputnik were 1073 and 772 respectively, and from those repeats, 191 sets of primers were designed using stringent conditions. All primers had annealing temperature 63 ± 1°C, runs ≤5 bp, 3′ GC clamp and length 22 ± 2 bp. A total of 527 contigs containing repeats were submitted to GenBank with accession numbers GQ872539 to GQ873066. To simplify the recording of repeat motifs, those repeats that were circular permutations and reverse complements of each other were grouped together as one type, i.e. AAC, ACA, CAA, GTT, TGT and TTG were recorded as AAC. The most abundant repeat motifs found in *C. rotundus* microsatellites were AG, AC and AAC (Table 1). In addition, repeat motifs of 9 bp or longer, normally considered minisatellites, were also detected in *C. rotundus*; both micro- and minisatellites with their corresponding frequencies are listed in Table 1.

Of the 191 markers tested, 49% were polymorphic, 22% were monomorphic and 29% of the markers had weak or no amplification. We report the 57 best SSR markers, based on the quality of the electropherograms, for nutsedge with the amplicon sizes observed across

Table 1 Microsatellite and minisatellite motifs isolated from *Cyperus* spp. SSR-enriched genomic libraries. The number of base pairs (bp) in the motifs is indicated at the top of the columns; frequency of each motif is in parenthesis

Microsatellites – Motif length					
2	3	4	5	6 & 7	8 bp
AG (445)	AAC (42)	ACAG (29)	AACCT (9)	AACATC (2)	ACATACAC (2)
AC (258)	AAG (35)	AACT (15)	AGAGG (2)	AGTGAG (1)	ATACACCC (1)
AT (18)	ACT (32)	AAAG (10)	AAAAG (1)	ACTACA (1)	ACTGGGAG (1)
CG (3)	ATG (22)	ATCT (9)	AAAAT (1)	ACCAGT (1)	GTCAGTCA (1)
	ACC (15)	ATAC (7)	AAGGG (1)	AGGGAC (1)	ACGGACAG (1)
	ATC (15)	ATGT (6)	AATGG (1)	ACTGCTG (1)	TCACTAAC (1)
	AGC (8)	AGGG (5)	AGGAG (1)	ACTCTC (1)	
	ACG (7)	AATG (4)	AGGGG (1)		
	AGG (7)	ACTG (4)	ATGGG (1)		

Minisatellites – Motif length	
9, 10, 11	12, 14, 20, 62 bp
CATCAACAC (1)	CCATACATCCAC (1)
CATCATCAA (1)	TCTTCGTTTTCT (1)
AACCACACCT (1)	AGGGACAAGCCGGG (1)
ACCTAACCAC (1)	GGAAAGTGATCAGATGGCTT (1)
ATTGTCAGTT (1)	
CACATACACA (1)	CTACAACCAACTTTTAAACCATTTCAACACACTACAACCTCATTTCAACACACAACACCACA (1)
TAACCACACC (1)	
AACCATCAAAC (1)	

SSR, simple sequence repeats.

accessions (Table 2). Numbers in the column of marker names correspond to the contig numbers submitted to GenBank. The top 40 markers reported in Table 2 amplified all 13 nutsedge accessions from around the world, whereas the other 17 markers listed amplified all accessions tested from continental USA. Another 79 markers (not listed), 75 of which were polymorphic, amplified almost exclusively accession N37-Mississippi_2, from which the SSR library was generated, and accession N36-Arkansas plus a few other accessions, including N41-Thailand and N26-Australia. To determine the resolving power of the markers for geographically distant accessions, the top 40 SSR markers that amplified all 13 *Cyperus* accessions were used in cluster analysis (Table 2, Fig. 1). Establishing an arbitrary threshold of 0.55 of genetic distance, we identified four clades (I, II, III and IV; Fig. 1). Clade I grouped all *C. rotundus* accessions from USA along with accessions from Australia, El Salvador, Thailand and Taiwan. In this clade (I), all accessions had few unique alleles (meaning alleles were not present for any other accession). Clades II, III and IV, on the contrary, had between 5 and 12 unique alleles per accession. The percentage of multiallelic loci for the 13 accessions ranged from 24% to 60%, with the lowest in N25-Brazil and the highest in N37-Mississippi-2 (Fig. 1). The maximum number of alleles per locus was 2 for N25-Brazil and between 3 and 6 for the rest of the accessions. In general, the number of alleles detected by each of these 40 markers across all the accessions was between 2 and 11.

BLASTx results of the contig sequences corresponding to the SSR markers had significant hits for 6 of the sequences (expected values 1.00E-51 to 1.0E-06); markers isolated from those sequences are indicated in bold face (Table 2). Markers StvCyR_197a and StvCyR_64a had significant hits in BLASTx corresponding to a phosphofructokinase (PFK) and a putative ATPase respectively. These two markers showed identical alleles across most of the samples in clade I and were polymorphic for the rest of the samples (Fig. 1). Among the markers that amplified samples from continental USA, StvCyR_260, StvCyR_391a and StvCyR_452a had significant hits on BLASTx, corresponding to a retrotransposon protein, DNAJ heat shock protein and a WD-repeat protein respectively.

Markers with UPIC scores different from zero are listed in Table 2. The number of samples uniquely discriminated by each marker is the UPIC score. Using UPIC software, we identified a minimum of four markers that were able to discriminate all 13 *Cyperus* accessions, StvCyR_45a, StvCyR_483a, StvCyR_116c and StvCyR_197a, and when combined could detect a total of 27 unique patterns or alleles (Table 2).

Chromosome number based on root tip counts showed that N12-Taiwan contained 2n = 24 chromosomes. Results from flow cytometry showed that 9 of the 12 accessions (N20-Indonesia was not analysed) were indistinguishable from N12-Taiwan, and two of the 12 produced peaks at two-thirds that of N12-Taiwan indicating a chromosome number of 2n = 16. These two accessions were N25-Brazil and *C. esculentus*. The remaining 10 *C. rotundus* genotypes, peaks indistinguishable from N12-Taiwan, have chromosome numbers of 2n = 24 (Table 3).

Discussion

This is the first report of microsatellite markers isolated from *C. rotundus*. No nucleotide sequences of *C. rotundus* were available in GenBank at the time this work started. Thus, data mining of existing DNA sequences could not be used to search for microsatellites. It is known that the frequency and distribution of microsatellites vary depending on the organism (Katti *et al.*, 2001) and also that certain motif lengths can be easier to score than others (Kumar *et al.*, 2002). Therefore, we generated SSR-enriched libraries using a variety of oligo repeats to increase the possibility of finding useful markers. Forty of these markers had transferability to the sample of *C. esculentus* tested, and five of them showed distinct alleles for this species. Additional samples of *C. esculentus* will need to be screened to determine whether these alleles are species specific.

Morphological variations have been documented for *C. rotundus* and *C. esculentus*, in some cases describing biotypes with overlapping characteristics, probably from natural hybridisation between species (Tayyar *et al.*, 2003). The large number of microsatellite markers provided here facilitates the testing of hypothetical hybridisations in *Cyperus* spp. Studies of *C. rotundus* using RAPDs have shown genetic variation among accessions from around the world, but little or no variation among accessions from the continental USA, suggesting that this species probably forms regional clones (Okoli *et al.*, 1997; Tayyar *et al.*, 2003; Molin *et al.*, 2009). In the present work, however, we report 47 microsatellite markers that detected genetic diversity among *C. rotundus* within the continental USA when using the same samples previously tested by RAPDs. In addition, the fact that 42% of the markers only amplified the accession from which they were isolated and few other samples suggests that there is a large genetic diversity in *C. rotundus*, even among accessions collected a few hundred miles apart. We attribute the lack of amplification on the rest of the accessions to an unusual abundance of 'null' alleles. The SSR markers reported herein provide better tools for the assessment

Table 2 Simple sequence repeats markers for *Cyperus* spp. Size: number of base pair (bp) in the amplicons. N Al.: number of alleles observed for a marker. Max Al./S: maximum number of alleles observed on individual samplesMarkers in bold font had significant hits on BLASTx (Expect value: 5.00E-51 to 1.00E-06). The top 40 markers listed amplified all samples tested. UPIC scores represent the number of accessions uniquely identified by each marker calculated by UPIC software (Arias *et al.*, 2009). Motif: correspond to the motif of the repeat that gave origin to the marker. C.e.: markers that distinguished *Cyperus esculentus* from the rest of the samples shown by a 'Y' symbol

Marker	5′ → 3′ Forward	5′ → 3′ Reverse	Size	UPIC score	Motif	N Al.	Max Al./S	C.e.
StvCyR_1a	GCATTTCGTCACCTTCCATTAAAC	TTACTTTGTTTGCAGTTGCAGAGG	193	0	AG	1	1	
StvCyR_27b	GAGTGAGGGAGTGAGAGAGGGAC	TATAGCAAAGTCAGCAGCGCAC	95–444	0	TGG	8	3	
StvCyR_32skd	CCACAAAAAACAACCCATATCATC	TAAGAGAGGAAGGAAAAAAAGGGG	176	0	ATATC	1	1	
StvCyR_45a	AAGGGAAGGTTCAAAGCTAAATGC	AATGAAATTTGAAACTCAGGCTCG	148–537	12	CT	11	4	
StvCyR_52a	CAGCCCGTATGGTCCCTAATCTAC	AATTTCACTATTCATGACGCCACC	104–285	0	ACAG	7	6	
StvCyR_64a	GTACAAACCACAAACCGTAGACCC	ACTCTCCTCCTCCATCGTAAGCTC	114–138	5	GA	7	4	
StvCyR_93ska	GCAGATATGCCTCTTCAGAGTTCAG	CAGGAGGACATTGTTGTAAGAGGG	164–174	4	AATGG	5	4	Y
StvCyR_95a	AACAGAAGATCTGCAAAAGATGCC	GATTGCTTCAATAAATTTGGCGAC	94–206	1	AAAC	3	2	
StvCyR_104a	GCTAGCATGTAACCCCTGTCTCTTAG	GTCTGTCTCCCTCCCTCTTCTCTC	114–129	6	AT	6	5	
StvCyR_116c	TTTATTGGATTTTTGTGGGGACTG	AACACTGTAAGAGGCTGCTATGGG	119–451	5	AG	6	3	
StvCyR_126a	GGCCTCCGTAAGAAGAAGAATGAC	AAGTTCAAGGCAGACGTTAAGCAC	165–166	0	GA	2	2	
StvCyR_142a	CACGGTAAATTAAACATCACACGG	TTGATCTAACACTTGTACTGCGCC	112	0	GA	1	1	
StvCyR_156a	TGCCCAGCTTTACATCTTAATTGC	CTGAACAACTGGCACATACAGAGC	182–191	1	AT	2	2	
StvCyR_181a	TCAATAGAAGAATCCCACTCAGCC	GTGGAGGTAAAGATCAGCAACCAG	135–157	5	GATA	5	2	
StvCyR_197a	TCGTGAAACATGGATACAATCAGG	GACCTGACCTGACCCAAAACC	219–251	7	AT	9	4	
StvCyR_218a	TCAAAAATCAAAACCAATCCATCC	GAGTTGGGAGAGGGAGTAGGAAAG	133–145	4	CT	4	3	
StvCyR_218b	CTTTCCTACTCCCTCTCCCAACTC	CGGGGGGATTACCTACCTACC	137–157	3	AGG	4	3	
StvCyR_254a	AATCATGAAGGTGATTGGACAAGG	CATCCATCCTCTTCTTTGTTCTCG	104–149	3	TA	2	2	
StvCyR_254c	TCTGTCTCTCATAAGCTCCTGCTC	TCAAATGTTTTTCAAGGAAGTTGC	99	0	ATGA	1	1	
StvCyR_327a	CATTAGACTTCGCTCTCATCTCTGG	GATTTATAGAGGGGAGGGGAGGAG	120–153	4	CT	8	2	
StvCyR_327ska	CATTAGACTTCGCTCTCATCTCTGG	AGGGGAGGAGTGGGATTTATAGAG	146–166	3	AGGGG	5	2	Y
StvCyR_327skb	TCCCCTCTATAAATCCCACTCCTC	TGATTCGTCATTGTCTGATGATCC	152	0	ACGG	1	1	
StvCyR_336a	TTCACCGGCTACTTTCACTAATCAC	GTGAAGATCCATATTTGGACGGAG	133–331	4	AAC	8	3	
StvCyR_344a	GTCTCCAAATCAGAGAAATGTGCC	GACTAACAAATAGATGGGAGGGGG	367–421	4	CT	5	3	
StvCyR_351a	GTCATTTGCAATGTCACTACCACC	GCATGTCTTTGTGAGCGTAGTCAG	90–169	3	AG	4	2	
StvCyR_376skb	CCTCTTCCTCCTGCTCCACC	TCGACGATCTTGAAGAGGATGAC	117–249	3	ATC	6	6	
StvCyR_437ska	CTTTCAGGTCAGGCCACTCAC	CGTGACAGGATCGGAGATTTG	112–154	2	ACAG	3	3	
StvCyR_444a	TTCTGAGTTGACTCGGAGGATTTC	TAGTTTTTTAACTCGGCTTCACGC	106–107	0	CT	2	2	
StvCyR_476a	CATGATAGTGTAGCCAAACGCAAC	AAAGGATAGTTTTGATTCACCGGC	83–363	5	TC	6	4	
StvCyR_483a	TGTGTTGTGTGGAAGAGAGGAGAG	AACAAACCTCAGAACTCAACTGCC	117–455	7	GA	8	5	
StvCyR_494ska	GACTATCCGTGAAGATATGTTTGGC	CAGTTGGCTACCTTGACCCTTG	128–356	4	AACTT	7	3	Y
StvCyR_497a	AGAAAGGAGTTCGGTCTAAATCGTC	AGTGGCAGGCAGAGACGAAG	113–462	0	CT	8	5	
StvCyR_551a	GAAATTTGGAGGTGATTCGATATG	AGTTTATGAGGAAGCAGAGAAACATC	230–282	6	AT	6	3	
StvCyR_566a	GTTGCTGGTACGGCTCTAGAAGTG	AGAGGAGAGGAGAGGAGAGTGTGG	113	1	AC	1	1	
StvCyR_572a	GACAAAGTGTGAGCATTTTTGGAC	TCTGTGGTTGTTATCCTCCTCCTC	180–360	2	GAAT	3	2	
StvCyR_577a	TTTGGTTACTTTGGTTTCAAGATAGAAG	ACCACAAATTCATGAGGTCTTCAG	171–176	4	TC	5	2	
StvCyR_590a	TCTGAGGGACTAAATGCTAATGTTTTAC	TCAGAAAAGATAACGCGTAGCAG	170–296	5	CTT	9	3	Y

Table 2 (*Continued*)

Marker	5′ → 3′ Forward	5′ → 3′ Reverse	Size	UPIC score	Motif	N Al.	Max Al./S	C.e.
StvCyR_604a	AAAACCCTTGAAGAAACCAAAACC	ACCTAGAGAGTCAGCTTCAGCACC	154–158	3	CT	4	3	Y
StvCyR_704ska	TTTTGGGTATTAACTGCTACGCTC	TGATTTTAATTATTGCTTACGACAACC	151–152	0	AAAT	2	1	
StvCyR_741a	GCTTCTTAACCTGACATACCTGACTG	TGATCTGGTGTGACTGCATGTTAG	99–132	2	AC	2	2	
Markers that amplified nutsedge accessions from continental USA and had 2–4 null alleles on other samples:								
StvCyR_3_a	TTTAAAAAATCCCCAATTCCCTTC	AATTGTGGATGAGGGTTGGTGTAG	132–189	1	TTC			
StvCyR_12_a	TAAGAAATAACAATTTCAGCCCGC	TGATCGATGACATAGAGTGAGAGGAG	148	0	TA			
StvCyR_53_a	GTTTGAGTTTAAAGGCGCTAGTGG	CATCCATTATCACCTCCTACCACC	134–379	4	GA			
StvCyR_63_a	ATCCAACTCCCTGGAATAATTTGC	CAACCTTCATGGATGCCTTTCTAC	116–349	2	ATGT			
StvCyR_161_b	CATTTGCCTCCGGTTTTTATGTAG	TCTAAGACACAGAACGGAATGACG	155–231	2	GA			
StvCyR_169_a	GGATATTAGTGATCTCCACCGGG	TCATCACTGCTTAAGTCCCATCAC	168–337	2	TGA			
StvCyR_185_a	ATCATCATGTTTGCAACTGATCCC	GTCAAGAGATGGTGAGAGGACGAG	104–105	0	CT			
StvCyR_244_b	ACCCTCTCTCGTTTCTCTTTCCTC	ATTATGGGTTGAGGTTGAAGTTGG	104–407	5	TC			
StvCyR_258_a	AACTGTTAAGAGGAGGGGTTGGTG	GGAGCCCACATTCTTTTACAGTCTC	161–172	2	GAAA			
StvCyR_260_a	ACCACACCACACTCATAGCAAGTG	CAATAGCAACAACAACTTCAACCG	320	0	TGT			
StvCyR_303_a	AGAATCGGTGCTAGAGAGAAGACG	CACTGGATTAACCGACTTCTGGAG	165–248	0	GA			
StvCyR_358_a	CAAAGAAATCCACAATCATGTTGAC	GTAGGTTCTAAATGTCATGCTGCC	202–478	6	GA			
StvCyR_391_a	ACGCGCAGTTGCCTTAAAAATAC	TTCTTTTTAGATGTTCCCGCACTC	222–395	3	AACT			
StvCyR_393_a	GTTTGTGAACAAGCACTGAAATGG	TTGTGCTACGAGTGAGTCTATGACAG	113–121	2	ATAC			
StvCyR_452_a	GATTCCTTCTCATTCTGGTGATCG	CCACAACTCACATTGGATTGCTC	124–262	3	TC			
StvCyR_509a	TTATTTCATTCGACAGACGAGCAG	GTTTAAACGATTGCATTCGAGGTC	153–325	6	GCGA			
StvCyR_713_a	TGCTGTAATAAAACATTGCCCAAC	CTTCTCCATTTCTGTCTCTCGCTC	189	0	GA			

UPIC, unique pattern informative combinations.

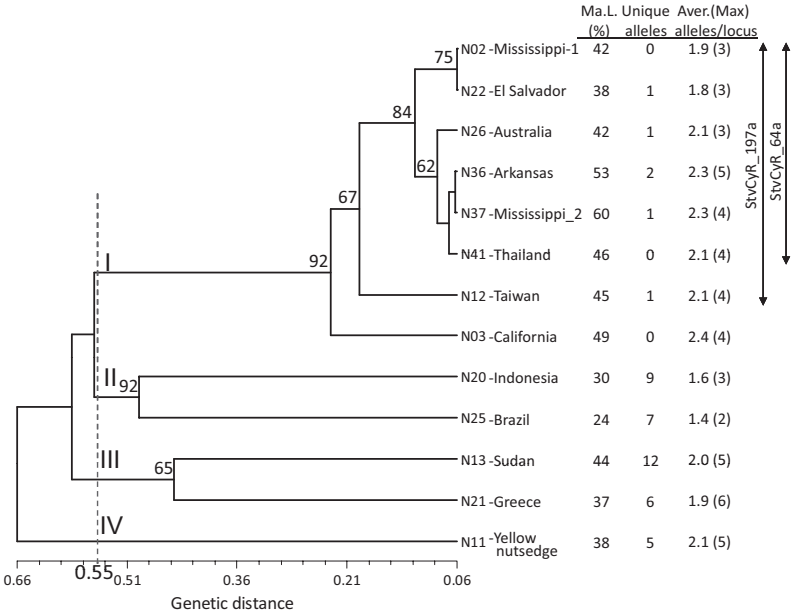

Fig. 1 Cluster analysis of 12 *Cyperus rotundus* and one *Cyperus esculentus* using 40 simple sequence repeats markers that amplified across all accessions. I, II, III and IV: clades separated at an arbitrary genetic distance of 0.55. Ma.L.: percentage of multiallelic loci calculated for each accession. Unique alleles: alleles not present in any other sample. Aver. (Max): average number of alleles per locus, in parenthesis is the maximum number of alleles detected in the accession. Bootstrap coefficients after 5000 bootstrap resampling are shown at the nodes, only values higher than 50 are represented. Continuous lines under StvCyR_197a and StvCyR_64a represent the group of samples for which these microsatellite markers had identical alleles.

Table 3 Chromosome number of 12 *Cyperus rotundus* and one *Cyperus esculentus* accessions based on results from flow cytometry

Plant	Peak number	Calculated ratio (N12-Taiwan)	Chromosome number
N02_Mississippi_1	1	1.00	2n = 24
N37-Mississippi_2	1	1.00	2n = 24
N22-El Salvador	1	1.00	2n = 24
N26-Australia	1	1.00	2n = 24
N36-Arkansas	1	1.00	2n = 24
N41-Thailand	1	1.00	2n = 24
N12-Taiwan*	1	Standard	2n = 24
N03-California	1	1.00	2n = 24
N13-Sudan	1	1.00	2n = 24
N21-Greece	1	1.00	2n = 24
N25-Brazil	2	0.67	2n = 16
N11-Yellow Nutsedge	2	0.68	2n = 16

*Used as internal standard, 2n = 24. Chromosome number based on somatic cells from root tips.

of genetic variability in this species, as microsatellites have several advantages over RAPDs as markers, including their reproducibility, transferability and co-dominant inheritance (Varshney *et al*., 2005; Sharma *et al*., 2008).

An interesting observation among the 12 accessions of *C. rotundus* studied here was the range in the percentage of multiallelic loci of the samples, ranging from 24% (N25-Brazil) to 60% (N37-Mississippi_2). To determine whether that variation was the result of different ploidy of the samples, we performed cytogenetic studies of the accessions. The unique distinction of *C. rotundus* samples from Brazil compared with the rest of the world had been previously observed by Okoli *et al*. (1997). The low percentage of multiallelic loci in N25-Brazil can be explained in part by the lower chromosome count (2n = 16) we observed in that accession. The number of chromosomes we observed in the *C. rotundus* samples (2n = 16, 24) is lower than observations of Bir *et al*. (1992) for *C. rotundus*. There are a few definitive reports of exact chromosome number of *C. esculentus* with Bennet and Smith (1991) reporting 128 and Heiser and Whitaker (1948) reporting 108, although the base number of chromosomes generally accepted for this genus is 8 (Bir *et al*., 1992). The overall heterozygosity of genomes is in general correlated with the levels observed for SSR markers (Aparicio *et al*., 2007), and high heterozygosity is considered a potential for adaptation to the environment (Hansson & Westerberg, 2002). Thus, with the availability of the microsatellite markers reported here, the hypothesis of correlation between multiallelic loci and adaptation to broad environmental conditions can now be tested in *C. rotundus*. In the family Cyperaceae, karyotype evolution has mainly resulted from agmatoploidy of holocentric chromosomes (Davies, 1956); thus, the ploidy status of *C. rotundus* is not clear. This precludes the classical use of tests for deviation of Hardy–Weinberg and linkage disequilibrium normally required to describe microsatellites.

The cluster analysis was performed to evaluate the resolving power of the markers and not to make phylogenetic inferences. However, several observations seem to indicate that the accessions in clade I do, indeed, have a greater genetic distance from the accessions in clades II and III. These observations are the presence of more abundant unique alleles in clades II and III compared with clade I in Fig. 1, the large number of 'null' alleles on 79 markers, which showed amplification

almost exclusively in the accession from where they were isolated, and the high bootstrap coefficients at the nodes of clades I, II and III. A larger number of accessions will need to be analysed to determine the genetic relatedness among accessions from various geographical areas.

Among the markers that amplified all accessions tested, and whose sequences had significant hits on BLASTx, two markers (StvCyr_197a and StvCyR_64a) were particularly interesting. The DNA sequence for StvCyr_197a had homology to a PFK, and StvCyR_64a had homology to a rice protein, a putative ATPase, indicating both could be involved in energy metabolism. PFK enzymes are known to be involved in anoxia tolerance in plants (Huang *et al*., 2008), and recently, ecotypes of *C. rotundus* have been described with adaptation to flooding, becoming a problem in lowland rice crops (Peña-Fronteras *et al*., 2009). In the cluster analysis, both markers mentioned were homogeneous in accessions within clade I, which contained *C. rotundus* accessions from USA, while being polymorphic in the rest of the accessions. It would be interesting to test any potential association of these two markers with flood adaptation.

A useful method to determine the utility of individual microsatellite markers is to calculate their UPIC score for a given set of DNA samples. UPIC scores indicate the number of accessions a marker can discriminate (Arias *et al*., 2009). We list the 31 markers that amplified all accessions and had UPIC scores ≥1, meaning that the markers could discriminate at least one accession. From this list, the minimum number of markers that discriminated all 13 *Cyperus* accessions was a combination of the four markers indicated in bold face in Table 2. The use of UPIC scores allows maximising the use of resources in future experiments, by choosing those that detect the maximum genetic variability.

The availability of the microsatellite markers reported here could benefit in the identification and categorisation of biotypes. Any work on biological control agents, such as the use of the phytopathogenic fungus *Puccinia canaliculata* (Schw.) for the control of *C. rotundus* (Okoli *et al*., 1997) or the search for herbivores to control this weed, may be biotype specific. For example, it has been observed that phytophagous mites used for the biocontrol of *Lygodium microphyllum* [(Cav.) R. Br.] were better able to feed on certain biotypes of the host plant than on others (Goolsby *et al*., 2006). Thus, if biotypes can be identified by these molecular markers, this could facilitate a more effective use of biocontrol agents. In addition, programmes for the development or testing of herbicide efficacy could use these SSR markers to identify biotypes with potentially different levels of response. For example, in *Aegilops cylindrica* Host (jointed goatgrass), direct association of microsatellites and herbicide resistance has been reported (Anderson *et al*., 2004). Furthermore, microsatellites can assist in tracing the movement and origin of invasive species (Baker *et al*., 2007; Okada *et al*., 2007). In addition, the SSRs developed here may also be useful as a taxonomic tool in determining the relationship among genera of sedges.

Acknowledgements

This work was supported by USDA-ARS project number 6402-21310-003-00. We thank Ms. Xiaofen (Fanny) Liu for library sequencing and Ms. Sheron A. Simpson for testing the SSR markers. No conflicts of interest have been declared.

References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–410.

Anderson JV (2008) Emerging technologies: an opportunity for weed biology research. *Weed Science* **56,** 281–282.

Anderson JA, Matthiesen L & Hegstad J (2004) Resistance to an imidazolinone herbicide is conferred by a gene on chromosome 6DL in the wheat line cv. 9804. *Weed Science* **52,** 83–90.

Aparicio JM, Ortego J & Cordero PJ (2007) Can a simple algebraic analysis predict markers-genome heterozygosity correlations? *Journal of Heredity* **98,** 93–96.

Arias RS, Ballard LE & Scheffler BE (2009) UPIC: perl scripts to determine the number of SSR markers to run. *Bioinformation* **3,** 353–361.

Baker J, Hidayat I & Preston C (2007) Molecular tools for understanding distribution and spread of weed genotypes. *Crop Protection* **26,** 198–205.

Bennet MD & Smith JB (1991) Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London Series B. Biological Sciences* **334,** 309–345.

Bir SS, Chatha GS & Sidhu M (1992) Intraspecific variation in Cyperaceae from Punjab Plain, India. *Willdenowia* **22,** 133–142.

Bodo Slotta TA (2008) What we know about weeds: insights from genetic markers. *Weed Science* **56,** 322–326.

Brownstein MJ, Carpten JD & Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20,** 1004–1006, 1008–1010.

Davies EW (1956) Cytology, evolution and origin of the aneuploid series in the genus Carex. *Hereditas* **42,** 349–365.

Dodet M, Petit RJ & Gasquez J (2008) Local spread of the invasive *Cyperus esculentus* (Cyperaceae) inferred using molecular genetic markers. *Weed Research* **48,** 19–27.

Edenfield MW, Brecke BJ, Colvin DL, Dusky JA & Shilling DG (2005) Purple Nutsedge (*Cyperus rotundus*) control with glyphosate in soybean and cotton. *Weed Technology* **19,** 947–953.

Efron B, Halloran E & Holmes S (1996) Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences USA* **93,** 13429–13434.

Felsenstein J (1985) Confidence-limits on phylogenies – an approach using the bootstrap. *Evolution* **39,** 783–791.

Glenn TC & Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology* **395,** 202–222.

Goolsby JA, de Barro PJ, Makinson JR, Pemberton RW, Hartley DM & Frohlich DR (2006) Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. *Molecular Ecology* **15,** 287–297.

Hansson B & Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11,** 2467–2474.

Hayden MJ, Good G & Sharp PJ (2002) Sequence tagged microsatellite profiling (STMP): improved isolation of DNA sequence flanking target SSRs. *Nucleic Acids Research* **30,** e129.

Heenan PB & de Lange PJ (2005) *Cyperus insularis* (Cyperaceae), a new species of sedge from northern New Zealand. *New Zealand Journal of Botany* **43,** 351–359.

Heiser CB & Whitaker TW (1948) Chromosome number, polyploidy, and growth habit in California weeds. *American Journal of Botany* **35,** 179–186.

Huang SB, Colmer TD & Millar AH (2008) Does anoxia tolerance involve altering the energy currency towards PPi? *Trends in Plant Science* **13,** 221–227.

Katti MV, Ranjekar PK & Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* **18,** 1161–1167.

Kijas JM, Fowler JC, Garbett CA & Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques* **16,** 656–660.

Kumar SV, Tan SG, Quah SC & Yusoff K (2002) Isolation and characterization of seven tetranucleotide microsatellite loci in mungbean, *Vigna radiata*. *Molecular Ecology Notes* **2,** 293–295.

Lu YQ, Baker J & Preston C (2007) The spread of resistance to acetolactate synthase inhibiting herbicides in a wind borne, self-pollinated weed species, *Lactuca serriola* L. *Theoretical and Applied Genetics* **115,** 443–450.

Molin WT, Ray JD, Scheffler BE, Kronfol RR & Bryson CT (2009) Genetic variation in purple nutsedge (*Cyperus rotundus*). WSSA Annual Meeting, Orlando, FL Abstract number 46.

Nei M (1972) Genetic distance between populations. *The American Naturalist* **106,** 283–292.

Okada M, Ahmad R & Jasieniuk M (2007) Microsatellite variation points to local landscape plantings as sources of invasive pampas grass (*Cortaderia selloana*) in California. *Molecular Ecology* **16,** 4956–4971.

Okoli CAN, Shilling DG, Smith RL & Bewick TA (1997) Genetic diversity in purple nutsedge (*Cyperus rotundus* L.) and yellow nutsedge (*Cyperus esculentus* L.). *Biological Control* **8,** 111–118.

Peña-Fronteras JT, Villalobos MC, Baltazar AM *et al.* (2009) Adaptation to flooding in upland and lowland ecotypes of *Cyperus rotundus*, a troublesome sedge weed of rice: tuber morphology and carbohydrate metabolism. *Annals of Botany* **103,** 295–302.

Reddy KN & Bryson CT (2009) In-crop and Autumn-applied glyphosate reduced purple nutsedge (*Cyperus rotundus*) density in no-till glyphosate-resistant corn and soybean. *Weed Technology* **23,** 384–390.

Richard GF, Kerrest A & Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews* **72,** 686–727.

Rozen S & Skaletsky H (2000) *Bioinformatics Methods and Protocols. In the series Methods in Molecular Biology*. Humana Press, Totowa, NJ, USA.

Sharma A, Namdeo AG & Mahadik KR (2008) Molecular markers: new prospects in plant genome analysis. *Pharmacognosy reviews* **2,** 23–34.

Sharopova N, McMullen MD, Schultz L *et al.* (2002) Development and mapping of SSR markers for maize. *Plant Molecular Biology* **48,** 463–481.

Tabacchi M, Mantegazza R, Spada A & Ferrero A (2006) Morphological traits and molecular markers for classification of *Echinochloa* species from Italian rice fields. *Weed Science* **54,** 1086–1093.

Tayyar RI, Nguyen JHT & Holt JS (2003) Genetic and morphological analysis of two novel nutsedge biotypes from California. *Weed Science* **51,** 731–739.

Techen N, Arias RS & Glynn NC (2010) Optimized construction of microsatellite-enriched libraries. *Molecular Ecology Resources* **10,** 508–515.

Varshney RK, Graner A & Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23,** 48–55.

Waldbieser GC, Quiniou SM & Karsi A (2003) Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. *BioTechniques* **35,** 976–979.

Wang G, McGiffen Jr ME & Ogbuchiekwe EJ (2008) Crop rotation effects on *Cyperus rotundus* and *C. esculentus* population dynamics in southern California vegetable production. *Weed Research* **48,** 420–428.

Weber JL (1990) Informativeness of human (dC-dA)n · (dG-dT)n polymorphisms. *Genomics* **7,** 524–530.

Wills GD (1998) Comparison of purple nutsedge (*Cyperus rotundus*) from around the world. *Weed Technology* **12,** 491–503.