MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2177–2193 doi: 10.1111/mec.13167

INVASION GENETICS: THE BAKER AND STEBBINS LEGACY

Exploring origins, invasion history and genetic diversity of *Imperata cylindrica* (L.) P. Beauv. (Cogongrass) in the United States using genotyping by sequencing

A. MILLIE BURRELL,* ALAN E. PEPPER,† GEORGE HODNETT,‡ JOHN A. GOOLSBY,§ WILLIAM A. OVERHOLT,¶ ALEXIS E. RACELIS,** RODRIGO DIAZ¶ and PATRICIA E. KLEIN*

*Institute for Plant Genomics and Biotechnology, Department of Horticultural Sciences, Texas A&M University, College Station, TX 77843-2123, USA, †Department of Biology, Texas A&M University, College Station, TX 77843-3258, USA, ‡Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474, USA, §Cattle Fever Tick Research Laboratory, USDA-ARS, Moore Air Base, Building 6419, Edinburg, TX 78541, USA, ¶Biological Control and Containment Laboratory, University of Florida, 2199 South Rock Road, Fort Pierce, FL 34945-3138, USA, **Department of Biology, University of Texas Pan American, 1201 West University Drive, Edinburg, TX 78539, USA*

Abstract

***Imperata cylindrica* (Cogongrass, Speargrass) is a diploid C4 grass that is a noxious weed in 73 countries and constitutes a significant threat to global biodiversity and sustainable agriculture. We used a cost-effective genotyping-by-sequencing (GBS) approach to identify the reproductive system, genetic diversity and geographic origins of invasions in the south-eastern United States. In this work, we demonstrated the advantage of employing the closely related, fully sequenced crop species *Sorghum bicolor* (L.) Moench as a proxy reference genome to identify a set of 2320 informative single nucleotide and insertion–deletion polymorphisms. Genetic analyses identified four clonal lineages of cogongrass and one clonal lineage of *Imperata brasiliensis* Trin. in the United States. Each lineage was highly homogeneous, and we found no evidence of hybridization among the different lineages, despite geographical overlap. We found evidence that at least three of these lineages showed clonal reproduction prior to introduction to the United States. These results indicate that cogongrass has limited evolutionary potential to adapt to novel environments and further suggest that upon arrival to its invaded range, this species did not require local adaptation through hybridization/introgression or selection of favourable alleles from a broad genetic base. Thus, cogongrass presents a clear case of broad invasive success, across a diversity of environments, in a clonal organism with limited genetic diversity.**

Keywords: clonal, comparative genomics, invasive species, Poaceae, rhizome, SNP molecular markers

Received 15 October 2014; revision received 16 March 2015; accepted 18 March 2015

Introduction

There are multiple genetic pathways to colonization success. In one well-established paradigm, standing genetic diversity, along with recombination, results in genotypes that are favourable for survival in the introduced range. A notable example comes from stickleback colonization of novel aquatic habitats (Hohenlohe *et al.* 2010; Jones *et al.* 2012). Another strategy involves hybridization of the introduced organism with related species present in the introduced range, resulting in genotypes favourable for survival in the new environment among the hybrid offspring, as demonstrated in *Helianthus* spp. (Whitney *et al.* 2006; Kane & Rieseberg 2007). In contrast, some organisms are believed to arrive in a new range as a 'general

Correspondence: A. Millie Burrell, Fax: +1 979 845 2891; E-mail: amburrell@tamu.edu

purpose genotype' (Baker 1965), pre-adapted to thrive, and not requiring recombination or selection for establishment and long-term persistence in novel environments.

Baker investigated the characteristics of the 'ideal weed', largely based on his concept of the 'general purpose genotype' (Baker 1965, 1974). A plant suitable for Baker's designation has characteristics that enable the following: (i) rapid growth, (ii) abundant reproduction of propagules, capable of rapid regeneration, (iii) ability to colonize a wide diversity of environmental conditions and (iv) distinctive traits contributing to its competitive success in its colonized environment (Baker 1965). The mode of reproduction (sexual or asexual) is inconsequential, providing that substantial numbers of rapidly regenerating structures are produced and possess traits ensuring dispersal over both long and short distances and successful colonization upon arrival.

For plants that do not require adaptation to the invaded range, the lack of an evolutionary lag time necessary for local adaptation facilitates rapid colonization, build-up of propagules and subsequent range expansion (Baker 1974; Barrett *et al.* 2008). In such plants, clonal reproduction has several advantages over sexual reproduction. A lack of sexual recombination would fix multilocus genotypes (MLGs) that are broadly adapted for survival in diverse environments (Barrett *et al.* 2008). For plants that utilize vegetative reproduction, the costs of sex are high in terms of allocation of photosynthate to reproductive organs. Further, adaptation by sexual recombination will require several reproductive cycles to produce a genotype suited for survival in the invaded range (Doncaster *et al.* 2000; Vallejo-Marin *et al.* 2010). Many plants utilize rhizomes as a means of colonizing new habitats, especially areas with heterogeneous soils and high levels of abiotic stress (You *et al.* 2014). Investing photosynthate into rhizomatous growth allows for both continued colonization as well as moisture and nutrient foraging in poor soils (Keser *et al.* 2014). A number of plant species with clonal reproductive strategies, and therefore limited genetic diversity, recombination and evolutionary potential, have emerged as significant biological invaders (Hollingsworth & Bailey 2000; Pappert *et al.* 2000; Zhang *et al.* 2010). In order to establish, displace and spread to novel habitats, it has been suggested that invasive clonal plants have characteristics approaching that of a 'Darwinian demon', a hypothetical organism that can reproduce early, produce infinitely many offspring and persist indefinitely (Law 1979; Rees 1993; Silvertown 2005).

The invasive Old World grass, *Imperata cylindrica*, exhibits many of the hallmark phenotypic characteristics of the 'ideal weed' and 'general purpose genotype' described by Baker (1965, 1974). This species exhibits both rapid vegetative growth and rapid production of numerous vegetative propagules with attributes enabling both short- and long-distance dispersal. Cogongrass also employs a number of classic competition mechanisms including the production of allelopathic compounds, dense 'choking' rhizomatous growth and ecological alteration of invaded habitats (Rodriguez *et al.* 2005; Daneshgar & Jose 2009; Xuan *et al.* 2009). Reports of widespread invasiveness began to surface in the early decades following its introduction to the United States (Wilcut *et al.* 1988; Lambert & Millar 1995; Peterson *et al.* 2003; Roderick & Navajas 2003; Lavergne & Molofsky 2007). The short time between the introductions (to Grand Bay, AL, McNeil, MS and Gainesville, FL) and published accounts warning of its invasiveness support the hypothesis that cogongrass was pre-adapted to its introduced range at the time of introduction, thereby not requiring an adaptive lag time necessary for effective colonization.

Other characteristics this species shares with Baker's 'ideal weed' include its C4 metabolism that, under some climatic conditions, broadens its ecological range in various soil types, sunlight levels and moisture regimes (Brown 1978; Patterson *et al.* 1980). Further, the piercing rhizomes of cogongrass form a dense, almost impenetrable underground matrix that typically comprises 80% of its total biomass (Aulakh *et al.* 2014). This species is not only tolerant of a wide range of soil compositions, but accumulates iron and shows no reduction in biomass accumulation in the presence of levels of heavy metals such as lead, which would be considered toxic to the majority of plants (Baker 1965; Paz-Alberto *et al.* 2007). The ability to remove iron from the soil, thereby altering soil chemistry, has been shown to be detrimental to native plant growth, which includes iron-dependent endemics (Rodriguez *et al.* 2005). Another competition mechanism is the ability of this species to alter the fire ecology of its invaded range. Low-temperature fires were once a natural component of the pine savannah in its US invaded range. The intense, high-temperature fires of cogongrass thatch (>500 °C) are sufficient to kill adult pine trees as well as destroy the viability of dormant seeds in the seed bank (Grace *et al.* 2001; Jose *et al.* 2002; Falk *et al.* 2007; Yager *et al.* 2010) (Fig. 1). Because this species has shown consistent resistance to herbicides, the overarching goal for this research is to develop effective host-specific biological controls for affected areas in the United States.

In this work, we used a population genomics approach utilizing genotyping by sequencing (GBS) and comparative genomics to address the following objectives: (i) elucidation of the reproductive mode of this organism, (ii) determination of genetic diversity within and among US populations, (iii) determination of the pattern of introductions to North America and, lastly,

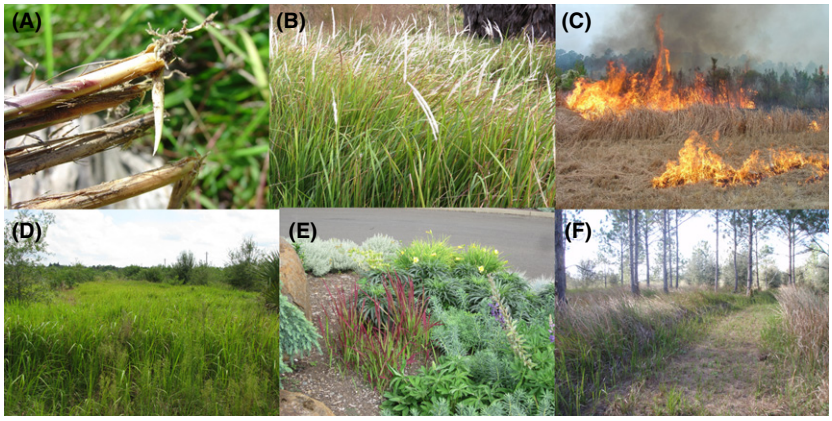

Fig. 1 Ecological features of *Imperata cylindrica* (Cogongrass) invasion. (A) Sharp, piercing rhizomatous growth, (B) monoculture of cogongrass in bloom. Each inflorescence can produce ~3000 seeds, (C) high-temperature grassfire of cogongrass, (D) a once productive citrus orchard, presently a monoculture of *I. cylindrica*, (E) horticultural variety, *I. cylindrica* '*rubra*' (Japanese blood grass). At the edges of the clump, green rhizomatous growth is beginning to spread away from parent plant, (F) regrowth of cogongrass following herbicide application. In the maintained right-of-way, rhizomatous growth is emerging. Photograph credits: Mike Murphrey, Texas A&M Forest Service; Charles Cook, Florida Department of Environmental Protection.

(iv) identification of potential international sites of origin (that would provide a framework in the search for possible biological control agents).

We provide compelling evidence that cogongrass has a highly clonal reproductive strategy and very limited genetic diversity within its invaded ranges. The lack of correlation between genetic distance and geographical distance of the accessions sampled highlights the continuing anthropogenic dispersal of this species. International sampling identified genotypes belonging to three of the five clonal lineages of *Imperata* spp. extant in the United States.

Materials and methods

Study system

As a member of the Poaceae family, *Imperata* spp. reside in the well-studied subtribe, Saccharinae (Clifton-Brown *et al.* 2008; Murray *et al.* 2008; Waclawovsky *et al.* 2010; Wang *et al.* 2013) that includes several species that are agronomic crops for food and fibre as well as bioenergy production such as sorghum, sugarcane and switchgrass. High economic value has driven the production of significant genomic resources for these taxa (Paterson *et al.* 2009; Sharma *et al.* 2012; de Setta *et al.* 2014). This subtribe also includes weedy and invasive pests (in particular, *Saccharum spontaneum* and *Sorghum halapense*) (Joo Kim *et al.* 2008; Craven *et al.* 2009; Rout *et al.* 2013). Several members of this group, including some considered invasive, possess the abilities to produce apomictic (asexual) seed as well as colonize through asexual rhizomatous growth (Duara & Stebbins 1952; Panje 1970; Warwick & Black 1983; Chapman 1992; Richard 1998). Similar to many of its Poaceae relatives, *I. cylindrica* has been classified as an obligate outcrossing species (Dozier *et al.* 1998; Stone 2002).

Cogongrass occurs on every continent except Antarctica, and its geographical centre of origin and 'native range' are unknown. With regard to the introduction of accessions to the United States, anecdotal records suggest that one introduction may have occurred in a shipment of satsuma orange bud-wood (*Citrus unshiu* (Swingle) Marcow.) from Japan to Mobile, Alabama circa 1911–1912 (MacDonald 2004; Capo-chichi *et al.* 2008; Barry 2009). The current areas of infestation in the United States are primarily in rural regions of the south-eastern states—a region with a long history of dependence on agriculture and timber production as a principal source of revenue (Willard *et al.* 1990; Yager *et al.* 2010; Larson *et al.* 2011) (Fig. 2). Due to its ability to produce significant biomass without irrigation and inputs of fertilizer, propagules of cogongrass (purportedly from The Philippine Islands) were later intentionally planted as a forage grass and soil stabilizer in McNeill, Mississippi and Gainesville, Florida (Hubbard *et al.* 1944; Dickens & Buchanan 1971; Dickens 1974; Dozier *et al.* 1998). To date, cogongrass is reported to occur in eight states as a noxious weed. Its range presently extends from North Carolina west to Texas (MacDonald 2004; Burrell, data collection).

The work described here utilizes a genomewide polymorphic molecular marker system based on next-generation DNA sequence data. Previous population-genetic studies have been undertaken to assess the population structure and diversity of cogongrass in the United States (Capo-chichi *et al.* 2008; Vergara *et al.* 2008; Lucardi *et al.* 2013, 2014). However, small sample sizes,

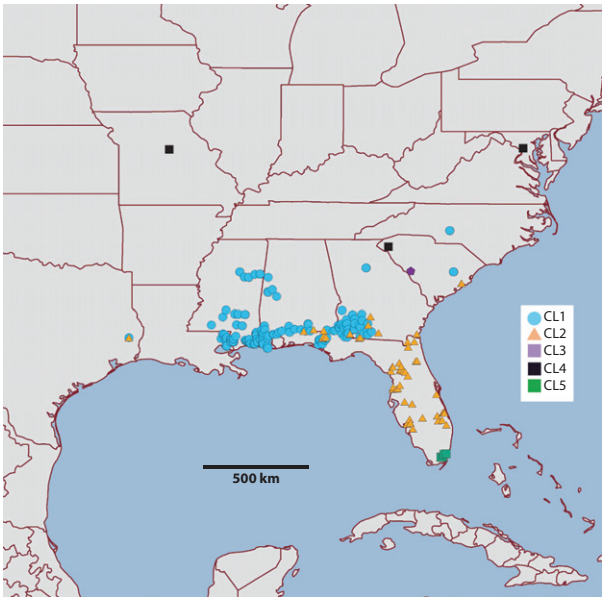

Fig. 2 Distribution of *Imperata* spp. in the United States. Teal circles: Clonal Lineage 1. Orange triangles: Clonal Lineage 2. Purple hexagons: Clonal Lineage 3. Black squares: Clonal Lineage 4 (*Imperata cylindrica 'rubra'*, also in Oregon). Green squares: Clonal Lineage 5 (*Imperata brasiliensis*). Scale bar shown is 500 km.

small numbers of markers and/or geographically limited sampling have reduced the informative value of these efforts. In addition, the preceding studies have employed dominant, non-sequence-based molecular markers such as intersimple sequence repeats (ISSRs) and amplified fragment length polymorphisms (AFLPs), which provide insufficient information on hybridization, heterozygosity and other population-genetic parameters (Mueller & Wolfenbarger 1999; Balloux *et al.* 2003). Some of these studies have concluded that significant genetic diversity is present in cogongrass and that it may hybridize with congener *Imperata brasiliensis* Trin. in the United States (Hall 1998; Bryson *et al.* 2010; Lucardi *et al.* 2013). There are two studies employing the use of sequence-based microsatellite markers (SSRs) on Japanese accessions (Maeda *et al.* 2009; Chiang *et al.* 2012). However, none of these sequence-based markers have been employed on US accessions of cogongrass.

Sampling of US and international accessions

A total of 449 *Imperata* spp. samples (including GPS coordinates for each sample) were collected from 94 populations in cooperation with a network of academic collaborators, state foresters and federal land managers. *Imperata* spp. sampled included accessions from the following US states: Alabama (AL), Florida (FL), Georgia (GA), Louisiana (LA), Maryland (MD), Mississippi (MS), Missouri (MO), North Carolina (NC), Oregon (OR), South Carolina (SC) and Texas (TX) (Fig. 2). Collection locations were sites that had been reported to be infested with cogongrass and are regularly herbicide-treated by land managers, who collected samples prior to scheduled herbicide application, as well as sites in natural areas and unreported populations that were not managed with herbicide application. At the request of private landowners, we agreed to publish collection location only at the county level. To explore potential geographical origins for introduced genotypes, as well as genetic similarities to US accessions, international samples were collected from locations historically considered to be potential sources of introduction of US populations of cogongrass, including Brazil, Japan and the Philippine Islands. We also included accessions of the cogongrass horticultural variety '*rubra*' (Japanese blood grass) in our study. Although not considered a noxious weed in most US states, this strain has been observed to shift to an aggressive invasive phenotype (material provided by Dr. Leland Cseke) (Bryson & Carter 1993; Cseke & Talley 2012). The species, *I. brasiliensis*, collected in both Dade County, FL, and multiple sites in Brazil was included in this study to determine whether hybridization between *I. brasiliensis* and cogongrass has occurred in the United States following introduction and potentially contributed to novel invasive genotypes (Ellstrand & Schierenbeck 2000). Collector data are hosted in the Dryad Digital Depository in data package doi:10.5061/dryad.k2d05.

A minimum of five samples was obtained from each location. Individual sampled plants were spaced at a minimum of 5 m apart but on average 10 m. Four to five leaves approximately 15 cm in length were collected from individual plants and placed into WhirlPak self-sealing plastic bags (part no. B01009WA; Nasco) filled with 8%, indicating silica bead desiccant (t.h.e Desiccant, part no. DX0014-1; EMD Millipore) and labelled with geospatial coordinates. Leaf samples were allowed to dry and were stored at room temperature in a sealed box prior to DNA extraction. A ~1 cm × 6 cm fragment of a single dried leaf was used for DNA extraction from each accession.

DNA extraction

Total genomic DNA was isolated using MP Biomedicals FastDNA Spin Kit in conjunction with the FastPrep FP120 instrument (Bio 101, Savant). DNA samples were purified in a 96-well format using the ZR-96 Genomic DNA Clean and Concentrator™-5 (Zymo Research) to reduce contaminants that would potentially interfere with restriction enzyme digestion. Due to a high concentration of silicates in the leaves of cogongrass, we

made minor modifications to the standard FastDNA Spin Kit protocol (Appendix S1, Supporting information). The resulting RNase-treated DNAs were quantified using a double-stranded DNA-specific fluorescent dye assay (AccuBlue Broad Range dsDNA Quantification Kit; Biotium) in 96-well plate format. DNA purity was assessed using the NanoDrop1000 instrument (Thermo Fisher Scientific). DNA samples from leaf tissue were re-extracted unless the OD 260/280 ratio was >1.7 and the absorbance maximum at 260 nm.

Genome size estimation

Estimates of genome size and composition are critical for the development of genotyping strategies based on restriction endonucleases (Cariou *et al.* 2013). For example, the amount of sequencing required to achieve a desired depth of coverage per multiplexed sample is dependent upon genome size and complexity. *Imperata* spp. plants used for genome size estimation were grown in a temperature- and humidity-controlled growth chamber, approved by APHIS for quarantine to prevent accidental release of propagules into the environment. Healthy young leaves from each plant along with an internal standard (*Oryza sativa* L. IR36 or *Sorghum bicolor* Tx3361) were chopped with a razor blade in ice-cold woody plant buffer (Loureiro *et al.* 2007). The concentration of Triton X-100 in the buffer was changed from a concentration of 1% (v/v) to 0.2%. Chopped leaves were filtered through a 54-μm nylon mesh filter. To the filtered buffer, propidium iodide was added to a final concentration of 50 ppm and placed on ice. Mean fluorescence of 12 samples from each *Imperata* spp. genotype was quantified using the BD Accuri™ C6 flow cytometer (BD Biosciences). The samples were run in triplicate on each of four independent runs. The 2C DNA content of each genotype was determined by comparing the mean fluorescence of each *Imperata* spp. sample to the internal standards.

Genotyping-by-sequencing methodology

DNA samples were genotyped using a restriction endonuclease-assisted, reduced-representation marker system called digital genotyping (DG) (Morishige *et al.* 2013). This method was developed specifically for use in C4 grasses on the Illumina GAIIx short read sequencing platform (Illumina). The method incorporates the use of the methylation-sensitive restriction enzyme *Fse*I (GGCCGG|CC) to achieve complexity reduction, enrich for hypomethylated euchromatic genomic regions and to limit genomic fragments from hypermethylated heterochromatic regions, such as repetitive DNA clustered around centromeres, pseudogenes, transposons and retrotransposons (Davey *et al.* 2011). Particularly in grass species, an abundance of methylation and repetitive sequences have been observed, which present bioinformatic analysis challenges (Larrinua & Belmar 2008). This significant hurdle can be bypassed using a methylation-sensitive restriction enzyme such as *Fse*I for the initial restriction enzyme digestion.

After restriction enzyme digestion, samples were ligated to a set of Illumina-compatible adapters with individual 4-bp in-line forward-read barcodes. Samples were pooled and randomly sheared to a 250–300 bp range using sonication (Covaris® S2) and size-selected on a 2% agarose gel. Following overhang fill-in, blunting and 3′ adenylation, the pools underwent ligation with an Illumina-specific adapter. Pools were PCR-amplified for 10 cycles using Phusion® High-Fidelity Polymerase. Single-strand products were obtained and then PCR-amplified for 8 cycles with Phusion® High-Fidelity Polymerase to incorporate the Illumina bridge amplification sequence. Using this multiplex approach, 48 samples were pooled in one lane of the Illumina GAIIx. Single-end sequencing was carried out for 78 cycles. Detailed information about modifications to this protocol for use in *Imperata* spp. is contained in Appendix S1 (Supporting information).

Bioinformatics data processing

FASTQ sequences were trimmed and sorted based on individual 4-bp barcodes and the partial *Fse*I restriction site. Only those sequences having 100% identity to the barcode and partial restriction site were retained. Although a reference sequence is not essential for GBS applications (Willing *et al.* 2011; Narum *et al.* 2013), we had established significant sequence similarity of *I. cylindrica* to close relative *S. bicolor* through BLAST (Altschul *et al.* 1990) comparisons with microsatellites (Burrell 2009). Additionally, phylogenetic evidence of this relationship was found in the existing literature (Al-Janabi *et al.* 1994; Hodkinson *et al.* 2002; Roodt-Wilding & Spies 2006; Bouchenak-Khelladi *et al.* 2008). Thus, the fully sequenced and well-annotated *S. bicolor* genome was employed as a proxy reference sequence to facilitate discovery of single nucleotide polymorphisms (SNPs) in the *Imperata* spp. genomes. Filtered reads for each individual accession were mapped to the *S. bicolor* genome sequence (Sb1.4) (Paterson *et al.* 2009) and analysed for SNPs and INDELs using the CLC Bio Genomics Workbench software (versions 5.0.2 and 6.5.1; CLC Bio). Read mapping parameters were set to insertion and deletion cost = 3, mismatch cost = 2, 50% minimum read length required to match the reference and a minimum of 80% similarity between the read and the reference sequence. Reads that did not align to the *S. bicolor*

genome and reads that aligned to more than one position in the *S. bicolor* genome (e.g. repetitive elements) were discarded.

One potential pitfall of using marker systems, such as GBS and RAD-seq (without a reference genome), AFLPs and RAPDs, to characterize samples collected from natural habitats was the risk of sample contamination from a variety of organisms, ranging from microbial epiphytes, endophytes and pathogens, to residues from invertebrates and vertebrates residing in the ecosystem. By employing the *S. bicolor* genome as a proxy reference genome, we could ensure that any contaminating sequences were excluded from the data set at this juncture of analysis.

For SNP detection in the CLC Bio Genomics Workbench, the parameters included: a neighbourhood radius of 5, a maximum gap and mismatch count of 3, a minimum quality of the SNP base of 20, a minimum average quality of the nucleotides surrounding the SNP of 15 and a minimum read coverage for a SNP of 10. This stringent minimum coverage for each SNP was applied as a means of differentiating a sequencing error from a legitimate SNP. To estimate error in sequencing and SNP calling, a single DNA was genotyped in triplicate on separate lanes on the Illumina GAIIx.

The data were exported from the CLC Genomics Workbench in the form of SAM files, and comma-separated value formatted tables containing SNPs and INDELs for each sample. These files were reformatted for downstream analysis using a series of custom perl and python scripts. The scripts and descriptions of their function can be found at the Dryad Digital Depository, doi:10.5061/dryad.k2d05. The final output file combined the data for all *Imperata* spp. samples and contained SNP/INDEL identity anchored to a specific location in the *S. bicolor* genome, as well as the number of reads for each variant in a .csv format that can be reformatted for any downstream data analysis software. At no point in any of the sequence analyses were data imputed.

Data analysis

Sequence data from 449 samples were included in the final data set for analysis to limit oversampling bias from heavily sampled populations ($n > 50$). Due to the rhizomatous growth habit and ability to regenerate from as small as a 2-mm^2 rhizome fragment, we were cognizant that clonal reproduction might play a significant role in the population-genetic structure of this species. Thus, the first step in the analysis was to obtain estimates of fit to Hardy–Weinberg equilibrium (HWE) (Table 1). If populations showed significant deviation from HWE, then widely utilized ancestry-inference model software programs such as STRUCTURE and ADMIXTURE would not be appropriate for the analysis of clonal data. These programs assume HWE for functionality (Pritchard *et al.* 2000; Alexander *et al.* 2009; J. K. Pritchard, personal communication, 2013). Because alleles are not independently assorting through random mating in clonal populations, the assumption of HWE is violated (Kamvar *et al.* 2014). In addition, loci will show linkage disequilibrium in clonal populations (de-Meeûs & Balloux 2004).

The software package, GENODIVE, which is broadly applicable to both diploid and polyploid species as well as clonal and sexual species (Meirmans & Van Tienderen 2004), was used for calculating genetic distance among samples, population-genetic statistics, principal coordinates analysis (PCoA), estimates of genetic distance vs. geographic distance correlation and measures of potential clonality. To assign individuals to clonal lineages, we employed a method that uses a frequency distribution of pairwise distances among individuals to empirically delineate a distance threshold for membership in the same clonal lineage (Rogstad *et al.* 2002; Douhovnikoff & Dodd 2003; Meirmans & Van Tienderen 2004; Arnaud-Haond *et al.* 2007a; Tarin *et al.* 2013). Because SNP data were employed, an infinite allele mutation model was used (Meirmans & Van Tienderen 2004). In all analyses, A, C, G, T and INDEL alleles were weighted equally. Missing data were ignored. Geographic mapping by genotype was visualized using ARC-GIS software (ESRI 2011).

To estimate the error rate due to sequencing and variant calling at all 2320 loci used in the SNP data set, one DNA sample from a single plant (Marion, OR) was sequenced in triplicate in separate lanes on the Illumina GAIIx instrument.

Results

Genome size of Imperata spp.

The genome sizes for the *Imperata* spp. accessions analysed (Table 2) were similar to each other (range 640–667 Mbp) with the exception of *I. cylindrica 'rubra'* (Japanese blood grass), which had a larger genome (745 Mbp). Thus, the genomes of *Imperata* spp. were smaller than but similar to that of *S. bicolor* (818 Mbp) (Johnston *et al.* 1999; Price *et al.* 2005). From this data, we concluded that the digital genotyping and multiplexed sample pooling strategies developed for *S. bicolor* were appropriate for use in this project.

SNP/INDEL marker characterization

Utilization of comparative genomics enabled us to identify reads with high sequence similarity to *S. bicolor*,

Table 1 Population-genetic parameters of representative cogongrass infestations

Population	Clonal lineage	Sample size	Mean alleles/locus	Effective alleles/locus	H_O	H_S	G_{IS}
Aiken SC	CL3	9	1.201	1.128	0.123	0.070	−0.748
Harrison MS	CL1	39	1.305	1.135	0.112	0.078	−0.445
Pearl River MS	CL1	26	1.250	1.134	0.109	0.077	−0.425
Washington, FL	CL2	10	1.280	1.149	0.113	0.089	−0.274
George, MS	CL1	25	1.229	1.138	0.111	0.079	−0.404
Jackson, MS	CL1	23	1.308	1.144	0.119	0.083	−0.432
Decatur, GA	CL1	7	1.268	1.149	0.115	0.091	−0.258
Early, GA	CL1	7	1.208	1.137	0.113	0.080	−0.413
Mobile, AL	CL1	25	1.282	1.139	0.114	0.080	−0.423
Oktibbeha, MS	CL1	7	1.191	1.135	0.121	0.076	−0.586
Dade, FL (*I. brasiliensis*)	CL5	8	1.140	1.089	0.079	0.050	−0.586
St. Lucie, FL	CL2	9	1.180	1.120	0.106	0.068	−0.559
Charlotte, FL	CL2	7	1.190	1.122	0.109	0.072	−0.527
Kuroshio, Japan	CL4	7	1.184	1.137	0.134	0.073	−0.826
Kochi, Japan	CL1	7	1.247	1.150	0.127	0.089	−0.424
Laguna, Phil.	n/a	14	1.310	1.147	0.133	0.084	−0.582
Sape, Brazil (*I. brasiliensis*)	CL5	7	1.190	1.101	0.091	0.061	−0.496

H_O, observed heterozygosity; H_S, expected heterozygosity; G_{IS}, inbreeding coefficient ($1 - H_O/H_S$).

Table 2 Genomic DNA content of representative plants of *Imperata cylindrica* and *Imperata brasiliensis* clonal lineages. *Sorghum bicolor* genotype Tx3361 (2C content = 1.67 pg) and *Oryza sativa* cultivar IR36 (2C content = 1.08 pg) were used as calibration standards

	2C DNA content			
Species	pg	SE	Mbp/1C*	Origin
I. cylindrica (CL1)	1.36	0.028	667	Washington, FL
I. cylindrica (CL2)	1.31	0.026	640	St. Lucie, FL
I. cylindrica (CL4)	1.52	0.031	745	Nursery Stock, OR
I. brasiliensis	1.36	0.03	664	Dade County, FL

*1 pg = 980 mega base pairs (Mbp) (Cavalier-Smith 1985).

which we could assume originated in the *Imperata* genome. This filter allowed us to eliminate any reads that were of nonplant DNA origin from field-collected samples. Our hypothesis that reads produced from sequencing wild-collected leaf tissues might contain non-*Imperata* reads was confirmed when individually barcoded sequences that did not map to *S. bicolor* were queried against the NCBI nonredundant nucleotide database using BLAST (Altschul *et al.* 1990). From one sample alone, collected near the western shore of Mobile Bay, AL, we observed an abundance of nonplant alignments. These included some notable examples of organisms one would expect in the coastal AL pine savannah ecosystem: halophilic soil microbes, saprophytic fungi, bacteria isolated from feral hog faeces, as well as plant pathogens affecting rice, sugarcane, corn and citrus (*e*-values $< E \times 10^{-7}$) (data not shown).

Approximately 30 Gbp of DNA sequence data was produced in this study. The average number of filter-passed reads containing both the partial restriction site and multiplex identifier barcode was 603 081 reads per sample. Filter-passed sequences from the *Imperata* data set mapped to 25 685 unique loci in the *S. bicolor* genome and were polymorphic with respect to the *S. bicolor* genome. A subset of 13 291 loci had 5% or less missing data across the entire sample set. These loci were further culled by removing all monomorphic *Imperata* loci, leaving a set of 2320 loci that were polymorphic among the *Imperata* accessions. These loci were distributed over each of the ten *S. bicolor* chromosomes. In the final data set, the average read depth per marker was 46.7 ± 15.0.

Estimate of error rate

Pairwise comparisons of the three resulting data sets of the sample sequenced in triplicate showed that among the 6960 (2320 × 3) pairwise nucleotide comparisons, there were 175 disagreements (data not shown). The overwhelming majority of these (172) were loci that were called as homozygous for a particular allele (e.g. T/T) in a subset of replicates and as heterozygous for that same allele (e.g. T/C) in the remaining replicate. Likely causes for this type of error include: (i) a true heterozygote is called as a homozygote due to inadequate sequencing depth, and (ii) substitution error during PCR amplification of the GBS library resulting in a true homozygote being called as a heterozygote. One locus (chr4_5016663) was homozygous for an INDEL in one sequencing replicate, but homozygous for a T in the other two replicates. Another locus (chr1_57965250)

was called as heterozygous (T/INDEL) in one replicate, but heterozygous for another allele (A/INDEL) in the other replicates. There were no homozygous to homozygous nucleotide substitutions (e.g. C/C to T/T). Across all loci, the overall technical error rate was 2.7 ± 0.7%. While exceedingly low, this technical error rate was not zero, and thus, it made a *positive* contribution to the observed genetic diversity seen among all samples in the study.

Population-genetic analyses

Across all 449 individuals and 2320 polymorphic loci in the final data set, the mean number of alleles per locus was 2.07 ± 0.011, indicating that most loci had two alleles, as is commonly the case for SNP and INDEL loci (Table S1, Supporting information). In our initial assessment of population-genetic parameters, we assigned individuals to populations (demes) based on the location of infestation. For these analyses, we used a subset of 17 designated populations with sample sizes of $n \geq 7$ that included 13 US sites along with four international sites (from potential source locations) for comparison (Table 1).

All of the selected populations showed highly significant deviations from HWE ($P < 0.0001$) as a result of heterozygote excess (inbreeding coefficient, G_{IS}, ranged from −0.274 to −0.826). An excess of observed heterozygotes (negative values for F_{IS} and G_{IS}) is predicted by simulations of clonal populations (Balloux *et al.* 2003) and is a common observation in natural populations that exhibit some degree of clonal reproduction (Pappert *et al.* 2000; Stoeckel *et al.* 2006; Prugnolle & De Meeûs 2008; Meloni *et al.* 2013; Hodoki *et al.* 2014), including the clonal invasive plants *Pueraria lobata* (Kudzu) (Pappert *et al.* 2000) and *Arundo donax* (Giant Cane) (Tarin *et al.* 2013). For these reasons, we examined the subset of 17 designated populations using measures of clonal diversity. All had a Nei's genetic diversity (*div*) of 1.0, suggesting that a single clone was present within each population (Nei 1987). As expected, the genotypic evenness (*eve*) in all populations was 1.0, and the Shannon's Corrected Index (*shc*) was 0.0 (Chao & Shen 2003). Although the sample sizes of some of the populations were quite small, our data showed a clear and consistent pattern indicative of a highly clonal mode of reproduction (i.e. vegetative reproduction) in both the US and international sites.

Assignment of individuals to clonal lineages

Given enough resolution, even purely clonally related individuals ('clone mates') will display measurable levels of genetic diversity as a result of technical sequencing errors, errors in the computational calling of alleles, and somatic mutation (Rogstad *et al.* 2002; Douhovnikoff & Dodd 2003; Meirmans & Van Tienderen 2004; Arnaud-Haond *et al.* 2007a). For this reason, we favour the use of the term 'clonal lineages' as has been suggested previously (Anderson & Kohn 1995; Balloux *et al.* 2003; Arnaud-Haond *et al.* 2007a,b) rather than 'clones' or 'MLGs'. Using the empirical method based on assignment of a clonal threshold to a histogram of pairwise genetic distances (Meirmans & Van Tienderen 2004), we found that all US samples fell into five clonal lineages (designated CL1–CL5), while the Japanese samples fell into two clonal lineages (CL1 and CL4), both present in the United States. Samples from the Philippine Islands fell into multiple clonal lineages, none of which were shared with US samples, and were therefore dropped from the frequency distribution of pairwise distances shown in Fig. 3.

Our distribution of pairwise distances was tri-modal. The leftmost peak was assumed to be distances between members of the same clonal lineage (the result of a composite of technical error and somatic mutation). Examples of accession comparisons in this peak include CL1 by CL1 and CL2 by CL2. The mean frequency of polymorphism among individuals in the leftmost peak was approximately ~8.7%, suggesting that the observed divergence due to somatic mutation was on the order of ~6% (when ~2.7% sequencing error was taken into consideration). The threshold for clonal assignment was set at a genetic distance of 341 (corresponding to a mean polymorphism rate of ~15%). Based on this assessment, the most geographically widespread clonal lineage in the United States (designated CL1) was distributed across the Gulf Coast states and the south-eastern United States from TX to NC. CL1 also included individuals sampled from the Japanese islands of Honshu, Kyushu and Shikoku (distribution in Japan not shown). A second clonal lineage (CL2) was found primarily in peninsular FL, but also sporadically in northern FL and neighbouring regions of GA and AL as well as in TX. The geographical distribution of CL2 overlapped with that of CL1 in eight of the populations sampled. A third apparent clonal lineage, CL3, was limited to a single infestation near the city of Aiken, SC. The fourth clonal lineage (CL4) included individuals that had been morphologically categorized by our collectors as invasive and noninvasive forms of Japanese blood grass from the island of Shikoku, Japan, and from a number of widespread locations across the United States. The fifth clonal lineage (CL5) corresponded to those individuals that had been taxonomically described as *I. brasiliensis*, obtained from both the presumed native range (Brazil) and the southern tip of the FL Peninsula.

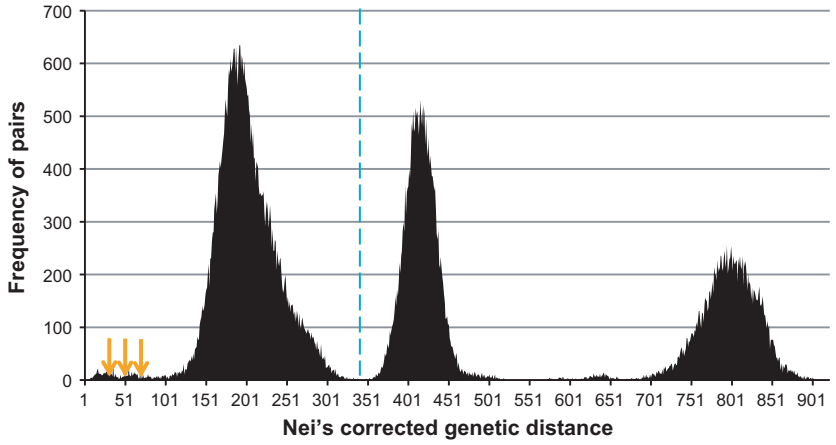

Fig. 3 Frequency distribution of pairwise genetic distances between *Imperata* samples. Genetic distances were calculated as described by Meirmans & Van Tienderen (2004), using an infinite allele model. This operation calculates the number of mutational steps to convert the genotype of one individual to the genotype of another individual. Histogram includes all study samples except those from the Philippine Islands. Orange arrows show the observed genetic distances between replicate sequencing trials from the same DNA sample (same individual, same tissue). Dashed line indicates threshold distance (D = 341) used for assignment of individuals to clonal lineages. The leftmost peak represents distances between members of the same clonal lineage (the result of a composite of technical error and somatic mutation). The mean frequency of polymorphism among individuals in the leftmost peak was approximately ~8.7%. The middle and rightmost peaks from the frequency distribution were distances between individuals of different clonal lineages. The middle peak represents comparisons between members of clonal lineages that were reasonably closely related. The rightmost peak represents members of clonal lineages that are highly dissimilar.

The middle and rightmost peaks from the frequency distribution (Fig. 3) were distances between individuals of different clonal lineages. The middle peak in Fig. 3 represents comparisons between members of clonal lineages that were reasonably closely related in the PCoA (Fig. 4). Examples included comparisons between CL1 and CL2 as well as comparisons between members of CL3 and CL4. The rightmost peak in Fig. 3 was populated by pairwise distances between highly dissimilar clones in the PCoA, such as CL2 and CL5 (*I. brasiliensis),* as well as CL1 and CL4 (Japanese blood grass), illustrated in Fig. 4.

Principal coordinates analysis

To further clarify the taxonomic relationships of *Imperata* accessions and to quantify the genetic differentiation among them, we performed a PCoA in GENODIVE, which neither imputes missing data nor sorts accessions into like groups based on missing data—as is often problematic with other PCoA estimation algorithms (Peakall & Smouse 2006). The first two coordinates explained 44% of the variance (Fig. 4). Our PCoA results consistently mirrored the clonal assignment data. Accessions from the Philippines were included in this analysis and are similar to but clearly distinct from CL2. The PCoA implies some degree of similarity in the Aiken, SC population (CL3) to the horticultural variety, Japanese blood grass (CL4). However, these were clearly two distinct clonal lineages.

Sites with mixed infestation

To investigate potential hybridization between clonal lineages, we examined 42 samples from eight sites within the geographic overlap of the distributions of CL1 and CL2 (Fig. 2) that had mixed infestations. These sites from northern FL, southern GA, south-eastern AL and Tyler County, TX, were examined using pairwise genetic divergence between individuals and PCoA. Pairwise genetic divergences between individuals and PCoA are shown in Figs S1 and S2 (Supporting information). In these analyses, we found no positive evidence for interclonal hybrids. Similarly, near the town of Kuroshio, Kochi Prefecture, Japan, infestations of CL1 and CL4 were located in close proximity, but no samples of hybrid origin or ancestry were observed.

Geographic pattern of genetic diversity within clonal lineages

To test the relationship between genetic distance and geographical distance, we employed Spearman's nonparametric regression (Conover & Iman 1981) and Mantel test (Smouse *et al.* 1986) as implemented in GENODIVE (Meirmans & Van Tienderen 2004). We obtained the following values for CL1: Mantel r = 0.019 (P = 0.297) and Spearman r = 0.026 (P = 0.223). For CL2, we obtained the following values: Mantel r = −0.034 (P = 0.332) and Spearman r = 0.037 (P = 0.280).

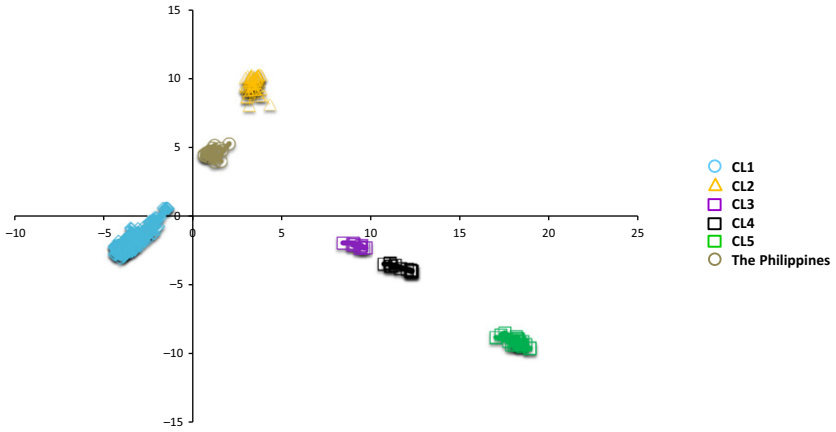

Fig. 4 Principal components analysis of 449 *Imperata* samples utilized in this study. Principal components 1 and 2 explain 44% of the observed variation. Axis 1 explains 22.52% of the observed variation. Axis 2 explains 18.77% of the observed variation. Teal samples were assigned to the Clonal Lineage 1, collected in AL, FL, GA, MS, NC, SC and Japan. Orange samples were assigned to the Clonal Lineage 2, collected in AL, FL, GA, SC and TX. Tan samples were assigned to accessions collected throughout the Philippine Islands. Purple samples were assigned to Clonal Lineage 3, collected in SC. Black samples were assigned to the Clonal Lineage 4, collected in MD, MO, OR, SC and Japan. Green samples were assigned to the *Imperata brasiliensis* Clonal Lineage 5, collected in FL and Brazil.

Within both clonal lineages (CL1 and CL2), we observed clusters of samples with highly similar MLGs, presumably the result of recent common ancestry, that were widely and discontinuously dispersed throughout their introduced ranges (Fig. 2). For example, samples from a recent infestation in Stanly, NC (reported May 2013), showed the highest genetic similarity to a set of 11 samples from Harrison County on the MS Gulf Coast (average genetic distance = 149.6 ± 2.6) and the three samples from Walton, GA (average genetic distance = 154.4 ± 2.5). These samples were spread across a south-west to north-east linear transect of ~920 km. Similarly, a CL1 sample collected in Tyler, TX, was most highly related to samples collected in Jackson County, MS (average genetic distance = 148.8 ± 5.74), approximately 560 km to the east.

Based on genetic distances, the CL2 component of infestation in Tyler, TX, was nested within a spatially noncontiguous sublineage of extremely closely related genotypes distributed sporadically across regions in northern FL, south-eastern AL and south-western GA. Genetic distances within this cluster averaged 45.5 ± 14.8, which is similar to the genetic distance seen among sequencing replicates of the single Marion, OR, sample (52.7 ± 13.9). The signature of this sublineage is prominent as the leftmost peak of Fig. S1 (Supporting information). Again, these results are consistent with a geographic dispersal pattern that is highly sporadic in nature, rather than some sort of 'front', 'stepping stone' or 'island model' of dispersal or gene flow.

Discussion

Clonal reproduction in Imperata cylindrica

One of our key findings using sequence-based markers was an excess in observed heterozygosity (H_o) over Hardy–Weinberg expectations. One explanation for excess H_o is clonal reproduction. At the point of a shift to asexual reproduction, any loci that are heterozygous in the sexual ancestor will be fixed in its asexual progeny. Further, any somatic mutations that occur during clonal reproduction will affect only one allele at each locus, leading to the gradual accumulation of heterozygous loci. Such heterozygous loci will be fixed in a clonal lineage, as heterozygosity cannot be eliminated through sexual processes such as segregation (in theory, such heterozygosity could be lost through mitotic recombination and gene conversion) (Klekowski 1997; Barrett *et al.* 2008). Indeed, Balloux *et al.* (2003) suggested that strict clonality in diploid populations can be easily detected by heterozygote excess.

Other potential sources of excess H_o include heterozygote advantage (overdominance), heterosis (which could include broad, allele-independent genomewide mechanisms) and disassortative mating (e.g. obligate outcrossing). However, overdominance at one or a few loci in a sexual population would not explain the genomewide distribution of excess H_o in our study. Further, overdominance, heterosis and disassortative mating in a sexual population will all result in more balanced polymorphism and a higher effective number of alleles than that observed here (Templeton 2006).

Excess in H_o can also occur as the result of a genetic bottleneck such as an invasion (Cornuet & Luikart 1996). However, this phenomenon is transient, and expected levels of heterozygosity are restored through subsequent sexual reproduction. In this study, values for G_{IS} were similar in putative source locations in Japan (−0.421, −0.826) and Brazil (−0.496) to what was found after invasion of the United States (Table 1). Thus, no signatures of presumed bottlenecks were discernable as differences in H_o.

Technical errors in sequencing and genotype calling could also lead to deviations from expected levels of heterozygosity. In contrast, novel sequences arising from PCR substitution error during the amplification phase of GBS library preparation would lead to seemingly novel alleles (and thus heterozygous genotype calls). To avert this possibility, we only retained alleles that were present in more than one DNA sample. Given the low intrinsic error rate of Phusion® High-Fidelity Polymerase (4.4×10^{-7} substitutions/base) and the limited rounds of PCR amplification during library preparation, a private allele shared by two (or more) samples is far more likely to reflect actual identity-by-descent than an artefact caused by parallel PCR errors ($\sim 2 \times 10^{-5}$ per locus). Thus, by employing the conservative measure of eliminating any alleles that were private to a single sample, we greatly reduced potential errors in overestimation of H_o (and possibly even induced errors leading to an underestimation in H_o).

Based on these considerations, the excess in H_o in all demes supports a hypothesis in which reproduction in cogongrass is highly asexual. Much of the observed genetic diversity among individuals within a purported clonal lineage may be due to the accumulation of somatic mutations over many mitotic generations of vegetative proliferation. We cannot make the conclusion that cogongrass reproduction is entirely asexual, as this study was limited to samples in the United States, Japan, the Philippine Islands and Brazil. In a population that is dominated by clonal reproduction, it is exceedingly difficult to detect sexual reproduction among members of the same clonal lineage (Halkett *et al.* 2005; Jacquemyn *et al.* 2006). However, theoretical considerations suggest that even a small amount of sexual reproduction would greatly reduce heterozygote excess observed (Balloux *et al.* 2003). In addition, we did not observe any hybridization between clonal lineages, which could be easily detected if present by PCoA (Hardig *et al.* 2000; Paul *et al.* 2010; Haselhorst & Buerkle 2013; Pujolar *et al.* 2014).

Cogongrass has two potential mechanisms for colonization: inflorescences containing thousands of plumed seeds and dense rhizomatous growth that is easily detached from a source plant (Zhang *et al.* 2010). For example, the horticultural variety, Japanese blood grass, which is facultatively invasive, is solely propagated via rhizomes (Coile & Shilling 1993; Cseke & Talley 2012). Over the years, seed viability of cogongrass has been reported in a spectrum ranging from 95% to 0% (Bryson & Carter 1993; King & Grace 2000; G. MacDonald, personal communication; Wilcut *et al.* 1988). A recent study showed significant reduction in seed head production and seed viability to be induced by repeated herbicide treatment (Aulakh *et al.* 2014), which is widely practiced in the United States. In addition, it is possible that the viable seeds that are produced arise asexually via apomixis, which is a process observed in its Saccharinae relatives. Additionally, cogongrass has been classified as an obligate outcrossing species—a characteristic often associated with clonality (Vallejo-Marin *et al.* 2010). These results, when considered along with our genetic data, strongly argue against sex as a primary mode of reproduction in cogongrass in its invaded US range as well as sampled international ranges.

Multiple clonal lineages of Imperata spp. in the United States

Our data showed that multiple invasive clonal lineages of cogongrass have been introduced into the United States. As of now, these remain limited primarily to the south-eastern region of the country, but show ecological potential for range expansion as evidenced by populations in NC, MD, MO and OR.

CL1 accessions were widely distributed across the invaded region ranging from NC as far west as TX, making it by far the most predominant of the five clonal lineages observed. When analysed with samples from the Honshu, Kyushu and Shikoku islands of Japan, both clonal assignment and PCoA indicate that these samples and those from CL1 in the United States share the same clonal lineage, supporting the anecdotal suggestions of southern Japan as the proximal origin of some introductions to the Gulf Coast states.

On the other hand, CL2, which is predominant in the peninsular region of FL and has more sporadic distribution in other states, was hypothesized from historical records to be an introduction from the Philippine Islands. While our PCoA showed that genotypes of CL2 do share some genetic similarity with samples from the Philippine Islands, the US and Filipino accessions are not of the same clonal lineage. However, this does not preclude that the proximal origin of CL2 is in the Philippine Islands at a site that has not yet been sampled.

US-collected accessions of CL4 share the same clonal lineage as samples collected on the island of Shikoku in Japan. Samples of members of CL1 and CL4 were collected from nearby infestations in the Kochi prefec-

ture of Shikoku Island, Japan. However, the PCoA data indicate that Japanese blood grass (CL4) is not likely a derivative of CL1. This finding is supported by major differences in genome size between Japanese blood grass and other cogongrass genotypes analysed (Table 2). These data indicate that multiple clonal lineages were likely introduced to Japan. This is supported by the deviation from HWE in a Japanese study employing 18 polymorphic microsatellites which resolved five clonal populations of *I. cylindrica* (Chiang *et al.* 2012). Like cogongrass, *I. brasiliensis* (CL5) appears to exhibit highly clonal reproduction. Further, *I. brasiliensis* accessions collected in southern FL, the state of Bahia in Brazil, and in the state of Espirito Santo (800 km to the south) are all elements of the same clonal lineage.

As we were able trace individual clonal lineages from suspected sites of origin in Japan and Brazil to the infested states in the United States, we concluded that based on the lack of genetic diversity within the introduced clonal lineages of North America and their international conspecifics, there is a great likelihood that the reproductive strategy of introduced propagules was clonal at the time of introduction and remained so upon range expansion in the United States.

Ongoing introductions and dispersal

It is difficult to enumerate how many introductions of cogongrass to the United States have occurred. New populations are reported each year, as various outreach programmes to make the public aware of this destructive species provide information on how to identify it, along with easy reporting mechanisms.

Relative to most previous studies of clonal organisms (e.g. using microsatellites), this work represents an orders of magnitude increase in the number of markers employed. As the number of markers and resolution increase, the traditional concepts of clones and MLGs lose their utility. For example, the number of MLGs and apparent clonal diversity both increase with the number of markers employed (Loxdale & Lushai 2003; Arnaud-Haond *et al.* 2005, 2007a). With the use of more than 2000 polymorphic markers, we obtained genetic resolution of every sampled individual (although ~2.7% of this variation was due to technical error). For this work, we used the term 'clonal lineage' to describe a group of highly related individuals, derived from a common ancestor through largely asexual processes. At this level of resolution, we were also able to discern possible sublineages of extremely high similarity and likely very recent common ancestry.

The values obtained by Mantel and Spearman analysis suggest that in both clonal lineages CL1 and CL2, genetic distances were not correlated with geographic distances. This finding is consistent with historical evidence of long-distance anthropogenic dispersal of asexual propagules (Baker 1974; Richards *et al.* 2006). Using our sequence-based data, we traced the dispersal of several of these highly similar sublineages, detected in discontiguous locations in Tyler County, TX (CL1, CL2) and Stanly County, NC (CL1). Based on the genetic similarity within these sublineages, we can presume that they are of very recent origin and have undergone rapid and extreme geographic dispersal. This pattern, along with the lack of correlation between genetic and geographic distances, is consistent with anthropogenic rather than natural modes of dispersal such as seed diffusion and underground growth of rhizomes (Lonsdale 1999). Large numbers of seeds and tough, sclerous rhizomes of cogongrass are known to be dispersed by mechanical equipment associated with activities such as mowing and timber harvesting (Ervin & Holly 2011). These propagules arrive in novel environments in which they can take advantage of the lack of established stands of cogongrass, typically the disturbed areas Baker described as environments favourable for weeds (Baker 1974; MacDonald 2004; Brewer 2008).

Due to limited genetic standing diversity within the US clonal lineages, it is doubtful that *Imperata* spp. possess the evolutionary potential to adapt to a novel and challenging environment. Because they appear to have overcome a number of environmental obstacles that would limit the growth of nonadapted plants (heavy metals, nutrient stress, heat, drought, saline soils, pathogens and mechanical disruption), they represent a truly 'general purpose [invasive] genotype' (Baker 1965).

Implications for biocontrol and management

Cogongrass is a significant threat to natural ecosystems, agriculture and silviculture throughout its introduced ranges (McDonald 2002; MacDonald 2004; Evans *et al.* 2007; Daneshgar *et al.* 2008). The resilience of this species in response to herbicides and other conventional control methods has made it the target of a search for biological controls. The best control agents are, by nature, highly host specific. Such agents (e.g. specialized phytophagous insects) are not only species specific, but are often ecotype and genotype specific. Thus, population-scale genetic analyses of an invasive species are necessary to determine its suitability as a candidate for biological control, and to identify effective and host-specific biological control agents (Arriola & Ellstrand 1996; Ellstrand & Schierenbeck 2000; Allendorf & Lundquist 2003). Experience has shown that the selection of highly specific, yet highly effective biological control agents requires careful matching of the agent to its target, using geographical, ecological and genetic data (Goo-

lsby *et al.* 2006; Friedman *et al.* 2008; Manrique *et al.* 2008; Tarin *et al.* 2013). The strong genetic evidence that cogongrass is highly clonal and has limited genetic diversity in its suspected ranges of origin makes it a promising candidate for biological control, particularly because it has limited evolutionary potential to develop resistance to control agents (Burdon & Marshall 1981). This work provides a foundation for genetically identifying the ultimate geographic sources of invasive cogongrass to aid in the search for effective biological control agents.

A case for comparative genomics for nonmodel organisms

It is imperative to distinguish sequencing reads generated from the organism of interest from sequencing reads obtained from potential contaminating organismal DNA, which has potential to be a significant source of error in the form of: (i) 'novel' polymorphisms as well as apparent private alleles in isolated populations, (ii) inflated estimates of genetic diversity and (iii) a source of missing data (assuming samples from different geographical locations would be subject to exposure to dissimilar organisms of endemic origin based on varying habitat conditions). With the *S. bicolor* genome, we were able to use comparative genomics to readily eliminate spurious marker data points from nontarget organisms. While a reference genome is not required for SNP discovery of GBS data, mapping of sample reads to related organisms is necessary in order to eliminate sources of error and subsequent erroneous conclusions, as evidenced in the non-*Imperata* reads obtained in our marker system.

Cogongrass as a model for clonal weed genomics

As a diploid plant with a comparatively compact genome, cogongrass has attributes to become a useful model organism for the study of clonal invasive plants (Stewart *et al.* 2009). The significant molecular and genomic resources of its well-studied relatives in the Saccharinae, which share conservation of gene order in both nuclear and organellar genomes (Moore *et al.* 1995; Kim *et al.* 2005; Paterson *et al.* 2009; Ma *et al.* 2012), will be of continued utility in the study of *Imperata*. Remarkably, the most invasive relatives of *I. cylindrica* are polyploid (*Sorghum halepense* (L.) Pers. and *Saccharum spontanuem* L.), with larger and more complex genomes, which pose greater genomic challenges. The polyploidization of these two species could be potentially the result of an adaptive evolutionary event to fix advantageous heterosis through a shift to asexual rhizomatous colonization (Ellstrand & Schierenbeck 2000; Grivet & Arruda 2002; Paterson 2008), whereas our data illustrate such adaptive events are absent in United States accessions of *I. cylindrica*. The study of the genetic architecture underlying its extreme invasive success will be valuable to understanding the functional characteristics of a strong candidate for Baker's 'ideal weed'.

Acknowledgements

The authors wish to acknowledge the extraordinary generosity in sample collection by members of the Alabama Forestry Commission, Clemson Department of Plant Industries, the Florida Forest Service, the Georgia Forestry Commission, the Louisiana Forestry Commission, the Mississippi Forestry Commission and the Texas A&M Forest Service. We thank Jim Walley, Wallace Allred, Leland Cseke and Daniel Tarin for helpful discussions and sample material. We thank Natalie Patterson for help with DNA extraction and Illumina template library preparation. The authors also thank L.J. Grauke (USDA-ARS, College Station) and R. Daniel Lineberger for help with arc-GIS figure preparation. We would also like to thank the three anonymous reviewers for their valuable comments and suggestions. This work was supported by USDA-NIFA grant #2012-67013-19340 and by a Texas A&M AgriLife Research Genomics Seed Grant.

References

Al-Janabi SM, McClelland M, Petersen C, Sobral BWS (1994) Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. *Theoretical and Applied Genetics*, **88**, 933–944.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Allendorf F, Lundquist L (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology*, **17**, 24–30.

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

Anderson JB, Kohn LM (1995) Clonality in soilborne, plant-pathogenic fungi. *Annual review of Phytopathology*, **33**, 369–391.

Arnaud-Haond S, Alberto F, Teixeira S *et al.* (2005) Assessing genetic diversity in clonal organisms: low diversity or low resolution? Combining power and cost efficiency in selecting markers. *Journal of Heredity*, **96**, 434–440.

Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA (2007a) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.

Arnaud-Haond S, Migliaccio M, Diaz-Almela E *et al.* (2007b) Vicariance patterns in the Mediterranean Sea: east–west cleavage and low dispersal in the endemic seagrass *Posidonia oceanica*. *Journal of Biogeography*, **34**, 963–976.

Arriola PE, Ellstrand NC (1996) Crop-to-weed gene flow in the genus Sorghum (Poaceae): spontaneous interspecific hybridization between johnsongrass, *Sorghum halepense*, and crop Sorghum, *S. bicolor*. *American Journal of Botany*, **83**, 1153–1159.

Aulakh JS, Enloe SF, Loewenstein NJ *et al.* (2014) Pushing toward cogongrass (Imperata cylindrica) patch eradication: the influence of herbicide treatment and application timing on cogongrass rhizome elimination. *Invasive Plant Science and Management*, **7**, 398–407.

Baker HG (1965) Characteristics and modes of origin of weeds. In: *The Genetics of Colonizing Species* (eds Baker HG, Stebbins GL), pp. 147–168. Academic Press Inc., New York.

Baker HG (1974) The evolution of weeds. *Annual Review of Ecology and Systematics*, **5**, 1–24.

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.

Barrett SCH, Colautti RI, Eckert CG (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology*, **17**, 373–383.

Barry D (2009) Weed heroes: the war on the invader cogongrass. *The New York Times*. The New York Times Company, New York.

Bouchenak-Khelladi Y, Salamin N, Savolainen V *et al.* (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution*, **47**, 488–505.

Brewer S (2008) Declines in plant species richness and endemic plant species in longleaf pine savannas invaded by *Imperata cylindrica*. *Biological Invasions*, **10**, 1257–1264.

Brown RH (1978) A difference in N use efficiency in C3 and C4 plants and its implications in adaptation and evolution. *Crop Science*, **18**, 93–98.

Bryson CT, Carter R (1993) Cogongrass, *Imperata cylindrica*, in the United States. *Weed Technology*, **7**, 1005–1009.

Bryson CT, Krutz LJ, Ervin GN, Reddy KN, Byrd JD (2010) Ecotype variability and edaphic characteristics for Cogongrass (*Imperata cylindrica*) populations in Mississippi. *Invasive Plant Science and Management*, **3**, 199–207.

Burdon JJ, Marshall DR (1981) Biological control and the reproductive mode of weeds. *Journal of Applied Ecology*, **18**, 649–658.

Burrell M (2009) Development of molecular resources for the noxious weed, *Imperata cylindrica* (L.) Beauv. In: *Texas Invasive Pest and Plant Conference*, San Antonio, Texas.

Capo-chichi LJA, Faircloth WH, Williamson AG *et al.* (2008) Invasion dynamics and genotypic diversity of cogongrass (*Imperata cylindrica*) at the point of introduction in the southeastern United States. *Invasive Plant Science and Management*, **1**, 133–141.

Cariou M, Duret L, Charlat S (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, **3**, 846–852.

Cavalier-Smith T (ed.) (1985) Eukaryote gene numbers, noncoding DNA and genome size. In: *The Evolution of Genome Size*, pp. 69–103. Wiley, Chichester.

Chao A, Shen T-J (2003) Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics*, **10**, 429–443.

Chapman G (1992) Apomixis and evolution. In: *Grass Evolution and Domestication* (ed. Chapman GP), pp. 138–155. Cambridge University Press, Cambridge.

Chiang Y-C, Tsai C-C, Hsu T-W, Chou C-H (2012) Characterization of 21 microsatellite markers from cogongrass, Imperata cylindrica (Poaceae), a weed species distributed worldwide. *American Journal of Botany*, **9**, 428–430.

Clifton-Brown J, Chiang Y-C, Hodkinson TR (2008) Miscanthus: genetic resources and breeding potential to enhance bioenergy production. In: *Genetic Improvement of Bioenergy Crops* (ed. Vernmerris W), pp. 273–294. Springer, New York.

Coile N, Shilling D (1993) Cogongrass, *Imperata cylindrica* (L.) Beauv.: a good grass gone bad. *Botany Circular*, **28**, 1–3.

Conover WJ, Iman RL (1981) Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, **35**, 124–129.

Cornuet JM, Luikart G (1996) Description and power of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Craven D, Hall J, Verjans J-M (2009) Impacts of herbicide application and mechanical cleanings on growth and mortality of two timber species in *Saccharum spontaneum* grasslands of the Panama Canal Watershed. *Restoration Ecology*, **17**, 751–761.

Cseke LJ, Talley SM (2012) A PCR-based genotyping method to distinguish between wld-type and ornamental varieties of *Imperata cylindrica*. *Journal of Visualized Experiments: JoVE*, **60**, 3625.

Daneshgar P, Jose S (2009) *Imperata cylindrica*, an alien invasive grass, maintains control over nitrogen availability in an establishing pine forest. *Plant and Soil*, **320**, 209–218.

Daneshgar P, Jose S, Collins A, Ramsey C (2008) Cogongrass (*Imperata cylindrica*), an alien invasive grass, reduces survival and productivity of an establishing pine forest. *Forest Science*, **54**, 579–587.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Dickens R (1974) Cogongrass in Alabama after sixty years. *Weed Science*, **22**, 177–179.

Dickens R, Buchanan G (1971) Old weed in a new home-that's cogongrass. In: *Highlights of Agricultural Research* (ed. University A), 4 pp. Auburn University, Auburn, Alabama.

Doncaster CP, Graeme EP, Cox SJ (2000) The ecological cost of sex. *Nature*, **404**, 281–285.

Douhovnikoff V, Dodd R (2003) Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. *Theoretical and Applied Genetics*, **106**, 1307–1315.

Dozier H, Gaffney JF, McDonald SK, Johnson ERRL, Shilling DG (1998) Cogongrass in the United States: history, ecology, impacts, and management. *Weed Technology*, **12**, 737–743.

Duara B, Stebbins G (1952) A polyhaploid obtained from a hybrid derivative of *Sorghum halepense* x *S. vulgare* var. *sudanense*. *Genetics*, **37**, 369.

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences*, **97**, 7043–7050.

Ervin G, Holly DC (2011) Anthropogenic dispersal corridors override large-scale natural disturbance in determining distribution of a widespread invasive grass (*Imperata cylindrica*). In: *Plant Invasions: Policies, Politics and Practices* (ed. Rindos E), pp. 60–69. Center for Invasive Plant Management, Shepherdstown, West Virginia.

ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, California.

Evans CW, Moorhead DJ, Bargeron CT, Douce GK (2007) Implementation of control and prevention strategies for man-

aging cogongrass (*Imperata cylindrica*) by the Georgia Invasive Species Task Force. *Natural Areas Journal*, **27**, 226–231.

Falk D, Miller C, McKenzie D, Black A (2007) Cross-scale analysis of fire regimes. *Ecosystems*, **10**, 809–823.

Friedman JM, Roelle JE, Gaskin JF, Pepper AE, Manhart JR (2008) Latitudinal variation in cold hardiness in introduced *Tamarix* and native *Populus*. *Evolutionary Applications*, **1**, 598–607.

Goolsby JA, De Barro PJ, Makinson JR *et al.* (2006) Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. *Molecular Ecology*, **15**, 287–297.

Grace JB, Smith M, Grace SL, Collins S, Stohlgren TJ (2001) Interactions between fire and invasive plants in temperate grasslands in North America. In: *Fire Conference 2000: The First National Congress on Fire, Ecology, Prevention and Management. Invasive Species Workshop: The Role of Fire in the Control and Spread of Invasive Species* (eds Galley K, Wilson T), pp. 40–65. Allen Press, Lawrence, Kansas.

Grivet L, Arruda P (2002) Sugarcane genomics: depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology*, **5**, 122–127.

Halkett F, Simon J-C, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, **20**, 194–201.

Hall DW (1998) Is cogongrass really an exotic? *Wildland Weeds*, **1**, 14.

Hardig TM, Brunsfeld SJ, Fritz RS *et al.* (2000) Morphological and molecular evidence for hybrdization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology*, **9**, 9–24.

Haselhorst MSH, Buerkle CA (2013) Population genetic structure of *Picea engelmannii*, *P. glauca* and their previously unrecognized hybrids in the central Rocky Mountains. *Tree Genetics and Genomes*, **9**, 669–681.

Hodkinson T, Chase M, Lledó D, Salamin N, Renvoize S (2002) Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL*, intron and *trnL-F* intergenic spacers. *Journal of Plant Research*, **115**, 381–392.

Hodoki Y, Ohbayashi K, Kunii H (2014) Analysis of population clonal diversity using microsatellite markers in the salt marsh sedge *Carex scabrifolia* in western Japan. *Landscape and Ecological Engineering*, **10**, 9–15.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in Threespine Stickleback using sequenced RAD Tags. *PLoS Genetics*, **6**, e1000862.

Hollingsworth ML, Bailey JP (2000) Evidence for massive clonal growth in the invasive weed *Fallopia japonica* (Japanese knotweed). *Botanical Journal of the Linnaean Society*, **133**, 463–472.

Hubbard CE, Whyte RO, Brown D, Gray AP (1944) *Imperata cylindrica*. Taxonomy, distribution, economic significance and control. Imperial Agricultural Bureaux Joint Publication No. 7, London, UK

Jacquemyn H, Brys R, Honnay O, Hermy M, Roldán-Ruiz I (2006) Sexual reproduction, clonal diversity and genetic differentiation in patchily distributed populations of the temperate forest herb *Paris quadrifolia* (Trilliaceae). *Oecologia*, **147**, 434–444.

Johnston JS, Bennett MD, Rayburn AL, Galbraith DW, Price HJ (1999) Reference standards for determination of DNA content of plant nuclei. *American Journal of Botany*, **86**, 609–613.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Joo Kim T, Montagnini F, Dent D (2008) Rehabilitating abandoned pastures in Panama: control of the invasive exotic grass, *Saccharum spontaneum* L., using artificial shade treatments. *Journal of Sustainable Forestry*, **26**, 192–203.

Jose S, Cox J, Miller DL *et al.* (2002) Alien plant invasions: the story of cogongrass in southeastern forests. *Journal of Forestry*, **100**, 41–44.

Kamvar ZN, Tabima JF, Grunwald NJ (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics*, **175**, 1823–1834.

Keser LH, Dawson W, Song Y-B *et al.* (2014) Invasive clonal plant species have a greater root-foraging plasticity than non-invasive ones. *Oecologia*, **174**, 1055–1064.

Kim J-S, Islam-Faridi M, Klein PE *et al.* (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. *Genetics*, **171**, 1963–1976.

King SE, Grace JB (2000) The effects of gap size and disturbance type on invasion of wet pine savanna by cogongrass, *Imperata cylindrica* (Poaceae). *American Journal of Botany*, **87**, 1279–1286.

Klekowski EJ (1997) Somatic mutation theory of clonality. In: *The Ecology and Evolution of Clonal Plants* (eds deKroon H, van Groenendael J), pp. 227–242. Backhuys Publishers, Leiden, the Netherlands.

Lambert DM, Millar CD (1995) DNA science and conservation. *Pacific Conservation Biology*, **2**, 21–38.

Larrinua IM, Belmar SB (2008) Bioinformatics and its relevance to weed science. *Weed Science*, **56**, 297–305.

Larson DL, Phillips-Mao L, Quiram G *et al.* (2011) A framework for sustainable invasive species management: environmental, social, and economic objectives. *Journal of Environmental Management*, **92**, 14–22.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences*, **104**, 3883–3888.

Law R (1979) The cost of reproduction in annual meadow grass. *American Naturalist*, **113**, 3–16.

Lonsdale WM (1999) Global patterns of plant invasions and the concept of invasibility. *Ecology*, **80**, 1522–1536.

Loureiro J, Rodriguez E, Doležel J *et al.* (2007) Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany*, **100**, 875–888.

Loxdale HD, Lushai G (2003) Rapid changes in clonal lines: the death of a 'sacred cow'. *Biological Journal of the Linnean Society*, **79**, 3–16.

Lucardi RD, Wallace LE, Ervin GN (2013) Invasion success in Cogongrass (*Imperata cylindrica*): a population genetic approach exploring genetic diversity and historical introductions. *Invasive Plant Science and Management*, **7**, 59–75.

Lucardi RD, Wallace LE, Ervin GN (2014) Evaluating hybridization as a potential facilitator of successful cogongrass (*Imperata cylindrica*) invasion in Florida, USA. *Biological Invasions*, **16**, 2147–2161.

Ma X-F, Jensen E, Alexandrov N *et al.* (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS ONE*, **7**, e33821.

MacDonald GE (2004) Cogongrass (*Imperata cylindrica*)–biology, ecology, and management. *Critical Reviews in Plant Sciences*, **23**, 367–380.

Maeda J, Kaneko S, Isagi Y *et al.* (2009) Isolation and characterization of polymorphic microsatellite loci for *Imperata cylindrica*, an invasive perennial grass. *Conservation Genetics Resources*, **1**, 127–129.

Manrique V, Cuda J, Overholt W *et al.* (2008) Effect of host-plant genotypes on the performance of three candidate biological control agents of *Schinus terebinthifolius* in Florida. *Biological Control*, **47**, 167–171.

McDonald C (2002) Cogongrass: interesting facts about the perfect weed. *Wildlife Issues*, **3**, 10–16.

deMeeûs T, Balloux F (2004) Clonal reproduction and linkage disquilibrium in diploids: a simulation study. *Infection, Genetics and Evolution*, **4**, 345–351.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Meloni M, Reid A, Caujapé-Castells J *et al.* (2013) Effects of clonality on the genetic variability of rare, insular species: the case of *Ruta microcarpa* from the Canary Islands. *Ecology and Evolution*, **3**, 1569–1579.

Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution: grasses, line up and form a circle. *Current Biology*, **5**, 737–739.

Morishige DT, Klein PE, Hilley JL *et al.* (2013) Digital genotyping of sorghum–a diverse plant species with a large repeat-rich genome. *BMC Genomics*, **14**, 448.

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution*, **14**, 389–394.

Murray SC, Sharma A, Rooney WL *et al.* (2008) Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Science*, **48**, 2165–2179.

Narum SR, Buerkle CA, Davey JW *et al.* (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Panje R (1970) The evolution of a weed. *PANS Pest Articles & News Summaries*, **16**, 590–595.

Pappert RA, Hamrick J, Donovan LA (2000) Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany*, **87**, 1240–1245.

Paterson AH (2008) Genomics of sorghum. *International Journal of Plant Genomics*, Article ID 362451, doi:10.1155/2008/362451.

Paterson AH, Bowers JE, Bruggmann R *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Patterson DT, Flint EP, Dickens R (1980) Effects of temperature, photoperiod, and population source on the growth of cogongrass (*Imperata cylindrica*). *Weed Science*, **28**, 505–509.

Paul J, Vachon N, Garroway CJ, Freeland JR (2010) Molecular data provide strong evidence of natural hybridization between natural and introduced lineages of *Phragmites australis* in North America. *Biological Invasions*, **9**, 2967–2973.

Paz-Alberto AM, Sigua GC, Baui BG, Prudente JA (2007) Phytoextraction of lead-contaminated soil using vetivergrass (*Vetiveria zizanioides* L.), cogongrass (*Imperata cylindrica* L.) and carabaograss (*Paspalum conjugatum* L.). *Environmental Science and Pollution Research-International*, **14**, 498–504.

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288.

Peterson AT, Papes M, Kluza DA (2003) Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*, **51**, 863–868.

Price HJ, Dillon SL, Hodnett G *et al.* (2005) Genome evolution in the genus *Sorghum* (Poaceae). *Annals of Botany*, **95**, 219–227.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prugnolle F, De Meeûs T (2008) The impact of clonality on parasite population genetic structure. *Parasite*, **15**, 455–457.

Pujolar JM, Jacobsen M, Als TD *et al.* (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity*, **112**, 627–637.

Rees M (1993) Trade-offs among dispersal strategies in British plants. *Nature*, **366**, 150–152.

Richard EP Jr (1998) Control of perennated bermudagrass (*Cynodon dactylon*) and Johnsongrass (*Sorghum halepense*) in sugarcane (*Saccharum spp.* hybrids). *Weed Technology*, **12**, 128–133.

Richards CL, Bossdorf O, Muth NZ (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters*, **9**, 981–993.

Roderick GK, Navajas M (2003) Genes in new environments: genetics and evolution in biological control. *Nature Reviews Genetics*, **4**, 889–899.

Rodriguez N, Menendez N, Tornero J, Amils R, De la Fuente V (2005) Internal iron biomineralization in *Imperata cylindrica*, a perennial grass: chemical composition, speciation and plant localization. *New Phytologist*, **165**, 781–789.

Rogstad SH, Keane B, Beresh J (2002) Genetic variation across VNTR loci in central North American *Taraxacum* surveyed at different spatial scales. *Plant Ecology*, **161**, 111–121.

Roodt-Wilding R, Spies JJ (2006) Phylogenetic relationships in Southern African chloridoid grasses (Poaceae) based on nuclear and chloroplast sequence data. *Systematics and Biodiversity*, **4**, 401–415.

Rout ME, Chrzanowski TH, Smith WK, Gough L (2013) Ecological impacts of the invasive grass *Sorghum halepense* on native tallgrass prairie. *Biological Invasions*, **15**, 327–339.

de Setta N, Monteiro-Vittorello CB, Metcalfe CJ *et al.* (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics*, **15**, 540.

Sharma MK, Sharma R, Cao P *et al.* (2012) A genome-wide survey of switchgrass genome structure and organization. *PLoS ONE*, doi:10.1371/journal.pone.0033892.

Silvertown J (2005) *Demons in Eden*. Chicago University Press, Chicago and London.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Stewart CN, Tranel PJ, Horvath DP *et al.* (2009) Evolution of weediness and invasiveness: charting the course for weed genomics. *Weed Science*, **57**, 451–462.

Stoeckel S, Grange J, Fernandez-Manjarres JF *et al.* (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species —*Prunus avium* L. *Molecular Ecology*, **15**, 2109–2118.

Stone JL (2002) Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *The Quarterly Review of Botany*, **77**, 17–32.

Tarin D, Pepper AE, Goolsby JA *et al.* (2013) Microsatellites uncover multiple introductions of clonal giant reed (*Arundo donax*). *Invasive Plant Science and Management*, **6**, 328–338.

Templeton AR (2006) *Population Genetics and Microevolutionary Theory*. Wiley, New York.

Vallejo-Marin M, Dorken ME, Barrett SCH (2010) The ecological and evolutionary consequences of clonality for plant mating. *Annual Review of Ecology Evolution and Systematics*, **41**, 193–213.

Vergara R, Minno MC, Minno M, Soltis DE, Soltis PS (2008) Preliminary study using ISSRs to differentiate *Imperata* Taxa (Poaceae: Andropogoneae) growing in the United States. *Southeastern Naturalist*, **7**, 267–276.

Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM (2010) Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal*, **8**, 263–276.

Wang Y-H, Acharya A, Burrell AM *et al.* (2013) Mapping and candidate genes associated with saccharification yield in sorghum. *Genome*, **56**, 659–665.

Warwick S, Black L (1983) The biology of Canadian weeds: 61. *Sorghum halepense* (L.) PERS. *Canadian Journal of Plant Science*, **63**, 997–1014.

Whitney KD, Randell RA, Rieseberg LH (2006) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist*, **167**, 794–807.

Wilcut JW, Roland RD, Truelove B, Davis DE (1988) Factors limiting the distribution of cogongrass, *Imperata cylindrica*, and torpedograss, *Panicum repens*. *Weed Science*, **36**, 577–582.

Willard TR, Hall DW, Shilling DG, Lewis JA, Currey WL (1990) Cogongrass (*Imperata cylindrica*) distribution on Florida highway rights-of-way. *Weed Technology*, **4**, 658–660.

Willing E-M, Hoffmann M, Klein JD, Weigel D, Dreyer C (2011) Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics*, **27**, 2187–2193.

Xuan TD, Toyama T, Fukuta M, Khanh TD, Tawata S (2009) Chemical interaction in the invasiveness of cogongrass (*Imperata cylindrica* (L.) Beauv.). *Journal of Agricultural and Food Chemistry*, **57**, 9448–9453.

Yager LY, Miller DL, Jones J (2010) Susceptibility of longleaf pine forest associations in south Mississippi to invasion by cogongrass [*Imperata cylindrica* (L.) Beauv.]. *Natural Areas Journal*, **30**, 226–232.

You W, Fen S, Yu D *et al.* (2014) An invasive clonal plant benefits from clonal integration more than a co-occurring native plant in nutrient-patchy and competitive enviroments. *PLoS ONE*, **9**, e97246.

Zhang YYE, Zhang DAY, Barrett SCH (2010) Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichornia crassipes*), a clonal acquatic plant. *Molecular Ecology*, **19**, 1774–1786.

M.B. conceived of and outlined the project with input from P.K., A.P. and J.G. M.B., A.P., G.H. and P.K. produced and analysed the data. M.B., P.K., J.G., A.R., R.D. and W.O. conducted field surveys and sampling. M.B. and A.P. drafted the manuscript. All authors contributed to and approved the final manuscript.

Data accessibility

SNP genotype data, accession collection data and the custom perl and python scripts used in the bioinformatic processing of this project can be found in the Dryad Digital Depository, doi:10.5061/dryad.k2d05.

Sequence files available at the NCBI Short Read Archive: BioProject SRP056365.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Stacked frequency distribution of pairwise genetic distances between *Imperata cylindrica* samples from eight sites of mixed infestation by CL1 and CL2.

Fig. S2 Principal Components Analysis of *Imperata cylindrica* samples from eight sites of mixed infestation by CL1 and CL2. Samples in teal are assigned to CL1.

Table S1 Study-wide analysis of population genetic parameters (*n* = 449 samples).

Appendix S1 Supporting technical information.