Citation

Muller MF, Barnes I, Kunene NT, Crampton BG, Bluhm BH, Phillips SM, Olivier
    NA, Berger DK (2016) *Cercospora zeina* from Maize in South Africa Exhibits
    High Genetic Diversity and Lack of Regional Population Differentiation.
    Phytopathology 106 (10):1194-1205.
    doi:10.1094/PHYTO-02-16-0084-FI
    http://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO-02-16-0084-FI

Cercospora zeina **from maize in South Africa exhibits high genetic diversity and lack of regional population differentiation**

Mischa F. Muller, Irene Barnes, Ncobile T. Kunene, Bridget G. Crampton, Burton H. Bluhm, Sonia M. Phillips, Nicholas A. Olivier and Dave K. Berger

First, third, fourth, sixth, seventh and eighth author(s): Department of Plant and Soil Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0028, South Africa; Second author: Department of Genetics, FABI, University of Pretoria, Pretoria 0028, South Africa. Fifth author: Department of Plant Pathology, University of Arkansas, USA. Seventh author: Centre for Bioinformatics and Computational Biology, Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa

ABSTRACT

Muller, M.F., Barnes, I., Kunene, N.T., Crampton, B.G., Bluhm, B.H., Phillips, S., Olivier, N.A. and Berger, D.K. 2016. *Cercospora zeina* from maize in South Africa exhibits high genetic diversity and lack of regional population differentiation. Phytopathology XX: XX-XX.

South Africa is one of the leading maize-producing countries in sub-Saharan Africa. Since the 1980s, *Cercospora zeina*, a causal agent of gray leaf spot (GLS) of maize, has become endemic in South Africa, and is responsible for substantial yield reductions. To assess genetic diversity and population structure of *C. zeina* in South Africa, 369 isolates were collected from commercial maize farms in three provinces (KwaZulu-Natal, Mpumalanga and North West). These isolates were evaluated with fourteen microsatellite markers and species-specific mating type markers that were designed from draft genome sequences of *C. zeina* isolates from Africa (CMW 25467) and USA (USPA-4). Sixty alleles were identified across 14 loci, and gene diversity values within each

province ranged from 0.18 to 0.35. High levels of gene flow were observed (Nm = 5.51), and in a few cases, identical multilocus haplotypes were found in different provinces. Overall, 242 unique multilocus haplotypes were identified with a low clonal fraction of 34%. No distinct population clusters were identified using STRUCTURE, Principle Co-ordinate analysis or Weir's theta θ statistic. The lack of population differentiation was supported with AMOVA analyses which indicated that only 2% of the variation was attributed to variability between populations from each province. Mating-type ratios of *MAT1-1* and *MAT1-2* idiomorphs from 335 isolates were not significantly different from a 1:1 ratio in all provinces, which provided evidence for sexual reproduction. The draft genome of *C. zeina* CMW 25467 exhibited a complete genomic copy of the *MAT1-1* idiomorph as well as exonic fragments of *MAT* genes from both idiomorphs. The high level of gene diversity, shared haplotypes at different geographical locations within South Africa, and presence of both *MAT* idiomorphs at all sites indicates widespread dispersal of *C. zeina* between maize fields in the country as well as evidence for sexual recombination. The outcomes of this genome-enabled study are important for disease management since the high diversity has implications for dispersal of fungicide resistance should it emerge and the need for diversified resistance breeding.

INTRODUCTION

Maize (*Zea mays*) is the most important staple crop in sub-Saharan Africa and the Americas (Ranum et al. 2014). Foliar diseases of maize result in yield losses that threaten food production, food security and agricultural economies on a global scale (Shiferaw et al. 2011). Disease management requires knowledge of farming practises, the environment, host resistance and the pathogen. Chemical control is an important component of disease management. Maize foliar fungicide applications have increased in the USA over the past decade (Mallowa et al. 2015), and anecdotal evidence suggests that applications have increased in other maize production areas around the globe, including South Africa. Widespread adoption of conservation tillage has allowed many fungal foliar pathogens to over-winter on maize debris, thus increasing the inoculum pressure in subsequent growing seasons (Ward et al. 1999). Maize foliar disease resistance breeding is ongoing (Berger et al. 2014), and this is particularly important for small-holder farmers in developing countries who currently have limited access to chemical control.

GLS is a globally important foliar disease of maize (Meisel et al. 2009; Ward et al. 1999). The disease was initially associated with two sibling species named *Cercospora zeae-maydis* type I and type II, which were distinguished by distinct AFLP patterns and had seven nucleotide point differences in their ITS sequences (Dunkle and Levy 2000; Wang et al. 1998). The two organisms were subsequently reclassified as *Cercospora zeae-maydis* Tehon & E.Y. Daniels and *Cercospora zeina* Crous & U.Braun, respectively, based on further molecular and morphological analysis (Crous et al. 2006), although the symptoms they cause on maize are indistinguishable. *Cercospora sorghi var.-maydis* Ellis & Everh has also been isolated occasionally from diseased maize leaves (Crous et al. 2006; Neves et al. 2015), but it is not clearly established whether this organism is a pathogen or a saprophyte in GLS lesions.

C. zeae-maydis and *C. zeina* have both been documented in the Americas and Asia, but only *C. zeina* has been identified in Africa to date (Crous et al. 2006; Dunkle and Levy 2000; Liu and Xu 2013; Meisel et al. 2009; Ward et al. 1999). *C. zeina* was first reported from Africa in 1988 in KwaZulu-Natal Province, South Africa, where yield losses of 65% were reported (Ward et al.

1999). Subsequently, *C. zeina* was reported from Uganda, Zimbabwe, Zambia and Kenya (Dunkle and Levy 2000; Goodwin et al. 2001). Isolates collected nearly a decade later in South Africa were confirmed to be *C. zeina*, and *C. zeae-maydis* was not found to be present (Meisel et al. 2009).

Cercospora zeae-maydis is considered to be the predominant pathogen in the USA, although it's range overlaps with *C. zeina* in the eastern cornbelt (Wang et al. 1998). All African isolates collected by Dunkle and Levy (2000) were identified as *C. zeina*, and AFLP analysis indicated high similarity with *C. zeina* isolates from USA. Both species have been reported from tropical and subtropical regions of Brazil (Brunelli et al. 2008; Neves et al. 2015), although *C. zeina* was most prevalent. In China, only *C. zeina* was isolated from GLS lesions of maize from 25 sites in the southern province of Yunnan (Liu and Xu 2013), although *C. zeae-maydis* was previously recorded from China (Dunkle and Levy 2000).

The small number of population genetics studies of *Cercospora* species completed to date have reported high genetic diversity. Dunkle and colleagues developed ten microsatellite markers from expressed sequence tag (EST) data of *C. zeae-maydis,* and observed high genetic diversity within a population from Indiana, USA (Molecular Ecology Resources Primer Development Consortium, 2009). Interestingly, the markers did not amplify in samples of *C. zeina*. *Cercospora beticola* is the causal agent of *Cercospora* leaf spot on sugar beet: five *C. beticola* microsatellite markers were found to distinguish between isolates from six populations in Europe, Iran and New Zealand (Groenewald et al. 2007). *Cercospora sojina* isolates from soybean in Arkansas, USA, exhibited high genotypic diversity and equal ratios of mating type genes, suggestive of cryptic sexual recombination (Kim et al. 2013).

Currently, there is little knowledge of the genetic diversity and population dynamics of *C. zeina* in Africa or other maize growing regions (Dunkle and Levy 2000). This is largely due to the lack of markers available to study the pathogen. Information about genetic diversity is important since resistance breeding, fungicide applications, and regional environmental conditions can have selective effects on pathogen populations. Improved control strategies can be implemented once

there is understanding of the genetic diversity, mode of reproduction, and life history on host plants (McDermott and McDonald 1993; McDonald 1997).

Fungi can reproduce sexually and/or asexually, with sexual reproduction being controlled by the mating-type (MAT) locus.. Ascomycetes have two idiomorphs of the MAT locus, *MAT1-1* or *MAT1-2*, characterized by *MAT1-1-1* or *MAT1-2-1* genes encoding proteins with an alpha box or HMG domain, respectively (Coppin et al. 1997). Heterothallic fungal species are defined as those that have both *MAT* idiomorphs but individuals have either one or the other *MAT* idiomorph, whereas species where individuals have both *MAT* idiomorphs are homothallic. In heterothallic fungi, the ratios of *MAT* idiomorphs within a population can be used to gauge the predominant mode of reproduction. Equal proportions of *MAT* idiomorphs are expected in sexually reproducing populations. Uneven *MAT* idiomorph ratios are expected in populations that are reproducing asexually, populations that have recently undergone a genetic bottleneck or recently introduced populations (Barnes et al. 2014; Milgroom 1996). Hypotheses for the presence of the MAT locus in fungi that appear to be reproducing asexually may be that the sexual stage has not yet been observed or that the locus confers a fitness benefit (Arie et al. 2000; Paoletti et al. 2005; Stergiopoulos et al. 2007).

As sexual structures have not been documented in *Cercospora* species, fungi in this genus were believed to reproduce asexually (Groenewald et al. 2006). However, in previous studies performed on *C. beticola*, *C. zeae-maydis, C zeina* and *C. sojina*, isolates have been shown to be heterothallic and exhibit equal proportions of both *MAT* idiomorphs in their respective populations (Bolton et al. 2012; Groenewald et al. 2006; Groenewald et al. 2008; Kim et al. 2013). Sexual reproduction recombines genetic variation within a population, which in turn enhances the fitness and environmental adaptation (Groenewald et al. 2008). Thus, the potential for sexual reproduction should be considered when developing disease management strategies.

This study was prompted by the increasing importance of GLS as a foliar disease of maize in South Africa (Meisel et al. 2009), coupled with the lack of information about the genetic diversity or population structure of *C. zeina*. A previous small-scale study of 49 isolates of *C. zeina* from

KwaZulu-Natal province inferred the potential for cryptic sexual reproduction based on *MAT* gene ratios (Groenewald et al. 2006), however neutral marker frequencies were not assessed. In the current study, we analysed a larger collection of isolates from multiple provinces, and assessed genetic diversity and the possibility of cryptic sexual recombination with microsatellite markers and *MAT* gene analysis. We hypothesized that this approach would provide greater resolution of population dynamics and would ultimately provide support for one of two possible theories underlying the relatively recent emergence of GLS as a disease of importance in South Africa. First, evidence for a predominantly clonal population across the country would support the hypothesis of a single introduction through Durban harbour in the 1980's (Ward et al. 1999). Alternatively, high levels of genetic diversity, regional differentiation, and equal *MAT* gene ratios would be more consistent with cryptic sexual recombination, multiple introductions into South Africa, or a more ancient origin of *C. zeina* on the African continent (Crous et al. 2006; Groenewald et al. 2006).

Genome sequence-enabled discovery of microsatellites holds advantages over selective amplification or genomic library-based methods (Santana et al. 2009), since it relies predominantly on bioinformatics approaches. Therefore, the aim of this study was to utilize recently constructed draft genome sequences for isolates of *C. zeina* from the USA and Africa to develop microsatellite and mating type markers for population genetic analyses of populations of *C. zeina* from South Africa.

MATERIALS AND METHODS

Sample collection, isolations and DNA extraction. Maize leaves presenting GLS symptoms were collected from maize growing areas in KwaZulu-Natal (Baynesfield, Cedara, Greytown and Winterton), Mpumalanga (Machadodorp) and the North West Province (Klerksdorp) in South Africa (Fig. 1), over three seasons (2011-2013). Two leaves each, from fifty plants per field, were collected using a diagonal sampling design method and stored individually in paper envelopes. The diagonal sampling design involved sampling in a cross formation at each field; starting at one

corner of the field and sampling from 25 plants at regular intervals until the opposite corner, and then repeating the process between the other two corners. Conidia, typical of *Cercospora species*, were identified from maize leaf material under a dissecting microscope, picked with a sterile hypodermic needle, and transferred to V8 agar plates containing 100ug/ml carbenicillin. Isolates were cultured for 10 days at 25°C in the dark to produce conidia. To obtain pure colonies, a single conidium per colony was transferred to a new V8 agar plate. After two weeks of growth at 25°C in the dark, half of each culture was used to make 15% glycerol stocks that were stored at -80°C. The rest of the culture was left to grow for 18 weeks with sub-culturing every three weeks at 25°C in the dark for DNA extraction.

To prepare samples for DNA extraction, mycelia were scraped from plates with a scalpel and freeze dried. The freeze-dried material was homogenised with metal beads and a homogeniser. One spatula of ground freeze-dried mycelium (<80 mg) was transferred to a 2 ml Eppendorf tube for DNA extraction. A modified CTAB DNA extraction protocol was performed (Stewart and Via 1993), which included adding 125 µl β-mercaptoethanol (BME) (Sigma-Aldrich) and 1 g of polyvinylpyrroilidone (Sigma-Aldrich) to 25 ml CTAB (Hexadecyltrimethyl ammonium bromide) (Sigma-Aldrich). Genomic DNA concentrations were determined with a Nano-Drop™ 1000 spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific Inc., USA). All genomic DNA was diluted to a working concentration of 30 ng/µl for subsequent PCRs.

Diagnostic histone PCR. A histone H3 specific diagnostic PCR test was used on all isolates to confirm the identity of the *Cercospora* species isolated from the maize leaves collected (Crous et al. 2006). The method was adapted as follows: PCR reactions were not multiplexed, and annealing temperatures were 55°C for the CYLH3F/CYLH3R primer pair, and 58°C for primer pairs *C. zeinaHIST*/CYLH3R, *C. zeaeHIST*/CYLH3R and CmaizeHIST/CYLH3R. The ex-type cultures of CPC11998 (*Cercospora zeina*) and CBS1177556 (*Cercospora zeae-maydis*) were used as positive controls.

Microsatellite discovery and primer design. The draft genome sequence of an African isolate of *C. zeina* (CMW 25467: Meisel et al. (2009) was sequenced using the Illumina platform at the

Purdue Genomics Core Facility (raw reads produced: paired end library - 3.1 Gb; 3 Kb mate pair library - 3.5 Gb; and 8 Kb mate pair library – 10 Gb). Reads of poor quality were filtered and trimmed using the Trimmomatic v. 0.30 package (Bolger et. al. 2014). The 8 Kb matepair library reads were treated as single-end and assembled using the Velvet *de novo* assembler (Zerbino & Birney 2008), with an optimum k-mer size (57nt) determined with the VelvetOptimizer script. Scaffolding of the assembled contigs was performed with SSPACE v. 2.0. (Boetzer et al. 2011) using the paired-end, 3 Kb and 8 Kb mate-pair library reads with an approximated 0.25 error in library size. Assembly gaps were filled using Gapfiller v. 1.11 (Nadalin et al. 2012) using the same libraries and error approximation. The draft assembly of 37 Mb consisted of 10,044 contigs >200 bp with an N50 = 161 Kb, and an average contig size = 4059 bp (DK Berger and NA Olivier, unpublished data).

MSATCOMMANDER software was used to identify *C. zeina* CMW 25467 contigs that contained dinucleotide, trinucleotide and tetranucleotide microsatellites that possessed eight motifs or more (Faircloth 2008). PRIMER3 was used to design primers flanking the microsatellite regions with the following primer parameters: 19 to 23 nucleotides in length, possess GC content higher than 50%, primers end with a guanine or cytosine at the 3' end, have an annealing temperature of 58°C, and amplify fragments between 100-450 bp (Rozen and Skaletsky 2000). Primers were designed flanking microsatellite regions from the largest contigs (>450 Kb) to reduce the possibility of linked loci (Rozen and Skaletsky 2000)..

The draft genome sequence of a *C. zeina* isolate from the United States, USPA-4 (OYPA) (Dunkle and Levy 2000; Wang et al. 1998), was generated using Illumina technology (paired end and mate pair libraries). The *C. zeina* USPA-4 Illumina reads were assembled using the VELVET software (Zerbino and Birney 2008) as described above to produce a draft genome assembly of 25.6 Mbp with 25,791 contigs (> 200 bp) and N50 = 1,466 bp (BH Bluhm, unpublished data). Simple Sequence Repeat Identification Tool (SSRIT) (Temnykh et al. 2001) was used to identify dinucleotide, trinucleotide and tetranucleotide repeats that possessed ten motifs or more from the draft genome of *C. zeina* USPA-4. Primers were designed as described above.

Different software packages (MSATCOMMANDER and SSRIT) were used for microsatellite discovery from the two different *C. zeina* isolates due to the routine use and availability of the software in each lab. However, both packages are based on similar principles for microsatellite identification and proved suitable.

To test primers for successful amplification and to determine if the loci they amplified were polymorphic, all microsatellite markers were tested on five geographically distinct isolates (USPA-4 from the USA, CMW 25467 from Africa, and one isolate each from KwaZulu-Natal (Greytown), Mpumalanga and the North West province). PCR reactions with 20 µl volumes consisted of 5X MyTaq™ reaction buffer (Bioline, Cape Town), 10µM of each primer (Integrated DNA Technologies), 1U of MyTaq™ polymerase (Bioline, Cape Town), 30ng gDNA and MilliQ water. The cycling parameters were constant for all steps but the annealing temperatures were optimised individually for each marker. Initial denaturation was performed at 94°C for 1 min followed by 30 cycles of 15 sec denaturation at 94°C, 15 sec annealing of primers at 54°C, 56°C or 58°C and 20 sec elongation at 72°C, and a final elongation step at 72°C for 1 min. All PCR products were verified on a 3% agarose gel in a 1X TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA at pH8.0), stained with 0.5µg/ml ethidium bromide and visualised under UV light (Gel Doc™ EZ Imager, Bio-Rad, Richmond, CA).

PCR amplicons of varied sizes for each primer pair were cleaned using ethanol-salt purification and precipitation and sequenced in both directions to confirm polymorphisms. Sequencing reactions were set up in total volumes of 10 µl as follows: 2 µl BigDye[®] Terminator v3.1 (Thermo Fisher Scientific Inc., USA), 1 µl BigDye[®] buffer (Thermo Fisher Scientific), 2 µl PCR product, 1.6 µl microsatellite primer (forward or reverse) and 3.4 µl MilliQ water. Sequencing cycling parameters started with initial denaturation at 95°C for 10 min followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 20 sec and elongation at 72°C for 30 sec. A final elongation step at 72°C for 7 min completed the PCR cycle. Microsatellite sequencing reactions were cleaned by an ethanol-salt purification and run on the ABI 3500xl Genetic Analyser (Thermo Fisher Scientific Inc., USA).

Sequences were analysed and assembled using CLC MAIN WORKBENCH v6.0 (CLC Bio, a QIAGEN® Company, USA). Microsatellite markers that displayed a length polymorphism in the microsatellite region between at least two of the five isolates were selected for further fragment analysis. Sequences of the microsatellite markers developed in this study (Table 1) corresponding to the alleles from *C. zeina* CMW 25467 (Meisel et al. 2009) have been deposited in GenBank with accession numbers shown in Table 1.

Microsatellite fragment analysis. The 5' end of each forward primer for all the polymorphic microsatellite markers was fluorescently labelled (Thermo Fisher Scientific Inc., USA) with one of four available fluorophores in the G5 filter set viz. 6-FAM™ (blue), VIC® (green), NED™ (yellow) or PET® (Red). Primer pair amplicons that could be pooled together into one lane for fragment analysis were arranged into either panel 1 or panel 2 (Table 1). Each panel was set up to contain amplicons between 100-450 bp. Primers that produced amplicons of the same size were either labelled with different fluorescent dyes or placed into different panels (Table 1).

PCR amplification and amplicon concentrations were optimised for fragment analysis on the ABI 3500xl Genetic Analyser (Thermo Fisher Scientific Inc., USA). Each marker was first amplified separately before pooling the amplicons in either panel 1 or 2 (Table 1). PCR reactions consisted of 5x MyTaq™ reaction buffer, 10 µM of each primer, 1U of MyTaq™ DNA polymerase, 10 ng gDNA and distilled MilliQ water to a total volume of 10 µl. The cycling conditions were 1 min initial denaturation at 94°C followed by 30 cycles of 15 sec denaturation at 94°C, 15 sec annealing of primers at 58°C or 59°C (only *CzSSR*10 and *CzSSR*12), and 20 sec elongation at 72°C, followed by a final elongation step of 1 min at 72°C.

Fluorescently labelled amplicons were analysed with the ABI 3500xl Genetic Analyser (Thermo Fisher Scientific Inc., Carlsbad, USA) as follows; 1µl pooled amplicons were loaded along with 0.14µl LIZ500 size standard (Thermo Fisher Scientific Inc., Carlsbad, USA) and 9.86µl Hi-Di formamide (Thermo Fisher Scientific). Alleles were detected using GENESCAN™ software, and were manually scored according to the fragment length using GENEMAPPER® software v4.1 (Thermo Fisher Scientific Inc., USA).

Genetic diversity analysis based on microsatellite data: All *C. zeina* isolates collected in South Africa were evaluated with the 14 microsatellite markers (Table 1). Isolates that contained the same haplotype were considered clones. Two datasets were generated: one dataset contained all the isolates sampled (non clone-corrected) and the other contained the unique haplotypes for each population (clone-corrected). Populations were defined by the geographic regions from where the isolates originated, namely KwaZulu-Natal, Mpumalanga and the North West province.

The non clone-corrected data set was used to calculate Stoddard and Taylor's measure of genotypic diversity: $G = 1/\sum p_i^2$, where p_i is the observed frequency of the ith haplotype in each of the non-clone-corrected populations of South Africa (Stoddart and Taylor 1988). The number of multilocus genotypes (MLG) and G for each population was determined using the statistical program R v. 3.1.2 (R Development Core Team, Vienna) and the R package poppr (Kamvar et al. 2014). To adjust for sample size discrepancies, the expected number of genotypes (eMLG) and genotypic diversity (G.est) was calculated with poppr using rarefaction to the smallest sample size of 14 (from North West province) (Grünwald et al. 2003). To validate whether the sample size and number of microsatellite markers used in the study is sufficient to support any statistical significance of haplotypic diversity, MULTILOCUS v1.3b was used to plot the haplotypic diversity against number of loci (Agapow and Burt 2001). The level of haplotypic diversity was assessed by calculating the clonal fraction, where clonal fraction (CF) was calculated as CF = 1 - (number of unique haplotypes/total number of isolates) (Zhan et al. 2003).

The clone-corrected data set was used to calculate all other measures of genetic diversity. The allelic frequency, number of alleles, number of effective alleles and Nei's gene diversity (Nei 1972) was calculated for each population using the software package POPGENE v1.31 (https://www.ualberta.ca/~fyeh/popgene.html (Yeh et al. 1999). Nei's gene diversity (H) was calculated using the equation $H = 1 - \sum x_\alpha^2$, where x_α is the frequency of the αth allele (Nei 1972). In order to compare the allelic richness and private allelic richness for each population, the population data sets were subjected to rarefaction with the smallest population containing a sample size of 14 (Table 3) (Szpiech et al. 2008).

Population structure analysis.   Clone-corrected data was used for population structure analyses. STRUCTURE v2.3.4 was used to determine the most likely number of clusters in the dataset (Falush et al. 2003). STRUCTURE uses a Bayesian Monte Carlo Markov Chain (MCMC) clustering algorithm to calculate the most likely number of clusters (K) from the allelic frequencies at each locus for each isolate (Falush et al. 2003). Both ancestry models (no admixture and admixture) were tested along with each allelic frequency model (independent and correlated), respectively. One hundred iterations were run for K ranging between one and ten, and were simulated for 200 000 MCMC runs after a burnin of 10 000. The optimal number of clusters (K) was determined by implementing the Evanno method, where delta K was generated using STRUCTURE HARVESTER (Earl and von Holdt 2011; Evanno et al. 2005). Delta K is a likelihood statistic based on the log probability change in relation to the number of K. The uppermost value for Delta K correlates to the most likely number of clusters (K) observed in the population. The STRUCTURE bar graphs were generated in CLUMPAK (Kopelman et al. 2015).

GENALEX v6.501 was used to calculate genetic distance (Da) for the data set, and Principle Coordinate Analysis (PCoA) was used to graphically visualise the genetic distance generated between isolates in each population (Peakall and Smouse 2012).

Population differentiation was calculated for all isolates using Weir's Theta (θ) statistic, executed in MULTILOCUS v1.3b (Agapow and Burt 2001). Weir's Theta (θ) statistic is a modified F_{ST} (Wright 1978), calculated by the following equation: $\theta = 2N_e\mu$, where N_e is the effective population size, and μ is the mutation rate. The significance was tested by 1000 randomisations of the data set. The null hypothesis of no population differentiation was rejected at $P<0.05$.

Molecular variation within and among populations was assessed by performing an AMOVA, in GENALEX v6.501 (Peakall and Smouse 2012). The significance of Φ (also termed PhiPT, a measure on a scale of 0 to 1 of population differentiation analogous to Fst, but more appropriate for haploid-microsatellite data) was tested with 1000 permutations (Peakall and Smouse 2012). The null hypothesis that Φ (PhiPT) = 0 was rejected if $P<0.05$.

To evaluate the level of gene flow between populations, the estimate of gene flow (Nm) was calculated using POPGENE v1.32 (https://www.ualberta.ca/~fyeh/popgene.html (Yeh et al. 1999). Gene flow (Nm) between populations was calculated using the estimate of genetic differentiation (G_{ST}), whereby $Nm = 0.5(1-G_{ST})/G_{ST}$, using POPGENE v1.32 (https://www.ualberta.ca/~fyeh/popgene.html; (Yeh et al. 1999).

Nei's standard genetic distance (Da) was calculated between all isolates for the clone-corrected data set in POPULATIONS v1.2.32 (Nei et al. 1983). To graphically represent the population structure of *C. zeina* within South Africa, a neighbour joining (NJ) tree was constructed using the genetic distance matrix generated with 1000 bootstrap replicates. The neighbor joining tree was further edited using the software package MEGA6 (Tamura et al. 2013), where isolates were colour-coded according to their geographic origin.

Mating type gene analysis and primer development. Mating type primers from Groenewald et al. (2006) were tested for amplification of the *MAT* genes in *C. zeina*, however, non-specific amplification was observed, possibly due to the degenerate nature of the MAT1-2 primer pair designed for several *Cercospora* species. *MAT1-1* (CPC 11998) and *MAT1-2* (CPC 11995) sequences (GenBank accession numbers DQ264748 and DQ264762, respectively) were queried against the CMW 25467 draft genome to identify *MAT* genes and to determine which *MAT* genes were present. Exon fragments from both mating types were manually reconstructed from the draft genome sequence using the alignment function in CLC MAIN WORKBENCH v6.0 (CLC Bio, a QIAGEN® Company, USA). *C. zeina* specific primers were re-designed based on the CPC 11998, CPC 11995 and the CMW 25467 sequences, using Primer 3. *MAT1-1-1* primers, CzMAT1F and CzMAT1R (Table 4), were designed 5' to the first exon and within the third exon of the predicted *MAT1-1-1* coding region, respectively (Fig. 4). This region includes two introns and the alpha box domain. *MAT1-2-1* primers, CzMAT2F and CzMAT2R (Table 4), were designed within the first and third exons of the predicted *MAT1-2-1* coding region, respectively (Fig. 4). This region includes part of the HMG domain. The gDNA sequence of *MAT1-1-1* from CMW25467, and the PCR product sequences of *MAT1-1-1* from USPA-4 (amplified with primers CzMAT1F and CzMAT1R)

and *MAT1-2-1* from *C. zeina* isolate 2011.GT11 (amplified with primers CzMAT2F and CzMAT2R) have been deposited in GenBank with accession numbers KU686678-KU686680, respectively.

Mating types were determined for *C. zeina* isolates in a multiplex PCR assay. PCR reactions were performed in a total volume of 25 μl, containing 1X NH4 Reaction Buffer, 2 mM MgCl$_2$, 0.2 mM dNTPs, 0.2 μM primer, 1.25U BIOTAQ™ DNA polymerase (Bioline, California, USA) and 30ng gDNA. PCR cycling conditions with an initial elongation step at 94°C for 3 min followed by 25 cycles of denaturation at 94°C for 20 sec, primer annealing at 60°C for 30 sec and elongation at 72°C for 40 sec, followed by a final elongation step for 5 min at 72°C. PCR products were separated on a 1.5% agarose gel (Lonza, USA) containing 0.5 μg/ml ethidium bromide, in a 1X TAE buffer, and visualized under UV light. Isolates were scored visually for their mating type based on amplicon size. The PCR amplicons were sequenced to confirm that the correct *MAT* idiomorphs were amplified. The PCR products were purified using Sephadex® G50 (Sigma-Aldrich, St Louis, USA) 20μm receiver columns (Macherey-Nagel, Düren, Germany) before and after sequencing reactions. Sequencing reactions were carried out in a final volume of 10 μl with 1 μl Big Dye®, 1.5 μl 5x sequencing buffer, 3.2 pmol of forward or reverse primer and 3 μl of PCR product. Cycling conditions included an initial denaturation step at 95°C for 30 sec followed by 25 cycles of 95°C for 30 sec, 58°C for 10 sec and 60°C for 4 min. The PCR amplicons were sequenced on an ABI 3500xl Genetic Analyser POP 7™ and the sequences were analysed using CLC MAIN WORKBENCH v6.0.

Distribution of mating types. Mating type ratios (*MAT1-1*/*MAT1-2*) were calculated for each population. A population in this instance was defined as isolates from the same field site. The Chi Square goodness of fit test was used with $p < 0.05$, to test the null hypothesis that there is no significant difference between the expected 1:1 ratio for random mating and observed *MAT* gene ratios.

RESULTS

Cercospora **species identification.** A total of 369 isolates were obtained from the three geographic regions in South Africa: 239 isolates from KwaZulu-Natal, 110 from Mpumalanga and 20 from the North West province (Fig. 1). For all isolates, primer set CYLH3F/CYLH3R generated fragments of 389bp, which indicated that all isolates were a *Cercospora* species. Primer set *C. zeinaHIST*/CYLH3R generated fragments of 284bp, indicating that all isolates were *C. zeina* (Meisel et al. 2009). For all isolates, primer sets *C. zeaeHIST*/CYLH3R and CmaizeHIST/CYLH3R did not generate any amplicons.

Microsatellite marker discovery. VELVET software was used to assemble the draft *C. zeina* (USPA-4) genome into 24,791 contigs (>200 bp). SSRIT was used to identify 36 microsatellites from the draft genome, from which primers were designed. Of the 36 microsatellite markers, 22 primer pairs generated amplicons and 12 primer pairs showed different amplicon sizes between USPA-4 and CMW 25467.

VELVET software was used to assemble the draft *C. zeina* (CMW 25467) genome into 48,969 contigs (Zerbino and Birney 2008). MSATCOMMANDER was used to identify 1295 contigs from the draft *C. zeina* (CMW 25467) genome that contained a total of 6389 SSRs (Faircloth 2008). Seventy-five microsatellite regions were chosen for primer design. Of the 75 designed primers 51 successfully amplified PCR products for USPA-4 and CMW 25467, and produced clear bands on an agarose gel. These 51 primer pairs were further tested on five geographically distinct isolates, generating varied amplicon sizes between the geographically distinct isolates for twenty-four primer pairs.

After sequencing, 19 primer pairs were confirmed polymorphic in the microsatellite region and fluorescently labelled with the G5 Filter Standard Dye Sets (Life Technologies). Five markers required further optimisation due to absence of peaks or presence of peak patterns that could not be scored, and thus were discarded to reduce protocol complexity. Fourteen (Table 1) of the 19 primer pairs were optimised and pooled into two panels for analysis on the ABI Genetic Analyser 3500xl (Life Technologies). From the two genomes, a total of 14 microsatellite markers were utilized for downstream analyses, ten from USPA-4 and four from CMW 25467 (Table 1).

Genetic diversity of *C. zeina* isolates. The 14 microsatellite markers produced 60 different alleles across the 369 *C. zeina* isolates collected in South Africa (Table 1). This represents an average of ~5 alleles per microsatellite marker. The most variable microsatellite marker was *CzSSR*07 with 12 alleles; and markers *CzSSR*13, *CzSSR*15 and *CzSSR*18 were the least variable with only 2 alleles each (Table 1). Seventeen alleles were restricted by geographical location (10 in KwaZulu-Natal and 7 in Mpumalanga) and confirmed as private alleles using HP-RARE (Kalinowski 2005) (Supplementary Table 1). Seven unique alleles were identified for USPA-4 with only seven alleles shared with the South African isolates. There were no unique alleles identified for CMW 25467, as all alleles observed were shared with South African isolates. Interestingly, none of the 14 *C. zeina* microsatellite primer pairs produced amplicons from gDNA of *C. zeae-maydis* (CBS1177556) (data not shown). Control histone PCR reactions showed that the *C. zeae-maydis* gDNA was of sufficient quality for PCR amplification.

Two hundred and forty two unique haplotypes were identified from a total of 369 isolates in South Africa (Table 2)(Supplementary Table 2). Four haplotypes were shared among KwaZulu-Natal, Mpumalanga and North West province populations. Eight haplotypes were shared between KwaZulu-Natal and Mpumalanga, and two haplotypes were shared between KwaZulu-Natal and the North West province which suggested some degree of direct movement of the pathogen between different provinces. There were 154 haplotypes identified in the KwaZulu-Natal population, 92 haplotypes identified in Mpumalanga and 14 in the North West province (Table 2). The clonal fraction for all South African isolates was 0.34, with 0.36 in KwaZulu-Natal, 0.16 in Mpumalanga and 0.30 in North West province (Table 3). The G and G.est values were 47.48/15.81 for KwaZulu-Natal, 77.56/18.83 for Mpumalanga, 7.69/7.69 for North West province, and 64.26/16.64 for the overall haplotypic diversity for South Africa (Table 2). MULTILOCUS v3.1b was used to plot the mean genotypic diversity against the number of microsatellite markers, which showed a plateau at 12 markers (Supplementary Fig. 1), indicating that the set of 14 microsatellite markers was sufficient to elucidate the genetic diversity of these *C. zeina* isolates collected in South Africa.

The unbiased gene diversity calculated with rarefaction (Hexp) was identical for RSA as a whole, KwaZulu-Natal and Mpumalanga at 0.35 (Table 2). The gene diversity for the North West Province was slightly lower at 0.18. Allelic richness ranged from 1.53 in the North West to 2.80 in KwaZulu-Natal; and the private allelic richness ranged from 0.15 in the North West to 0.24 in KwaZulu-Natal (Table 3).

Population Structure. The most probable number of clusters in the dataset, defined using STRUCTUE analyses, was K=2 for the no admixture model and K= 3 for the admixture model using either the independent or correlated allelic frequency models. In all cases, however, no individual was assigned to a single cluster and no clear population structure could be observed based on isolates from a particular geographic location (KwaZulu-Natal, Mpumalanga and North West). Instead, a high degree of admixture was observed where isolates had membership in all clusters represented at either K=2, K=3 or K=4 (Fig. 2). Such admixture in a population is indicative of frequent recombination events.

PCoA analysis graphically corroborated the lack of population sub-structures observed from STRUCTURE analysis (Fig. 3). Isolates from KwaZulu-Natal, Mpumalanga and the North West did not show separate groupings based on province, since isolates from different provinces clustered together (Fig. 3). The first principle co-ordinate accounted for 90.2% of the variation, as determined using Nei's genetic distance.

The value of theta at $\theta = 0.03$ did not fall within the predicted expected range of 1000 randomisations, where θ ranged between -0.005 to 0.009, at P<0.001. This indicated a lack of evidence to support regional population differentiation for *C. zeina* within South Africa. No significant population differentiation was determined from AMOVA analyses. AMOVA indicated 98% of the variance was attributed to variability between individuals, and only 2% variance between regional populations, producing a low measure of population differentiation Φ (PhiPT) = 0.021 with P=0.001 (Peakall and Smouse 2012). In this study, the estimate of gene flow (Nm) was calculated as 5.51. This strong evidence of gene flow between KwaZulu-Natal,

Mpumalanga and the North West populations supports the lack of evidence for population differentiation presented by the theta (θ) statistic and AMOVA.

The NJ tree presents all *C. zeina* isolates within the clone-corrected dataset used in this study (Supplementary Fig. 2). No clear geographical structure or sub-structures were observed since isolates from KwaZulu-Natal, Mpumalanga and the North West province were randomly distributed throughout the tree (Supplementary Fig. 2). In addition, the overall low genetic distance (Da) between isolates and lack of bootstrap support observed in the NJ tree (bootstrap values for branches <10: Supplementary Fig. 2) further corroborates the lack of population differentiation between the isolates. These results support those obtained with STRUCTURE, PCoA and AMOVA indicating that the isolates of *C. zeina* from South Africa represent a panmictic (randomly recombining) population.

***C. zeina MAT* gene analysis.** The draft genome sequence of *C. zeina* CMW 25467 facilitated the search for *MAT* gene sequences and design of a multiplex *MAT* PCR assay. The aim was to determine which *MAT* idiomorph was present in the genome of *C. zeina* CMW 25467, and if the isolate contained additional fragments of both *MAT* genes as exonic fragments, as reported for *C. beticola* and *C. zeae-maydis* (Bolton et al. 2014). BLAST analysis and sequence alignment revealed that isolate CMW 25467 contained on contig 24 the complete *MAT1-1* idiomorph (deposited in Genbank with accession number KU686678), which had the same gDNA sequence as the *MAT1-1* idiomorph from *C. zeina* CPC 11998 (DQ264748)(Fig. 4). No *MAT1-2* idiomorph was identified in the draft genome sequence of *C. zeina* CMW 25467, consistent with the heterothallic nature of this species (Crous et al. 2006).

BLAST analysis identified additional exonic fragments of both *MAT1-1-1* and *MAT1-2-1* on several contigs of the draft genome sequence of *C. zeina* CMW 25467 (Fig.4). These exonic fragments showed characteristics observed in *C. beticola* and *C. zeae-maydis*, namely lack of intron sequences and proximity of exons of both *MAT* genes on the same contig for five of the contigs (Fig. 4). We reconstructed the CMW 25467 *MAT1-1-1* exonic fragments from the five contigs (Fig. 4A), and the resultant *MAT1-1-1* sequence had high identity to the predicted open reading frame

(ORF) from the *MAT1-1* idiomorphs from *C. zeina* CPC 11998 (DQ264748) and CMW 25467 (KU686678)(95% identity, E-value 2e-06). Likewise, the *MAT1-2-1* sequence reconstructed from the six contigs (Fig. 4B) had 90% identity (E-value 1e-176) to the predicted ORF from the *C. zeina* CPC 11995 *MAT1-2* idiomorph (DQ264762).

A *MAT* gene multiplex PCR assay for screening populations of *C. zeina* was developed from the *MAT1-1* idiomorph sequence and *MAT1-2-1* reconstructed exonic sequence from the draft genome of *C. zeina* CMW 25467. Primers were designed to ensure that only the single *MAT* idiomorph would be amplified from an isolate, and not exonic fragments (Fig. 4). The forward primer for *MAT1-1-1* is positioned 5' to the first exon, and thus is not present in any of the exonic fragments (Fig. 4A). The reverse primer is in the third exon present additionally in contig 135, but this contig lacks sequences corresponding to the forward primer (Fig. 4A). These primers produce a specific *MAT1-1* idiomorph product of 631 bp (Fig. 4A, Fig. 5). The forward and reverse primers for *MAT1-2-1* were designed to exon 1 and exon 2, respectively, to specifically amplify a *MAT1-2* idiomorph product of 409 bp (Fig. 4B, Fig. 5). No amplification was expected to occur from the *MAT1-2-1* exonic fragments since the forward and reverse primers are on different contigs (Fig. 4B).

Distribution of *C. zeina MAT* gene idiomorphs in South Africa. The *C. zeina MAT* multiplex PCR assay developed in this study clearly distinguished isolates with the *MAT1-1* or the *MAT1-2* idiomorphs (Fig. 5). A PCR product of 631bp was produced from *MAT1-1* isolates (Fig. 5, lanes 5 & 6), whereas a PCR product of 409bp identified *MAT1-2* isolates (Fig. 5, lanes 2, 3 and 4). *C. zeina* USPA-4 and CMW 25467 were *MAT1-1*. All 335 South African isolates of *C. zeina* that were screened possessed one or the other *MAT* idiomorph but not both (Table 5)(Supplementary Table 2). Both mating types were observed in each geographic location (Table 5). Most field sites did not show significant deviations from the expected 1:1 ratio of *MAT* idiomorphs, except for Greytown and Baynesfield in KwaZulu-Natal (Table 5).

DISCUSSION

Management of GLS in South Africa and worldwide faces several key challenges. These challenges include: (i) the expansion of GLS due to increasing adoption of conservation tillage, facilitating over-wintering of the pathogen on maize debris (Ward et al. 1999); (ii) the conundrum that indistinguishable symptoms can be caused by two distinct fungal species − *C. zeae-maydis* and *C. zeina* (Wang et al. 1998); (iii) the lack of knowledge of the genetic diversity or mode of reproduction of these fungi; and (iv) the reliance on fungicides for control in commercial farms and the risk of fungicide resistance (Bradley and Pedersen 2011).

In this study, we illustrated how the availability of draft genome sequence for a fungal pathogen, which can be generated relatively rapidly, can be used to address disease management challenges (ii) and (iii). First, DNA sequence information has facilitated the design of rapid species-specific PCR assays which can be used to screen many isolates from GLS lesions to identify the causal species of *Cercospora* (Crous et al. 2006; Korsman et al. 2012; Meisel et al. 2009). Second, microsatellite markers could be designed from the genome sequence, overcoming the need for tedious and inefficient laboratory approaches for marker discovery (Santana et al. 2009). Third, the genome sequence revealed the problem of additional copies of *MAT* gene exons found in some Dothidiomycetes (Bolton et al. 2014), and allowed the design of a *MAT* idiomorph specific multiplex PCR assay. Finally, in the event that fungicide resistance is observed in *C. zeina* in future (challenge (iv)), draft genome sequence information can be used to design PCR primers to likely targets for mutations in resistant isolates, such as cytochrome B genes (Zeng et al. 2015).

Draft genome sequences of two isolates of *C. zeina* were mined for microsatellite sequences and primer pairs were designed to the flanking regions. A comprehensive set of 369 isolates of *C. zeina* were collected from three maize growing provinces in South Africa, and subjected to molecular marker analysis using the 14 polymorphic microsatellite markers designed in this study. This is the first set of microsatellite markers developed for *C. zeina*, which is a useful contribution towards future population genetics studies of this important maize pathogen worldwide. Our analysis revealed high levels of genetic diversity and gene flow between *C. zeina* isolates, and lack of

regional population differentiation between the provinces in South Africa. This indicates that there is widespread dispersal of the pathogen across maize growing areas in the region.

This study covered provinces in South Africa where GLS is of most importance in commercial maize farming (KwaZulu-Natal, North-West and Mpumalanga). Ad hoc field surveys indicated that gray leaf spot disease of maize is absent or rare from the drier northern and western commercial maize production areas of South Africa (Limpopo, Free State and Gauteng provinces)(FJ Kloppers, PANNAR SEED Pty Ltd, personal communication). GLS is occasionally observed on irrigated maize along the Orange river (Northern Cape), however we failed to obtain isolates during this study (2011 – 2013). The Eastern Cape province was not included, since commercial maize production is limited, and thus GLS from the predominant small-holder grown maize would be part of a future study.

Mining of the *C. zeina* genome sequence of an African isolate (CMW 25467) revealed the presence of an intact genomic copy of the *MAT1-1* idiomorph plus exonic fragments of *MAT* genes from both idiomorphs, as observed in other *Cercospora* spp (Bolton et al. 2014). Genome sequence data allowed the design of a multiplex PCR assay for detection of either intact *MAT* idiomorphs, but not the additional exonic fragments. This assay was used to screen the *C. zeina* collection from South Africa, and confirmed that this species is heterothallic, since each isolate had only one of the two intact *MAT* idiomorphs. *C. zeina* isolates from all three provinces showed *MAT* gene ratios that were not significantly different from 1:1, apart from two locations in KwaZulu-Natal. These results support the hypothesis that *C. zeina* in South Africa is undergoing cryptic sexual recombination, and have implications for disease management, as outlined below.

Previous studies had indicated that only *C. zeina* was the causal agent of GLS of maize in South Africa and other African countries (Dunkle and Levy 2000; Meisel et al. 2009), however only a small number of isolates were surveyed from each region. We report here a comprehensive survey of more than 350 isolates over several seasons from South Africa, confirming that there is no evidence for *C. zeae-maydis* in the region. Two additional *Cercospora* species associated with GLS of maize, namely *Cercospora* sp. CPC 12062 (Crous et al. 2006) and *C. sorghi var. maydis* (Carson

et al. 2002; Neves et al. 2015) were not observed in this survey in South Africa. However, it is not clear whether these species are causal agents of GLS or saprophytes associated with GLS lesions caused by *C. zeina* or *C. zeae-maydis*. Further work on the importance of these related sibling species will be facilitated by DNA-based diagnostic methods. These include the histone H3 PCR which can be implemented directly from GLS lesions (Meisel et al. 2009), and the *cpr1* quantitative PCR assay, which can distinguish *C. zeina* from *C. zeae-maydis* based on the Tm of the amplicon (Korsman et al. 2012).

Microsatellite markers were designed from genomes of both a USA and African isolate of *C. zeina*, and the selected 14 markers produced amplicons from all 369 *C. zeina* isolates from South Africa. Our success in using genome sequence information from a USA isolate for African isolates is consistent with the results of Dunkle and Levy (2000), who compared nine USA *C. zeina* isolates with 30 African *C. zeina* isolates (from Uganda and Zimbabwe). PCR-RFLP of ITS and rDNA with five restriction enzymes gave identical patterns for USA and African isolates, which were distinct from the patterns for *C. zeae-maydis* (Dunkle and Levy 2000). In addition, AFLP analysis showed 97% similarity between the USA and Uganda/Zimbabwe *C. zeina* isolates. The authors included additional isolates from South Africa and Zambia, and concluded that isolates from all four African countries were con-specific with *C. zeina* in USA and distinct from *C. zeae-maydis* (Dunkle and Levy 2000). Evidence for genomic differences between *C. zeina* and *C. zeae-maydis* is given by the fact that none of the 14 microsatellite marker primer pairs developed from the genome sequences of *C. zeina* in this study were able to produce amplicons from *C. zeae-maydis*. Similarly, ten microsatellite primer pairs developed from ESTs of *C. zeae-maydis* were not amplifiable in *C. zeina* (Molecular Ecology Resources Primer Development Consortium, 2009).

In a previous study, Wang et al. (1998) conducted AFLP analysis of 12 isolates of *C. zeina* from different states in USA, and scored 111 loci to find 94% genetic similarity on average. A larger number (79) of *C. zeae-maydis* isolates revealed a similar high genetic similarity (93%) from the same AFLP study. The same AFLP primers were used to score 104 loci in 41 *C. zeae-maydis* isolates and 28 *C. zeina* isolates from different regions of Brazil, and similar high levels (>92%) of

genetic similarity within species was observed (Brunelli et al. 2008). In contrast, Dunkle and collegues screened 80 *C. zeae-maydis* isolates from a single site in Illinois using 10 microsatellite markers and reported high genetic diversity (79 haplotypes; within-marker gene diversity of 0.61-0.91) (Molecular Ecology Resources Primer Development Consortium, 2009). In the current study, microsatellite analysis of the 369 isolates from South Africa revealed high genetic diversity, more in line with the latter microsatellite results. There were a high number of unique haplotypes (242) and an overall population gene diversity of 0.35 (Table 2). It appears that the microsatellite markers were better able to detect genetic diversity than AFLP markers, and this could be due to multiple microsatellite alleles per locus (averages of 7 and 4 alleles for USA and South African microsatellite studies, respectively). High genetic diversity was also observed in microsatellite analysis of a collection of 132 *C. sojina* isolates from Arkansas, USA (Kim et al. 2013). Eight microsatellite markers (average of four alleles/marker) produced high gene diversity (H) values of 0.42 – 0.58 across collection sites. Similarly, high genetic diversity was also observed in populations of *C. beticola* from Europe (Groenewald et al. 2008).

The South African *C. zeina* isolates in this study were collected from three different provinces separated by several hundred kilometres (Fig. 1). However, analysis of the microsatellite data in several different ways did not reveal regional population differentiation (G_{ST} = 0.08, Table 3). Bayesian clustering implemented in STRUCTURE did not reveal regional populations (Fig. 2), neither did Weir's theta statistic (theta = 0.03), or Principal Co-ordinate analysis which showed no clustering of region-specific haplotypes (Fig. 3). AMOVA corroborated this result, since only 2% of the variation was explained by differences between regional populations. Only 17 of the 60 microsatellite alleles were confined to a province (Supplementary Table 1). Additionally, gene flow between provinces was also calculated to be high (Nm = 5.51) (Table 2). Ward et al. (1999) noted that *Cercospora* spp. spores can travel up to 40 km at wind speeds of 10 m/s, and this may be one factor that accounts for the distribution of genotypes between regions (Ward et al. 1999).

In a previous study, regional population differentiation was also not seen in *C. sojina* isolates collected from soybean in six regions in Arkansas, USA (G_{ST} = 0.084), and the authors inferred

high levels of genetic exchange between the regions (Kim et al. 2013). Five Eurasian and one New

Zealand population of *C. beticola* were assessed with AFLP markers (Groenewald et al. 2008).

Population differentiation was low within the Eurasian populations (theta = 0.07) with high gene

flow (Nm = 5.8), but there was high differentiation between each of the Eurasian populations and

the New Zealand population (theta = 0.33 − 0.41). These results from populations of *C. zeina* in

South Africa, *C. sojina* in USA and *C. beticola* in Europe present a common pattern of lack of

regional population differentiation and high gene flow in geographically connected regions.

Intriguingly, BLAST analysis with the *MAT* gene idiomorphs showed that the genome of

C. zeina isolate CMW 25467 contained exonic fragments of both *MAT* genes in addition to the

MAT1 locus (Fig. 4). Bolton et al. (2014) observed this in *C. beticola* and *C. zeae-maydis*, which

indicates that the exonic duplication of *MAT* genes preceded the divergence of *C. zeina* and these

species. Mating type analysis showed that the two *MAT* idiomorphs were evenly distributed at all

sites sampled in South Africa, except for two fields (Table 5). Both mating types could be found in

all fields, suggesting that there was opportunity for sexual reproduction to occur. Both *MAT*

idiomorphs were identified within the same plant (data not shown), which suggests that sexual

recombination may occur under optimal conditions. This was also observed for *C. beticola* where

they discovered both mating types within a single leaf (Bolton et al. 2012).

These results above are consistent with the study by Groenewald et al. (2006) where equal

numbers of *MAT* idiomorphs in a small sample of *C. zeina* isolates from KwaZulu-Natal province

was observed. Although the sexual state of *C. zeina* has not been observed in nature, these results

suggest that sexual reproduction is occurring within *C. zeina* populations in South Africa. Studies

on other *Cercospora* species have shown that mating type genes are equally distributed (Bolton et

al. 2012; Groenewald et al. 2006; Groenewald et al. 2008; Kim et al. 2013). *C. beticola* and *C.*

sojina were shown to be genetically diverse and *MAT* genes were of approximately equal

proportions which is an indirect indication that sexual reproduction may be occurring (Bolton et al.

2012; Kim et al. 2013). These observations were inconsistent with populations solely undergoing

asexual reproduction which lead the authors to suggest that these species may have a cryptic sexual

cycle (Bolton et al. 2012; Groenewald et al. 2006; Kim et al. 2013). It has been suggested for *C. beticola* and *C. sojina* that sexual reproduction may occur in plant debris at the end of the growing season (Bolton et al. 2012; Kim et al. 2013) and this may also be true for *C. zeina.*

Management of GLS in commercial maize in USA and South Africa is conducted by foliar fungicide treatments (Mallowa et al. 2015; Ward et al. 1999), resulting in high selection pressure for fungicide resistance in the pathogens. Our study has indicated high levels of genetic diversity in *C. zeina* from South African maize growing regions, which suggests mechanisms for mutation or recombination within the population, with implications for the development of fungicide resistance. Quinone outside inhibitor (QoI) fungicides (i.e. strobilurins) are popular foliar fungicides due to perceived additional plant health benefits (Mallowa et al. 2015), however development of resistance by *Cercospora* species has been documented. Isolates of *C. sojina* from soybean have exhibited reduced sensitivity to QoI fungicides in several USA states (Standish et al. 2015; Zeng et al. 2015; Zhang et al. 2012), and QoI resistance in *C. beticola* from sugarbeet has been reported from USA and Italy (Birla et al. 2012;   Bolton et al. 2013). QoI resistance has not been reported from *Cercospora* pathogens of maize, however baseline sensitivities have been established for monitoring purposes (Bradley and Pedersen 2011). To date, there have been no published reports to our knowledge of fungicide resistance in *Cercospora* species causing GLS in maize fields. The availability of genome sequence information for *C. zeina* in future would allow the design of diagnostic assays for common mutations that lead to QoI resistance, such as the G143A mutation in the cytochrome B gene reported in other *Cercospora* spp. (Zeng et al. 2015).

Resistance of maize germplasm to GLS is quantitative in nature with many QTL identified in different maize growing countries, although very few studies have defined the causal species (Berger et al. 2014). The pathosystem is not characterized by pathogen races that are recognised by cognate resistance genes (Balint-Kurti et al. 2008), however the high diversity and evidence for sexual recombination reported in this study reveal a potential for differences in virulence between *C. zeina* isolates. Currently, it is thought that QTL in maize germplasm are equally effective against *C. zeina* and *C. zeae-maydis* (Carson et al. 2002; Wang et al. 1998), however this has not been

tested systematically. In the case that *Cercospora* species-specific maize resistance genes are discovered in future, this would have implications for resistance breeding in Africa in three ways: (i) germplasm screening in disease nurseries should be conducted locally, since breeding for resistance in a region where *C. zeae-maydis* is dominant may not produce effective resistance against *C. zeina*; (ii) breeding programs need to take into account the previously unknown level of genetic diversity, for example by inoculating with diverse isolates of *C. zeina*, and (iii) maize breeders would need to prioritize *C. zeina*-specific resistance genes/mechanisms. Resistance breeding is of paramount important to small-holder maize farmers in South Africa, who generally do not implement fungicide control of GLS.

In this paper we have illustrated how an application of genomics (i.e. genome sequencing) can be used to rapidly gain information about a pathogen's genetic diversity and potential mode of reproduction, providing valuable information and tools for crop protection agencies in South Africa and worldwide where maize production is constrained by GLS.

LITERATURE CITED

Agapow, P.-M., and Burt, A. 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1:101-102.

Arie, T., Kaneko, I., Yoshida, T., Noguchi, M., Nomura, Y., and Yamaguchi, I. 2000. Mating-type genes from asexual phytopathogenic ascomycetes *Fusarium oxysporum* and *Alternaria alternata*. Mol. Plant-Microbe In. 13:1330-1339.

Balint-Kurti, P. J., Wisser, R., and Zwonitzer, J. C. 2008. Use of an advanced intercross line population for precise mapping of quantitative trait loci for gray leaf spot resistance in maize. Crop Sci. 48:1696-1704.

Barnes, I., Wingfield, M. J., Carbone, I., Kirisits, T., Wingfield, B. D. 2014. Population structure and diversity of an invasive pine needle pathogen reflects anthropogenic activity. Ecol. Evol. 4: 3642-3661.

Berger, D. K., Carstens, M., Korsman, J. N., Middleton, F., Kloppers, F. J., Tongoona, P., and Myburg, A. A. 2014. Mapping QTL conferring resistance in maize to gray leaf spot disease caused by *Cercospora zeina*. BMC Genetics 15:60.

Birla, K., Rivera-Varas, V., Secor, G. A., Khan, M. F. R., and Bolton, M. D. 2012. Characterization of cytochrome b from European field isolates of *Cercospora beticola* with quinone outside inhibitor resistance. Eur. J Plant Pathol. 134:475-488.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578-579.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bolton, M. D., Rivera, V., and Secor, G. 2013. Identification of the G143A mutation associated with QoI resistance in *Cercospora beticola* field isolates from Michigan, United States. Pest Manag. Sci. 69:35-39.

Bolton, M. D., Secor, G. A., Rivera, V., Weiland, J. J., Rudolph, K., Birla, K., Rengifo, J., and Campbell, L. G. 2012. Evaluation of the potential for sexual reproduction in field populations of *Cercospora beticola* from USA. Fungal Biol. 116:511-521.

Bolton, M. D., de Jonge, R., Inderbitzin, P., Liu, Z., Birla, K., Van de Peer, Y., Subbaroa, K.V., Thomma, B.P., and Secor, G. A. 2014. The heterothallic sugarbeet pathogen *Cercospora beticola* contains exon fragments of both MAT genes that are homogenized by concerted evolution. Fungal Genet. Biol. 62:43-54.

Bradley, C. A., and Pedersen, D. K. 2011. Baseline sensitivity of *Cercospora zeae-maydi*s to quinone outside inhibitor fungicides. Plant Dis. 95:189-194.

Brunelli, K. R., Dunkle, L. D., Sobrinho, C. A., Fazza, A. C., and Camargo, L. E. A. 2008. Molecular variability in the maize grey leaf spot pathogens in Brazil. Genet. Mol. Biol. 31:938-942.

Carson, M. L., Goodman, M. M., and Williamson, S. M. 2002. Variation in aggressiveness among isolates of *Cercospora* from maize as a potential cause of genotype-environment interaction in gray leaf spot trials. Plant Dis. 86:1089-1093.

Coppin, E., Debuchy, R., Arnaise, S., and Picard, M. 1997. Mating types and sexual development in filamentous ascomycetes. Microbiol. Mol. Biol R 61:411-428.

Crous, P., Groenewald, J., Groenewald, M., Caldwell, P., Braun, U., and Harrington, T. 2006. Species of *Cercospora* associated with grey leaf spot of maize. Stud.Mycol. 55:189 - 197.

Dunkle, L., and Levy, M. 2000. Genetic relatedness of African and United States populations of *Cercospora zeae-maydis*. Phytopathology 90:486 - 490.

Earl, D. A., and vonHoldt, B. M. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359-361.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol.Ecol. 14:2611-2620.

Faircloth, B. C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8:92-94.

Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Goodwin, S. B., Dunkle, L. D., and Zismann, V. L. 2001. Phylogenetic analysis of *Cercospora* and *Mycosphaerella* based on the internal transcribed spacer region of ribosomal DNA. Phytopathology 91:648-658.

Groenewald, M., Groenewald, J. Z., Linde, C. C., and Crous, P. W. 2007. Development of polymorphic microsatellite and single nucleotide polymorphism markers for *Cercospora beticola* (*Mycosphaerellaceae*). Mol. Ecol. Notes 7:890-892.

Groenewald, M., Linde, C. C., Groenewald, J. Z., and Crous, P. W. 2008. Indirect evidence for sexual reproduction in *Cercospora beticola* populations from sugar beet. Plant Pathol. 57:25-32.

Groenewald, M., Groenewald, J. Z., Harrington, T. C., Abeln, E. C., and Crous, P. W. 2006. Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. Fungal Genet. Biol. 43:813-825.

Grünwald, N. J., Goodwin, S. B., Milgroom, M. G., and Fry, W. E. 2003. Analysis of genotypic diversity data for populations of microorganisms. Phytopathology 93:738-746.

Hurlbert, S. H. 1971. The nonconcept of species diversity: A critique and alternative parameters. Ecology 52:577-586.

Kalinowski, S. T. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes 5:187-189.

Kamvar, Z. N., Tabima, J. F., and Grunwald, N. J. 2014. POPPR: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Peer J 2:e281.

Kim, H., Newell, A. D., Cota-Sieckmeyer, R. G., Rupe, J. C., Fakhoury, A. M., and Bluhm, B. H. 2013. Mating-type distribution and genetic diversity of *Cercospora sojina* populations on soybean from Arkansas: evidence for potential sexual reproduction. Phytopathology 103:1045-1051.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15:1179-1191.

Korsman, J., Meisel, B., Kloppers, F. J., Crampton, B. G., and Berger, D. K. 2012. Quantitative phenotyping of grey leaf spot disease in maize using real-time PCR. Eur. J. Plant Pathol. 133:461-471.

Liu, K. J., and Xu, X. D. 2013. First report of gray leaf spot of maize caused by *Cercospora zeina* in China. Plant Dis. 97:1656.

Mallowa, S. O., Esker, P. D., Paul, P. A., Bradley, C. A., Chapara, V. R., Conley, S. P., and Robertson, A. E. 2015. Effect of maize hybrid and foliar fungicides on yield under low foliar disease severity conditions. Phytopathology 105:1080-1089.

McDermott, J. M., and McDonald, B. A. 1993. Gene flow in plant pathosystems. Annu. Rev. Phytopathol. 31:353-373.

McDonald, B. A. 1997. The population genetics of fungi: tools and techniques. Phytopathology 87:448-453.

Meisel, B., Korsman, J., Kloppers, F., and Berger, D. 2009. *Cercospora zeina* is the causal agent of grey leaf spot disease of maize in southern Africa. Eur. J. Plant Pathol. 124:577-583.

Milgroom, M. G. 1996. Recombination and the multilocus structure of fungal populations. Annu. Rev. Phytopathol. 34:457-477.

Molecular Ecology Resources Primer Development Consortium. 2009. Permanent Genetic Resources added to Molecular Ecology Resources database 1 January 2009–30 April 2009. *Cercospora zeae-maydis* http://tomato.bio.trinity.edu/manuscripts/9-5/mer-09-0045.pdf Mol. Ecol. Resour. 9:1375-1379.

Nadalin, F., Vezzi, F., and Policriti, A. 2012. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics 13 Suppl 14:S8.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283-292.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nei, M., Tajima, F., and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19:153-170.

Neves, D., L., Silva, C., N., Pereira, C., B., Campos, H., D., and Tessmann, D., J. 2015. *Cercospora zeina* is the main species causing gray leaf spot in southern and central Brazilian maize regions. Trop. Plant Pathol. 40:368-374.

Paoletti, M., Rydholm, C., Schwier, E. U., Anderson, M. J., Szakacs, G., Lutzoni, F., Debeaupuis, J-P., Latge, J-P., Denning, D.W., and Dyer, P. S. 2005. Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. Curr. Biol. 15:1242-1248.

Peakall, R., and Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28:2537-2539.

Ranum, P., Pena-Rosas, J. P., and Garcia-Casal, M. N. 2014. Global maize production, utilization, and consumption. Ann. NY Acad. Sci. 1312:105-112.

Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Method Mol. Biol. 132:365-386.

Santana, Q., Coetzee, M., Steenkamp, E., Mlonyeni, O., Hammond, G., Wingfield, M., and Wingfield, B. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. BioTechniques 46:217-223.

Shiferaw, B., Prasanna, B. M., Hellin, J., and M., B. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Secur. 3:307.

Standish, J. R., Tomaso-Peterson, M., Allen, T. W., Sabanadzovic, S., and Aboughanem-Sabanadzovic, N. 2015. Occurrence of QoI fungicide resistance in *Cercospora sojina* from Mississippi soybean. Plant Dis. 99:1347-1352.

Stergiopoulos, I., Groenewald, M., Staats, M., Lindhout, P., Crous, P. W., and De Wit, P. J. 2007. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. Fungal Genet. Biol. 44:415-429.

Stewart, C. N., Jr., and Via, L. E. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. BioTechniques 14:748-750.

Stoddart, J. A., and Taylor, J. F. 1988. Genotypic diversity: estimation and prediction in samples. Genetics 118:705-711.

Szpiech, Z. A., Jakobsson, M., and Rosenberg, N. A. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24:2498-2504.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30:2725-2729.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11:1441-1452.

Wang, J., Levy, M., and Dunkle, L. 1998. Sibling species of *Cercospora* associated with Gray Leaf Spot of Maize. Phytopathology 88:1269 - 1275.

Ward, J. M. J., Stromberg, E. L., Nowell, D. C., and Nutter, F. W. 1999. Gray leaf spot: a disease of global importance in maize production. Plant Dis. 83:884-895.

Wright, S. 1978. Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations. University of Chicago Press.

Yeh, F., Yang, R.-C., Boyle , T.B.J., Ye, Z-H.., and Mao, J. X.. 1999. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

Zeng, F., Arnao, E., Zhang, G., Olaya, G., Wullschleger, J., Sierotzki, H., Ming, R., Bluhm, B.H., Bong, J.P., Fakhoury, J.M., and Bradley, C. A. 2015. Characterization of quinone outside inhibitor fungicide resistance in *Cercospora sojina* and development of diagnostic tools for its identification. Plant Dis. 99:544-550.

Zerbino, D. R., and Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

Zhan, J., Pettway, R. E., and McDonald, B. A. 2003. The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. Fungal Genet. Biol. 38:286-297.

Zhang, G. R., Newman, M. A., and Bradley, C. A. 2012. First report of the soybean frogeye leaf spot fungus (*Cercospora sojina*) resistant to quinone outside inhibitor fungicides in North America. Plant Dis. 96:767.

TABLE 1. Microsatellite markers designed in this study from the draft genome sequences of *Cercospora zeina*

Locus Name	Primer Sequence (5'-3')	Motif repeat	Tag	Panel [#]	No. of Alleles	GenBank Accession No.
*CzSSR*01[α]	F: AATTAATCGTAAGCACGACGA	TGT^15	6-FAM	2	4	KP015832
	R: CTCCCTCCACAACCACAACT					
*CzSSR*02[α]	F: ACGTCTCGGAATCTCTTTCT	TGT^9	6-FAM	2	3	KP015833
	R: TGGGATCTGGAACCCAAATA					
*CzSSR*04[α]	F: GGTTAGCGTGTAGCCGAGTT	GTT^26	VIC	2	4	KP015834
	R: CGACCAAGTGCTTGTCAAC					
*CzSSR*05[α]	F: CTTCGACTACGTTGCGTTGA	ATG^16	VIC	1	3	KP015835
	R: AGCCCTTGACAGCACTGACT					
*CzSSR*06[α]	F: CAGAAAGAAGGCACCAAAGC	TCT^15	6-FAM	2	4	KP015836
	R: GAGCAGGTTTAGTCGGAGGA					
*CzSSR*07[α]	F: CAAGAATGCCAATGATGCTG	ATG^23	PET®	2	12	KP015837
	R: GTCTCCTTTCTGGCGAAGTG					
*CzSSR*08[α]	F: GTAACTCCGCGAGATTCCTG	TGT^20	VIC	2	4	KP015838
	R: AGCAGCAGCAGCAGTAACAA					
*CzSSR*10[α]	F: GCGTTACTTCGAAGGTGCTT	ATG^20	6-FAM	1	3	KP015839
	R: GTTGGTCGTTTGTTTTGTCCT					
*CzSSR*11[α]	F: CGAGACTAATAGCGACGAGAG	TTTA^24	PET®	2	9	KP015840
	R: GAGGCTTTTCTCTCGCAAAT					
*CzSSR*12[α]	F: GAAGGCTTTTCTCTCGCAAA	AATA^19	NED	1	5	KP015841
	R: TTGTCCCTCGGTCGCTTAT					
*CzSSR*13[β]	F: GAGAGATAGTTGCGGCGT	ACAGC^9	NED	1	2	KP015842
	R: GATGATGATTTGAGGAGTGTTG					
*CzSSR*15[β]	F: CATTCTTTGTCCGCGTTC	GGC^10	PET®	2	2	KP015844
	R: CACTCACTTCCCACATAC					
*CzSSR*17[β]	F: GGTGTTTGTCGATAACCTGTG	CCT^10	6-FAM	1	3	KP015846
	R: CAAAATTCCAAAGAACCTCGT					
*CzSSR*18[β]	F: ATGCGTCAAAATCACACTTTC	CTC^8	VIC	1	2	KP015847
	R: AAAGCGTCTCCTCATCGATAC					

[α] = marker designed from the genome sequence of USPA-4.
[β] = marker designed from the genome sequence of CMW 25467
# Amplicons were pooled into two panels (Panel 1 or Panel 2) prior to fragment analysis.

TABLE 2. Overall diversity statistics of *C. zeina* isolates from South Africa

Region	N	n	MLG	eMLG	Hexp	G	G.est	θ	G_{ST}	Nm
RSA	369	260	242	18.3 ± 1.31	0.35	64.26	16.64	0.03	0.08	5.51
KwaZulu-Natal	239	154	154	17.8 ± 1.45	0.35	47.48	15.81	-	-	-
Mpumalanga	110	92	92	19.3 ± 0.80	0.35	77.56	18.83	-	-	-
North West	20	14	14	14.0 ± 0.00	0.18	7.69	7.69	-	-	-

N = Number of isolates from each province

n = Number of haplotypes observed from each province

MLG = Number of multilocus genotypes (MLG) observed. Note that 18 MLG were observed in more than one province and counted for each province, however these duplicates were not counted for the RSA total.

eMLG = The number of expected MLG at the smallest sample size ≥ 10 based on rarefaction (Hurlbert 1971), and standard error based on the eMLG

Hexp = Nei's unbiased gene diversity (Nei 1978)

G = Stoddart and Taylor's Index of MLG diversity (Stoddart and Taylor 1988)

G.est = Stoddart and Taylor's Index of MLG diversity (Stoddart and Taylor 1988) based on rarefaction (Hurlbert 1971)

θ = Weir's Theta (θ) statistic for range (-0.005<θ<0.009), P<0.001

G_{ST} = Estimate of genetic differentiation within populations determined from Nei's gene diversity

Nm = Estimate of gene flow from G_{ST} such that Nm = 0.5 (1-G_{ST})/G_{ST} (McDermott and McDonald 1993)

TABLE 3. Diversity indices for clone-corrected data of *C. zeina* isolates from three provinces in South Africa.

Province	NCC	CC	CF	Na	Ne	$A_R \pm SE$	$PA_R \pm SE$
KwaZulu-Natal	239	154	0.36	3.21±1.84	1.81±0.48	2.80±0.36	0.24±0.07
Mpumalanga	110	92	0.16	3.00±1.80	1.73±0.52	2.45±0.28	0.19±0.06
North West	20	14	0.3	1.64±0.63	1.33±0.39	1.53±0.12	0.15±0.06

NCC = The number of samples in the non-clone-corrected data set

CC = The number of samples in the clone-corrected data set

CF = Clonal Fraction as defined by: CF = 1 - (number of unique genotypes/total number of isolates)

Na = The average number of alleles per marker (clone-corrected dataset)

Ne = The effective number of alleles per marker (clone-corrected dataset)

A_R = Allelic Richness, standardised to a sample size of 14 (clone-corrected dataset)

PA_R = Private Allelic Richness, standardised to a sample size of 14 (clone-corrected dataset)

TABLE 4. Multiplex PCR primer sequences used to identify *MAT1-1* or *MAT1-2* isolates of *C. zeina*.

MAT idiomorph	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Amplicon Size
MAT1-1	TCACCCTTTCACCGTACCCA	CACCTGCCATCCCATCATCTC	631 bp
MAT1-2	CGATGTCACGGAGGACCTGA	GTGGAGGTCGAGACGGTAGA	409 bp

TABLE 5. Distribution of *MAT1-1* and *MAT1-2* idiomorphs of *C. zeina* isolates from commercial farms in South Africa.

Location (Province)	No. of isolates	*MAT1-1*	*MAT1-2*	Ratio[a]	χ^2 [b]	*P*-value[c]
Baynesfield (KZN)	34	10	24	0.4	5.77	0.016*
Cedara (KZN)	54	22	32	0.7	1.85	0.17
Greytown (KZN)	119	42	77	0.6	10.29	0.0013*
Winterton (KZN)	20	11	9	1.2	0.20	0.65
Machadodorp (MP)	88	51	37	1.4	2.23	0.14
Klerksdorp (NW)	20	9	11	0.8	0.20	0.65

a   Ratio of *MAT1-1*/*MAT1-2*

b χ^2 value calculated for the deviation from the expected 1:1 ratio.

c   Probability of a greater χ^2 value under the null hypothesis of a 1:1 ratio (with 1 degree of freedom).

*   Mating type departs from the null hypothesis of a 1:1 ratio at $P < 0.05$.

Figure Legends

Fig. 1. Map showing collection sites of *C. zeina* from maize fields in KwaZulu-Natal, Mpumalanga and North West Provinces of South Africa. The inset shows the position of the provinces in South Africa.

Fig. 2. Population structure of *C. zeina* isolates from South Africa based on STRUCTURE analyses using 14 microsatellite markers. Individuals are sorted by province (KwaZulu-Natal, Mpumalanga, North West) from which they were collected. The proportion of inferred ancestry of each individual (vertical bar) to either one of two, three or four clusters (K=2, 3, 4), as represented by different colours, is shown. Bar plots indicate that there is a high level of admixture in the South African isolates of *C. zeina*.

Fig. 3. Principle coordinate analysis (PCoA) of haplotypes of *C. zeina* isolates shows lack of regional population differentiation. Nei's genetic distance was calculated between the SSR-based haplotypes of isolates from the three provinces in South Africa (clone-corrected dataset), and the relationships determined using PCoA with GENALEX software. Isolates were collected from KwaZulu-Natal (○), Mpumalanga (Δ), and North West Province (X).

Fig. 4. The *C. zeina* CMW 25467 genome sequence contains the MAT1-1 idiomorph and additional exonic fragments of *MAT1-1-1* and *MAT1-2-1* genes, which facilitated design of a multiplex PCR *MAT* assay for *C. zeina*. **A:** *C. zeina* CMW 25467 has the intact *MAT1-1* idiomorph and exonic fragments of *MAT1-1-1*. The CMW 25467 *MAT1-1* idiomorph on contig 24 and predicted open reading frame (ORF) (i.e. exons) are identical to those for the *MAT1-1* idiomorph from *C. zeina* CPC 11998 (DQ264748; Groenewald et al. 2006). The CMW 25467 draft genome contains additional exonic fragments of *MAT1-1-1* on five contigs.. **B**: CMW 25467 draft genome lacks a *MAT1-2* idiomorph, but contains *MAT1-2-1* exonic fragments on six contigs. Reconstruction of a *MAT1-2-1* ORF from these exonic fragments produces a sequence with high identity to the

predicted ORF (exons) of CPC 11995, which contains an intact *MAT1-2* idiomorph (DQ264762; Groenewald et al. 2006). *C. zeina MAT* primer pairs (**A** (*MAT1-1-1*): CercosporaMat1f & CercosporaMat1r; **B** (*MAT1-2-1*): CercosporaMat2f & CercosporaMat2r) developed by Groenewald et al. (2006) for *MAT* idiomorph screening are indicated by green arrows, however these were not found to amplify optimally under the conditions tested in this study. Therefore alternative non-degenerate *C. zeina* specific *MAT* primers specific to *MAT* idiomorphs and that would not amplify from *MAT* exonic fragments were developed (blue arrows; **A** (*MAT1-1-1*): CzMAT1F & CzMAT1R; **B** (*MAT1-2-1*): CzMAT2F & CzMAT2R). Diagnostic PCR products for *MAT1-1-1* (631 bp) and *MAT1-2-1* (409 bp) are shown with horizontal blue lines in panel **A** and **B**, respectively. Vertical blue dotted lines give the 5'end of each primer to illustrate that the primers are specific for intact *MAT* idiomorphs and would not amplify from additional *MAT* exonic fragments. The red and purple arrows represent the alpha box and the HMG box domains in *MAT1-1-1* and *MAT1-2-1*, respectively. The numbers on the right hand side refer to DNA sizes in bp.

Fig. 5. Representative multiplex *MAT* gene PCR used to screen *C. zeina* isolates. Multiplex PCR reactions contained the primers CzMAT1F, CzMAT1R, CzMAT2F and CzMAT2R. Lane 1, no template negative control, Lane 2 (isolate 2011.CD01), lane 3 (isolate 2012.GT14) and lane 4 (isolate 2013.MP02) indicate isolates that contain *MAT1-2-1* (409bp). Lane 5 (isolate 2012.CD07) and lane 6 (isolate 2013.GT12) represent isolates containing *MAT1-1-1* (631bp). M; O'Gene Ruler 100 bp DNA ladder (Thermo Scientific).

Suppl. Table 1. The number of private alleles per microsatellite marker observed for *Cercospora zeina* in each Province in South Africa

Suppl. Table 2. Microsatellite marker alleles and mating type for collection of *Cercospora zeina* isolates from South Africa.

Suppl. Fig. 1. Mean genotypic diversity plotted against the number of microsatellite markers, which shows a plateau at 12 markers.

Suppl. Fig. 2. A neighbour joining tree showing the genetic distance (Da) (Nei 1983) between 242 unique haplotypes of *Cercospora zeina* isolates from South Africa based on data obtained using 14 microsatellite markers in this study. Isolates from the province of Mpumalanga are represented in green, North West in purple and KwaZulu-Natal in black. Taxon markers are indicated in the key. More than one isolate from the same geographical region that branched together were further compressed. Isolates from all three geographic locations are randomly distributed throughout the tree and no significant bootstrap values for branch support were obtained from 1000 re-sampling replicates. The NJ tree indicates significant recombination and gene flow between isolates, and the lack of boostrap support is consistent with the absence of regional population differentiation.
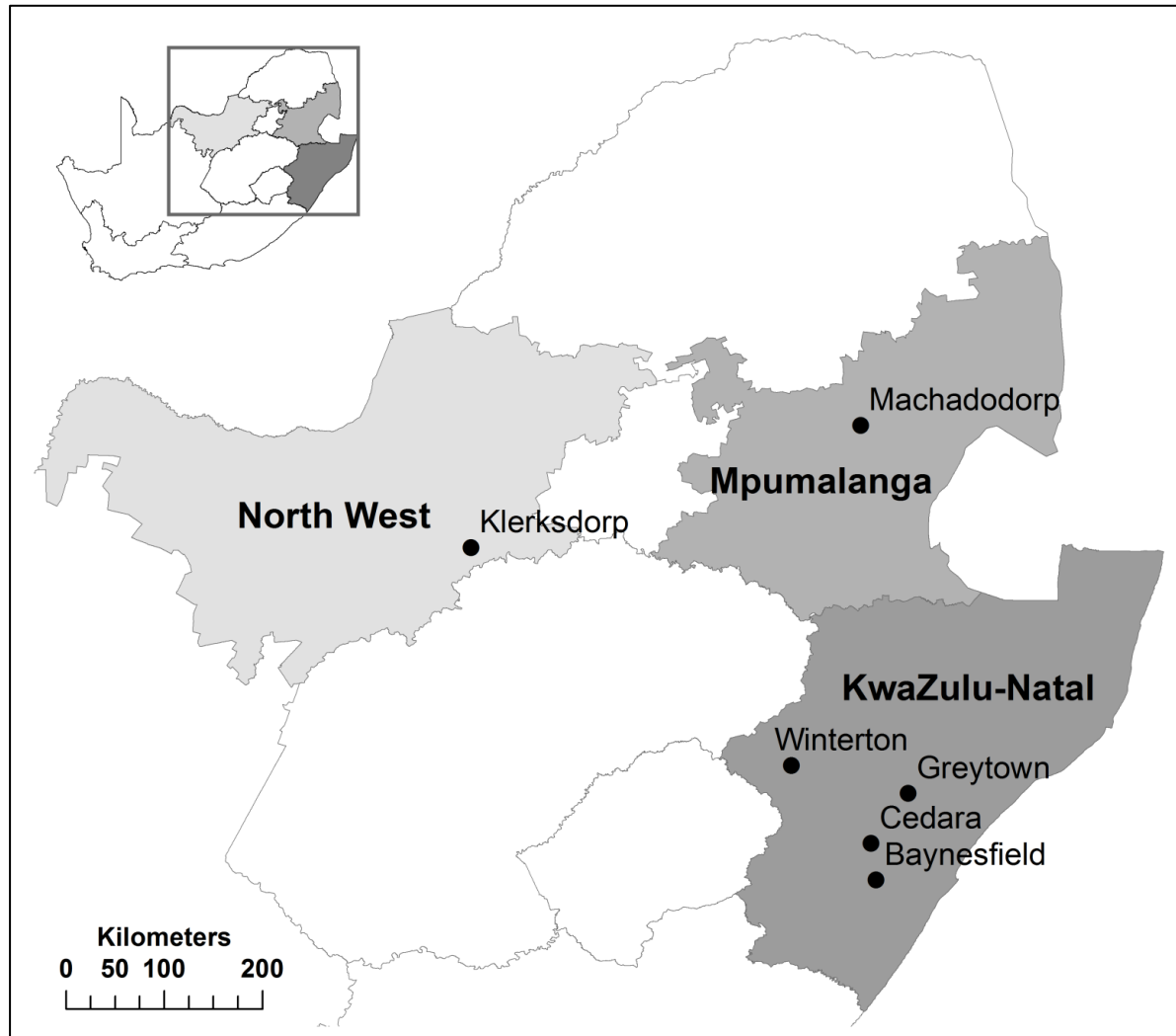
Fig. 1.   Muller et al.   *Phytopathology*

Fig. 1. Map showing collection sites of *C. zeina* from maize fields in KwaZulu-Natal, Mpumalanga and North West Provinces of South Africa,. The inset shows the position of the provinces in South Africa.
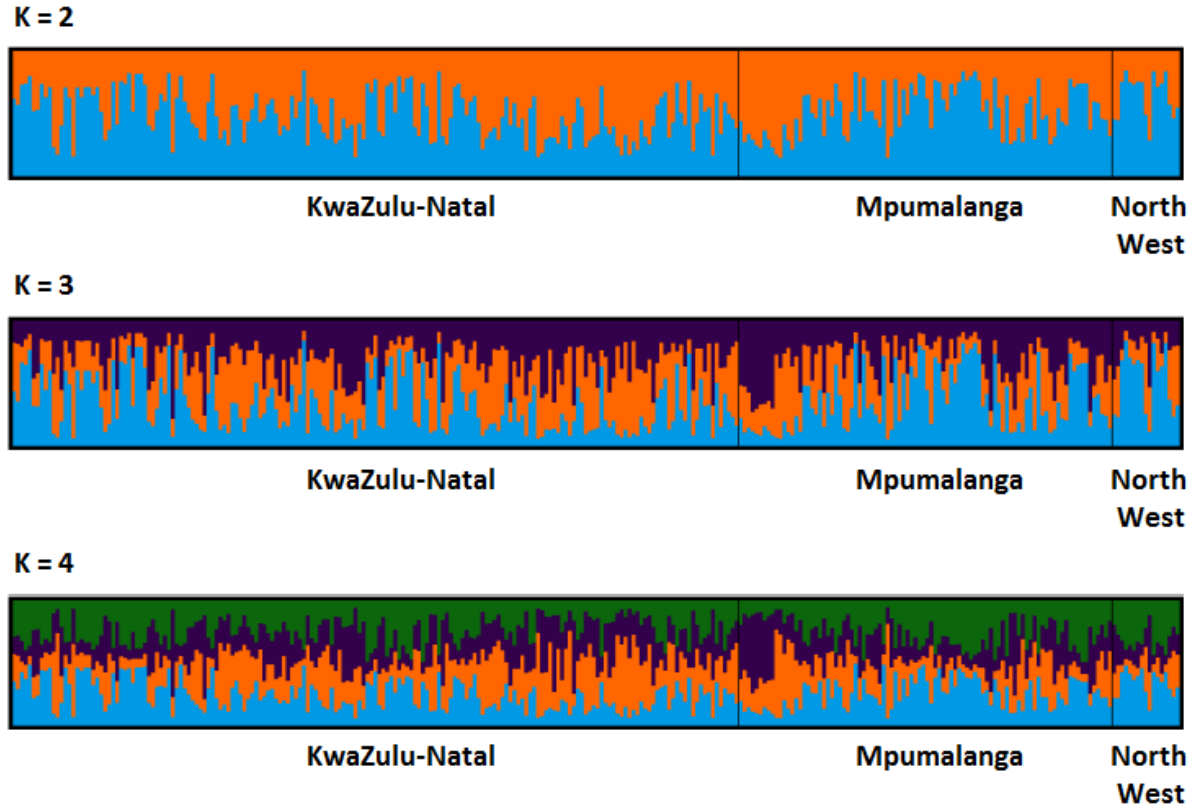
Fig. 2.   Muller et al.   *Phytopathology*

Fig. 2. Population structure of *C. zeina* isolates from South Africa based on STRUCTURE analyses using 14 microsatellite markers. Individuals are sorted by province (KwaZulu-Natal, Mpumalanga, North West) from which they were collected. The proportion of inferred ancestry of each individual (vertical bar) to either one of two, three or four clusters (K=2, 3, 4), as represented by different colours, is shown. Bar plots indicate that there is a high level of admixture in the South African isolates of *C. zeina*.

Fig. 3.   Muller et al.   *Phytopathology*

Fig. 3. Principle coordinate analysis (PCoA) of haplotypes of *C. zeina* isolates shows lack of regional population differentiation. Nei's genetic distance was calculated between the SSR-based haplotypes of isolates from the three provinces in South Africa (clone-corrected dataset), and the relationships determined using PCoA with GENALEX software. Isolates were collected from KwaZulu-Natal (○), Mpumalanga (△), and North West Province (✗).
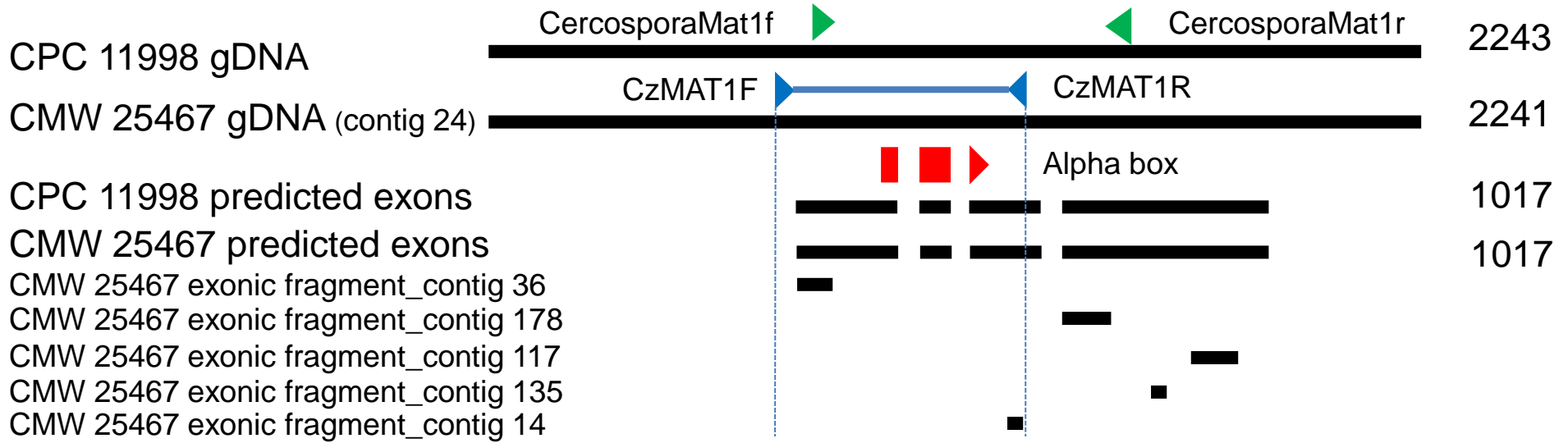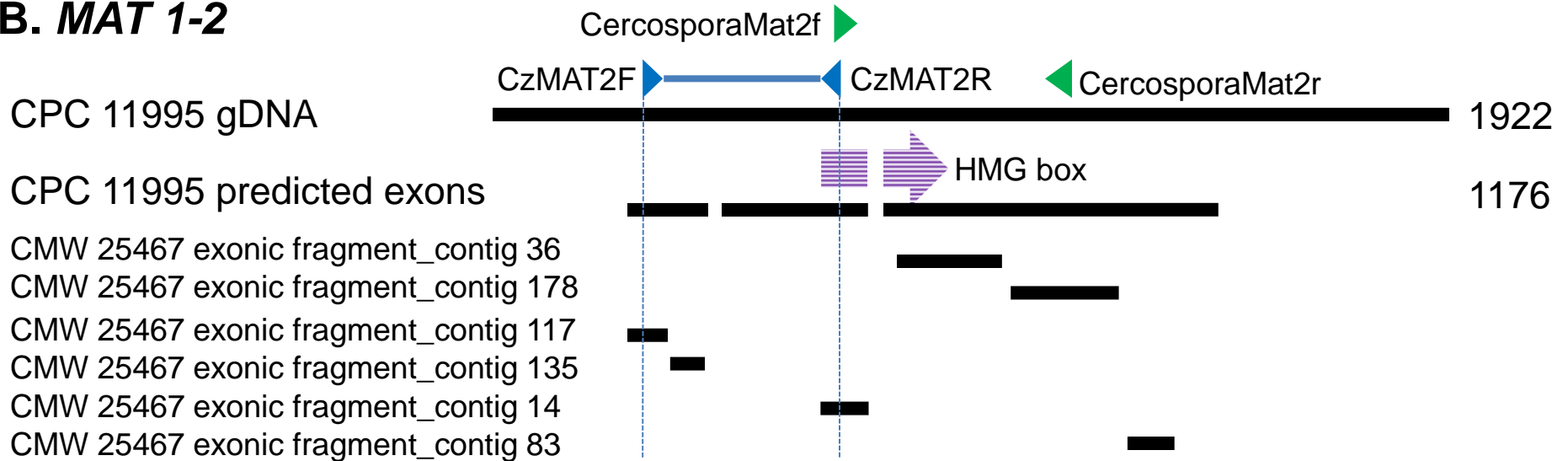
Fig. 4. The *C. zeina* CMW 25467 genome sequence contains the *MAT1-1* idiomorph and exonic fragments of *MAT1-1-1* and *MAT1-2-1* genes, which facilitated design of a multiplex PCR *MAT* assay for *C. zeina*.
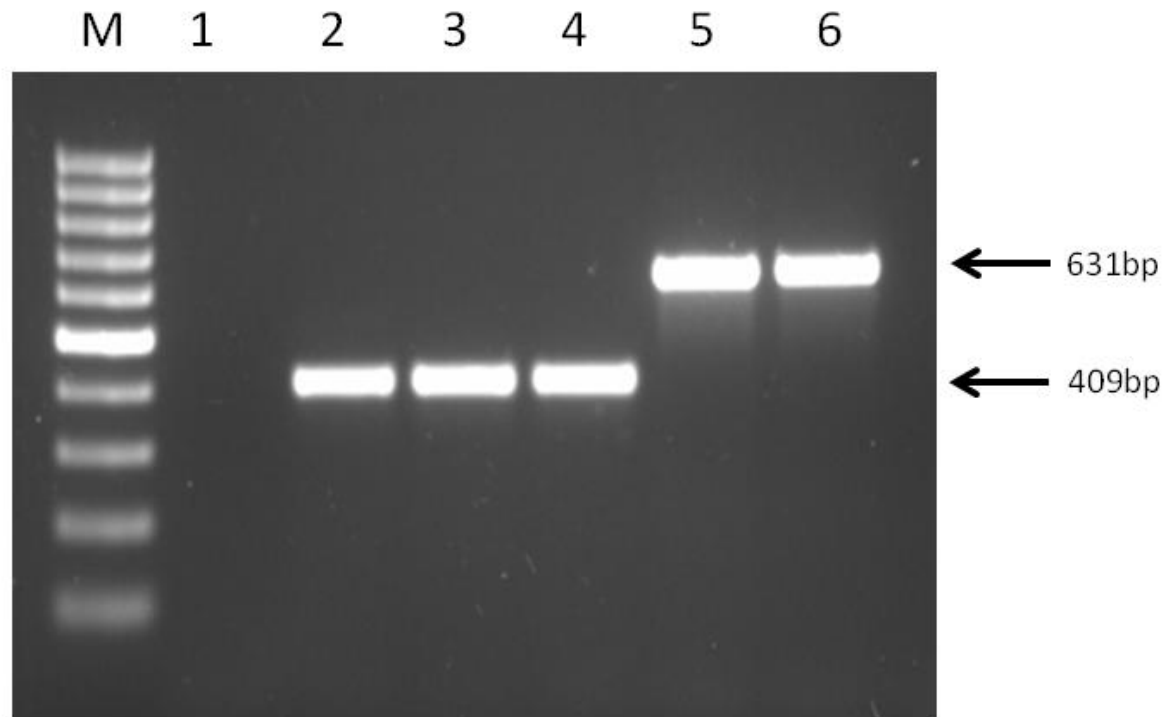
Fig. 5. Muller et al. *Phytopathology*

Fig. 5. Representative multiplex *MAT* gene PCR used to screen *C. zeina* isolates. Multiplex PCR reactions contained the primers CzMAT1F, CzMAT1R, CzMAT2F and CzMAT2R. Lane 1, no template negative control, Lane 2 (isolate 2011.CD01), lane 3 (isolate 2012.GT14) and lane 4 (isolate 2013.MP02) indicate isolates that contain *MAT1-2-1* (409bp). Lane 5 (isolate 2012.CD07) and lane 6 (isolate 2013.GT12) represent isolates containing *MAT1-1-1* (631bp). M; O'Gene Ruler 100 bp DNA ladder (Thermo Scientific).