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

DOI: 10.1111/wre.12260

Genetic diversity and population structure of *Striga hermonthica* populations from Kenya and Nigeria

N N UNACHUKWU*†, A MENKIR*, I Y RABBI*, M OLUOCH*, S MURANAKA‡, A ELZEIN*§, G ODHIAMBO¶, E O FAROMBI† & M GEDIL*

*International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, †Department of Biochemistry, University of Ibadan, Ibadan, Nigeria, ‡Japan International Research Center for Agricultural Sciences (JIRCAS), c/o International Institute of Tropical Agriculture, Ibadan, Nigeria, §Department of Botany and Agricultural Biotechnology, University of Khartoum, Khartoum, Sudan, and ¶Department of Applied Plant Science, School of Agriculture and Food Security, Maseno University, Maseno, Kenya

Received 20 December 2016
Revised version accepted 8 May 2017
Subject Editor: Maurizio Vurro, CNR, Bari, Italy

Summary

Striga hermonthica is a parasitic weed that poses a serious threat to the production of economically important cereals in sub-Saharan Africa. The existence of genetic diversity within and between *S. hermonthica* populations presents a challenge to the successful development and deployment of effective control technologies against this parasitic weed. Understanding the extent of diversity between *S. hermonthica* populations will facilitate the design and deployment of effective control technologies against the parasite. In the present study, *S. hermonthica* plants collected from different locations and host crops in Kenya and Nigeria were genotyped using single nucleotide polymorphisms. Statistically significant genetic differentiation (F_{ST} = 0.15, P = 0.001) was uncovered between populations collected from the two countries. Also, the populations collected in Nigeria formed three distinct subgroups. Unique loci undergoing selection were observed between the Kenyan and Nigerian populations and among the three subgroups found in Nigeria. *Striga hermonthica* populations parasitising rice in Kenya appeared to be genetically distinct from those parasitising maize and sorghum. The presence of distinct populations in East and West Africa and in different regions in Nigeria highlights the importance of developing and testing *Striga* control technologies in multiple locations, including locations representing the geographic regions in Nigeria where genetically distinct subpopulations of the parasite were found. Efforts should also be made to develop relevant control technologies for areas infested with 'rice-specific' *Striga* spp. populations in Kenya.

Keywords: weed biology, population genetics, outlier analysis, positive selection, ecotypes.

UNACHUKWU NN, MENKIR A, RABBI IY, OLUOCH M, MURANAKA S, ELZEIN A, ODHIAMBO G, FAROMBI EO & GEDIL M (2017). Genetic diversity and population structure of *Striga hermonthica* populations from Kenya and Nigeria. *Weed Research*. https://doi.org/10.1111/wre.12260.

Introduction

Striga hermonthica (Delile) Benth. is an obligate hemiparasitic plant that belongs to the family Orobanchaceae. Globally, it is considered to be the most economically important parasitic weed (Parker & Riches, 1993). It infests 57% of the total area in sub-Saharan Africa under cereal production (Sauerborn, 1991) and has shown a steady increase in geographic distribution and infestation level, particularly in sub-Saharan Africa (Ejeta & Gressel, 2007), due to the dispersal mechanisms of *S. hermonthica* seeds and lack of

Correspondence: M Gedil, International Institute of Tropical Agriculture (IITA), Headquarters & West Africa Hub, PMB 5320, Oyo Road, Ibadan 200001, Nigeria. Tel: (+234) 700 800 4482 ext. 2703; Fax: +44 208 7113785; E-mail: M.Gedil@cgiar.org

knowledge and means to properly control the parasite (Berner *et al.*, 1994). Changes in climatic conditions may further increase the geographic distribution and invasive potential of *Striga,* as habitats suitable for the parasite's growth are predicted to expand (Mohamed *et al.*, 2006). *Striga hermonthica* parasitises cereals, causing significant reduction in growth and performance of these crops. Each *S. hermonthica* plant produces a large number of seeds that go through a phase called 'pre-conditioning', which is characterised by moisture and high temperature for 7–14 days (Parker & Riches, 1993), after which they germinate in response to germination stimulants exuded from a suitable host. On contact with the host root, it attaches to the root and penetrates the root cortex using haustoria. The scanning electron microscopy study by Dorr (1997) showed that invading *Striga* spp. cells perforate the host vascular system using a specialised structure, the osculum, to form xylem-to-xylem connections. The parasite then abstracts water and nutrients from the host roots via the haustorium. Infected host plants show stunting, chlorosis and death in severe cases (Dorr, 1997). A single plant of *S. hermonthica* can inflict an approximately 5% loss in yield on a host plant (Parker & Riches, 1993), and high infestations can lead to total crop failure. Average losses of maize in Kenya are estimated at 57%, while estimates of maize crop area infested are 30–40% in Togo, Mali and Nigeria (De Groote *et al.*, 2008).

A number of control options for *S. hermonthica* exist. Host plant resistance is considered to be the most practical and economical approach for resource-poor smallholder farmers (Kim, 1994). However, breeding host plants with resistance to the parasite is complicated by the ability of *S. hermonthica* to overcome host resistance due to high levels of genetic variation within and between populations (Koyama, 2000). This genetic variation also undermines other control methods, leading to seasonal and geographic variability in the effectiveness of *S. hermonthica* control (Hearne, 2009).

Striga hermonthica is expected to have enormous within- and between-population genetic variation (Bozkurt *et al.*, 2015), which can enable the parasite to evolve and adapt to changing environmental conditions (Koyama, 2000). Therefore, knowledge of the distribution of *S. hermonthica* genetic diversity can facilitate selection of representative testing sites for host resistance screening and evaluation of the viability of other control options. This will allow for the development of genotypes with durable resistance and other control options for effective control across the prevalent broad range of ecotypes of the parasite.

The different diversity studies conducted in various locations across Africa using a variety of markers agree on the presence of relationships between geographic distance and genetic distance among the parasite populations, but disagree on the presence or absence of host-specific *S. hermonthica* populations (Koyama, 2000; Gethi *et al.*, 2005; Estep *et al.*, 2011; Welsh & Mohamed, 2011; Bozkurt *et al.*, 2015). This highlights the need for more extensive studies involving populations collected from different geographic areas and an array of host crops. In the present study, *S. hermonthica* parasitising the major staple crops in agro-ecological zones in two countries were collected and characterised using genotyping-by-sequencing derived single nucleotide polymorphisms (SNP). The populations were collected from maize, rice, pearl millet and sorghum. The main objectives of this study were to characterise the extent of diversity of *S. hermonthica* populations collected in Nigeria and Western Kenya and to determine the level of genetic differentiation between populations collected within and between the two countries.

Methods

Sample collection

Striga hermonthica plants were collected from infested areas in Western Kenya and Northern Nigeria. Leaf samples were collected from infested farms (Fig. 1 and Figure S1, Table S1). In each site, *S. hermonthica* was simultaneously collected from different hosts including pearl millet, sorghum, maize and rice when present in the farming system. Leaves from randomly selected *S. hermonthica* plants were collected in each farm on hosts found at least 1 m away from each other; each sample came from a single S*triga* plant growing on a different individual host within the population. The samples were placed in separate sealable plastic bags containing sufficient silica gel to dry them.

DNA extraction and GBS genotyping

High-quality DNA (showing high molecular weight bands on agarose gel) was extracted from *S. hermonthica* leaf samples using a CTAB extraction method modified from Doyle and Doyle (1987). The extracted DNA samples were digested using EcoR1 to determine whether the extracted DNA can be completely digested, indicating the absence of digest inhibitors. The DNA was then lyophilised and sent to Cornell University for genotyping-by-sequencing (GBS). Genotyping-by-sequencing was performed as described in

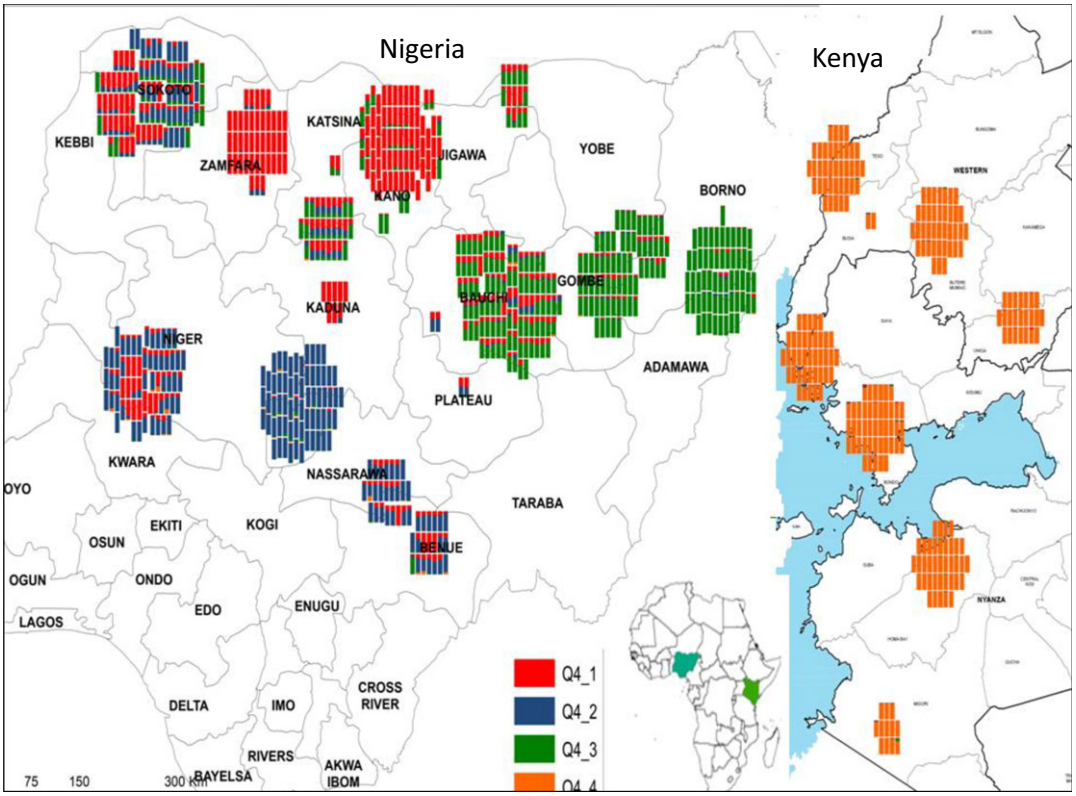

Fig. 1 Geographic distribution of *Striga hermonthica* plants in Nigeria (left) and Kenya (right) analysed in this study. The middle inset is the map of Africa showing the location of Nigeria and Kenya. The right inset highlights the study region in Kenya. Each rectangle represents a plant, and the colour scheme represents the ancestral group(s) of each plant as determined by the study and shown in Fig. 2 at $K = 4$.

Elshire *et al.* (2011). In brief, genomic DNA samples of individuals were digested separately with a selective restriction enzyme known as *Ape*KI, a type II restriction endonuclease that recognises a degenerate 5 bp sequence (GCWGC, where W can be A or T), creates a 3 bp 5′ overhang and is partially methylation sensitive. The digested DNA fragments were then ligated to two types of adapters: common adapters and barcoded adapters. The DNA samples were pooled and amplified using primers complementary to the adapter' sequences. DNA fragments having adapter combinations of common–common or barcoded–barcoded were not amplified, but those with a combination of common–barcoded fragments are selectively amplified. The PCR products were then purified and sequenced using Illumina Hiseq 2500 (Illumina, USA), after which the SNPs were called using the UNEAK (Universal Network Enabled Analysis Kit) GBS pipeline (Lu *et al.*, 2013), which is part of the TASSEL 3.0 bioinformatics analysis package (Bradbury *et al.*, 2007) (Version: 3.0.166 Date: 17 April 2014). This method does not require a reference sequence; SNP discovery is performed directly within pairs of matched sequence tags and filtered through network analysis (See Lu *et al.*, 2013 for an exhaustive discussion and description of the method). Also, a GBS discovery pipeline, using *S. hermonthica* transcriptome, available on Tassel (Version: 3.0.166 Date: 17 April 2014) was used. In this instance, sequence reads were mapped to the *S. hermonthica* transcriptome.

Genetic diversity and population structure

The GBS-derived SNPs were further filtered using the TASSEL software (Bradbury *et al.*, 2007) to retain only polymorphic SNPs with a maximum of 10% missing values and a minimum and maximum allele frequency of 0.05 and 0.95 respectively. The final filtered data comprised 1029 individuals and 1576 SNP markers for further analysis**.** Basic diversity indices for each population were calculated using PowerMarker (Liu & Muse, 2005) and GenAlex version 6.41 (Peakall & Smouse, 2006). The pairwise genetic distance (identity-by-state, IBS) matrix was calculated among all individuals using PLINK (Purcell *et al.*, 2007). A Ward's minimum variance hierarchical cluster dendrogram was built from the IBS matrix using the analyses of phylogenetics and evolution (ape) package (Paradis

et al., 2004) implemented in R (R Core Team, 2015). Population structures of the *S. hermonthica* samples from Nigeria and Kenya were determined using two complementary approaches: (i) a model-based maximum likelihood estimation of ancestral subpopulations using *ADMIXTURE* (Alexander *et al.*, 2009) and (ii) discriminant analysis of principal components (DAPC) (Jombart *et al.*, 2010). The *ADMIXTURE* method assumes linkage equilibrium among loci and Hardy–Weinberg equilibrium within ancestral populations (Alexander *et al.*, 2009). For *ADMIXTURE* analysis, the number of subpopulations, *K*, varied from 1 to 10. The most appropriate *K* value was selected after considering the 10-fold cross-validations, whereby the best *K* exhibits low cross-validation error compared with other *K* and correspondence of the results with the clustering pattern obtained by the hierarchical cluster. The clustering pattern obtained from *ADMIXTURE* was validated using discriminant analysis of principal components (DAPC) using the R package 'adegenet' (Jombart, 2008). DAPC involves first using *K*-means analysis to infer the optimal number of clusters of PCA-transformed SNP data by varying the possible number of clusters from 2 to 40 and then assessing the best supported model by Bayesian information criterion. Then, DAPC was carried out on the identified clusters using the first 55 principal components. The membership probabilities of each individual for the different groups were obtained from DAPC, and the results of DAPC analysis, *ADMIXTURE* and the hierarchical tree are compared.

The fixation index (F_{ST}) and standardised F_{ST} (F'_{ST}) of the observed populations were assessed using analysis of molecular variance (AMOVA) implemented in GenAlex 6.41. These analyses were also performed for plants collected in Nigeria and Kenya partitioned based on their host plants and geographic locations. Correlations between pairwise linearised F_{ST} values and geographic distance matrices were calculated using the Mantel test, after 1000 random iterations, as implemented in GenAlex software version 6.41.

To detect whether there were markers under selection within the population structure observed above, we used the hierarchical Bayesian method described in Beaumont and Balding (2004) as implemented in BayeScan 2.1 software (Foll & Gaggiotti, 2008). BayeScan assumes that allele frequencies within populations follow a multinomial Dirichlet distribution (Balding & Nichols, 1995) with F_{ST} parameters being a function of population-specific components shared among all loci (β) and of locus-specific components shared among all populations (α). For a given locus, departure from neutrality is assumed when the locus-specific component is required to explain the observed pattern of diversity. BayeScan directly infers the posterior probability of each locus to be under the effect of selection by defining and comparing two alternative models: one model includes the locus-specific component, while the other excludes it. The ratio of the model posterior probabilities is used to calculate the posterior odds (PO), which measures how much more likely the model with selection is, compared with the model without selection (Balding & Nichols, 1995). The estimation of model parameters was automatically tuned on the basis of short pilot runs (10 pilot runs, length 5000 and burn in 50 000). The sample size was set to 5000 and the thinning interval to 10 resulting in a total chain length of 100 000 iterations. False discovery rate (FDR) was used to control for multiple testing. To identify loci under selection, the posterior distribution of α_i was used; a positive value suggests that locus '*i*' is subject to directional selection, whereas a negative value suggests that stabilising selection is tending to homogenise allele frequencies over the populations. Loci were then ranked according to their estimated posterior probabilities. An R function (as provided in BAYESCAN) was used to identify and plot outlier loci using different criteria.

Results

Genetic diversity across countries

A total of 1576 SNPs from GBS was used to investigate population structure in 254 *S. hermonthica* plants collected in Kenya and 775 plants collected in Nigeria. The plants from Kenya had a higher genetic diversity as measured by effective number of alleles, observed and expected heterozygosity and Shannon's information index in comparison with the Nigerian samples (Table 1). In contrast, we found more private alleles in the Nigerian plants (61) compared with the plants from Kenya (11) (Table 1).

Table 1 Mean allelic patterns across populations of *Striga hermonthica* collected in Kenya and Nigeria

Population	Kenya	Nigeria
Ne	1.405 (0.009)	1.315 (0.008)
I	0.380 (0.006)	0.324 (0.006)
Private Alleles	11	161
Ho	0.284	0.210
He	0.245 (0.005)	0.200 (0.004)

Standard error is given in parenthesis, Ne = no. of effective alleles I = Shannon's information index = −1* Sum (pi * Ln (pi)), private alleles = no. of alleles unique to a population/group, Ho = observed heterozygosity, He = expected heterozygosity.

The hierarchical cluster analysis separated all the plants into two major groups with the first group containing all the plants from Kenya and the second group comprising plants from Nigeria (Fig. 2). The Nigerian plants were further split into three major subclusters (Fig. 2). *ADMIXTURE* analysis confirmed the results of hierarchical cluster analysis. At $K = 2$, the populations were separated by country of origin, while at $K = 4$, the plants from Nigeria were further divided into three genetic groups (Fig. 2). DAPC analysis also revealed the presence of four genetic groups (Fig. 3), one comprising the Kenyan plants and the other three the Nigerian plants. The results from these three analyses were consistent and showed good correspondence (Fig. 2), indicating that the population structure within the plants had been correctly identified.

Analysis of molecular variance (AMOVA) revealed a moderate to high (Hamrick, 1982) level ($F_{ST} = 0.15$) of genetic differentiation between *S. hermonthica* plants collected from Nigeria and plants from Kenya, which was statistically significant ($P = 0.001$) (Table 2).

Genetic structure within Kenya

Analysis of molecular variance of the Kenyan plants revealed an F_{ST} value of 0.021 ($P = 0.001$) among populations collected from different locations (Table 2), and Mantel's test detected a strong relationship between the geographic distance between sampling locations and their pairwise linearised F_{ST} values ($R^2 = 0.33$, $P = 0.01$, Figure S2). When the *Striga* plants were grouped based on host plants, a low level of differentiation ($F_{ST} = 0.02$) was observed (Table 2). The pairwise F_{ST} value between *Striga* plants collected from rice and maize (0.049) and rice and sorghum (0.047) was statistically significant ($P = 0.001$), whereas the value between plants collected from maize and sorghum was not statistically significant (0.002, $P = 0.108$) (Table 3). DAPC was used to further explore the extent of genetic diversity and differentiation between plants collected from different locations and different host crops. The Kenyan plants did not form clusters based on collection sites, but when clustering was performed based on hosts, *S. hermonthica* plants with rice as hosts clustered away from those with maize and sorghum as hosts (Fig. 4).

Genetic structure within Nigeria

The three clustering methods used to investigate the *S. hermonthica* plants collected in Nigeria showed the presence of three genetic groups (Fig. 2). Mantel's test

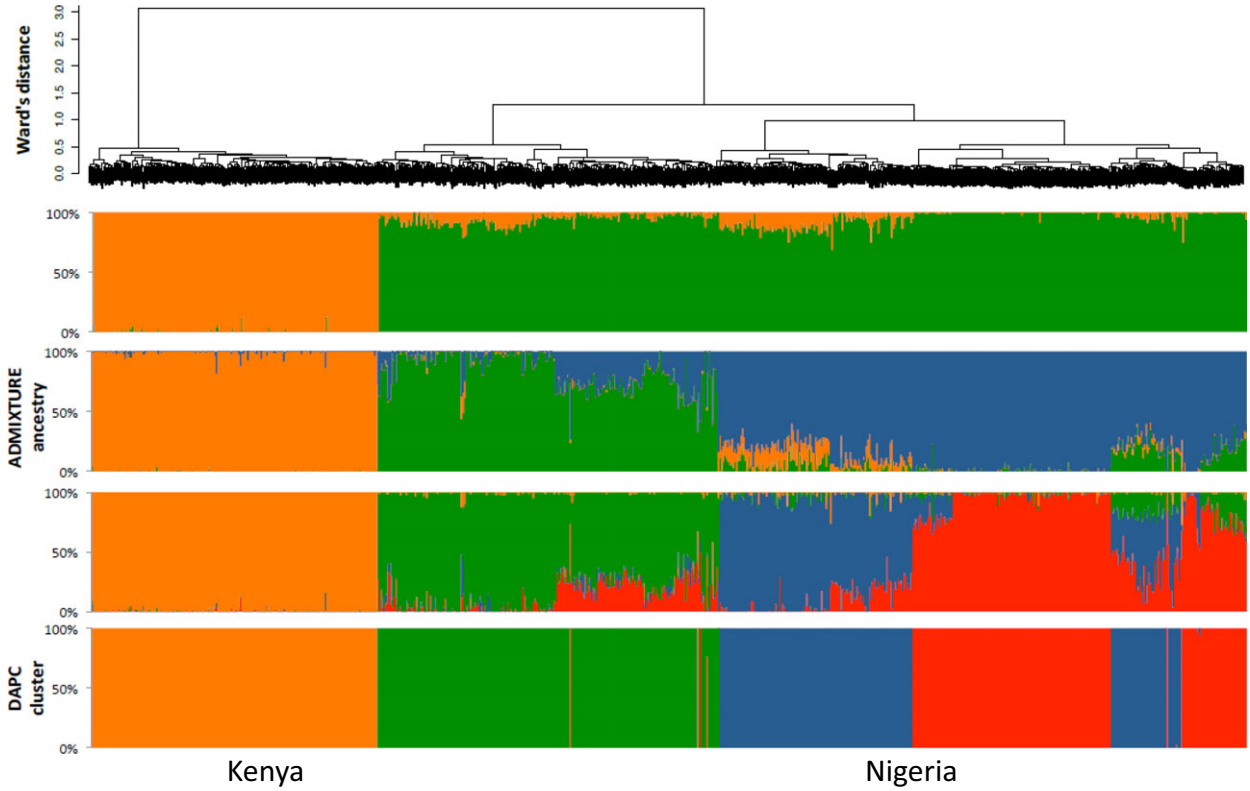

Fig. 2 Combined Dendrogram, *ADMIXTURE** and DAPC plots showing population structure correspondence between the three plots. (**ADMIXTURE* at $k = 2$, 3 and 4 progressively separates the entire samples into Nigerian and Kenyan plants and then splits the Nigerian plants into three groups. (The groups of the DAPC plot and the arms of the dendrogram correspond to $K = 4$ of the *ADMIXTURE* plot)

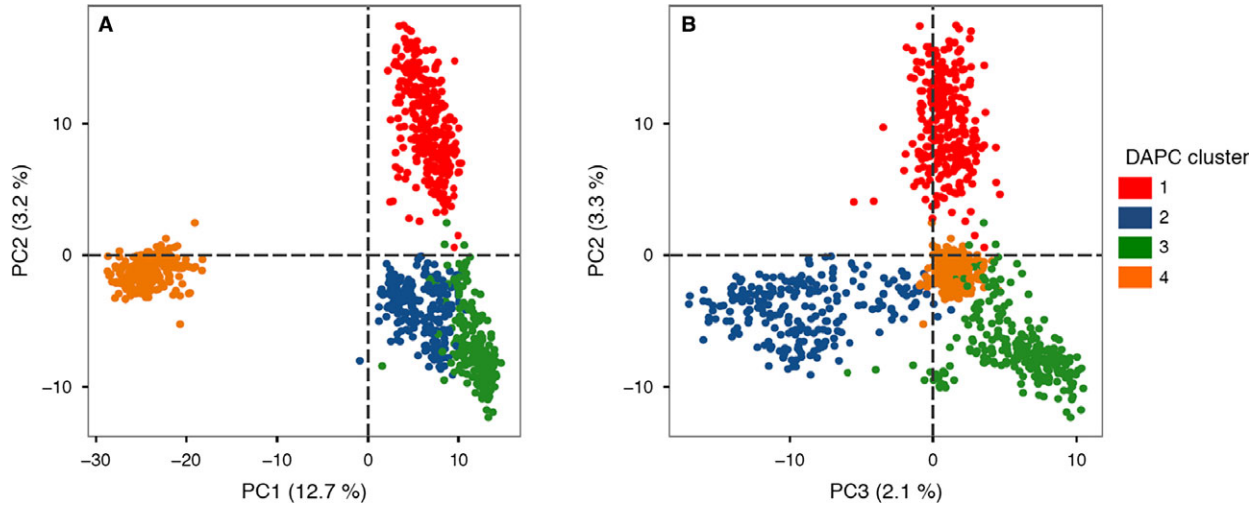

Fig. 3 DAPC plots of all the *Striga hermonthica* plants sampled. Points are plotted along PC1 and PC2 (A) and PC 2 and PC3 (B). The clusters in each set are consistent with the population structure (at K 4) from Fig 2. (Groups 1, 2 and 3 = Nigerian populations, group 4 = Kenyan population).

detected a very weak relationship between the geographic distance between sampling locations and their pairwise linearised F_{ST} values ($R = 0.030$, $P = 0.05$) (Figure S3). The pairwise F_{ST} among the three genetic groups varied from 0.045 to 0.054 and was statistically significant ($P = 0.001$) (Table 4).

Table 2 Summary of AMOVA results

Striga hermonthica population	F_{ST}	P	F'_{ST}
Between Nigerian and Kenyan populations	0.147	0.001	0.203
Among Kenyan plants from different locations	0.021	0.001	0.028
Among Kenyan population on different hosts	0.019	0.003	0.026
Among the three observed Nigerian groups	0.053	0.001	0.072
Among Nigerian population on different hosts	0.009	0.001	0.012

F_{ST} = fixation index = (variance among populations/total variance), F'_{ST} = Standardised F_{ST} = (F_{ST}/F_{ST} max), P = significance.

Table 3 Pairwise F_{ST}, P and F'_{ST} of *Striga hermonthica* plants collected from different hosts in Kenya

Hosts	F_{ST}	P	F'_{ST}
Between maize and rice	0.049	0.001	0.067
Between maize and sorghum	0.002	0.108	0.003
Between rice and sorghum	0.047	0.001	0.066

F_{ST} = fixation index = (variance among populations/total variance), F'_{ST} = standardised F_{ST} = (F_{ST}/F_{ST} max), P = significance.

There was significant correlation between PCA1 and longitude ($P < 2 \times 10^{-16}$, Pearson's $r = -0.62$), and PC 2 and latitude ($P < 2 \times 10^{-16}$, Pearson's $r = 0.54$) (Table 5). The latitudinal stratification is also observed in that a plot of co-ordinates of the sample collection sites shows that the distinct *S. hermonthica* groups are found in three separate areas within Nigeria (Fig. 1). Genetic analysis of the groups reveal that group 3 which comprised *Striga* plants collected mostly in north-eastern Nigeria was the most genetically diverse, followed by group 2 that consisted of plants collected

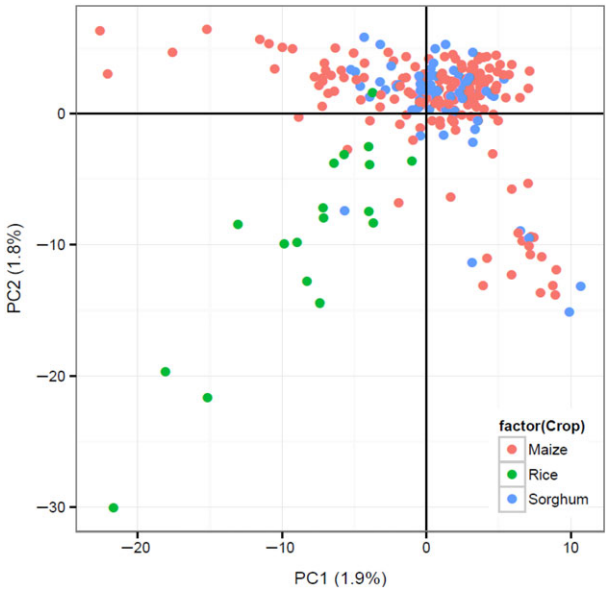

Fig. 4 Plot of *Striga hermonthica* plants collected from Kenya. Plants from rice (in green) cluster separately from those from sorghum and maize.

Table 4 Pairwise F_{ST}, P and F'_{ST} of the three populations of *Striga hermonthica* plants observed in Nigeria

Populations	F_{ST}	*P*	F'_{ST}
Between 1 (north-west) and 2 (central region)	0.057	0.001	0.077
Between 1 (north-west) and 3 (north-east)	0.061	0.001	0.081
Between 2 (central region) and 3 (north-east)	0.043	0.001	0.061

F_{ST} = fixation index = (variance among populations/total variance), F'_{ST} =standardised F_{ST} = (F_{ST}/ F_{ST} max), *P* = significance.

Table 5 Correlation coefficients (Pearson's *r*) between collection location, given as latitude and longitude, and SNP PCA of the Nigerian *Striga hermonthica* collection

	PCA1	PCA2	PCA3
Latitude	0.1(9.74 e-08)	0.54(2.2 e-16)	−0.2(3.99 e-09)
Longitude	−0.62(2.2 e-16)	0.39(2.2 e-16)	0.1(3.0 e-05)

P value is given in parenthesis.

Table 6 Mean allelic patterns across populations observed in Nigeria (1 in the north-west, 2 in the central region and 3 in the north-east)

Population	1	2	3
Ne	1.27(0.013)	1.31(0.012)	1.34(0.012)
I	0.26(0.006)	0.31(0.006)	0.3(0.006)
Private alleles	10	46	19
Ho	0.193	0.2	0.229
He	0.164(0.005)	0.195(0.004)	0.209(0.004)

Standard error is given in parenthesis, Ne = no. of effective alleles, I = Shannon's information index = −1* Sum (pi * Ln (pi)), Ho = observed heterozygosity, private alleles = no. of alleles unique to a population/group, He = expected heterozygosity.

in the central part of the country, while group 1 that contained plants collected in the north-western region had the least genetic diversity (Table 6). Group 2 had the highest number of rare alleles, while group 1 had the least. It was observed that *S. hermonthica* collected from pearl millet in Region 2 clustered with group 1. A low level of genetic differentiation was observed when host plants were used to group the plants (F_{ST} = 0.009, *P* = 0.001) (Table 2). The pairwise F_{ST} values varied between maize and pearl millet (F_{ST} = 0.02), and between maize and sorghum (F_{ST} = 0.009) and between sorghum and pearl millet (F_{ST} = 0.07).

Outlier analysis

The BayeScan analysis was used to determine potential markers under selection in the genetic groups detected by population structure analysis: (i) one Kenyan and three Nigerian populations, (ii) the three Nigerian genetic groups observed by population structure analysis only, (iii) the *S. hermonthica* plants parasitising sorghum, rice and maize in Kenya and (iv) the entire Nigerian *S. hermonthica* collection taken as a whole and the Kenyan population (Figure S3). The Nigerian versus the Kenyan collection showed nine markers to be potentially under selection, FDR ≤ 3%, and log (PO) ≥ 1.05 corresponding to 'strong selection and above' on the Jeffery's scale. All the nine markers also showed positive α values and their F_{ST} values were high (mean F_{ST} = 0.45). Among the four genetic groups observed between Nigeria and Kenya, based on population structure analysis, 23 markers with FDR ≤ 2% and log (PO) ≥ 1.6 corresponding to 'very strong selection and above' on Jeffery's scale of evidence were identified as potentially under selection. All of the 23 markers showed positive α values and relatively high F_{ST} (mean F_{ST} = 0.39) indicative of positive selection, and also the three genetic groups detected within Nigeria revealed three markers to be potentially under selection, FDR ≤ 4% and log (PO) ≥ 1, corresponding to 'strong selection and above' on the Jeffery's scale, all of which also showed positive α values and relatively high F_{ST} (mean F_{ST} = 0.28) indicative of positive selection. Analysis of the Kenyan *S. hermonthica* plants parasitising maize, sorghum and rice revealed two markers to be potentially under selection (FDR ≤ 4%, and log (PO)) ≥ 1 corresponding to 'strong selection and above'. The two markers also showed positive α values; however, their F_{ST} values were low (mean F_{ST} = 0.08) (Figure S4, Table S2).

Discussion

Genetic diversity

In this study, *S. hermonthica* plants, collected in Kenya and Nigeria, were characterised using SNPs to determine the extent of genetic diversity existing within and between the two countries. The *S. hermonthica* populations exhibited a high level of genetic diversity, and the Nigerian and Kenyan populations were observed to be two genetically distinct groups, indicating that they have had limited exchange of genetic material. This is consistent with the results of Bozkurt *et al.* (2015).

Population structure

The Kenyan population showed little or no population structure and a low-level genetic differentiation that correlated with the distance between the sampling sites; this suggests that the sampled *S. hermonthica* plants are

interconnected by a stepwise exchange of genetic material among adjacent populations resulting in an isolation-by-distance pattern. The Nigerian plants, on the other hand, showed significant structuring. The *S. hermonthica* plants collected from Nigeria formed three groups that are similar in distribution to the distribution of three *Striga gesnerioides* (Willd.) Vatke biotypes observed in Nigeria by Lane *et al.* (1996). Our study suggests the presence of two biotypes in Kenya, one adapted to rice and the other to both maize and sorghum. This is, however, not conclusive, as the observed differentiation might be due to isolation by distance. *Striga hermonthica* plants attached to maize, pearl millet and sorghum, collected in Nigeria, were not clearly separated. Our results strongly suggest that positive selection played a role in the divergence of the Kenyan and Nigerian populations of *S. hermonthica* and also in the divergence of the Nigerian populations. Information on the Kenyan maize, sorghum and rice population is, however, not conclusive because, while outlier tests indicated diversifying selection, Mantel's test also indicated the presence of isolation by distance.

In studies on other parasitic plants like *Striga gesnerioides* (Lane *et al.*, 1996) and *Viscum album* L. (Zuber & Widmer, 2000), host adaptation was suggested to drive race formation. However, Botanga and Timko (2006) suggested that in addition to host adaptation, geographic isolation was also a critical factor in race formation. Geography appears to be the major element structuring genetic variation and differentiation in this study, and our results suggest that *S. hermonthica* populations retain a rather broad host range. The rotation of crop cultivars and species, through mixed cropping, relay cropping and crop rotation systems, that is common in *Striga* spp. infested areas in Nigeria and Kenya (Ajeigbe *et al.*, 2010), could provide an explanation for the maintenance of genetic variability for host range in *S. hermonthica* populations (Huang *et al.*, 2012). This is because changing the crop varieties and species planted in a particular location frequently will prevent tight adaptation of *Striga* spp. to any one of them (Huang *et al.*, 2011). The subpopulations observed may therefore have arisen as result of differential adaptation to environmental conditions prevalent across the locations where they are found.

The results in this study show that Kenyan and Nigerian populations of *S. hermonthica* represent distinct ecotypes, with the Nigerian *S. hermonthica* divided into three genetic groups that exist in different regions in Nigeria. The results also show some host-based differentiation in the Kenyan population. It may, therefore, be useful to characterise these populations phenotypically, to determine whether they exhibit variations in their virulence characteristics and use these characteristics to identify and pyramid resistance genes that will make cereal varieties resistant to multiple ecotypes from every region and host crop in the country and across the countries. As noted by Bozkurt *et al.* (2015), factors driving the formation and continued existence of these ecotypes and their subpopulations need to be determined and included in breeding efforts.

Testing of *S. hermonthica* control technologies in Nigeria should be performed at sites representing the areas of collections of the three genetic groups. This is particularly important for breeding crop varieties with broad-based resistance against different *S. hermonthica* populations, as the gene flow that exists within the three regions will make it difficult to develop control options that will be specifically adapted to each subpopulation.

Acknowledgements

We thank Busayo Odunaya for DNA extraction. We acknowledge the research and technical staff of Maseno University, Kenya and IITA Kano Station, Nigeria, for their help with sample collection. The assistance of Tunrayo Alabi of IITA GIS Unit in generation of the *S. hermonthica* distribution maps is acknowledged. We sincerely appreciate Sharon Mitchell and her team at the Genomic Diversity Facility, Institute of Biotechnology, Cornell University, USA, for the genotyping service. We are very grateful for the inputs of two anonymous reviewers. This study was carried out through the project, 'Achieving sustainable Striga control for poor farmers in Africa' (aka Integrated Striga Management in Africa) and funded by the Bill and Melinda Gates Foundation through Contract no OPP1006185 to the International Institute of Tropical Agriculture, Ibadan. The authors declare that we have no conflicts of interest.

References

Ajeigbe HA, Singh BB, Musa A, Adeosun JO, Adamu RS & Chikoye D (2010) Improved cowpea–cereal cropping systems: Cereal–double cowpea system for the northern Guinea savanna zone, *International Institute of Tropical Agriculture (IITA)*, p. 17.

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

Balding DJ & Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**, 3–12.

Beaumont MA & Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969–980.

Berner DK, Cardwell KF, Faturoti BO, Ikie FO & Williams OA. (1994) Relative roles of wind, crop seeds, and cattle in dispersal of *Striga* spp. *Plant Disease* **78**, 402–406.

Botanga CJ & Timko MP (2006) Phenetic relationships among different races of Striga gesnerioides (Willd.) Vatke from West Africa. *Genome* **49**, 1351–1365.

Bozkurt ML, Muth P, Parzies HK & Haussmann BIG (2015) Genetic diversity of East and West African *Striga hermonthica* populations and virulence effects on a contrasting set of sorghum cultivars. *Weed Research* **55**, 71–81.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens T M, Ramdoss Y & Buckler S (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

De Groote H, Wangare L, Kanampiu F *et al*. (2008) The potential of herbicide resistant maize technology for *Striga* control in Africa. *Agricultural Systems* **97**, 83–94.

Dorr I (1997) How *Striga* parasitizes its host: a TEM and SEM study. *Annals of Botany* **79**, 463–472.

Doyle JJ & Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15.

Ejeta G (2007) The *Striga* scourge in Africa: a growing pandemic. In: *Integrating New Technologies for Striga Control: Towards ending the witch-hunt*. (eds G Ejeta & J Gressel), 3–16. World Scientific Publishing Co. Pte. Ltd., Singapore.

Elshire RJ, Glaubitz JC, Sun Q *et al*. (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

Estep MC, Van Mourik TA, Muth P *et al*. (2011) Genetic diversity of a parasitic weed, *Striga hermonthica*, on sorghum and pearl millet in Mali. *Tropical Plant Biology* **4**, 91–98.

Foll M & Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993.

Gethi JG, Smith ME, Mitchell SE & Kresovich S (2005) Genetic diversity of *Striga hermonthica* and *Striga asiatica* population in Kenya. *Weed Research* **45**, 64–73.

Hamrick J (1982) Plant population genetics and evolution. *American Journal of Botany*, **69** 1685–1693.

Hearne SJ (2009) Control - the *Striga* conundrum. *Pest Management Science* **65**, 603–614.

Huang K, Whitlock R, Press MC & Scholes JC (2012) Variation for host range within and among populations of the parasitic plant *Striga hermonthica*. *Heredity* **108**, 96–104.

Jombart T (2008) Adegenet an R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405.

Jombart T, Devillard S & Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 1–15.

Kim SK (1994) Genetics of maize tolerance of *Striga hermonthica*. *Crop Science* **34**, 900–990.

Koyama ML (2000) Molecular markers for studying pathogen variability: implications for breeding for resistance to *Striga hermonthica*. In: *Breeding for Striga Resistance in Cereals* (eds BIG Haussmann, DE Hess, ML Koyama, L Grivet, HFW Rattunde & HH Geiger), 227–254. Margraf Verlag, Weikersheim, Germany.

Lane JA, Moore THM, Child DV & Cardwell KF (1996) Characterization of virulence and geographic distribution of *Striga gesnerioides* on cowpea in West Africa. *Plant Disease* **80**, 299–301.

Liu K & Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* **21**, 2128–2129.

Lu F, Lipka AE, Glaubitz J *et al*. (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* **9**, e1003215.

Mohamed KI, Papes M, Williams R, Benz BW & Peterson TA (2006) Global invasive potential of 10 parasitic witchweeds and related Orobanchaceae. *Ambio* **35**, 281–288.

Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290.

Parker C & Riches CR (1993) *Parasitic Weeds of the World: Biology and Control*. CAB International, Wallingford, UK.

Peakall R & Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295.

Purcell S, Neale B, Todd-Brown K *et al*. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**, 559–575.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sauerborn J (1991) The economic importance of the phytoparasites *Orobanche* and *Striga*. In: *Proceedings of the Fifth International Symposium of Parasitic Weeds*, (eds JK Ransom, LJ Musselman, AD Worsham & C Parker), 137–143. CIMMYT (International Maize and Wheat Improvement Center), Nairobi, Kenya.

Welsh AB & Mohamed KI (2011) Genetic diversity of *Striga hermonthica* populations in Ethiopia: evaluating the role of geography and host specificity in shaping population. *International Journal of Plant Sciences* **172**, 777–782.

Zuber D & Widmer A (2000) Genetic evidence for host specificity in the hemi-parasitic *Viscum album* L (Viscaceae). *Molecular Ecology* **9**, 1069–1073.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Maps of Nigeria and Western Kenya showing *Striga hermonthica* collection sites and host crops.

Figure S2 A plot of geographic distance between sampling locations in Kenya and linearised F_{ST} values

of Striga samples collected from those locations. ($R^2 = 0.33$, $P = 0.01$, $r = 0.57$).

Figure S3 A plot of geographic distance between sampling locations in Nigeria and linearised F_{ST} values of Striga samples collected from those locations. ($R^2 = 0.004$, $P = 0.2$, $r = 0.064$).

Figure S4 Results of BayeScan analysis for 1576 SNPs genotyped in *Striga hermonthica*. Marker-specific F_{ST} is plotted against the posterior odds (PO) of being under selection.

Table S1 Passport data showing the location, host and GPS co-ordinates of the *Striga hermonthica* plants collected.

Table S2 Outlier SNPs indicating positive selection in a genome scan of 1578 SNP markers using the method implemented in Bayescan 2.0.