Genetic Diversity and Population Structure in Annual Ryegrass (*Lolium multiflorum* Lam.)

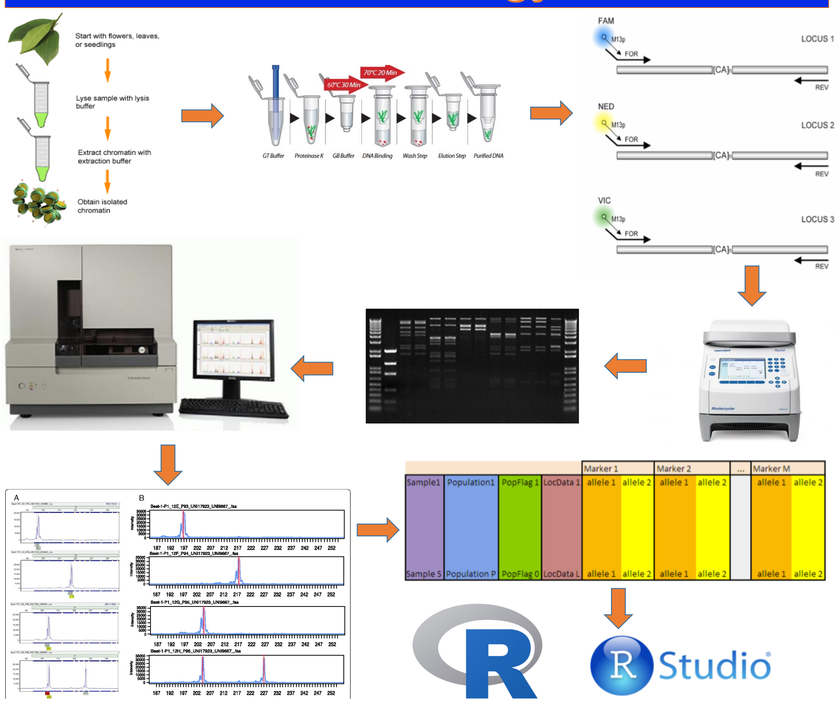
Juan Guillermo Perez[1], Malay Saha[2], Patricio Munoz[3] and Esteban Rios[1]

[1] University of Florida, Agronomy [2] Noble Research Institute [3] University of Florida, Horticultural Sciences

Introduction

Annual ryegrass (*Lolium multiflorum* Lam.) is a forage species that shows high yield, high nutritional value, and wide adaptation. The species is diploid (2n = 2x = 14), except for some improved tetraploid cultivars (2n = 4x = 28). It is important for breeding and germplasm conservation purposes to perform an assessment of the genetic diversity present in the USDA-NPGS annual ryegrass collection. Genetic improvement of this crop is important to produce valuable information about genetic structure and variability of populations using molecular marker techniques such as ETS-SSR. The objective of the current study was to assess the genetic diversity and relationship among/within 139 populations from the USDA-NPGS annual ryegrass collection, 15 experimental lines from the UF Annual Ryegrass Breeding Program and 13 commercial cultivars using 38 microsatellite markers (EST-SSR). Our findings likely will be useful for forage breeding and germplasm conservation.

Methodology



A total of 960 annual ryegrass individuals from 167 populations with five genetic statuses [cultivar (cv), cultivated, experimental, unknown and wild] were analyzed. Analysis of molecular variance (AMOVA), *Fst-value,* Shannon Weiner index and minor allele frequency (MAF) was performed to determine the genetic diversity among and within populations and genetic status. Population structure was calculated with STRUCTURE and DAPC, which uses a Bayesian approach to infer the optimal number of population within the collection using only genotypes scores as a priori information. The optimal K value was determined following Evanno et al. (2005) using pophelper web. Finally, the cluster analysis was evaluated using correspondence analysis and neighbor joinning tree (NJ) as an additional means of validation through R libraries (adegenet, poppr., etc.).

Bibliography

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

Guan, X ., et al. 2017. Genetic Diversity and Structure of Lolium Species Surveyed on Nuclear Simple Sequence Repeat and Cytoplasmic Markers. Front. Plant Sci. 8:584.

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

Viera, E.A., et al. 2004. Genetic structure of annual ryegrass (*Lolium multiflorum*) populations estimated by rapd. Sci. Agric. (Piracicaba, Braz.), v.61, n.4, p.407-413, July/August 2004

Results

This study revealed a high level of polymorphism, with 1013 alleles in total. The average MAF was 0.2 overall. The average number of observed alleles was 25.66 for the 38 loci. AMOVA showed that 82.7% of the total variation is found within populations, while the fixation index (*Fst* = <0.05) corroborates that the populations are not well differentiated (table 1). EST-SSR markers revealed that 7.11% of the total variation was due to differences among status, with the remaining 10.2% due to differences within populations and 82.74% due to differences within populations (Table 1). The Shannon and Simpson indexes are measures to compare diversity and dominance among communities respectively (Table 2).

Table 1. AMOVA for 960 annual ryegrass individuals from 167 populations and five genetic statuses, analyzed with 38 EST-SSR markers.

Source of variation	Df	Sum Sq	Mean Sq	%
Among status	4	80.9	20.2	7.1
Among individuals within populations	157	331.7	2.1	10.2
Within populations	798	976.5	1.2	82.7
Total	959	1389.1	1.5	100.0

Table 2. Genetic Diversity in annual ryegrasss. H, Shannon Weiner diversity index; S Simpson's index. Nei, Nei's gene diversity; Ia, Index of Association for each population. D, Standardized Index of Association.

Status	samples	H	S	Nei	Ia	D
Cultivated	42	3.74	0.97	0.82	0.540	0.065
cv	341	5.83	0.99	0.81	0.147	0.027
Experimental	68	4.22	0.98	0.78	0.022	0.016
Unknown	317	5.76	0.99	0.79	0.208	0.016
Wild	192	5.26	0.99	0.85	0.295	0.014
Total	960	6.87	0.99	0.82	0.155	0.027

The Bayesian model based approach using STRUCTURE showed the hierarchical organization and genetic relatedness of 960 individuals. Determining the basic structural features of the data across to the Evano's test (ΔK), we assigned individuals to a subpopulation based on status. In order to explore deeper population structure, structure analyses were also carried out for each status for k= 3 and k=5. The number of subpopulations that best defined the entire population was K=5, where each individual can be defined relative to the percentage of each subpopulation in its ancestry (Figure 1).

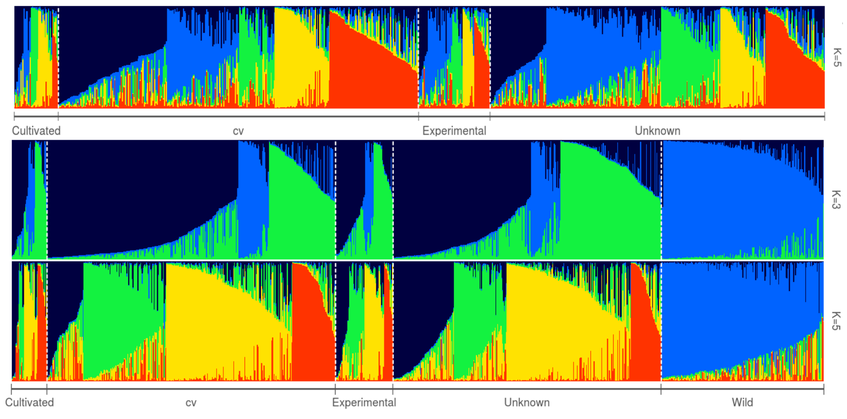


Figure 1. STRUCTURE analysis assigned groupings to the 960 individuals with 5 genetic statuses. Vertical bars indicate the estimated membership coefficients (Q) of each individual for each population cluster. Bar plot was sorted by Q-values in a single line. The top graph represents a K=5 cluster without considering wild-type individuals. Middle and bottom graphs include wild-type individuals for K=3 and K=5 respectively.

The neighbor joining tree, based on distance between individuals, showed minimum evolutionary steps in annual ryegrass and the population could be composed of five primary clusters based on genetic status (Figure 2). Pair-wise estimates of the Wright statistics (*Fst*-value) exhibits the degree of variation between subpopulations. These estimates showed little genetic divergence between unknown to cv (0.0089) and experimental (0.0095), but a moderate genetic divergence occurred between wild to cv and cultivated (0.05). This is also seen with DPAC genetic distances, which describes diversity between pre-defined groups (Figure 3).

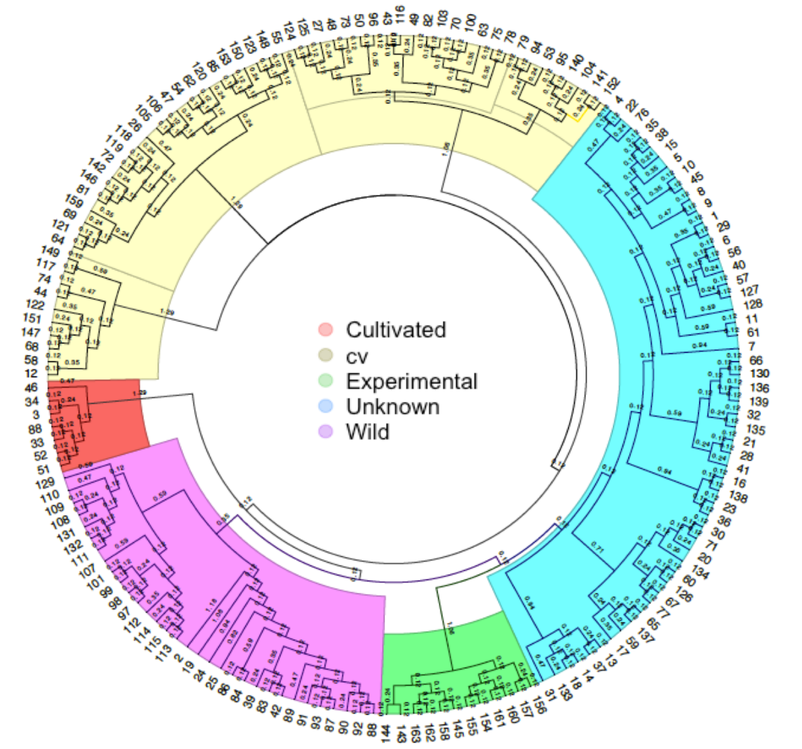


Figure 2. Neighbor joining tree of diversity based on Nei (1978) distance and bootstrap method to evaluate the parsimony and reliability of tree topology with 100 replicates. Each branch is color-coded according to membership into the K = 5 groups identified by STRUCTURE.

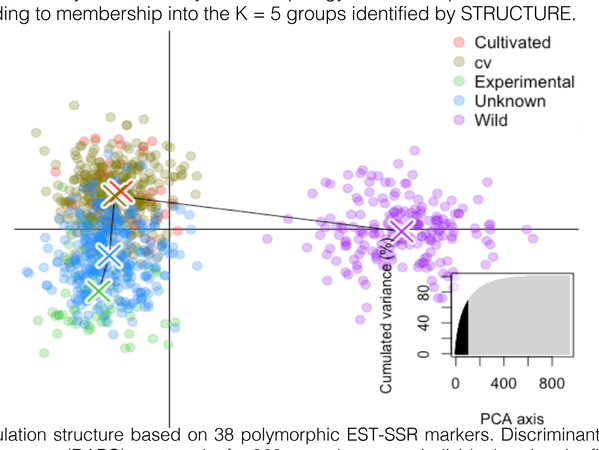


Figure 3. Population structure based on 38 polymorphic EST-SSR markers. Discriminant Analysis of Principal Components (DAPC) scatter plot for 960 annual ryegrass individuals using the first two PCs. Colors indicates the number of PCs retained to show the relationship between genetic status (K=5).

Discussion

The level of genetic diversity estimated using ETS-SSR markers is higher than previous reports for other DNA markers in annual ryegrass (Vieira et al., 2004). The total genetic diversity (Nei=0.82) in annual ryegrass (Table 2), is close (0.84) to reports for the species by Guan et al. (2017). Wild and unknown status exhibited a greater genetic diversity than the other genetic statuses. These high genetic diversity could be explained by the allogamous nature of the species and the relatively short period of cultivation and artificial selection (Breese and Hayward, 1972).

Conclusion

ETS-SSR markers have enough discriminant power to assess diversity in annual ryegrass. The molecular data indicated that the cultivated, cv, experimental and unknown are diverse, and all these groups are largely distant to wild germplasm. These results will be valuable for the use of annual ryegrass germplasm improvement and increase the effectiveness of ryegrass breeding programs.

Molecular data generated in this study will be used in future studies to find associations between molecular markers and phenotypic data.

IFTBC International Forage & Turf Breeding Conference
A Global Vision for Innovation

NAAIC
NAAIC, Trifolium, & Grass Breeders
June 4-6, 2018 · Logan, Utah