Genetic Studies on Kernel Zinc in Tropical Maize Germplasm

Project Description

Micronutrient deficiencies have become prevalent leading to considerable morbidity and mortality worldwide. Approximately 30% of the world's population is reported to suffer from inadequate zinc intakes (Gannon et al., 2017). Inadequate Zn intake leads to reduced growth and development, impaired immune system, altered reproductive biology, gastro intestinal problems and adverse pregnancy outcomes (Institute of Medicine (US) and Panel on Micronutrients, 2001).

Efforts to address the problem of micronutrient deficiency have focused on micronutrient supplementation, food fortification and diet diversification. However, these methods are deemed inaccessible and expensive, especially in developing countries where people do not have a strong purchasing power. To complement these approaches, the use of plant breeding to naturally increase micronutrients density in commonly consumed staple food crops, through genetic biofortification is considered viable from the view point of its cost effectiveness and sustainability. To develop crop varieties high in zinc micronutrient, it necessary to identify germplasm high in Zinc concentration, assess their potential for grain yield and other agronomic traits, understand the mechanism underlying kernel zinc inheritance and finally develop knowledge-based strategies that can be used for selecting lines that are high in grain Zn.

The goal of this project is to improve maize for kernel Zn through classical breeding and by using modern breeding techniques. In this project, we propose to Assess the mode of inheritance for kernel Zn and the feasibility of using genomic prediction to select maize lines high in kernel Zn. The methodologies and analyses during this project will generate data that will be helpful in breeding programs. This is because, in a breeding program, a continuous improvement of breeding strategies as well as the potential to use molecular markers to accelerate the genetic gain is very important.

Expected data type

The proposed research project will generate experimental data. We will generate both phenotypic (from field and laboratory) and genotypic (from sequencing) data. Phenotypic data from field experiments and from different field locations will comprise measurements on agronomic traits such as, plant height, flowering time and grain yield. From the laboratory, phenotypic data will comprise measurement of grain Zn concentration. Genotypic data will be generated from Genotyping by sequencing (GBS) which is one of the several techniques used to genotype populations using high throughput sequencing (HTS). For the field experiments, we will also record information including soil type, soil pH, precipitation and other general observations. The data generated from the field and laboratory experiments will be stored in a spreadsheet (Microsoft excel). Files containing the data will be named according to the format site_ ddmmyyyy_season.csv. Genotypic data (GBS) will be stored in HapMap and the reference genome will be included.

Data Format

Data from this proposed project will be comprised of Ms. Excel (.xls) and HapMap (. hmp) formats. These formats were chosen because they are easy to work with, accessible and readable. For example, HapMap which is a text based file format for storing sequence data keeps all the

information for a series of SNPs as well as the germplasm lines is in one file. The first row contains the header labels, and each additional row contains all the information associated with a single SNP. For HapMap formatted files, the first 11 columns will describe attributes of the SNP, while the following columns will describe the SNP value for a single germplasm line.

Data storage and preservation
Data generated from this project will be well organized using google drive/docs as a version control. These files will be stored in well labeled folders. With the assistance from the IT department at CIMMYT, monthly and seasonal backups of the data will be retained at the CIMMYT's website (https://data.cimmyt.org/dataverse/root).. Therefore, all raw and summarized data will be stored and backed up in the CIMMYTs server and in the cloud-based backup systems. In the long term, we are committed to generate data that is of high quality and that can be easily accessible by other researchers and policy makers anytime. Raw data will therefore be made available indefinitely as it will be deposited to Dryad (http://datadryad.org/) part of the Data One network and which is committed to long-term data preservation and access. For data preservation, we will include the .csv and .xls files along with the associated metadata files. We will also submit an abstract with the datasets that describe their original context and any potentially relevant project information.

Data sharing, publishing and dissemination
After publication of manuscripts based on the data we collect, we will share our data and metadata with the publicly via data updates at the CIMMYT's and HarvestPlus websites (https://data.cimmyt.org/dataverse/root and (https://harvestplus.com) as .xls and .csv files. Preliminary results from this project will be presented at relevant national and international meetings and conferences as posters or talks. Access to associated software tools and scripts generated under this project will be made available for research and educational purposes. A GitHub repository and access will be provided as appropriate. Promising and desirable inbred lines identified under the project will be disseminated and shared in accordance with CIMMYT's policies for example using material transfer agreement.

Roles and responsibilities
All the data collected as part of this project will be owned by CIMMYT. The project's PI will act as a steward for the data while they are being collected, processed, and analyzed. The PI will also take responsibility for the for the management and sharing of the research data. He will also be responsible for preparing data for long term preservation and for updating contact information for investigators. Day to day quality assessment of the data will be the responsibility of every research person involved in this project