**COLLEEN F. YANARELLA**
**LAWRENCE-DILL PLANT INFORMATICS AND COMPUTATION LAB (DILL-PICL)**
**JULY 12, 2022**

# SPEECH-BASED GENOTYPE TO PHENOTYPE ANALYSIS FOR ASSOCIATION GENETICS IN MAIZE: A PROOF OF CONCEPT

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
Department of Agronomy

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
Department of Agronomy

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
Department of Agronomy

# OUTLINE

I. **How we perform association studies**

II. How we collected spoken phenotypes
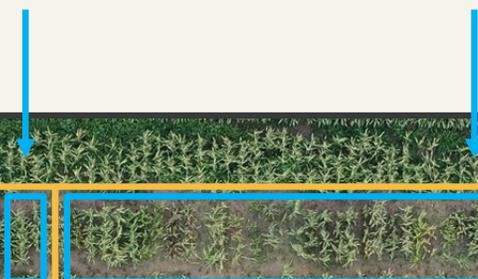
III. Overview of our pipeline

IV. Next steps

I'M AN
AGRONOMIST

IOWA STATE UNIVERSITY
**Department of Agronomy**

Many generations of intercrossing, introgressions, and mutations

Recreated and Adapted from: Pasam, R.K., Sharma, R. (2014). Association Mapping: A New Paradigm for Dissection of Complex Traits in Crops. In: P.B.K., Bandopadhyay, R., Suravajhala, P. (eds) Agricultural Bioinformatics. Springer, New Delhi.

686 WiDiv Lines
8 B73 Environmental Checks
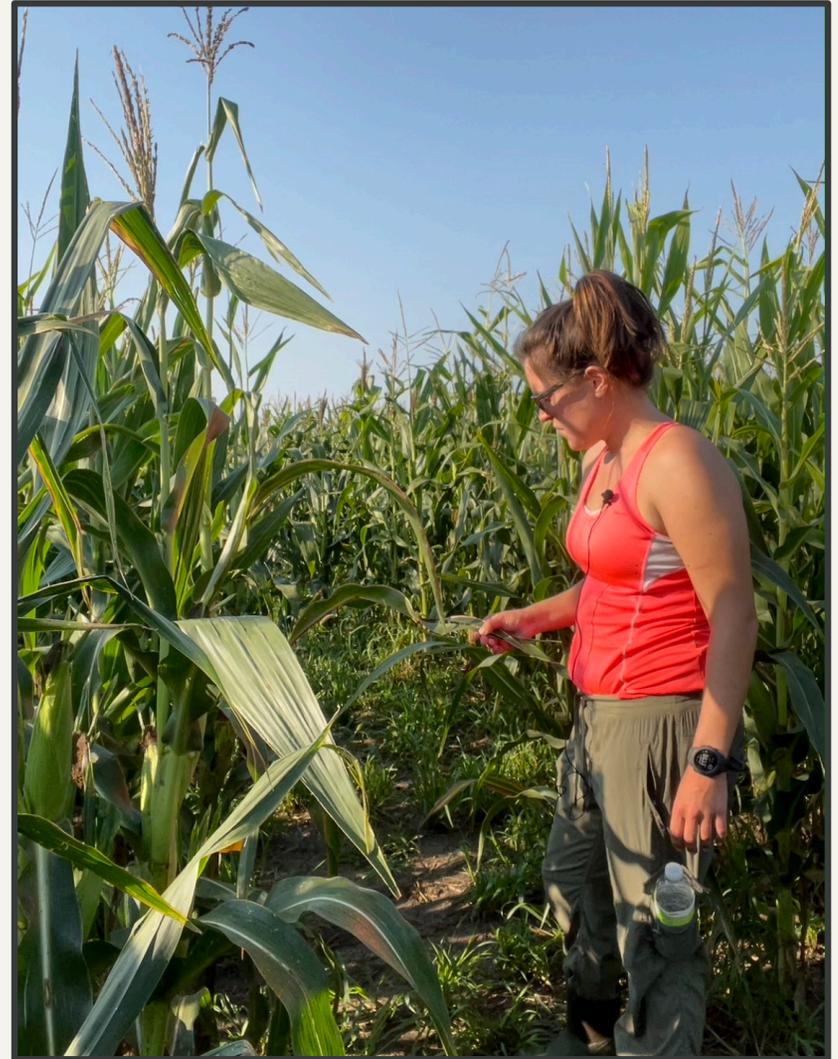25 Positive Controls

Row

Range

Rep 1

Rep 2

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
Department of Agronomy

# OUTLINE

I.    How we perform association studies

II.   **How we collected spoken phenotypes**

III.  Overview of our pipeline

IV.  Next steps

I'M AN **AGRONOMIST**

**IOWA STATE UNIVERSITY**
**Department of Agronomy**
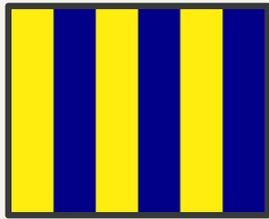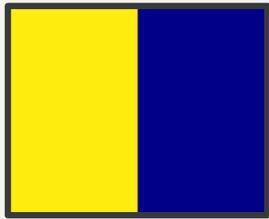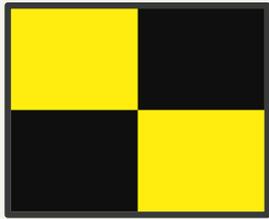
**I'M AN AGRONOMIST**

**IOWA STATE UNIVERSITY**
**Department of Agronomy**

Delta  Golf  Kilo
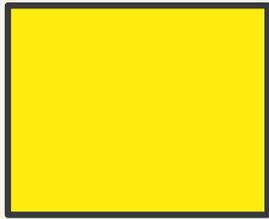
Lima  Mike  Quebec

Victor  Yankee  Zulu

https://www.flyingcolours.org/

**IOWA STATE UNIVERSITY**
Office of Research Ethics

IRB Study 21-179-00 Exempt Status
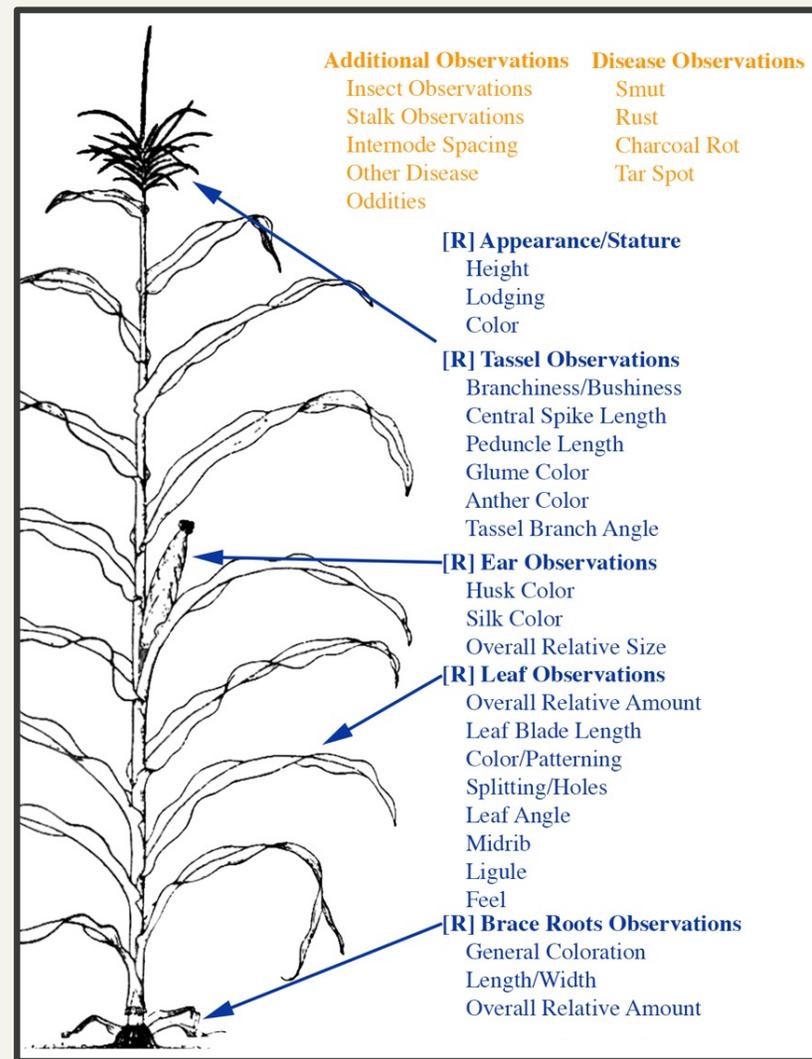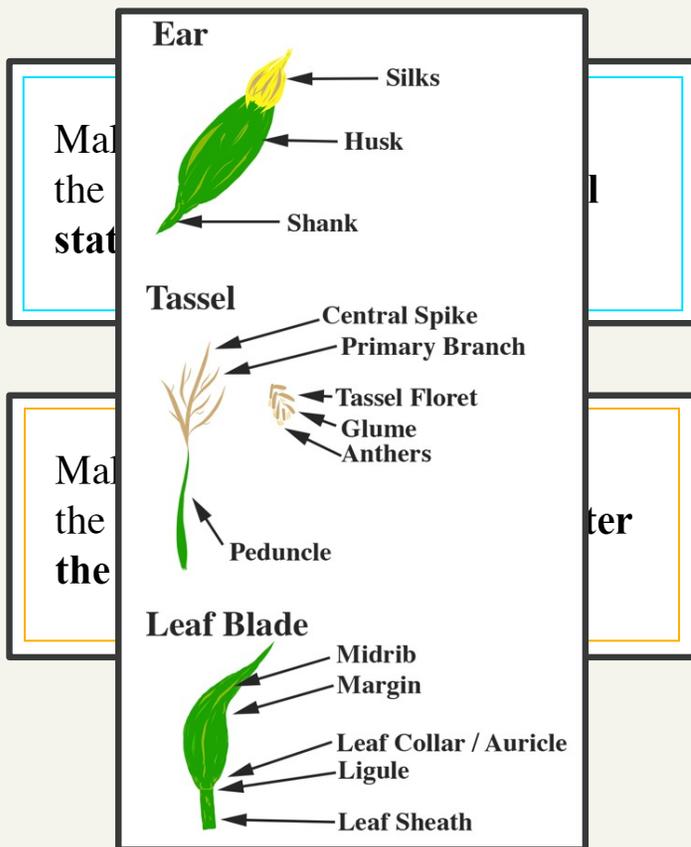
F indable  A ccessible  I nteroperable  R eusable

https://upload.wikimedia.org/wikipedia/commons/a/aa/FAIR_data_principles.jpg

Whiskey

I'M AN **AGRONOMIST**

**IOWA STATE UNIVERSITY**
**Department of Agronomy**

# Observation Guide Card
## REMEMBER TO PRESS RECORD
1. Start with **NATO code name** and **row tag number**
2. Make observations about the required phenotypes [if present and visible] (designated with [R] below)
3. Make observations about the additional phenotypes and disease presence [if present and visible]



**Ear**
- Silks
- Husk
- Shank

**Tassel**
- Central Spike
- Primary Branch
- Tassel Floret
- Glume
- Anthers
- Peduncle

**Leaf Blade**
- Midrib
- Margin
- Leaf Collar / Auricle
- Ligule
- Leaf Sheath



**Additional Observations**
- Insect Observations
- Stalk Observations
- Internode Spacing
- Other Disease
- Oddities

**Disease Observations**
- Smut
- Rust
- Charcoal Rot
- Tar Spot

**[R] Appearance/Stature**
- Height
- Lodging
- Color

**[R] Tassel Observations**
- Branchiness/Bushiness
- Central Spike Length
- Peduncle Length
- Glume Color
- Anther Color
- Tassel Branch Angle

**[R] Ear Observations**
- Husk Color
- Silk Color
- Overall Relative Size

**[R] Leaf Observations**
- Overall Relative Amount
- Leaf Blade Length
- Color/Patterning
- Splitting/Holes
- Leaf Angle
- Midrib
- Ligule
- Feel

**[R] Brace Roots Observations**
- General Coloration
- Length/Width
- Overall Relative Amount

Adapted from Figure 12 (p.12): Kiesselbach, T. A., "The Structure and Reproduction of Corn" (1949). Historical Research Bulletins of the Nebraska Agricultural Experiment Station. 284.

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
**Department of Agronomy**

I'M AN
AGRONOMIST

IOWA STATE UNIVERSITY
Department of Agronomy

# OUTLINE

I. How we perform association studies

II. How we collected spoken phenotypes

III. **Overview of our pipeline**

IV. Next steps

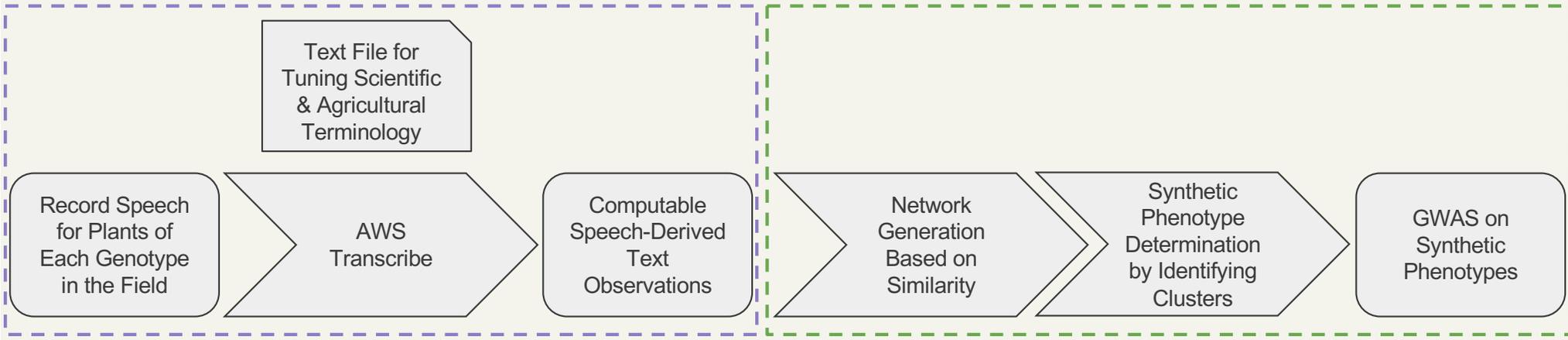I'M AN
**AGRONOMIST**

**IOWA STATE UNIVERSITY**
**Department of Agronomy**

# Process

## Audio Processing Component

Text File for Tuning Scientific & Agricultural Terminology

Record Speech for Plants of Each Genotype in the Field → AWS Transcribe → Computable Speech-Derived Text Observations

## Association Study Component

Network Generation Based on Similarity → Synthetic Phenotype Determination by Identifying Clusters → GWAS on Synthetic Phenotypes

# Example

## Audio Processing Component
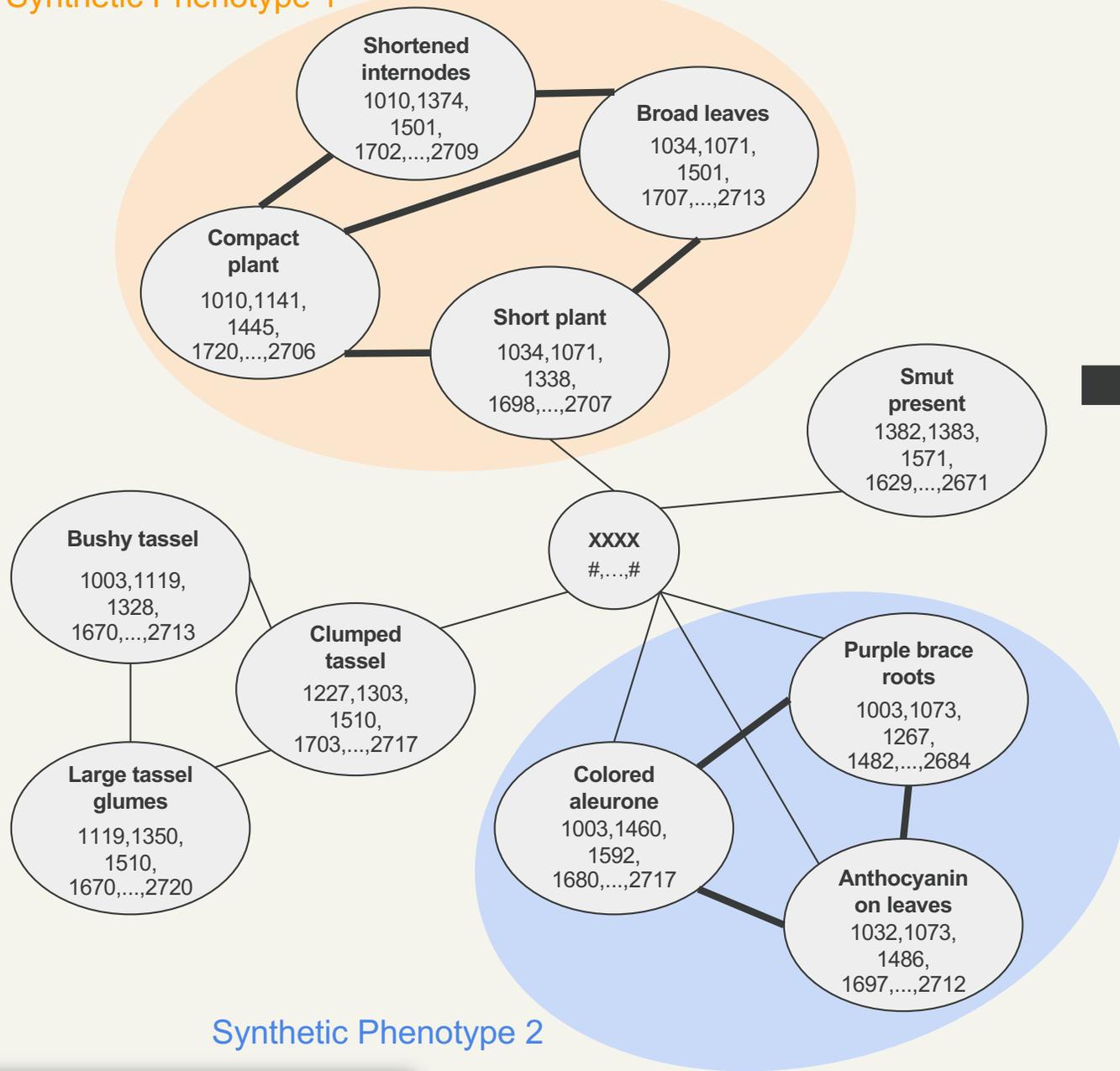
Whiskey 1574…
Extremely tall…
Lodged…
Leaves feel veiny...

→

```
Whiskey 1574…
Extremely tall…
Lodged…
Leaves feel veiny…
```

Whiskey 1575…
Medium height…
Upright plants…
Large tassels...

→

```
Whiskey 1575…
Medium height…
Upright plants…
Large tassels…
```

## Association Study Component

Synthetic Phenotype 1

Synthetic Phenotype 2

**Generate Phenotype Network**

**Determine Synthetic Phenotypes**

Trait Model

$-\log_{10}p$

Chromosome

**Associate Synthetic Phenotypes to SNPs**

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
**Department of Agronomy**

# OUTLINE

I.    How we perform association studies

II.   How we collected spoken phenotypes

III.  Overview of our pipeline

IV.  **Next steps**

IOWA STATE UNIVERSITY
**Department of Agronomy**

Mona Mazaheri[1,2], Marlies Heckwolf[1,2], Brieanne Vaillancourt[3,4], Joseph L. Gage[1], Brett Burdo[1], Sven Heckwolf[5], Kerrie Barry[6], Anna Lipzen[6], Camila Bastos Ribeiro[7], Thomas J. Y. Kono[8,10], Heidi F. Kaeppler[1,2], Edgar P. Spalding[5], Candice N. Hirsch[8], C. Robin Buell[3,4,9], Natalia de Leon[1,2] and Shawn M. Kaeppler[1,2*]

**942** WiDiv Lines
**899,784** SNPs



Plant Height
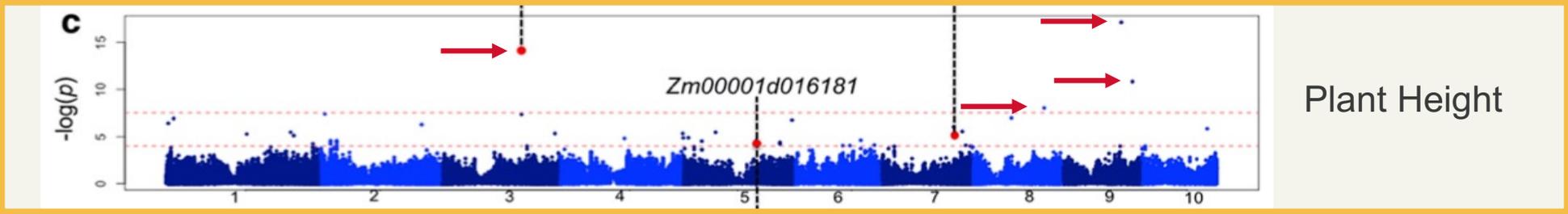
Brian Dilkes

Rajdeep Khangura

Amanpreet Kaur

PURDUE UNIVERSITY

Construction of the third-generation *Zea mays* haplotype map

Robert Bukowski[1], Xiaosen Guo[2,3], Yanli Lu[4], Cheng Zou[5], Bing He[2], Zhengqin Rong[2], Bo Wang[2], Dawen Xu[2], Bicheng Yang[2], Chuanxiao Xie[5], Longjiang Fan[6], Shibin Gao[4], Xun Xu[2], Gengyun Zhang[2], Yingrui Li[2], Yinping Jiao[7], John F. Doebley[8], Jeffrey Ross-Ibarra[9], Anne Lorant[9], Vince Buffalo[9], M. Cinta Romay[10], Edward S. Buckler[10,11], Doreen Ware[7], Jinsheng Lai[13], Qi Sun[1,*] and Yunbi Xu[5,12,*]

**942** WiDiv Lines
**HapMap3 positions**
          **~ 55 million** imputed SNPs
**MAF >= 0.05**
          **~22 million** imputed SNPs

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
**Department of Agronomy**

# ACKNOWLEDGEMENTS

I'M AN AGRONOMIST

IOWA STATE UNIVERSITY
**Department of Agronomy**