Systemic Analysis of Biological Data from an Isogenic Maize Line

Merritt Bryer Burch, Boris Shmagin, Vivek Shrestha, Donald Auger

Department of Biology and Microbiology, South Dakota State University, Brookings SD 57006, USA

Introduction & Background

Doubled-Haploids (DH)

- Monoploid plants can chemically or naturally diploidize to create completely homozygous progeny so that both parent and progeny are identical genotypically and phenotypically
- This can change quickly. Sprague et al. (1960) demonstrated that polymorphisms in quantitative traits among DH maize emerged faster than the known rate of discrete mutations

Factor Analysis (FA)

- Factor analysis is a statistical technique used to group observed variables into unique components based on their correlation strength, variance, and amount of phenotypic variability they explain
- These components are combinations of measured traits whose variability explains a majority of the relationships in our data
- Systematic patterns of variation can arise and be seen in multiple generations
- These analyses provide an alternative look at complex data sets and are useful when attempting to explain heritability patterns typically found with ANOVAs or mixed effect models
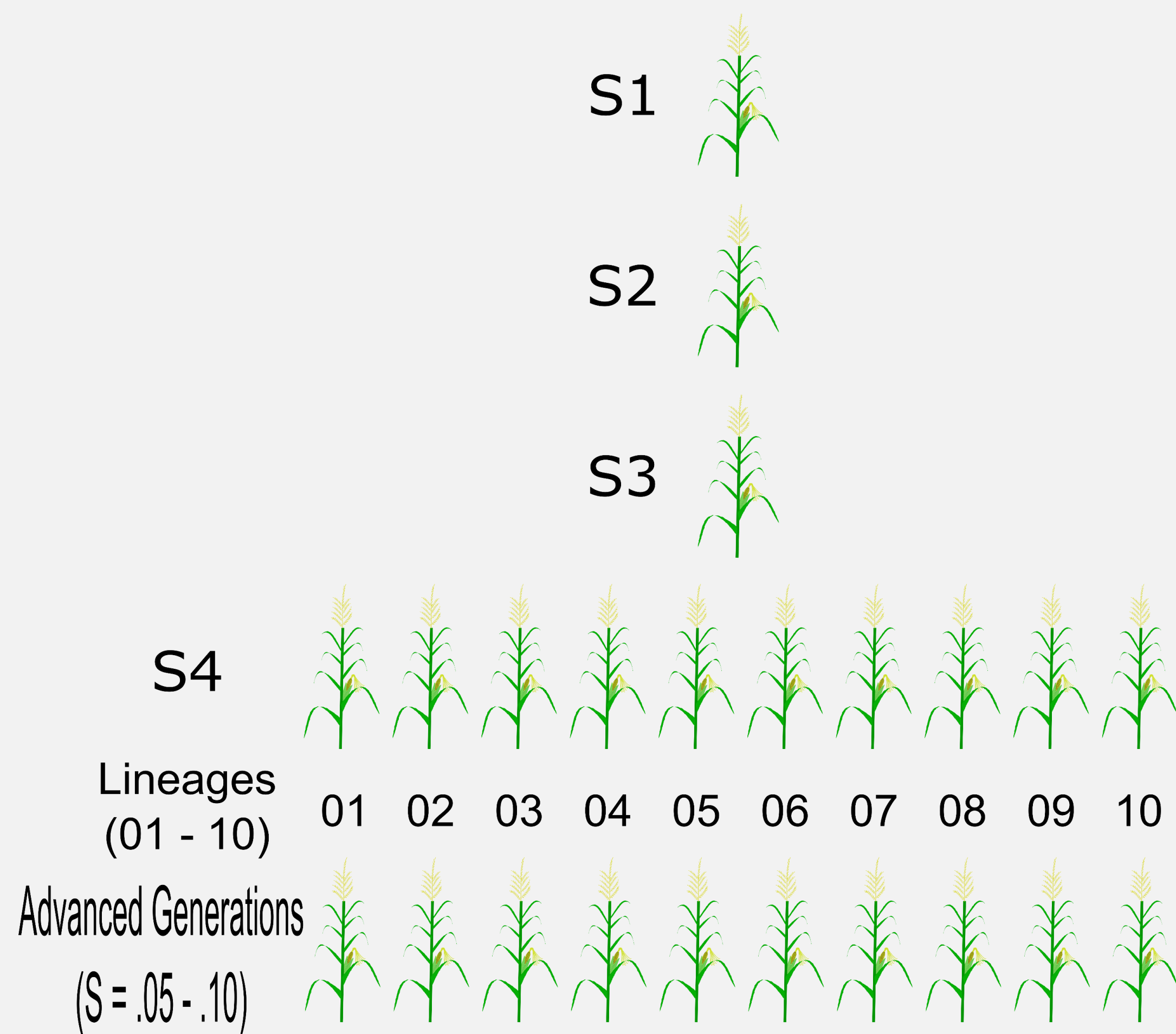
Materials and Methods

Fig. 1: A single-seed descent line created from a B73 doubled-haploid plant was sequentially self-pollinated. Generation 3 (S3) seeds were randomly selected to create ten separate DH lineages. These lineages were advanced and assessed for variation in quantitative traits.

Fig 2: A typical maize plant showing the location of ten quantitative traits measured. A total of thirteen traits were measured and analyzed.
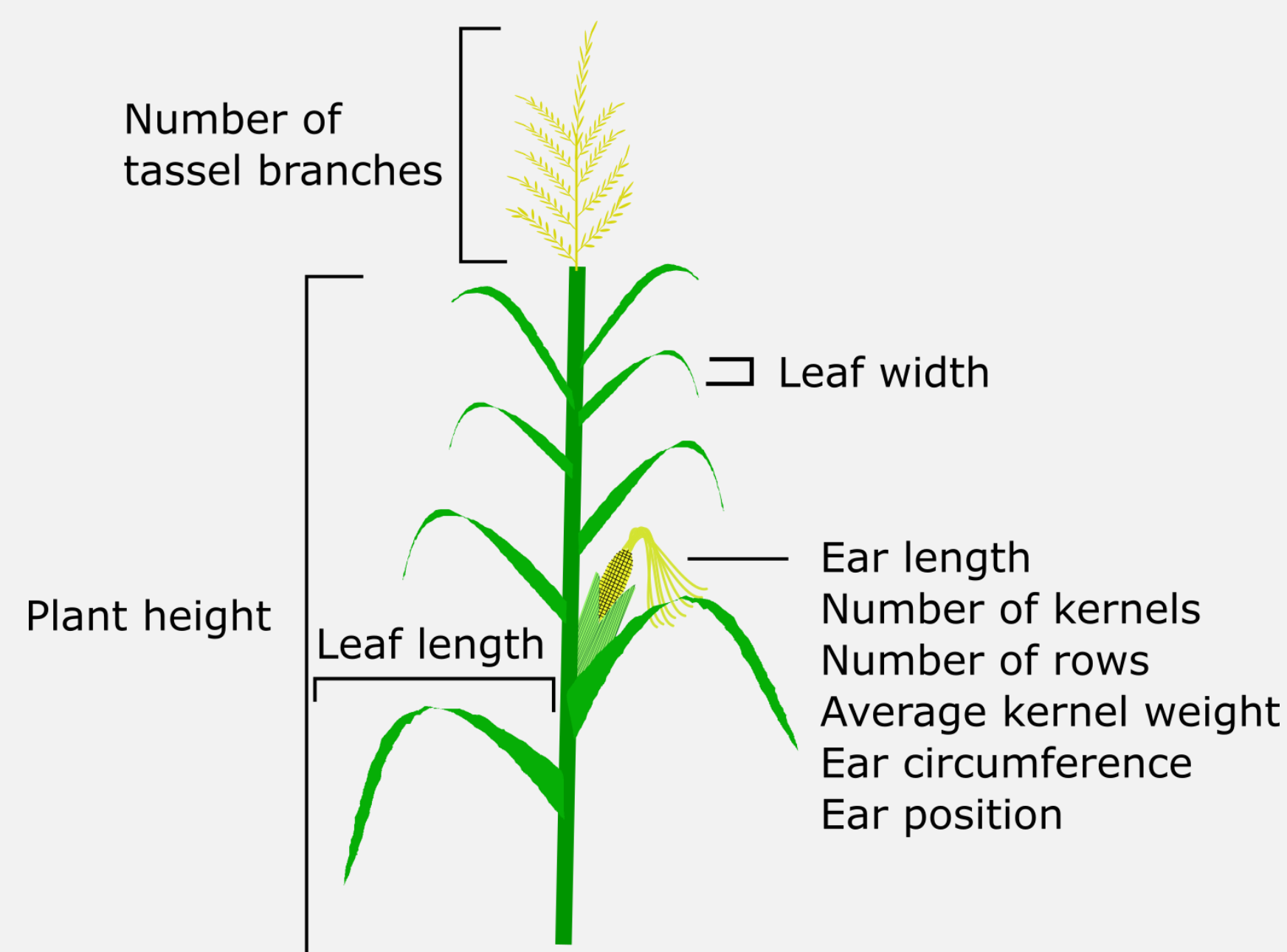
- In the summer of 2014, 2015, and 2017 all ten lineages with two or more generations, were grown at SDSU in a randomized complete block design.
- Factor analysis on was done on five factors with an orthogonal rotation
- Data visualization and analysis was performed in the psych, ggplot2, and factoextra packages in R Studio (V: 3.4.2)
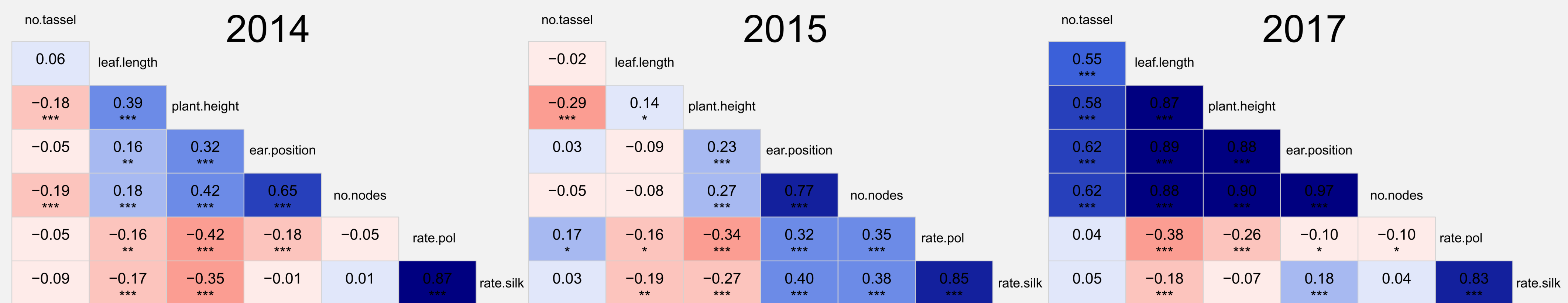
Results

Fig. 3: Correlation Matrices for DH data collected in 2014, 2015, and 2017. Numerous moderate and strong relationships are conserved throughout the years, especially between the rate of pollen shed and rate of silk emergence along with the relationship between ear position and the number of nodes.
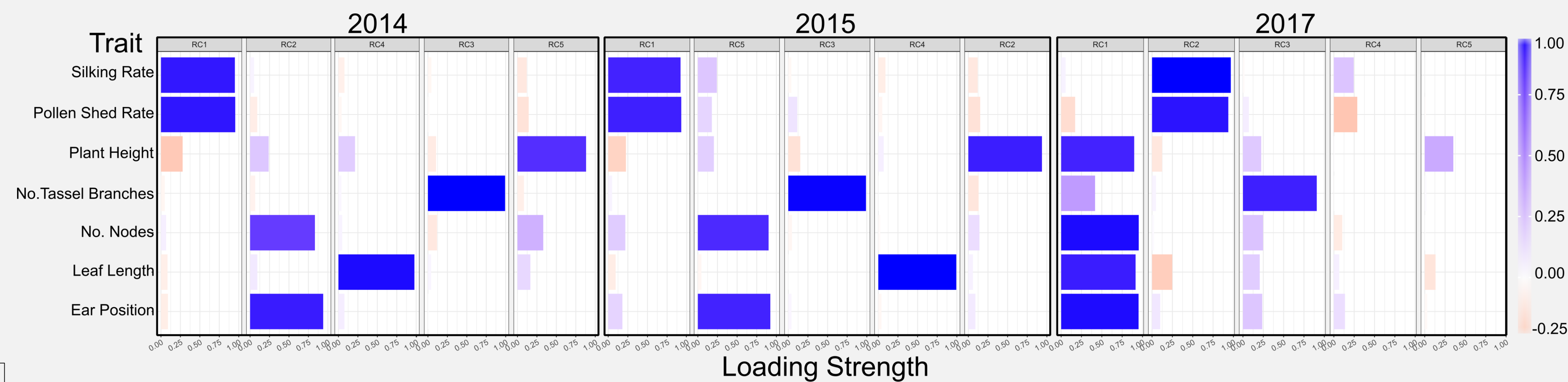

Fig. 4: Factor analysis results showing the factor loading scores of the five-selected rotated components in three-years of DH data. Data are sorted based on the highest to lowest proportional variance explained by each rotated component. The five factors explained 85%, 87%, and 99% of the variance in 2014, 2015, and 2017 respectively. Consistent patterns in the rate of silk emergence and pollen shed, the number of nodes and ear position, and plant height reoccur in all three years. In 2017, many more traits clustered into the first rotated component, potentially due to the increased number of individual DH lineages measured compared to 2014 or 2015.
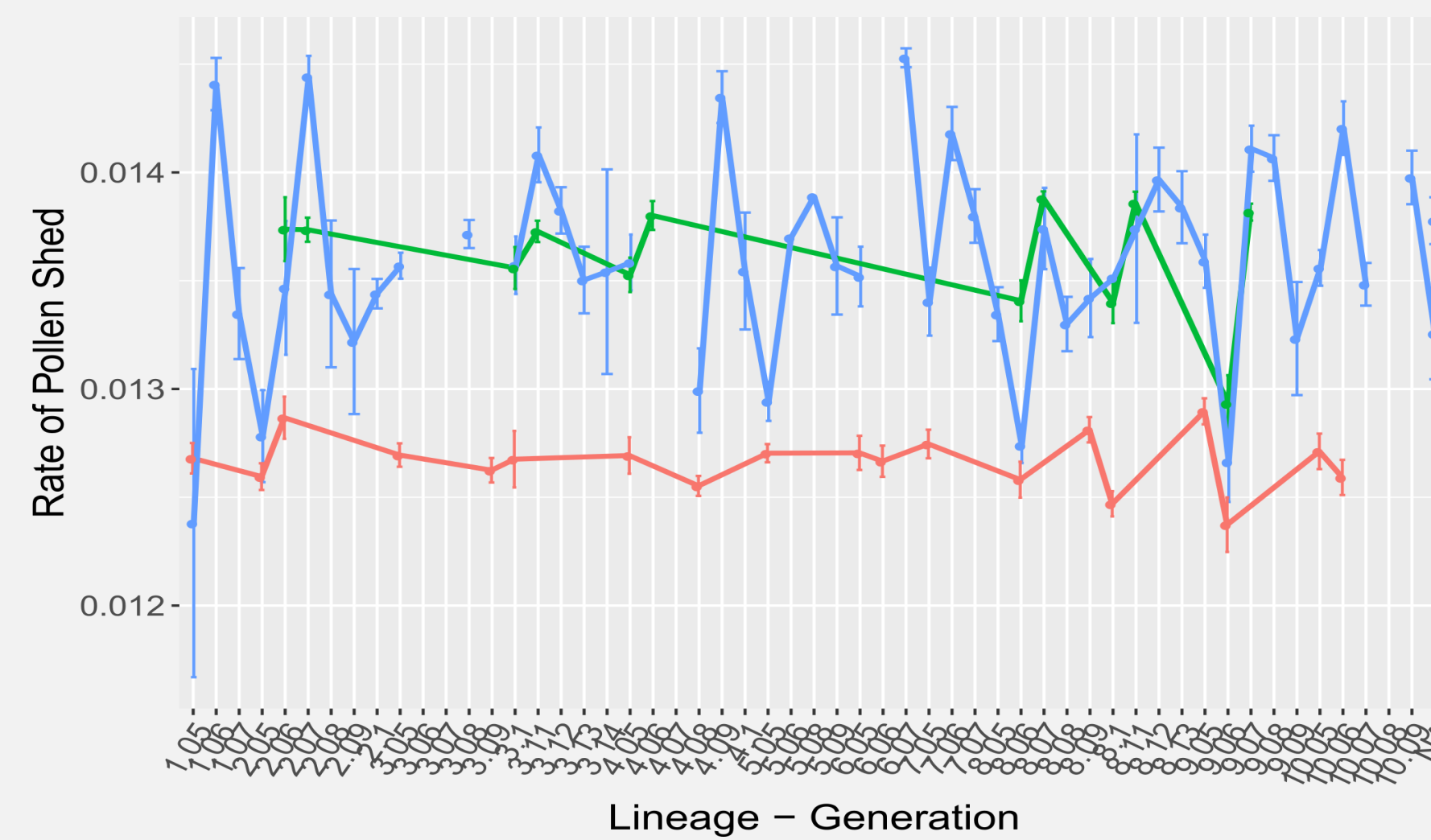

Fig. 5: Line plot showing the mean and SEM for the rate of pollen shed over three years of data collection. Regardless of the year, some similar, but not as intense patterns can be observed between individual lineages as compared to more visible patterns observed with factor analysis.
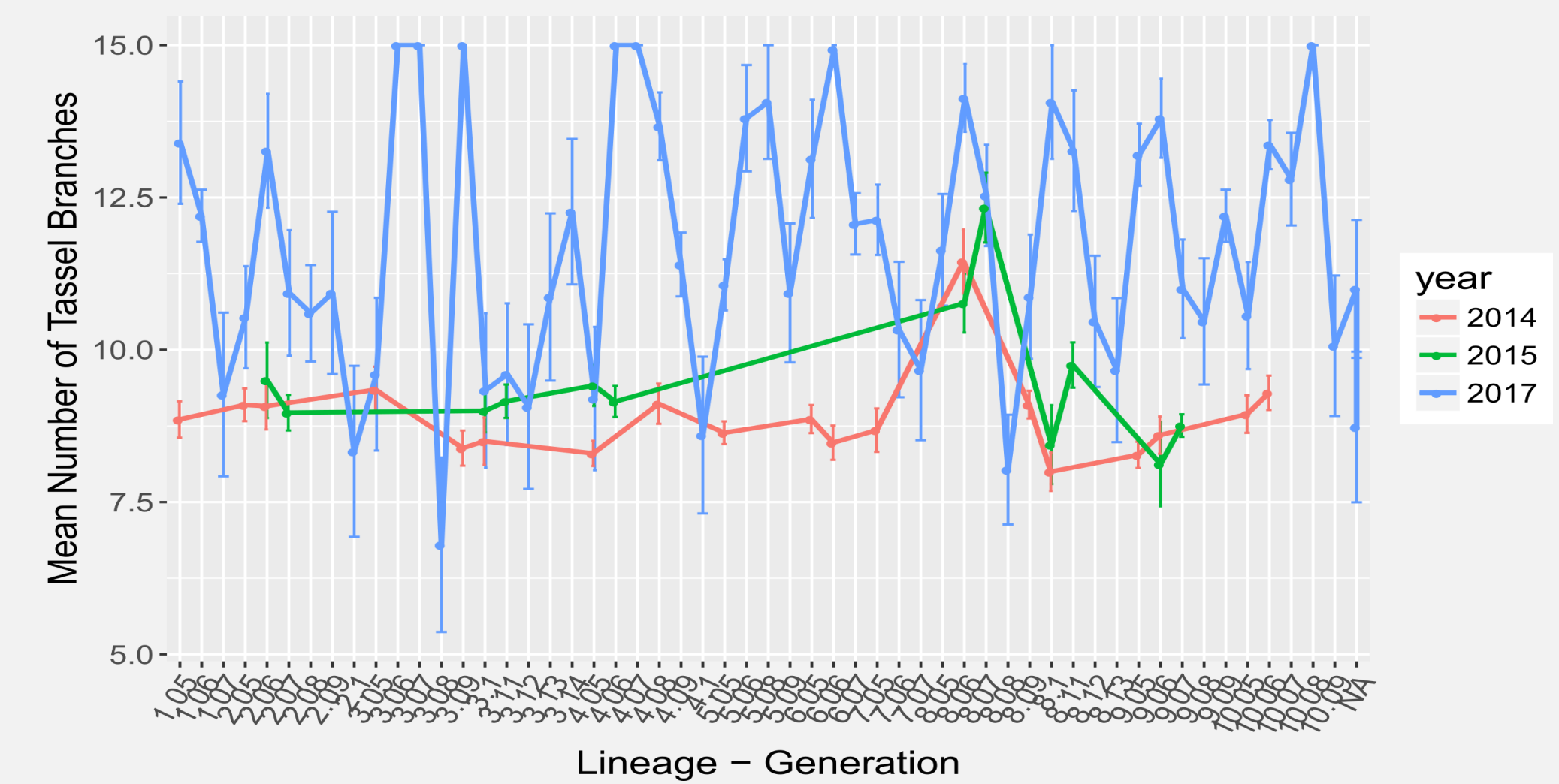

Fig. 6: Line plot showing the mean and SEM for the number of tassel branches over three years of data collection. Data from 2015 shows such erratic patterns due to the limited number of DH lineages measured.

Conclusions

- Although insightful for certain purposes, correlations between single variables or comparisons of population means cannot fully explain the biological complexity beyond the yearly environmental interactions (Fig. 3, 5, 6)
- Factor analysis was used in this study to discover previously unseen patterns within the data that would have otherwise been missed due to environmental noise (Fig. 4)
- FA showed that certain traits over multiple years cluster with each other (ex., rate of pollen shed and rate of silk emergence), this would potentially suggest that similar regulatory or genetic factors control groups of traits
- Data from 2017 does not contain many of the same five components as the 2014 and 2015 data, this may be do to the increased DH lineages measured.

Future Work

- Continue using FA with the DH data, removing lineages that were not measured in each year, this may lead to all the FA loading scores to show consistent results.
- Sequencing these lineages by single-trait RNAseq or GBS may shed more light on specific genes or enriched pathways behaving differently between lineages
- Determine if parental effects from the initial creation of this DH line is possibility causing this heritable variability

Acknowledgements