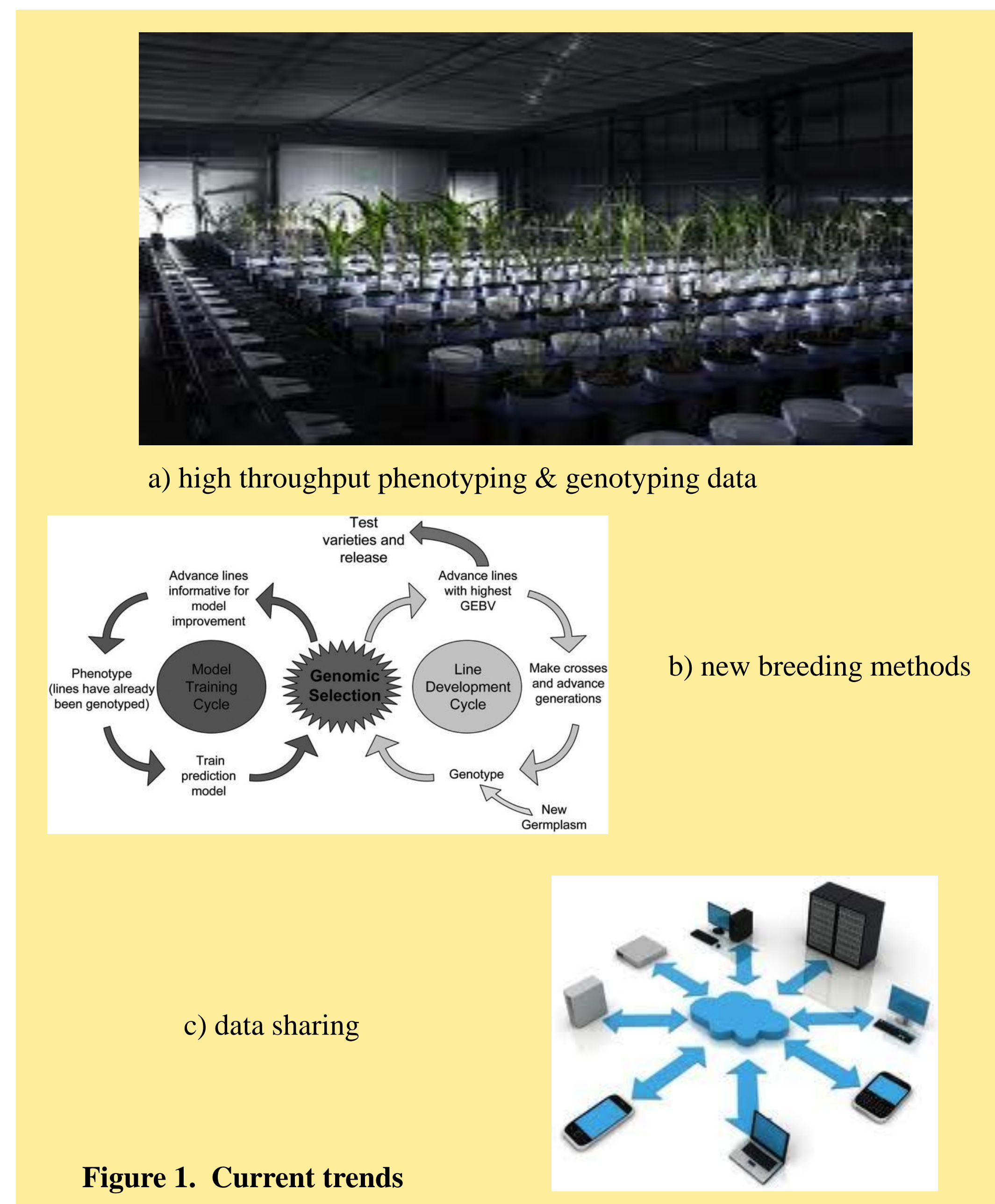


Abstract

Potato and sweetpotato breeding involves not only handling large amounts of plant materials, locations and respective environmental and management conditions but also the associated phenotypic traits like yield, resistance levels and nutritional values. Molecular markers have been introduced to facilitate the selection process – these current portfolio of breeding information already poses challenges for an effective information platform. However, already new high-throughput phenotyping and genotyping technologies are on the horizon providing magnitudes more information. Also, other advances in breeding methods and information and knowledge management have become apparent adding additional requirements. In short, a current breeding information platform will need to address the following principal challenges: a) raw data from latest high-throughput phenotyping and genotyping platforms are typically already too big for standard laptops or standard PCs; b) new breeding methods like genome wide selection have the promise of integrating information from a diverse set of sources given a common repository to predict breeding values; c) following best practices to share those data with the wider scientific community as soon as possible. Here, we outline a new architecture for a breeding information platform to address these challenges based on currently already available open source software tools. Principal components of the proposed platform are standardized set of traits (ontologies), a cloud based data repository, a near-real time information discovery tool, an online/offline data analysis and visualization platform, and supporting tools for data auditing and knowledge management.

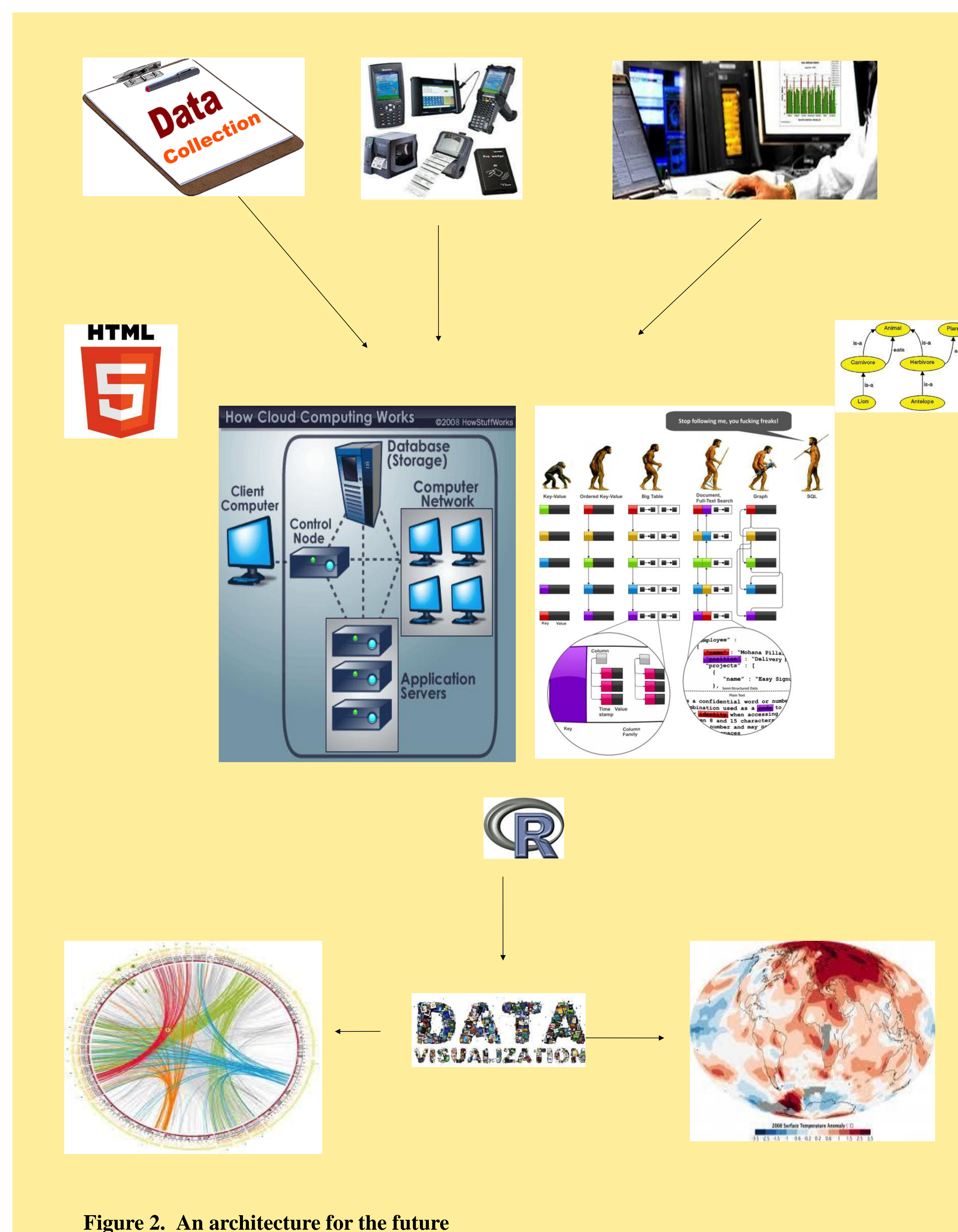


Motivation

Matching the increasing amount of genomic data with phenomic data is inhibited by various factors including the lack of a public infrastructure and an unwillingness to disclose raw phenotypic data or list the identifiers of tested genotypes (Zamir, 2013). Other hurdles include that current management tools for germplasm and breeding data still are designed as relational database schemes which do not perform well for really big datasets or highly visited websites. Review of current cutting edge technologies from other industries however shows that there are tools available to address these shortcomings for plant breeding data.

Elements of the solution

Schemaless databases address the problem of performance and allow quick access to and discovery of data. New standards like HTML5 allow for seamless offline/online integration as well as for mobile data gathering. Semantic standards like ontologies facilitate the integration of data from different sources. Cloud-based computing permits the storage of huge datasets as well as easy access for analysing these datasets.



References

- D Zamir (2013). Where have all the crop phenotypes gone? PLoS Biology 11(6): e1001595. doi:10.1371/journal.pbio.1001595

