1 Introduction

1.1 Motivation

As world population and food demand increases, and the looming threats of climate change intensify, conserving the genetic diversity of both our crop plants as well as related wild species becomes increasingly imperative. Access to genetically diverse plant species allows plant breeders to produce crop varieties that are more resilient and better adapted to their habitat, resulting in greater crop yield (Plucknett & Nigel, 1987, p. 3). As such, the genetic information preserved within each plant species remains a tremendously valuable natural resource.

With growing necessity and urgency placed on conserving the genetic information of crop species comes the very practical question of preservation. While crop diversity has historically been maintained informally through cultivation within farm communities, seed banks have increasingly become the answer to limitations in space and resources. They provide a centralized repository for collecting and storing germplasm and allows for the protection of present crop varieties as well as antique or rare plant species that are no longer cultivated (Plucknett & Nigel, 1987, p. 4).

As of 2006, about six million seed samples have been stored in about 1,300 gene banks throughout the world (Rajasekharan, 2015, p.1). Local seed diversity has long been protected and preserved through community seed banks, which may take the form of families or community groups gathering and storing seeds which represent the local environment and are best adapted to withstand native pests, soil composition, diseases and climatic variations. Recently, these banks have also begun to take the form of large global repositories, such as the Svalbard Global Seed Vault, which holds samples from around the world in underground chambers within a thick layer of permafrost and rock (Van 't Wout et al., 2014, p.1-7). Global seed banks offer the opportunity to create secure backups of local seed repositories from the around the world, and can sometimes provide benefits that may not be available to smaller community seed banks, such as technical expertise, sophisticated storage facilities, and assistance in monitoring viability and regenerating materials when needed (Fowler, 2008).

As seed storage and preservation grows in both necessity and popularity, with it grows the need to understand and evaluate the impact and efficacy of these banks. One of the primary incentives for storing seeds is preserving genetic diversity, and consequently characterizing the present state of seed repositories may offer the ability to observe trends and identify phenotypes or growth conditions that are over and underrepresented in current collections.

Characterizing the biodiversity within these banks is no small task, and there are many potential approaches to doing so. Seed banks are most often used as functional repositories for plant breeders, and so maintaining basic information about specific samples is an important step in the archival process. Species name as well as the date and location of collection are recorded in what is called the seed's "passport data" (Plucknett & Nigel, 1987, p. 72). However, this information alone is not sufficient for the purposes of our analysis. As large scale genomic testing has yet to be performed (**CITATION?), characterization based on species traits provides a more accessible path.

The pages that follow offer an attempt to not only characterize the diversity represented in one large scale seed bank, but outlines a visual solution to the complex problem of meaningfully interpreting this. While data mining and processing can bridge archival information and species characteristics, the scope and complexity of this data and these relationships is obscured without the ability to appropriately interact with them. By focusing on the Svalbard Global Seed Bank, currently considered to have the most diverse collection on earth, and using the Max Planck Institute for Biogeochemistry's TRY database – the world's largest database on plant's functional properties – this thesis outlines a process for representing the genetic diversity within one seed bank and investigating how rich a representation of the planet's plant biodiversity these archives provide.

1.2 The Data

1.2.1 Svalbard Global Seed Vault

1.2.2. TRY Global Database

Bibliography

Fowler, C. (2008). The Svalbard Seed Vault and Crop Security. *BioScience, 58*(3), 190-191. Retrieved March 1, 2016, from http://bioscience.oxfordjournals.org/content/58/3/190.full.pdf html

Peres, S. (2016). Saving the gene pool for the future: Seed banks as archives. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 55*, 96-104. Retrieved March 2, 2016, from http://www.sciencedirect.com/science/article/pii/S1369848615001284

Plucknett, Donald L., and Nigel J.H. Smith. Gene Banks and the World's Food. New Jersey: Princeton UP., 1987. Print.

Rajasekharan, P. E. (2015). Gene Banking for Ex Situ Conservation of Plant Genetic Resources. Plant Biology and Biotechnology, 445-459. Retrieved February 29, 2016, from http://link.springer.com/chapter/10.1007/978-81-322-2283-5_23

Van 't Wout, T., & Dalla Valle, F. (2014). *Community Seed Banks* (Rep.). Retrieved March 1, 2016, from Food And Agriculture Organization Of The United Nations website: http://www.fao.org/fileadmin/user_upload/fao_ilo/pdf/Other_docs/FAO/Community_Seed_Banks.pdf

Weiher, E., Werf, A. V., Thompson, K., Roderick, M., Garnier, E., & Eriksson, O. (1999). Challenging Theophrastus: A common core list of plant traits for functional ecology. *Journal of Vegetation Science, 10*(5), 609-620. Retrieved March 2, 2016, from http://onlinelibrary.wiley.com/doi/10.2307/3237076/abstract