RES-FOR HIGHLIGHT #9

November 2020

Opening the Social Dimensions of Genomic Selection Research

Overview

Genomic selection (GS) is a diagnostic approach that links genetic and phenotypic data through complex biological information and sophisticated mathematical models to improve plant breeding (Grattapaglia 2017). In forestry, to supplement traditional tree breeding, leading researchers are carefully and meticulously developing GS applications that can select trees at early stages in otherwise long breeding cycles. The application of GS to tree breeding can enhance traditional traits, like wood quality. Another application is to select drought resilient and disease/pest resistant trees in response to challenges posed by climate change.

Due to its complexity, the social dimensions of GS can easily be *'black boxed'*, meaning that the assumptions and values shaping its development are not open to scrutiny. Collaboration with social science researchers presents a unique opportunity to identify the social dimensions of technology that otherwise may not be acknowledged. Transparency about assumptions can enhance social learning (the process of bringing diverse perspectives and interests together to address problems), and improve communication and engagement, all of which are essential features of responsible research and innovation (Genus and Stirling 2018).

Research Objective

The objective of this MSc research was to examine how a genomic selection model is made. Drawing on the field of science and technology studies, this research traces the processes and assumptions informing the development and application of GS for forestry (Bates et al. 2016; Latour 1987; 1999). This MSc research contributes to the broader GE[3]LS[*] objective of the RES-FOR[**] project to enhance reflexivity in the development of genomic selection. Reflexivity – a dimension of responsible research and innovation – requires that the assumptions, values, processes, and explanatory models underlying technological development are made explicit throughout the lifecycle of innovation.

Methods

Data was gathered through interviews with RES-FOR researchers (N=9) who collected and analyzed data for RES-FOR's GS models. Respondents were at various stages of their academic careers, ranging from PhD candidates to senior scientists, and of various disciplinary backgrounds, including genetics, bioinformatics, plant physiology, chemical ecology, plant molecular biology, and analytical chemistry. Data was also collected through site visit observations at laboratories, greenhouses, field sites, team meetings, and team workshops. Qualitative approaches, such as the one undertaken in this study, focused on gaining depth of understanding with a relatively small sample size, as compared with the large sample sizes required by statistical, quantitative methodological approaches.

Interview questions covered research contributions to, and challenges within, the RES-FOR project, as well as assessment of uncertainty and understanding of genomic selection in forestry, climate change, and associated complexities as it relates to genomics. Interview transcripts, field notes, and photographs were uploaded into the qualitative software package NVivo[TM] for analysis. Findings were corroborated by a broader set of interviews with genomic and forestry researchers (N=38) conducted by the GE[3]LS team at scientific conferences.

Select Findings

1. GS translates biological processes into mathematical data through a complex network

One common theme that emerged is that GS is an extension of conventional tree breeding practices. However, findings also revealed that GS differs from conventional breeding practices in terms of its reliance on a complex, geographically disparate network of sites for data collection and analysis in which complex biological processes are translated into molecular and digital formats (Figure 1). At each translation site, choices are made that are informed by the values and assumptions of different research disciplines.

Mapping the research networks involved with the RES-FOR project shows that genetic data, in both physical and digital forms, crossed international borders. Digital DNA data moves, is accessed, stored, and is governed differently than physical DNA and the plants themselves. Recognizing this important difference can create space to address broader questions of ownership, control, access, and benefit sharing.
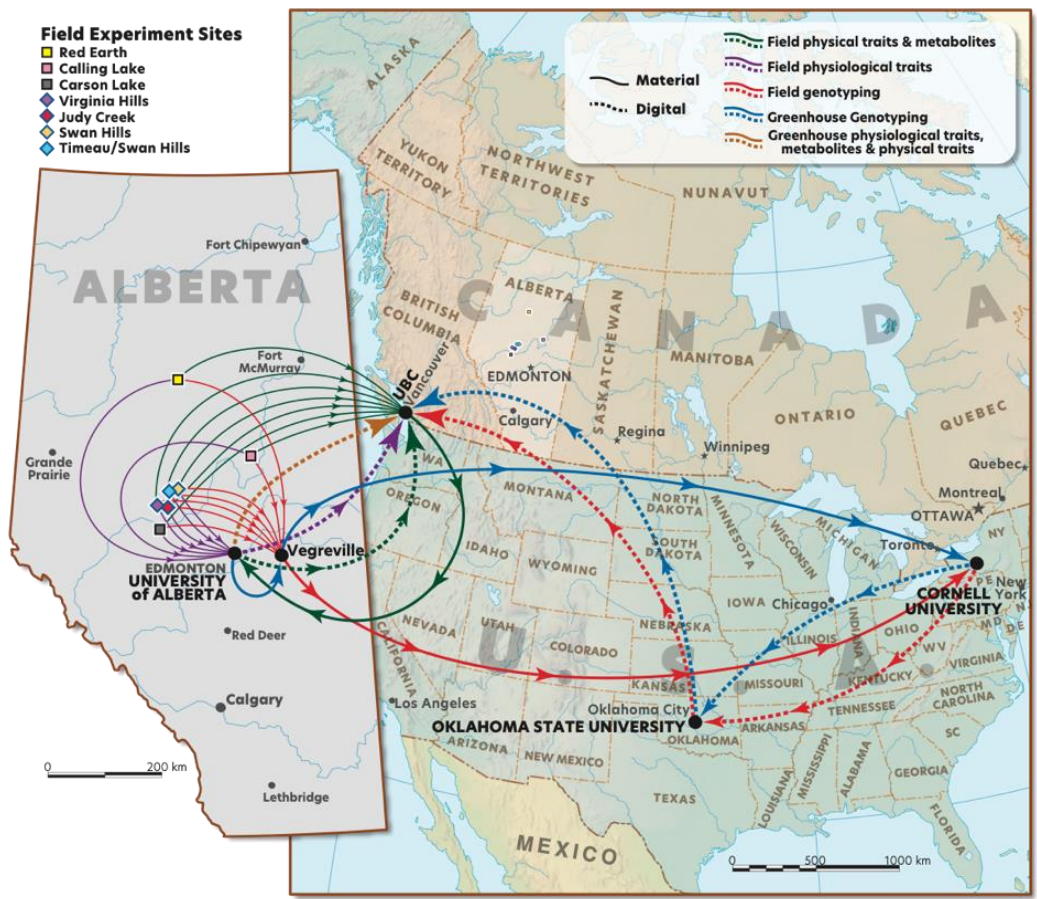
2. Assumption: More data leads to better predictions

Another commonly-held assumption was that large data sets can be equated with better knowledge about a tree's growth and natural capacity for drought and pest resistance. Better knowledge is associated with better predictions. To make these predictions, GS technologies rely on large amounts of data for both genetic markers and phenotypic traits.

Implications

The vast amounts of data required for GS predictions reveals GS' reliance on algorithms. At various stages of GS technology development, RES-FOR researchers rely on algorithms to mine, parse, sort, configure, and visualize data. Algorithms are themselves black boxes to those without the expertise of the developers. Although many RES-FOR researchers dedicate considerable time to discuss and debate the use of algorithms, these conversations may not be visible to others. Moreover, algorithms are embedded with both statistical and social assumptions that may affect decision-making in unintended ways. While statistical assumptions are well characterized, more discussion about the social aspects and contexts of these algorithms can open up broader conversations about the complexities of GS prediction and selection.

Figure 1: Data flows of RES-FOR GS development

Conclusion

Mapping the development of GS, understanding the processes, and identifying assumptions can, without undermining the importance of using genomic data to select better (eg: more resilient) trees, highlight other issues that warrant discussion. Such assessments are particularly relevant when science and technology are being asked to provide answers for the anticipated problems of tomorrow. End users and decision-makers often only see the products of GS (eg: genomic estimated breeding values) as they lack the expertise to conduct the analysis themselves. It is important to communicate GS development processes and assumptions, in addition to its benefits. Communicating the complexities of GS development opens up new opportunities to reflect on the technologies' advantages, potential limitations, and broader social implications.

References: Bates et al. (2016) *Big Data & Soc* 1-12; Genus and Stirling (2018) *Res Pol* **47**:61-69; Grattapaglia (2017) *Springer* 199-249; Latour 1987 *Cambridge* 1-288, 1999 *Harv U Press* 1-324.

*GE[3]LS: Genomics, Ethical, Environmental, Economic, Legal and Social Aspects
**RES-FOR: Resilient Forests: Climate, Pests and Policy – Genomic Applications