



UMEÅ UNIVERSITET

Embracing the Data Flood

Integrating Diverse Data to Improve Phenotype Association Discovery in Forest Trees

Bastian Schiffthaler

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för
avläggande av filosofie doktorsexamen framläggs till offentligt
försvar i Carl Kempe Salen, byggnad Linnaeus väg 6,

Fredagen den 12. Juni, kl. 13:00.

Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Prof., Marek Mutwil,
Nanyang Technological University, Singapore.

Dept. of Plant Physiology

Organization

Umeå University
Dept. of Plant Physiology

Document type

Doctoral thesis

Date of publication

20th May 2020

Author

Bastian Schiffthaler

Title

Embracing the Data Flood
Integrating Diverse Data to Improve Phenotype Association Discovery in Forest Trees

Abstract

Complex traits represent valuable research targets as many highly desirable properties of plants and animals (such as growth rate and height) fall into this group. However, associating biological markers with these traits is incredibly challenging, in part due to their small effect sizes. For the two species at the core of our research, European aspen (*Populus tremula*) and Norway spruce (*Picea abies*), association studies are even more challenging, primarily due to the fragmented state of their genome assemblies. These assemblies represent the gene space well, but poorly represented inter-genic regions hinder variant discovery and large-scale association studies.

In this thesis, I present my work to improve association discovery of complex traits in forest trees. Firstly, to overcome the issues with assembly fragmentation, I have created an updated version of the *P. tremula* genome, which is highly contiguous and anchored in full chromosomes. To calculate the dense linkage map required to order and orient the aspen assembly, I developed “BatchMap”, a parallel implementation of linkage mapping software. BatchMap has been successfully applied to several dense linkage maps, including aspen and Norway spruce, and was essential to the progress in improving the aspen genome assembly. Further, I developed seidr, which represents a starting point in multi-layer, network-based systems biology, an analysis technique with promising prospects for complex trait association analysis. As a case study, I applied some of the methods developed to the analysis of leaf shape in natural populations of European aspen, a complex, omnigenic trait.

The multi-layer model of systems biology and related analysis techniques offer promise in the analysis of complex traits, and this thesis represents a starting point toward an intricate, holistic model of systems biology that may help to unravel the overwhelmingly complicated nature of complex traits.

Keywords

genomics, transcriptomics, forest trees, gene network, genetic map, European aspen, Norway spruce, genome assembly, genome wide association, machine learning

Language

English

ISBN

print: 978-91-7855-273-3
PDF: 978-91-7855-274-0

Number of pages

86 + 5 papers