

# MrIML: Multi-response interpretable machine learning to map genomic landscapes

Nichola Fountain-Jones<sup>1</sup>, Christopher Kozakiewicz<sup>2</sup>, Brenna Forester<sup>3</sup>, Erin Landguth<sup>4</sup>, Scott Carver<sup>1</sup>, Michael Charleston<sup>1</sup>, Roderick Gagne<sup>5</sup>, Brandon Greenwell<sup>6</sup>, Simona Kraberger<sup>3</sup>, Daryl Trumbo<sup>3</sup>, Michael Mayer<sup>7</sup>, Nicholas Clark<sup>8</sup>, and Gustavo Machado<sup>9</sup>

<sup>1</sup>University of Tasmania

<sup>2</sup>Washington State University

<sup>3</sup>Colorado State University

<sup>4</sup>University of Montana

<sup>5</sup>Affiliation not available

<sup>6</sup>University of Cincinnati

<sup>7</sup>Actuarial department

<sup>8</sup>University of Queensland

<sup>9</sup>North Carolina State University

December 21, 2020

## Abstract

We introduce a new R package ‘MrIML’ (Multi-response Interpretable Machine Learning). MrIML provides a powerful and interpretable framework that enables users to harness recent advances in machine learning to map multi-locus genomic relationships, to identify loci of interest for future landscape genetics studies and to gain new insights into adaptation across environmental gradients. Relationships between genetic change and environment are often non-linear, interactive and autocorrelated. Our package helps capture this complexity and offers functions that construct, fit and conduct inference on a wide range of highly flexible models that are routinely used for single-locus landscape genetics studies but are rarely extended to estimate response functions for multiple loci. To demonstrate the package’s broad functionality, we test its ability to recover landscape relationships from simulated genomic data. We also apply the package to two empirical case studies. In the first we estimate variation in the population-level genetic composition of North American balsam poplar (*Populus balsamifera*, Salicaceae) and in the second we recover individual-level landscapes while estimating host drivers of feline immunodeficiency virus genetic spread in bobcats (*Lynx rufus*). The ability to model thousands of loci collectively and compare models from linear regression to extreme gradient boosting, within the same analytical framework, has the potential to be transformative. The MrIML framework is also extendable and not limited to mapping genetic change, for example, it can be used to quantify the environmental driver of microbiomes and coinfection dynamics.

## Hosted file

Manuscript draft Dec11\_FINAL.pdf available at <https://authorea.com/users/384112/articles/499905-mriml-multi-response-interpretable-machine-learning-to-map-genomic-landscapes>