Feature Selection from Microbial Profiles via a Genetic Algorithm

Nisha Puthiyedth1, Steven Mamel2, Steven Siciliano2, Michael Horsch1, Kevin Stanley3, Anthony Kusakil, Matthew Links1,3*

1 Department of Computer Science, 2 Department of Soil Science, 3 Department of Animal and Poultry Science *matthew.links@usask.ca

University of Saskatchewan

Discussion and Conclusions

• The consortium of OTUs identified by the genetic algorithm from the epiphytic seed data shows a clear separation between wheat and Brassica samples (Figure 2). This consortium of OTU would be sufficient to classify samples as having derived from the respective crops.

• 64% of the OTUs from the genetic algorithm consortium match the results reported by Links et al.1.

• There were 4 OTUs (5622, 845, 3024, 4081) present in the genetic algorithm results that were also shown by Links et al. to be biologically linked with epiphytic microbial abundance.

• Analysis of the smooth brome data demonstrated that the genetic algorithm is capable of identifying similar consortium to the work published by Mamel et al.2. There was significant (50%) overlap between the result of genetic algorithm and the previously published results. However OTU membership in the consortium is clearly dependent on the fitness function used.

• Areas of future investigation:

  • Assess the stability of genetic algorithm results.

  • Characterize OTUs that are uniquely identified by the genetic algorithm.

  • Assess whether there are phylogenetic patterns in the results identified by the genetic algorithm versus.

  • Further develop the use of genetic algorithms for jointly studying microbiome dynamics as well host genotypes.

References


Acknowledgements

This work is supported by The Plant Phenotyping and Imaging Research Centre (p2irc.usask.ca), The Global Institute for Food Security (gif.ca) serves as lead for the Canada First Research Excellence Fund award (cfref-apogee.gc.ca/home-accelr-eng.aspx) that supports this work.

Materials and Methods

Data from 2 previously published microbiome studies are used.

1. Epiphytic seed – 5477 features and 11 samples
2. Smooth brome – 6747 features and 109 samples
4. Basic steps/phases in GA:

   - Initial Population
   - Inversion
   - Crossover
   - Mutation
   - Evaluation Fitness
   - Termination

Table 1: Fitness function and parameters used within the GA. The final crossover rate for each test is varied from 0.6 – 1% and the final mutation rate is varied from 70-80% to achieve the maximum fitness.