

FOR OFFICE USE ONLY

DATE RECEIVED: **03/13/2025**

DATE APPROVED: **03/13/2025**

FILE NUMBER: **#25-045**

APPLICATION FOR STATISTICAL CONSULTING

LAST NAME: **Li**

FIRST NAME: **Ying**

DEPARTMENT (full name): **HLA**

CAMPUS MAILING ADDRESS: **HORT**

PHONE: **7654940104**

EMAIL ADDRESS: **li2627@purdue.edu**

YOUR PRIMARY POSITION AT PURDUE: **Faculty**

Other:

(if a student) MAJOR PROFESSOR LAST NAME: FIRST NAME:

PHONE NUMBER:

MAJOR PROFESSOR CAMPUS ADDRESS (BLDG & DEPT): /

MAJOR PROFESSOR EMAIL:

HOW DID YOU FIND US: **I've used the consulting service before**

LIST STATISTICS COURSES TAKEN AND STATISTICAL COMPUTING EXPERIENCE: **I have been using bioinformatics and machine learning approaches in my research with genomics data. Experienced with R**

STAGE OF RESEARCH: **Design (no data collected yet)**

IF DESIGN STAGE IS COMPLETE, WAS A STATISTICIAN CONSULTED FOR DESIGN?

PREVIOUS CONSULTANT – INSTITUTION/DEPARTMENT:

ESTIMATED NUMBER OF CONSULTING HOURS NEEDED THIS SEMESTER: **<5 hours**

EXPECTED COMPLETION DATE OF PROJECT: **3/28/2025**

IMPORTANT DEADLINE OR DUE DATES RELATED TO YOUR PROJECT:

THE RESULTS OF THIS RESEARCH WILL PROBABLY BE PUBLISHED AS:

Journal Article, Grant Proposal

IS THIS RESEARCH SUPPORTED BY A GRANT OR CONTRACT? **Yes**

If so, give grant/contract title: **Optimizing photosynthesis-nitrogen relationships for future climates: improving the predictive accuracy of gene networks with epigenetic data**

GIVE A BRIEF DESCRIPTION OF YOUR RESEARCH INCLUDING:

PURPOSE:

infer gene regulatory networks important for integrating CO2 availability and nitrogen supply in common bean.

DESCRIPTION OF VARIABLES TO BE MEASURED:

The study design is 3 different nitrogen supply with 2 different CO2 concentration, for two different varieties of beans. Transcriptome data and chromatin data will be collected.

RESEARCH QUESTIONS THAT YOU WANT TO ADDRESS USING STATISTICAL METHODS:

Transcriptome data will be used to infer a gene regulatory network using a random forest based approaches called GENIE 3 (<https://bioconductor.org/packages/release/bioc/html/GENIE3.html>).

STATISTICAL ISSUES:

We have experience with GENIE3, however, we want to adapt GENIE3 to include the chromatin data, and would like statistics expertise's opinion on how to add it (in the sense that which layers/parameter to change in the random forest).

ADDITIONAL INFORMATION YOU THINK WOULD BE HELPFUL:

ATTACHMENTS:

No Attachment

