APS-OIP Global Experience Award Report, 2019

Prepared and submitted by

Shankar K. Shakya (Ph.D.)

Bioinformatics associate scientist

Ball Horticultural Company

West Chicago, IL

Workshop: Population genetics and data analyses workshop using R programming language in Nepal

Summary

Population genetics and data analyses workshop was conducted in Nepal with the support from APS-OIP Global and Oregon State University. The three-day workshop was focused on teaching the basics of population genetics and using R to analyze microsatellite marker data. We reproduced the graphs and table from a recent study by Shakya et al. 2018. The introduction to R tutorial developed by Sydney Everhart's lab at University of Nebraska, Lincoln was also presented to cover the basics of R. A total of 26 undergraduate students (11 males, 15 females), two assistant professors from Tribhuwan University (TU), and one scientist from Nepal Agriculture Research Council (NARC, equivalent of USDA) attended the workshop. The workshop was conducted in collaboration with Paklihawa campus Tribhuwan University (TU). Subodh Khanal, assistant professor (TU), helped organize the workshop as host country coordinator.

Photo 1: Group photo after the successful completion of the workshop

Project justification

Tribhuwan university is one of the few universities in Nepal that offers degrees in agricultural science (B.S., M.S., and Ph.D.), including a major in plant pathology. However, the development of plant pathology science in Nepal is very limited, either due to lack of human resources or infrastructure, such as labs and equipment. Students with a major in plant pathology hardly get exposure to the concept of population genetics and data analysis however this is starting to change. Most students are now aware of open source software like R for data analysis. As Nepal is transitioning from applied plant pathology to molecular science, this becomes an opportune time to introduce concepts of population genetics and R to undergraduate students. Even though Nepal doesn't have its own sequencing center, Nepali scientists can send their samples for sequencing to other core sequencing facilities and analyze the data themselves.

Workshop participants

We were expecting 15-20 participants for the workshop. But because of the high interest we ended up with 29 participants (26 undergraduate students, 2 assistant professors and 1 scientist from NARC). A Google survey form was sent out before the workshop to assess the experience of participants which was then used to form a group (3-5 individuals) for data analysis.

Workshop location and facility

The workshop was conducted at Paklihawa Campus, Tribhuvan University. The campus is in the city of Bhairahawa which is 170 miles from the capital city Kathmandu. The campus offers undergraduate degrees in agriculture with majors including plant pathology, plant breeding, entomology, and horticulture. The campus is well equipped with seminar halls and also has free wi-fi for students, faculty and staff. Host country coordinator (Subodh Khanal, assistant professor) made arrangements for the seminar hall, projector and screen and other materials needed for the success of the workshop.

Workshop impact

Our goal of this workshop was to expose students to the concept of population genetics and types of data that are associated with along with its analysis. We succeeded in doing that as workshop participants were able to reproduce the graphs and tables from Shakya et al, 2018 study using R package poppr. Data mining, manipulation, filtering and plotting were the key features that workshop participants invested most of their time and showed great enthusiasm in learning statistics associated with it.

Workshop schedule

Day 1: Introduction to concept of population genetics

9am - 12pm: Allele frequencies, Hardy-Weinberg equilibrium, linkage, diversity measurement

12pm – 1pm: Lunch break

1pm - 4pm: Introduction to molecular markers and genotyping, Simple Sequence Repeat (SSR), Single Nucleotide Polymorphism (SNPs)

Day 2: Introduction to R and R packages

9am - 12pm: Introduction to R (Materials from Sydney Everhart's lab)

12pm – 1pm: Lunch break

1pm - 4pm: Data visualization in R (ggplot basics)

Day 3: Data analysis in R (Group work)

9am- 12pm: Paper discussion (Shakya et al. 2018) and SSR data preparation

12pm – 1pm: Lunch break

1pm - 4pm: Group work using using R package poppr to analyze the data and reproduce figures and tables presented in Shakya et al. 2018

Photo 2. Participants learning to import and visualize data in R.

Photo 3. Participant (Lalit BC) having a discussionn about STRUCTURE plot.

Photo 4. Sample certificate that was distributed to the workshop participants after the completion of the workshop.

References:

- Shakya, S. K., Larsen, M. M., Cuenca-Condoy, M. M., Lozoya-Saldaña, H., and Grünwald, N. J. 2018. Variation in genetic diversity of *Phytophthora infestans* populations in Mexico from the center of origin outwards. Plant Dis. 102:1534–1540

- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281