

Hilmar Lapp^{1,2}, and Participants in the Population Genetics in R Hackathon

(1) US National Evolutionary Synthesis Center (NESCent), and (2) Center for Genomic and Computational Biology (GCB), Duke University, Durham, NC, USA

Motivation

The broad and inexpensive availability of modern next-generation sequencing and genotyping technologies has led to a wealth of data and analytical methods for population genetics research. There are now dozens of packages available for analyzing and visualizing population genetics data in the popular statistical and mathematical computing platform R. However, this organically grown wealth of methods and packages, combined with the exponential growth of datasets, has also created challenges for researchers to take full advantage of these resources. It can be difficult to know which R packages are best used, and many packages do not interoperate well. A common base class that provides efficient storage of genetic data and promotes interoperability remains lacking even though the need was identified years ago. Algorithm implementations often do not scale well to the kind of large volume datasets that are increasingly common. Creating complex analysis workflows that need to pass data, metadata, and other state information from one package to another can be challenging.



To address these gaps, the Population Genetics in R Hackathon was sponsored by and held at the National Evolutionary Synthesis Center (NESCent) on March 16-20, 2015. The event targeted interoperability, scalability, workflow gaps, and gaps in end-user documentation. Its goal was ultimately to help foster an interoperating ecosystem of tools and resources for both users and researcher-developers. For more details, see <https://github.com/NESCent/r-popgen-hackathon>

