# Raman Babu, PhD Utilizes SVS to Handle Large Datasets with Quick Results

Dr. Raman Babu, a Maize Molecular Breeder at the International Maize and Wheat Improvement Center (CIMMYT, *www.cimmyt.org*) headquartered in Mexico, uses molecular marker technology to develop maize lines with increased crop yields and nutritional quality. The completion of the maize genome sequence combined with the development of low-cost, high-throughput sequencing technologies have generated a need for an analytical tool to correlate high density SNP marker information with the range of traits that CIMMYT focuses on in their breeding programs. Dr. Babu found that SNP & Variation Suite (SVS) met most of those needs. "SVS can handle a large amount of data with relative ease," he says.

The mission of CIMMYT is "to sustainably increase the productivity of maize and wheat systems to ensure global food security and reduce poverty". This goal is reached, in part, by targeting specific traits for improvement and will ultimately lead to higher-yielding and nutritionally enriched crop varieties. Babu conducts his SNP association mapping studies within CIMMYT's Global Maize Program (*www.cimmyt.org/en/programs-and-units/global-maize-program*). Using high-density marker data coupled with phenotype data, Babu aims to identify and validate the molecular markers associated with the complex trait genetics underlying the specific targets for trait improvement – for example, drought tolerance and improving nitrogen use efficiency.

Until recently, Babu's work mostly involved molecular markers for simple traits – phenotypes that are controlled by only a few genes. As CIMMYT began to employ high-throughput genotyping platforms, such as the Infinium platform from Illumina and the genotyping-by-sequencing (GBS) platform developed at Cornell University,

the data ranges increased from 200 to 300 SNPs per sample to between 5,000 and 50,000 or 500,000 and 800,000 SNPs per sample depending on the genotyping platform used. According to Babu, "That's where SVS comes into the picture."

Although Babu often uses software that is publicly available to the plant genetics community, he says there are limitations to using these programs exclusively. For instance, these programs are not updated on a regular basis, and they were developed in the pre-SNP era. Thus, they are not designed to handle large datasets. Babu appreciates a number of features that SVS has to offer. One is SVS' superior graphics as compared with other software packages available in the public domain. Also, "the regression capabilities are quite robust," he says. Babu finds the genetic association testing offered by SVS to be of particular relevance. "I like the association genetics," says Babu, "it has some built-in options for correcting the association results for population structure, and it does it quickly through

## QUICK FACTS

**Name:** Dr. Raman Babu

**Title:** Maize Molecular Breeder

**Institution:** International Maize and Wheat Improvement Center (CIMMYT)

**Location:** Mexico

**Studying:** Maize

**Data:** Illumina Infinium and GBS – Cornell

**Number of SNPs:** 50,000 – 500,000/800,000 per sample

**Analysis:** SNP and NGS

**SVS:** DNA-Seq Package

*Maize is the most important cereal crop for food in sub-Saharan Africa and Latin America.*

the principal component analysis approach, which is very efficient." He also states that SVS produces results much quicker compared with other available software. The time-saving aspect is important, as Babu does most of the analysis work himself.

Additionally, Babu has employed custom scripts from Golden Helix. His initial perception of GHI was that the company catered to the human genetics research community. "In the plant science research community, SVS is not commonly used," Babu says. "The nature of the research as well as the challenges in the plant science community are slightly different." Babu says that the GHI team is "very responsive, and I have been able to get a few custom scripts, and I am very happy about that." In particular, the GHI team is working with


*Dr. Raman Babu*

Babu to build some custom scripts for haplotype analysis that will allow the results of association analyses based on haplotypes to be corrected for population structure – which is very important to the plant genetics community.

Babu summarizes: "being able to predict the phenotypic performance of a plant without testing it in the field using the high-density marker information is at the heart of the research within the plant breeding community."

Since implementing SVS into his research less than a year ago, Babu is currently working on two or three publications involving the software, and they will be coming out soon. With all of the data that is being generated, Babu is also

using SVS as a database tool. As CIMMYT expands, they will be building databases to store the large amount of data that is being generated – via another support program called the Crop Research Informatics Laboratory (*www.cimmyt.org/en/programs-and-units/units-a-labs/crop-research-informatics-laboratory*). For now, however, "we find it quite convenient to store the data using SVS," he says "as we don't have many built in databases."

Babu states that no currently available software can answer all of his team's questions. However, SVS has the ability to export data in a variety of formats, which is very useful in Babu's research. The features that are currently available with SVS, coupled with the willingness of the Golden Helix team to help meet the ever-changing needs of their customers, provides Babu with the foundation to work toward the goal of shortening the time it takes to develop improved crop varieties and to make the whole process more economical.

### About Golden Helix

*Founded in 1998, Golden Helix is known for helping genetic research groups working with large-scale DNA-sequencing or microarray data overcome the frustration and challenges of bioinformatic roadblocks: delayed projects, lack of quality findings, and low productivity. By empowering researchers with highly effective software tools, world-class support, and an array of complementary analytic services, we refute the notion that analysis has to be difficult or time consuming. Golden Helix's flagship software product, SNP & Variation Suite (SVS), is an integrated collection of powerful data management, quality assurance, visualization, and tertiary analysis tools for genetic data. SVS is delivered in a user-friendly, scalable platform and is supported by a team of highly trained bioinformaticians, statistical geneticists, and computer scientists that together make advanced statistical and bioinformatic methods accessible to scientists of all levels.*