

## S-PLUS SOFTWARE MINES GENOMIC DATA

**Unix Reviewer:** Todd Wood, Director of Bioinformatics, Genomics Institute

**Background:** Cereals are the most important food crops in the world and determining the entire genomic sequence of a model cereal, such as rice is critical to meeting our future nutritional demands and food security needs. Rice is the single most important food crop in the world feeding over half of the world's population. The Genomics Institute uses sophisticated data mining techniques to using S-PLUS software to mine genomic data generated by rice.

**Problem Solved:** Researchers at the Genomics Institute were interested in studying genomic patterns in rice. Discoveries are made by mining genomic data using mapping and sequencing techniques. Todd Wood was interested in using powerful data analysis software to query his genomic data to make important discoveries. He selected S-PLUS 5.1 for Unix because it provided him with powerful analytical tools and unique Trellis graphics. "I selected S-PLUS for UNIX 5.1 because the software is based on the powerful next generation object-oriented language from Lucent Technologies. The S language has always been regarded as the premier language for data analysis and statistical modeling," says Wood. "With S-PLUS we benefit from superior memory resourcing allowing us to process larger data sets faster. We can pre-process our data and analyze gigabytes of data with modest computer resources."

**Product Functionality:** "The product is an invaluable tool for accessing, analyzing and visualizing data. S-PLUS supports sequential processing through block reads and writes, allowing us to analyze arbitrarily large data sets. We have the tools to handle big problems, from megabytes to gigabytes.

**Strengths:** The product makes it easy to read data from virtually any source. The comprehensive import/export capabilities reduce time spent moving data from source to source, allowing us to focus on our analysis. S-PLUS also offers a comprehensive set of traditional and modern methods. The software package also offers new statistical techniques including bootstrap, jackknife, robust MM regression, and missing data methods for variance and correlations, giving us industry leading tools.

**Selection Criteria:** I selected this product because the company has a history of developing and delivering powerful analytical tools. I believe that this product has a significant advantage over other competitors due to its powerful analytics and visualization tools.

**Deliverables:** The product allows us to query genomic data and to discover significant patterns. The unique Trellis graphics provide easy-to-read reports for communicating results to colleagues and industry representatives. Further, the graphics allow us to study our data effectively.

**Vendor Support:** The vendor support has been excellent.

**Documentation:** Yes, the documentation is complete and easy-to-read

*Source: Insightful Corp., 2007*