Next Generation sequencing technologies, namely Illumina/Solexa, Roche/454 and ABI/SOLiD provide a qualitative leap in sequencing output but each technology has its own significant sequencing error rates and systematic errors. Our current NextGen research projects include evaluation of sequence assembly algorithm performance using benchmarking on complex, eukaryotic data and guided combinations of assemblies. To fill the growing gap in Bioinformatics capacity, we are developing an automated pipeline tool (Aqwa – Automated Query and Workflow Agent) specifically designed for large-scale NextGen sequencing projects with multiple-user groups based on a generic design allowing for the incorporation of any Linux/Unix Bioinformatics tool that can be run on a single workstation or on a large-scale HPC cluster. In an effort to reduce sequencing cost and improve yield, sequence capture is used to select certain portions of a genome as inputs for NextGen sequencing. We are currently working on a whole-exome (180,000 exons, 20,000 genes) capture study with a commercial vendor (Roche/454 and Nimblegen, a Roche subsidiary) incorporating pedigree information for the verification of SNPs. In addition, we are developing a transcriptome analysis method using Illumina/Solexa sequencing in collaboration with the Miami Project for the Cure for Paralysis (Vance Lemmon Lab).

Aqwa provides a complete environment for data input and management, custom workflow generation and data sharing. Users can manage workflows (create/delete/copy workflows, add applications, set application parameters) and run them as automated pipelines that are saved to their project folders. Access to each project folder is configurable by the individual users. Output such as predicted SNPs, reference alignments and other genome-related data can be filtered using the configurable Report tool and viewed in the Ajax-based genome viewer designed to cope with large volumes of NextGen data and allow interactive interrogation of the data to determine patterns and relations with previously identified genomic features.