

Facility Description of the Bioinformatic Core

The 900 sqft bioinformatic facility is located in [1007 Noel T. Keen Hall](#). It consists of a state-of-the-art server room, one office and a multifunctional bioinformatics lab that can accommodate eight programmers/students and hands-on workshops with 14 participants.

For programming and workshops, the bioinformatic laboratory has 8 triple-bootable Windows/Apple/Linux workstations and one molecular modeling workstation from Silicon Graphics. Two 16 CPU production server systems with 15 TB storage space manages large databases, data storage and multi-user activity. Network intensive web services, such as public databases, are managed by a separate 8 CPU web/database server with an attached 12 TB storage area network (SAN). All hardware components on this high-availability SAN system are duplicated in order to avoid downtime during hardware failures. Two supercomputing Linux clusters are fully integrated into this hardware infrastructure. An older 64-CPU Linux cluster and a new one with 256 CPU cores (64x Xeon Quad Core, Harpertown, 12MB cache), 640GB of total RAM and the latest Infiniband interconnect. For memory intensive tasks, a high-memory node is available with 64GB of usable RAM for a single process. The data from all workstations and servers are automatically replicated every night onto a 24 TB backup server system that is located in a geographically separated server room. A strong focus on remote access systems maximizes the availability of all hardware and software resources for many simultaneous users from any networked location on or off campus. Online information on this hardware infrastructure, including its usage, is available at: <http://faculty.ucr.edu/~tgirke/Tools.htm>

The facility is strongly committed to maintaining a comprehensive open-source and open-access software infrastructure for Linux- and Unix-based operating systems. This approach offers access to the widest spectrum of software tools with the most advanced algorithms; and it maximizes the freedom to operate in a highly diverse and multidisciplinary academic research environments. The Debian Linux distribution is used on all workstations and servers to synchronize and automate software and OS updates on all machines of the facility. In addition, a centralized file system maximizes the efficiency and security of the data availability and access. Currently, the facility maintains over 300 open-source bioinformatics software packages for sequence/genome analysis, data mining, molecular modeling, cheminformatics, evolutionary biology, ecology, statistical analysis, etc. In areas that depend on industrial software applications, the facility also owns various commercial licenses. The most important commercial tools are the GCG package for traditional sequence analyses; the Insight II environment for protein modeling and ligand docking; the Catalyst software for pharmacophore modeling and pharmacophore-based database searching; and the Cerius2 package for large-scale small molecule mining and QSAR analyses. Online information on this software infrastructure, including its usage, is available at: <http://faculty.ucr.edu/~tgirke/Tools.htm>. As much as possible, the facility provides access to various web-based data analysis tools (e.g. EMBOSS, local BLAST server). Some of these tools were developed by the staff or the students of the facility (e.g. online miRNA prediction or compound clustering). All available online tools can be accessed through this site: <http://faculty.ucr.edu/~tgirke/LocalBLAST.htm>.