IIGB Bioinformatic Facility and Equipment Description

IIGB’s Bioinformatics Facility is located in 1207/1208 Genomics Building. The facility consists of a dedicated office for two systems administrators, one office for up to seven programmers and one for the director of the facility. A multifunctional bioinformatics lab is part of this infrastructure. It can accommodate twelve programmers and students. In addition, the facility has a dedicated 600 sqft server room with a state-of-the-art raised floor AC system and uninterrupted power supplied by a large UPS and a backup power generator. A bioinformatics laboratory is available for workshops with 10 triple-bootable Windows/OSX/Linux workstations and one molecular modeling workstation from Silicon Graphics. The most relevant HPC resources include the following components. Twelve 16-48 CPU core production server systems with 200TB storage space manages large databases, data storage and multi-user activities. Network intensive web services, such as public databases, are managed by four separate 8-48 CPU web/database servers with an attached 48TB storage area network (SAN). All hardware components of this high-availability SAN system are redundant in order to avoid downtime during hardware failures. Three supercomputing Linux clusters are fully integrated into this hardware infrastructure. An older 64-CPU Linux cluster and a new one with 348 CPU cores, 3TB of total RAM and the latest Infiniband interconnect. A dedicated cluster for memory intensive tasks consists of nodes with a total of 160 CPU cores and 256-512GB of RAM per node. The data from all workstations and servers are automatically replicated every night onto a 250TB 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. Online information on this hardware infrastructure, including its usage, is available here: http://manuals.bioinformatics.ucr.edu/home/hpc. 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 supports 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 environment. 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 data management tasks. Currently, the facility maintains over 300 open-source bioinformatics/cheminformatics software packages for high-throughput data 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 software tools. The most important commercial tools are 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. As much as possible, the facility provides access to various web-based data analysis tools. Many of these tools were
developed by the IIGB’s Bioinformatics Facility.