

NBBnet – The National Biotechnology and Bioinformatics Network: A Malaysian initiative towards a national infrastructure for bioinformatics

Mohd Firdaus Raih

School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
E-mail: mfirm@cgat.ukm.my

Sharr Azni Harmin

National Biotechnology Directorate (BIOTEK)
Ministry of Science, Technology and the Environment
Parcel C, Level 3 Block C4, Federal Government Administrative Centre
62662 Putrajaya Malaysia
E-mail: sharr@moste.gov.my

Hafiza Aida Ahmad

Interim Laboratory
The National Institute for Genomics and Molecular Biology
UKM-MTDC Smart Technology Centre
43600 UKM Bangi, Malaysia
E-mail: hafiza@cgat.ukm.my

Mohd Noor Mat Isa

Interim Laboratory
The National Institute for Genomics and Molecular Biology
UKM-MTDC Smart Technology Centre
43600 UKM Bangi, Malaysia
E-mail: emno@cgat.ukm.my

Nor Muhammad Mahadi

Interim Laboratory
The National Institute for Genomics and Molecular Biology
UKM-MTDC Smart Technology Centre
43600 UKM Bangi, Malaysia
E-mail: mahadi@ukm.my

Abdul Latif Ibrahim

National Biotechnology Directorate (BIOTEK)
Ministry of Science, Technology and the Environment
Parcel C, Level 3 Block C4, Federal Government Administrative Centre
62662 Putrajaya Malaysia
E-mail: latif@moste.gov.my

Rahmah Mohamed*

Interim Laboratory
The National Institute for Genomics and Molecular Biology
UKM-MTDC Smart Technology Centre
43600 UKM Bangi, Malaysia
Tel: 603 89267446
Fax: 603 89267972
E-mail: ram@ukm.my

Bioinformatics is a necessary technology and tool, which current molecular biologists cannot do without in light of the genome data explosion and advent of fields such as genomics and proteomics. Yet, even with this in mind, there is still a limited pool of human resource to support and develop bioinformatics particularly in

poorer developing nations. The highly multidisciplinary nature of this field can be an obstacle in converting just any molecular biologist into a proficient and capable bioinformaticist. Coupled with the high costs of software, hardware, technical maintenance and systems support; many developing and poor countries may not

* Corresponding author

be able to fully realize their biotech potential either academically or economically due to the inability to fully utilize bioinformatics as a technology to support and co-develop with molecular biology oriented research programmes. Malaysia, a developing South East Asian nation took steps to counter these problems and develop an infrastructure for bioinformatics to support nationally important biotech research agendas, particularly government initiatives in genomics and proteomics, with the setting up of NBBnet – The National Biotechnology and Bioinformatics network. We further propose the setting up of similar models in other developing countries, an endeavour which we are willing to share our experiences and resources. We hope that these networks may in the future evolve into a super network of national or regional networks on biotechnology and bioinformatics.

Genesis and evolution

Developing nations with limited financial and human resources will most likely have many factors constraining biotechnology related research. This is unfortunate as many of these developing nations, such as those with tropical rain forest climates, are also endowed with a rich diversity of flora and fauna, the raw materials of many biotech related ventures. Seeing the vast economic potential in its own backyard, Malaysia has set out to mine these virgin resources from her rainforests and seas while maintaining minimal impact, imbalance and intrusion to the environment and local ecosystems. Technologies such as genomics and proteomics with molecular biology as the basic driving science were selected to spearhead these efforts to harness Malaysia's biological diversity. Genomics and proteomics themselves, being already expensive forays, was still incomplete without bioinformatics. Malaysia, like many other developing nations faced problems such as a shortage in human resource (bioinformaticists) for this niche as well as other aspects of bioinformatics such as hardware and software.

Realizing this, the idea of a nation wide network and information technology infrastructure supporting biotechnology and bioinformatics research in Malaysia was mooted in 1997. This was followed with the mustering of a team to build resources, maintain and support as well as to study the future directions and mode of operations of the mooted national network. By the end of 1999, a modest operation was running providing rudimentary technical support services particularly to help molecular biologists unfamiliar with bioinformatics tools. This original team also set out to assimilate and retool as many Malaysian molecular biologists as possible with training on the usage of bioinformatics applications and softwares. The main objective of this retooling exercise was mainly to build up a critical mass of users who in turn will encounter problems

in the course of their research. These problems will therefore act as the driving force for the network to solve and therefore transform from a community of users, to a community of bioinformatics developers starting with the core team. The National Biotechnology and Bioinformatics network or its acronym NBBnet was officially launched in 1999 (<http://www.nbbnet.gov.my>).

The next evolution of the national network involved a more niche oriented field: protein structure prediction and modeling. A network of labs was set up to provide tools for protein structure prediction and modeling. These labs were equipped mainly with Silicon Graphics workstations and were running MSI's (now Accelrys Inc.) InsightII protein modelling suite. Undoubtedly the cost of these software and hardware were considered steep for the standard of Malaysian Research and Development expenditure. To enable more users to benefit from these investments, the hardware and accompanying software were placed at central geographic locations to allow easy and unlimited access for any interested party. This initiative was named the NBBnet Virtual Protein laboratory and has since then expanded to include more distributed user labs from the initial four. This expansion also included beefing up existing capabilities with more hardware and a wider software range as user needs and demands also grew. Training and retooling exercises remained a constant factor in the network's development plans throughout these efforts.

Operations and services

A dedicated development and support unit was then set up in 2000 and was operationally supporting the Malaysian biotechnology Research and Development community by 2001. This particular evolution saw the objectives and direction of many NBBnet initiatives diverge from a purely support and applications usage role to an active role in research and development as well as infrastructure management.

NBBnet maintains a core skeletal staff mainly tasked with maintenance, development and services operations and are physically based at the NBBnet Bioinformatics Research and Development unit at the interim laboratory of the National Institute for Genomics and Molecular Biology, Malaysia (<http://cgat.ukm.my/genomicslab/>). Since NBBnet also operates on a concept of distributed user centers, these centers are further maintained by minimal staff, mainly to oversee the maintenance and operations of the services provided. The distributed user centers are usually overseen by one academic or senior research staff and assisted by a couple of postgraduates or research officers. These staff are in effect employed by the partnering or hosting institution for the user center or lab. By this type of partnership, the NBBnet infrastructure put in place, was able to operate with

just three full time personnel who are administratively attached to NBBnet and based at the National Biotechnology Directorate (<http://www.biotek.gov.my>). All other personnel, including the development personnel at the Research and Development lab, are attached to their respective institutions which in turn act as the physical host or distributed user centers. Another advantage to this arrangement was that the personnel involved, particularly on the development side, were full time scientists or academics with specific research interests and problems to solve. This in turn fuels the development process as well as drive the operations side as these personnel are also active users of the infrastructure and are therefore able to troubleshoot and act as technical consultants for other NBBnet users.

The NBBnet infrastructure also included virtual or online assets which were developed within NBBnet for internal usage. These included a personal resource management portal from which users have access to remote accessible softwares, personal database management, BLAST database management, personalized directory of web accessible bioinformatics resources as well as forums, message boards and other similarly oriented media for electronic discussion and information sharing. All the software resources developed in-house, were built at zero financial cost save for the salaries and time of the development personnel. This interface, named “Bioinformatics Tools @ NBBnet” is accessible to NBBnet members via a web enabled interface (<http://www.nbbnet.gov.my/tools/>; <http://cgat.ukm.my/tools/>). Commercial software packages were also included for use by NBBnet members. Amongst these are the GCG package (Accelrys Inc, San Diego, CA) for sequence analysis and gRNA – genomics research network architecture, an API package for bioinformatics and genomics analysis and development (Helixense Pte. Ltd., Singapore – <http://www.helixense.com>). The “Bioinformatics Tools @ NBBnet” portal also serves as a bioinformatics applications directory and search engine with unrestricted access for any interested user.

The research and development team are also involved in installing and running other freely available, academically licensed softwares (to NBBnet) and open source software packages such as HMMer, Gromacs, the Staden package, Artemis, EMBOSS and other similar software. Another important role played by this virtual infrastructure concept is the hosting of databases and research groups websites. NBBnet members and affiliated research groups can have their databases (relational and BLAST) hosted on NBBnet servers. This eliminates the research group from having to maintain hardware, administer network and security as well as avoid down times occurring on less robust system architectures. The database hosting capability is operated via a Linux Virtual Server (LVS) running on a parallel

cluster. This architecture enables load balancing as well as a high availability capability to the hosted resources. Users are therefore able to manage and upload their own research related databases which is accessible 24 hours a day, 7 days a week. Access to the databases hosted are exclusively controlled by the respective data owners with NBBnet acting only as a host. This has resulted in many of the databases being proprietary to the developing party and its members.

From the short amount of time that these services have been online, the response and usage has generally been favorable with many users welcoming the lessened responsibility and hassles on their part in operating and maintaining databases crucial to their research. Such matters may seem trivial to better equipped labs and more financially privileged nations, but it was seen that for a country like Malaysia, the infrastructure put in place is hoped to have helped narrow the gap in bioinformatics between Malaysia and other better equipped nations. It is hoped that research output will increase by minimizing the time spent on system administrative matters and concentrating on the subject matters that each researcher knows best and therefore further developing those niche areas to benefit Malaysia and the world at large. The other obvious benefit of this resource sharing initiative is of course in cost cutting. Instead of the need to acquire many servers and the staff to maintain them, NBBnet was able to acquire just a centralized hosting resource. The staff responsible for maintaining and operating this resource being already familiar with the task at hand and at the same time developing new tools and resources for their own use as well as other NBBnet members. In our opinion, it is this symbiotic relationship that has enabled NBBnet to survive and strive forward.

Research and development initiatives

As with many scientific and technological ventures, research and development remains an important factor. This rings true for the NBBnet program as well. Amongst the primary activities for NBBnet is development of resources, resource management systems and interfaces. Since its inception, the majority of NBBnet’s Research and Development efforts have been geared towards information management and database or knowledgebase development. The problems that arised during daily research were mainly centered around management of sequence data, information acquisition from these data. As a result amongst the first initiatives was the setting up of a functional genomics knowledgebase for *Burkholderia pseudomallei*. One notable foray into a niche knowledgebase concept was a bacterial proteases database acronymed as ProLysED (Prokaryotic Lysis Enzymes Database – <http://cgat.ukm.my/prolyses/>).

To date, NBBnet has developed and is hosting a diverse range of databases such as those on bacterial comparative genomics, *Eimeria tenella* functional genomics, *Burkholderia pseudomallei* fuctional genomics, seaweed

genomics, fish (*Lates calcarifer*) genomics as well as databases on bacterial proteases and tropical animal diseases to name a few. As mentioned earlier, many of these resources still have restricted usage and access by non group members. Development of NBBnet databases and applications have mainly utilized open source resources such as the MySQL relational database management system (<http://www.mysql.com/>) and the Linux operating system (<http://www.linux.org/>). Coding and scripts used or written were also open source friendly such as perl (<http://www.perl.org>; bioperl – <http://www.bioperl.org/>), python (<http://www.pyhton.org>; biopython – <http://www.biopython.org>), PHP (PHP hypertext processor – <http://www.php.net/>) in addition to more complex languages such as C. We see the web integration of databases, as discussed by Xia et al. 2002, as an important avenue for biologists to share data and information. As a result, all NBBnet developed or hosted databases are interfaced for web operations.

Our research and development efforts also concentrate on the development of non-platform specific as well as unifying tools which we hope will bridge between the many applications and formats available and in use in bioinformatics today. As discussed by Stein, 2002, we fully believe that biological data and information should be able to be fully exploited by all whenever possible. In doing so, we hope to exploit and develop open source technologies for bioinformatics applications as well as promote and integrate the usage of current formats and standards. Our model of operations tries to concur with the guidelines as listed out by Stein, 2002 whenever possible.

We have noted earlier the constant training and retooling exercise which we believe to be vital to the development of bioinformatics expertise. This training is usually done through periodic workshops on either bioinformatics in general or focused towards certain niches within bioinformatics such as programming and database development, protein modeling and genome informatics. A significant portion of the “Bioinformatics Tools @ NBBnet” is dedicated to online tutorials and what we have termed as a bioinformatics-biotechnology infoportal. This online tutorial and self learning part of the portal, includes the S-Star Alliance (<http://s-star.org>) online lectures bioinformatics mirror site. Another interesting concept for development work used was attachments of biologists to the research and development lab. These biologists came from the labs of other NBBnet members and had specific problems to solve or specific tools to develop but were not equipped with the technical skills or know how to do so. Via these attachments, they were able to acquire the skills needed for the mission critical purposes and at the same time have access to the hardware and network infrastructure provided under the NBBnet program. During these

attachments, the participants are exposed to programming and development through class instruction and practical sessions. After acquiring a degree of competency, they then start with developing their research specific tools or databases while still in the attachments periods. Upon return to their respective labs, they would usually have gotten their respective projects of the ground with consultative assistance from the Research and Development team and continue to administer the databases or tools that they have developed and hosted on NBBnet servers remotely from their institutions. These attachments periods normally last from a fortnight to a month.

Conclusions and future directions

The NBBnet program and initiatives can be seen as a working model to provide bioinformatics infrastructure using shared resources over dispersed geographical locations for a developing country. The funding and policy doctrine for this infrastructure was provided by the Malaysian government through the Ministry of Science, Technology and the Environment, while the input for development and operations was provided by the Malaysian academic and biotechnology research sector such as universities and research institutes. This symbiosis of resources can be seen as the major factor in the ability of NBBnet to take off and continue operating. While the network still has far to go, the objective to provide Malaysian biotech researchers with resources in bioinformatics regardless of their geographical locations and financial or research grant status can be said as having been achieved. Our aims and objectives were simply to provide skills, tools and outlets in terms of infrastructure for bioinformatics oriented research to the Malaysian biotechnology community.

We envision the future of the network to progress to a full integration of national Research and Development resources. This would include the integration of laboratory management systems and data analysis systems over a nation wide grid. Such integration will allow users to distribute computational jobs to utilize idle CPU resources. This connectivity will also enable better resource and data management, be it computational resources or laboratory based resources. As an example, planning and development is under way to enable information sharing and control of major laboratory equipment. An NBBnet member will be able to check where a certain resource is located, when it will be free for use, when the results are expected and from there on channel the results to analytical stages of this pipeline. As a further development, similar national networks can liase with bigger regional networks such as APBionet (Asia Pacific Bioinformatics Network – <http://www.apbionet.org>) and from there connect to a global network of biotechnology and bioinformatics resources.

Acknowledgements

We thank the Ministry of Science, Technology and the Environment, Malaysia for the funding and support to make NBBnet operational. Our gratitude is also directed to Dato' Dr. Islahudin Baba for his untiring support to get NBBnet off the ground. The authors would also like to thank Khairil, Fuad, R. Murzaferi and Rashdi who form the core of the development and operation team as well; for maintaining the resource databases and M. Yusof Radzuan Saad, for overseeing the mirrors.

References

STEIN, Lincoln. Creating a bioinformatics nation. *Nature*, May 2002, vol. 417, p. 119-120.

XIA, Yulu; STINNER, Roland E. and CHU, Ping-Chu. Database integration with the web for biologists to share data and information. *Electronic Journal of Biotechnology* [online] August 15 2002, vol. 5, no. 2, [cited February 28 2003]. Available from: <http://www.ejbiotechnology.info/content/vol5/issue2/full/8/index.html>. ISSN: 0717-3458.