Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Update of microbial genome programs for bacteria and archaea

Patricia Borges Santos Celestino[1,3], Lydston Rodrigues de Carvalho[2], Leandro Martins de Freitas[3], Fernanda Alves Dorella[3], Natalia Florêncio Martins[4], Luiz Gustavo Carvalho Pacheco[3], Anderson Miyoshi[3] and Vasco Azevedo[3]

[1]Department of Botany and Microbiology and Department of Zoology, Ohio Wesleyan University, Delaware, OH, USA
[2]Departamento de Bacteriologia, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil
[3]Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[4]Laboratório de Bioinformática, Embrapa Recursos Genéticos e Biotecnologia, Brasilia, DF, Brasil
Corresponding author: V. Azevedo
E-mail: vasco@mono.icb.ufmg.br

Genet. Mol. Res. 3 (3): 421-431 (2004)
Received August 2, 2004
Accepted August 31, 2004
Published September 30, 2004

ABSTRACT. Since the *Haemophilus influenzae* genome sequence was completed in 1995, 172 other prokaryotic genomes have been completely sequenced, while 508 projects are underway. Besides pathogens, organisms important in several other fields, such as biotechnology and bioremediation, have also been sequenced. Institutions choose the organisms they wish to sequence according to the importance that these species represent to them, the availability of the microbes, and based on the similarity of a species of interest with others that have been sequenced previously. Improvements in sequencing techniques and in associated methodologies have been achieved; however, scientists need to continue working on the development of this field. In Brazil, a multicentered, centrally coordinated and research-focused network was adopted and successfully used for the sequencing of several important organisms. We analyzed the current status of microbial genomes, the trends for

criteria used to choose new sequencing projects, the future of microbial sequencing, and the Brazilian genome network.

Key words: Microbial genome project, Biotechnology, Bioinformatics, Genomics, Prokaryotes, Brazil

INTRODUCTION

Prokaryotic microorganisms comprise the largest part of the planet's total biomass. This group contains a vast array of species, with enormous genetic, metabolic, physiological and behavioral diversity; however, less than 1% of them have been cultured. Despite their ubiquity, little is known about their fundamental properties, about their range of diversity, about how they interact with the environment, about their evolution, and about the roles they play in global biogeochemical cycles (Rodrìguez-Valera, 2004). It is believed that progress towards filling these knowledge gaps will advance significantly when more whole genome sequences become available.

The bacteria have long been the subject of scientific study due to their ability to cause disease in humans (Lederberg, 2000). One of the major advances in the health and well-being of human civilizations was the development of antibiotics. Although the introduction of antibiotics has had an enormous impact on the ability to treat bacterial infections, bacteria continue to be the leading cause of death worldwide. Moreover, the effectiveness of antibiotics has been eroded by the appearance of pathogenic strains that are resistant to nearly all classes of antibiotics, coupled with the fact that only one new class of antibiotics has been introduced by the pharmaceutical industry since the mid-1970s (Binder et al., 1999).

Clearly, the discovery of new therapies against diseases caused by bacterial pathogens is a critical necessity of the 21st century. Over the past decade, the field of genomics has revolutionized both basic research, and particularly the pharmaceutical industry. The field of genetics was also fundamentally affected by bacterial genetic research. Starting in 1928, studies on transformation of pneumococcus by Griffith established a new critical concept in genetics: that DNA was the genetic material of life. Over the next few decades, other genetic breakthroughs, including determination of the mechanisms of replication, transcription and translation of the genetic code, and of the structure and expression of genes, were made available through microbiological research using bacteriophages and the workhorse bacterium *E. coli*. In addition, a number of molecular tools were discovered in bacteria, such as DNA ligases and restriction enzymes. However, it was not until the landmark work of Cohen (1993) that these enzymes were used together along with plasmid replicons to enable the cloning of DNA fragments (Lederberg, 2000). This led to the birth of the field of molecular biology, which had a profound effect on drug discovery and development. Instead of using brute force protein purification to isolate targets for small molecular compounds or therapeutic proteins, cloning and expression technologies allowed these entities to be supplied in bulk. The age of molecular biology transformed the pharmaceutical industry, and the newly spawned biotechnology industry, on an unprecedented scale, perhaps only matched by the recent breakthroughs in genomics.

The current availability of bacterial genome information, originated from molecular biol-

ogy accomplishments, has allowed hundreds of protein-protein interactions to be predicted, based solely on sequence comparisons. Moreover, genome sequence information can now be coupled with other experimental data (structures, domain shuffling, expression patterns, and gene adjacency in genomes) to allow new approaches to determining gene function. Nowadays, genomics, and especially metagenomics approaches, represent an advance in knowledge and understanding of microbial biology, since it is not possible to transform a bacterial strain, delete gene information or manipulate any level of protein expression of a non-culturable bacteria using traditional classical genetics techniques.

We examined the current status of microbial genomics, analyzed the trends in this field, looked at some criteria that can be used to elect a microorganism to be involved in future genome projects, speculated on consequences and future applications derived from this knowledge, and examined the networks for genome sequencing in Brazil.

CURRENT STATUS

Among other biological sciences, microbiology has been one of the greatest beneficiaries of the breakthrough in genomics and bioinformatics technologies that followed after the first whole prokaryotic genome sequence was published in July 1995 - that of *Haemophilus influenzae* (Fraser et al., 2002). Since then, more investments were made in this technique. Up to June 2004, 172 prokaryotic genome projects had been completed and 508 projects were in progress (GOLD[TM], 2004). The major focus on pathogens (53.3% of all genomes completed) now shares interest with a few model microorganisms and a few unusual organisms, such as *Deinococcus radiodurans*, a microorganism known to be the most radiation-resistant of all (Nelson et al, 2000; GOLD List, 2004).

In the past decade, the progress in DNA sequencing and assembly, the faster generation of shotgun sequences, and the use of sophisticated methods for annotation have reduced the time required for each stage of a genome project and the cost per base pair, resulting in a finished product of higher quality (Nelson, 2003; Simpson et al., 2004b). The improvements in sequencing have been accompanied by free access to these sequences in public databases (Table 1). These public databases can aid scientists in isolating genes, comparing genomes, relating species evolutionarily, and speculating on the presence and function of genes, and consequently of the proteins that genes code for.

Table 1. Four commonly used public sequence databases.

Databases	Webpage
GOLD™ Genomes Online Database	http://www.genomesonline.org
The Institute for Genomic Research - Microbial Database (TIGR)	http://www.tigr.org/tdb/mdb/mdbinprogress.html
National Center for Biotechnology Information - Microbial Genome (NCBI)	http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi
The Wellcome Trust Sanger Institute - Microbial Genome	http://www.sanger.ac.uk/Projects/Microbes/

The information derived from whole-genome sequences following their comparative analysis can be used in studies that search for novel aspects of biochemistry, physiology, and metabolism of these organisms to investigate the roles microorganisms play in complex ecosystems and in global geochemical cycles, to study their diversity, to predict the impact microorganisms have on the productivity and sustainability of agriculture and forestry and on the safety and quality of food supply. Also, new genome sequences can be used to infer phylogenetic relationships among prokaryotes that deal with the organization and evolution of microbial genomes, mechanisms of transmission, exchange and reshuffling of genetic information (Koonin, 1997; MGSP, 2003).

A phylogenetic tree was designed based on the 16S rRNA sequences of genera of prokaryotes that have been sequenced or with sequencing in progress (Figure 1). The tree was limited due to the use of genera instead of species in its construction. However, although members of different, traditionally defined species within some genera contain distinct gene sequences, this does not always hold true (Fox et al., 1992). By analyzing this tree, it is possible to observe a bigger concentration of genera in the Gamma Proteobacteria group - including some that contain pathogens such as *Escherichia*, *Haemophilus*, *Vibrio*, and *Salmonella* - followed by the Alpha Proteobacteria, Firmacutes, and Actinobacteria. Overall, the figure shows how the availability of more sequences can allow scientists to understand evolution and pathogenicity through the distribution of genome sequences of the tree of life, to discover new or different relationships among prokaryotic organisms, and to perform comparative genome studies, including analysis of genome composition, gene organization, and gene families within and across the domains (Relman, 1994). However, for easier access, comparison and study of large numbers of genomes, it is necessary for databases to standardize their genome annotation formats, and for us to develop computer programs that are capable of analyzing larger groups of sequences, while making these programs available to most scientists (Nelson et al., 2000).

CRITERIA FOR CHOOSING NEW PROJECTS

When the first microorganisms started to be sequenced, preference was given to those microbes that were important for human, animal, or plant health. Examples are the first microbe sequenced, *Haemophilus influenzae*, followed by others such as *Xylella fastidiosa* and *Brucella melitensis* (Simpson et al., 2000; DelVecchio et al., 2002; Fraser et al., 2002). However, soon after, some scientists started to focus on the sequencing of microbial genomes that were significant to their own interests and institutions, including species important for veterinary application, plant pathology, study models, and biotechnological uses (GOLD List, 2004). For example, the U.S. Department of Energy (DOE) started in 1994 a Microbial Genome Program (MGP) focused on sequencing nonpathogenic microbes that appeared to show some importance to their activities, such as research in bioremediation, biotechnology, global climate change, energy production, ecology, and evolution. Thus, they chose prokaryotes of the archaea group, such as *Methanococcus jannaschii*, *Archaeglobus fulgidus* and *Thermotoga maritma*, important for evolutionary studies and environmental remediation, respectively, and the bacterium *Shewanella putrefaciens*, important for remediation (DOE MGP Report, 2000; U.S. DOE Microbial Genome, 2003).

As the DOE case exemplifies, the current trend for the criteria used in industry, academic and governmental institutions to choose new microbial genomes to be sequenced seems

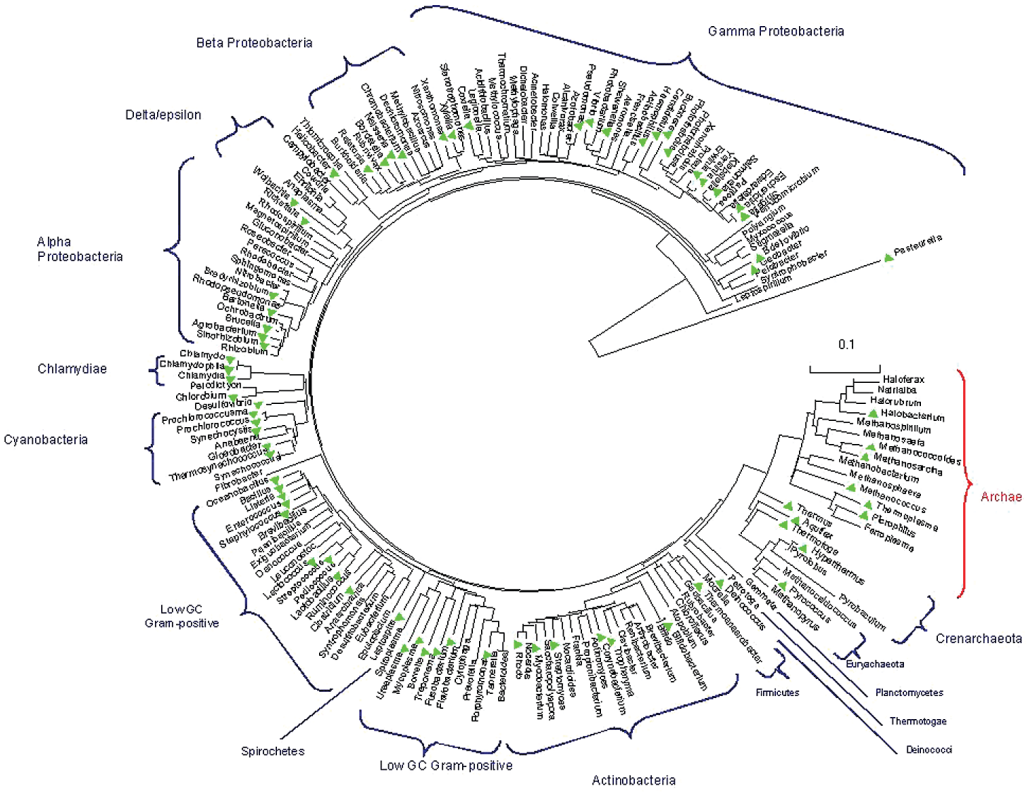

Figure 1. Phylogenetic tree based on completed and ongoing 16S rRNA genomic sequences. Unrooted phylogenetic tree based on 16S rRNA sequences for prokaryotic organisms obtained by the neighbor-joining method (Saitou and Nei, 1987) after alignment with CLUSTAL W (1.82) multiple sequence alignment (Thompson et al., 1994). The phylogenetic distances were calculated by the software MEGA 3 (Kumar et al., 2004) with consistency of data tested by bootstraping the alignments 500 times. Triangles represent complete genomes at the GenBank (updated on July 23, 2004). The phylogenetic groupings are indicated by the keys (Archaea in red, Eubacteria in blue).

to follow the relevance that each microbe has for the particular institution (U.S. DOE Microbial Genome, 2003). This trend was criticized by Barry Bloom (Harvard School for Public Health, USA) in the 14th Genome Sequencing and Analysis Conference in Boston, USA, who brought up the fact that even today, all over the world, people suffer and die of infectious diseases (Kemmer and Fraser, 2002). Therefore, we see that today, some still defend the idea that scientists should continue to prioritize the use of genome-sequencing technologies towards microbes that directly affect human lives. This is not to say that other fields should be overlooked. However, a substantial effort should continue to be made, especially by governmental agencies, in the sequencing of human pathogens and the application of the data collected from these projects, while other microbes involved in fields that include animal and plant health, and industrial and environmental applications, should receive support proportional to their importance. For example, after the terrorist attacks, the United States government started to invest more on biodefense research (Fraser, 2004). The availability of genome sequences is one strategy that can help scientists understand organisms that could be used in bioterrorist attacks and thus develop ways to prevent and fight diseases. On the other hand, for private institutions, this may not be the case.

The problems with criteria not only depend on what microbes are important, but also on which ones are available. For instance, the speed at which microbial genomes are being sequenced is not proportional within the prokaryote group. Up to now, of the 172 complete prokaryotic genomes, 19 belong to the Archaea domain (see Figure 2) (GOLD[TM], 2004). Some archaea are known for their ability to live in extremes of salt concentration, temperature, pressure, etc. Although these characteristics may be beneficial for their utilization in industry, environment, biochemistry, and biotechnology, they may also make these microorganisms hard to be grown in laboratory, and could result in difficulties in manipulation prior to sequencing; the relative paucity of basic information on the biology of archaea is also a problem (DOE MGP Report, 2000). However, many of the archaea representatives have been identified recently and are now being characterized, as we learn ways to culture them by supplying their special needs (Nelson et al., 2000).

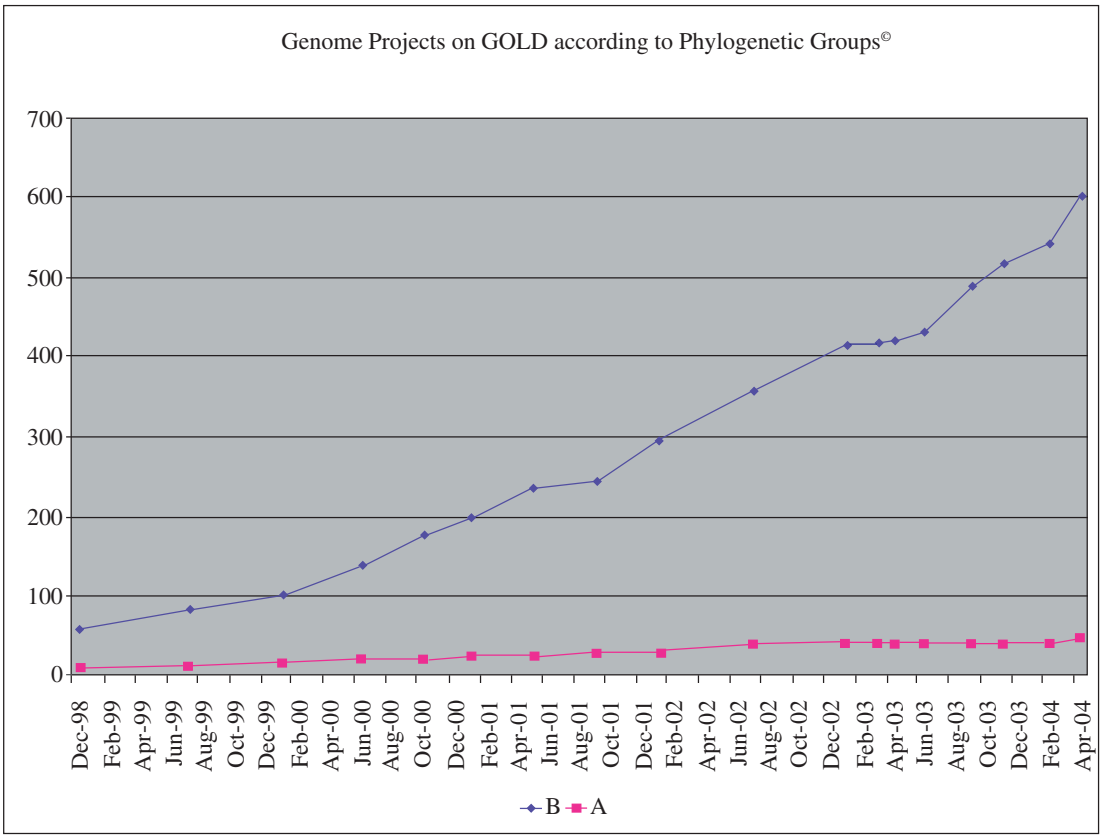

Figure 2. Update on progress of prokaryotic sequences. B = Bacteria. A = Archaea. Source: modified from GOLD[TM], 2004.

The ability to culture a microbe certainly assists the sequencing of genomes. Some laboratories have already developed techniques to sequence organisms without ever culturing them (Kemmer and Fraser, 2002). This technique is important for the case of organisms that are not well understood or that live in very complex environments and can be done by obtaining DNA directly from the environment, for example, from the sea, soil, as well as the human oral

cavity and the gastrointestinal tract (Nelson, 2003). This technique, if explored, will allow scientists to discover new enzymes, antibiotics, and other microbial products useful in bioremediation, biotechnology, medicine, industry, etc. Another application of sequencing directly from environments is in the better understanding of the soil metagenome and of the metagenome of a healthy versus a diseased individual (Nelson, 2003). The gene pools present in a prokaryotic species can be orders of magnitude larger that the genome of a single strain. Contrasting with eukaryotic genomes, the repertoire of genes present in a prokaryotic cell does not correlate stringently with its taxonomic identity. Hence, gene catalogues from a particular environment might provide more meaningful information than the classical species catalogues. Metagenomics, or microbial environmental genomics, provides a different tool that focuses on the habitat rather than the species. It could, therefore, be the right tool to complement organismal genomics, and better understand microbial ecology, and prokaryotic diversity and evolution (Rodrìguez-Valera, 2004).

A third problem with the criteria used for choosing genomes to be sequenced is in deciding whether organisms that have similar relatives already sequenced should be given priority over little studied organisms. Again, the answer depends on the necessity, on the interest of the particular institution involved, and on the research objectives. In some cases, one organism in a group may answer a particular question, while in other cases, it is necessary to have several samples of a group sequenced and studied for a more complete answer (Kemmer and Fraser, 2002).

FUTURE DIRECTIONS

Every genome that is sequenced provides information that can be used for close relatives of that species and for insight into gene functions, biological processes, evolution, and possible applications of the microorganisms through comparative tools and sequence databases now available (Nelson, 2003; Thomson et al., 2003). For example, scientists found out that *Mycobacterium bovis* and *M. tuberculosis* have 99.95% of their DNA identical, and therefore, the ability to infect bovines or humans must be due to differences in gene expression by each species (Thomson et al., 2003). However, until 2003, there had not been a prokaryote that had all its gene functions known, and in addition, among the microbial genome sequences completed until 2003, it was estimated that about 40% of the potential genes coded for proteins whose functions were unknown to scientists. This fact indicates that scientists need to continue to work with individual genomes, to investigate unknown genes present in larger groups through the use of proteomics, and to study how genes are expressed and how proteins and nucleic acids interact. One step was given through the use of microarrays, which can both detect the presence of genes in an organism and study gene expression (Nelson, 2003).

Although the number of finished microbial genome projects is getting larger, there is still a lack of basic information concerning microorganisms of environmental and veterinary interest. In this case, the finished microbial genomes can be used, in the future, in comparative studies to complete genome sequences for closely related organisms, and also as a scaffold to order and orient contigs, to name a few applications (Fraser et al., 2002). Presently, we are seeing a strong movement towards the sequencing of genomes of species that have already been sequenced but differ phenotypically from strain to strain. These differences could be due to mutations in the DNA, which can be detected through genome sequencing. Some scientists are also using genome sequences of non-pathogenic bacteria that are similar to pathogens to gain insight into the study of the properties of the latter (Holden et al., 2004). Therefore, be-

cause different species and organisms of the same species are important and potential candidates for future genome sequencing, there is a present need to develop software able to work with enormous numbers of databases to perform post-genomic analysis. The analysis of these upcoming genomes will help us solve problems brought by imperfect genomes and incomplete databases that were concluded on the beginning of the genomics era (Fraser et al., 2002).

The future holds a continued effort to drive the costs of genome projects down with the acknowledgment of the importance of funding future genome projects, updating already assembled facilities and creating new ones. The future also holds more efficient applications of the knowledge that can be brought through the sequencing of microbes, for example, in understanding the nature and function of bacteria that cause diseases, developing vaccines, identifying better ways to cure diseases with specific targeting of microbes, better isolating particular organisms or genes, better understanding host-pathogen interactions, finding new evolutionary relationships, and recognizing, understanding and wisely utilizing the diversity of life on Earth. Scientific communities and governments need to consider how much there is to learn about the diversity of microbial life on our planet, and what this knowledge represents in terms of development, health and economics. They will then realize that funding of this field of research is one of the best investments that will ever be made (Fraser et al., 2002).

THE BRAZILIAN PROGRESS IN MICROBIAL GENOME SEQUENCING

In Brazil, the principal incentive for participation in genomic sequencing projects came from the need to improve technology in the country so as not to be dependent on richer nations (Simpson et al., 2004b). In 1997, motivated by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), a governmental foundation that supports scientific research in São Paulo, the state's scientific community, focused on biotechnological development through genomics. The genomic program was supported not only for the benefit of sequencing new organisms but also for the benefit of the scientists to become familiar with new biotechnological research and procedures. A network of laboratories known as REDE ONSA (Organização para Sequenciamento e Análise de Nucleotídeos) was organized (Simpson et al., 2004b). This alternative method was composed of a multicenter, centrally coordinated and research-focused network that Brazil used in sequencing genomes. Most countries in the world do this type of activity in large, high-tech centers built for genome sequencing (Camargo and Simpson, 2003; Simpson et al., 2004a). The Brazilian system was interconnected through the Internet, and each of the centers had the benefit of receiving financial support and new equipment for the joint completion of the project (Camargo and Simpson, 2003).

The decision to sequence the genome of *Xyllela fastidiosa*, chosen after the network was organized and funding was available, was based on the relevance of the organism for scientific knowledge and for the Brazilian economy, especially for the State of São Paulo (Simpson et al., 2004b). This bacterium causes citrus variegated chlorosis, which affects oranges and coffee, compromising orange plantations in São Paulo that employ 400,000 workers and generate US$1.4 billion per year through the exportation of juice (Camargo and Simpson, 2003; Simpson et al., 2004b).

After the successful work done with the sequencing of *Xylella fastidiosa* through the ONSA network, a national sequencing network, called BRGene, was developed with incentive from the Ministry of Science and Technology and CNPq (Conselho Nacional de Desenvolvi-

mento Científico e Tecnológico) along with other regional networks. To date, a number of networks have been involved in genome projects in the country, including the sequencing of the bacteria *Xanthomonas citri*, *X. campestris*, and *Chromobacterium violaceum* (Camargo and Simpson, 2003).

Simpson and collaborators (2004a) have recognized the success of the Brazilian network, not only because of the rapid sequencing and the successful organization and utilization of a new method for development of scientific research, but also as a means to engage in complex research. This type of network research is seen as a better way to complete a research project and directly contribute to society in opposition to research performed by an individual or by a small group. This is true because individual research programs are limited by the number of scientists working and by time and by the size of the project (Simpson et al., 2004a). Another success of the program was that it made Brazil competitive internationally in the genomic era, without the need of waiting for the construction of large, specialized centers for genome sequencing.

CONCLUSIONS

Overall, for the development of science and for better understanding of all types of organisms, it is mandatory that scientists continue to sequence genomes of a greater diversity of organisms, and continue to invest in the development of new and in the improvement of older techniques to analyze and make use of the sequences. In general, the choice of microbes to be sequenced seems to follow an individualistic trend, in which choices are made depending on the institution and on the environment in which it is situated. However, it is still necessary that, for the benefit of human kind, some institutions keep on focusing on the deeper understanding of organisms that directly affect our lives.

A case for us to observe and question is the Brazilian genome sequencing network. Should collaborations continue to be encouraged in Brazil? Should other countries adopt this system? In which situations? As the REDE ONSA and other national networks have shown, Brazil should continue to invest in this successful technique and encourage the adoption of this method by other developing countries, especially those that have a weak biotechnological background but are capable of financing their progress in this field without the need to wait for the construction of very large and specialized centers. For richer countries, this system should be considered for testing, without compromising their already established research system. On the other hand, although the exchange of information and the widespread cooperation in a scientific project may be of great benefit for the growth and development of science and scientists, some may oppose initiatives to use this system due to financial and institutional competition.

ACKNOWLEDGMENTS

We thank Dr. Lawrence E. Young Award of Ohio Wesleyan University and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for funding. We are also indebted to Dr. Leon Rabinovitch from IOC/FIOCRUZ for advice, friendship and trust. We give special acknowledgments to the "Genética de Microorganismos Procariotos" lecture graduate students ICB/UFMG, 1st semester 2003 class, for encouragement and productive discussions at the beginning of this project.

REFERENCES

Binder, S., Levitt, A.M., Sacks, J.J. and **Hughes, J.M.** (1999). Emerging infectious diseases: public health issues for the 21st century. *Science 284*: 1311-1313.

Camargo, A.A. and **Simpson, A.J.** (2003). Collaborative research networks work. *J. Clin. Invest. 112*: 468-471.

DelVecchio, V.G., Kapatral, V., Elzer, P., Patra, G. and **Mujer, C.V.** (2002). The genome of *Brucella melitensis*. *Vet. Microbiol*. *90*: 587-592.

DOE MGP Report (2000). Microbial genome report program. Available at http://www.ornl.gov/sci/techresources/Human_genome/publicat/microbial/. Accessed on June 29, 2004.

Fox, G.E., Wisotzkey, J.D. and **Jurtshuk, P.J.** (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol. 42*: 166-170.

Fraser, C.M. (2004). A genomics-based approach to biodefence preparedness. *Nat. Rev. Genet. 5*: 23-33.

Fraser, C.M., Eisen, J.A., Nelson, K.E., Paulsen, I.T. and **Salzberg, S.L.** (2002). The value of complete microbial genome sequencing (you get what you pay for). *J. Bacteriol*. *184*: 6403-6405.

GOLD[TM] (2004). Genomes online database. Available at http://www.genomesonline.org/. Accessed on July 6, 2004.

GOLD List (2004). Complete genome list. Available at http://www.genomesonline.org/CompleteGenomesList.html. Accessed on July 12, 2004.

Holden, M., Crossman, L., Cerdeno-Tarraga, A. and **Parkhill, J.** (2004). Pathogenomics of non-pathogens. *Nat. Rev. Microbiol. 2*: 91-92.

Kemmer, D. and **Fraser, A.** (2002). Whose genome is next? *Genome Biol. 3*: 4037.1-4037.3.

Koonin, E.V. (1997). Big time for small genomes. *Genome Res. 7*: 418-421.

Kumar, S., Tamura, K. and **Nei, M.** (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*. *5*: 150-163.

Lederberg, J. (2000). Infectious history. *Science 288*: 287-293.

MGSP (2003). Microbial genome sequencing program. Available at http://www.nsf.gov/pubs/2003/nsf03526/nsf03526.htm. Accessed on July 1st, 2003.

Nelson, K.E. (2003). The future of microbial genomes. *Environ. Microbiol*. *5*: 1223-1225.

Nelson, K.E., Paulsen, I.T., Heidelberg, J.F. and **Fraser, C.M.** (2000). Status of genome projects for nonpathogenic bacteria and archaea. *Nat. Biotechnol. 18*: 1049-1054.

Relman, D.A. (1994). Phylogenetic diversity of microbial pathogens. In: *Molecular Genetics of Bacterial Pathogenesis* (Miller, V.L., Kaper, J.B., Portnoy, D.A. and Isberg, R.R., eds.). ASM Press, Washington, DC, USA, pp. 507-517.

Rodrìguez-Valera, F. (2004). Environmental genomics, the big picture? *FEMS Microbiol. Lett*. *231*: 153-158.

Saitou, N. and **Nei, M.** (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. *4*: 406-425.

Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R., Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C., Lemos, E.G., Lemos, M.V., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M., Madeira, H.M., Marino, C.L., Marques, M.V., Martins, E.A., Martins, E.M., Matsukuma, A.Y., Menck, C.F., Miracca, E.C., Miyaki, C.Y., Monteriro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L., Netto, L.E., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C., da Silva, A.M., da Silva, F.R., da Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsuhako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and **Setubal, J.C.** (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature 406*: 151-157.

Simpson, A.J.G., Camargo, A.A., Ferro, J.A., Parra, J. and **Vasconcelos, A.T.** (2004a). Coordinated, network-based research as a strategic component of science in Brazil. *Genet. Mol. Res. 3*: 18-25.

Simpson, A.J.G., Parra, J.C. and **Perez, J.F.** (2004b). Projetos genoma no Brasil: revolução na biotecnologia brasileira. In: *Genômica* (Mir, L., ed.). Atheneu, São Paulo, SP, Brazil, pp. 1091-1100.

Thompson, J.D., Higgins, D.G. and **Gibson, T.J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*: 4673-4680.

Thomson, N., Sebaihia, M., Cerdeño-Tárraga, A., Bentley, S., Crossman, L. and **Parkhill, J.** (2003). The value of comparison. *Nat. Rev. Microbiol. 1*: 11-12.

U.S. DOE Microbial Genome (2003). Microbial genome program. Available at http://www.ornl.gov/sci/microbialgenomes/mgp.shtml. Accessed on June 29, 2004.