Collaborative research networks work

Anamaria A. Camargo and Andrew J.G. Simpson

Ludwig Institute for Cancer Research, New York, New York, USA

Brazil was heralded for completion of the first genome sequence of a plant pathogen following the development of a virtual research center — a collaborative network of laboratories throughout the state of São Paulo, drawing on the expertise of a dispersed and diverse scientific community and on investment from both the government and the private sector. Strategies key to the success of this model are discussed here in the context of continuing collaborative scientific endeavors in both developed and developing countries.

J. Clin. Invest. **112**:468–471 (2003). doi:10.1172/JCI200319520.

Biomedical research can be approached in a piecemeal manner or attacked in large swaths. In molecular genetics, for example, genes can be identified, cloned, sequenced, and functionally characterized one at a time, or this task can be simultaneously undertaken for all the genes in a particular organism's genome. The latter approach is known as genomics, which encompasses the generation, analysis, and utilization of global sets of genetic data. The individual-gene approach fits perfectly within the traditional structural organization of scientific research, but genomics, especially when applied to complex organisms, does not. For example, single-gene cloning projects can be assigned to individual students or postdoctoral fellows within a research group, but genome projects are more complex and challenging and must be approached quite differently.

Address correspondence to: Andrew J.G. Simpson, Ludwig Institute for Cancer Research, 605 Third Avenue, New York, New York 10158, USA.
Phone: (212) 450-1556; Fax: (212) 450-1555; E-mail: asimpson@licr.org.

Conflict of interest: The authors have declared that no conflict of interest exists.

Nonstandard abbreviations used: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); Organization for Nucleotide Sequencing and Analysis (ONSA); expressed sequence tag (EST).

Networks: Brazil's strategy for genome sequencing

The most commonly adopted strategy for genomics, and in particular genome sequencing, has been to create large, stand-apart genome centers where the means to generate a vast amount of data are concentrated, recognizing that this kind of science has no real place in the average traditional research laboratory (1). In Brazil, however, a developing country with a reasonably well developed scientific infrastructure and with aspirations to integrate science and technology into its further economic growth, a different strategy was used — that of the collaborative research network (1, 2). In this approach, the necessary hardware, consumables, financing, and expertise are distributed within the existing scientific community, constituting a virtual research center that is, in reality, a large number of well-equipped laboratories linked by the Internet.

Such networks can be set up rapidly and can strengthen the general scientific community by disseminating financing, novel equipment, and expertise. Moreover, they can undertake complex and socially relevant projects that would be impossible within the traditional model of modern science organized around an individual researcher and his or her immediate colleagues; thus they can attain a higher level of scientific achievement. This network structure may not only enable developing countries like Brazil to move into more challenging areas of research but is also germane to the international community.

Brazil's initial genome sequencing project: *Xylella fastidiosa*

It was in early 1997 that the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the governmental agency responsible for the support of scientific research in the state of São Paulo, first contemplated an organized genome project (2). This was done with the view, not that Brazil urgently needed the genome sequence of any particular organism, but rather that expertise in this important field would greatly benefit the research community. It was also perceived that increasing expertise in genomics should simultaneously increase expertise in more basic molecular genetics, which was poorly represented relative to other areas of biology. In May of that year a plan of action was proposed by FAPESP, in September the deadlines for participant applications were defined, and by the end of the year the practical work had begun.

The decision made by FAPESP was that a project would be funded that was aimed at enabling the community to become acquainted with the complexities of genomics by generating fundamental data in the form of a complete, computer-annotated sequence of a bacterial genome. At that time, this was a significant challenge, since fewer than ten such projects had been completed anywhere in the world. Moreover, none of the participants (including the coordinators) had any experience in genome sequencing or hands-on bioinformatics. Most members of the network had never cloned a gene or sequenced DNA at all! FAPESP decided from the start that the project would be undertaken by a network of laboratories,

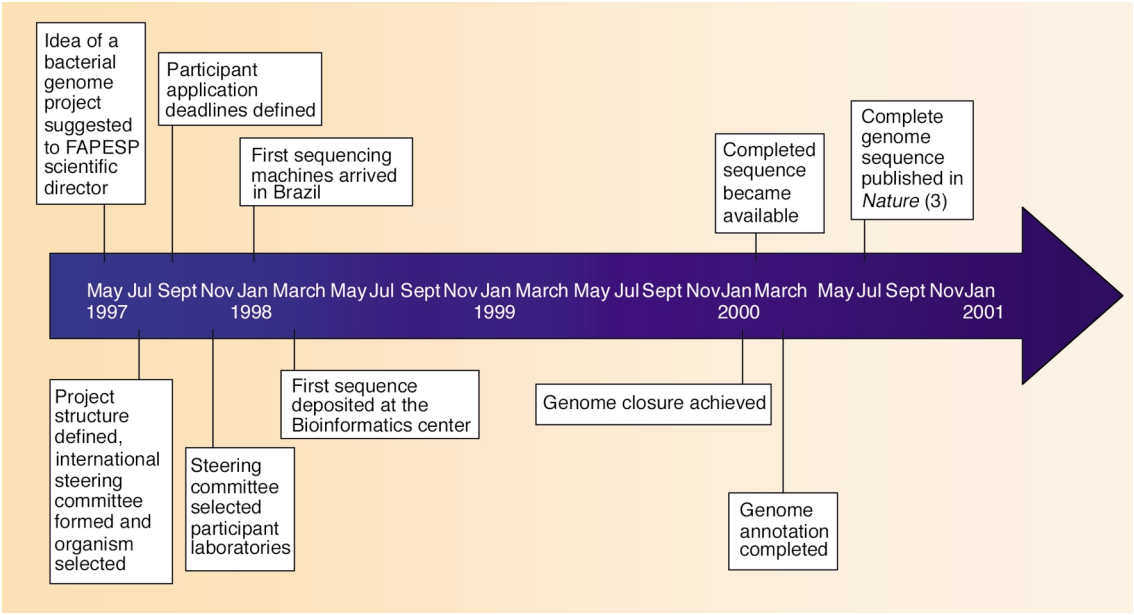

Figure 1
Chronological landmarks of the *X. fastidiosa* genome sequencing project.

rather than waiting for the construction of a dedicated genomic center. In fact, it was the genome to be sequenced that was determined last. As FAPESP, the agency that both conceived of and underwrote the initiative, receives its funds directly from taxpayers in a society with many economic challenges, it was crucial that the investment be perceived as a tool to raise not only the scientific but also the economic competitiveness of Brazil. Thus, the organism *Xylella fastidiosa* was selected. This insect-borne phytopathogen causes citrus variegated chlorosis, a disease that primarily affects oranges and results in significant economic loss to the state of São Paulo.

Overnight, the *X. fastidiosa* genome project became the largest and most widely known scientific project in Brazil, receiving financial support of approximately US$12 million and involving laboratories spread throughout the state of São Paulo. The network, termed the Organization for Nucleotide Sequencing and Analysis (ONSA), was composed of one coordination laboratory, one bioinformatics laboratory, and 30 sequencing centers functionally linked by the Internet with a central website (http://watson.fapesp.br/onsa/Genoma3.htm). Despite the group's lack of experience and the complexity of its task, the project was completed ahead of schedule (Figure 1), and the annotated sequence of the *X. fastidiosa* genome was published in a cover story in *Nature* in July 2000 (3). Brazil thus became the first country outside the US, Europe, and Japan to complete the sequencing of a bacterial genome. As a result, the project received wide attention from the national and international media (for a list of related news articles see http://watson.fapesp.br/imprensa/press.htm). *The Economist* began an article with "Samba, football and . . . genomics. The list of things for which Brazil is renowned has suddenly got longer" (4).

Consolidation of the original network

Inspired by the success of the *X. fastidiosa* genome project and by very positive responses from the scientific, financial, and commercial communities, ONSA immediately embarked on several other sequencing projects, including (a) the FAPESP/LICR Human Cancer Genome Project (jointly run and funded with the Ludwig Institute for Cancer Research), which has to date generated and deposited in GenBank almost 1 million DNA sequences of transcripts derived from human tumors and normal tissues (5); (b) the Sugar Cane EST Project (SUCEST), which has produced over 300,000 expressed sequence tags (ESTs) from sugar cane; and (c) the *Schistosoma mansoni* genome project, which has generated 160,000 ESTs from this important parasitic trematode. Thereafter, several other bacterial genome sequencing projects were also launched, many of which have already been completed and published (Table 1). Most gratifyingly, a number of private industries and non-Brazilian agencies came forward to join with FAPESP in funding the novel genome projects, with total investments approaching US$10 million. The disparate projects were upgraded into a fully fledged state research program, with a further US$15 million set aside by FAPESP for investment. As a result of this program, a total of 65 laboratories are now equipped for high-throughput DNA sequencing, and over 450 researchers have been trained to generate and, more importantly, to explore whole-genome data. Six major papers have been published (with several more submitted at the time of writing), over 1.5 million ESTs have been generated from various organisms, and over 20 Mb of high-quality, annotated bacterial genome sequence has been produced (1, 6–10). Furthermore, two start-up biotechnology companies have now been founded by

former ONSA members and are currently hiring many of the young researchers trained. Thus, we believe that, by any criteria, the initiative to develop genomics via collaborative research networks in the state of São Paulo must be considered a success.

The establishment of a nationwide genome network

In addition to the ONSA network, which operates in the relatively restricted area of the state of São Paulo (about the size of France), a distinct National Sequencing Network (BRGene; http://www.lncc.br/~brgene/) was also established under the auspices of the Ministério da Ciência e Tecnologia and the Conselho Nacional de Desenvolvimento Científico e Tecnológico. The national network was inaugurated in 2000 and incorporated 25 sequencing laboratories from as far apart as Manaus and Belém in the Amazon, and Porto Alegre near the borders of Uruguay. BRGene has now completed the sequencing of the genome of *Chromobacterium violaceum*, a free-living bacteria that is very abundant in the waters of the Rio Negro in the Amazon basin, and that is potentially important to the biotechnology industry. At present, the network is also in the final sequencing phase for the genome of *Mycoplasma synoviae*, a chicken pathogen. It is notable that the same spirit of enthusiastic collaboration that was the hallmark of ONSA also emerged within BRGene, even though distance allowed only infrequent meetings of the group as a whole and almost all interactions occurred via e-mail and the Internet. In addition to this nationwide initiative, further regional networks are currently undertaking both small-genome sequencing and EST projects (Table 1).

We have found that the greatest strength of sequencing networks comes from the interaction of the large number of scientists involved. For every sequencing machine there is at least one established scientist together with students and technicians (typically we funded the hiring of at least one technician per machine to do the daily work of generating data). Thus, extensive expertise is available for the analysis and annotation of the genome data, and also for the solution of the inevitable technical problems that occur usually during the final stages of a project. The network also contributes strongly to the community in which it is established. This contribution is

Table 1
Summary of genome sequencing projects undertaken by Brazilian genome networks

Date	Project/Organism	Related disease or economic effect	Network	Status
1997	*Xylella fastidiosa*	Citrus variegated chlorosis	ONSA, São Paulo	Completed (3)
1998	Human Cancer Genome Project	Tumors of high incidence in Brazil	ONSA, São Paulo	Completed (6–8)
	SUCEST – sugar cane	NA	ONSA, São Paulo	Completed[A]
1999	*Xanthomonas citri*	Citrus bacterial canker	ONSA, São Paulo	Completed (9)
	Xanthomonas campestris	Citrus bacterial canker	ONSA, São Paulo	Completed (9)
2000	Virus Genome Project	HIV, HCV, HPV	ONSA, São Paulo	In progress
	Chromobacterium violaceum	Free-living organism from the Amazon river with biotechnological potential	National Sequencing Network BRGene	Completed[A]
	Leifsonia xyli	Ratoon stunting disease of sugar cane	ONSA, São Paulo	Completed[A]
2001	*Schistosoma mansoni*	Schistosomiasis	ONSA, São Paulo; and regional network, Minas Gerais	Completed[A]
	Paracoccidioides brasiliensis	Paracoccidioidomycosis	Regional network, central/west	In progress
	Leptospira interrogans	Leptospirosis	ONSA, São Paulo	In progress
	Xylella fastidiosa (grape isolate)	Pierce disease of grapevine	ONSA, São Paulo	Completed (10)
	Leishmania chagasi	Leishmaniasis	Regional network, northeast	In progress
	Herbaspirillum seropedicae	Free-living organism related to N_2 fixation	Regional network, Paraná	In progress
	Gluconacetobacter diazotrophicus	Endophyte isolated from sugar cane and related to N_2 fixation	Regional network, Rio de Janeiro	In progress
2002	Eucalyptus	Product of major importance to the Brazilian economy (essential oils and paper production)	ONSA, São Paulo	Completed[A]
	Coffea arabica	Product of major importance to the Brazilian economy (coffee)	ONSA, São Paulo	In progress
	Paullinia cupana	Guaraná tree that is a creeping shrub native to theAmazon and is used as the basis of a popular soft drink	Regional network, north	In progress
	Mycoplasma synoviae	Poultry disease	National Sequencing Network BRGene	In progress
	Mycoplasma hypopneumoniae	Mycoplasmal pneumonia	Regional network, south	In progress
2003	*Bos taurus*	Product of major importance to the Brazilian economy (cattle breed)	ONSA, São Paulo	In progress
	Litopenaeus vannamei	Product of major importance to the Brazilian economy (white shrimp)	Regional network, various states	In progress

HCV, hepatitis C virus; HPV, human papilloma virus. [A]Related manuscripts are either submitted or in preparation.

achieved through the widespread availability of the sequencing equipment, which can be employed in whatever other projects the participating groups are involved in, and through the practical experience gained from dealing with large DNA data sets. Perhaps more important, however, is the experience of working within a coordinated, collaborative network and the realization of the power that comes from collective endeavors.

Key factors in the success of genome networks in Brazil

In retrospect, several factors were crucial to the success of the sequencing networks. First, many of the participants were young and not yet well established in a specific line of research; thus, the limits to their collaborative spirit were few. To attract young, motivated groups, the networks refrained from requiring either previous experience in genomics or full-time dedication. Also, to stimulate groups without a solid background in genetics to join the project, funding levels were set at a generous amount so that all participating laboratories could purchase an automated sequencer and improve their laboratories. A second key factor was the commitment of governmental research-funding agencies. Most Brazilian funding agencies are autonomous and have regular budgets for investment in training and research infrastructure. Through these agencies, hundreds of MD and PhD fellowships were made available for genomics research. Third, it was crucial for the projects to have outside steering committees, which selected and evaluated the groups, monitored progress, and were prepared to recommend removal of groups, though no group has been removed to date.

The speed and quality of work achieved by all the networks that were set up in Brazil probably surprised even the participants. Fundamentally, this impressive showing demonstrates that, when appropriately nurtured and supported, group-based research is tremendously effective. Without the anxiety engendered by excessive competition for both funding and experienced staff, the various strengths, talents, and areas of expertise within a group can complement one another for the good of all. Subsequent to the projects we have described, several other collaborative research networks have been established in Brazil. These include a functional genomics network to study phytopathogens, an initiative to complete the coding sequences of expressed human genes, a viral-genetics network to monitor the genotype of endemic viral pathogens, and, in a completely different field, a huge effort to record and register the complete range of fauna and flora in the state of São Paulo (http://www.biota.org.br/). In Brazil, the research network is now a firmly established model, throughout the country and in many different fields. This represents a major change in the structure of Brazilian research.

The worldwide adoption of the network structure

In the Northern Hemisphere, we have come to feel that biomedical science should be an essentially individualistic activity, reaching its finest form when the greatest minds constantly vie for funds and prestige. However, the problems we face, particularly in the genomics era, are too complex to be tackled effectively within the traditional research paradigm based on standalone laboratories. This is especially clear in the crucially important quest to move science from mere discovery to concrete contributions to society. Science must, in the long term, add to the richness of human life, reduce human suffering, and stimulate economic growth if it is to continue to receive the long-term support from governmental and private agencies that scientists have come to expect. For example, in our own field, an individual researcher may identify a gene that contributes to the molecular distortion of an epithelial cell, leading to its malignant transformation. However, it takes large coordinated teams to then discover effective strategies for inhibiting the gene product, to produce the required inhibiting substance under appropriate conditions, to undertake clinical trials, and ultimately to make the drug available to patients. Thus, the tendency toward individualistic approaches may contribute to the paradox of modern cancer research — our tremendous progress in understanding the disease but modest progress in improving treatment. The Brazilian demonstration of the efficacy of disseminated academic research networks that are focused on specific questions provides an alternative model for the organization of science that could be used to enhance its practical contributions.

We strongly advocate the adoption of network structures, similar to that which we developed in Brazil, in other developing countries that are embracing the challenge of sustained investment in science and technology. We also urge that collaborative networks be more vigorously pursued in developed countries, so that academic research might become more multidimensional and capable of tackling complex problems. We realize that this would depend on some fundamental changes in the manner in which individual scientists are evaluated and research funds dispersed. Nevertheless, we feel that we have amply demonstrated that research networks work, and that their implementation within the wealthy research communities of the US, Europe, and Japan could begin a whole new chapter in modern biomedical research, enhancing the translation of improved knowledge into concrete contributions to society.

1. Simpson, A.J. 2001. Genome sequencing networks. *Nat. Rev. Genet.* **2**:979–983.
2. Simpson, A.J., and Perez, J.F. 1998. ONSA, the Sao Paulo Virtual Genomics Institute. Organization for Nucleotide Sequencing and Analysis. *Nat. Biotechnol.* **16**:795–796.
3. Simpson, A.J., et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature.* **13**:151–157.
4. Collins, P. 2000. Fruits of co-operation. *The Economist*, 22–29 July.
5. Bonalume Neto, R. 1999. Brazilian scientists team up for cancer genome project. *Nature.* **398**:450.
6. Dias Neto, E., et al. 2000. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* **97**:3491–3496.
7. de Souza, S.J., et al. 2000. Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* **97**:12690–12693.
8. Camargo, A.A., et al. 2001. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* **98**:12103–12108.
9. da Silva, A.C., et al. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature.* **417**:459–463.
10. Van Sluys, M.A., et al. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J. Bacteriol.* **185**:1018–1026.