ISSN 1018-5593

European Commission

An assessment of progress in Human Genome Programmes worldwide

(A support study for the evaluation of the EC Human Genome Analysis Programme)

Research evaluation

Report
EUR 15412 EN

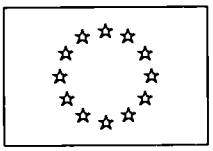
European Commission

An assessment of progress in Human Genome Programmes worldwide

(A support study for the evaluation of the EC Human Genome Analysis Programme)

Authors: **B.R. Jordan**

1994



Published by the

EUROPEAN COMMISSION

DIRECTORATE GENERAL XIII
Telecommunications, Information Market and Exploitation of Research
L-2920 Luxembourg

LEGAL NOTICE

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information

Cataloguing data can be found at the end of this publication

Luxembourg: Office for Official Publications of the European Communities, 1994

ISBN 92-826-8226-9

Printed in Belgium

TABLE OF CONTENTS

FOREWORD

A realistic evaluation of progress under the EEC Human Genome analysis programme must take into account how similar projects have developed elsewhere, so as to highlight in a comparative way strengths and weaknesses of the European enterprise. First-hand knowledge of actual progress in other countries is here essential, since the quality and efficiency of research cannot be accurately assessed from documents alone, especially in such a rapidly developing and evolving field. I provide here such an assessment, based on direct contacts and on-site visits to a number of genome centres world-wide. The groundwork for this was laid by a one-year study I performed in 1991, during which I visited in depth approximately 100 major genome laboratories all over the world. This has been updated through further visits and discussions with a number of scientists in these countries, and the resulting document is a personal, probably biased but at least candid account based on both facts and impressions. Only references to recent publications (1993, 1992, and in a few cases 1991) are given.

Bertrand R. Jordan, March 1993

I. INTRODUCTION

Serious discussion about launching a human genome sequencing programme began in the USA in the mid-eighties. By 1986-87, the Department of Energy had committed several tens of millions of dollars to this effort, shortly followed by NIH. Over succeeding years funding grew to a total of close to 200 million dollars per year, while goals were redefined to focus on genetic and physical mapping with a smaller component devoted to DNA sequencing. Other countries followed suit, notably Japan (with, however, four or five different programs whose coordination took time to become effective) and Great Britain with a well-integrated Human Genome Mapping Project. The Soviet Union started a project in 1989, although efforts were hampered by lack of hard currency, communication problems and by the subsequent disintegration of the Union; France has also been present in the field with both an official genome programme and a very successful project run by non-governmental organizations (CEPH, AFM and Généthon). Each of these non-US national projects has its distinctive flavour, and most appear more strongly focused on functional genes than the US enterprise - although the latter turned out to be the first to set up on a large scale the very effective partial cDNA sequencing approach. Relative strengths of these countries in the field, in 1990-1991, can be summarized as shown in Figure 1; the different indicators used give roughly equivalent results, with the USA a clear leader, Great Britain a strong second and France a more distant third, slightly ahead of Japan; on a continental basis the EEC (Britain included) is almost equivalent to the US. Recent successes in Europe (sequencing yeast chromosome III) and notably in France (the Généthon genetic and physical mapping programmes) are increasing the share of Europe in this field. Figure 2 indicates approximate funding levels in 1993 for some of these national projects.

In spite of the short time since Genome programmes were initiated, it is abundantly clear that they have numerous consequences. They push recombinant DNA technology towards more powerful and efficient procedures, which can benefit many other fields of research; they have important potential commercial implications (in terms of diagnosis and therapy), as evidenced by the - now almost settled? - controversy on patenting of partial cDNA sequences; and they raise socioeconomic as well as ethical problems that are not easy to address.

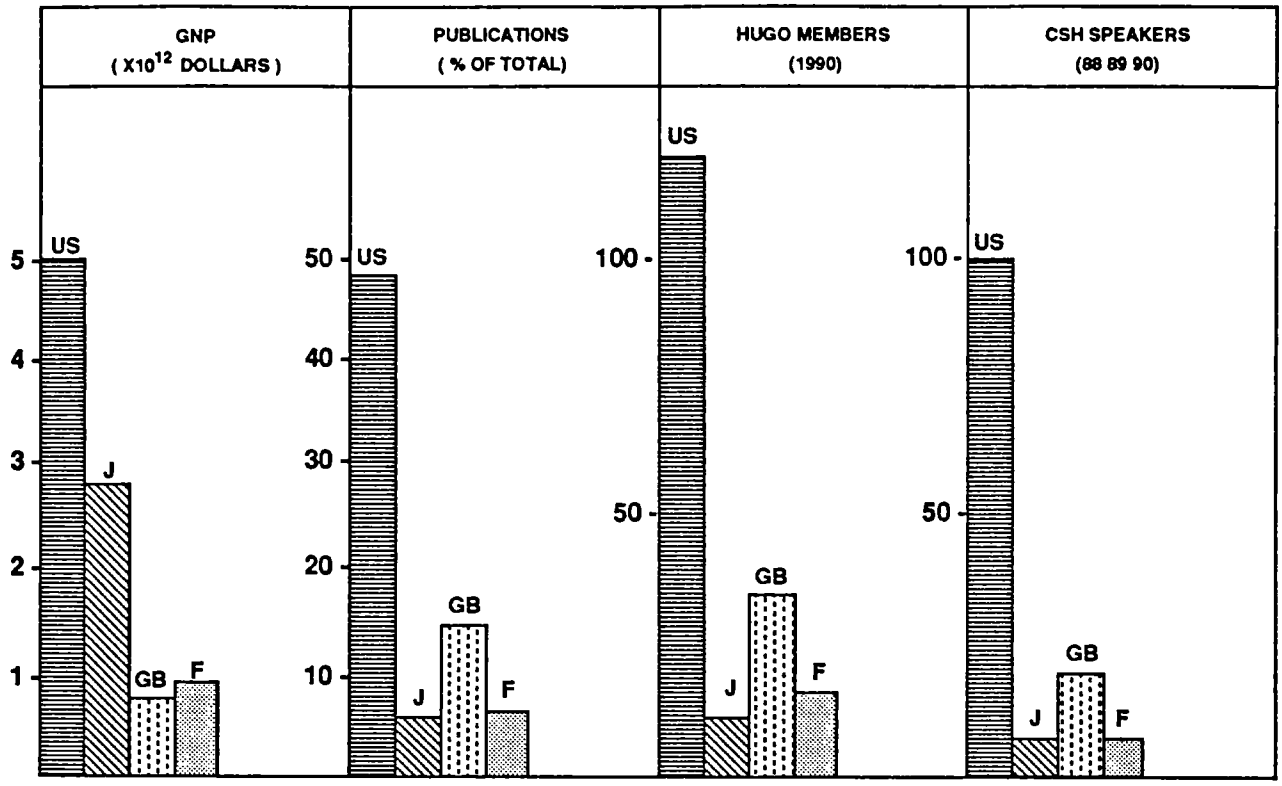


FIGURE 1

Relative strengths of the USA, Japan, Great Britain and France. From left to right, GNP in 10^{12} US dollars; share of human genome-related publications as evaluated by the ESF in 1991; number of members elected to HUGO; cumulated number of speakers at the 1988, 1989 and 1990 Cold Spring Harbor Genome Mapping and Sequencing meeting. It is clear that the three indicators relative to genome research give very consistent results.

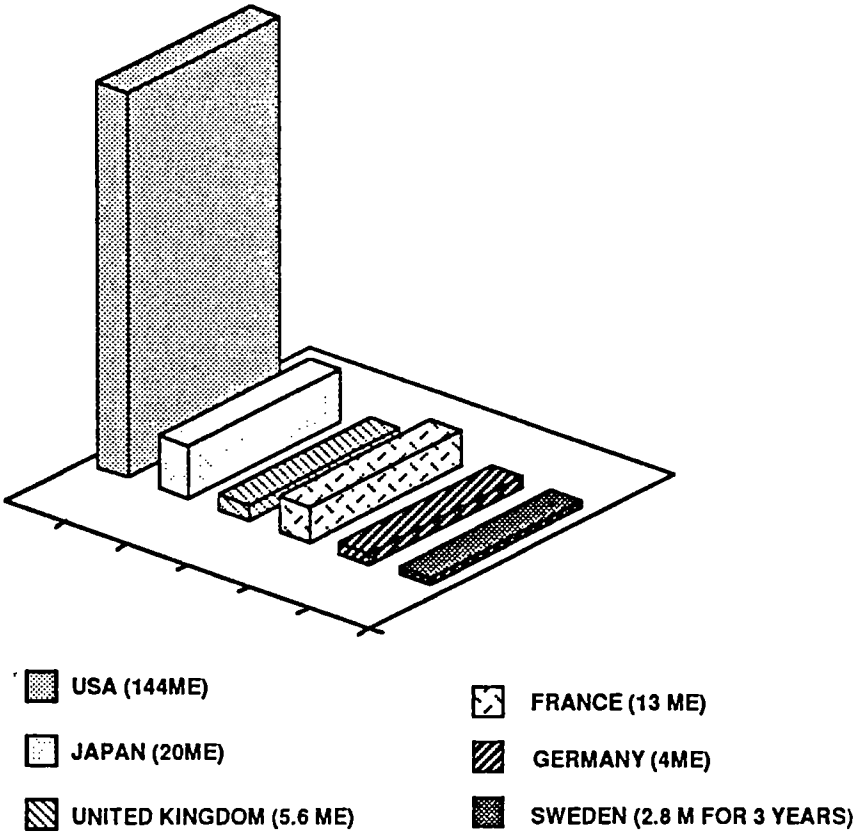


FIGURE 2

Yearly funding in 1993 (in million ECUs) for some major government genome programmes. These figures provide an order of magnitude, but they must be taken with caution since accounting rules differ: for example, salaries may or may not be included in the amounts indicated.

II. REVIEW OF MAJOR NATIONAL PROGRAMMES

II A. UNITED STATES

II A 1. Current plan and major participants

The US genome programme has suffered a period of uncertainty following the resignation of Jim Watson, head of the NIH component, over disagreements with the director of NIH on, in particular, patenting of cDNA sequences. An interim director (Michael Gottesman) had been appointed, but clearly future directions will depend very much on Watson's successor, Francis Collins, whose appointment has recently been announced (Thompson, 1993). The present five year plan is now beginning its third year. Its major goals are: completion of a 2 to 5 centimorgan genetic map in which each marker is identified by an STS; construction of STS-based physical maps with 100 kilobase average spacing and progress towards genome coverage with large (>2 megabase) contigs; improvement of DNA sequencing methods and determination of approximately 10 megabases of sequence in large stretches; derivation of a detailed genetic map of the mouse as well as sequence data and physical maps; and, finally, development of software, databases and algorithms.

Some of these goals now appear too modest in the face of recent developments, a point we will discuss in the general conclusion.

II A 2. Assessment of progress

A caveat

The following comments, as well as those I will provide for other countries, are both qualitative and personal. While based on site visits to the centres quoted and on extended discussions, they are also highly subjective and their conclusions are certainly debatable. I feel however that they represent the most valuable contribution I can make since they give a first-hand feeling about ongoing research as experienced by a scientist in the same field. They should usefully complement objective data on results and publications. Any value judgments made are entirely my responsibility.

The DOE juggernaut

DOE runs three major genome centres, at Lawrence Livermore, Los Alamos and Lawrence Berkeley. The latter has been without a director for a long time, following Charlie Cantor's "sideways promotion" at fall 1990; the acting director, Jasper Rine, has now been confirmed, but many scientists left or drifted apart during the long interim period, and I do not feel the centre has really resumed effective operations - especially since its major topic, chromosome 21, has been effectively taken up by others, in particular Daniel Cohen's group at CEPH and Généthon (Chumakov et al, 1992 b).

The other centres, Los Alamos and Lawrence Livermore, both have responsibility for providing chromosome-specific libraries, which they have done on the whole with success, moving progressively from small insert to large insert lambda libraries and now to cosmids; these libraries have been tremendously useful to the whole community. Efforts at Los Alamos to extend this approach to YAC libraries derived from flow-sorted chromosomes appeared promising but are superseded by the more efficient method of "extracting" chromosome-specific collections from whole genome YAC libraries (Chumakov et al, 1992 a; Ross et al,

1992). The two centres have also undertaken the construction of whole chromosome physical maps (respectively chromosomes 16 and 19) through cosmid contigs assembled by two variants of a fingerprinting method. This heroic effort, started just when YACs were appearing, has achieved relatively good but still incomplete coverage (Stallings et al, 1992; Trask et al, 1992): at this time the cosmid contigs span only slightly more than half of chromosome 16 and 19. The physical map is now being completed with the help of YACs; there is much to be said, in fact, for a physical map based on both YACs and cosmids, since the latter are usually the most useful for any group interested in a particular region. Both centres are run in a professional way, with good organization and stable personnel; this is particularly the case at Lawrence Livermore. Both suffer from lack of sufficient flexibility, limited contact with other laboratories (few post-docs and no students) and from the fact that serious physical mapping may have been started too early so that major technical options were frozen based on technology which quickly became obsolete.

Wholly YAC-based physical efforts have now proved very effective (Chumakov et al, 1992 b; Bellanne-Chantelot et al, 1992; Foote et al, 1992; Green et al, 1991 a) and reevaluation of the DOE programmes becomes necessary. DOE contractor's meetings, attended by all groups receiving DOE funding, gave me a cross section of this effort which is of good quality and has a strong technological orientation. On the whole my impression of DOE (apart from the Lawrence Berkeley centre, beset by serious political problems) is that of a technically competent operation, with reasonable goals and qualified personnel, but lacking some of the flexibility and brilliance necessary in such a rapidly evolving field.

The NIH kaleidoscope

NIH is much more difficult to characterize because of its great diversity. Genome centres funded by this agency range from heavyweight physical mapping groups operating almost exclusively with technicians (e.g. Glen Evans, Salk Institute, for chromosome 11) or with a more traditional mix of post- docs, students and technicians (e.g. David Schlessinger and Maynard Olson, Washington University of Saint-Louis, for chromosomes 7 and X) to departments mostly involved in disease-oriented research (Tom Caskey, Baylor Institute at Houston, with a partial focus on chromosomes 17 and X).

In the latter case the genome grant is used to set up very effective "core services" (for YAC screening, sequencing, informatics...) used by the research groups in the departments as well - in principle - as by outside users. In contrast, the operation at Saint-Louis is completely devoted to developing a complete physical map of the two chromosomes in question, using a very robust STS-based contig building approach. This programme is pursued with great dedication and a commitment to avoid sidestepping into lateral projects - even though appealing opportunities may arise. Needless to say, there is a fair amount of tension between groups which are funded by the same programme but pursue such widely differing objectives.

On the whole work carried out in the NIH framework appears more trendy, more open to new concepts and more innovative than the research performed under DOE. On the other hand, it seems often less well organized; it benefits from the presence of good students and post-docs but suffers from the instability of this personnel which lacks dedication to long-term, systematical objectives and tends to rush into "exciting" sidelines.

II A 3. General comments (scientific)

Genetic maps

The human genetic map appears to be progressing well, after a period of disillusionment and scepticism in 1989-1990. The shift from RFLPs to CA or other simple sequence repeats is general, although a number of technical problems remain: on one hand, the tendency of these repeats to yield multiple bands complicates scoring of the results, and difficulties persist in streamlining the procedure and developing automatic data capture. Sperm mapping does not appear to have caught on as an effective fine-scale mapping method, although new developments appear very promising (Zhang et al, 1992). The scientists involved appear confident that a map with an average marker density of one marker per two centimorgans (not all ordered) can be produced by the end of the current 5 year plan (1995).

A "NIH/CEPH" comprehensive linkage map including 1416 loci has been obtained by a constellation of collaborating laboratories using the CEPH panel of families (NIH/CEPH collaborative mapping group, 1992). In spite of the magnitude of this achievement, the usefulness of this map is limited by the fact that most loci correspond to moderately informative markers, RFLPs rather than microsatellites. In addition, it appears likely that genetic distances are somewhat overevaluated because the data (collected by many groups of non-uniform expertise) includes some typing errors - that tend to increase the apparent recombination rate and thus the map distances. In both these respects the genetic map recently obtained at Généthon by Jean Weissenbach's group (Weissenbach et al, 1992) appears superior as it is solely based on microsatellites and relies on data of uniform (and high) quality.

Physical maps

The advancement of physical mapping appears more uneven. As indicated above, the early start of DOE cosmid contig efforts means that much work has been done using a technology that is now largely obsolete; on the other hand these contigs, with the addition of selective YAC coverage, were the first to show that a physical map of a whole human chromosome was feasible. More recent attempts such as those implemented in Saint-Louis have every chance of success if they are pursued to the end; but these approaches are superseded by the much more efficient YAC fingerprinting techniques used, in particular, at Généthon in France (Bellane-Chantelot et al, 1992). The successful mapping project on the Y chromosome (Foote et al, 1992) has been set up for the most part outside the "official" genome projects. The whole programme will obviously have to be reevaluated.

Cytogenetics

Cytogenetics, once thought to be almost a dying art, has enjoyed an astonishing revival thanks to the development of very effective non-radioactive in situ hybridization techniques. Metaphase in situ has become a highly efficient and quick way of mapping cosmids and YACs to a precision of one chromosome band; interphase in situ with multicolour technology can provide ordering of clones separated by a hundred kilobases or even less. In addition these developments have direct application in clinical genetics, using chromosome painting or two-colour interphase in situ for the detection of translocations. The two world leaders in the field are probably David Ward (Yale) and Barbara Trask (Lawrence Livermore), but the technology has now spread to a large number of laboratories all over the USA and is extremely effective.

Sequencing

Large scale sequencing of genomic DNA is not a major objective of the present DOE and NIH programmes - not unless cost can be brought down to much less than one US dollar a base: the current 5 year plan envisions a total of 10 megabases of human DNA and twice as much for model organisms. Recently completed genomic segments include a 106 kilobase segment from chromosome 19 (Martin-Gallardo et al, 1992), a 95 kilobase region containing part of the mouse T cell receptor alpha/delta locus at Caltech (Wilson et al, 1992) and the Nematode sequencing work carried out in Saint-Louis in collaborating with the British group (Sulston et al, 1992).

New technologies are still under investigation, although enthusiasm for direct sequence reading by STM or AFM microscopy has very much died down. The single-molecule detection system proposed by R. Keller (Los Alamos) is nearing the feasibility demonstration stage (and is being developed in collaboration with Gibco-BRL) but practical applications are probably still far in the future. More evolutionary developments such as ultrathin gels, capillary electrophoresis or the use of new enzymes and reporter groups may significantly improve the performance of present technology. DNA sequencing by hybridization, as originally proposed by soviet (Pevzner et al, 1991) and Yugoslav scientists, is now considered a serious possibility (Cantor et al, 1992 a); it is mainly explored, in the USA, by Crenjakov's group at Argonne (DOE).

cDNA sequencing has been taken up very effectively by Craig Venter's group (although not with genome funding) and others (Adams et al, 1991; Adams et al, 1992; Khan et al, 1992), and attempts at patenting the partial cDNA sequences have caused a well-known national and international controversy. This approach may become a major focus of the US genome effort; Jim Watson was opposed to this choice, but his successor may well hold a different opinion.

II A 4. General comments (organizational)

DOE vs NIH

As indicated above, the two agencies have definitely different styles. They are to a certain extent complementary, but my impression is that this complementarity is not sufficiently exploited. There is interagency coordination at the higher levels, but day-to-day, lab-to-lab contact between DOE and NIH centres is not very intensive, in part because the major DOE centres are in outlying locations to which, in addition, access is sometimes restricted for security reasons. Thus contacts between groups are not sufficient, and some potential synergisms are not put into effect.

Manpower

Genome centres find themselves with the dilemma of requiring some very qualified scientists, if only for definition of experimental strategy and for trouble-shooting procedures, and yet assigning to them a majority of systematic and rather uninspiring tasks. Various centres try to tackle this contradiction in different ways, but the balance appears difficult to find. The status of many technicians in the USA, i.e. the fact that they are often former students who after a couple of years as technician will resume their studies and work for a PhD is helpful here, since it provides a flow of young technicians - in contrast with the usual situation in Europe where being a technician is a lifelong occupation.

Informatics and instrumentation

Although these two fields receive much attention and sizeable funding they continue to be rather badly integrated with actual everyday laboratory work. There are exceptions, in particular in DOE centres, but generally the gap between biologists and computer people is not completely bridged. Instrumentation is another weak point, in spite of much-publicized efforts: nowhere, for example, is a colony-picking robot actually in everyday laboratory use, and robotics in Genome centres rarely go beyond one or two pipeting robots, usually Beckman Biomek or TECAM machines. Some DOE centres (notably Lawrence Livermore) have had more success in breaking down cultural barriers between biologists, software experts and robotics engineers, possibly because the stable employment structure of genome centers in this agency makes possible long-term contact between specialists from different fields.

II A 5. Recent developments

The most important recent result from the US is probably the physical map of the Y chromosome recently published by David Page's group (Foote et al, 1992), as well as rapid progress on the mouse genome in Eric Lander's group. The other chromosome-specific mapping projects will have to be reevaluated; in fact a shift to whole-genome mapping has already taken place with the announcement of a large centre in Cambridge, Massachusetts, headed by Eric Lander and concentrating on large-scale STS content physical mapping - in collaboration with Daniel Cohen's group and using their "MegaYacs". This is funded at the level of 24 million US dollars for five years. Another large centre in Iowa (15 million dollars for four years) will concentrate on generation of large numbers of microsatellite markers and genetic mapping (Anderson, 1992 b; Roberts, 1992). Thus the US programme - at least its NIH component - is moving away from its previous single-chromosome approach. Meanwhile Craig Venter has resigned his NIH post in July to head up a new "Institute for Genomic Research" (Anderson, 1992 a), funded by a corporate partner, Human Genome Sciences, Inc. This new institute has ordered 50 DNA sequencers and will go for cDNA sequencing on a very large scale. The company will retain commercial rights to discoveries made in the Institute (which it funds to the tune of 70 million dollars), but results will be published in the normal way. Corporate interest in human genome research and its applications has grown substantially during the last year and a number of ventures are underway (Figure 3 and Anderson, 1993 a).

SOME NEW AND PLANNED GENOME-RELATED COMPANIES

Company Name	Research Focus	Main scientists	Main financial backing	Status
SEQ Ltd Priceton, NJ	DNA sequencing, technology development	Kevin Ulmer	Johnston Associates Inc.	Incorporated 1987, main funding 1992
Inoyte Pharmaceuticals Inc. Palo Alto, CA	cDNA sequencing	Randy Scott	Schroder Ventures Phoenix Partners	Incorporated 1991
(Operating under Darwin name; see below) Seattle, WA	cDNA sequencing	Leroy Hood	Frederic Bourke	Inc. 1992, planned merger with Darwin 1993
The Institute for Genomic Research Rockville, MD	cDNA sequencing	Craig Venter, Mark Adams, Chris Fields	Human Genome Sciences Inc.	Incorporated 1992
Myriad Genetics Inc. Salt Lake City, UT	Cancer genes	Mark Skolnick, Walter Gilbert	Eli Lilly & Co. Spencer Trask Inc.	Incorporated 1990 main funding 1992
Mercator Genetics Inc. San Francisco area	Disease genes	David Cox, Richard Myers, Dennis Drayna	Robertson Stephens & Co.	Incorporated 1992
Sequana Therapeutics Inc. San Diego area	Polygenic disease genes	Peter Goodfellow, Anthony Monaco, Hans Lehrach	Avalon Medical Partners D. Blech & Co.	Incorporated 1992
(Not yet decided) Cambridge, MA	Disease genes	Eric Lander, Daniel Cohen, Jeffrey Friedman	Mayfield Fund, Kleiner, Perkins, Caufield & Byers	Under discussion
Human Genome Sciences Inc. Rockville, MD	Disease genes	Craig Rosen William Haseltine	Health Care Investment Corp.	Incorporated 1992
Darwin Molecular Technologies Inc. Seattle, WA	Applied molecular evolution; cancer, inflammatory disease genes	Mark Pearson Gerald Joyce	George Rathmann Ronald Cape	Incorporated 1992
Nanotronics Inc. San Diego, CA	High-speed sequencers	Glen Evens Michael Heller	Bimdorf Biotechnology Dev., Enterprise Partners	Incorporated 1992
(Not yet decided) Long Island aera	High-speed sequencers	William Studier Thomas Marr	Long Island Venture Fund	Under discussion
Genomix Inc. San Francisco, CA	Long read-length sequencers	Thomas Brennan	Genentech Inc.	Inc. 1988, planned private placement, 1993
Genome Systems Inc. St Louis, MS	Gene library screening	David Smaller	Gold Biotechnology Inc.	Incorporated 1992

FIGURE 3

A list, compiled by Science (Anderson 1993 a) of commercial ventures in the US involving human genome research.

II B. JAPAN

II B 1. Current plan and major participants

There are several genome programmes in Japan, corresponding to the different agencies involved and funded for a total of approximately 20 million US dollars in 1992. The project of the Ministry of Education (Monbusho), headed by Kenichi Matsubara, a very well-respected scientist, is centred on fundamental research including cDNAs, general genetic and physical mapping (without focus on a particular chromosome), informatics, technical developments and sequencing of DNA from model organisms (Schizosaccharomyces pombe, Bacillus subtilis...). 1992 funding for Monbusho was around 6 million US dollars. The Science and Technology Agency (STA) has a more technological programme focussed on chromosomes 3, 11 and 21, funded at the level of 8 million; the ministry of Health has a smaller project (3 million) on genetic diseases. In addition the Ministry of Agriculture funds research on the rice genome (3 million) and the Ministry for International Trade and Industry (MITI) in involved in setting up sequencing institutes with industries and local governments. A committee headed by senator Mori has been set up since 1991 to coordinate these various efforts.

II B 2. Assessment of progress

STA is reminiscent of DOE

There are more than superficial analogies between the two agencies, since STA, like DOE, is a government agency primarily involved in energy research and general technological development. Its biology has therefore a fairly instrumental and systematic aspect - as is the case for DOE. Two main STA centres are involved in genome work: the RIKEN life sciences centre in Tsukuba, and a division within the National Institute of Radiological Sciences in Chiba (both close to Tokyo). RIKEN houses the "HUGA sequencing factory" (Endo et al, 1991) which is the outcome of Akiyoshi Wada's efforts, started more than ten years ago. This is a complete set of robots performing each step of (conventional) sequencing, linked into a system in which transport of samples from one machine to the next is taken care of more or less automatically. A very impressive and unique set up demonstrating the technological expertise of the Japanese and their capacity for long-term investment. Actual results from use of this set-up are awaited with interest: the theoretical throughput is more than 100,000 bases of raw sequence per 24 hours. Not surprisingly, teething troubles have hit HUGA, and it is not yet in actual production; it nevertheless represents a very interesting and unique effort.

The Chiba group is a more conventional cytogenetics laboratory which localizes cosmids by non isotopic in situ hybridization with good results and efficiency, as shown by the recent localization of nearly three hundred chromosome 3 cosmids (Takahashi et al, 1992). It is closely associated with the group of Yusuke Nakamura (formerly at Salt Lake City) who studies chromosomes 3 and 11 in the context of a private cancer institute, but with heavy funding from STA. This is a very impressive group which has recently scored some spectacular hits and is a very serious competitor for the best groups in the West. Overall STA appears well funded, highly technological and well-connected with industry, but sometimes not close enough to biologists - although this has changed, and STA grants are now being obtained even by university groups, a feat previously unheard of.

Monbusho is more academic

There are a number of excellent molecular biology groups in Japan, such as Honjo (Immunoglobulins), Okayama (cDNAs and the cell cycle), Yanagida (S. pombe) as well as Matsubara and others. The Monbusho programme attempts to bring together the expertise from these groups in a genome context but without setting too stringent criteria on the topic. Awareness and expertise in up to date technology has undoubtedly increased, for example in pulse field gel work and use of YACs, and the "soft" coordination promoted by Matsubara appears to be effective. Of particular interest is the tack taken by Japanese groups with regard to cDNA studies: they have opted for the construction of cDNA libraries reflecting as closely as possible the mRNA content of the corresponding tissue, i.e. the opposite direction to those who invest in "normalization" procedures. The plan is to sequence 1000 cDNA clones for each of the 200 basic cell types, and thus to obtain a sample of the genes expressed in these tissues as well as of their level of transcription (Okubo et al, 1992).

II B 3. General comments (scientific)

A late but effective start

Compared with the state of affairs three or four years ago, Japanese involvement in human genome research has undoubtedly taken off. This is quite visible in international meetings, where presentations from Japanese scientists in this field tended to be somewhat second rate, and are now often of excellent scientific and technical quality. Some of the work done in university laboratories is superb (e.g. Matsuda's coverage of the Ig region by YACs, in Honjo's group), while the level of STA research has definitely improved.

Emphasis on sequencing

There does not appear to be any systematic genetic mapping (except for the collaborative work involving Nakamura), and it is doubtful whether physical mapping of whole chromosomes will be seriously attempted (chromosome 21 was announced, but this in now essentially obsolete). On the other hand, sequencing seems to receive a lot of attention. cDNA sequencing projects (using the "signature" approach) are underway in several groups including Matsubara's; as indicated previously, the Japanese approach stresses the use of this method as a way to assess the expression patterns in different tissues. In addition, large-scale genomic sequencing (not necessarily of human DNA) is being prepared for not only by STA at RIKEN but in one or two other special-purpose institutes funded by industry, local governments and MITI. This may become quite a strong Japanese contribution to the genome project.

II B 4. General comments (organizational)

A special context

There are several reasons for Japan's modest position in genome research in the recent past. For cultural reasons (inherited diseases are looked upon as shameful) clinical genetics has remained underdeveloped, and has not served (as in other countries) as foundation and spur for genome research. In addition, strong disagreements between competing agencies have reduced the efficiency of research and delayed the initiation of proper genome programmes. However these difficulties appear to be largely overcome, and Japan has definitely got its act together, thanks in part to the influence of Kenichi Matsubara. The Japanese capacity for long-term investment (as for the HUGA sequencing factory) is also beginning to pay off.

Progress in instrumentation is particularly striking, and the Japanese industry may well market in the near future instruments such as budget-priced "personal sequencers" which could be a great hit.

II B 5. Recent developments

The 1993 budget includes healthy increases for genome research: total funding is around 3,000 million yen (i. e., close to 30 million US dollars) with 7.5 million (US dollars) to Monbusho, 10 to STA, 7.5 to the Ministry of Agriculture and Fisheries (rice genome) and 5 to the Ministry of Health.

Informatics and databases are still the subject of some discussions between Monbusho and STA, but a Genome centre largely dealing with informatics (directed by Kanehisa and funded by Monbusho) is being set up at the Institute of Medical Science (Tokyo University). STA has responsibility and should eventually - when some bureaucratic hurdles are cleared - get funds for organizing the Japanese GDB node. Reliable accounts indicate, on the other hand, that the HUGA sequencing factory has not reached actual operation and may never do so...

II C. GREAT BRITAIN

II C 1. Current plan and major participants

After extensive discussions between the two major players in this field, namely the Medical Research Council and Imperial Cancer Research Fund, the Human Genome Mapping Project was officially initiated in 1989, with separate funding of 11 million pounds for three years, thereafter to be continued at the level of 4 million per year (but at that later stage incorporated into the MRC funding baseline, in open competition with other fields). Major goals were genetic mapping, localized physical mapping, cDNA studies and model organism sequencing, notably the Nematode. In addition a Resource Centre was set up to provide logistic help to groups (probes, oligonucleotides, YAC library screening, computer services...). The various aspects of the HGMP have been implemented, and information is provided by an regular newsletter, "G-nome news".

II C 2. Assessment of progress

The "Resource Centre" is indeed service-oriented

The HGMP resource centre, located in the western suburbs of London, has developed into a facility providing a range of services to many laboratories. YAC screening, probe and primer provision are particularly popular, as well as the computer services which provide access to the major genome databases (GDB, OMIM, GBASE...) as well as to a variety of software. The cDNA sequencing activity is more self-centred at this stage, but will provide information and reagents to the outside world (the confusion arising from the reaction of MRC to cDNA patents is now cleared up), and the mouse backcross should prove extremely useful to groups wishing to map quickly a few sequences on the mouse genome. To my knowledge this is one of the few cases in which a facility set up and advertised as a service lab actually functions almost entirely as such!

Highly "genomic" laboratories

The number of groups involved in systematic, large-scale studies is not large. Two of them are particularly noteworthy: Hans Lehrach's "genome analysis laboratory" at ICRF, and Alan Coulson and John Sulston's group at LMB, Cambridge. The ICRF group has proposed, developed and implemented a number of highly original and efficient technical approaches to map large genomes, notably the extensive use of high-density gridded filters, contig building by oligonucleotide hybridization and the "reference library" concept (Nizetic et al, 1991). These have resulted in significant achievements such as the assembly of cosmid contigs over the S. pombe genome (Maier et al, 1992), but the whole potential of these highly intelligent technologies has not been realized so far, possibly because the number of topics and organisms studied is too large. To a certain extent, Hans Lehrach's objective was to achieve one of the major goals of the genome programme (physical mapping of the whole genome) ahead of the US project, thanks to a more intelligent and well-organized strategy. At this time (early 1993), it appears that this effort has not really succeeded, while Daniel Cohen's groups, at CEPH and Généthon, is on the way to achieving this with somewhat less astute but more seriously implemented methods.

The LMB team has completed the physical map of the Nematode and successfully initiated large-scale sequencing, using a very rational and flexible approach which involves both ABI and Pharmacia sequencers at different stages of the process (Sulston et al, 1992).

Integration of clinical genetics and genome research

Many groups, laboratories and research units in Great Britain combine some genomic work with focussed clinical genetics, using very up to date technology. Examples abound: several groups at ICRF, in Oxford (e.g. Tony Monaco, Kay Davies and others), and institutes like the MRC Human Genetics Unit in Edinburgh show a high degree of quality and integration. These groups appear to use the HGMP framework and services quite effectively to advance their goals.

II C 3. General comments (scientific)

Different degrees of emphasis

Genetic mapping does not seem to be particularly emphasized in Great Britain, although it is very competently done in the context of focussed (disease-oriented) projects. Physical mapping is more a strong point, rooted both in the tradition of cosmid fingerprinting pioneered by Alan Coulson and John Sulston and pursued by others, and in the elegant methods set up by Hans Lehrach. Sequencing is definitely a major aspect in this country, both with elaborate manual methods (Bart Barrell) and with an efficient combination of different machines, as indicated above, for the Nematode project.

Originality and inventiveness

A number of original methods or approaches are developed: Ed Southern's "tethered oligonucleotides" system (Southern et al, 1992) which should have a major impact on diagnostics and maybe on sequencing, clever manipulation of chromosomes to obtain ordered sets of progressively shortened chromosomes in Howard Cooke and Peter Goodfellow's groups etc...

II C 4. General comments (organizational)

Good information flow and coordination

Communications within the genome community and with clinical geneticists appear quite good; the HGMP newsletter (G-nome News) provides an effective link, contains a high proportion of practical information and does not avoid discussion of delicate issues (viz Tony Vicker's features about instrument manufacturers or the MRC stand on cDNA patenting). This is in contrast with the US newsletter which is much more formal and usually avoids controversial topics (such as Cantor's removal, Watson's resignation or, in the last issue, success on the chromosome 21 map in France). Many other channels of communication and in particular a multitude of small meetings also improve information flow.

Good value for money

In general genome research in Britain appears highly inventive, possibly in reaction to a general lack of funds; this leads to a good ratio of results to funding. Operations like the LMB nematode group are particularly impressive in this respect. On the other hand financial limitations may sometimes imperil highly innovative but money-intensive approaches like Hans Lehrach's - which would probably have flourished much better in an environment like "Généthon" (see the section about France). Finally the excellence of clinical genetics in Britain provides good research opportunities and speedy application of new results. This positive picture may not last if funding problems persist, as even now this really hampers a number of very good projects;

II C 5. Recent developments

The most noteworthy is the definite plans for setting up the "Sanger Institute", funded by the Wellcome trust and probably located in Cambridge, which will be devoted to megabase sequencing of the nematode under John Sulston as well as to other large scale sequencing projects.

II D. FRANCE

II D 1. Current plan and major participants

The French situation with respect to genome research is complex and requires specific discussion. Although this is a rather delicate matter, I will try to provide a candid assessment.

Academic groups vs CEPH

There are a number of good groups operating within INSERM and CNRS, the two major French research agencies. These enjoy stable funding and generally reasonable working conditions, but have very little flexibility in terms of large budget increases for specific projects or hiring of personnel (since it is largely a civil servant system, and hiring a person - except a foreign post-doc - on grant money is almost impossible). Consequently these groups have generally remained close to medical genetics and have not been able to launch large-scale (and expensive) genomic programmes. On the other hand CEPH, which is a private foundation ("association loi de 1901") but draws about half of its funding from the State, has much more leeway in terms of budget, personnel and generally methods of operation. CEPH has built on its success as coordinator of genetic mapping through the provision of DNA from a large panel of well-characterised families, and has ventured successfully into structural genomic studies. In particular the large-insert YAC library constructed at CEPH appears to be the best available at this time.

AFM vs the "public sector"

The "Association Française contre les Myopathies", headed by a very dynamic president (Bernard Barataud) has become a major player in the genome field. It has collected between 250 and 300 million French francs (30 to 40 MEcus) at each of its yearly telethons, and has invested a large fraction of these funds in research on genetic diseases. Part of these funds have been distributed through several grant programmes, but a major share has been used to set up and run, jointly with CEPH, the "Généthon" laboratory, which is partly a service lab but mostly a research institute. This is where most of the heavyweight genomic work is done in France.

The genome programme of the Ministry of Research

A French Genome programme was announced in October 1990, with 50 million (French Francs) new funding for 1991 and 100 million for 1992. Priorities stressed were cDNA studies, model organisms, informatics and some physical mapping. The actual implementation of this programme has been fraught with difficulties. Some granting decisions were made late in 1991, but these grants were honored only by late spring or summer 1992; a new scientific committee, headed by Piotr Slonimski, published a call for proposals in May 1992 and allocated approximately 100 million to close to a hundred projects (some of them within CEPH or Généthon). Most of these grants have been finally approved and a further 80-odd million francs are slated for 1993.

A complex situation

As could be expected, there are considerable tensions between these various foundations and agencies which have very contrasting motivations and also quite different time

constants. The slow start of the "official" genome programme is particularly unfortunate in this respect, since it could have been used to federate the different components involved in this work.

II D 2. Assessment of progress

Good quality academic groups

Good research groups operate in INSERM or CNRS laboratories, such as Jean-Loup Mandel in Strasbourg (Fragile X, adenolencodystrophy), Judith Melki and Arnold Munnich in Paris (SMA) and a number of others. Recent successes in mapping or cloning a disease gene within these groups often involved help from CEPH and/or Généthon: use of the extensive Southern blotting facilities at Généthon, of the CEPH YAC library, of Généthon blood and cell line banks...

CEPH and Généthon

CEPH has built on the "semi-industrial" expertise acquired while administering the genetic mapping panel to venture into heavyweight genomic work. Some endeavours (such as sequencing the MHC region) were unsuccessful (as other similar sequencing programs at the time), others such as the YAC library ended up, after initial difficulties, with excellent results. CEPH also became involved with ICRF and two firms, Amersham and Bertin, in the Eureka "Labimap" programme to develop a line of instruments for molecular biology. CEPH and Bertin are now the main, if not the sole, partners in this venture that has had a complex and conflictual history; the first instrument, a Southern blotting machine, has been introduced on the market in October 1992. In addition, CEPH continues to be very much involved in genetic mapping, both as coordinator, e.g. for the "NIH/CEPH map" (NIH/CEPH, 1992) and for in-house work (Mark Lathrop's group).

Généthon is a large facility (130 employees, mostly technicians, and a total yearly budget of 74 million French Francs) located just south of Paris. It has both service and research components, with large cell banking facilities and a set of twenty blotting machines on one hand, and several research groups on the other. Major ongoing projects include YAC contig building over the whole human genome (Daniel Cohen), generation and assignment of a very large number of microsatellite markers (Jean Weissenbach), signature cDNA sequencing (Charles Auffray) and others. Each of these projects involves something like twenty technicians and engineers, 15 million francs per year and heavy equipment. It is now clear that some of these endeavors are highly successful. The most visible is the physical mapping enterprise directed by Daniel Cohen which in 1992 demonstrated efficient selection of chromosome-specific YAC from a whole genome library (Chumakov et al, 1992 a), provided a continuous YAC contig map of chromosome 21 (Chumakov et al, 1992 b) and is on the way to a general YAC contig map of the whole human genome (Bellanne-Chantelot et al, 1992). The microsatellite generation and genetic mapping effort led by Jean Weissenbach is also highly productive, with more than a thousand markers already generated and organized in an effective "second-generation" genetic map (Weissenbach et al, 1992), while the cDNA project has already placed more than two thousand new partial sequences in public databases.

III D 3. General comments (scientific)

Genetic mapping

The unique position of CEPH (sole provider world-wide of an extensive panel of individuals as DNA) provides a strong focus for genetic mapping in France. This is likely to be reinforced by the influx of highly polymorphic CA repeat markers originating from Jean Weissenbach's work at Généthon. Genetic mapping is also performed by several groups in the context of their disease-oriented programmes.

Physical mapping

Physical mapping is, again, performed very competently by several groups on an ad hoc basis; in terms of large-scale approaches, the Alu fingerprinting procedure implemented by Daniel Cohen at Généthon (using the local automated blotting facilities and extensive computer facilities) contributes very significantly to the total mapping of the genome. This rapid and, for many, unexpected advance will induce major reassessment of physical mapping programmes elsewhere; it also raises the question of the distribution of the YACs which constitute the map: this is absolutely necessary but will require substantial resources as well as excellent organization.

Sequencing

There is substantial French participation to several collaborative sequencing programmes targeted on model organisms, notably yeast, Bacillus subtilis and Arabidopsis thaliana, and several cDNA sequencing projects are underway: at Généthon (Charles Auffray), but also in several other laboratories.

Informatics

This field is one of the foci of the programme from the Ministry of Research. French groups have been quite active in this field, with the "ACNUC" DNA sequence database set up in Lyon, one of the first to handle such objects, the "GENATLAS" system developed for the HGM 9 meeting in Paris by Jean Frezal's group (a general genome database which is now to some extent a competitor of GDB) as well as "BISANCE", a general molecular biology utility system set up by Philippe Dessen and accessed by numerous users. However most of these endeavors have failed to reach the necessary critical size, and, in spite of the recognized strength of software development in France, their practical impact on genome research has so far remained too modest, although this may change with the growing popularity of ACeDB (see below).

II D 4. General comments (organizational)

Many comments have already been made in the introductory section. It is to be hoped that effective implementation of the Ministry of Research genome programme will ease the present tensions and unify somewhat a community which at this time experiences serious tensions; meanwhile the important role of CEPH, the success of the "Généthon experiment" and the massive support of AFM to genome research should all be acknowledged and taken into account.

II D 5. Recent developments

Efforts are underway to organize all the Généthon data into a database format and to make it a generally available to the community. The format chosen is that of the very popular Nematode database, ACeDB, already used for the Arabidopsis project world-wide and for the Heidelberg Integrated Genome Database front end. Both the CEPH-Généthon physical maps and the cDNA project have attracted some criticism (Anderson, 1993 b): the reliability of the YACs (as faithful representations of genomic DNA) and of the maps is questioned, while it is claimed that many of the cDNAs sequenced in the Généthon project do not correspond to human sequences. Only time, and further experiments, will tell how serious these problems are.

II E. REST OF CONTINENTAL EUROPE

II E 1. Current plans and major participants

Excluding Great Britain and France, Europe accounts for a little more than 10% of the world output in human and general genetics, i.e. approximately twice as much as Japan for both fields. Its contribution is therefore far from negligible, but fairly dispersed and often difficult to differentiate from medical genetics. The four main producers of data are Germany, Italy, the Netherlands and the Nordic nations (Sweden, Norway, Finland and Denmark).

Germany, by far the wealthier of these countries, has a relatively moderate involvement, with a programme for "Analysis of the human genome by molecular biological methods" funded at a yearly level of approximately 4 million Ecus, in addition to substantial investment in bioinformatics. Human gene mapping is also included in various molecular biology projects, as well as the development of genome technology through informatics and instrumentation groups. Italy has had a human genome programme since 1987, centred on physical and genetic studies of the distal half of the long arm of the X chromosome, with an extensive network of collaborating laboratories. The Netherlands have a strong tradition of clinical genetics and molecular biology, and a newly started national genome programme; finally the Nordic countries have launched a coordinating "Nordic initiative", and a national programme is active in Sweden since mid-1992 (25 million crowns, i.e. 2.8 million ECUs, over three years).

II E 2. Assessment of progress

Branching out from clinical genetics to genome work

A number of excellent laboratories were originally medical genetics groups and have moved towards genomic studies. Some obvious and successful examples are the department of Medical Genetics in Leyden (previously led by Peter Pearson, now by Gert-Jan Van Ommen) that has become a major centre for sophisticated applications of non-isotopic in situ hybridization as well as of YAC technology, or Magnus Nordenskjöld's department in Uppsala. Many other examples could be given. More rarely, some institutes have moved into the Genome field from general molecular biology, as is the case for the International Institute of Genetics and Biophysics in Naples with Michele D'Urso. The Italian case is rather unique as it is the only national project to focus explicitly on a particular chromosome region; this appears to have worked rather well, in particular thanks to close ties between several Italian groups and the Saint-Louis laboratory in the USA that established the first YAC library for this region.

European instrumentation

The existence of several centres developing original instrumentation or methods relevant to genome work is noteworthy. The most important is probably the instrumentation group of Wilhelm Ansorge at EMBL which has produced the only serious competitor to date for the ABI DNA sequencer (whose marketing was unfortunately delayed by almost two years because of the takeover of LKB by Pharmacia, otherwise it might have captured a much bigger market share). This group has developed additional instruments, such as a very high-throughput DNA synthesizer and several robotic systems, which could have a significant impact. Other teams develop elegant methods (the OLA system in Uppsala, for example) which have great potential if they are taken up by competent industrial groups.

II E 3. General comments (scientific)

Contributions to the advancement of the human genome programme are very diverse and often stem from work done in the context of a particular genetic disease.

Genome informatics is a potential strong point in Europe, but again the integration between biologists and software experts still leaves to be desired. A central Bioinformatics Institute like that recently proposed by EMBL might be of great help if it was set up in a way ensuring good connection with, and input from, practising biologists.

II E 4. General comments (organizational)

Few heavyweight structured centres

In contrast to Great Britain and France, there are few large centres focussed on human genetics at a genomic level in the rest of Europe.

USA vs the EEC

The effect of the EEC genome programme in federating these groups is beginning to be felt, but - as is the general case for Biology in Europe - connections are often better with the USA than with other European countries. This is slowly changing, but, for example, the preferred country for a post-doc remains the USA.

II E 5. Recent developments

One of the most important recent development in Europe is the successful completion of the first yeast chromosome sequencing project, which has resulted in the complete sequence of chromosome III. This landmark in DNA sequencing is scientifically important because of the wealth of new genes revealed and the reevaluations thus brought about. However its psychological impact is even more important: it shows that consortia of numerous European laboratories can be effective (although expensive) in spite of their cultural and technical differences, and, together with the striking recent successes of Généthon, it has strongly bolstered the morale of European groups and given them confidence that the USA is not the only environment in which large scientific projects can be successfully carried out. The successes of Généthon groups in the fields of genetic and physical mapping are also very important factors in the emergence of Europe as a major "genomic power".

II F. Former Soviet Union

II F 1. Current plans and major participants

1989: A grand USSR Human Genome Programme

At the initiative of several Russian scientists, in particular Alexander Bayev who liaised with Gorbatchev on this matter, a Human Genome programme was announced and launched in 1989. It was well funded: approximately 25 million roubles per year (at a time where one rouble was supposedly equivalent to one US dollar), with a specific allocation of 5 million roubles in foreign exchange. This represented quite a significant amount, to be compared, for example, with a total yearly budget of a few million roubles for a large research institute. The programme included all the usual categories of topics, with emphasis on specific chromosomes (3, 5, 13, 19). It also covered functional studies and medical genetics as well as instrument and reagent development. Several hundred grants were provided to scientists in a score of institutes, the average grant being relatively small (10 to 20,000 roubles). The visibility of the programme at the international level was high, with Andrei Mirzabekov (director of the Institute of Molecular Biology, Moscow) becoming vice-president of HUGO and the planned opening of a first HUGO office in Moscow.

Genome research in the present Russian environment

The programme is now essentially a Russian one. It is in fact considered as the most active biological research project in this country, and has received in 1992 a total of 130 million roubles ([1]). The project is run from an office located at the Moscow Institute of Molecular Biology, and the funds are dispensed through this Institute. The whole programme is administered in a rather autonomous way by a scientific council and seven committees, one for each of the specific fields designated:

- cell culture collection, chromosome isolation
- clone libraries, physical mapping of chromosomes
- human genome sequencing
- structural-functional analysis
- medical genetic mapping, gene therapy
- software
- instruments, reagents and probes

As mentioned previously, individual grants are rather small: the whole budget is divided into ca. 750 units (rather quaintly called "genes"), a typical grant being one or two "genes"; some bigger projects may be funded at the level of five or ten "genes", but there has been no attempt to set up large specialised genome centers as in the US or elsewhere.

1.It is extremely difficult to estimate how much this represents in actual purchasing power. If the sum is translated into US dollars using the going rate early 1993 (500 roubles to one dollar), the amount is almost ridiculous: less than 300,000 $. If however one uses as a yardstick the price of a Russian-made mid-range PCR machine (50, OOO roubles), whose equivalent in the West is worth maybe 4,000 dollars, this budget comes out as more than 10 million $...

Thus the main aim of the project appears to be stimulating and coordinating research related to genome studies in existing groups.

II F 2. Assessment of progress

Off to a flying start

Laboratory visits indicate that the Russian (or originally soviet) genome programme has had a significant impact. It is clear that the major laboratories (mostly belonging to the Academy of Sciences) went on a buying spree in 1989/1990, and the level of equipment is very good: DNA sequencers, Biomek robots, PC computers abound. Genome funding represents a high percentage of lab income is some cases, and a large number of research projects have been started two or three years ago - especially so since the scope of the project was quite wide and could cover a good fraction of molecular biology.

Difficulties linked to the collapse of the USSR

Times have changed however, and the demise of the USSR has created many problems: loss of buying power because of the very high inflation, lack of foreign exchange, severance of ties with former Soviet republics that were significant scientific partners (such as Ukraine) or provided reagents payable in roubles (such as Lithuania for restriction enzymes). A number of scientists have left for Western laboratories, in principle for limited periods, but in fact permanently for a good proportion of them. In this context, the birth of a new program has been difficult, and the results so far relatively modest - although there are centers of excellence and groups operating at a very good level.

II F 3. General comments (scientific)

A strong DOE flavour

Discussions with Russian scientists involved in Genome research make it clear that the predominant conceptual influence comes from DOE groups and leaders such as Charles Cantor, Anthony Carrano or Bob Moysis. This may be accidental, but it could possibly reflect some similarities between DOE organization and soviet structures - at least when compared with the more free-wheeling NIH style! Contacts with Europe (including Great-Britain) appear less developed, and there seems to be little collaboration with CEPH or Généthon groups.

Approaches to specific chromosomes

There does not seem to be extensive genetic mapping efforts in the country, apart from local mapping in the context of a given disease. This type of research is furthermore hampered by restricted access to patient populations now belonging to different, sometimes antagonistic countries. There is however a concerted physical mapping project centred on four chromosomes (3, 5, 13, 19). Part of this effort is invested in chromosome jumping and linking clone schemes rather similar to those advocated by Charles Cantor and Cassandra Smith, a couple of years ago. A set of apparently very efficient vectors has been developed to construct the corresponding jumping and linking libraries; the work is pursued, in particular on chromosome 3, in collaboration with George Klein's group in Sweden and involves the Moscow Molecular Biology Institute (Lev Kisselev). Interest on chromosome 13 is largely centred on the Wilson disease locus (13q14-q21). Studies on chromosome 19 involve groups at the Moscow Molecular Biology Institute (Alexander Zelenin), at the Shemiakin Bioorganic Chemistry Institute and at the Institute of Molecular Genetics (both in Moscow) with Eugene

Sverdlov. The latter studies are to a large extent centred on expressed sequences, using several methods to specifically recover cDNA clones corresponding to genes present on this chromosome. Svedlov's group has initiated collaborations with Anthony Carrano's Genome center at Lawrence Livermore (USA). These various efforts appear to be of medium to high quality, but they are probably somewhat obsolescent now with recent progress on the whole-genome physical map.

DNA sequencing

DNA sequencing is relatively well developed; Applied Biosystems or Pharmacia machines are installed and used in several laboratories; STS sequencing is done as well as sequencing of the jumping and linking clones mentioned above to find their overlaps. A large sequencing center (using conventional radioactive methods) is apparently active in Novosibirsk (Siberia), and performs contract sequencing for various institutions. In addition, Andrei Mirzabekov's group is one of the pioneers of sequencing by hybridization. They have developed a variation of the method solving one of its major problems (the differential stability of different sequences) and have made progress in the miniaturization of the "sequencing chips", in collaboration with a former military research laboratory.

II F 4. General comments (organizational)

Foreign exchange and brain drain

Genome research suffers, as does the rest of Russian science, from the recent scarcity of foreign exchange. The present exchange rate (which does not reflect the actual cost of living within the country) makes any foreign reagent prohibitively expensive; thus well-equipped and - until recently - well-staffed laboratories have been considerably slowed down. In addition academic salaries have not kept up with inflation. A senior scientist is paid (at the going rate) ten or twenty dollars per month. The actual buying power is at least ten times higher - still very insufficient, leading these professionals to spend part of their time in unrelated occupations. These two factors have promoted a significant brain drain, leaving some laboratories heavily understaffed. A few, such as the Moscow Molecular Biology Institute, have succeeded so far in limiting the problem by encouraging applications for Russian and foreign grants (which now make up 70% of the Institute's income), giving young scientist group leadership and allowing trebling of salaries from grant funds. Even there, however, the situation remains very fragile. Yet there still is a lot of potential in these groups, which could be released by funds provided in almost catalytic amounts.

Self-reliance and "reinventing the wheel"

The relative isolation of Russian scientists in the past has forced them to be very self-sufficient. To a certain extent this characteristic remains helpful in the present situation: lack of foreign exchange is partially compensated by local construction or modification of instruments, and several groups, for example, synthesize their own reagents for oligonucleotide assembly. This tendency can however be overdone and lead to efforts amounting to reinventing the wheel, discernible in the development of PC software packages for handling sequence data or for some vector developments. Contact with Western science is still restricted (for monetary rather than political reasons now), access to "grey" literature and grapevine information very limited, with the result that somewhat outdated or unpromising avenues of research are sometimes pursued too long.

II F 5. Recent developments

A very fluid situation

The whole country is in a situation of flux, with very high inflation (20% per week) and looming political uncertainties. Understandably, science is not a first priority under these conditions, but the risk is that lasting damage can occur if this situation continues (or worsens), if more scientists leave the country while the presently abundant and still functional equipment breaks down or becomes obsolete. Genome research, although privileged compared to other types of biological science, is also in danger. Yet the whole environment is probably too chaotic and unstable to envision large-scale, long-term projects.

Need for modest but quick input

As mentioned previously, some institutes are still very functional, with good equipment and qualified scientists. They only require very modest amounts of cash to, on one hand, purchase some vital supplies unobtainable within Russia, and on the other hand, provide salary supplements allowing researchers to stay in Russia and to do research full-time. The sums involved are very modest: hoped-for salary supplements are in the range of a hundred US dollars a month, and reagent requirements for a large Institute have been estimated at a total of 20,000 dollars per year. As an interim measure, a number of grants of this size for studies involving some collaboration with European groups, quickly awarded, could make a significant contribution to the survival of Russian science while providing useful genome data.

III. MAJOR ISSUES

After this review of national programmes and progress, I will give an outlook (again, a highly personal one) on the present state in the four major components of genome programmes (genetic mapping, physical mapping, sequencing and informatics), and indicate how I see the evolution of these topics.

III A. GENETIC MAPPING

Initial successes and great expectations followed by difficulties

The concept of a general genetic map of the human genome based on DNA polymorphisms is only a little over ten years old, but the construction of such a map has made tremendous progress since its first proposal by Botstein. Optimism was high in the mid to late eighties, with the publication in 1987 of the first general genetic map having an average spacing of 10 to 20 centimorgans. Prediction of, and planning for, a 2 centimorgan map by 1991 or 92 was incorporated, for example, in the US genome programme. Progress turned to be more difficult than anticipated, both because of the sheer amount of work required (with constant technology, a 2 centimorgan map requires a hundred times as much effort as a 20 centimorgan map since ten times as many individuals need to be studied with ten times as many probes), because of the rather uninspiring nature of the day to day effort involved and of a certain lack of interest by funding agencies. It was also felt, in some circles, that large-scale physical maps based on pulsed field gel experiments and using "linking clones" were around the corner, and that they would make much of genetic mapping obsolete.

Progress is resumed thanks to new technology

However it was soon realized that physical maps would have to be based on contigs of cloned DNA segments to be really useful - and that this was not going to be easy. It also became apparent that in any case a complete and detailed genetic map with high quality landmarks was, and would remain, essential to the positional cloning ("reverse genetics") approaches. More polymorphic markers were found, first minisatellites or VNTRs, later CA repeats or microsatellites - so that by the time Southern blotting was more or less automated (e.g. at Généthon) RFLPs were no longer the main genetic mapping tool. Generation of large numbers of CA repeats is going ahead in a number of laboratories, and the first whole-chromosome, and even whole genome, genetic maps based solely on CA repeats have already been published (Hazan et al, 1992; Weissenbach et al, 1992). Data capture and analysis using this system remain too cumbersome, and ways are being found to efficiently multiplex the assay of CA polymorphism and to call the results more or less automatically. Throughout this stage, as well as the preceding one, the general availability of the CEPH panel of families as DNA samples has played an important and very positive role - even though the required commitment of recipients to test the whole panel if a polymorphism was found was sometimes felt to be quite stringent.

Two general maps have been published late in 1992, the one produced by collaborating groups using the CEPH panel (NIH/CEPH collaborative mapping group, 1992) and a "pure" microsatellite map obtained by Jean Weissenbach's laboratory at Généthon (Weissenbach et al, 1992). Although less detailed (814 markers instead of 1416), the Généthon map is both more reliable (being constructed by a single group rather than assembled from many different sets of data) and more useful since it only incorporates highly polymorphic markers.

The importance of software

The software available to perform genetic analysis has made good progress (aided of course by the very large increase in computing power per dollar) and is now able to tackle difficult problems such as the search for genetic component in multifactorial diseases such as diabetes or hypertension, as well as in characteristics such as adult height or longevity. Input of data, however, remains a largely manual and error prone procedure which should be automated if at all possible or at least made much more user-friendly.

The near future

It now appears quite feasible to generate a density of markers (most of them highly polymorphic) equivalent to one per centimorgan. This does not mean it will be easy to order all of them with respect to each other, since at such small recombination fractions the search for individuals with recombinations in the interval becomes the dominant factor. It may however not be necessary to do this by genetic means, since efficient cytogenetic methods (interphase mapping, or even in situ hybridization to highly decondensed chromosomes) may be able to do this much more quickly. A recent variant of "sperm mapping", in which the whole DNA of a single sperm is first amplified with random primers (Zhang et al, 1992), thus making it possible in later stages to assay many markers on the same haploid DNA complement, may also become very significant. Thus the outlook for a high-resolution genetic map appears quite bright.

III B. PHYSICAL MAPPING

Substantial vs unsubstantial maps

At the beginning of genome programmes the concept of whole chromosome physical maps based on pulsed field gel analysis of Not I cut genomic DNA complemented by a complete set of the corresponding linking clones (segments of DNA containing the rare Not I sites and a few kilobases of DNA on either side) was entertained by several groups. It soon became clear that this approach suffered from serious technical difficulties and shortcomings, and that a really reliable and useful map of a whole human chromosome had to based on a complete (or almost complete) set of overlapping cloned DNA segments - as done by Kohara et al for the E. coli genome in 1987.

Are cosmids completely superseded by YACs?

Contig building across human chromosomes was initiated, in particular for chromosomes 16 and 19 at Los Alamos and Lawrence Livermore. This rather heroic undertaking (it takes 5,000 to 10,000 cosmids to cover such a piece of DNA) was pursued until coverage reached 60 to 70%, and benefited from the arrival of YACs which promise to make closure possible, as has been the case for the nematode genome. The extensive investment in cosmids was not a dead loss since, in practice, it turns out to be very useful to have a "cosmid layer" underlying a YAC contig map: cosmids provide DNA in a form which is very amenable to further experiments such as looking for genes.

Chromosome-specific YAC libraries

Chromosome-specific YAC libraries are in principle an important resource for the construction of a physical map: alignment of the few hundred large YACs needed to cover an average-sized chromosome should be almost trivial compared to the equivalent endeavour with ten thousand cosmids. They can be obtained the hard way, by making a library from a suitable human-hamster hybrid cell line and subsequently screening for the few per cent of human-specific clones, as done in Saint-Louis for the Xq24 to Xq28 library. Attempts to make such libraries from sorted chromosomes have given some promising results but have so far failed to provide complete libraries. The alternative route of "extracting" chromosome-specific components from a whole-genome YAC library is much more appealing, and recently two groups have reported success in doing this using as a mixed probe sequences derived from a chromosome-specific cosmid library (Ross et al, 1992) or from a somatic hybrid (Chumakov et al, 1992 a).

What is the best way to obtain YAC contigs?

It is now generally accepted that the primary whole-chromosome physical maps will be based on YACs, in spite of the problems with chimeric clones (Green et al, 1991 b), at least until some better vector comes along. Ab initio physical mapping of chromosomes 7 and X has been pursued in Saint-Louis, using "STS content mapping" (Green et al, 1991 a) and the whole genome YAC library. This approach relies on defining about one thousand STS per chromosome and then assaying YACs with the corresponding oligonucleotides. Positive YACs usually contain two or three such STS and can be overlapped with their neighbours by looking for the common STS. This very robust approach is work-intensive but reliable. Daniel Cohen's group at Généthon have used a somewhat similar STS content method to construct the complete chromosome 21 map (Chumakov et al, 1992 a), and a more adventurous fingerprinting technique to build contigs simultaneously over the whole genome. This project,

which uses the high-capacity blotting machines installed at Généthon, appears well on the way to produce a nearly continuous contig map of the whole genome (Bellanne-Chantelot et al, 1992).

The near future: complete - and available? - YAC contigs

Thus complete (or almost complete) YAC-based physical maps of many whole chromosomes will soon be obtained. These maps can be tremendously useful to the whole community, particularly so if they, and the underlying reagents, are made available in a convenient format. In this respect, the example set by the nematode community is worth considering. The LMB laboratory provides so-called "polytene filters" which contain a complete set of YACs ordered as they occur along the six Nematode chromosomes. The user simply hybridizes his probe to this filter and obtains two or three positive signals on adjacent spots, whose position on the filter immediately provide chromosome assignment and localization to within a hundred kilobases; the corresponding YACs and cosmids can then be obtained from the centre. This is one of the schemes which could be used to make available maps and reagents for each of the human chromosomes.

III C. SEQUENCING

No breakthrough in methods (so far)

When Genome programmes were first discussed in the mid-eighties, it was generally expected that sequencing technology would progress rapidly and that megabase sequencing at prices well below one dollar per base was just around the corner. Such progress has not occurred; instead several megabase sequencing projects initiated in the late eighties and aiming confidently at obtaining one or several megabases of sequence have failed to deliver the expected results: viz. the E. coli and Salmonella sequence, the human MHC region, or the Xq28 band. The traditional sequencing method has been automated in part or (in Japan) in toto, but throughput in the latter case cannot yet be assessed; multiplex sequencing, a very rational approach, has run into serious difficulties with automating capture of the huge amounts of data produced. Exotic methods such as STM and AFM have generated much initial excitement but have not yet demonstrated feasibility; the same is true for very fast techniques based on single molecule degradation and extremely sensitive analysis of single bases. Sequencing by hybridization is closer to a real-life feasibility demonstration; it certainly holds a lot of promise for detection of mutations in known genes and may well develop into the megasequencing method of the late nineties (Cantor et al, 1992; Pevzner et al, 1991). Efforts must be continued, since only radically new technologies are capable of bringing speed up, and cost down, by the one or two orders of magnitude necessary to make human genome sequencing a realistic task.

The potential of well organized "classical" sequencing operations

Current results show that it is feasible with present technology to sequence a few hundred kilobases, up to at least one megabase, at a cost of the order of one dollar per base. In this respect the model for near-future megabase sequencing appears to be the Nematode operation (Sulston et al, 1992) - rather than the EEC yeast effort (Oliver et al, 1992) which has been useful and worthwhile but very expensive (total cost 2.6 MEcus for 315 kilobases, i.e. close to ten dollars per base): sequencing will in the future have to be divided in larger chunks between a smaller number of centres. Sequencing at this cost is certainly worthwhile for model organisms with small genomes such as E. coli, Bacillus subtilis or even yeast; as genomes get larger (e.g. the 100 megabase Nematode genome) and less densely populated with genes, the answer becomes less obvious. In the human system megabase sequencing can only at this point be considered as a model experiment: sequencing a few one or two megabase long, particularly "interesting" regions, to find out how such sequencing actually goes in a genome very rich in repetitive sequences, how efficient present algorithms are at finding genes in such DNA... and whether anything unexpected comes out of such massive experiments. It will then be easier to decide whether, and how, to proceed.

cDNA sequencing

Massive and partial cDNA sequencing was proposed several years ago, in particular by Sydney Brenner, but the most publicized implementation of this strategy occurred in the USA with Craig Venter at NIH. The approach has turned out to be quite effective at finding "new" genes, and provides partial sequence data on several thousands of hitherto unknown genes. It is practiced in a number of laboratories outside the USA, including the HGMP resource centre in Harrow, the Genethon (Charles Auffray), the Osaka Institute of Molecular and Cellular Biology (Kenichi Matsubara) and others. Some scientific questions remain open: how effective is the approach at finding all, or even a significant fraction, of the existing

genes? For example, one may wonder how many of the 182 chromosome III genes in yeast would have escaped detection by this approach. Another issue is the localization of the "new" genes revealed by this approach: even simple chromosome assignment is, at this time, very cumbersome (Polymeropoulos et al, 1992), and new methods (a high throughput non radioactive in situ hybridization technique routinely applicable to 0.5 kilobase segments, or preferably wide distribution of human YAC "polytene filters") are badly needed. As these exercises continue, the question of real time data availability becomes more and more important, to avoid duplication of effort. This has two aspects: data must be comparable, i.e. laboratories should agree to sequence the same region of cDNAs (5' or 3'), and it should be deposited and made accessible very quickly. Finally, of course, the vexing problem generated by attempts to patent these partial sequences should be resolved, ideally by a general decision not to indulge into such silly exercises. The recent negative ruling by the US patent office is naturally a big step in the right direction.

Genomic DNA versus cDNA sequencing

Discussion on the relative merits of these two approaches are heavily dependent on the organism considered and on the cost of large-scale sequencing. Under present conditions (one to two dollars per base) sequencing genomic DNA definitely makes sense for small genomes, up to a few megabases: a 5 megabase (E coli) or even 15 megabase (yeast) sequence is clearly feasible, and the wealth of information provided (as shown by the recent nematode and yeast results) offsets the relatively high price of the exercise. The high gene density found in these organisms (one gene per two kilobases on average in yeast) reduces the saving which could be achieved by going the cDNA way.

For larger genomes with intron-containing genes (such as the 100 megabase nematode genome) the question becomes more debatable. For the human genome, if one takes the presently accepted upper limit for the number of genes of 100,000, and allows 1.5 Kilobases of coding sequence per gene, one ends up with an exon content of 5% - a figure which does make cDNA approaches attractive. So far large human DNA sequences (the HPRT locus, and the T cell receptor region (Wilson et al, 1992)) do not appear to contradict current estimates of gene density - but the sample is too small to be representative. Indeed, if the Nematode uses 15,000 genes (the present estimate) for its relatively simple lifestyle and the coordination of its 959 cells, it may well be expected that a human being needs more than ten times that number: gene number and gene density would thus go up, and so would the attractiveness of genomic sequencing. In any case, if cost can be brought down to below 0.1 dollars a base by new techniques or by very efficient implementation of present methods, a whole chromosome could be sequenced for a sum of the order of 10 million dollars, i.e. roughly the amount previously budgeted for the physical map of such an entity in the US genome programme: large-scale sequencing would become irresistible.

III D. INFORMATICS

Informatics has been recognized from the start as a very important component of genome programmes, and has been funded accordingly, in particular in the USA. Significant progress has been made, but the situation is still far from satisfactory because of the exponential growth in the amount of data and the bewildering variety of tasks which need to be addressed.

III D 1. General genome databases

Content and timeliness: the issue of unverified data

Several genome databases have been established to cater to the growing needs of genome mappers. In addition to the familiar DNA sequence databases (EMBL and Genbank), there are systems storing all kinds of human mapping data (GDB), information on the mouse genome (GBASE), clinical description with mapping results (OMIM, GENATLAS) to name but a few. In principle these systems only store "public" data, either already published or adequately verified by expert inspection and released for general distribution. Thus the data is in most cases quite reliable, but often lags behind recent work; for example databases contain only a fraction of the microsatellite primers which have been defined over the world (Pearson et al, 1991). It would probably be very useful to display provisional data (flagged as such), since such non-verified results often provide very useful hints to others.

Structure and user (un-) friendliness

Databases have been developed over the years using a variety of database structures : IRX for OMIM, SYBASE for GDB and GENATLAS, INGRESS for GBASE... Each of these systems has its own query structure and its particular quirks ; their only standard feature is their lack of user friendliness, which makes it relatively difficult for the average biologist to use them. A rising star in this world of databases is the Nematode system, ACeDB ("A Cenorhabditis elegans Data Base) which is winning over scientists from other fields thanks to its performance and ease of use: the system is now used for the Arabidopsis programme (AAtDB), by a some Drosophila groups, and it constitutes the front end of the Integrated Genome Database (IGD) being developed at the DKFZ in Heidelbgerg.

Connection problems

Because of their complex nature, of the variety of data stored and of the many ways in which links can be established between different kinds of information, these databases are normally used on-line (the only exception being the sequence banks in their present - and transient - simple flat-file format). Accessing them implies not only a local computer (which is easy), but also a high-speed link to the database itself or to one of its secondary nodes, and sufficient local network expertise to actually establish the connection in a reliable way. While the high-speed link situation is relatively satisfactory in most of the USA, it is definitely not so in the rest of the world, yet using a database such as GDB with transmission speeds below 9600 baud is essentially hopeless; local experts able to puzzle out an efficient connection are also very scarce.

The growing power and dropping price of microcomputers and Unix workstations, in conjunction with these transmission problems, is causing a shift towards system in which a local (periodically refreshed) version of the database is used.

III D 2. Laboratory notebooks and local databases

Computerized laboratory notebook systems are often discussed, but few groups actually implement and use them. Yet the need is obvious: even partial automation leads to a large increase in the number of objects handled in the laboratory, thus rigourous and efficient data organization and storage become absolutely essential to avoid disaster. But the development of such a system for the average group is not easy, as it makes great demands on flexibility (objects, techniques and nature of data change quite often) and user friendliness (every scientist, student and technician should use the system daily), not to mention raw computing power and storage space (as images often have to be included in the data). Only fairly large groups organized around a common task seem to succeed in this endeavour, for example the Lawrence Livermore group which has set up a sophisticated and relatively easy to use system to record and follow progress on its cosmid fingerprinting and contig building work (Stallings et al, 1992), or the Généthon team for their whole-genome contig building task (Bellanne-Chantelot et al, 1992). Such large groups end up building local databases which must then be able to communicate with the "general" databases: this is best taken into account from the start, as it involves difficult informatics problems in addition to delicate questions of data ownership and restricted or general availability of the information.

III D 3. Software for comparing and interpreting sequences

Comparing a sequence to all known sequences is a task whose complexity increases exponentially with the number of known sequences. In spite of the very rapid growth of computing power per dollar over the past decade, it is not certain that computers will keep up with the increasing flow of data. New algorithms are being perfected, and massively parallel machines, using dedicated hardware chips, may be the answer. Finding exons in a large genomic DNA sequence is still a vexing problem, as shown by the difficulties encountered in interpreting a recent 106 Kilobase genomic sequence from chromosome 19 (Martin-Gallardo et al, 1992): the algorithm used found less than half of the exons subsequently detected by other methods. More work is clearly needed here.

III D 4. Communication problems

Communication between biologists and computer scientists remains imperfect: they still belong, by and large, to two quite different cultures. Both fields have been evolving very quickly in recent years, and few individuals have real dual competence. An increase in their numbers, through attractive fellowships and positions, is probably the only way of easing this difficulty.

III E. INSTRUMENTATION

A necessity...

Newcomers to recombinant DNA laboratories, and even to genome centres, are often appalled by the predominance of manual work. Most procedures are performed by hand, machines being used only for a few specific steps. Even DNA "sequencers" only automate a small - albeit vital - part of the sequencing procedure. It appears obvious that to increase throughput and reproducibility the hands of the student, the technician or the post-doc must be replaced by machines - a feat which should not be impossible in this age of sophisticated electromechanical devices driven by cheap and powerful microprocessors.

... Yet automated labs are few and far between

In fact, even in specialized genome centres, manual work still predominates. Few laboratories, for example, have completely automated cosmid preparations; construction of YAC libraries still involves lengthy and tedious hand picking of clones into microtiter plates; and apart from a few pipeting robots (in most cases the Beckman Biomek machine, originally developed for Elisa assays) there are very few robots around. Exceptions exist, the foremost being the "Genethon" centre set up near Paris by AFM and CEPH; another case - however not yet functional - is the HUGA "sequencing factory" in Tsukuba. But on the whole the penetration of instrumentation and robotics has been disappointingly slow in laboratories implementing human genome programmes (of course, it is even slower elsewhere).

Understandable difficulties

This unfortunate state of affairs reflects some very real problems. Recombinant DNA technology is still in a state of flux: thus a manufacturer's decision to develop an instrument automating a certain procedure entails a very serious risk: the probability is high that by the time the instrument is ready to be marketed (several years later) the procedure will have become obsolete. The "Labimap" Southern blotting machine has just been marketed, more than four years after the programme started, and by now PCR has replaced Southern blotting in most of its applications; likewise the HUGA set-up is based in part on obsolescent sequencing technology. In addition genome research, even today, is a relatively small sector: the market for a hypothetical high-throughput sequencing machine capable of reading 500,000 bases per working day is likely to be quite small - unless the resulting sequence is very affordable, a cent or less per base. Other, more "cultural" reasons hinder progress: most biologists are not very instrument-literate, and the race to a publishable result does not lend itself to careful investment in new technology; and communication between robotics specialists and biologists suffers some problems - as with informatics.

Future directions

There are however some signs of progress. ABI has sold more than 800 sequencers world-wide, even though they are not all in actual use; semi-automated procedures based on the 96-well microtiter plate (or multiples thereof) are becoming more common; and instrument literacy, as well as computer knowledge, is increasing among biologists. Genome programmes *have* had a positive effect, by providing both the incentive and the funds for at least partial automation; and the demonstrated success of some groups in performing some routines by machine is now tempting smaller, more conventional laboratories to do the same. Concurrent increases in computing power per dollar for small computers make it easier to automate complex procedures: mechanical precision (always very expensive) can be substituted by

software sophistication - an increasingly available commodity. Numerous opportunities exist in this field, both for research groups to operate more effectively and for astute companies to generate cash flow and profits.

IV. CONCLUSIONS AND PERSPECTIVES

It is difficult to draw conclusions in such a study at a time when events are moving so rapidly. Clearly the balance has shifted in recent months towards Europe, with the success of enterprises such as yeast whole-chromosome sequencing, whole genome physical maps and refined genetic maps. Agencies in the US have, as usual, been quick to assess the new situation and to launch new centers incorporating some of the features of Généthon (Roberts, 1992; Anderson, 1993) and in attracting the CEPH Mega Yac library to the other side of the Atlantic.

Even though I have deliberately avoided discussing the EEC programme as such (noting only the areas in which its effects have been felt), it is clear that it has a great role to play in capitalizing on this - possibly temporary - success to establish a more balanced relationship with the USA. Making the results of whole genome physical mapping readily available to the community will be an extensive but necessary task; it must be tackled by informatic means (an accessible, complete and user-friendly database), but also at the level of the distribution of YAC clones, a very labour-intensive task. Making these reagents readily accessible to all European scientists (e.g.. by wide distribution of "polytene filters" containing ordered YACs for all human chromosome) would be an essential step in this direction.

In view of very widespread criticism directed to GDB, it becomes useful to envision alternative databases or at least different database access systems: in this respect, the progress on the Heidelberg IGD appears quite promising and the distributed model advocated on this group, using a locally running ACeDB system fed with data from the other genome databases, appears particularly promising.

Finally attention, and some funding, should be devoted to improving interfaces between the "genome world" and the general biology community, including not only medical genetics but also, for example, groups performing functional research in the mouse system so that structure and function are brought closer together.

REFERENCES

ADAMS MD et al, 1991: Complementary DNA sequencing: Expressed sequence tags and human genome project. Science 252: 1651-1656

ADAMS MD et al, 1992: Sequence identification of 2,375 human brains genes. Nature 355: 632-634

ANDERSON C, 1992 a: Controversial NIH genome researcher leaves for new $70-million institute. Nature 358: 95

ANDERSON C, 1992 b: US genome project does it the French way, conceding that size matters after all. Nature 360: 401

ANDERSON C, 1993 a: Genome project goes commercial. Science 259: 300-302

ANDERSON C, 1993 b: Genome shortcut leads to problems. Science 259 : 1684-1687

BAYEV AA, 1990: The Human Genome Project in the USSR. Biomed. Sci. 1: 106-107

BELLANNE-CHANTELOT C et al, 1992: Mapping the whole human genome by fingerprinting yeast artificial chromosomes. Cell 70: 1059-1068

CANTOR CR et al, 1992: Report on the sequencing by hybridization workshop (special feature, meeting report). Genomics 13: 1378-1383

CHUMAKOV IM et al, 1992 a: Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. Nature Genetics 1: 222-225

CHUMAKOV IM et al, 1992 b: A continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359: 380-387

ENDO I et al, 1991: Human genome analysis system. Nature 352: 89-90

FOOTE S et al, 1992: The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. Science 258: 50-66

GREEN ED et al, 1991 a: Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. Genomics 11: 548-564

GREEN ED et al, 1991 b: Detection and characterization of chimeric yeast artificial chromosome clones. Genomics 11: 658-669

HAZAN et al, 1992: A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. Genomics 12: 183-189

KHAN AS et al, 1992: Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genetics 2: 182-185

MAIER E et al, 1992: Complete coverage of the schizosaccharomyces pombe genome in yeast artificial chromosomes. Nature Genetics 1: 273-277

MARTIN-GALLARDO A et al, 1992: Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nature Genetics 1: 34-39

NIH/CEPH collaborative mapping group, 1992: A comprehensive genetic linkage map of the human genome. Science 258: 67-80

NIZETIC D et al, 1991: Construction, arraying, and high-density screening of large insert libraries of human chromosome X and 21: Their potential use as reference libraries. Proc. Natl. Acad. Sci. (USA) 88: 3233-3237

OKUBO K et al, 1992: Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genetics 2: 173-179

OLIVER SG et al, 1992: The complete DNA sequence of yeast chromosome III. Nature 357: 38-46

PEARSON PL et al, 1991: The Human Genome initiative - Do databases reflect current progress? (Perspective) Science 254: 214

PEVZNER PA et al, 1991: Improved chips for sequencing by hybridization. J. BIOMOL. STRUCT. DYN. 9: 399

POLYMEROPOULOS MH et al, 1992: Chromosomal assignment of 46 brain cDNAs. Genomics 12: 492-496

ROBERTS L, 1992: NIH takes new tack on gene mapping. Science 258: 1573

ROSS MT et al, 1992: Selection of a human chromosome 21 enriched YAC sub-library using a chromosome-specific composite probe. Nature Genetics 1: 284-290

SOUTHERN EM et al, 1992: Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics 13: 1008-1017

STALLINGS RL et al, 1992: Evaluation of a cosmid contig physical map of human chromosome 16. Genomics 13: 1031-1039

SULSTON J et al, 1992: The C. elegans genome sequencing project: a beginning. Nature 336: 37-41

TAKAHASHI E-I et al, 1992: A high-resolution cytogenetic map of human chromosome 3: Localization of 291 new cosmid markers by direct R-banding fluorescence in situ hybridization. Genomics 13: 1047-1055
THOMPSON L, 1993: Healy and Collins strike a deal. Science 259: 22-23

TRASK B et al, 1992: Fluorescence in situ hybridization mapping of human chromosome 19: Mapping and verification of cosmid contigs formed by random restriction enzyme fingerprinting. Genomics 14: 162-167

WEISSENBACH F et al, 1992: A second generation linkage map of the human genome based on highly informative microsatellite loci. Nature 359: 794-801

WILSON RK et al, 1992: Nucleotide sequence analysis of 95 kb near the 3' end of the murine T cell receptor alpha/delta chain locus: Strategy and methodology. Genomics 13: 1198-1208

ZHANG L et al, 1992: Whole genome amplification from a single cell: Implications for genetic analysis. Proc. Natl. Acad. Sci. (USA) 89: 5847-5851

European Commission

EUR 15412 - An assessment of progress in Human Genome Programmes worldwide.
(A support study for the evaluation of the EC Human Genome Analysis Programme)

B.R. Jordan

Luxembourg: Office for Official Publications of the European Communities

1994 – X, 43 pp. – 21.0 x 29.7 cm

Science and Technology policy series

ISBN 92-826-8226-9

Price (excluding VAT) in Luxembourg: ECU 7

This report, based on a one-year survey in approximately one hundred genome research laboratories all over the world and on continuing contact with their leaders, attempts to provide a first-hand yet synthetic and reasonably up to date picture of developments, trends and problems in this field.

The situation in each of the major countries is reviewed, indicating in each case the major agencies and/or foundations involved, an assessment of overall progress towards the advertised goals, specific scientific and organizational comments, and finally a note on the most recent developments. Broadly speaking this indicates continuing strength in the USA, fairly recent but very serious efforts in Japan, excellent results in Great-Britain in spite of rather modest funding, major achievements in France obtained in a fairly unconventional way, uneven but often interesting performance in other European countries and serious difficulties in Russia although the potential remains significant.

The second part of the document is devoted to a more synthetic and in-depth discussion of the major issues in genome research. This covers genetic mapping, stressing the contribution of microsatellite-based approaches and the importance of data acquisition and informatics, physical mapping with YAC contigs and a discussion of how to make them widely available, genomic and cDNA sequencing, informatics both at the level of general databases and at that of laboratory notebooks, and finally instrumentation, a still relatively slow-moving field in actual practice.

Conclusions, perspectives and a few recent references (1992 and 1993 only) round off the document.

Venta y suscripciones • Salg og abonnement • Verkauf und Abonnement • Πωλήσεις και συνδρομές
Sales and subscriptions • Vente et abonnements • Vendita e abbonamenti
Verkoop en abonnementen • Venda e assinaturas

BELGIQUE / BELGIË

Moniteur belge / Belgisch Staatsblad
Rue de Louvain 42 / Leuvenseweg 42
1000 Bruxelles / 1000 Brussel
Tél. (02) 512 00 26
Fax 511 01 84
CCP / Postrekening 000-2005502-27

Autres distributeurs /
Overige verkooppunten

Librairie européenne/ Europese Boekhandel
Avenue Albert Jonnart 50 /
Albert Jonnartlaan 50
1200 Bruxelles / 1200 Brussel
Tél. (02) 734 02 81
Fax 735 08 60

Jean De Lannoy
Avenue du Roi 202 /Koningslaan 202
1060 Bruxelles / 1060 Brussel
Tél. (02) 538 51 69
Télex 63220 UNBOOK B
Fax (02) 538 08 41

CREDOC
Rue de la Montagne 34 / Bergstraat 34
Bte 11 / Bus 11
1000 Bruxelles / 1000 Brussel

DANMARK

J. H. Schultz Information A/S
EF-Publikationer
Ottiliavej 18
2500 Valby
Tlf. 36 44 22 66
Fax 38 44 01 41
Girokonto 8 00 08 86

BR DEUTSCHLAND

Bundesanzeiger Verlag
Breite Straße
Postfach 10 60 06
5000 Köln 1
Tel. (02 21) 20 29-0
Telex ANZEIGER BONN 8 882 595
Fax 20 29 278

GREECE/ΕΛΛΑΔΑ

G.C. Eleftheroudakis SA
International Bookstore
Nikis Street 4
10583 Athens
Tel. (01) 322 63 23
Telex 219410 ELEF
Fax 323 98 21

ESPAÑA

Boletín Oficial del Estado
Trafalgar, 27
28010 Madrid
Tel. (91) 44 82 135

Mundi-Prensa Libros, S.A.
Castelló, 37
28001 Madrid
Tel. (91) 431 33 99 (Libros)
431 32 22 (Suscripciones)
435 38 37 (Dirección)
Télex 49370-MPLI-E
Fax (91) 575 39 98

Sucursal:

Librería Internacional AEDOS
Consejo de Ciento, 391
08009 Barcelona
Tel. (93) 301 86 15
Fax (93) 317 01 41

Librería de la Generalitat de Catalunya
Rambla dels Estudis, 118 (Palau Moja)
08002 Barcelona
Tel. (93) 302 68 35
302 84 62
Fax (93) 302 12 99

FRANCE

Journal officiel
Service des publications des Communautés européennes
26, rue Desaix
75727 Paris Cedex 15
Tél. (1) 40 58 75 00
Fax (1) 40 58 75 74

IRELAND

Government Supplies Agency
4-5 Harcourt Road
Dublin 2
Tel. (1) 81 31 11
Fax (1) 78 06 45

ITALIA

Licosa Spa
Via Duca di Calabria, 1/1
Casella postale 552
50125 Firenze
Tel. (055) 64 54 15
Fax 64 12 57
Telex 570466 LICOSA I
CCP 343 509

GRAND-DUCHÉ DE LUXEMBOURG

Messageries Paul Kraus
11, rue Christophe Plantin
2339 Luxembourg
Tél. 499 88 88
Télex 2515
Fax 499 88 64 44
CCP 49242-63

NEDERLAND

SDU Overheidsinformatie
Externe Fondsen
Postbus 20014
2500 EA 's-Gravenhage
Tel. (070) 37 89 911
Fax (070) 34 75 778

PORTUGAL

Imprensa Nacional
Casa da Moeda, EP
Rua D. Francisco Manuel de Melo, 5
1092 Lisboa Codex
Tel. (01) 69 34 14

Distribuidora de Livros Bertrand, Ld.ª
Grupo Bertrand, SA
Rua das Terras dos Vales, 4-A
Apartado 37
2700 Amadora Codex
Tel. (01) 49 59 050
Telex 15798 BERDIS
Fax 49 60 255

UNITED KINGDOM

HMSO Books (PC 16)
HMSO Publications Centre
51 Nine Elms Lane
London SW8 5DR
Tel. (071) 873 2000
Fax GP3 873 8463
Telex 29 71 138

ÖSTERREICH

Manz'sche Verlags- und Universitätsbuchhandlung
Kohlmarkt 16
1014 Wien
Tel. (0222) 531 61-0
Telex 11 25 00 BOX A
Fax (0222) 531 61-39

SUOMI

Akateeminen Kirjakauppa
Keskuskatu 1
PO Box 128
00101 Helsinki
Tel. (0) 121 41
Fax (0) 121 44 41

NORGE

Narvesen information center
Bertrand Narvesens vei 2
PO Box 6125 Etterstad
0602 Oslo 6
Tel. (2) 57 33 00
Telex 79668 NIC N
Fax (2) 68 19 01

SVERIGE

BTJ
Box 200
22100 Lund
Tel. (046) 18 00 00
Fax (046) 18 01 25

SCHWEIZ / SUISSE / SVIZZERA

OSEC
Stampfenbachstraße 85
8035 Zürich
Tel. (01) 365 54 49
Fax (01) 365 54 11

CESKOSLOVENSKO

NIS
Havelkova 22
13000 Praha 3
Tel. (02) 235 84 46
Fax 42-2-264775

MAGYARORSZÁG

Euro-Info-Service
Budapest I. Kir.
Attila út 93
1012 Budapest
Tel. (1) 56 82 11
Telex (22) 4717 AGINF H-61
Fax (1) 17 59 031

POLSKA

Business Foundation
ul. Krucza 38/42
00-512 Warszawa
Tel. (22) 21 99 93, 628-28-82
International Fax&Phone
(0-39) 12-00-77

JUGOSLAVIJA

Privredni Vjesnik
Bulevar Lenjina 171/XIV
11070 Beograd
Tel. (11) 123 23 40

CYPRUS

Cyprus Chamber of Commerce & Industry
Chamber Building
38 Grivas Dhigenis Ave
3 Deligiorgis Street
PO Box 1455
Nicosia
Tel. (2) 449500/462312
Fax (2) 458630

TÜRKIYE

Pres Gazete Kitap Dergi Pazarlama Dağitim Ticaret ve san AŞ
Narlibahçe Sokak N. 15
Istanbul-Cağaloğlu
Tel. (1) 520 92 96 - 528 55 66
Fax 520 64 57
Telex 23822 DSVO-TR

CANADA

Renouf Publishing Co. Ltd
Mail orders — Head Office:
1294 Algoma Road
Ottawa, Ontario K1B 3W8
Tel. (613) 741 43 33
Fax (613) 741 54 39
Telex 0534783

Ottawa Store:
61 Sparks Street
Tel. (813) 238 89 85

Toronto Store:
211 Yonge Street
Tel. (416) 383 31 71

UNITED STATES OF AMERICA

UNIPUB
4611-F Assembly Drive
Lanham, MD 20706-4391
Tel. Toll Free (800) 274 4888
Fax (301) 459 0056

AUSTRALIA

Hunter Publications
58A Gipps Street
Collingwood
Victoria 3066

JAPAN

Kinokuniya Company Ltd
17-7 Shinjuku 3-Chome
Shinjuku-ku
Tokyo 160-91
Tel. (03) 3439-0121

Journal Department
PO Box 55 Chitose
Tokyo 156
Tel. (03) 3439-0124

AUTRES PAYS
OTHER COUNTRIES
ANDERE LÄNDER

Office des publications officielles des Communautés européennes
2, rue Mercier
2985 Luxembourg
Tél. 49 92 81
Télex PUBOF LU 1324 b
Fax 48 85 73/48 88 17
CC bancaire BIL 8-109/6003/700

6 426709 410066 426709 410066 426709 410066

NOTICE TO THE READER

All scientific and technical reports published by the European Commission are announced in the monthly periodical **'euro abstracts'**. For subscription (1 year: ECU 60) please write to the address below.

15 CG-NA-15412-EN-C

Price (excluding VAT) in Luxembourg: ECU 7

ISBN 92-826-8226-9

OFFICE FOR OFFICIAL PUBLICATIONS
OF THE EUROPEAN COMMUNITIES

L-2985 Luxembourg