

What are the main factors responsible for this dramatic acceleration in genome sequencing projects? Two dominant themes are apparent: access to massively parallel systems in both sequencing and computing and fundamental changes in the way genome projects are conceived and organised (Venter et al 2003). Early genome projects were large and multi-site in nature whereas many of today's successful projects are conducted in quality-controlled facilities that have smaller more flexible and highly experienced teams.

Apart from *Arabidopsis* and rice there is considerable interest in sequencing other important genomes. At present there is sufficient sequencing capacity available and the realisation of this objective is almost entirely dependent on availability of funds. The abundance of long terminal repeat retrotransposons in plant genomes may present some technical challenges for sequence assembly **but should the CGIAR have a flagship genome sequencing project?**

Expressed –sequence tags (ESTs) of randomly picked complementary DNA (cDNA) clones are widely used as a gene discovery platform particularly for plant scientists (Figure 2). Such material provides the template for genetic marker discovery e.g. SSRs and SNPs and EST are a very important resource for genetic diversity studies. **Gaining access to ESTs is therefore essential.**

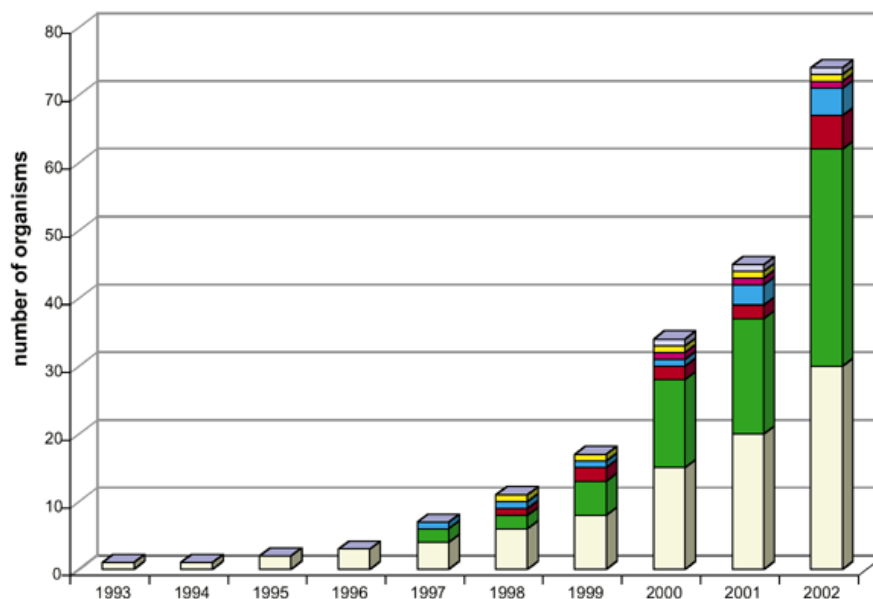


Figure 2. The growth of EST sequencing since 1992.

Stacked columns indicate the cumulative number of organisms with over 10,000 ESTs in GenBank. Animals shown in white, plants in green, fungi in red, alveolate in blue, euglenozoa in purple, mycetozoa (slime mold) in yellow and red algae in grey.

As of the 14th March 2003, dbEST contained 15,897,028 public entries, providing a wonderful resource for gene discovery and exploitation. However, this does not present the whole picture. It may be pertinent to consider the following table (Table 1), which details the Future Harvest Genebank collections held in trust for the world community.

Table 1. Future Harvest Genebank Collections held in-trust for the world community based on agreement with FAO (2002)

CENTRE	CROP	ESTs
CIAT	Cassava	851
	Forages	
	Bean	1399
CIMMYT	Maize	200,799
	Wheat	415,642
CIP	Andean roots & tuber	
	Sweet potato	4,153
	Potato	94,423
ICARDA	Barley	314,884
	Chickpea	469
	Faba Bean	1
	Wheat	415,642
	Forages	
	Lentil	9
ICRAF	<i>Sesbania</i>	
ICRISAT	Chickpea	469
	Groundnut	105
	Pearl millet	93
	Pigeon pea	23
	Sorghum	84,712
	Minor millets	
IITA	Bambara groundnut	
	Cassava	851
	Cowpea	45
	Soybean	308,564
	Wild <i>Vigna</i>	17
	Yam	7
ILRI	Forages	
IPGRI	<i>Musa</i>	437
IRRI	Rice	130,812
WARDA	Rice	

The most striking feature of this table is the disparity between the various crops in relation to available DNA sequence information. This clearly reflects the efforts of the private and public sector which have focussed on a restricted range of species. Based on these preliminary observations there are at least three points worthy of consideration: first, for those species with large amounts of sequence information e.g. maize, wheat, rice, potato etc. there is an opportunity to quantify intra-species sequence diversity to provide a knowledge base that is unrivalled in the history of plant science; comparative genomics/informatics may provide important technology leverage and thirdly there is a clear need for targeted gene discovery for other crop species which are under represented. This could be extended to

include a ‘horizontal genomics’ programme that focuses on non-crop plants which harbour completely novel genes that could have utility in agriculture and biotechnology.

It might be useful to carry out a similar analysis for forestry and fisheries. A cursory scan of dbEST reveals the following entries:

<i>Danio rerio</i> (zebrafish)	294,479
<i>Bos taurus</i> (cattle)	265,998
<i>Sus scrofa</i> (pig)	116,668
<i>Oryzias latipes</i> (Japanese medaka)	103,098
<i>Pinus taeda</i> (loblolly pine)	60,226
<i>Populus tremula</i> & <i>P. tremuloides</i>	56,013

There is an opportunity to raise the value and profile of genetic resources that is unique to the CGIAR.

During the past decade advances in technology have contributed to an unprecedented plethora of data and approaches on which to build investigative studies of biological processes. These include gene-function and validation approaches together with a wealth of information on gene expression and protein function. The plant biology 2010 project (Sommerville & Dangl 2000) outlines a series of representative goals for *Arabidopsis* in which every gene (25,900) will have been subjected to one or more experiments in which the gene is inactivated or over-expressed. Accessing this information, interpreting and relating to putative homologues from other species will be important together with the design of smart experiments to exploit this information in organisms of direct relevance. Furthermore, as more genome information becomes available the tools of comparative genomics will become increasingly important to analyse gene structure, function and evolution in different organismal kingdoms.

The CGIAR centres should consider positioning themselves to be global leaders in the exploitation of this information, to be the agricultural knowledge base for innovative solutions to global problems and providing the experimental framework for rigorous hypothesis testing particularly in relation to quantitative traits that require access to genetic resources, ability to accurately quantify phenotypic and genetic effects and the ability to validate putative quantitative trait loci (QTL) in different genetic backgrounds.

It is predicted that the next decade will lead to acquisition of knowledge rather than information. Organisations such as the Biotechnology and Biological Sciences Research Council (BBSRC) in the UK propose to build biological sciences around three inter-related areas: experimental data, models and tools from which a more ‘predictive biology’ will emerge.

As emphasised in the BBSRC ten-year vision (http://www.bbsrc.ac.uk/about/strategic_plan/), the biological sciences is now extremely ‘data rich’ and this is changing the way biologists work and more importantly think. The quantitative nature of modern biology demands closer collaboration between biologists and bioinformatics experts. A more collaborative approach is needed to capture emerging opportunities but also to gain access to large-scale expensive resources.

Availability of this data has a number of implications at both a strategic and an operational level.

- A need for a different strategic overview and different approaches to biological problems.
- Attaining a balance between laboratory/field-based research programmes and computational biology.
- New skills and competencies.
- Technology interaction, integration and leverage to benefit from existing data.
- New funding relationships and scientific networks.
- A willingness to change the emphasis of research to ensure that new scientific discoveries can be converted into new scientific opportunities for agriculture fisheries and forestry.
- Information and knowledge management to facilitate co-ordination, alignment and integration of R & D.
- A need to attain a balance between discovery science and hypothesis driven research.

More emphasis is also being placed on inter-disciplinary research with exciting discoveries being made at the interface of biology, chemistry, physics, mathematics and engineering. This will require scientists to take a much broader view of biology and be prepared to establish collaborations and engage researchers working outwith their immediate disciplines. Most importantly this will also require strong leadership and a strategic plan that will build on existing strengths, identify key competencies and a focus on acquiring or building new skills. Clearly a major opportunity for the 21st Century will be to integrate the reductionist view of biology with genome wide approaches and whole organism biology. The traditional strengths of the CGIAR centers in the integrative approaches to biology provide a sound framework for future growth but this must be connected to the new concepts, technologies and potential of modern biology. This means that there is also a need for critical mass and scientific depth of effort in important areas particularly genome science. It is also clear that organisations cannot operate in isolation and new relationships will be required to bring new and additional funding, velocity, visibility and impact.

Many countries and organisations are ensuring integration of activities to create 'biotechnology pipelines' that reflect the waves of technology that has permeated contemporary bio-science. Figure 3 shows the path from genotype to phenotype and the role of genomics, proteomics/structural biology, metabolomics and bioinformatics.

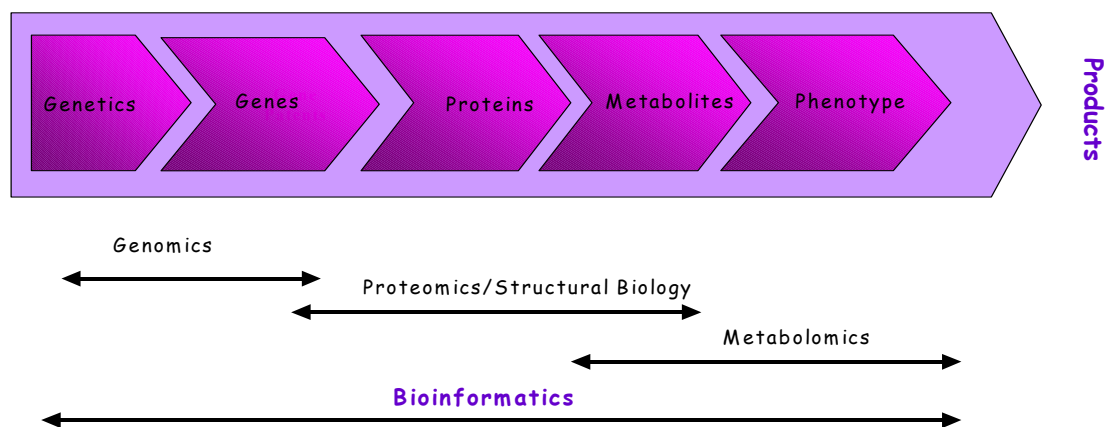


Figure 3. The Biotechnology Pipeline

IMPLICATIONS

Based on this brief analysis a number of clear points emerge that may influence and impact the framework for future developments. These include the need for the CGIAR to clearly identify points of differentiation that will allow the portfolio of activities to add-value. It is also relevant to note that value capture models for Agriculture are different to that developed for medical and pharmaceutical enterprises, presenting both opportunities and challenges. Seeds, cultivars and new agricultural products are the integrators for genome technologies. Capturing the value from genomics will therefore require strong competency in plant and animal breeding and whole organism biology. Opportunities therefore exist for the integration of genomics with breeding to create technology platforms for the conversion of traits to products. **Seeds or propagules provide the delivery system by which innovation in biotechnology can be captured and transferred to the major beneficiaries.**

- The science and technology being pursued is immensely powerful but resource demanding. This will require new forms of scientific collaboration but also new funding relationships and models.
- Large scale well funded programmes have the ability to assume greater risk and a potential for greater return.
- Prioritisation based on a clear understanding of the opportunities and the limitations of the technology. There will be a need for focus and selectivity together with a recognition that resources will need to be targeted to tackle major problems using the best scientific tools. Put simply this will mean addressing fewer problems with more resource to ensure greater impact.
- High quality scientific labour force.

- A research and operational infrastructure that is capable of contributing to and capturing the opportunities emerging from genomics and related disciplines.
- As more organisms are sequenced more emphasis will be placed on understanding genetic diversity and the relationship between sequence diversity and heritable phenotypic differences. Converting the vast amount of information into knowledge represents a major opportunity.

OPTIONS

- Exploit genomes as a source of genes for breeders, biotechnologists and biologists.
- Molecular and biochemical diversity and the modification of the nutritional composition of plants.
- Biodiversity provides the basic raw materials for new discoveries in Agriculture and related areas. In addition in the post-genomics era genetic resources (which represent mutations conducted in nature) will play an important role in validating gene function (functional genomics).
- Creating technology supply chains based on genetics and biotechnology.
- Conventional and enhanced breeding methods are vital.
- Developing Ag-biotech strategies for the delivery of products; comprehensive genetics capability from molecule to seed/propagule to product.
- Reinforcing the concept of the CGIAR enterprise being an ‘end to end’ provider.

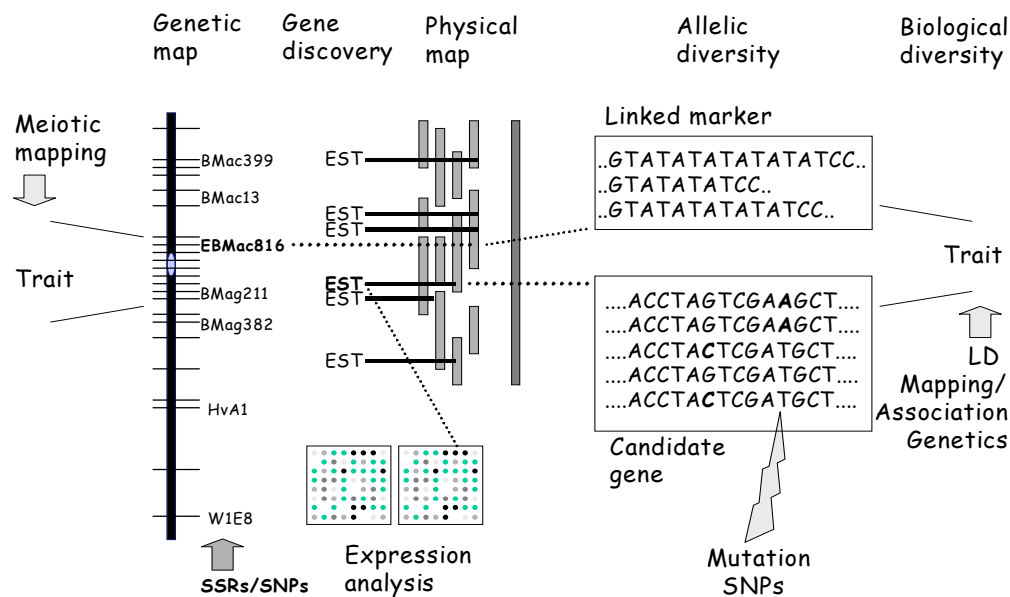


Figure 4. A view of the genome as a source of genes for breeders, biotechnologists and biologists. It stresses the importance of genetics, integrating technology and the vital role of genetic resources in validating gene function and providing novel alleles and genes.

STRATEGIC ISSUES.

- **Increased and co-ordinated efforts in bioinformatics.**
 - o Need to handle mass data volumes and make these available over a distributed geography.
 - o Integrated platform to link all genomic data with agronomic data (phenotypes, genotypes, pedigrees). Visualisation programs.
 - o Create platforms for segregation and population based trait mapping.

- **Strategic resource allocation and capital investment to ensure International leadership in:**
 - o DNA diagnostics programs to support molecular breeding strategies for trait introgression.
 - o Understanding the significance and magnitude of intra-specific sequence diversity and its connection to heritable phenotypic differences.
 - Haplotype based analysis and linkage disequilibrium based association genetics.
 - SNP based association genetics is likely to drive gene discovery for complex traits.
 - o Genomes as a source of genes for breeders, biologists and biotechnologists.
 - Population development to underpin map based gene discovery and genomics.
 - Connecting physical and meiotic maps.
 - Map-based identification of gene function.
 - o EST programs for targeted gene discovery (Horizontal genomics).
 - o New scientific capabilities to ensure that the CGIAR is participating, and in some case leading, globally competitive science and technology deployment.

- **Ag-biotechnology/Genomics needs breeding to capture opportunities. A need for strategic product orientated research.**
 - o Plant and animal breeding as technology platforms for the delivery of products.
 - o Ensure that there are mechanisms/skills to deliver products to those that should and can benefit.
 - o Create technology supply chains that are built around enabling science, germplasm and delivery of public goods.

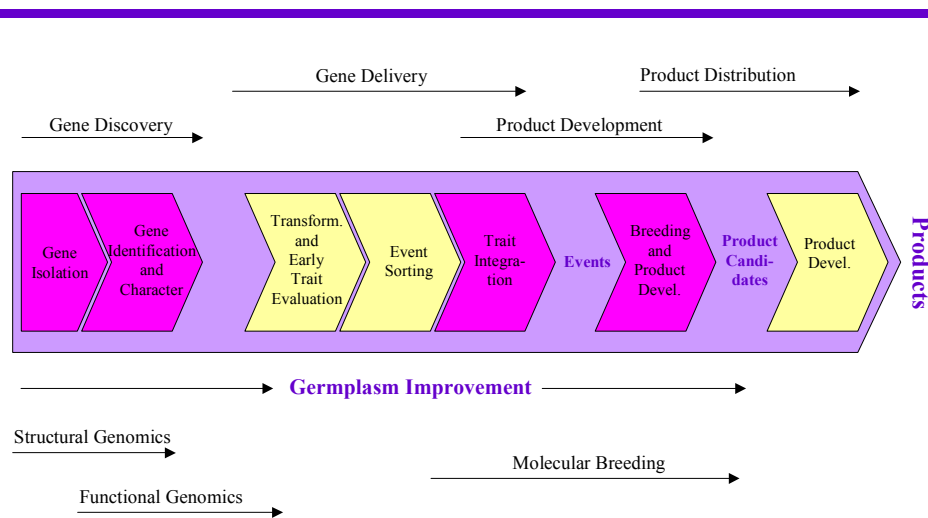


Figure 4. Technology supply chain for the delivery of new products will require integration of different technologies that includes transgenic and non-transgenic approaches.

- **Information based systems will drive future gene discovery and exploitation paths. Hence knowledge of the species and experimental systems will be paramount.**
- **Attracting World-class scientists and training.**
 - o Raising the profile of science and technology and its critical role in solving global problems.
 - o The CG system provides a wonderful example of a ‘reduction to practice lab’.
- **Ensuring that the potency of genome science is connected to the goals of natural resource management and productivity and profitability of smallholder farmers.**
 - o This means that there will be a need to understand the potential and limitations of the technology, giving science and technology more visibility and a full understanding of targets.

CONCLUSIONS

This is the most exciting time to be involved in agricultural sciences (in its broadest context). To capture and develop these opportunities we need to reinforce the importance of science, technology and people. I would advocate that the CGIAR considers a dual track strategy of providing the knowledge base for innovation in agriculture, forestry and fisheries and creating the science and technology supply chains that are necessary to deliver products and solutions.

To quote the Nobel Laureate Sydney Brennan “which type of science to fund is simple: all science is problem driven and should be judged by the importance of the problems posed and the quality of the solution provided.”

REFERENCES

Sommerville, C. & Dangl, L. 2000. Genomics – Plant biology in 2010. *Science* **290**, 2077-2078.

Venter, J.C., Levy, S., Stockwell, T., Remington, K. & Halpern A. 2003. Massive parallelism randomness and genomic advances. *Nature Genetics* **33**, 219-277