CGIAR System-wide Genetic Resources Programme (SGRP)

Report on the International Workshop on Genebanks and Comparative Genetics

Exploiting genome homoeology to enhance utilization of
ex situ *germplasm collections*

held at ISNAR, The Hague, the Netherlands, August 24-27, 1999

Comparative genetics can enhance exploitation of genebank collections. The CGIAR must not lose an opportunity of becoming an important player in the field, by exploiting its own comparative advantages of germplasm management and enhancement. Jointly, the CGIAR and advanced laboratories now have an historic opportunity of working together using the latest and most efficient molecular approaches to ensure greater food security that will lead towards the alleviation of poverty.

Agenda and participants

1. The application of comparative genetics to enhance use and management of germplasm collections was the focus of the international workshop co-organized by IRRI and ISNAR through the System-Wide Genetic Resources Program (SGRP), and hosted by ISNAR. The workshop brought together more than 50 participants (Annex 1): experts in comparative genetics and molecular biology from universities and research institutes in Brazil, China, France, Germany, India, Japan, Malawi, Malaysia, the Netherlands, UK, and USA, and germplasm specialists, molecular biologists, and biotechnologists from 12 CGIAR centers.

2. Workshop sessions were devoted to the application and a technical needs analysis of comparative genetics, and the traits and pathways that would benefit from this approach (Annex 2). Working groups in the cereals, legumes and root and tuber crops discussed research agenda to apply comparative genetics for the use and management of germplasm. Plenary sessions on the management of intellectual property rights and bioinformatics provided background for these discussions.

3. A summary of workshop expenses is shown in Annex 3.

Major findings

4. The overwhelming message from the workshop was that the CGIAR must take advantage now of the latest technologies in genomics research to apply comparative genetics to the germplasm collections that they hold in trust.

5. The principal conclusions from the workshop were:
   - comparative genetics can provide a more precise, unambiguous and comprehensive tool for germplasm characterization than any other.
   - cross-species comparisons will allow identification of the germplasm sources of superior, potentially optimal alleles for specific traits, thus motivating the use of that germplasm.
   - comparative genetics will provide a multi-lateral flow of knowledge between major and minor crops.
   - the CGIAR centers will have to take the initiative to develop comparative genetics research for several crops within their mandates, such as minor millets and legumes. The alternative

will be an ever-accelerating growth in the knowledge (and subsequent use) gap between major and minor crops.
- molecular genetic tools developed for minor crops will attract additional users from the scientific community, leading to a snowballing of knowledge, use and tool development.
- use of comparative genetics will help reposition the CGIAR genebanks for the future and enhance use of germplasm.
- the strong comparative advantage of the CGIAR centers to conserve, phenotype, and use germplasm should be linked with expertise in comparative genetics existing in many laboratories worldwide.
- the investment of the CGIAR centers in comparative genetics will ensure that the results and benefits will continue as international public goods.
- the CGIAR centers should make a system-wide investment in bioinformatics, using a SINGER or CGNET model, to ensure access to major, public bioinformatics databases and tools, and to encourage the standardization that will be required for effective comparative genetics studies.

6. It was agreed that the potential of comparative genetics could best be demonstrated with traits where gene action is simple and well understood, such as resistance to some pests and diseases, submergence tolerance, starch accumulation, nutritional qualities, phosphate uptake, resistance to soil toxicity, weed competitiveness, and flowering response.

7. The participants agreed that comparative genetics would facilitate:
   - the systematic search for useful alleles in germplasm accessions without having to discover the genes for each crop.
   - identification of genetic resources containing useful allele combinations.
   - the enhancement of minor crops based on knowledge and tools developed for the major crops.
   - understanding of the genetics underlying traits (so-called genetic dissection of traits).
   - better understanding of the structure of diversity that will enhance management of germplasm collections.

Recommendations

8. However, the generation of comparative genetics information will require, among others:
   - the development of mapping populations for some minor species, and their curation by the CGIAR centers (mapping populations for major crops already exist).
   - identification of cross-species anchor markers and the development of framework maps for minor species.
   - discovery of gene sequences or expressed sequence tags (ESTs).
   - discovery of allelic variants with primers developed from the gene sequences.
   - investigation of the phenotypic variance for traits associated with allelic variants.
   - use of heterologous diagnostic markers to search for traits across crops.
   - a strong and coordinated bioinformatics base.

9. Comparative genetics provides the potential for trait extrapolation from a species where the genetic control is well understood, and for which there are molecular markers, to a species which has a limited amount of information. For example, rice is regarded as a model for cereal genomics because of its small genome. The similarity of cereal genomes means that the genetic and physical maps of rice can be used as reference points for the exploration of the much larger and more difficult genomes of the other major cereal crops, and be applied to the minor cereals. Conversely, decades of breeding work and molecular analysis of maize, wheat and barley can

now find direct application in rice improvement. Comparative genetics can also be used to locate desirable alleles in genepools close to the target crop so that transfer can be achieved by conventional methods. Once the research base has been developed for the root and tuber crops, and the legumes, the same applications are expected.

10. The research agenda for the different groups of crop species are very similar. However, the opportunities to apply comparative genetics now are furthest advanced in the cereals in which considerable research investment has already been made in crops such as rice, wheat, maize, and sorghum. Research among the legume species is probably the most limited at the present time, but the workshop provided an excellent opportunity for researchers to meet for the first time as a group to address these issues. Outputs from the three crop working groups are listed in Annexes 4, 5, and 6.

11. Without significant investment in the short term, the research gap between the CGIAR centers and laboratories already heavily involved in comparative genetics will widen. The CGIAR will have fewer opportunities thereafter to influence a research agenda to focus on its mandate crop species. Collaboration with advanced laboratories is essential to exploit fully the potential of comparative genetics. However, investment by the CGIAR in comparative genetics will further support technology transfer to and biotechnology in developing countries.

12. It was agreed that pilot projects in comparative genetics should be supported. For cereals, a project focussing on finger millet as a minor crop would potentially demonstrate the power of comparative genetics to build a resource base in under-utilized crops. Steps should also be taken to improve the research base for the root and tuber crops and legumes in terms of comparative genetics.

Follow-up actions

13. A proposal for a system-wide bioinformatics capability was prepared by a sub-group of participants in consultation with centers and advanced laboratories (Annex 7), and a follow-up meeting at the NCGR planned for February 4, 2000 with commitments from CIMMYT, CIP, and IRRI, and CIAT and IPGRI invited.

14. At the Annual Meeting of the Crop Science Society of America held in Salt Lake City in November 1999, there was an informal meeting of some of the workshop participants (Steve Kresovich, Andy Paterson, Tom Hash, and Mike Jackson) and other interested persons such as Masa Iwanaga (IPGRI) and Rob Bertram (USAID). It was agreed that Steve Kresovich and Rob Bertram would look into possible USDA/USAID interest in supporting US institute-CGIAR collaboration in this area. Individual centers also need to take the lead for particular groups of crops, and this would be discussed at the ICWG-GR meeting. It was expected that there would be additional follow-up by some workshop participants at the Plant and Animal Genome Meeting in San Diego in January 2000.

15. The ICWG-GR should analyze the outcome of the workshop and make further recommendations concerning specific projects. The workshop highlighted the advantages of inter-center collaboration to exploit germplasm more effectively. It is important for the CGIAR to find mechanisms to benefit from the technological developments in comparative genetics that support breeding activities across crops.

Annex 1

Workshop participants

Universities	
Villoo Morawala-Patell	Bangalore, India
John Newbury	Birmingham, UK
Stephen Kresovich	Cornell
Michael Thomson	Cornell
Bikram Gill	Kansas State
Robert Zeigler	Kansas State
Günter Kahl	Frankfurt
Peter Winter	Frankfurt
Andrew Paterson	Georgia
Chong-Lek Koh	Malaya
Nevin Young	Minnesota
Jeffrey Bennetzen	Purdue
Richard Visser	Wageningen
David Spooner	Wisconsin
Research Institutes	
Jizeng Jia	Chinese Academy of Agricultural Sciences, Beijing
Jean-Christophe Glaszmann	CIRAD, France
Theo van Hintum	Centre for Genetic Resources the Netherlands, Wageningen
Marcio Ferreira	EMBRAPA-CENARGEN, Brasilia
Katrien Devos	John Innes Centre, Norwich, UK
Jo Dicks	John Innes Centre, Norwich, UK
Noel Ellis	John Innes Centre, Norwich, UK
Rowland Chirwa	Min. of Agriculture & Irrigation, Malawi
Bruno Sobral	National Center for Genome Resources, Santa Fe, New Mexico, USA
Nori Kurata	National Institute of Genetics, Japan
Robbie Waugh	Scottish Crop Research Institute, Dundee, Scotland, UK
CGIAR	
Daniel Debouck	CIAT
Joseph Tohme	CIAT
David Hoisington	CIMMYT
Bent Skovmand	CIMMYT
Suketoshi Taba	CIMMYT
Meredith Bonierbale	CIP
Marc Ghislain	CIP
Michael Baum	ICARDA
Kamel Chabane	ICARDA
Sripada Udupa	ICARDA
Ian Dawson	ICRAF
Paula Bramel-Cox	ICRISAT
Thomas Hash	ICRISAT
Jacob Mignouna	IITA
Quat Ng	IITA

CGIAR cont'd	
Jean Hanson	ILRI
Harmanjeet Jamnadass	ILRI
Nina Dudnik	IPGRI
Geoffrey Hawtin	IPGRI
Toby Hodgkin	IPGRI
Michael Jackson	IRRI
Graham McLaren	IRRI
Joel Cohen	ISNAR
John Komen	ISNAR
Pierre-Louis Amoussou	WARDA
Robert Guei	WARDA
Maria Antonia Fernández M.	TAC

Annex 2

Program of the International Workshop on *Genebanks and Comparative Genetics*

Genebanks and Comparative Genetics
An International Workshop
ISNAR, The Hague, the Netherlands
August 24-27, 1999

TUESDAY 24 AUGUST	
08:30-10:00	Getting started Welcome to ISNAR Dr. Howard Elliott, Deputy Director General Background to the Workshop Local arrangements Dr. Joel Cohen (ISNAR) Chair – Inter-Center Working Group in Genetic Resources (ICWG) Goals & Objectives of the Workshop Dr. Mike Jackson (IRRI) The SGRP and CGIAR genebank collections Dr. Geoff Hawtin , Director General (IPGRI) Leader – System-wide Genetic Resources Program (SGRP)
10:00-10:30	Coffee break
10:30-12:00	**KEYNOTE ADDRESS AND DISCUSSION** Dr. Jeff Bennetzen (Purdue) *Comparative genetics/genomics for the discovery, study and use of allelic diversity and 'optimal' alleles*
12:00-13:30	Lunch
13:30-17:30	**EXPLOITING COMPARATIVE GENETICS** Concurrent Working Group discussions on the same issues: Objectives Potential benefits Key unknown benefits Technical needs Limitations Traits Pathways Working Group Chairpersons: Group 1 - Dr. Merideth Bonierbale (CIP) Group 2 - Dr. Katrien Devos (John Innes Centre) Group 3 - Dr. Joe Tohme (CIAT) Group 4 - Dr. Steve Kresovich (Cornell) Coffee break 15:15-15:45
17:45	Cocktail, co-hosted by ISNAR and the SGRP

WEDNESDAY 25 AUGUST	
08:30-10:15	***EXPLOITING COMPARATIVE GENETICS***
	Continuation of Working Groups 1 - 4
10:15-10:45	Coffee break
10:45-12:30	***REPORTS FROM WORKING GROUPS - PLENARY***
12:30-13:30	Lunch
13:30-15:15	***BIOINFORMATICS PANEL***
	Panel members: Dr. Jo Dicks (JIC), Dr. Graham McLaren (IRRI), and Dr. Bruno Sobral (NGRC)
	Database standardization
	Nomenclature
	Technical needs and capabilities
15:15-15:45	Coffee break
	IPR ISSUES AND MANAGEMENT OF COMPARATIVE GENETICS – PLENARY DISCUSSION
	Discussion Leader: Dr. Joel Cohen (ISNAR)
	Synthesis of IPR issues to consider when developing concept notes
19:30	Workshop dinner
THURSDAY 26 AUGUST	
08:30-12:30	**DEFINITION OF RESEARCH AGENDA FOR APPLYING COMPARATIVE GENETICS TO THE MANAGEMENT AND USE OF GENEBANK COLLECTIONS**
	Crop working groups
	Cereals — Dr. Bob Zeigler (Kansas) - chair
	Solanaceae — Dr. Robbie Waugh (SCRI) - chair
	Legumes — Dr. Günter Kahl (Frankfurt) – chair
	Coffee break 10:15-10:45
12:30-13:30	Lunch
13:30-17:30	**DEVELOPMENT OF CONCEPT NOTES FOR RESEARCH PROJECTS AND COLLABORATION**
	Continuation of crop working groups
	Coffee break 15:15-15:45
FRIDAY 27 AUGUST	
08:30-09:45	**REPORTS FROM CROP WORKING GROUPS**
09:45-10:15	***Coffee break***
10:15-11:45	**SYNTHESIS AND NEXT STEPS**
11:45-12:00	**CLOSURE**

Annex 3

Workshop on Genebanks and Comparative Genetics
ISNAR, The Hague, the Netherlands
24-27 August 1999

Summary of Expenses

Particulars	**Amount US$)**
▪ Airtickets	50,258
▪ Hotel accommodation	16,003
▪ Per diem and local travel	11,409
▪ Miscellaneous expenses *Recording system, printing costs, group picture* *Telephone calls, support staff services, transcriber fee* *Hospitality costs*	10,242
▪ Biodiversity in Trust (50 copies)	2,719
Total Expenses	**90,631**
Total Budget	***100,000.00***
Balance	**9,369**

Annex 4

Cereals Working Group		
Chair:	Robert Zeigler	
Members:	Pierre-Louis Amoussou	Thomas Hash
	Jeffrey Bennetzen	Toby Hodgkin
	Kamel Chabane	Michael Jackson
	Katrien Devos	Jizeng Jia
	Jo Dicks	Chong-Lek Koh
	Marcio Ferreira	Villoo Morawala-Patell
	Bikram Gill	Andrew Paterson
	Jean-Christophe Glaszmann	Bent Skovmand
	Robert Guei	Suketoshi Taba
	Graham McLaren	Michael Thomson
	Jean Hanson	

Summary of discussion:

I. Basic Science Questions

1. Genetic distance
2. Markers for traits
3. Extrapolate marker information from mapping populations to accessions
4. Integration of mapping information to target-specific traits
5. Study diversity of neutral markers as an indication of allelic diversity gene evolution across spp.
6. Comparative mapping – to follow trait markers across genomes
7. Ease of additional mapping – SSR markers
8. Access to standard mapping populations
9. Nature of allelic diversity and its relationship to functional diversity
10. DNA polymorphism distributed across loci
11. Linkage disequilibrium mapping (very dense maps) info on population structure phenotype info.
12. From cloned genes can we look for additional diversity in accessions?
13. EST libraries (trait-based)
14. Genotyping accessions with clones genes of known function

II. Genebank Questions

1. Guide conservation decisions (information on extent and distribution of diversity) - emphasis on minor crops and wild relative
2. Relationship between diversity measures from neutral markers and trait-based markers
3. Research strategies for finding useful alleles.
4. Quantify variation for phenotypic characters and research genetic control
5. Knowledge of population structure for efficient conservation decisions
6. Guide collecting strategies
7. New genes and novel traits
8. Understanding gene regulations by comparing across species
9. Tools to improve genebanks
10. Increase accessibility of genebanks
    - more better information on useful variation

- different views of germplasm structure
- evolutionary studies to predict regions of useful variation

11. Increase use of germplasm in collections
12. Information on heterotic groupings – support breeding strategies (pre breeding) within and across spp.

III. Strategy

A common key to many of the questions outlined in the first two sections is the identification and quantification of useful allelic diversity in a way that allows comparison across species. The following steps are required to achieve this:

a) Using a set of common anchor probes and new or existing mapping populations, develop a map that is comparable to other grass species.
b) Increase the density of the map with a variety of convenient marker types including markers developed from genes cloned for grass species, markers known to be associated with important traits in any grass species, expressed sequence tags and anonymous markers spanning the genome. These markers may be species specific but should be cheap and easy to locate.
c) Develop, maintain and curate these mapping populations in conjunction with the germplasm collections.
d) Fingerprint subsets of the collection for each species using the set of convenient markers. Subsets being carefully chosen to represent the range of genetic diversity in the collection.
e) Publish the mapping and fingerprinting information as an accessible bioinformatics resource.

Once this resource has been developed it can be used for allele mining, extrapolation of trait variation, identification of putative QTLs and analysis of diversity. Allele mining would be achieved by looking for allelic variation in clone-derived markers and phenotyping the variants. Trait variation can be extrapolated to new species by phenotyping accessions that are polymorphic for markers near trait-based markers linked across the consensus map. Putative QTLs could be identified by looking for linkage between traits and markers in the target species and then checking for known QTLs in other species in homologous regions. Diversity studies, used to target collection and assist in management, would be based on analysis of the anonymous markers.

Bioinformatics requirements

An efficient and flexible model for the bioinformatics component would be to develop a common database structure for the fingerprinting data that would be curated independently by each collection. Then to develop (in partnership) or acquire access to a common interface to international bioinformatics resources, including the genome databases where the species maps would be maintained. Access to, and development of software for comparative genomics would be a central, but long term component of the project requiring coordinated collaboration with ARIs working in the field.

Hypothetical application

Several traits were considered as candidates for the development of a pilot project. These included:

a) submergence tolerance
b) P uptake efficiency
c) Seedling vigor
d) Disease resistance

e) Tolerance to aluminum toxicity
f) Drought/salinity tolerance
g) Photoperiod response
h) Micronutrient response
i) Grain quality
j) Brown midrib

The most promising and visible of these were thought to be **submergence tolerance**, **photoperiod response**, and **blast resistance** for **finger millet** (*Eleusine coracana*). This led to the following proposed hypothetical study:

Use knowledge of genetic variation in one species to detect potentially useful allelic variation in a target species.

a) Select test traits which are of interest across a wide range of grass species and where genetic variation is known and understood in one or more 'source' species. Examples would be photoperiod sensitivity, submergence tolerance and blast resistance.
b) Using markers developed from cloned genes or known QTLs from source species, fingerprint subsets of several target grass species including finger millet as an example where little information is currently available. Design a database structure to manage this information and load it into separate databases for each test species.
c) The discovery of polymorphism for these markers in target species would be an indication of transferability of information on genetic information across species.
d) Phenotyping of variants with respect to test traits could demonstrate potentially useful allelic variation.
e) More useful information would be available in target species where mapping had been done so that regions showing allelic polymorphism could be compared with homologous regions in other species where QTLs for the test traits may have been identified. Hence a useful extension of the pilot study would be to undertake the mapping exercise in finger millet for the anchor probes, the test markers and other convenient markers chosen to cover the target genome.
f) An analysis of molecular and morphological diversity could be used to indicate the efficacy of morphology based subsets and could indicate collection as management strategies for target species.

Identification and quantification of useful allelic diversity (using a common set of both neutral and location-based markers and function-based markers)

1. Common anchor markers on each species map
2. Increase map density with convenient markers (could be species-specific)
   - SSRs for within-species analyses
   - ESTs for trait-based markers and across species
3. Develop/curate and maintain mapping population (Curate includes automatic map improvement functions)
4. Fingerprint subsets (representative) of the collection with the chosen markers
5. Bioinformatics data
   - Common structure for new data
   - Common interface to international bioinformatics resources (access to and from)
   - Develop/access user-driven software for comparative genetic analyses

Annex 5

Solanaceae and Other Roots and Tubers Working Group		
Chair:	Robbie Waugh	
Members:	Maria Antonia Fernandez	Jacob Mignouna
	Meredith Bonierbale	David Spooner
	Nina Dudnik	Joseph Tohme
	Marc Ghislain	Richard Visser
	Stephen Kresovich	

Summary of discussion:

Options

- Common maps and mapping populations, but
  - ⇒ what cross?
  - ⇒ what ploidy?
  - ⇒ single, two or multiple.
    - BAC libraries derived from mapping population.
- CG centers to hold a standard population and also distribute germplasm and DNA.
- Identify sets of anchored markers for mapping.
- Characterizing the germplasm in as many ways as possible and making all this information consolidated and easily accessible. How do we bring together all the data that is accumulating to get something of value and something that will be an aid for using the germplasm collections? Need an important system that makes all the information available.

First steps

Option 1

- Map information markers that can be compared across species.
- 3-D map at each locus has allele frequencies from all the germplasm.
- Locate major genes and QTLs on this same map, and use it to screen germplasm.
- Next, identify haplotypes; identify by descent; locate ESTs on the map instead of re-screening all the materials for variation at that locus, and use haplotypes/variation in the flanking area.
  - ⇒ mapping populations
  - ⇒ loci, markers
  - ⇒ diversity at marker loci.

Option 2

- Germplasm collection (the material that displays the phenotype)
- Germplasm → phenotype → maps
- Have to characterize the germplasm for phenotype → ESTs → model systems
- Expression association of gene data system with phenotypes
- Hypothesis

Goals

- To make collections more accessible, usable, and valuable.

- Who is the end-user? This leads to many other questions and priorities that differ by each user.
-

For example:

- Breeder: - allelic diversity, linkage information, positions on maps
- Natural products seeker – cares little about maps

Which markers are most informative?

- One viewpoint: robust, repeatable results – sequence-based, not hybridization or amplification-based; SNPs or multi-allele SNPs or SSR
- Another viewpoint: SSRs are too hyper-variable don't cross species well on even cross-sample in diverse germplasm

- Criteria for choosing ANCHOR markers (genotypic markers, non-trait specific from any allele tissues):
  ⇒ map distribution
  ⇒ informativeness – polymorphism

Hypotheses/suggested priorities

- maps are valuable
- CG centers maintain and distribute mapping populations
- maps should contain a given number of anchored markers: sequence-based, usable across a broad range of species, linkable to growing genomics information, have a definite level of polymorphism.
- use these maps and markers to study genetic loci, place ESTs, identify shortcuts to finding best material for further study.

Disease resistance

Phytophthora: genes we know are induced information from other systems

Most of the traits we are interested in are not single-gene traits. How do we construct an approach to allele mining?

⇒ trait ESTs/markers that are sequence-based
⇒ mapped
⇒ variation at these sequences
⇒ a way to structure → superimpose → genotype information → collection → QTL data and linkage disequilibrium information.

Phenotype information → Genes → biology

Always add new markers, when you clone a new gene, add it in and then get allelic variation data.

ESTs (and maps) for minor/orphan crops

Include the parents of mapping population to generate SNPs.

1. Develop, maintain, distribute "core" consensus mapping populations at an international center
   (a) Potato: BCT (130 + 130 lines)
   SH x RH (250)

Others: Piurae)
PD) keep parents at CIP as well
CE)
Sterling cross)

(b) Sweet potato
(c) Cassava
(d) Yam (tomato and other Solanaceae if eventually becomes relevant.)

2. Develop, maintain, acquire, map gene/sequence-based anchor markers which are
   a. Distributed
   b. Informative
   c. Single-copy
   - Identify mapped EST/DNA sets
   - Identify reference sets of genotypes
   - Gene cloned, sequenced, alleles identified
   - cDNA – develop primers and test amplification over wide range of species; amplify products from a reference sets, sequence, a defined sets of alleles.
   - Keep sequencing new ones until new alleles plateau
   - Is there a potential problem with work done by CIP using pre-CBD germplasm of Peru claiming control/ownership?
   - IPR related to genes used to develop markers?

3. Genotype the collection
   - Using high-throughput methods not necessarily the entire collection, maybe a structural set.
   - Keep in mind other (non-SNP) tools (such as organellar markers, etc.) that maybe more appropriate for certain questions, or differentiation at particular taxonomic levels.

4. Incorporate 'trait' alleles into the genotyping sets. This also includes the 'allele mining' idea.
   - Integrate information on mapping population: SolGenes, Wageningen, SINGER. Also be able to compare genetic physical map, include raw data.
   - ESTs → generated for orphan crops → get all the raw data too → annotate (Blast, e.g.) generated by some other high-throughput lab → Where and how to store? (publicly-available data; fed into system? Where are they and in what format?).

5. Generate SNPs and map them
   - multiple sequence alignment viewer which links into EST sequence
   - link to map location

6. What is available? Current resources (recent papers on colinearity, published in Genetics)
   - Solanaceae:

   <u>*Genomic resources*</u>

   Well-established genetic colinearity tomato-potato-pepper

   Soon-to-be: 100,000 tomato ESTs
   55,000 potato ESTs
   300 tuber development ESTs (Wageningen)
   500 tuber ESTs (SCRI)
   200 ESTs of P. infestans and Erwinia-infected leaves (SCRI)
   eggplant ESTs (Tanksley)

   Soon: expression arrays
   also ESTs of *P. infestans* itself

BAC libraries: potato, tomato
Micro-colinearity with *Arabidopsis* (Tanskley)
Clones genes from pathways, e.g. from starch breakdown/resynthesis
R gene clusters/loci cloned and analyzed
PTO in tomato, RX potato

Information resources

SINGER
SOLGENES
CROPNET
GRIN
APIC (Assoc. Potato Intergenebank Coll.)
ICIS for pedigree management
TGSC (Tomato Genetic Stocks Center, Davis, CA)

Genetic tools

Maps
Major gene and QTL positions

Genetic stocks – mapping populations

- Potato, *Alstroemeria*; arrowroot; tomato; peppers; eggplant; cassava; yam; sweet potato
  Mutant lines (tomato, potato)

- Cassava - synteny with rubber
  Map
  ESTs: post harvest deteriorated material
  BAC library (+2)
  Cassava Db
  SINGER
  Some genes and QTLs mapped
  Genomic and cDNA library
  Cloned PTO-like genes from rubber
  RG-like genes from rubber

- Yams
  Rudimentary map – AFLPs, isozymes
  Genomic and cDNA libraries
  SINGER

- Sweet Potato
  1 published map, low density, RAPD
  2 AFLP higher-density maps in progress @ CIP
  Genes and QTL
  Mapping genes to anchor maps

Annex 6

Legumes Working Group		
Chair:	Günter Kahl	
Members:	Michael Baum	Harmanjeet Jamnadass
	Paula Bramel-Cox	John Newbury
	Joel Cohen	Quat Ng
	Rowland Chirwa	Sripada Udupa
	Ian Dawson	Peter Winter
	Daniel Debouck	Nevin Young
	Noel Ellis	

Summary of discussion:

- Survey of legume species important in agriculture (crops, forage, trees) – ICARDA, CIAT, IITA, ICRISAT.
- Survey of traits – cold, disease, drought, architecture, pest photoperiod, salt, nutritional quality, yield, N-fixation, P-deficiencies.
- Question target, "one-center", vs. global system-wide proposal? (Genomics vs. all-legumes….) What about funding?
- What is common, important, required throughout legumes? Common markers with common map(s) for every crop. Diseases important throughout legumes. Comparability through ESTs and/or SSRs other?
- Connecting: Centers → Advanced Labs
  - Genebanks → Molecular Techniques
  - Lines
  - Phenotypic/Scoring.
- Comparative advantages, division of labor, timing of linked action.
- Should "centers" take a lead role in molecular work?
- Examine new options for marker techniques.

"Proposal"

- Generate 100 STMS = SSRs (per species); 20 transferable (some already available)
- Test for synteny (RFLPs cross species well; but rarely polymorphic).
- Markers for diversity
- Experimental observations – go out to wild, 2 chickpea, SSR markers fall off to ~50% amplication; pea – 48%, lentil – 38%, common to 29% (Winter).
- Localize RGA onto maps (like RKs, NBS-LRRs) through conserved domains.
- Follow up with create physical libraries (BAC0 as gene repository for major legumes (several) do at one for efficiency (with emphasis on resistance genes).
- Should emphasis be focused toward characterizing germplasm collections (rather than focus on maps) (Ellis).
- Facilitating transferring knowledge from models to "minor" species should be goals (Newbury).
- Taxonomy (Ellis).

- What is important in genebank that which is different?

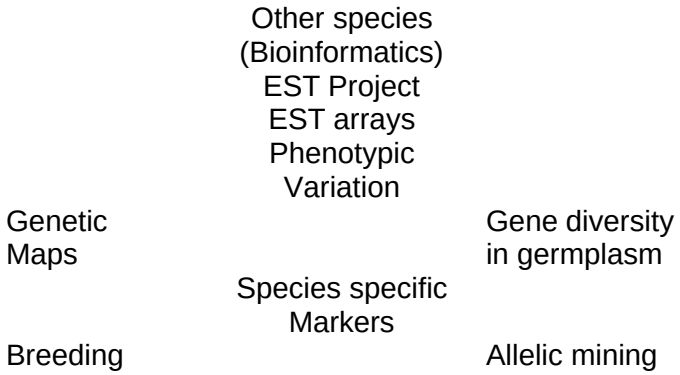
- RGAs in chickpea. Could be useful candidate markers among different legume (in combination with SSRs) (Winter).
- Between “large” vs. “small” project need to know traits (Dawson).
- No, lining up the maps is the key, then information will flow (Newbury).
- Can’t we do both? (global and individual components).
- Need to emphasize the comparative advantage of big group working together.
- Need to remember centers to commit to individual project(s) often get subtracted from core funding.
- In the past – simply making a comparative map is not a priority.
- This meeting is novel in sense of getting to see people from other CG-Centers.
- System-wide initiative (SGRPs) can potentially address concerns about legumes.
- (Lack of) crop specific tools (so, rely on spillover).
- Already has seen benefits from sorghum and millet as part of cereal genomics. Legumes are under-researched – gives strong rationale.
- Linking centers is desirable; Visibility of legume-group coming together; Must be funded over/above core funding (Ng).
- “Integrated Gene Management” is major focus. Interfacing with other initiatives in CG.
- Examples of “successful” comparative genomics; So many labs now interested in sorghum, developing maps etc. (Bramel-Cox).

Comparative genetics among *Striga* resistance in taxa

- EST libraries, comparative at level of sequence. Every crop is going to need reliable markers to use for germplasm/mapping – coordinate through informatics. Really ends up being a “cluster” of projects (Ellis).
- Agree – comparative through informatics; but also, by having a common “public” face and single agency for negociating (sequencing) project.
- ESTs not very useful for maps (SSRs are better) (Winter).
- Will need to have core markers, one way or another (Ellis).
- Less emphasis on genetic mapping, more on ESTs to use (Udupa).
- Maybe it would be possible to “sell” just the importance of developing tools for legumes (not emphasize traits) so legume genetics don’t fall further behind (Bramel-Cox).
- Don’t forget – “Gene Bank” – what can we use genomic tools to apply to genebanks, especially for legumes (distmet from cereals). How can we find “interesting” genes within collection (Ng).
- Let’s not concentrate on details of particular marker technology. Importance – what are the tasks /goals. Goal is allelic mining; Task is how you do it. Lets draw up scheme how to put it all together.
- Could activity be maps in groundnut, pigeon peas; is transformation working? What defines a land race? Look at commonalties among mandate legumes; across crops, cross-feeding from information; connect to databases, get researchers on soyabean interested; survey collections; areas that have under-representation; enhancing transfer from wild species (including MAS) (Bramel-Cox).
- Connect these great markers together with population that are already being developed (Baum).
- Where does comparative mapping fit in? Remember to emphasize system-wide component? Out of individual agendas (opened by marker) toward global comparative legume-mapping (Debouk).
- Won’t know exactly what will come from comparative genetics until you see it (Bramel-Cox).

EST Management / Bioinformatic Populations

- One genomic system that can readily be used involves resistance genes (candidates) (Winter).
- Resistance gene sequences (globally) would be a useful tool to include among legume-genomic tools (Young).
- Already have developed Cicer NBS-LRR (also *Phaseolus*, *Glycine*, *Medicago*) – common R-gene markers (Kahl / Winter).
- Lab's interest in leaf development (connect to flower development). Two successful candidate gene. RGAs could be involved in N-fixation and mycorrhizal. Be realistic – linkage is easy, actually cloning is difficult (Ellis).
- Could use legume-interesting gene types that form basis for which genes to pursue (Winter)
- "Feedback Loop" where genebank informs development of DNA markers informs genebank, etc. (Young).
- Once we decide what to do, find the expert for each area; comparative advantage; need to set up management structure for project/program (Winter).
- Present a "scheme" describing the way elements integrate (Ellis)



EST Projects

- Value of EST libraries, EST (gene) arrays to look in germplasm collection, segregating populations (Ellis)
  - ⇒ EST-basically an analytic tool (expression pattern)
  - ⇒ Starting point for searching for allelic
  - ⇒ Starting point for useful mapping markers.
- Example of "Expression Fingerprint" for some phenotype (race-independent resistance) then use to screen germplasm, even transfer fingerprint pattern to (expression pattern) in related species (Young).
- Idea of use ESTs as anchor points (on map) and targeting related families of genes (e.g. stress-related) (Winter).
- Need to choose – one from different branch; could be used to find species-specific genes (could then screen) Potentially, you can target specific stages treatments for mRNA-isolation (counter-argument, don't know enough about biology, should take everything to be sure). (Young).
- ESTs from each clade – test the usefulness and feasibility – temporary state of affairs – will eventually need to move to species-specific ESTs in medium/long term (Ellis).
- Need to stay aware of difficulties in cross-species interpretation (Dawson).

- May be necessary to add (*Arachis* groundnut) EST library because it is so important. The importance of choosing the accessions; which tissues to sample (Ellis).
- Importance of standardization in preparing cDNA libraries – can they be made at the CG-Centers (Young).
- Common crops between centers (e.g. lentil in ICARD & ICRISAT) (Baum).

Populations

- Defining populations - population size, traits, etc. (Kahl).
- Easier to work on population more diverse that breeders are used to using (interspecific). Use RILs whenever possible. Segregate for most important pathogen(s) for example; other interesting traits (Winter).
- Wary about "requiring" interspecific. Perhaps germplasm can help guide choice of parents' diversity (Ellis).
- Markers (eventually) need to work for breeders (Udupa).
- Compromise – wide-cross but intraspecific (Baum).
- Centers already have several pops (including disease resistance) or drought tolerance (Debouck).
- How about a "common" BC-inbred project – shared among different legumes – that focus on a common story (drought) (Young).
- May be difficult to apply in legumes (because it is so difficult to gets lots of seeds) (Baum).
- In terms of population, better hope that best choices were made in past. The particular crop will also impact the nature of the crosses/parents that ought to be made (Bramel-Cox).
- Diverse, but not too diverse (Dawson).
- Which traits (if any) should be targeted. Seems like consensus on drought. Disease resistance (Kahl).
- Still interested in N-fixation. More and more N-mutants (Baum).
- Of course, *Cajanus* utilization is drought-tolerant, so it is not an issue here. (Of course, it would be interesting to look at genes that make *Cajanus* unique) – compare *Cajanus* expression vs. *Glycine* (Bramel-Cox)
- Other group (cereal) – discussed drought stress and concluded that this is difficult to score/study. More impact from social/economic approaches (Cohen).
- Leave Centers/Breeders to decide on which population traits (Kahl).
- Can't we agree on 1-2 traits universally scored and uniformly measured – as basis for comparative mapping (Debouck).
- How about earliness (easy to measure and important) (Winter).
- Comparative map will develop from use of common/anchor markers – loaded into data warehouse and coordinated through (Young).
- The importance that everyone agrees to phenotype, including common methods, tested in multiple locations (Young).
- This also emphasized the system-wide character of the project (Debouck).
- Could be an opportunity to involve national program in an interesting, advance program (Winter).
- Most appropriate to plant lines in ICRISAT, ICARDA, CIAT & IITA (some sharing may not be possible) (Baum).
- How about yield? (Winter).
- Importance to bring in physiology/physiologists into the process (Newbury).
- How about using microarray-expression data as basis for QTL mapping (Young).

- It will be important that EST-based screening of a germplasm collection be simple, non-gel based (Ellis).
- Importance of a strong training component, visiting lab programs. Can all CG-labs carry out the effort of things like making cDNA libraries (Jamnadass).

Management

- Who else, among advanced labs, would be appropriate to bring in? (Kahl).
- Honenherher (cereals) – H. Geiger (ICRISAT), Canada / Montreal; ACIAR, Queensland, Belgians, Paul Gepts (Bramel-Cox).
- U.C. Davis, University Paris (CIAT) (Debouck).
- ACIAR (*Acacia*) – University of Queensland; molecular biology of relevant microbes (Jamnadass).
- Germany, WSU (Muehlbauer) (ICARDA) (Baum).
- JI and cowpea; SSR for cowpea (in "Canada"); agreement with NGO-Cambia (R. Jefferson) (Ng).
- How about a "pre-grant" to carry out better more extensive planning. (But then what have we accomplished at this meeting so far. Will CG pay for additional planning?) (Young).
- You might want to plan in an early review to bring group together, potentially bring additional folks into project (e.g. Muehlbauer) (Ellis).
- What is the intended outcome of this process: One proposal? Three? (Debouck).
- Immediate Action vs. Medium-Term: for example, a CG center, status of populations, etc. Everyone could e-mail or distribute their expertise. Date to have material distributed. Potential referees? Potential funding agency? (Winter).
- Makes it difficult to come up with concrete plan; timeline; budget (Kahl).
- People go back, to check on agree-upon issues, and then distribute information to entire group, proposal-author (Debouck).
- What report do we make at this meeting. Inter-Center Group, TAC in October? (Ng).
- How about 5-page executive summary with short time line (Young).
- Sort a vertical 'technology' for legumes we will work on. The "horizontal" dimension needs to be explored further (in the proposal). Strengthen activities (based on comparative genetics) applied to legume crops (Debouck).
- Utilize information from "model" in less studied species/crop (Winter).
- Seems like a lot of focus on tools – rather than applications to germplasm – suggest a stronger component about applications and enhancement. Create a target of what we want to accomplish (*vis-à-vis* genebanks in the end) (Paterson).

Conclusion

Titles:

- Coordinated mapping and gene discovery in legumes using comparative genomics.
- Accelerating gene discovery and utilization in legumes using comparative genomics for food security.
- Improved exploitation of the legume genepool by comparative genomics.
- Food security requires improved gene discovery in legumes.
- Comparative genomics and research networks: enhancing the management and utilization of legume genetic resources in the CG system.

Output:

- Tools for generating information-rich germplasm; allele discovery (accession identification/description).

Tool Box Components:

- Syntenic relations
- Commonality of genetic pathways (traits)
- Transferable skills
- Predictive gene discovery
- Technologies to facilitate legume improvement

Background

- To facilitate breeding
- Highly informative, user-friendly legume-wide markers for genetic and diversity analysis
- High-throughput, species-specific markers
- Clade-specific EST libraries
- Harvest legume resistance gene candidates
- Trained personnel and genomic expertise at CG centers
- Comparative mapping and coordinated gene discovery in mandated legumes.
- Central data warehouse for comparative legume informatics
- One big project training component – to transfer cDNA libraries made by Centers (?); another kind of planning.

Annex 7

Draft Concept Note

Bioinformatics

The CGIAR centers have gathered a huge resource of phenotypic data around their germplasm collections and crop improvement programs. Research in molecular biology, genome sequencing, functional genomics, and comparative genetics are producing mountains of new genomic data. Bioinformatics is essential for the management, integration and analysis of phenotypic and genomic data if the promise of molecular biology is to be realized.

New discoveries in comparative genetics indicate a high degree of conservation of genetic material across the genomes of many species. This applies in terms of gene order and gene structure and has important implications for our ability to translate findings in molecular biology in one species to others. This process will not be possible unless the bioinformatics tools are also compatible across species.

Numerous research initiatives all over the world are collecting genomic data. These are often made available for bioinformatic analysis in public databases. The task of linking these data resources, integrating our own contributions and analyzing the product is too great for any center to handle. Furthermore, people with skill and experience in this new and rapidly changing field are rare and dispersed.

The CGIAR centers have a unique role to play in the design and deployment of a bioinformatics system due to the their mandate to extend research and technology to developing countries for the benefit of resource poor farmers and consumers. If our research partners in NARS are to bring the benefits of molecular biology to their own work, the CGIAR needs to make sure that bioinformatics tools are available to them.

Compatibility, access to resources, operational efficiency and the CGIAR mandate require that the centers work together with advanced research institutes and NARS partners to develop, deploy and extend an integrated bioinformatics system.

The Project

A collaborative project between CGIAR centers, and the National Center for Genome Research (NCGR) is proposed with the following aims:

- Develop bioinformatics capacity and capabilities of centers and NARS partners.
- Coordinate development of compatible bioinformatic infrastructure across centers.
- Establish links to NCGR data warehouses for integration, analysis and publication of CGIAR data.
- Promote use and develop state-of-the-art bioinformatic tools for all molecular biology and genomics initiatives at participating CG centers and NARS partners.

Outputs

- Laboratory information management system in each participating center including integration with the International Crop Information System and the System Wide Information Network for Genetic Resources.
- Management and analysis of gene expression data including development of a policy and strategy on standards of annotation. Analysis and visualization tools for expression

array data are already being developed at NCGR and we would have an opportunity to use these tools and influence their development.

- Coordinate access to genome sequence data. Develop links with international sequencing projects and present acceptable solutions to technical constraints and commission development of linkage tools if necessary.
- Interfaces to link CG data to public data warehouses where they become accessible to bioinformatic analysis. This includes policy and technical components since it depends on timing and methodology of publication.
- Coordinate basic training of molecular biologists. Hosted through biotechnology networks with NCGR providing advice and resource persons as required.
- Project-based, specialist training for CG staff to visit NCGR to learn specific methodologies and develop new tools with NCGR staff. Examples would be learning to use new gene product and pathway tools or expression array viewers.

The Partners

The project would ideally involve all centers with molecular biology or genomics research. However, recognizing the different states and priorities of centers, it could start with as few as four centers. CIAT, CIMMYT, CIP and IRRI have recognized the need and urgency for this collaboration. NCGR is a world leader in bioinformatics and computational biology with a wealth of experience, and human and physical resources. It has a mandate to develop bioinformatic resources but does not conduct biological research. They concentrate on information technology and seek partnerships with institutions collecting and managing biological data. Hence, their contribution is ideally compatible with that of the CG centers. Most of their products are publicly accessible, as would be any resulting from this project, and hence, there is a cumulative development of compatible resources from many different partnerships.

The CG centers have developed strong networks with NARS partners in molecular biology and genomics such as the Asian Rice Biotechnology Network and the Asian Maize Biotechnology Network.

These networks are now in need of access to the bioinformatics skills and resources that will be supplied through this project.

Size and Scope

The project should commence with an initial phase of 3 years. Initial investment should be at a level of US$ 50,000 per center per year. CG centers can join at any time, but a core of four centers is required for critical mass.

Activities

The project will start with a joint technical workshop in order to

- Identify and prioritize common bioinformatic needs
- Decide which of these needs can be accommodated with the current project and appoint teams to address them
- Develop project activities and appoint project management teams to seek resources to address needs outside the scope of the current project.
- Produce a training and extension strategy to develop bioinformatics capacity and skills in centers and NARS partners.