DRAFT FOR DISCUSSION

White paper on GCP research components: Cloned genes

21 AUGUST 2012

Table of Contents

Acronyms, short names and abbreviations – Cloned genes

Al	aluminium
Alt_{SB}	major Al tolerance gene in sorghum cross BR007 × SC283
CG (see CGIAR)	
CGIAR	No longer an acronym (*formerly* Consultative Group on International Agricultural Research)
CornellU	Cornell University, USA
CRPs	CGIAR Research Programmes
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária (*Brazilian Agricultural Research Corporation*)
GCP	Generation Challenge Programme (of the CGIAR)
GRiSP	Global Rice Science Partnership (a CGIAR initiative led by IRRI)
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IP	intellectual property
IRRI	International Rice Research Institute
JIRCAS	Japan International Research Center for Agricultural Sciences
KARI	Kenya Agricultural Research Institute
KASPar assay	KBioscience Competitive Allele-Specific PCR SNP genotyping system
MAB	marker-assisted breeding
MAGIC	**multiparent** advanced generation intercross
MATE	family of multidrug and toxin extrusion transporters
MB	molecular breeding
P	phosphorus
Pup1	phosphorus uptake 1 (gene)
qRT-PCR	quantitative real-time polymerase chain reaction
QTL	quantitative trait locus
RI	Research Initiative (of GCP), *formerly* Challenge Initiative (CI)
RNA	ribonucleic acid
Sb	***Sorghum bicolor* L**
SNP	single-nucleotide polymorphism
USA	United States of America
USD	United States dollar
USDA	United States Department of Agriculture, USA
Zm	*Zea mays* L

Background and process

A series of white papers are being drafted by the Generation Challenge Programme (GCP) team in collaboration with external experts. The goals are to communicate the outputs and deliverables from each research component during 2004–2014 and to explore options for enabling and ensuring that the potential benefits of these components will be fully realised in the future. At this stage, the white papers are really a first analysis for internal use.[1] They are expected to evolve over time, shaped by progress made during GCP's remaining time and by the evolution of international agricultural research for development, particularly in terms of the 'moving landscape' of socio-economic, political and environmental issues in which operate the research portfolios of the CGIAR Consortium of International Agricultural Research Centres and related CGIAR Research Programmes (CRPs). Each white paper is designed to contribute to GCP's orderly closure in 2014 by considering the following three questions.

1. What research assets will be completed by the end of GCP's lifetime in December 2014?

2. What research assets can best continue as integral components of the new CGIAR Research Programmes (CRPs) or elsewhere?

3. What research assets may not fit within existing institutions or programmes and may require alternative implementation mechanisms?

This paper focuses on the outputs and options for GCP's cloned genes component. Outputs have been achieved through (a) collaborative work among three sets of actors: a broad network of partners in regional and country research programmes, the CGIAR and academia; and (b) through capacity enhancement to assist developing-world researchers to tap into new genetic diversity and access modern breeding tools and services. GCP research activities have produced the research products described below[2].

Introduction and rationale

During Phase I (2004–2008), GCP's efforts towards gene cloning focused on aluminium tolerance in sorghum (*Alt* genes) and phosphorus-uptake efficiency in rice (*Pup* genes). Phase II (2009–2014) builds on that initial effort by conducting molecular breeding (MB) for those two traits. In addition, orthologous genes, such as the sorghum Al tolerance gene (Alt_{SB}) in maize ~~and rice,~~ and rice *Pup1* in maize and sorghum, were identified. The gene-

[1] This GCP white paper, like the others in this series, is not a conclusive, static document. Instead, it will continue to grow and evolve as the processes of evaluation and deliberation advance toward GCP's end in 2014.

[2] GCP is supported by generous funding from an array of donor organisations listed at http://www.generationcp.org/network/funders. See also descriptions of products at http://www.generationcp.org/impact/product-catalogue and of the institutions that generated them at http://www.generationcp.org/research/research-projects.

cloning projects, search for orthologous genes in other crops and application in breeding programmes together required an average expenditure of about 5% of a total GCP research budget of USD 150 million spread over 11 years.

Aluminium (Al) toxicity and low phosphorus (P) are major factors that hamper cereal productivity in acid soils, especially in sub-Saharan Africa and South America. Aluminium toxicity is a primary constraint for crop production in arable lands, affecting 38% of farmland in Southeast Asia, 31% in Latin America and 20% in East Asia, sub-Saharan Africa and North America. Phosphorus is the second most important inorganic plant nutrient after nitrogen. It is also one of the least available nutrients because of its tendency to be immobilised in Al and iron complexes in the soil. Almost 50% of rice-lands are currently P deficient. Both Al toxicity and low P cause problems of food security throughout the world and, as abiotic constraints to crop production, are exceeded only by drought.

Tolerance of these stresses in different crops is generally derived from the action of major genes, which explain a large percentage of phenotypic variance. Major genes are extremely valuable for MB, because they are relatively easy to pyramid, compared with the large numbers of minor genes involved in the regulation of highly polygenic traits such as drought tolerance. Cloning these genes can enhance understanding of mechanisms behind tolerance of Al and low P in cereals, and facilitate the search for homologous genes across (orthologs) or within species (paralogs). Gene cloning allows the development of highly precise and reliable 'on the gene' molecular markers making the introgression of the target gene easier (reduced linkage drag) compared to link markers.

In Phase I, GCP's gene-cloning activities were conducted mainly through competitive grants (six projects). In Phase II, they were carried out within the Comparative Genomics Research Initiative (five projects) aimed to improve cereal yields in high-Al and low-P soils.

Objectives of the gene-cloning research component are as follows:

1. Identify, characterise and clone major genes for Al tolerance and low-P uptake efficiency.

2. Provide means for comparative functional analysis, enabling the identification of homologous genes for tolerance of Al and low-P within and across cereal species.

3. Develop molecular markers and introgress favourable alleles of key genes into susceptible elite germplasm.

4. Provide evidence that gene cloning, as a component of genomics, is useful to crop breeding.

Project activities and outputs

The same sequential approach was followed for the two major initiatives conducted through successive projects over GCP's ten years. These efforts resulted in gene cloning and the

development and use of markers to breed for Al tolerance and low P-uptake efficiency, as follows:

- First, the screening of diverse germplasm allowed the identification of genotypes with contrasting response to specific experimental conditions, that is, for Al tolerance, using hydroponics, where roots grow in acid solutions; and, for P efficiency, using low-P concentrations.

- Second, segregating populations were developed for further phenotyping and fine-mapping of genomic regions involved in the expression of target traits.

- Third, following in some cases a candidate gene approach, the mapping and then cloning of the target genes were accomplished.

- Fourth, the cloned gene was validated through approaches such as over-expression (transgenesis), quantitative RNA expression analysis (qRT-PCR), genetic analysis (of near isogenic lines) and verification (evaluation of genetic effects).

- Fifth, contrasting germplasm was screened to identify elite alleles and enable their introgression into elite but susceptible germplasm for the target trait.

In parallel to the introgression of favourable alleles using molecular markers, significant efforts were made to identify orthologous genes in other cereal crops.

In an early GCP-supported project, the gene Alt_{SB}, verified to confer high Al tolerance in sorghum, was cloned by collaborating researchers at Cornell University (USA) and Embrapa Maize & Sorghum (Brazil). This important work resulted in the identification, using positional cloning, of *SbMATE*, a gene encoding a member of the multidrug and toxic compound extrusion (MATE) family in sorghum. *SbMATE* is an Al-activated citrate transporter and is responsible for the major sorghum Al-tolerance locus (Alt_{SB}). Results were published in *Nature Genetics* in 2007[3].

Polymorphisms in regulatory regions of Alt_{SB} contribute large allelic effects, acting to increase Alt_{SB} expression in the root apex of tolerant genotypes. These findings allowed scientists to identify superior Alt_{SB} haplotypes and to use them, via MB, in acid-soil breeding programmes, not only in Brazil but also in Mali and Niger. Thus, they helped increase crop yields in developing countries where acid soils predominate.

Information gathered from this accomplishment stimulated projects to identify orthologs of this gene in maize and rice (Comparative Genomics RI). *ZmMATE1* was confirmed as an

[3] Magalhães JV, Liu J, Guimarães CT, Lana UGP, Alves VMC and others. 2007. A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature Genetics* doi:10.1038/ng2074.

orthologous gene underlying a QTL that accounts for a large percentage of Al tolerance in maize. Results were published in *Plant Journal* (2010)[4].

A candidate gene in maize (*ZmMATE1*) has been identified and is now being validated. Sequencing this gene is an integral activity of this process, assisting the development of markers to use for introgressing the associated QTL into elite Al-susceptible lines for validation studies in the field.

Similarly, *Pup1* (now designated as *Pstol1*), a gene that enhances yield under P-deficient field conditions, has been cloned in rice, revealing a novel protein kinase gene. Results have been published in *Nature* (2012). The effect of this gene on P-uptake efficiency has been validated, using transgenic 'Nipponbare' plants (T1). Phenotyping of five independent IR64 T1 lines with one or two copies of the 35S:*Pup1*-kinase construct revealed a large effect of this gene on early growth vigour. Plants developed a significantly higher tiller number, root biomass and grain yield when grown under P-deficient and drought conditions. An effort was made to identify orthologous genes of *Pup1* in the maize and sorghum genomes, using comparative genomics approaches. Many genes in sorghum and maize showed homology with the rice *Pup1* gene. A detailed analysis is currently underway.

Although some genes being studied have yet to be cloned, the major products generated from GCP-supported projects are as follows:

Aluminium tolerance

1. Alt_{SB} gene underlying the major QTL for Al tolerance in sorghum – cloning completed. Elite alleles and 'on the gene' markers are identified. Introgression of the QTL into elite susceptible lines for Brazil and Niger is now accomplished.

2. *ZmMATE1*, an orthologous gene of Alt_{SB}, underlying a major QTL for Al tolerance in maize – cloning completed. Markers for the QTL for introgressing the trait have been developed and are now being used in marker-assisted breeding (MAB).

Phosphorus-uptake efficiency

3. *Pup1*, a gene underlying the major QTL for P-uptake efficiency in rice – cloning completed. Linked and 'on the gene' markers are available and MB activities have been successfully conducted in Indonesia (improved lines already registered) and other Asian countries, deploying rice populations developed through multiparent advanced generation intercrossing (MAGIC).

[4] Maron LG, Piñeros MA, Guimarães CT, Magalhães JV, Pleiman JK, Mao C, Shaff J, Belicuas SN, Kochian LV. 2010. Two functionally distinct members of the MATE (multi-drug and toxic compound extrusion) family of transporters potentially underlie two major aluminum tolerance QTLs in maize. *Plant J* 61(5):728–724.

Aluminium tolerance and phosphorus efficiency

4. The Comparative Genomics RI is identifying (a) Al-tolerance genes or QTLs and developed linked markers in maize and rice, and (b) is currently developing P-uptake efficiency genes or QTLs and related markers in sorghum and maize.

Gene cloning today functions as an integral component of two Comparative Genomics RI projects that have clear breeding objectives. Gene cloning has been a relatively small but significant cost in GCP's research budget. Gene-cloning projects, the search for orthologous genes in other crops and application in breeding programmes together required an average expenditure of about 5% of a total research budget of USD 150 million spread over ten years.

Post-GCP sustainability and projected impact

Gene cloning for *Alt$_{SB}$* in sorghum and *ZmMATE1* in maize was conducted at Cornell University and Embrapa Maize & Sorghum, whereas *Pup1* was cloned in rice by scientists at the International Rice Research Institute (IRRI) in collaboration with the Japan International Research Centre for Agricultural Sciences (JIRCAS). The *Alt$_{SB}$* and *ZmMATE1* genes are maintained by CornellU and Embrapa Maize & Sorghum as plasmids in bacterial cultures stored in glycerol at -80 °C. The *Pup1* gene cloned recently is held at IRRI under similar conditions. Furthermore, gene sequences with causal nucleotide variation tightly linked to the QTL in all three cases are published. Other homologous genes for Al tolerance (in rice) and P-uptake efficiency (in sorghum and maize) that may be cloned during GCP's remaining period will be maintained under the same conditions and the gene sequences will also be made public under the GCP IP rules.

Considering results so far, the Programme's gene-cloning objectives could be claimed as having been attained for both Al tolerance in sorghum and maize and P-uptake efficiency in rice. Major genes have been cloned, molecular markers are available and improved germplasm has been developed in collaboration with scientists from developing countries. As this knowledge is publically available, these products (ie, genes and markers) are available today in a sustainable way. Improved germplasm has been generated through those different research activities, introgressing using marker-assisted selection those major genes into recipient elite or popular germplasm. Considering the breeding value of those lines when grown in acid soils compared with existing elite germplasm, their maintenance and distribution by respective CG Centres or country programmes should not be an issue.

More recent projects, embedded in the Comparative Genomics RI, permitted identification of the Al tolerance loci in maize and although identification of the P-uptake efficiency genes is in its preliminary stages for maize and sorghum, results are quite promising. In maize colleagues at Embrapa reported a pretty nice co-localization between the top *Pup1* hit and a QTL for root traits and validation work is now ongoing in the field. For sorghum, several SNPs that are specific for 7 *Pup1* candidates were identified, converted into the KASPar system and genotyped on the GCP sorghum association panel demonstrating some level of associations with root traits for most of these candidates. Although on one hand some early GCP projects demonstrated some limitations of the comparative genomics approach for

gene discovery when targeting very polygenic traits such as drought tolerance, on the other hand comparative genomics projects, based on major genes to identify functional orthologs or at least orthologs involved in the expression of comparable response mechanisms, such as root growth in the case of Pup1, are showing interesting results from a breeding perspective.

An example of an *ex ante* impact analysis is the recent study of the economic benefits of MB that had used markers developed by GCP. New rice varieties that tolerate salinity (*Salt1*) and P deficiency (*Pup1*) had been developed (IS IT IN PROGRESS OR ALREADY DONE??) for Bangladesh, India, Indonesia and the Philippines. Encompassing a broad set of economic parameters, the study concluded that the MB approach saved an estimated minimum of two to three years in varietal development time. This acceleration of the process resulted in significant incremental benefits, ranging between USD 300 million and 800 million, depending on the country, extent of abiotic stress and lag for conventional breeding[5].

Two and half years remain before GCP closes, during which time the shape and number of products to be delivered can evolve. Considering the upstream nature of GCP research activities, it is also fair to say that impact shall be measured based on concrete indicators 3–5 years after the Programme ends (see *Transition strategy* for these indicators at http://www.generationcp.org/about-us/strategies). Therefore, instead of anticipating a relative or absolute value – the equivalent of gazing into a crystal ball – a relative scale from 1 to 5 can preferably be used, where 5 is the largest impact across all GCP products, regardless of activity or crop, and 0 is no impact. With this approach, the impact of cloning genes and the improved germplasm resulting from MB activities, pyramiding elite alleles and cloned genes is estimated to have an impact factor of 3. Such a score indicates that new markers for major genes will have a significant impact on plant breeding efficiency in developing countries.

Analysing the post-GCP placement of the cloned genes component

GCP has primarily a research and capacity-building function, but development investment must follow if outputs are to reach farmers in developing countries. GCP's goal was always to hand over those projects initiated during its Phase II (including 'on the gene' markers for cloned genes) to CGIAR partners, country partners or local private companies so they may develop or deliver new varieties to farmers.

What will be finished by December 2014

Clear products have been generated and are available in a sustainable way. The cloning of Alt_{SB} in sorghum, *ZmMATE1* in maize and *Pup1* in rice has been achieved, MB has been either successfully conducted or is underway, and the cloning of some corresponding

[5] Ismail AM, Heuer S, Thomson MJ, Wissuwa M. 2007. Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Mol Biol* 4:547–570.
Alpuerto VE, Norton GW, Alwang J, Ismail AM. 2009. Economic impact analysis of marker-assisted breeding for tolerance to salinity and phosphorous deficiency in rice. *Rev Agric Econ* 31:779–792.

orthologous genes should be achieved by December 2014. GCP's gene-cloning activities are expected to be completed and related objectives met by December 2014.

Extending activities to CRPs, Centres or other institutions

As for all GCP research activities, the projects of the Comparative Genomics RI are embedded in respective crop CRPs. Further development of improved germplasm, using the 'on the gene markers' so far developed with GCP partners, may therefore be conducted effectively in those CRP breeding activities. After GCP's projects are completed, the CRPs may choose to use the resulting products according to their individual assessment of value.

Sorghum

Although not directly involved in gene cloning, ICRISAT has been involved in improving sorghum, using *Alt*$_{SB}$ (and putative orthologous *Pup1* gene). The Centre will take up the responsibility of managing the sorghum projects as part of leading the CRP on Dryland Cereals, which includes sorghum breeding.

Rice

IRRI has been the lead partner for cloning the *Pup1* gene and should therefore manage the follow-up on the rice project through the Global Rice Science Partnership (GRiSP), a CGIAR Research Programme led by IRRI. The Institute could also take up the research on the discovery and application of Al-tolerance genes.

Maize

The maize projects would fall to CIMMYT, which, however, has not been a partner for any of these projects. This would make the extension of the maize activities beyond December 2014 challenging. However, if clear added value emerges in extending these activities, mainly breeding for acid soil using *ZmMATE1* and the other possible P-uptake efficiency gene, then such extension should be considered because EMBRAPA, Cornell and some partner country programmes such as KARI in Kenya are also designated partners of the Maize CRP.

If desirable, other gene-cloning or follow-up activities may also be continued by non-CGIAR partners. However, the viability of these activities in non-CGIAR institutions would depend entirely on the impact the extension of these activities would have on related breeding programmes, both at research and application (MB) levels, and on either local or grant funds being available to these institutions. Gene cloning activities related to Al tolerance and other abiotic stress shall continue both within USDA at Cornell University and Embrapa. Thus, there is an opportunity for the CRPs to engage into a collaborative effort with mutual benefit, if the mechanisms to do that are put in place within the CRPs.

Embedding the work in a new entity as a research activity

Clearly, then, the CRPs and many non-CGIAR institutions such as Cornell have the expertise and resources to conduct gene cloning, given sufficient interest and funding. There is, therefore, no reason to consider a new entity for such gene-cloning activities or for the breeding applications that may result from discovering new genes.

Conclusion

The gene-cloning component has yielded important genes for Al tolerance and P-uptake efficiency, within and across genomes; and the molecular markers to improve elite but susceptible germplasm of sorghum, maize and rice. These products are available in a sustainable way, their application in breeding programmes can be conducted in routine, and GCP gene-cloning activities can be considered as completed by December 2014. Further searching for genes and QTLs for Al tolerance and P-uptake efficiency appears suitable, considering the importance of the target traits and the need for more genes or QTLs to be characterised. We, therefore, feel that the CRPs – who have already incorporated comparative genomics projects into their respective workplans – will be in the best position to decide on their continued use and priority in their research plans after December 2014. Grant-based partnerships with non-CGIAR institutions that are highly practised in the science such as Cornell, as well as lead country partners such as EMBRAPA or KARI, could also be an efficient and beneficial arrangement for all parties involved.

GCP remains committed to its mission and community to the last day of the Programme and will work with partners along the delivery chain to maximize successful implementation of the delivery plans developed for each Research Initiative. GCP will also closely engage with its partners until its very sunset to ensure – as far as will be possible – the integration, extension, and expansion of activities as may be required. The Programme will even help initiate related new activities that build on GCP's achievements, should there be clear added value and demand for such activities. In this way, the Programme is working to secure a broad and sustainable use of its products well beyond 2014 while mitigating against the loss of gains made this far.