GSDR 2015 Brief

Thinking a Global Open Genome Sequence Data Framework for Sustainable Development

*By Norman Warthmann and Claudio Chiarolla**

Introduction

The cost of genome sequencing has fallen one-million fold in the past several years.[1] The technology is widely accessible and it is now inexpensive to quickly produce genome sequence information for large numbers of individuals.[2] A 'genomics revolution' is underway, which is transforming the life sciences, including biomedicine and animal and plant breeding.

The UN Secretary-General's High Level Panel on post-2015 development goals has recently called for "a New Data Revolution" for sustainable development.[3] However, genomic data does not squarely fit within the narrow statistical focus described by the Panel. Critical gaps concerning the governance of genomics data need to be filled for the promotion of science as a global public good. Main focus of our contribution is on plant breeding, but similar cases can be made for animal breeding and (human) biomedicine.

The goal: unrestricted sharing of genomic data

"Big data comes with big opportunities to tackle big global challenges."[4] This is particularly true in the life sciences, however, the possibility to harness gene sequence data in a proprietary manner and to claim molecules, plant traits and DNA constructs, including ones previously existing in nature, may impinge on the ability of others to undertake fundamental and/or applied research.[5] Inexpensive DNA sequencing is transforming genetic resources into a promising source of appropriable information. The hereditary information found in DNA, coupled with the identification of gene function, can constitute the basis of monopoly rights.

Transparent, universal access to data enables innovation. In a situation where we need to sustainably intensify agricultural production for a growing human population and a changing climate, we must accelerate plant breeding, and the sharing of genomic information of PGRFA is a key component.

Faster progress through sharing and use of genomic data

A genome is the blueprint and the code of an organism; the genome determines the traits. In most plants and animals parental genomes recombine with each other and so create the genomes of the offspring. A genome can hence be viewed as composed of building blocks from

---

* The views and opinions expressed are the authors' and do not represent those of the Secretariat of the United Nations. Online publication or dissemination does not imply endorsement by the United Nations. Norman Warthmann, Lecturer, Research School of Biology, The Australian National University, Canberra, Australia. norman@warthmann.com; Claudio Chiarolla, Research Fellow, International Governance of Biodiversity, Institut du Développement Durable et des Relations Internationales (IDDRI) & Lecturer in Biodiversity Law & Policy at the Paris School of International Affairs (PSIA) Sciences Po. claudio.chiarolla@iddri.org

[1] http://www.genome.gov/sequencingcosts/

[2] http://www.planttreaty.org/sites/default/files/ITPGRFA_BS005e.pdf

[3] http://www.post2015hlp.org/wp-content/uploads/2013/05/UN-Report.pdf

[4] https://www.croptrust.org/content/speech-experts-big-data-and-crop-diversity

[5] C Chiarolla, *Intellectual Property, Agriculture and Global Food Security: The Privatisation of Crop Diversity*, Cheltenham, UK & Northampton, MA, USA: Edward Elgar (2011).

its ancestors, and the genomes of modern cultivars of our crop plants are, with rare exceptions, combinations of building blocks inherited from landraces and wild ancestors. Breeding aims to identify offspring with superior characteristics; a process that is mainly based on selecting offspring exhibiting desired traits: a slow trial-and-error approach. But because these traits are encoded in the genome, knowledge of associations between traits and corresponding genomic locations (loci) holds the promise to make the breeding outcome predictable: With genomic information and tools, conservation efforts of germplasm will become more comprehensive, pre-breeding will become more efficient and precise, and the breeding process itself will be greatly accelerated[6].

Due to the combinatorial nature and the diversity of the genomic building blocks, establishing robust causal links between traits and genome requires large numbers of samples, and traits measured in different environments. Datasets of sufficient size can only be achieved by aggregating and comparing data across experiments and genotypes. It is hence crucial that data is available, shareable and shared. Given a comprehensive catalogue, genome sequence information can then be used to make informed decisions on suitable PGRFA for further breeding, greatly reducing the need for field evaluations.

In order to establish such catalogues, large-scale efforts are currently underway to characterise the genomes of germplasm in public seed banks[7], starting with the ones under the public goods mandate of the Consultative Group on International Agricultural Research (CGIAR). This endeavour is timely and IRRI, CIMMYT, and ICRISAT are ploughing through thousands of their varieties already. Broad access to this data will be a game changer with vast implications for breeding and food security. Notwithstanding the proclaimed intent of this[8] and other important ongoing scientific[9] and political initiatives[10] to promote the public good through open data, no mechanism is yet in place to ensure that such data will be systematically treated as a public good for the benefits of mankind.

A public license for genomic information on crop germplasm

Publicly available genomic information with the freedom to operate is valuable for scientific progress, breeding, biodiversity conservation and food security. PGRFA under the multilateral framework established by the FAO's International Treaty are to be treated as a 'global public good'. It is now time to think about how to also secure their genomic sequences as a public good, and, at the same time, encourage more data accumulation and sharing. Towards a **Global Open Genome Sequence Data Framework** we propose a **License for genomic information on germplasm** (Annex I).

Starting with the CGIAR seed banks, and in addition to the requirement to provide information to the global information system of

---

[6] http://www.planttreaty.org/sites/default/files/ITPGRFA_BS005e.pdf

[7] www.divseek.org

[8] DivSeek's objective to "harness [...] the power of crop diversity to feed the future", see: http://www.nature.com/nature/journal/v499/n7456/full/499023a.html

[9] In the field of biomedical research, the most important initiative of this kind is the 'Global Alliance for Genomic and Health' (http://genomicsandhealth.org). It is in the process of creating a "Global Alliance to enable responsible sharing of genomic and clinical data." http://genomicsandhealth.org/about-the-global-alliance/key-documents/framework-responsible-sharing-genomic-and-health-related-data

[10] The conclusions of the Council of the European Union establishing the mandate for the negotiations at the twelfth meeting of the Conference of the Parties to the UN Convention on Biological Diversity also highlighted the importance of improving the knowledge management of research results and open access to the data and results. Doc 9411/1/14 REV 1, para. 29.

the FAO Treaty[11], this license aims to ensure that genome sequence information and related data on genetic diversity from publicly funded repositories and other stakeholders is *free* and made available as a public good. Anyone who sequences such germplasm must be allowed to make the resulting information publicly available as is without prior curation, and without being exposed to the risk of infringing someone else's rights on the material and/or related information, including patents or copyright. This license reaches through – i.e. the genomic information of descendants of material subject to this license is again subject to the same terms and conditions.

Conclusions & Recommendations

This paper emphasises the need to promote the sharing of genomic information for sustainable development. We suggest considering copyleft-type licensing mechanisms, inspired by the *free* and open-source computer software movement[12]. Open access to, and freedom to operate with genomic data will enable innovation. It can also be viewed as component of non-monetary benefit-sharing[13]. We suggest that the UN should play an active role in developing and adopting a **Global Open Genome Sequence Data Framework** by framing an enabling legislative environment and by promoting relevant technology developments.

Annex I – Proposal for a Public License for Information Concerning Crop Germplasm

Copyleft-type licenses in their various forms (e.g., *The GNU General Public License, GPL*) have transformed software development and industry by requiring source code to be open and by imposing the same license onto any new software produced based on *copyleft license*d software. Software published under the GPL is *free*[14], its code is *open source* and can be used by anybody, even for commercial purposes and products, and descendent code must in turn again be *free*, i.e., it must be published under the very same license. The below draft license develops a similar licensing scheme for information related to crop germplasm, including full genome sequence data.

All Information concerning crop genetic material covered by this license shall be treated as a public good. Such information is distributed under the <*yet to be named*> Public License and its use is subject to the following conditions:

1) The use and/or publication of information about *the <named crop genetic material>* and all its descendants cannot be restricted in any way. This includes, but is not restricted to, genetic information (i.e. whole genome sequences) and phenotypic measurements.

2) At the time *the <named material>*, including descendants and/or derived materials, are released or transferred, full genome sequences shall, in principle, be publicly available.

3) In case of information obtained from PGRFA within the purview of the Multilateral System of the FAO International Treaty such information shall be treated as non-confidential information and it shall be made available in accordance with SMTA Articles 5(b) and 6.9.[15]

---

[11] As described in Article 17 of the ITPGRFA.

[12] https://www.gnu.org/philosophy/open-source-misses-the-point.html

[13] On benefit sharing in general see: http://biodiversity-l.iisd.org/guest-articles/the-promises-of-benefit-sharing/

[14] *Free* as in *free speech*, not *free beer.* See: https://www.gnu.org/philosophy/open-source-misses-the-point.en.html

[15] SMTA Article 5(b) states that: "All available passport data and, subject to applicable law, any other associated available non-confidential descriptive information, shall be made available with the Plant Genetic Resources for Food and Agriculture provided." Article 6.9 states that: "The Recipient shall make available to the Multilateral System, through the

4) In case genome sequence information is released to the public, such information shall be traceable to the sequenced material.

5) Each time information covered by this license is released or transferred to a third party the recipient automatically receives a license from the original licensor(s), which is identical to this License.

6) All descendants and/or derived materials that contain information covered by this license must be traceable to *the <named material>* and are subject to this <*yet to be named*> Public License.

7) No further restrictions on the exercise of the rights granted or affirmed under this License may be imposed. For example, the recipient may not impose a license fee, royalty, or other charge for exercise of rights granted under this License. The recipient may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing a PGRFA in relation to any use of information concerning crop genetic material covered by this License.

---

information system provided for in Article 17 of the Treaty, all non-confidential information that results from research and development carried out on the Material."