

CANADA'S SUBMISSION ON DSI/GSD TO THE INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES  
FOR FOOD AND AGRICULTURE

During its 7<sup>th</sup> Session of the Governing Body of the International Treaty of Plant Genetic Resources for Food and Agriculture held in Kigali, Rwanda – 30 October to 3 November 2017, the GB adopted resolution 13/2017 (Multi-Year Programme of Work of the Governing Body of the International Treaty) in which Para 4 is to invite contracting parties, other governments, relevant stakeholders and individuals with relevant expertise on the matter to provide information to the GB related to “Digital Sequence Information” on Plant Genetic Resources for Food and Agriculture (PGRFA) and the types and extent of uses of digital information on PGRFA. On Dec. 3, 2018, the Secretary sent a notification to invite contracting parties to submit views and other information on “Digital Sequence Information” with a deadline of March 31, 2019. The present document includes Canada’s information on the terminology used in this area, actors involved with DSI on PGRFA, the types and extent of uses of DSI on PGRFA, such as characterization; breeding and genetic improvement; conservation; identification of PGRFA and the relevance of DSI on PGRFA for food security and nutrition, the use and potential implications of “digital sequence information” on genetic resources for food and agriculture, building on what Canada’s contribution submitted to the Executive Secretary of the Convention on Biological Diversity (CBD), set by decision CBD COP XIII/16 as well as to the Commission of Genetic Resources for Food and Agriculture for review at its 17<sup>th</sup> Session held in Rome – 17 February to 22 February, 2019. In Jan.3 0-Feb.3, 2017, at its 16<sup>th</sup> Session, the Commission established a new work stream (CGRFA-16/17/Report Rev.1, para.86) on “digital sequence information” on genetic resources for food and agriculture. At its 16<sup>th</sup> Regular Session, the Commission requested the Secretary to invite member countries (para.88) to submit information on the use of “digital sequence information” on GRFA and potential implications for the conservation and sustainable use of GRFA, including exchange, access and fair and equitable sharing of the benefits arising from their use and to compile the information. Document (**CGRFA-17/19/4/Inf.1**) highlights submissions by member countries and observers. Canada submitted views on “DSI” to the Secretary of the Commission on 18 September 2017 (see: **CGRFA-17/19/4/Inf.1**, page 10-16).

**Introduction**

Canada has always been of the opinion that access and benefit-sharing under the CBD and under the ITPGRFA, involves genetic resources, which does not include “digital sequence information”. It is Canada’s position that “digital sequence information” has the potential to bring numerous benefits to the Canadian and global societies. This includes potential solutions to food shortages and diseases like malaria. Canada has taken a firm position in many fora (ITPGRFA, CGRFA, CBD, and WHO) that DSI is not equivalent to genetic resources and continue to strongly resist any language that implies an equivalency between “DSI” and “Genetic Resources” as defined by the Convention on Biological Diversity or the International Treaty on Plant Genetic Resources for Food and Agriculture.

Agriculture, fisheries, and forestry, as well as food and non-food product value chains, depend on living organisms; this includes pollinators, naturally-occurring organisms that support ecosystem services (e.g. soil microorganisms), and natural enemies that can be used as biocontrol agents to manage pests. Biodiversity is the pool of resources upon which production systems are based. A vast pool of genetic

diversity has been lost through domestication and is in urgent need of effective conservation and sustainable use strategies around the world. Therefore, efforts to understand, preserve and utilize biological diversity for agriculture, forestry and fisheries have been and continue to be essential to bio-based industries and to sustain the enormous improvements science has brought to production, processing, conservation efforts and utilization of genetic resources.

The study of biodiversity has become central to Canada's research activities as science started uncovering the important interconnections between biodiversity, productivity and resiliency in agriculture, fisheries and forestry. The study of the biological diversity of crops, animals, forests, wild relatives, beneficial species, pests, invasive and alien species, and native species enables advances in areas such as genetic improvement of forest populations and breeds of farm animals, improvement of pest and nutrient management methods, development of new crops or crop attributes, threat identification and risk mitigation, and identification and management of invasive alien species. It supports resiliency of productive systems by enabling the development of beneficial management practices to manage the effects of intensive management practices on populations of beneficial organisms and the ecosystem services they provide.

Furthermore, in an increasingly globalized world, the movement of pests and diseases is a constant threat to production and could result in catastrophic losses of major economic importance for the agriculture, fisheries and forestry sector, as well as catastrophic damage to biodiversity. Therefore, the capacity to identify and characterize organisms that may threaten its ecosystems and biological resource base is vital.

Given the importance to agriculture of understanding the various organisms at play, their life cycles and the conditions they need in order to thrive, the study of agrobiodiversity and bio-resources has always been a priority research area supported by Agriculture and Agri-Food Canada. Its scientific collections, as well as the biosystematic study of species important for agriculture, are as old as research in this department (128 years). Over the years, AAFC has built a critical mass of knowledge, expertise and science capacity to study taxa of importance for agriculture. AAFC shares information on its holdings of plant and animal genetic resources collections using Internet websites and other tools. Biodiversity in agricultural systems can be studied at the various levels of organization, and all these levels are important: alleles, genes, populations, species, communities in an agricultural landscape and the agro-ecosystem as a whole. This ranges from genetic variation that provides the basis for valuable traits in a crop or affects the behaviour of a pest species, to populations of soil microorganisms and their importance for nutrient or carbon cycling or productivity of agricultural crops dependant on pollinator populations.

The Government of Canada is committed to openness, transparency and information sharing through the Open Government Partnership. Since joining the Open Government Partnership, a multilateral initiative to foster greater transparency and accountability, Canada developed three national action plans aimed at enhanced access to information and expanded open data, among other things. The Ministry of Environment and Climate Change Canada is the federal lead on the Open Science element of Canada's 2016-2018 Action Plan on Open Government. Sharing data related to genetic sequences is not new – the GenBank initiative (<https://www.ncbi.nlm.nih.gov/genbank/>) has made genetic sequence data available worldwide since 1982. Full sequences of genotypes of various crops (wheat in 2012, soybean in 2010 and

others) and annotations of animal genomes carried out by various countries (<http://www.ensembl.org/info/about/index.html>) have also been published and made available online. Canada is also a voting participant of the Global Biodiversity Information Facility.

As a federal government department, the Ministry of Agriculture and Agri-Food Canada (AAFC) is also committed to providing access to data produced, collected and used to the Canadian public, researchers and industry communities. AAFC is a contributor to the Government of Canada's Open Data Portal. On March 22, 2016, the Government of Canada announced investment in AAFC of \$30 million CDN over 6 years, starting in 2016-2017 to accelerate the DNA analysis, and digital recording of the department's collections of over 19 million specimens of insects, plants, fungi, bacteria and nematodes. This initiative also includes generating DNA sequence data (Genetic Sequence Data ,GSD) on germplasm accessions held by the Canadian national gene bank for plant genetic resource for food and agriculture, Plant Gene Resources of Canada (PGRC) in addition to improving digital passport data records. It was intended that the data capture and imaging of specimens and molecular sequences from the collection be freely shared on the Federal Government Open Data Portal. By 2022, label data capture of six collections will be 100% completed as well as 25% of the Canadian National Collection of Insects, Arachnids and Nematodes. More than 7.6 million data records will be digitized. This initiative will increase 10-fold the number of specimen-based sequences that AAFC currently has. Through this initiative, Canada is also providing global leadership by also including molecular characterization (DNA sequence data). This data will be integrated to improve public accessibility to AAFC's collections and supports research in priority areas including climate change, and the rapid identification and prevention of biological threats to agriculture. To date, high-throughput protocols for molecular characterization were developed, the high-performance computing environment and storage was procured, installed and is being used for genomics analytics, the high-throughput digitization unit was purchased and installed to replace static digitization units to accelerate data capture and imaging of collection specimens.

The Federal Government Genomics Research and Development Initiative (GRDI) was implemented in 1999 to support genomics research in eight federal science departments and agencies according to their mandates and the priorities of the Government of Canada and is now in its sixth funding cycle. Each year, GRDI provides \$19.9 million to Agriculture and Agri-Food Canada (AAFC), the Canadian Food Inspection Agency (CFIA), Environment and Climate Change Canada (ECCC), Fisheries and Oceans Canada (DFO), Health Canada (HC), the National Research Council of Canada (NRC), Natural Resources Canada (NRCan), and the Public Health Agency of Canada (PHAC). GRDI funding has been crucial to the development and maintenance of Canadian Crop Genomics and supported the development of sustainable food and feed supplies from cereal, pulses, and oilseeds, and the introduction of crop platforms for bioproducts. AAFC has been receiving 4.4M\$ per year for the last 20 years. Funding was renewed for on Feb. 28, 2019 and AAFC will continue to receive an annual funding of \$4,440,000 for the years 2019-2010 to 2023-2024, and provide ongoing funding thereafter. Agriculture and Agri-Food Canada is currently using its GRDI funding (4.4M\$) to advance Canadian Crop Genomics Initiative priorities as well as enable industry to take advantage of innovations. Activities fall under two broad themes: 1) **biodiversity, gene mining and functional analysis**—develop value-added traits (e.g., seed quality) for the highly competitive marketplace as well as make Canada's crop production more resilient against potentially catastrophic

abiotic and biotic stresses to maximize sector profitability; 2) **improved access to biological materials and data sets**—make plant breeding more efficient, laying the scientific foundation for major advances in priority trait development and delivery for industry.

### **Terminology**

Canada remains uncomfortable with the term “digital sequence information” to characterize this data. First, the sequences in question are genetic sequences, not (for example) mathematical sequences or sequences of events. Therefore, to avoid confusion, the word “genetic” should be part of the term identifying these data.

Second, the fact that genetic sequences can be conveyed digitally is of secondary importance; the “digital” aspect only refers to a particular mode of storage or mode of transmission. Genetic sequences listed on a sheet of paper would still be genetic sequences. Therefore “digital” should be removed from the term.

Third, the term “information” is not appropriate. “Data” would be more accurate, because data can be codified and is transmissible, which is the case for genetic sequences, but not always the case for information.

Therefore, Canada would prefer to use the term “genetic sequence data” (GSD), which we shall use in the rest of this document.

In the background study 68 commissioned by the Commission on Genetic Resources for Food and Agriculture it is mentioned that “DSI” is not a term being used by the scientific community.

#### **Genetic Sequence Data are not Genetic Resources**

It has been proposed that GSD should be treated as if they were genetic resources. Canada does not agree; we are of the view that GSD are not genetic resources.

The definition of genetic resources in three legally-binding international instruments makes that clear.

In the Convention on Biological Diversity, Article 2:

*“Genetic resources”* means genetic material of actual or potential value

*“Genetic material”* means any material of plant, animal, microbial or other origin containing functional units of heredity.

In the International Treaty on Plant Genetic Resources for Food and Agriculture, Article 2:

*“Plant genetic resources for food and agriculture”* means any genetic material of plant origin of actual or potential value for food and agriculture.

*“Genetic material”* means any material of plant origin, including reproductive and vegetative propagating material, containing functional units of heredity.

In the Nagoya Protocol on Access to Genetic Resources and Benefit-Sharing, Article 2:

The terms defined in Article 2 of the Convention shall apply to this Protocol.

These definitions are consistent and aligned. They state clearly that genetic resources “contain” “functional units of heredity” like DNA. The functional units of heredity are not themselves genetic resources, and even less so the base sequences within these functional units. Genetic sequence data, by their nature, do not and cannot contain functional units of heredity like DNA. Genetic sequence data do not accord with the legal definition of genetic resources in these international instruments. They therefore provide no legal basis to treat GSD like genetic resources.

There are also risks incurred in treating GSD as genetic resources. Genetic sequence data are not permanent and would change over the years; the sequences will be modified by punctual, random mutations. It is also evident in other articles of these instruments that they were not intended to relate to genetic sequence data. For example, provisions in these international instruments to conserve genetic resources certainly do not apply to strands of DNA.

Canada prefers to use the term “genetic sequence data” (GSD) which is consistent with Canada’s submission to the CBD and to the Commission in fall of 2017 and is aligned with the WHO terminology. From the Canadian national genebank perspective, the response to the request by the ITPGRFA Secretariat is with the strict understanding that *Digital Sequence Information (DSI) refers to genetic sequence data of DNA and RNA sequences*. Therefore, the term GSD should be used consistently rather than the term DSI which is not a commonly used term by the scientific community. There is a lack of a common understanding of what does DSI include. Attempts to define DSI universally have not been successful so far.

### **Actors involved with GSD on PGRFA**

In the context of Plant Gene Resources of Canada (PGRC), there are several research scientists closely associated with PGRC at the Science and Technology Branch of Agriculture and Agriculture Canada (AAFC) that have produced and analyzed GSD or are involved in ongoing projects that will do so. Other PGRC clients, outside of AAFC, have also conducted such work or indicated that they will do so, but so far this has been limited. However, there is an increasing demand on PGRC germplasm for such purpose by public research institutions, because the cost of generating GSD have decreased considerably in recent years.

### **Types and Extent of Uses of GSD on PGRFA**

In this section, Canada provides examples of research done that require simple and ready access to genetic sequence data of Canadian and international origin.

#### **Characterization**

Starting about 2010, GSD data on PGRC accessions has been produced on subsets of germplasm of oat (including wild oat species), barley (including wild barley species), wheat, flax, maize, soybean, and oilseed

brassica species. The number of samples (PGRC accessions) analyzed has been limited in most cases to less than 200 accessions of pre-selected subsamples that represent the diversity in the species. The purpose was in all cases to better characterize the genetic diversity in order to enhance *ex situ* germplasm management, or in order to clarify phylogenetic relationships within the taxa considered. In wild oat, wild barley, wild emmer, crested wheatgrass and northern wheatgrass larger samples have been analyzed to assess the genetic diversity for conservation purposes including *in situ* conservation.

Applied RNA-sequencing technology has been used to identify differentially expressed genes for traits of the breeding target late maturity in crested wheatgrass and sainfoin germplasm in collaboration with a Canadian public breeding programme. GSD has been used to screen 600 soybean accessions to support public breeding for early maturity

### **Breeding and Genetic Improvement**

Breeding programs are part of Agriculture and Agri-Food Canada holistic approach to crop production challenges.

**Pre-breeding:** Much work is taking place at the pre-breeding stage, where efforts are concentrating on identifying desirable characteristics and/or genes from unadapted materials that cannot be used directly in applied breeding and to transfer these traits to intermediate materials that breeders can use further in producing new varieties to be released and used by farmers. There are necessary steps in the use of diversity contained in wild relatives and landraces and allow breeders to broaden the genetic base of crops. Characteristics such as cold, heat, and drought tolerance as well as increasing resistance to fungal diseases of importance to Canada are the main focus of pre-breeding programs.

**Phenotyping:** Phenotypic characterization continues to be the basis for plant selection in Canada's breeders' work. Phenotypic evaluation remains by far the most expensive and time-consuming activity for breeding programs. Investment effort in Canada has been in Phenomics.

**Gene Sequence Information:** Genotyping and genome sequencing information are increasingly available for use in pre-breeding and breeding. For most target crops, DNA sequencing of thousands of accessions in our national genebank (Plant Gene Resources of Canada) and elsewhere in the world have generated a critical mass of data to create more accurate crop phylogenies and link genomic regions to traits. This research has shed new light on the relationships among crop varieties, landraces and wild relatives in collections and has helped breeders identify materials with desired traits.

Genomic information has been extensively used and combined with classic breeding for guiding selection, crossing, and evaluation in ongoing breeding programs for several crops. Single nucleotide polymorphism (SNP) and other types of nucleotide variations are being used to develop molecular markers linked to genes and quantitative trait loci (QTL). These markers allow plant breeders to use time and resources more efficiently. These markers are also useful for phylogenomic analysis and will aid in the detection of novel genetic variants amongst oat wild relatives.

Oat researchers at AAFC have developed sequencing-based bioinformatics analysis software (Tinker *et al.* 2016) that can generate thousands of reproducible markers from short-read DNA sequences. More than 4,600 cultivated oat lines have been analyzed using this method and over 200,000 potential markers identified. Of these markers, 30,000 have known genetic positions on the publicly available hexaploid oat consensus map (Bekele *et al.* 2018a). Recently, the oat genomics group designed the first whole genome exon (exome) sequence capture, targeting more than 25,000 target genes and sequenced gene coding sequences of more than 120 accessions of cultivated and wild oat species. The exome capture data will facilitate the annotation of *denovo*-assembled genomes, as well as aid in the development of more specific gene-linked molecular markers.

This work has allowed for the identification of genes that control important traits. For example, marker-assisted selection is being used to follow genes for crown and stem rust resistance in oat breeding lines. Single genes for resistance in a line can be identified easily without the use of genetic sequence information, but single genes are also easily overcome by evolution of the pathogen population. To ensure that multiple rust resistance genes have been incorporated into a line, the use of molecular markers is required. Targets for improvement using genomic selection include yield and beta-glucan, traits essential for food security and better health. Sequencing-based genome-wide markers are being used to calibrate selection models for the two traits, as well as to predict the performance of lines that have not been subjected to replicated field trials and seed quality assessment (Bekele *et al.* 2018b).

Similarly AAFC has contributed to marker development in *Brassica napus*, an economically important crop for Canada that supports a \$27B industry. AAFC led the development of a 60,000 SNP array for *B. napus* through the capture of genetic sequence variation from multiple lines and the careful scrutiny and analyses of these data (Clarke *et al.*, 2016, TAG 129: 1887-1899). The so-called 60K Brassica SNP array is now used globally as a highly effective tool for identifying and following positive alleles for traits of interest in breeding programs (Mason *et al.*, 2017, TAG 130: 621-633).

Genomic work is now included to varying degrees in Canada's breeding work and involves many partners from many countries. Most genomics sequencing information and genomics tools are made publicly available through online platforms and databases.

Genomic work is directly linked to and feeds into breeding programs at AAFC, in Canada and in partners organizations. One of the first and most important databases for DNA sequence and related information is GenBank® (<https://www.ncbi.nlm.nih.gov/genbank/>). The National Center for Biotechnology Information (NCBI) in the United States is responsible for GenBank, which contains millions of DNA sequences submitted by individual laboratories and international organizations such as the European Molecular Biology Laboratory (EMBL) and the DNA Database of Japan (DDBJ). Many peer-reviewed scientific journals will not allow papers to be published unless their sequencing data have been submitted to GenBank. Continuing with the example of oat, a search for '*Avena*' in the nucleotide database returns 137,681 hits.

This number does not include much of the recent molecular marker sequence data from oat, as other public databases are sometimes more suitable and may also be used as repositories. The oat community

is fortunate to have two such databases, GrainGenes and T3/Oat, both maintained by the United States Department of Agriculture (USDA). GrainGenes (<https://wheat.pw.usda.gov/GG3/>) manages genomic, genetic, germplasm, and phenotypic data, curating peer-reviewed datasets for researchers working on wheat, barley, rye, and oat. The T3/Oat database is part of the Triticeae Toolbox system of databases (<https://triticeaetoolbox.org/>) for wheat, barley, and oat. These databases contain published and unpublished data of the same type as GrainGenes, while also providing tools to enable breeders and others to map genotypes to phenotypes for use in, for example, marker-assisted and genomic selection programs.

Genomic work has allowed to identify genes that control important traits. For example GWAS (Genome Wide Association Study) has identified genetic regions controlling maturity and seed composition (Plant Biotechnology Journal 13:211-221, 2015) in Canadian soybean. Examples of databases are a comprehensive description of genome wide nucleotide and structural variation in short-season soya bean (Plant Biotechnology Journal 16:749-759, 2018).

Examples or particular cases where the generation and use of genomic sequence information has contributed to the conservation and sustainable use of plant genetic resources and sharing benefits associated with the use of those resources are highlighted below.

Genome Canada and several stakeholders of the flax industry in Canada including provincial governments, Flax Council of Canada and producer organizations funded the TUFGEN (*Total Utilization Flax GENomics*) project from 2009-2014 in part to assess the genotypic and phenotypic diversity of the flax collection preserved by Plant Gene Resources Canada (PGRC) as it pertained to characterization, conservation, breeding and genetic improvement. The reference sequence of the flax genome published in 2012 was, at the time, the 12<sup>th</sup> plant genome sequenced worldwide but the first to be sequenced by a Canadian team. Subsequently, a core collection of 400 accessions representing most of the genetic variability of the 3,000+ accessions of flax maintained by PGRC was sequenced and phenotyped for agronomic, disease and quality traits at multiple locations for 4-5 years. The 1.7M single nucleotide polymorphisms generated clearly illustrated the breadth of the genetic diversity of the collection. Accessions from the Abyssinian, Indian and Pakistani regions harbored the greatest amount of diversity, suggesting to conservationists better geographic sampling to capture the full extent of diversity. Several genome-wide association analyses were conducted and identified variants associated with 27 important traits including yield, seed size, straw strength and disease resistance. More than 125 scientific articles and book chapters were published by the TUFGEN team during that time and at least 50 additional contributions stemming from the research data have been published since, including a gold standard revision of the reference genome sequence into its 15 chromosomes. A database housing pedigree information and all genotypic and phenotypic data currently fuels research and breeding activities. Canadian breeders have benefited greatly from the outcome of TUFGEN and have now implemented marker-assisted and genomic selections in their programs.

Wheat contributes 20% of the proteins and 20% of the calories of the world population, making it the most important crop for food security. With a farm gate value of \$6B and a value-added industry nearing

\$15B, wheat is the most important Canadian crop. Genome Canada and several stakeholders of the wheat industry in Canada including provincial governments, producer organizations and the seed industry, funded two consecutive large-scale projects starting in 2010 on wheat: Canadian Triticum Advancement through Genomics (CTAG and CTAG2). The Canadian team led the international effort to sequence the large and complex polyploid wheat genome that culminated with landmark publications in Science in 2018 (doi: 10.1126/science.aar7191) and the sequencing of 15 additional accessions ([www.10wheatgenomes.com/](http://www.10wheatgenomes.com/)), including the Canadian cultivars CDC Stanley and CDC Landmark. High density genotyping for 250 lines from the Plant Gene Resources of Canada wheat collection are being anchored to these genome sequences. The PGRC germplasm was carefully selected to capture novel diversity that is absent in Canadian wheat breeding germplasm. Genome-wide association mapping studies are in progress, using extensive phenotypic data for biotic and abiotic stresses collected in western Canadian environments. The CTAG2 project was also instrumental in developing new DNA markers for resistance to Fusarium head blight, the most devastating disease of wheat in Canada. Several resistance genes for disease and insect resistance were also cloned. A new genotyping system tailored to capture the diversity in Canadian wheat breeding programs is currently being used as a new tool to support genomic selection experiments in wheat. A total of 75 peer-reviewed scientific articles, most in top-tier journals, were published from CTAG and CTAG2. Together, these technologies pave the way to improve the efficiency of wheat breeding programs in Canada. CTAG and CTAG2 provided the foundation for 4DWheat, as new pan-Canadian proposal which will delve even further into germplasm collection characterization and applications through its four pillars: Diversity, Domestication, Discovery and Delivery.

GSD has also been used to understand gene flow related to herbicide resistance. Since the beginning of agriculture, controlling weeds has been a major concern for farmers. Initially these were managed through labour intensive hand weeding. Currently, over 450 weed species globally have evolved resistance to at least one, but often many, herbicides. Within Canada, just over 60 species are herbicide resistant and in the United States there are over 150, some of which are spreading or will spread into Canada. Furthermore, with the introduction of crops modified to have herbicide resistance, the genes that confer these traits could move into wild relatives, providing additional challenges for weed management on the farm. Herbicide resistant weeds, whether they evolved or received this resistance through gene flow, are a clear threat to agricultural production within Canada and globally. Within the *Brassicaceae*, it has been documented that transgenes for herbicide resistance can move from crops (e.g. canola, *Brassica napus*) into weeds (e.g. Bird rape, *Brassica rapa*). This creates herbicide resistant weeds that are difficult to control. Further, if canola crops with different herbicide resistance genes are grown, it is possible for these traits to combine resulting in weeds with multiple resistances. As a result, it is important to assess the potential of transgenes escaping from *Brassicaceae* crops

The research conducted at Agriculture and Agri-Food Canada to address this threat relies at every stage on free and open access to genomic resources, including genetic sequence data. As a first step, we use freely available information on genes from species around the world to determine which weed species are most closely related to novel crops and, therefore, most likely to receive transgenes. This enables efficient allocation of resources to where the risk of gene flow is the highest. Second, we use similar information to develop molecular markers to detect hybrids between crops and their wild relatives,

allowing determination of the rate at which this hybridization is likely to occur. This information is needed by the Plant Biosafety Assessment Office at the Canadian Food Inspection Agency to make informed, science-based decisions on the unconfined release of enhanced cultivars. Finally, the availability of accessible published genetic sequence data allows us to create the tools needed to dissect the consequences of hybridization so that an informed risk benefit analysis can be made including: 1) pinpointing the potential frequency of this hybridization in nature, 2) determining the portions of the genome that are most likely to be exchanged, and 3) quantifying how quickly a transgene may spread once introduced. For example, a glyphosate resistant (e.g. Roundup resistant) ecotype of the weed *Kochia scoparia*, has recently invaded Canada from the USA and the gene that confers this resistance is spreading through Canadian populations. This species has the potential to cause devastating losses in wheat and soybean. Its genome has just been made available allowing us unprecedented ability to understand how seed and pollen movement contribute to the spread of these genes through populations. This information will contribute to the development of more effective and sustainable weed management strategies for this species.

This work supports the mainstreaming of biodiversity in agricultural production, and the promotion of “awareness, use and sustainable use of agricultural ecosystems; improve its productivity and diversification; integrate positive incentives for biodiversity in field production and reduce agricultural pollution”.

#### **Relevance of DSI on PGRFA for Food Security and Nutrition**

Work has been carried out on soybean where a genetic sequence was associated with reduced cadmium accumulation in seed ([Theor Appl Genet.](#) 2010 121:283-94). Molecular markers allow soybean breeders to select against Cd accumulation.

#### **Public availability of GSD generated on PGRC germplasm**

Upon completion of a project generating and analysing GSD on the germplasm of PGRC, all generated sequence data and some of derived sequence data were/are deposited into public databases such as National Centre for Biotechnology Information (NCBI) and Figshare (<https://figshare.com/>) which are universally accessible. Evidence for public use of some GSD generated on PGRC material was also noted. For an ongoing project generating GSD data on PGRC germplasm, the database structures to provide public access to GSD are currently being developed within AAFC. The GRIN-Global gene bank information system used by PGRC has no capacity to hold GSD data and the PGRC accession numbers will be used to point to the various databases that can hold GSD on PGRC gene bank accessions. Work on the Global Information System (GLIS) is focussing on the development of Digital Object Identifiers (DOIs) as permanent unique identifiers for PGRFA. Canada is in the process of migrating to GRIN-Global and both GRIN-Global and Genesys (the two genebank platforms) are currently being enhanced to accommodate DOIs.

## Capacity building in GSD by AAFC

AAFC trains on an ongoing base students, including students from developing world countries, in generating, analyzing and applying GSD on PGRFA for usage in conservation and plant breeding. Specifically, training on GSD on PGRFA has been provided to foreign scientists from Chile, China, Columbia Ethiopia, Mexico, Nigeria and Pakistan.

## Concluding Remarks

The various applications of genetic sequence data make important contributions to the utilization of genetic resources for food and agriculture, and to practical market-based or producer-oriented options that can enhance improvement of crops and facilitate their effective conservation.

Restricting the dissemination of genetic sequence data would limit the benefits that can be generated from the use of genetic resources and jeopardize their efficient conservation. Sharing information on genetic sequences is the corner-stone of entire scientific disciplines such as phylogeny, phylogenetics, molecular biology, molecular genetics, and many more, contributing to the global understanding of life in all its forms. Understanding diversity within a species is needed to maintain the viability of ecosystems and ecosystem services. Genetic sequence data are not genetic resources and should not be treated as such.

Genetic sequence data has the potential to bring numerous benefits to Canadian and global society. In this document, we provided examples of research projects that directly support the objectives of the international agreements on biodiversity and genetic resources. Such projects require a flow of genetic sequence data among countries that is not complicated or expensive in order to advance these objectives.

## References:

Bekele, W.A., Wight, C.P., Chao, S., Howarth, C.J., and Tinker, N.A., (2018a) Haplotype based genotyping-by-sequencing in oat genome research. *Plant Biotechnology Journal* **16**: 1452-1463.

<https://doi.org/10.1111/pbi.12888>

Bekele, W., Itaya, A., Wight, C., Frégeau-Reid, J., Yan, W., and Tinker, N. (2018b) Genomics-assisted breeding of oat at AAFC, Ottawa. Proceedings of the American Oat Workers' Conference, Seattle, USA, June 19-21. <https://oats.ce.byu.edu/sites/oats.ce.byu.edu/files/bekele.pdf> (Slide presentation: [https://oatnews.org/oatnews\\_pdfs/2018etc/AOWC2018\\_22\\_Bekele.pdf](https://oatnews.org/oatnews_pdfs/2018etc/AOWC2018_22_Bekele.pdf))

Tinker, N.A., Bekele, W.A., and Hattori, J. (2016) Haplotag: Software for Haplotype-Based Genotyping-by-Sequencing Analysis. *G3 (Bethesda)* **6**: 857-63. <https://doi.org/10.1534/g3.115.024596>