

Open access to genetic sequence data maximizes value to scientists, farmers, and society

Jim Gaffney^{a,*}, Redeat Tibebe^b, Rebecca Bart^c, Getu Beyene^c, Dejene Girma^d,
Ndjido Ardo Kane^e, Emma S. Mace^f, Todd Mockler^c, Thomas E. Nickson^g, Nigel Taylor^c,
Gina Zastrow-Hayes^a

^a Corteva Agriscience, USA

^b College Biological Sciences, University of Minnesota, USA

^c Donald Danforth Plant Science Center, USA

^d National Agricultural Biotechnology Research Center, Ethiopian Institute of Agricultural Research, Ethiopia

^e Institut Sénégalais de Recherches Agricoles, Centre d'Étude Régional pour l'Adaptation à la Sécheresse, Senegal

^f Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Australia

^g Nickson Research and Consulting, USA

ARTICLE INFO

Keywords:

Nucleotide sequences
Digital sequence information
Genetic sequence data
Open access
Convention on biodiversity

ABSTRACT

Open access to genetic sequence data, often referred to as Digital Sequence Information, has been available since genome sequencing became possible and creates both monetary and nonmonetary value. Nonmonetary value is created when scientists access sequence data for discovery, collaboration, and innovation. Monetary value is created when genetic variability is leveraged to develop more robust and resilient crop plants, vibrant seed systems, more sustainable agriculture, and food security for consumers. Millions of dollars have been invested in curating and creating access to sequence databases and scientists from almost every country in the world have accessed these databases, free of charge. This access may now be threatened by well-meaning policy-makers who have not consulted with the scientific community. Monetizing or creating greater regulation of genetic sequence data would create barriers to innovation, partnering, and problem-solving.

1. Introduction

The challenges of agriculture now and in coming decades will be to meet multiple demands of sustainability: food security and improved nutrition, conservation of biodiversity and healthy soils, and to create greater prosperity for small-holder farmers under the pressure of climate change and a growing and more prosperous global population (Gaffney et al., 2019; Sayers et al., 2019; United Nations Sustainable Development Programme, 2016). This will rely, in part, on a combination of approaches including traditional breeding methods, genomic selection, genome editing and other biotechnological approaches to improve agricultural sustainability under both biotic and abiotic stressors.

Progress in these areas relies on availability of foundational genomic resources which provides researchers with greater depth and breadth of data and information than previously available. Genetic sequence data (GSD), referred to in the context of ongoing international negotiations as Digital Sequence Information (DSI), has been historically made

available to scientists around the world, unencumbered and free of charge, through multiple, publicly available databases. As of 2016, there has been pressure from certain parties to the Convention on Biodiversity (CBD) and other influencers, to include GSD in an access and benefits sharing (ABS) scheme, in a manner similar to that attempted with plant genetic resources for food and agriculture (Halewood et al., 2018a) (Aubry, 2019; Laird et al., 2020). The issue has become more urgent as policy-makers "under the auspices of the United Nations" have been considering including GSD in an ABS scheme with little to no consultation of the scientific community (Laird et al., 2020). Negotiations under the CBD, Nagoya Protocol and International Treaty have reached a stalemate and policy makers may use GSD as a "bargaining chip" or compromise, further threatening the objectives of the CBD while destroying the true value of GSD.

The CBD, which was adopted in 1993, has three objectives: 1) conservation of biological diversity; 2) sustainable use of biological diversity; and 3) fair and equitable sharing of benefits from the use of genetic resources (Brink and van Hintum, 2020; UN2020, 2020).

* Corresponding author.

E-mail address: jim.gaffney@corteva.com (J. Gaffney).

<https://doi.org/10.1016/j.gfs.2020.100411>

Received 1 June 2020; Received in revised form 13 July 2020; Accepted 14 July 2020

Available online 31 July 2020

2211-9124/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations

ABS	access and benefit sharing
CBD	Convention on Biodiversity
DDBJ	Bioinformation and DNA Data Bank of Japan
DOE	U.S. Department of Energy
DSI	digital sequence information
EMBL-EBI	European Bioinformatics Institute
ENA	European Nucleotide Archive
GRSD	genetic resource sequence data
GSD	genetic sequence data
GR	Green Revolution
INSDC	International Nucleotide Sequence Database Collaboration;
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
JGI	Joint Genome Institute
NCBI	National Center for Biotechnology Information
NSD	nucleotide sequence data
PPD	postharvest physiological deterioration
SDGs	United Nations Sustainable Development Goals
SNP	single nucleotide polymorphism
SSA	Sub-Saharan Africa.

Proponents for inclusion of GSD in an ABS scheme believe that benefits derived from the use of GSD are not shared in a fair, balanced, or equitable manner and that monetization and regulation of GSD would better serve the third objective. Various proposals have been advanced, including development of guiding principles and self-reporting mechanics or extending the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and Nagoya Protocol to include GSD in an ABS scheme (Halewood et al., 2018a, 2018b) (Aubry, 2019). When discussing the impact of regulating GSD, it is important to understand better the current state of availability of GSD, how it is accessed and by whom, where value is derived from GSD, and how use of GSD is critical to conservation of agricultural biodiversity, food security and advancing science equitably. Likewise, full understanding of all benefits generated and shared directly or indirectly is critical to any ABS system.

Historical parallels can be drawn between GSD and the value and benefits resulting from sharing of plant genetic resources. For approximately 10,000 years, the movement of crop varieties (or germplasm) between regions and countries has contributed to agricultural productivity, farmer prosperity, food security, and to improving the environment (Cassaday et al., 2001; Jorasch, 2019). Perhaps the most famous example of near unabated movement and sharing of germplasm was during the Green Revolution (GR), when “shuttle-breeding” programs, took advantage of breeding selection in multiple environments around the world to accelerate development of highly productive wheat and rice varieties. Productivity gains from the GR saved up to 27 million hectares from agricultural production, thereby conserving natural habitat and biodiversity, lowering the environmental footprint of agriculture, and reducing poverty and hunger (Pingali, 2012; Stevenson et al., 2013). The greatest beneficiaries of the GR were consumers, especially in low income countries. Farmers benefited by becoming more resilient in the face of demands for intensification of production. Scientific research also benefited when a new, global community of plant breeders and agronomists was created, armed with the latest research tools and technology that had previously only been available to select scientists within western industry and academia. As M.S. Swaminathan stated, “The concept of shuttle-breeding transcended continental boundaries and a global college of wheat breeders emerged” (Swaminathan, 2009).

An example of the value, complexity, and contributions of the GR are

summed up in the wheat variety *Sonalika* (Fowler et al., 2001). Farmer developed landraces from 17 countries and breeding lines from 14 countries across six continents contributed to the final variety, released in India in 1966. A similar story is found in rice variety IR72, which included backgrounds traced to 22 landraces from seven countries, including five rice varieties bred in the U.S. (Ziegler, undated). Determining the value and contributions of each parent to the final product in these examples would be practically impossible. The individual landraces and breeding lines would have had little value compared to that available to farmers at the time. Through a combined effort of investment, cooperation and access to germplasm, tremendous value was created for farmers, consumers, and the scientific community. Environmental benefits should also be acknowledged. By making farmers more productive, land is spared from cultivation, greater biodiversity is conserved, soil health more likely to improve, and greater opportunities generated to reduce greenhouse gas emissions (Balmford et al., 2018; Gaffney et al., 2019; Poffenbarger et al., 2017; Rosegrant et al., 2014; Tilman et al., 2011).

The value of genetic sequence data holds similar or even greater promise as the GR, but only if the same level of open access and sharing remain available to scientists. Policy- and decision-makers must exercise care and thoughtfulness to create greater cooperation and help meet the demands placed on agriculture now and in the future. The objectives of this paper are to 1) define GSD more precisely; 2) provide examples of the current investment in open access databases; 3) identify where value is created through use of these databases to enable open access, exchange and use of GSD, and 4) develop a productive vision for the future.

2. Defining genetic sequence data (GSD) and relevant terminology related to Digital Sequence Information (DSI)

GSD has also been referred to as Digital Sequence Information (DSI) (Heinemann et al., 2018) (Aubry, 2019; CBD/COP/DEC/XIII/16, 2016), and confusion exists on what DSI encompasses and what is the actual value and who benefits from the sharing of DSI. The term DSI has been understood to refer to electronically stored and exchanged DNA sequence information. Since initial use of the term in the context of the CBD, the scope of the term has expanded and discussions on this subject matter are now confounded by multiple interpretations. This contributes to confusion concerning the value of DSI, what “information” is actual conveyed by DSI, and how much plant scientists understand DSI. As Laird et al. (2020) state, DSI is “a negotiated placeholder, the meaning and scope of which remain in dispute.” This paper proposes the term “Genetic Sequence Data” or “GSD”. Reference has also been made to the term “Genetic Resource Sequence Data” or “GRSD” as proposed by the International Chamber of Commerce, and to “Nucleotide Sequence Data” or “NSD”, as proposed by many scientists. The term GSD is clearer than DSI, scientifically-precise and facilitates a more fact- and science-based discussion. It also better describes the output used in developing scientific knowledge and beneficial products. The scientific community considers GSD to include “the spectrum of data raw reads, through alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations” (<http://www.insdc.org/>) that is made available in public access databases. We would argue that sequence data is not “information” and has limited to no value until time, money, expertise, organizational skills, ingenuity and dedicated human effort are invested in deciphering sequences to generate knowledge.

Proponents of greater governance and regulation of GSD often justify their position based on the notion that DSI affords developers a way to avoid benefit sharing. They argue that there is a “real” possibility of scientists “digitizing genetic resources and synthesizing the required nucleic acid fragments with the use of openly accessible DSI” (Laird et al., 2020). This argument of “dematerializing” sequence data and “rematerializing” something of value is overstated. In fact, converting sequence data into information or a more productive crop plant requires

significant time and effort. Generating the sequence data itself is only the starting point and represents a fraction of the effort required to create real value further downstream. A GSD for a single sample is a unique representation of recombination events influenced by breeder selection, domestication events, evolution, and the environment in which a plant grew and continues to grow. However, when viewed in isolation, both in terms of samples and other data types, GSD themselves hold limited value, as it is the integration of data from multiple sources that enables specific features of the sequence to be characterized and associated with specific functionality. Hence to capture value from GSD, not only are additional data types (e.g. phenotypic, biochemical data) required, but it is also critical to make comparisons across multiple accessions and/or populations to identify sequence variations that are associated with specific target traits. For many traits of relevance to breeding programs, these are quantitative in nature and as such are controlled by a large and complex interacting gene network. Small, isolated GSD would not provide the power to detect such complexity, and hence open access and exchange of GSD would greatly increase value to individual datasets. The complexity of extracting value in a meaningful biological context also underscores the difficulty of attributing specific GSD to regions or countries. Origins of, or contributions to, the GSD are not likely traceable, especially considering the millennia of evolution that have created the sequences and the vast amounts of GSD currently stored, curated, and made available to scientists in open-access databases. Examples of these databases follow.

3. Open-access sequence database – examples and policies

Genetic sequence databases were created to provide free and unrestricted access to data archives for scientists anywhere and numerous GSD databases are now available. The value of open access was fundamental to the founding of these databases.

The largest database available is the collaborative framework of the [International Nucleotide Sequence Database Collaboration \(INSDC\)](http://www.insdc.org/) (<http://www.insdc.org/>). The INSDC includes the European Bioinformatics Institute (EMBL-EBI) and its European Nucleotide Archive (ENA), the Bioinformatics and DNA Data Bank of Japan (DDBJ), and the National Center for Biotechnology Information (NCBI) and GenBank of the United States (Table 1). Collectively, the partnering organizations strive for the same goals: free and unrestricted access to data archives, including access to a comprehensive resource of nucleotide sequences. As an example, the INSDC has a uniform policy of free and unrestricted access to all data records in their database; that no use restrictions or licensing requirements will be included in any sequence data records; nor restrictions be placed on redistribution or use of the databases. Once database records are submitted to the INSDC, they remain permanently accessible. This policy was established so that any scientist, anywhere, “can access these data to plan experiments or publish any analysis or critique” (www.insdc.org/policy/html).

The three contributing organizations to the INSDC databases provide ever-expanding resources and infrastructure for improved data management and access. Examples include a new supercomputer at DDBJ ();

Table 1

The three databases of the International Nucleotide Sequence Database Collaboration (INSDC), respective websites and recent publications with updates on the databases.Database.

	Country/ Region	Website	See also:
European Nucleotide Archive (ENA)	Europe	https://www.ebi.ac.uk/	Cook et al. (2019)
Bioinformatics and DNA Data Bank of Japan (DDBJ)	Japan	https://www.ddbj.nig.ac.jp/index-e.html	Ogasawara et al. (2020)
GenBank	United States	www.ncbi.nlm.nih.gov/genbank/	Sayers et al. (2019)

the BLAST program for sequence-based searches, and the “Entrez” system to aid data submission and retrieval, YouTube tutorials, handbook and help manuals (Sayers et al., 2019). All are easily accessible via the NCRI home page (ncbi.nlm.nih.gov). New resources, search features, and training at EMBL-EBI are continually added with a stated role to “collate, integrate, curate, and make freely available to the public the world’s scientific data” (Cook et al., 2019). The growth of these databases has been impressive, especially since publication of the human genome in the year 2000. Fig. 1 depicts select milestones in the growth and use of GenBank over the past 30 years. GenBank contains over 7.72 trillion bases (NCBI Insights 2020) from over 450,000 species (Sayer et al., 2020). While the focus of GenBank is on human health, it also includes over 59 billion base pairs derived from plant species. GenBank activity is included within the U.S. National Institute of Health, with a budget of \$456.9 million in 2020 (breakouts of specific budget numbers devoted to GenBank are not available).

Examples of databases focused on crops include Phytozome, the Rice Genome Annotation Project, GrainGenes, MaizeGDB, SoyBase, Legume Information System, and EnsemblPlants (Table 2). The Phytozome database is within the U.S. Department of Energy (DOE) Joint Genome Institute (JGI). Its mission is “to provide the global research community with access to the most advanced integrative genome science capabilities in support of the DOE’s research mission”. Funded by the National Science Foundation and an international effort with the help of Japan, The Rice Genome Annotation Project provides genome sequence from the Nipponbare subspecies of rice and annotation of the 12 rice chromosomes. The United States Department of Agriculture/Agricultural Research Service (<https://www.ars.usda.gov/>) is home to GrainGenes, MaizeGDB, SoyBase and Legume Information System. GrainGenes, featuring wheat, barley, rye, and oat holds data such as primer sequences, polymorphism descriptions, genotype and trait scoring data, experimental protocols used, and photographs of marker polymorphisms, disease symptoms and mutant phenotypes. MaizeGDB includes a number of inbreds and open-pollinated varieties and states that it is “a community-oriented, long-term, federally funded informatics service to researchers focused on the crop plant and model organism *Zea mays*.” SoyBase contains soybean genome sequences, supporting gene sequences, and is supported by the SoyBase sequence browser. The Legume Information System (LIS) is a community resource for crop improvement across legume species (Dash et al., 2015). While the U.S., China, and India are the most frequent users, nearly every country in the world has accessed data from MaizeGDB, SoyBase, or LIS, with over 100,000 global users annually (Fig. 2). One of the most diverse databases is EnsemblPlants, with new genomes (Release 47) including pineapple, pistachio, almond, and olive tree (Howe et al., 2020). The Ensembl genome annotation system was initially developed for vertebrate genomes by the European Bioinformatics Institute and Wellcome Trust Sanger Institute, and since 2009 has included bacteria, protists, fungi, plants, and metazoan. Funding agencies for Ensembl have included the European Union, the Bill and Melinda Gates Foundation, Biosciences for the Future, National Human Genome Research Institute, and the National Science Foundation.

The databases given as examples here have been maintained and enhanced with every effort made to ensure open access and exchange of GSD and represent a critically important global investment and public good. The dividends of this investment are observed in the value created downstream, including discovery, innovation, and communities of scientists using GSD to solve agricultural challenges of productivity and sustainability.

4. Creating value from GSD

This investment in, and unencumbered access to use of GSD benefits society, consumers, farmers, and the global research community. The need and potential value of open access to GSD is perhaps best evidenced in Africa where there is a diversity of crops and cropping environments,

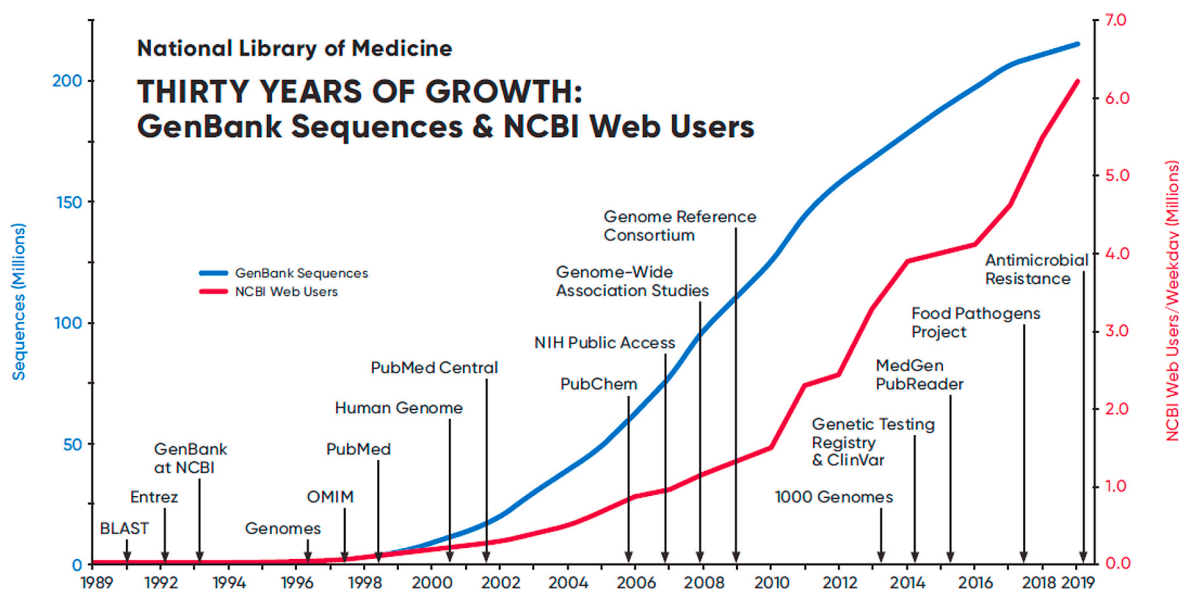


Fig. 1. Growth of GenBank sequences and NCBI web users through 2019. Figure borrowed from the Department of Health and Human Services National Institutes of Health. Website accessed 03 March 2020 (https://www.nlm.nih.gov/about/2021CJ_NLM.pdf).

Table 2

Examples of crop databases, respective websites and citations, and fiscal year 2020 budget estimate (if available) in support of the database.

Database	Website/Citations	Fiscal Year 2020 Budget Estimate (USD)
Phytozome	https://phytozome.jgi.doe.gov/pz/portal.html Goodstein et al. (2014)	NA
EnsemblPlants	https://plants.ensembl.org/index.html Howe et al. (2020)	NA
Rice Genome Annotation Project	http://rice.plantbiology.msu.edu/https://rapdb.dna.affrc.go.jp/ Kawahara et al. (2013) Ouyang et al. (2007)	Unfunded ^a
GrainGenes	https://wheat.pw.usda.gov/GG3 Matthews, D.E. 2003	\$1,224,000 ^b
MaizeGDB	https://www.maizegdb.org/ Lawrence et al. (2007)	\$1,117,000 ^b
Soybase/Legume Information System	https://soybase.org/sb_about.php https://legumeinfo.org/ Grant et al., (2010)	\$2,450,000 ^b

^a The Rice Genome Annotation Project at MSU (<http://rice.plantbiology.msu.edu/>), Personal communication with Dr. C. Robin Buell, Michigan State University.

^b Personal communication with Jack Okamoto, [USDA/ARS](https://www.usda.gov/).

vast biodiversity, a burgeoning scientific community, and a rapidly growing population. The smallholder farmer crops tef, Sorghum, pearl millet, and cassava provide specific examples of how value could be, and in some cases already is, being derived when GSD is openly available.

Tef (*Eragrostis tef*) is the most important cereal crop in Ethiopia with regard to cultivation area, economy and national consumption; accounting for 22% of agricultural land, generating the highest revenue of any crop to 6.6 million smallholder farmers and is the primary diet for the majority of the Ethiopian and Eritrean population (FAO, 2020) (Minten et al., 2018). Despite this, tef remains one of the least productive cereals with an average yield of 1.5 tons/hectare. This contrasts with maize, the second most important crop in the country, that yields 3.3 tons/hectare in Ethiopia (Minten et al., 2018). This low productivity

is mainly due to a lack of advanced tef research focused on plant improvement, with more than 80% of farmers still cultivating traditional landraces (Smith et al., 2012). The major biotic limitations of tef are plant lodging (bending or breaking of plant stem due to wind or rain), small seed size and susceptibility to acidic soils prevalent in the tropics (Girma et al., 2014; Mebratu et al., 2016). The leading cause of low yield is plant lodging and can account for up to 30% yield loss as well as adding to the difficulty of both manual and mechanical harvesting (Ketema, 1997). Although wide phenotypic diversity exists within tef germplasm collections, lack of full and annotated genome (GSD) makes it difficult to decipher and exploit the genetic diversity within these accessions. Tef breeders must rely primarily on direct phenotypic selection and conventional breeding. The availability of openly accessible GSD on farmer-preferred tef varieties would revolutionize tef improvement by allowing application of molecular breeding and advanced plant breeding approaches for achieving enhanced yield potential and beneficial agronomic characteristics.

Numerous agroecological zones in Ethiopia (Fig. 3) present a challenge to plant breeders because each zone requires focus on unique abiotic and biotic stressors, and each zone requires breeders to account for genetic by environmental by management interactions, with potentially different breeding strategies for each zone. However, knowledge of genetic diversity within tef provided by GSD could be coupled with genotype by environment interaction studies to provide farmers with varieties best adapted to specific agroecological zones. Additionally, the wide array and ever-increasing research data and techniques available in related cereal crops such as rice and wheat present opportunities for tef improvement via technologies such as molecular breeding and genome editing. Molecular breeding; a technique that uses genetic markers rather than phenotypic variability for selecting high-performing varieties, can accelerate breeding and increase yield potential. Comparative genomic analysis can be used to identify gene orthologs - genes possessing similar sequence and functions across related plant species. This approach can aid in identifying tef gene orthologs known to regulate yield-associated traits in major cereal crops. For example, Green Revolution semi-dwarf varieties of rice and wheat developed via breeding have been shown to possess unique mutations (nucleotide changes) in two genes controlling plant height (Peng et al., 1999; Spielmeier et al., 2002). Lodging is a major constraint for tef farmers, but no such dwarf varieties exist in the crop. Knowledge gained from rice and wheat can be used to identify the causal gene mutations

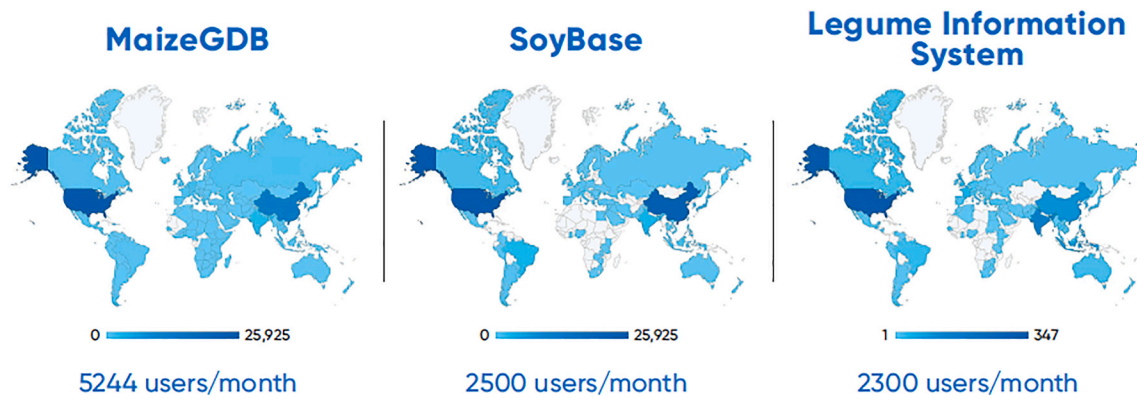


Fig. 2. Users by country and by month of select crop genomics databases. Figure compliments of Carson Andorf and Lisa Hartman, USDA/ARS.

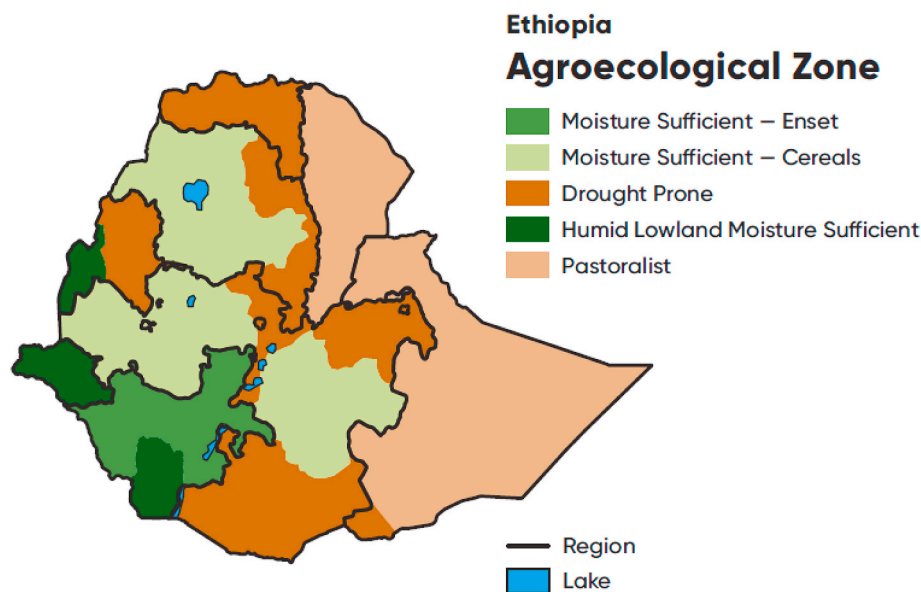


Fig. 3. Agroecological zones in Ethiopia (Figure borrowed from Minten et al., 2018.).

and generate the analogous changes within the tef genome via gene editing, but only if the genomic information for all these crops is available to researchers. Open and accessible GSD therefore holds significant value for genomic-supported tef research, which in turn will improve seed systems in Ethiopia leading to enhanced farmer well-being and increased chances of meeting the United Nations Sustainable Development Goals (SDGs) set for 2030, including “zero hunger” as formulated under SDG 2, in an integrated and efficient manner.

Sorghum (*Sorghum bicolor* (L.) Moench) has amongst the largest libraries of GSD available of any African indigenous crop, including multiple reference genome assemblies (hosted at Phytozome; Cooper et al., 2019; McCormick et al., 2018; Paterson et al., 2009), multiple re-sequenced accessions (Mace et al., 2013; Zheng et al., 2011), in addition to transcriptomics, SNP and QTL databases (Makita et al 2015) (Luo et al., 2016; Mace et al., 2019). Amongst the cereals, Sorghum is one of the best adapted to drought and high temperatures, and globally produces approximately 70 million metric tons of grain, making it the world’s fifth most important crop after maize, wheat, rice and barley. More than 90% of total global Sorghum harvested area is from Africa and Asia, feeding around half a billion people in sub-Saharan Africa and Asia alone. However sorghum remains below its yield potential in Africa and Asia with multiple biotic and abiotic challenges negatively impacting productivity, with Africa only accounting for 41% of global sorghum production and Asia only accounting for 18% (Mundia et al.,

2019). GSD has an important role to play in characterizing the genetic variation available for crop improvement programs to develop superior varieties. This diversity has been characterized genetically in numerous studies and germplasm collections, including the Sorghum Association Panel (e.g. Casa et al., 2008; Morris et al., 2013) and a set of diverse resequenced sorghum lines (Mace et al., 2013), and the data has been made available publicly through the Sequence Read Archive at NCBI. Studies to identify functional sequence variations associated with important agronomic traits have made use of the publicly available GSD and have investigated sequence variation in candidate genes for numerous traits including grain size (Tao et al., 2017), starch content (Gilding et al., 2013; Campbell et al., 2016), nitrogen use efficiency (Diatloff et al., 2017; Massel et al., 2016), heat tolerance (Chen et al., 2017) and anthracnose resistance (Cuevas et al., 2018). These studies have been able to integrate data from multiple sources to further mine and extract value from publicly available GSD and to generate critical information for sorghum breeders and researchers to enhance the rate of genetic gain and deliver superior sorghum varieties to farmers.

Yet another argument for open access to GSD is the tropical root crop cassava (*Manihot esculenta* Crantz). Cassava is widely grown in the tropics and sub-tropics, with an estimated annual production of over 292 million tonnes (MT) in 2017. Cassava is grown mainly as a staple food by hundreds of millions of subsistence farmers, used as feed for animals and increasingly as a source of starch for industrial and bioethanol

applications. Sub-Saharan Africa (SSA) accounts for more than 60% of world cassava production (FAOSTAT, accessed 04/09/2020), but yields are threatened by biotic factors such as cassava mosaic disease, cassava brown streak disease (CBSD), cassava bacterial blight and cassava green mites (Bart and Taylor, 2017; Bull et al., 2011). Annual production losses worth US\$1 billion are estimated due to mosaic disease and bacterial blight alone (Patil et al., 2015). Cassava utilization is also limited by inherent susceptibility of the storage roots to rapid post-harvest physiological deterioration which commences immediately after harvest (Beeching et al., 1998). Conventional breeding in cassava is constrained by the crops high degree of heterozygosity and strong inbreeding depression. Multifaceted innovations are key to combating constraints to cassava production, and require effective modern plant breeding and genotyping technologies and resources (<https://cassava-base.org/>, Fernandez-Pozo et al., 2015), high quality genome and genomic and transcriptome databases (<https://phytozome.jgi.doe.gov/pz/portal.html>, Bredeson et al., 2016), (http://shiny.danforthcenter.org/cassava_atlas/, Wilson et al., 2017), efficacious genetic transformation technologies (Chauhan et al., 2015; Taylor et al., 2012) and genome editing (Bull et al., 2018; Chauhan et al., 2015; Gomez et al., 2019). The value of these innovations and their application have already resulted in improvements in starch quality traits (Bull et al., 2018; Raemakers et al., 2005), discovery of single nucleotide polymorphism (SNP) responsible for provitamin A accumulation in orange fleshed cassava varieties (Welsch et al., 2010), development of iron and zinc biofortified cassava (Narayanan et al., 2019), and brown streak resistance (Wagaba et al., 2017) (Gomez et al., 2019). These resources are indispensable for enhancing discovery and innovation to improve productivity and sustainable utilization of this crop by smallholder farmers. Two specific examples are highlighted. Firstly, the majority of widely deployed cassava landraces and improved varieties in Africa carry a single dominant loci that confers robust resistance to mosaic disease. While earlier studies have mapped this dominant locus to chromosome 12 (Rabbi et al., 2014; Wolfe et al., 2016) this resistance gene is still not identified, nor the molecular mechanism of mosaic disease resistance understood. Knowledge of the gene/sequence that imparts resistance is critical for durable use of this important trait. Secondly, despite many years of research, loss of harvested cassava to postharvest deterioration remains an ongoing challenge, limiting the market potential and value of this crop. A recent report shows the existence of tight relationship between cassava storage root carbohydrate (starch and soluble sugars) and cassava storage root PPD (Beyene et al., 2020). Unencumbered open access to GSD is required to enable this and other important research to continue in cassava.

Many efforts have been made to generate knowledge and resources to cope against these limiting factors. The International Pearl Millet Genome Sequencing Consortium, launched in 2013 in Hyderabad, India, decrypted the genome of Tift23D2B1 inbred and a year later, a first draft of the sequence was made available. In 2017, the reference genome made from a panel of 994 lines capturing the whole diversity of this crop, cultivated as wild progenitor (*P. glaucum* subsp. *monodii*, syn. *Cenchrus americanus* ssp. *monodii*) was reported (Varshney et al., 2017), with an estimated 38,579 genes annotated to establish genomic-assisted breeding. Data sets (https://cegresources.icrisat.org/data_public/PearlMillet_Genome/) can be harnessed not only by the IPMGSC, but by all scientists across in the world.

Unlocking and making public the genome of pearl millet paves the way to new insights and potential to improve efficiency and precision in breeding. Specific genes associated with important agronomic traits such as grain yield, stem and leaf biomass as well as the exceptional tolerance of millet to high temperatures and drought (Varshney et al., 2017) have been identified. Re-sequencing a large number of germplasm lines and several population genomic studies has provided valuable insight into population structure, genetic diversity and domestication history of the crop. For example, the initial genomic information of pearl millet was essential for Burgarella et al. (2018) to more precisely infer

the origin of the domestication of millet 4900 years ago. Numerous marker-trait associations have also been linked to root traits, plant height, yield, grain quality, tolerance to water deficit and heat, resistance to insects and diseases. All are now available to the community to augment research with additional information of genomic variation across a panel of diverse individuals. Genome datasets are being used to accelerate genetic gain for specific traits and improve pearl millet production through hybrid breeding, genomic selection and in population genomics (Kane and Berthouly-Salazar and ref therein). Through improving yield stability and food security of this major dietary staple, the wider benefits will be to contribute to scientific knowledge, agricultural production enhancement, and economic growth for the Sahel region. This focus is highly relevant to the global challenge area of secure, inclusive and resilient food systems.

5. Vision for the future

Unencumbered, open access to and exchange of publicly available GSD contributes to the very essence of many of the United Nations Sustainable Development Goals (SDGs) (United Nations Sustainable Development Programme, 2016) as a critical tool in the conservation and sustainable use of genetic resources. The use of GSD is currently leading to many societal benefits and contributing to innovative products for improving food security and human health.

Allowing the imposition of additional Access and Benefit Sharing (ABS) obligations for the use of GSD - other than through the existing mechanism of mutually agreed terms - will have a significant negative impact on the future of biological research and the benefits resulting from it and policy-makers should include scientists in discussions before settling on any ABS scheme. As Laird et al. (2020) stated, "ABS is a particularly poor policy fit for regulating access to DSI." and "research practices and concepts of ethics and benefit sharing associated with DSI that have evolved in recent decades within the scientific community emphasize openness, transparency, networks and free exchange".

The need for greater innovation and investment in agricultural research – often enabled by open access to and exchange of GSD – could never be greater. The African population will reach 2.5 billion by 2050 (Pew Research Center, 2020). Many countries are land-poor with no further opportunities for expanded cultivation (Jayne et al., 2014). Soils in areas of the tropics and subtropics are becoming degraded beyond repair (Tittonell and Giller, 2013), with severe soil macro- and micro-nutrient imbalance (Pasley et al., 2019) and continual mining of nitrogen from the soil profile (Pasley et al., 2020). Estimates are that climate change will have a greater negative impact on Sub-Saharan Africa (SSA) than on temperate regions (Parry et al., 2004, 2005; Pingali, 2012). Subjecting smallholder farmers to unimproved landrace varieties developed for an environment that no longer exists and is changing rapidly is neither economically nor environmentally sustainable, nor is it morally acceptable. As Pingali (2012) stated, "the need for continued investments in agricultural innovation and productivity growth is as important today as it was in the early years of the GR".

Yet a GR-like focus on only two or three crops is unlikely to deliver the same results as the original GR delivered in Asia due to the diversity of crops cultivated in Africa (Pingali, 2012) and even greater diversity of agro-ecological zones. What is likely to be more helpful is an opportunity to utilize diverse GSD of the many crops to enable yield, yield stability, and quality improvements across the numerous cropping environments. Open access to and exchange of GSD is more likely to create long-term value on par with the GR than if GSD is encumbered with regulation or attempts made to charge for the data. In any case, the revenue generated from any scheme monetizing access and/or use of GSD would pale in comparison to the value creation of higher yield, more resilient crops, and vibrant seed systems developed with these crops through the broad availability of GSD. It is also likely to pale in comparison to the past and future investments in open access databases. The non-monetary benefits of open access – collaborative research

projects, public-private partnerships, scientific discovery, and building networks of scientists – should be an obvious indicator that the current system of open access is working. Current policies governing the open access databases are sufficient to guide the appropriate use of GSD.

Consideration and caution must be given to the risks inherent in the over-regulation and monetization of open access and exchange of GSD, and public research has the most to lose. Administrative, financial, and legal barriers will quickly become the focus of research organizations making determinations as to GSD value, what GSD is accessible and with whom negotiations for access must be conducted (a difficult if not impossible determination in many situations). Research organizations will be entering an administrative and legal quagmire rather than focusing on research. Innovation will be discouraged. Existing disparities between those able to access or use GSD and those who cannot will widen. Conservation of biodiversity, food security, and sustainable agriculture will suffer, especially in those countries most in need. Public-private-partnerships will grind to a halt as each partner ponders administrative and legal issues rather than thinking creatively about technological and scientific innovation; capacity will suffer. Research on minor crops, crop wild relatives, and non-agricultural species, for which we often have limited knowledge, will become more problematic and the value of these species will remain hidden.

The benefits to society of GSD often cannot be monetized, and appear only after hard-fought resources such as time, money, thought, organizational skills, and ingenuity have been invested. The long-term benefit of open access to and use of GSD cannot be over-estimated, and far outweighs any short-term gains to individual organizations or countries through regulation or monetization of GSD. The free access and use of public sources of GSD benefit everyone from conservationists to farmers, from researchers to consumers, and anyone concerned about climate change.

Greater equity in use of GSD will not come from ABS schemes, but through greater cooperation and capacity building. In this context, it is key to assess the needs of developing countries to ensure an effective access to and use of GSD, which should be based on a strategic plan defining priorities and involving all relevant (public and private) stakeholders. Building global communities and networks of scientists to rival the GR is possible and happening. One example is Bellis et al. (2020), in which scientists from the public and private sector and around the world – France, Kenya, Mali, Sweden, the United Kingdom, and the United States – joined forces to understand better the parasitic weed *Striga*, which has plagued farmers for centuries. Other inclusive initiatives are underway, documented in the examples in this paper, and are aided by the history of open access to and use of GSD and cooperation among diverse partners.

We have thus far focused on the importance of open access GSD for plants and agriculture, but human health is also directly impacted. The rapid development of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) diagnostic tests was a direct benefit of the open accessibility of GSD databases to global scientists (Sheridan, 2020). In the early days of the recent coronavirus outbreak, scientists were able to quickly and easily submit and access viral GSD to multiple open access databases, with freedom to operate and limited legal implications. This access allowed a rapid understanding of the cause of the virus, rapid development of diagnostic tools, and a focus on developing solutions. This is but one illustration of the importance of information sharing and collaboration amongst the global scientific community and should be kept in mind when considering any measures that could restrict access to such information.

Maintaining open access to and use of GSD will serve to enable research collaborations, maximize returns on investments and ensure the creation of value to manage societal issues, as well as move scientific discovery more quickly than if GSD is regulated and monetized. A global network of plant scientists will thrive and expand. The greatest value and benefits from GSD will be created if open-access and sharing are maintained and strengthened. Crops like tef, Sorghum, pearl millet and

cassava will be included in the next wave of the GR. Norman Borlaug and M.S. Swaminathan helped create the GR and led scientists and political leaders down one of the most productive paths in the history of agricultural science. The opportunity exists to create a new era of discovery and investment with GSD.

Funding

No funding was provided for the preparation of this manuscript.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jim Gaffney and Gina Zastrow-Hayes are employees of Corteva Agriscience. Tom Nickson is an independent consultant. The authors have no other conflict of interest to declare.

Acknowledgements

The authors wish to acknowledge Kelly Albee and Nancy Wilmeth for their help with final preparation, editing, and formatting of the manuscript.

References

- Aubry, S., 2019. The future of digital sequence information for plant genetic resources for food and agriculture. *Front. Plant Sci.* 10 <https://doi.org/10.3389/fpls.2019.01046>.
- Balmford, A., Amano, T., Bartlett, H., Chadwick, D., Collins, A., Edwards, D., Field, R., Garnsworthy, P., Green, R., Smith, P., Waters, H., Whitmore, A., Broom, D.M., Chara, J., Finch, T., Garnett, E., Gathorne-Hardy, A., Hernandez-Medrano, J., Herrero, M., Hua, F., Latawiec, A., Misselbrook, T., Phalan, B., Simmons, B.L., Takahashi, T., Vause, J., zu Ermgassen, E., Eisner, R., 2018. The environmental costs and benefits of high-yield farming. *Nat Sustain* 1, 477–485. <https://doi.org/10.1038/s41893-018-0138-5>.
- Bart, R.S., Taylor, N.J., 2017. New opportunities and challenges to engineer disease resistance in cassava, a staple food of African small-holder farmers. *PLoS Pathog.* 13, e1006287 <https://doi.org/10.1371/journal.ppat.1006287>.
- Beeching, J.R., Han, Y., Gómez-Vásquez, R., Day, R.C., Cooper, R.M., 1998. Wound and defense responses in cassava as related to post-harvest physiological deterioration. In: Romeo, J.T., et al. (Eds.), *Phytochemical Signals and Plant-Microbe Interactions*. Springer US, Boston, MA, pp. 231–248.
- Bellis, S., Kelly, E.A., Lorts, C.M., Gao, H., DeLeo, V., et al., 2020. Genomics of Sorghum local adaptation to a parasitic plant. *Proc. Natl. Acad. Sci. Unit. States Am.* 117, 4243–4251. <https://doi.org/10.1073/pnas.1908707117>.
- Beyene, G., Chauhan, R.D., Gehan, J., Siritunga, D., Taylor, N., 2020. Cassava shrunken-2 homolog MeAPL3 determines storage root starch and dry matter content and modulates storage root postharvest physiological deterioration. *Plant Mol. Biol.* <https://doi.org/10.1007/s11103-020-00995-z>.
- Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I.Y., Egese, C., Nauluvula, P., Lebot, V., Ndunguru, J., Mkamilo, G., Bart, R.S., Setter, T.L., Gleadow, R.M., Kulakov, P., Ferguson, M.E., Rounsley, S., Rokhsar, D.S., 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. <https://doi.org/10.1038/nbt.3535>.
- Brink, M., van Hintum, T., 2020. Genebank operation in the arena of access and benefit-sharing policies. *Front. Plant Sci.* 10, 1712. <https://doi.org/10.3389/fpls.2019.01712>.
- Bull, S.E., Ndunguru, J., Gruissem, W., Beeching, J.R., Vanderschuren, H., 2011. Cassava: constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Rep.* 30, 779–787. <https://doi.org/10.1007/s00299-010-0986-6>.
- Bull, S.E., Seung, D., Chanez, C., Mehta, D., Kuon, J.-E., Truernit, E., Hochmuth, A., Zurkirchen, I., Zeeman, S.C., Gruissem, W., Vanderschuren, H., 2018. Accelerated ex situ breeding of GBSS- and PTST1- edited cassava for modified starch. *Sci Adv* 4, eaat6086. <https://doi.org/10.1126/sciadv.aat6086>.
- Burgarella, C., Cubry, P., Kane, N.A., Varshney, R.K., Mariac, C., Liu, X., Shi, C., Thudi, M., Couderc, M., Xu, X., Chitikineni, A., Scarcelli, N., Barnaud, A., Rhoné, B., Dupuy, C., François, O., Berthouly-Salazar, C., Vigouroux, Y., 2018. A Western Sahara centre of domestication inferred from pearl millet genomes. *Nat Ecol Evol* 2, 1377–1380. <https://doi.org/10.1038/s41559-018-0643-y>.
- Campbell, B.C., Gilding, E.K., Mace, E.S., Tai, S., Tao, Y., Prentis, P.J., Thomelin, P., Jordan, D.R., Godwin, I.D., 2016. Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol. J* 14, 2240–2253. <https://doi.org/10.1111/pbi.12578>.
- Casa, A.M., Pressoir, G., Brown, P.J., Mitchell, S.E., Rooney, W.L., Tuinstra, M.R., Franks, C.D., Kresovich, S., 2008. Community resources and strategies for

- association mapping in Sorghum. *Crop Sci.* 48, 30–40. <https://doi.org/10.2135/cropsci2007.02.0080>.
- Cassaday, K., Smale, M., Fowler, C., Heisey, P.W., 2001. Benefits from giving and receiving genetic resources: the case of wheat. *Plant Genet. Resour. Newsl.* 127, 1–10.
- CBD/COP/DEC/XIII/16, 2016. Decision Adopted by the Conference of the Parties to the Convention on Biological Diversity. Convention of Biological Diversity. <http://www.cbd.int/doc/decisions/cop-13/cop-13-dec-16-en.pdf>.
- Chauhan, R.D., Beyene, G., Kalyaeva, M., Fauquet, C.M., Taylor, N., 2015. Improvements in Agrobacterium-mediated transformation of cassava (*Manihot esculenta* Crantz) for large-scale production of transgenic plants. *Plant Cell Tiss Org* 121, 591–603. <https://doi.org/10.1007/s11240-015-0729-z>.
- Chen, J., Chopra, R., Hayes, C., Morris, G., Marla, S., Burke, J., Xin, Z., Burrow, G., 2017. Genome-wide association study of developing leaves' heat tolerance during vegetative growth stages in a sorghum association panel. *The Plant Genome* 10, 10. <https://doi.org/10.1007/s11240-016-0909-1>.
- Cook, C.E., Stroe, O., Cochrane, G., Birney, E., Apweiler, R., 2019. The European Bioinformatics Institute in 2020: building a global infrastructure of interconnected data resources for the life sciences. *Nucleic Acids Res.* 48, D17–D23. <https://doi.org/10.1093/nar/gkz1033>.
- Cooper, E.A., Brenton, Z.W., Flinn, B.S., Jenkins, J., Shu, S., Flowers, D., Luo, F., Wang, Y., Xia, P., Barry, K., Daum, C., Lipzen, A., Yoshinaga, Y., Schmutz, J., Saski, C., Vermerris, W., Kresovich, S., 2019. A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism. *BMC Genom.* 20, 420. <https://doi.org/10.1186/s12864-019-5734-x>.
- Cuevas, H.E., Prom, L.K., Cooper, E.A., Knoll, J.E., Ni, X., 2018. Genome-wide association mapping of anthracnose (*Colletotrichum sublineolum*) resistance in the U. S. Sorghum association panel. *The Plant Genome* 11, 170099. <https://doi.org/10.3835/plantgenome2017.11.0099>.
- Dash, S., Campbell, J.D., Cannon, E.K.S., Cleary, A.M., Huang, W., Kalberer, S.R., Karingula, V., Rice, A.G., Singh, J., Umale, P.E., Weeks, N.T., Wilkey, A.P., Farmer, A.D., Cannon, S.B., 2015. Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44, D1181–D1188. <https://doi.org/10.1093/nar/gkv1159>.
- Diatloff, E., Jordan, D.R., Mace, E.S., Godwin, I.D., Tai, S., Filleur, S., Schmidt, S., 2017. The vegetative nitrogen response of sorghum lines containing different alleles for nitrate reductase and glutamate synthase. *Mol. Breed.* 37, 138. <https://doi.org/10.1007/s11032-017-0738-1>.
- Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Teclé, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., Mueller, L.A., 2015. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res.* 43, D1036–D1041. <https://doi.org/10.1093/nar/gku1195>.
- Food and Agriculture Organization of the United States. Tef. website accessed. <http://www.fao.org/traditional-crops/teff/en/>. (Accessed 19 May 2020).
- Fowler, C., Smale, M., Gaiji, S., 2001. Unequal exchange? Recent transfers of agricultural resources and their implications for developing countries. *Dev. Pol. Rev.* 19, 181–204. <https://doi.org/10.1111/1467-7679.00130>.
- Gaffney, J., Bing, J., Byrne, P.F., Cassman, K.G., Ciampitti, I., Delmer, D., Habben, J., Lafitte, H.R., Lidstrom, U.E., Porter, D.O., Sawyer, J.E., Schussler, J., Setter, T., Sharp, R.E., Vyn, T.J., Warner, D., 2019. Science-based intensive agriculture: sustainability, food security, and the role of technology. *Global Food Security* 23, 236–244. <https://doi.org/10.1016/j.gfs.2019.08.003>.
- Gilding, E., Frère, C., Cruickshank, A., et al., 2013. Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. *Nat. Commun.* 4, 1483. <https://doi.org/10.1038/ncomms2450>.
- Girma, D., Assefa, K., Chanyalew, S., Cannarozzi, G., Kuhlemeier, C., Tadele, Z., 2014. The origins and progress of genomics research on Tef (*Eragrostis tef*). *Plant Biotechnol. J* 12, 534–540. <https://doi.org/10.1111/pbi.12199>.
- Gomez, M.A., Lin, Z.D., Moll, T., Chauhan, R.D., Hayden, L., Renninger, K., Beyene, G., Taylor, N.J., Carrington, J.C., Staskawicz, B.J., Bart, R.S., 2019. Simultaneous CRISPR/Cas9-mediated editing of cassava *eIF4E* isoforms *nCBP-1* and *nCBP-2* reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J* 17, 421–434. <https://doi.org/10.1111/pbi.12987>.
- Goodstein, D., Batra, S., Carlson, J., Hayes, R., Phillips, J., Shu, S., et al., 2014. Phytosome Comparative Plant Genomics Portal. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7048E. Retrieved from. <https://escholarship.org/uc/item/22k9d6k9>.
- Grant, D., Nelson, R.T., Cannon, S.B., Shoemaker, R.C., 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38 (Suppl. 1), D843–D846. <https://doi.org/10.1093/nar/gkp798>.
- Halewood, M., Lopez Moreiga, I., Ellis, D., Roa, C., Rouard, M., Sackville Hamilton, R., 2018a. Using genomic sequence information to increase conservation and sustainable use of crop diversity and benefit-sharing. *Biopreservation and Biobanking* 16, 368–376. <https://doi.org/10.1089/bio.2018.0043>.
- Halewood, M., Chiurugwi, T., Sackville Hamilton, R., Durtz, B., Marden, E., Welch, E., Michiels, F., Mosafari, J., Sabran, M., Patron, N., Kersey, P.O., Bastow, R., Sorius, S., Dias, S., McCouch, S., Powell, W., 2018b. Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. *New Phytol.* 217, 1407–1419. <https://doi.org/10.1111/nph.14993>.
- Heinemann, J., Coray, D., Thaler, D., 2018. Fact-finding and scoping study on “digital sequence information” on genetic resources for food and agriculture. <https://www.cbd.int/doc/notifications/2017/nf-2017-115-abs-en.pdf>.
- Howe, K.L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D.M., Cambell, L., Carbajo, M., Chakiachvili, M., Christensen, M., Cummins, C., Cuzick, A., Davis, P., Fexova, S., Gall, A., George, N., Gil, L., Gupta, P., Hammond-Kosack, K.E., Haskell, E., Hunt, S. E., Jaiswal, P., Janacek, S.H., Kersey, P.J., Langridge, N., Maheswari, U., Maurel, T., McDowell, M.D., Moore, B., Muffato, M., Naamati, G., Naithani, S., Olson, A., Papatheodorou, I., Patricio, M., Paulini, M., Pedro, H., Perry, E., Preece, J., Rosello, M., Russel, M., Sitnik, V., Staines, D.M., Stein, J., Tello-Ruiz, M.K., Trevanion, S.J., Urban, M., Wei, S., Ware, D., Williams, G., Yates, A.D., Flicek, P., 2020. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* 48. <https://doi.org/10.1093/nar/gkz890>.
- International Nucleotide Sequence Database Collaboration. <http://www.insdc.org/>.
- Jayne, T.S., Chamberlin, J., Headey, D.D., 2014. Land pressures, the evolution of farming systems, and development strategies in Africa: a synthesis. *Food Pol.* 48, 1–17. <https://doi.org/10.1016/j.foodpol.2014.05.014>.
- Jorasch, P., 2019. The global need for plant breeding innovation. *Transgenic Res.* 28, 81–86. <https://doi.org/10.1007/s11248-019-00138-1>.
- Kane, N.A., Berthouly-Salazar, C., 2020. Population genomics of pearl millet. In: Rajora, O.P. (Ed.), *Population Genomics: Crop Plants*. Springer Nature AG, Switzerland. https://doi.org/10.1007/13836_2020_80.
- Kawahara, Y., de la Bastide, M., Hamilton, J.P., et al., 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 4. <https://doi.org/10.1186/1939-8433-6-4>.
- Ketema, S., 1997. Promoting the conservation and use of underutilized and neglected crops. 12 Tef *Eragrostis tef* (Zucc.) Trotter. *Int. Plant Genet. Resour. Inst.* 12.
- Laird, S., Wynberg, R., Rourke, M., Humphries, F., Muller, M.R., Lawson, C., 2020. Rethink the expansion of access and benefit sharing. *Science* 367, 1200–1202. <https://doi.org/10.1126/science.aba9609>.
- Lawrence, C.J., Schaeffer, M.L., Seigfried, T.E., Campbell, D.A., Harper, L.C., 2007. MaizeGDB's new data types, resources and activities. *Nucleic Acids Res.* 35, D895–D900. <https://doi.org/10.1093/nar/gkl1048>.
- Luo, H., Zhao, W., Wang, Y., Xia, Y., Wu, X., Zhang, L., Tang, B., Zhu, J., Fang, L., Du, Z., Bekele, W.A., Tai, S., Jordan, D.R., Godwin, I.D., Snowdon, R.J., Mace, E.S., Luo, J., Jing, H.-C., 2016. SorGSD: a sorghum genome SNP database. *Biotechnol. Biofuels* 9, 6. <https://doi.org/10.1186/s13068-015-0415-8>.
- Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., Cruickshank, A., Dai, C., Frère, C., Zhang, H., Hunt, C.H., Wang, X., Shatte, T., Wang, M., Su, Z., Li, J., Lin, X., Godwin, I.D., Jordan, D.R., Wang, J., 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* 4, 2320. <https://doi.org/10.1038/ncomms3320>.
- Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., Hassall, M., Hathorn, A., Jordan, D., 2019. The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theor. Appl. Genet.* 132, 751–766. <https://doi.org/10.1007/s00122-018-3212-5>.
- Makita, Y., Shimada, S., Kawashima, M., Kondou-Kuiyama, T., Toyoda, T., Matsui, M., 2015. MOROKOSHI: transcriptome database in Sorghum bicolor. *Plant Cell Physiol.* 56 (1) <https://doi.org/10.1093/pcp/pcu187> e6(1–8).
- Massel, K., Campbell, B.C., Mace, E.S., Tai, S., Tao, Y., Worland, B.G., Jordan, D.R., Botella, J.R., Godwin, I.D., 2016. Whole genome sequencing reveals potential new targets for improving nitrogen uptake and utilization in *Sorghum bicolor*. *Front. Plant Sci.* 7 <https://doi.org/10.3389/fpls.2016.01544>.
- McCormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B.D., McKinley, B., Mattison, A., Morishige, D.T., Grimwood, J., Schmutz, J., Mullet, J.E., 2018. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. <https://doi.org/10.1111/tpj.13781>.
- Mebratu, Y., Raghavaiah, C., Ashagre, H., 2016. Production potential of tef (*Eragrostis tef* (Zucc.) Trotter) genotypes in relation to integrated nutrient management on vertisols of mid high lands of Oromia Region of Ethiopia, East Africa. *Adv Crop Sci Tech* 4, 249. <https://doi.org/10.4172/2329-8863.1000249>.
- Minten, B., Taffesse, A., Brown, P., 2018. The Economics of Tef: Exploring Ethiopia's Biggest Cash Crop. International Food Policy Research Institute, Washington, DC, p. 448.
- Morris, G.P., Ramu, P., Deshpande, S.P., Hash, C.T., Shah, T., Upadhyaya, H.D., Riera-Lizarazu, O., Brown, P.J., Acharya, C.B., Mitchell, S.E., Harriman, J., Glaubitz, J.C., Buckler, E.S., Kresovich, S., 2013. Population genomic and genome-wide association studies of agroclimatic traits in Sorghum. *Proc. Natl. Acad. Sci.* 110, 453–458. <https://doi.org/10.1073/pnas.1215985110>.
- Mundia, C.W., Secchi, S., Akamani, K., Wang, G., 2019. A regional comparison of factors affecting global sorghum production: the case of North America, Asia and Africa's Sahel. *Sustainability* 11, 2135. <https://doi.org/10.3390/su11072135>.
- Narayanan, N., Beyene, G., Chauhan, R.D., Gaitán-Solís, E., Gehan, J., Butts, P., Siritunga, D., Okwuonu, I., Woll, A., Jiménez-Aguilar, D.M., Boy, E., Grusak, M.A., Anderson, P., Taylor, N.J., 2019. Biofortification of field-grown cassava by engineering expression of an iron transporter and ferritin. *Nat. Biotechnol.* 37, 144–151. <https://doi.org/10.1038/s41587-018-0002-1>.
- Ogasawara, O., Kodama, Y., Mashima, J., Kosuge, T., Fujisawa, T., 2020. DDBJ Database updates and computational infrastructure enhancement. *Nucleic Acids Res.* 48, D45–D50. <https://doi.org/10.1093/nar/gkz982>.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, Haining, L., Campbell, M., et al., 2007. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887. <https://doi.org/10.1093/nar/gkl976>.
- Parry, M.L., Rosenzweig, C., Iglesias, A., Livermore, M., Fischer, G., 2004. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Global Environ. Change* 14, 53–67. <https://doi.org/10.1016/j.gloenvcha.2003.10.008>.

- Parry, M., Rosenzweig, C., Livermore, M., 2005. Climate change, global food supply and risk of hunger. *Phil Trans R Soc B* 360, 2125–2138. <https://doi.org/10.1098/rstb.2005.1751>.
- Pasley, H.R., Cairns, J.E., Camberato, J.J., Vyn, T.J., 2019. Nitrogen fertilizer rate increases plant uptake and soil availability of essential nutrients in continuous maize production in Kenya and Zimbabwe. *Nutrient Cycl. Agroecosyst.* 115, 373–389. <https://doi.org/10.1007/s10705-019-10016-1>.
- Pasley, H.R., Camberato, J.J., Cairns, J.E., Zaman-Allah, M., Das, B., Vyn, T.J., 2020. Nitrogen rate impacts on tropical maize nitrogen use efficiency and soil nitrogen depletion in eastern and southern Africa. *Nutrient Cycl. Agroecosyst.* 116, 397–408. <https://doi.org/10.1007/s10705-020-10049-x>.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberler, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otiillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., Rokhsar, D.S., 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. <https://doi.org/10.1038/nature07723>.
- Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: a threat to food security in Africa. *J. Gen. Virol.* 96, 956–968. <https://doi.org/10.1099/vir.0.000014>.
- Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M.D., Harberd, N.P., 1999. ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* 400, 256–261. <https://doi.org/10.1038/22307>.
- Pew Research Center, 2020. Pew center research fact tank. <https://www.pewresearch.org/fact-tank/2019/06/17/worlds-population-is-projected-to-nearly-stop-growing-by-the-end-of-the-century/>.
- Pingali, P.L., 2012. Green Revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci.* 109, 12302–12308. <https://doi.org/10.1073/pnas.0912953109>.
- Poffenbarger, H.J., Barker, D.W., Helmers, M.J., Miguez, F.E., Olk, D.C., Sawyer, J.E., Six, J., Castellano, M.J., 2017. Maximum soil organic carbon storage in Midwest U.S. cropping systems when crops are optimally nitrogen-fertilized. *PLoS One* 12, e0172293. <https://doi.org/10.1371/journal.pone.0172293>.
- Rabbi, I.Y., Hamblin, M.T., Kumar, P.L., Gedil, M.A., Ikpan, A.S., Jannink, J.-L., Kulakow, P.A., 2014. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186, 87–96. <https://doi.org/10.1016/j.virusres.2013.12.028>.
- Raemakers, K., Schreuder, M., Suurs, L., Furrer-Verhorst, H., Vincken, J.-P., de Vetten, N., Jacobsen, E., Visser, R.G.F., 2005. Improved cassava starch by antisense inhibition of granule-bound starch synthase I. *Mol. Breed.* 16, 163–172. <https://doi.org/10.1007/s11032-005-7874-8>.
- Rosegrant, M.W., Koo, J., Cenacchi, N., Ringler, C., Robertson, R.D., Fisher, M., Cox, C. M., Garrett, K., Perez, N.D., Sabbagh, P., 2014. Food Security in a World of Natural Resource Scarcity: the Role of Agricultural Technologies. *International Food Policy Research Institute*, Washington, DC, p. 154.
- Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I., 2019. GenBank. *Nucleic Acids Res.* 48, D84–D86. <https://doi.org/10.1093/nar/gkz956>.
- Sheridan, C., 2020. Coronavirus and the race to distribute reliable diagnostics. *Nat. Biotechnol.* 38, 382–384. <https://doi.org/10.1038/d41587-020-00002-2>.
- Smith, S.M., Yuan, Y., Doust, A.N., Bennetzen, J.L., 2012. Haplotype Analysis and linkage disequilibrium at five loci in *Eragrostis tef*. *G3: Genes|Genomes|Genetics.* 2, 407–419. <https://doi.org/10.1534/g3.111.001511>.
- Spielmeier, W., Ellis, M.H., Chandler, P.M., 2002. Semidwarf (*sd-1*), ‘green revolution’ rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci.* 99, 9043–9048. <https://doi.org/10.1073/pnas.132266399>.
- Stevenson, J.R., Villoria, N., Byerlee, D., Kelley, T., Maredia, M., 2013. Green Revolution research saved an estimated 18 to 27 million hectares from being brought into agricultural production. *Proc. Natl. Acad. Sci.* 110, 8363–8368. <https://doi.org/10.1073/pnas.1208065110>.
- Swaminathan, M.S., 2009. Science and shaping our agricultural future. In: Jha, R. (Ed.), *The First Ten K R Narayanan Orations: Essays by Eminent Persons on the Rapidly Transforming Indian Economy*. The Australian National University E Press, pp. 155–183.
- Tao, Y., Mace, E.S., Tai, S., Cruickshank, A., Campbell, B.C., Zhao, X., Van Oosterom, E. J., Godwin, I.D., Botella, J.R., Jordan, D.R., 2017. Whole-genome analysis of candidate genes associated with seed size and weight in *Sorghum bicolor* reveals signatures of artificial selection and insights into parallel domestication in cereal crops. *Front. Plant Sci.* 8, 1237. <https://doi.org/10.3389/fpls.2017.01237>.
- Taylor, N., Gaitán-Solís, E., Moll, T., Trauterman, B., Jones, T., Pranjali, A., Trembley, C., Abernathy, V., Corbin, D., Fauquet, C.M., 2012. A High-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Trop Plant Biol.* 5, 127–139. <https://doi.org/10.1007/s12042-012-9099-4>.
- Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108, 20260–20264. <https://doi.org/10.1073/pnas.1116437108>.
- Tittonell, P., Giller, K.E., 2013. When yield gaps are poverty traps: the paradigm of ecological intensification in African smallholder agriculture. *Field Crop. Res.* 143, 76–90. <https://doi.org/10.1016/j.fcr.2012.10.007>.
- UN2020, 2020. Timeline to 2020: Convention on Biological Diversity (CBD). <http://un2020.org/timeline/timeline-cbd/>.
- United Nations Sustainable Development Programme, 2016. Sustainable Development Goals. United Nations Sustainable Development Programme. <https://www.undp.org/content/undp/en/home/sustainable-development-goals.html>.
- United States Department of Agriculture/Agriculture Research Service (USDA/ARS). <https://www.ars.usda.gov/>.
- Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., Srivastava, R.K., Chitikineni, A., Fan, G., Bajaj, P., Punnuri, S., Gupta, S.K., Wang, H., Jiang, Y., Couderc, M., Katta, M.A.V.S.K., Paudel, D.R., Mungra, K.D., Chen, W., Harris-Shultz, K.R., Garg, V., Desai, N., Doddamani, D., Kane, N.A., Conner, J.A., Ghatak, A., Chaturvedi, P., Subramaniam, S., Yadav, O.P., Berthouly-Salazar, C., Hamidou, F., Wang, J., Liang, X., Cloutault, J., Upadhyaya, H. D., Cubry, P., Rhoné, B., Gueye, M.C., Sunkar, R., Dupuy, C., Sparvoli, F., Cheng, S., Mahala, R.S., Singh, B., Yadav, R.S., Lyons, E., Datta, S.K., Hash, C.T., Devos, K.M., Buckler, E., Bennetzen, J.L., Paterson, A.H., Ozias-Akins, P., Grand, S., Wang, J., Mohapatra, T., Weckwerth, W., Reif, J.C., Liu, X., Vigouroux, Y., Xu, X., 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* 35, 969–976. <https://doi.org/10.1038/nbt.3943>.
- Wagaba, H., Beyene, G., Aleu, J., Odipio, J., Okao-Okuja, G., Deepika, C.R., et al., 2017. Field level RNAi-mediated resistance to Cassava brown streak disease across multiple cropping cycles and diverse East African agro-ecological locations. *Front. Plant Sci.* 7, 2060. <https://doi.org/10.3389/fpls.2016.02060>.
- Welsch, R., Arango, J., Bär, C., Salazar, B., Al-Babili, S., Beltrán, J., Chavarriaga, P., Ceballos, H., Tohme, J., Beyer, P., 2010. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *The Plant Cell* 22, 3348–3356. <https://doi.org/10.1105/tpc.110.077560>.
- Wilson, M.C., Mutka, A.M., Hummel, A.W., Berry, J., Chauhan, R.D., Vijayaraghavan, A., et al., 2017. Gene expression atlas for the food security crop cassava. *New Phytol.* 213, 1632–1641. <https://doi.org/10.1111/nph.14443>.
- Wolfe, M.D., Rabbi, I.Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D.P.D., Ramu, P., Jannink, J.-L., 2016. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *The Plant Genome* 9, 1–13. <https://doi.org/10.3835/plantgenome2015.11.0118>.
- Zheng, L.-Y., Guo, X.-S., He, B., Sun, L.-J., Peng, Y., Dong, S.-S., Liu, T.-F., Jiang, S., Ramachandran, S., Liu, C.-M., Jing, H.-C., 2011. Genome-wide patterns of genetic variation in sweet and grain Sorghum (*Sorghum bicolor*). *Genome Biol.* 12, R114. <https://doi.org/10.1186/gb-2011-12-11-r114>.
- Ziegler, R.S. Undated. Support for International Agricultural Research: Can the Pipeline Deliver? International Rice Research Institute. www.pitt.edu/~super1/PPT/35551.htm.