



Maximizing value of genetic sequence data requires an enabling environment and urgency

Jim Gaffney^{a,*}, Dejene Girma^b, Ndjido Ardo Kane^c, Victor Llaca^d, Emma Mace^e, Nigel Taylor^f, Redeat Tibebu^g

^a United States Agency for International Development, USA

^b National Agricultural Biotechnology Research Center, Ethiopian Institute for Ag Research, Ethiopia

^c Institut Senegalais de Recherches Agricoles, Centre d' Etude Regional pour l'Amelioration de l'Adaptation a la Secheresse, Senegal

^d Corteva Agriscience, USA

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

^f Donald Danforth Plant Science Center, USA

^g College of Biological Sciences, University of Minnesota, USA

ARTICLE INFO

Keywords:

Nucleotide sequence
Genetic sequence data
COVID-19 vaccine
Regulatory harmonization
Agricultural research and development investment
Science-based decision-making

ABSTRACT

Severe price spikes of the major grain commodities and rapid expansion of cultivated area in the past two decades are symptoms of a severely stressed global food supply. Scientific discovery and improved agricultural productivity are needed and are enabled by unencumbered access to, and use of, genetic sequence data. In the same way the world witnessed rapid development of vaccines for COVID-19, genetic sequence data afford enormous opportunities to improve crop production. In addition to an enabling regulatory environment that allowed for the sharing of genetic sequence data, robust funding fostered the rapid development of coronavirus diagnostics and COVID-19 vaccines. A similar level of commitment, collaboration, and cooperation is needed for agriculture.

1. Introduction

Access to, and use of, genetic sequence data (GSD) is a valuable public good that accelerates discovery, builds scientific capacity, and creates opportunities for increased agricultural productivity. Future access to GSD for all users must be ensured to realize its full potential. In July 2020, a global group of public and private sector authors explained the value of GSD and why open access and use of GSD is critical to the multiple demands of sustainable agriculture (Gaffney et al., 2020). The authors described the potential of GSD utilization for:

- improving crop productivity and sustainability;
- conservation of biodiversity and crop wild relatives;
- building capacity in the global scientific community;
- ensuring a level playing field among scientists regardless of location or organization.

Since the Gaffney et al. (2020) article, the value of open access and utilization of GSD has been clearly demonstrated through its role in the

rapid deployment of coronavirus diagnostic technologies (Kituyi, M., 2020), development of effective COVID-19 vaccines, and publication of the Moderna and Pfizer vaccine sequences in open access databases (Winter, 2021). Investment in the generation, sharing, and use of GSD available through open access has likewise allowed scientific discoveries in crop plants to move quickly, generating progress and value for the four species described in the review article – sorghum (*Sorghum bicolor*), cassava (*Manihot esculenta*), pearl millet (*Pennisetum glaucum*), and tef (*Eragrostis tef*). Vast differences exist between funding for the COVID response and that of agricultural research and development, even as hunger and malnutrition claim more lives than COVID. This letter is an update on how utilization of GSD is helping meet the multiple demands of food security, especially through creative collaborations. It provides a comparison of funding committed to the COVID response and investment in agricultural research and development (R&D) and requests that international bodies and individual countries work to maintain open access to and use of GSD so that it can be accessed and utilized by all scientists in all countries to enable agricultural advances.

Hunger, malnutrition, and related illnesses kill an estimated 2

* Corresponding author.

E-mail address: gaffneyjfg@gmail.com (J. Gaffney).

million children under the age of five annually (Alston et al., 2021). Global food security is “on a razor’s edge of sufficiency” (Cassman and Grassini, 2020). After approximately 100 years of stable, or falling, commodity prices paid to farmers (Sumner, 2009; Zulauf, 2016), three price spikes have occurred since 2000 (Cassman and Grassini, 2020). In 2021, a fourth price spike was underway with severe implications for regions in which food accounts for 40% or more of household spending (World Economic Forum, 2016).

Science-based decision-making is critical to addressing challenges of food security. The scientific community in the public and private sectors must combine forces to optimize a globally collaborative research environment. Urgency is needed from policy-makers and regulatory bodies to ensure an even playing field, enabling all scientists, farmers and actors in the food chain to have access to the latest technologies. While agricultural R&D will never receive funding at the urgency and level of the COVID response, access to and sharing of technology will allow the agricultural research community to “punch above its weight”. Updates on recent developments and use of GSD in the four crops featured in the original publication offer examples for the cooperation, partnering, and capacity building needed for near-term agricultural productivity and food security.

2. Examples of genetic sequence data value in neglected and under-utilized crops

Sorghum. Recently published research in sorghum demonstrates how GSD continues to improve productivity in a crop with a high level of diversity (Fig. 1). Tao et al. (2021) analyzed and assembled 13 genomes of cultivated and wild relatives of sorghum and combined them with three additional, publicly available genomes. These integrated data were used to create a pan-genome of 44,079 gene families, with 222.6 Mb of new sequence identified, enabling whole-genome comparisons across the most diverse genetics ever assembled. Genes responsible for grain shatter, seed dormancy, grain size, and a host of biotic and abiotic stressors were identified. This work demonstrates the value and need for broad and inclusive sequencing of crop species and their wild relatives, and for providing access and utilization of this data. The GSD has been made publicly available (China National GeneBank database, 2021), offering plant breeders, biotechnologists, and eventually farmers, more options for improving sorghum productivity under rapidly changing

growing conditions. Because much of the genome is conserved across species, Tao et al. have also provided guide posts for how to proceed with sequencing and analysis efforts in other crops.

Muleta et al., (2022) provide a further example of the value of genetic diversity and GSD in sorghum. The emergence of an aggressive biotype of sugarcane aphid (SCA) (*Melanaphis sacchari*) as a global pest has required greater use of synthetic insecticides in U.S. sorghum production. In Haiti, sorghum production had become near impossible, with crop losses reaching 30–70%. Over 50 years of globally shared plant material and knowledge among breeders, molecular biologists, and agronomists culminated in the identification of alleles conferring SCA resistance to sorghum. Resistant varieties are now widely available, and sorghum production is nearing pre-SCA resistant levels in Haiti. Importantly, this work also represents a salvaging of valuable genetics and conservation of biodiversity via “evolutionary rescue” (Alexander et al., 2014), only made possible by shared use of germplasm and GSD.

Cassava. Perhaps no crop can better benefit from using GSD than the staple tropical root crop cassava. Breeding cassava to enhance quality and productivity traits is challenging due to the crop’s high degree of heterozygosity and inbreeding depression. Cassava improvement has suffered from underinvestment, meaning that well-targeted investments can bring significant improvements in traits important to cassava farmers, processors, and consumers. Despite recent investments the cassava research and development community remains relatively small and a critical mass of intellectual capacity can only be reached if GSD and other resources are freely accessed by all. Recent and important advances have been made towards these goals. The first publicly available cassava reference genome assembly, released in 2009, has benefited from continuous improvement. Version 7 has been available since 2019, with Version 8 presently under assembly. At each stage, the reference genome has become a more powerful tool for researchers to access and query as they seek to discover genes and gene pathways responsible for traits such as disease resistance, storage root bulking, flowering and post-harvest physiology. This resource has remained available in the public domain, housed within Phytozome at Plant Comparative Genomics portal of the US Department of Energy’s Joint Genome Institute (<https://phytozome.jgi.doe.gov/pz/portal.html>).

GSD is only valuable if it can be used by researchers and breeders to develop enhanced varieties and to uncover new biological information about the crop. Cassava researchers have established open access



Fig. 1. Access to and use of genetic sequence data from highly diverse germplasm creates opportunities for conservation of biodiversity and innovation for biotic and abiotic stress management.

resources to achieve this. These include the International Cassava Genetic Map Consortium which has developed a reference map using more than 22,400 genetic markers (Bredeson et al., 2016) and placed these on the AM560 reference genome. A second example is Cassavabase. Cassavabase brings together 30 years of breeders' field data, allowing this wealth of information to be queried against GSD using publicly available digital tools (<https://www.cassavabase.org/>).

Cassava is highly heterozygous, meaning that significant differences in GSD exist for the same genes inherited from the two parents. Important traits, such as disease resistance, are often coded by only one of the parental copies (allele). In such cases it is necessary to assemble GSD for each parental copy (haplotype) separately to provide the resolution needed to identify a gene or genes responsible for a specific trait. Although technically challenging, recent advances in bioinformatics have allowed this to be achieved, including haplotype resolved genomes for select varieties (Kuon et al. (2019) and Mansfield et al. (2021); <https://www.biorxiv.org/content/10.1101/2021.06.25.450005v1>).

These publicly available resources were critical to ongoing investigation into gene(s) responsible for resistance to cassava mosaic disease and cassava brown streak disease, two virus diseases that threaten food and economic security for cassava farmers in sub-Saharan Africa. They are also critical for enhancing gene editing capacities for the crop. An example is genome editing to understand and control flowering (Fig. 2), where multiple genes controlling flowering in cassava can be modified to synchronize flower production, offering breeders potential to perform sexual crosses of elite varieties at frequencies not previously possible.

Pearl Millet. Pearl millet is an important crop in low-input farming systems in some of the hottest and driest agro-ecologies in the world. Harnessing GSD and other crop improvement tools can accelerate genetic gain and productivity in these drought-prone regions (Kumar et al., 2021). Genetic sequencing and knowledge development from diverse cultivars, landraces, and mapping populations of pearl millet breeding lines from 27 countries (Sehgal et al., 2015) have been brought together within the "Pearl Millet inbred Germplasm Association Panel", and provide valuable material for genome sequencing (Varshney et al., 2017). By accessing these publicly available resources (Cegresources.



Fig. 2. Flowering in cassava plants, shown in this photo, is inconsistent and male and female flowering is often not synchronized, thus preventing breeders from making crosses. Sharing of genetic sequence data has helped the global cassava research community identify control mechanisms which improve synchrony of flowering, giving breeders greater opportunities for yield and quality improvements.

[icrisat.org](https://www.icrisat.org), Kumar et al. (2021) identified gene groupings for plant height, flowering time, panicle length, and grain weight and quality useful for accelerating yield gain. This analysis will be useful for genomic-assisted breeding and crop development, an essential breeding strategy to improve production and enhance the adaptability of pearl millet in low-input farming systems.

Building further on these efforts, pearl millet inbred lines were recently sequenced as part of a collaboration between the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and Corteva Agriscience. The new sequencing data will be made publicly available to assist breeding programs and pearl millet growers everywhere. The references and the annotation currently in progress will be the foundation of collaboration, which also includes the sharing of CRISPR-Cas gene editing technology and expertise.

Tef. Tef is arguably one of the most neglected and under-utilized, semi-domesticated cereal species. It is, however, the most important crop in Ethiopia, with research steadily growing in recent years. The national tef improvement program at the Ethiopian Institute of Agricultural Research (EIAR), in collaboration with international partners, is developing high density linkage maps, an association panel, and genomic fingerprints for 49 recently released varieties. These research activities provide valuable, cross-referenced varietal genomic information for the genomics and phenotyping centers at Holeta and Debre-Zeit, Ethiopia, to generate basic information on the core tef collection. And to establish the genetic and molecular basis underlying adaptive traits in tef, a panel of 382 tef accessions have been re-sequenced at Colorado State University, USA, with the help of international public and private sector partners. Thirty years of climate data have been integrated with the passport data of some of these accessions, and both the GSD and background information will be shared in a public database and publication, respectively, representing another example of how open access to and use of GSD creates value.

The International Tef Research Consortium (ITRC) was established in 2019 to create a collaborative platform for tef researchers. Through this initiative, high-quality genomic sequence data for an improved tef cultivar "Tsedey" (DZ-Cr-37) has been generated through a collaboration between the EIAR and Corteva Agriscience. High-quality reference genome sequence was completed utilizing the latest genomics technology and provides near gapless contiguity sequence. This GSD is available to all ITRC members and will soon be deposited in a public database. With the availability of this full reference GSD, genes controlling plant stature and herbicide tolerance have been identified in tef via comparative genomic approaches.

Applied in conjunction with genome editing tools, this GSD offers new possibilities for rapid productivity and quality gains greater than possible via traditional breeding alone. A recent example is ongoing efforts to produce semi-dwarf tef through gene editing of the "Green Revolution" genes to enhance resistance to lodging, thereby delivering a trait long sought by tef breeders.

Collaboration for generating GSD across tef accessions is also taking place between Bahir Dar University, Ethiopia and the National Institute of Agricultural Botany (NIAB), UK (Matthew Milner, personal communication). This work is focused on understanding natural variation among tef accessions for nutritional traits such as zinc and protein content. Findings from this work will increase understanding of genotypic diversity in tef and have value in increasing tef productivity across different agroecological zones, and allow researchers to identify and improve traits in a manner similar to Tao et al. (2021).

3. Open access to GSD facilitates productive partnerships

In the four crops profiled above, we highlight investment and capacity building in neglected and underutilized species, describe productive collaborations between the public and private sectors and between the global North and South. New ways of conserving biodiversity have been identified through use of GSD and via evolutionary

rescue of plant genetic resources. New opportunities for understanding biotic and abiotic stress resistance (Massel et al., 2021), for improving nutrition and lowering the environmental impact of crop production are being identified (Bate et al., 2021; Eshed and Lippman, 2019). In each example, value created from the investments of time, money, and human capital can only occur through access and utilization of unregulated and non-monetized GSD. For this to continue, a globally harmonized regulatory and enabling environment is needed to ensure that all countries, regions, scientists, farmers, and consumers, benefit from good science and evidence-based decision making.

Investment in R&D is critically important. The value of investment and open access to GSD has never been more evident than during the ongoing COVID-19 pandemic, in which GSD continues to be widely shared. As of June 2020, nearly \$22 Trillion of funding had been committed to the COVID response (Cornish, 2021) with over \$50 billion devoted to vaccine R&D (Knowledgeportalia.org, 2021). Funding differences between the COVID response and agricultural R&D, however, are stark. Global public funding for agricultural R&D declined after the global financial crisis of 2008-09 – “the first sustained drop in over 50 years” (Heisy and Fuglie, 2018), and in 2015 was estimated at \$46.8 billion annually (Alston et al., 2021). The latest update on private sector agricultural R&D investment was estimated at \$15.6 billion annually (Fuglie, 2016). While funding in the trillions of dollars for agricultural research are unlikely to materialize, similar levels of collaboration, cooperation, global capacity building, and urgency observed in response to COVID-19 are needed for agriculture and should be expected. Free access and utilization of crop GSD is a critical component of such an effort. In his book, "Hunger" (2015), Martín Caparrós documents the number of lives lost to hunger-related illness at 25,000 per day globally and asks the question: “How do we manage to live with ourselves knowing that these things are happening?”

The scientific community continues to develop the tools needed to improve agricultural productivity, reduce hunger, and provide solutions for human health. For COVID, limited investment in public health services has prevented more effective control of the disease. For agriculture, delivery of impact is impaired due to long-term neglect of extension services and policy which limits the deployment of good science. A general misunderstanding and distrust of science is slowing delivery of results to those most in need for both agricultural R&D and the COVID response. Policy makers, and society, must now focus on creating an environment in which science thrives and is enabled to leverage and deliver the full potential of scientific discovery.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The views and opinions expressed in this paper are those of the authors and not necessarily the views and opinions of the United States Agency for International Development.

Acknowledgment

Special thanks to Tom Nickson, Nickson Research and Environmental Consulting, for contributing ideas and editing skills to the manuscript.

References

Alexander, H.K., Martin, G., Martin, O.Y., Bonhoeffer, S., 2014. Evolutionary rescue: linking theory for conservation and medicine. *Evol. Appl.* 7, 1161–1179. <https://doi.org/10.1111/eva.12221>.

- Alston, J.M., Pardey, P.G., Rao, X., 2021. Rekindling the slow magic of agricultural R&D. *Issues Sci. Technol.* (May 3, 2021) <https://issues.org/rekindling-magic-agricultural-research-development-alston-pardey-rao>. (Accessed 20 December 2021).
- Bate, N.J., Darlick, C.D., de Maagd, R.A., Williams, R.W., 2021. Opportunities and challenges applying gene editing to specialty crops. In: *Vitro Cellular and Development Biology – Plant*. <https://doi.org/10.1007/s11627-021-10208-x>.
- Bredeson, J.V., Lyons, J.B., Prochnik, S.E., et al., 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34 <https://doi.org/10.1038/nbt.3535>.
- Caparrós, M., 2015. *Hunger*, 46 John. Melvin House Publishing, St. Brooklyn, NY 11201, ISBN 978-1-61219-804-0, p. 512.
- Cassman, K.G., Grassini, P., 2020. A global perspective on sustainable intensification research. *Nat. Sustain.* 3, 262–268. <https://doi.org/10.1038/s41893-020-050708>.
- Cegresourcesicrisat.org, 2021. v1.1 - cegresources.icrisat.Org > Data public > PearlMillet Genome > v1.1.
- China National GeneBank database, 2021. <https://db.cngb.org> under project CNP0001440.
- Cornish, L., 2021. Interactive: who's funding the COVID 19 response and what are the priorities? <https://www.devex.com/news/interactive-who-s-funding-the-covid-19-response-and-what-are-the-priorities-96833>. (Accessed 20 December 2021).
- Eshed, Y., Lippman, Z.B., 2019. Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science*. ISSN: 0036-8075 366 (6466). <https://doi.org/10.1126/science.aax0025>.
- Fuglie, K., 2016. The growing role of the private sector in agricultural research and development world-wide. *Global Food Security* 10, 29–38. <https://doi.org/10.1016/j.gfs.2016.07.005>.
- Gaffney, J., Tibebu, R., Bart, R., Beyene, G., Girma, D., Kane, N.A., Mace, E.S., Mockler, T., Nickson, T.E., Taylor, N., Zastrow-Hayes, G., 2020. Open access to genetic sequence data maximizes value to scientists, farmers, and society. *Glob.Food. Sec.* 26, 100041. <https://doi.org/10.1016/j.gfs.2020.100041>.
- Heisy, P., Fuglie, K., 2018. USDA ERS - Agricultural Research Investment and Policy Reform in High-Income Countries Economic Research Report No. (ERR-249), p. 109.
- Kituyi, M., 2020. Secretary general, united nations conference on trade and development. In: COVID-19: Collaboration Is the Engine of Global Science – Especially for Developing Countries | World Economic Forum (weforum.Org). (Accessed 5 September 2021).
- Knowledgeportalia.org, 2021. COVID-19 vaccine R and D investments. <https://www.knowledgeportalia.org/covid19-r-d-funding>. (Accessed 20 December 2021).
- Kumar, S., Hash, C.T., Singh, G., Nepolean, T., Srivastava, R.K., 2021. Mapping QTLs for important agronomic traits in an Indian-derived immortal population of pearl millet. *Biotech. Notes.* 2, 26–32. <https://doi.org/10.1016/j.biotno.2021.06.001>.
- Kuon, J.-E., Qi, W., Schlapfer, P., et al., 2019. Haplotype-resolved genomes of geminivirus-resistant and geminivirus-susceptible African cassava cultivars. *BMC Biol.* 17, 75. <https://doi.org/10.1186/s12915-019-0697-6>.
- Mansfield, B.N., Boyher, A., Berry, J.C., et al., 2021. Large structural variations in the haplotype-resolved African cassava genome. <https://www.biorxiv.org/content/10.1101/2021.06.25.450005>.
- Massel, K., Lam, Y., Wong, A.C.S., et al., 2021. Hotter, drier, CRISPR: the latest edit on climate change. *Theor. Appl. Genet.* 134, 1691–1709. <https://doi.org/10.1007/s00122-020-03764-0>.
- Milner, Matthew, Specialist, Senior, National Institute of Agricultural Botany, August 2021. Personal communication. www.niab.com/about-us/people/staff.
- Muleta, K.T., Felderhoff, T., Winans, N., et al., 2022. The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange. *Sci. Adv.* <https://doi.org/10.1126/sciadv.abj4633>. Epub 2022 Feb 9.
- Phytozome, 2021. <https://phytozome.jgi.doe.gov/pz/portal.html>.
- Sehgal, D., Skot, L., Singh, R., Srivastava, R.K., Das, S.P., Taunk, J., Yadav, R.S., 2015. Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One* 10 (5). <https://doi.org/10.1371/journal.pone.0122165>.
- Sumner, D.A., 2009. Recent commodity price movements in historical perspective. *Am. J. Agric. Econ.* 91, 1250–1256. <https://doi.org/10.1111/j.1467-8276.2009.01292.x>.
- Tao, Y., Luo, H., Xu, J., et al., 2021. Extensive variation within the pan-genome of cultivated and wild sorghum. *Native Plants* 7, 766–773. <https://doi.org/10.1038/s41477-021-00925-x>.
- Varshney, R.K., Shi, C., Thusid, M., et al., 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>.
- Winter, L., 2021. Scientists Reverse Engineer mRNA Sequence of Moderna Vaccine | the Scientist Magazine® (the-scientist.Com). (Accessed 19 September 2021).
- World Economic Forum, 2016. Which countries spend the most on food? <https://www.weforum.org/agenda/2016/12/this-map-show-how-much-each-country-spends-on-food> Website. (Accessed 15 October 2021).
- Zulauf, C., 2016. Real deflated prices of corn, soybeans, wheat, and upland cotton: reemergence of the historical downward trend?. In: *Farmdoc Daily*, vol. 6. Department of Agricultural and consumer Economics, University of Illinois at Urbana-Champaign, p. 189. (Accessed 19 September 2016).