



COLLEGE OF  
AGRICULTURE AND  
LIFE SCIENCES  
VIRGINIA TECH.

2020

GAP  
REPORT®



GLOBAL AGRICULTURAL  
PRODUCTIVITY REPORT®



OCTOBER 2020

# OPEN ACCESS TO GENETIC SEQUENCE DATA MAXIMIZES VALUE TO SCIENTISTS, FARMERS, AND SOCIETY

## Authors:

Jim Gaffney, Corteva Agriscience  
Redeat Tibebe, University of Minnesota

## ABSTRACT

Over the last 20 years, genome sequencing has become increasingly faster and affordable, with genome sequences of thousands of species – plants, animals, insects, bacteria, viruses – now curated in multiple, open access databases available to all scientists worldwide with internet access. The open access database system, as it stands today, offers fair and equitable sharing of GSD, promotes sustainable use of biodiversity, while also contributing to conservation of biological diversity. Yet this free and unencumbered access to GSD may now be threatened by policy-makers looking to regulate and monetize GSD in the hopes of creating more “fair and equitable sharing of benefits from the use of genetic resources”. Unencumbered, open access to GSD contributes directly to improved breeding programs, more vibrant seed systems, greater agricultural productivity, and greater resilience of crops and farmers in the face of changing environmental conditions. We ask that policy makers and politicians move with great care and caution when considering regulation and monetization of GSD, beyond what currently exists for international sharing and partnering of GSD. Public research and emerging economies have the most to lose if GSD is further controlled, especially if the global scientific community is not consulted.

## INTRODUCTION

Combinations of the mononucleotides cytosine (C), guanine (G), adenine (A), and thymine (T) are the building blocks of life, lining up in unique orders, or sequences, thereby forming the genome, or genetic sequence data (GSD), specific to each organism. Over the last 20 years, genome sequencing has become increasingly faster and affordable, with genome sequences of thousands of species – plants, animals, insects, bacteria, viruses – now curated in multiple, open access databases available to all scientists worldwide with internet access. For example,

just three of these plant-focused databases, Legume Information System, SoyBase, and Maize GDB, serve 2,300 to over 5,200 users each month, accessed from nearly every country. (Figure 1). In a paper recently published in the journal *Global Food Security*, a group of international authors explain the value of GSD, and why the current system of open access to GSD is critical to discovery, innovation, and more sustainable global agriculture and food systems.<sup>1</sup>



**Figure 1:** Users by country and by month of select crop genomics databases. Figure compliments of Carson Andorf and Lisa Hartman, USDA/ARS.

Genomic sequence data, and broad, open, and unencumbered access to it, allows scientists all over the world the freedom to leverage the vast genetic diversity of crop varieties and their wild and weedy relatives, to enable a more thorough, precise, and rapid breeding process. The combination of GSD and new breeding technologies under development will continue to create opportunities for achieving greater food security, a balanced diet and more climate-resilient crops. The open access database system, as it stands today, offers fair and equitable sharing of GSD, promotes sustainable use of biodiversity, while also contributing to conservation of biological diversity – the three objectives of the Conventions on Biodiversity (CBD).<sup>2</sup>

Yet this free and unencumbered access to GSD may now be threatened by policy-makers looking to regulate and monetize GSD in the hopes of creating more “fair and equitable sharing of benefits from the use of genetic resources.”<sup>3,4,5</sup> 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. In so doing, they may further threaten achieving the objectives of the CBD and limit the value of GSD.

Funded by governments and philanthropic donor agencies, the open access databases have grown in size and sophistication. Curators, most often in public research and academia, have developed training tools and easy-to-use search applications to make the sequence data widely available. Scientists from both the public and private sectors have been the contributors of sequence data. The value of GSD, resides not in the raw nucleotide sequences provided in the databases, but in the evaluation, interpretation, and application of these data to generate information which is then applied to develop improved crops, vibrant seed systems, and greater food security. Developing this value requires creativity, resources, ingenuity, and importantly, collaboration amongst scientists. Value is created when more productive and resilient varieties of crops are integrated into a breeding pipeline and subsequently delivered to farmers. More rapid delivery of improved crops and cropping systems is one of our best opportunities to mitigate and adapt to climate change and help farmers become more resilient in the face of ever-changing cropping environments.

The value created through a paid access and benefit sharing scheme for GSD, however well-meaning, would pale in comparison to the monetary and non-monetary value created when GSD is unencumbered by regulation and broadly available for innovation and discovery.

## EXAMPLES OF VALUE

Numerous examples exist of how GSD is being utilized in crops important to Africa and Asia and serve to reinforce our points made above. Cassava, with 60 percent of global production occurring in Sub-Saharan Africa (SSA), is threatened by viral and bacterial disease and has numerous “anti-quality” traits, storage and nutritional problems.<sup>6,7,8</sup> The availability of sequence data for this crop combined with other plant breeding innovations has resulted in improved starch quality, higher levels of vitamin A in the roots, iron and zinc biofortification, and brown streak disease resistance.<sup>9,10,11,12</sup> Work continues on other diseases and to improve storage of cassava roots.

Sorghum is likewise an important food security crop, feeding around half a billion people in SSA and Asia alone. Sorghum has amongst the largest libraries of GSD available of any African indigenous crop.<sup>13,14,15,16,17,18,19</sup> These data have been used to evaluate numerous agronomic and quality traits, including grain size, heat tolerance, starch content, nitrogen use efficiency, and anthracnose resistance.<sup>20,21,22,23,24,25,26</sup>

Pearl millet, like sorghum, is indigenous to Africa and a resilient crop in the face of high heat and drought conditions. The reference genome of pearl millet was published with 38,579 genes and made available to plant scientists anywhere.<sup>27</sup> The information extracted from this database has been used to establish marker assisted breeding. Marker-trait associations have been linked to root traits, plant height, yield, grain quality, tolerance to water deficit and heat, and resistance to an ever-changing disease and insect pest spectrum.<sup>28</sup>

A final example is provided by how GSD could transform breeding efforts in Ethiopia’s national crop, *Eragrostis tef*. When considering area under cultivation, the number of farmers growing tef, and its importance to the diet, tef is far and away the most important crop in Ethiopia.<sup>29</sup> Yet it lags behind most other cereals in Ethiopia in terms of productivity and yield and has numerous constraining agronomic characteristics.<sup>30,31,32</sup> Until very recently, a draft or completely sequenced genome (GSD) was unavailable for tef, and tef improvement programs had to rely on direct phenotypic selection and conventional breeding, which is slow and cumbersome. The valuable genetic diversity of tef has therefore remained relatively hidden and unknown, and while helpful, the sequenced variety “*Dabbi*” is one of hundreds of farmer-used varieties sown in Ethiopia.<sup>33</sup> The availability of openly accessible GSD on greater numbers of 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. Ultimately, millions of farmers would achieve greater productivity and sustainability, over 100 million Ethiopians would have a more reliable and lower cost supply of tef, and pressure on land resources would be reduced.

## VISION FOR THE FUTURE

Unencumbered, open access to GSD contributes directly to improved breeding programs, more vibrant seed systems, greater agricultural productivity, and greater resilience of crops and farmers in the face of changing environmental conditions.

We ask that policy makers and politicians move with great care and caution when considering regulation and monetization of GSD, beyond what currently exists for international sharing and partnering of GSD. Public research and emerging economies have the most to lose if GSD is further controlled, especially if the global scientific community is not consulted. Administrative, financial, and legal barriers will quickly weigh down research organizations making determinations as to GSD value, point of origin, accessibility and with whom negotiations for access must be conducted. Many of these tasks would take time and be difficult, if not impossible. It is imperative to find ways to help scientists innovate while lowering legal and financial barriers to innovation, which will contribute to the most productive path forward to equitable use of genetic resources and GSD.

While we have focused this discussion on plants and agriculture, human health and public safety is also directly impacted by open access to GSD. In the early days of the current coronavirus pandemic, scientists quickly and easily submitted and accessed viral GSD to multiple open access databases, with freedom to operate and limited legal implications.<sup>34</sup> This access allowed scientists to rapidly gain a better understanding of the virus, develop diagnostic tests, and envision a path on how best to develop vaccines and other preventive measures.

Maintaining open access and the use of GSD will serve to enable research collaborations, maximize returns on investments and ensure creation of value to scientists, farmers, and society. This also allows for greater scientific collaboration, enabling and strengthening a global network of scientists. Crops like tef, sorghum, cassava, and pearl millet will become more productive in regions where food security is most at risk. The opportunity exists to create a new era of discovery and investment with GSD.

## ENDNOTES

- 1 Gaffney, J., et al. (2020). Open access to genetic sequence data maximizes value to scientists, farmers, and society. *Global Food Security*, 20. <https://doi.org/10.1016/j.gfs.2020.100411>
- 2 Brink, M., van Hintum, T., (2020). Genebank operation in the arena of access and benefit-sharing policies. *Front Plant Sci.* 10, <https://doi.org/10.3389/fpls.2019.01712>
- 3 Halewood, M., Noriega, I. L., Ellis, D., Roa, C., Rouard, M., & Hamilton, R. S. (2018). Using Genomic Sequence Information to Increase Conservation and Sustainable Use of Crop Diversity and Benefit-Sharing. *Biopreservation and Biobanking*, 16(5). <https://doi.org/10.1089/bio.2018.0043>
- 4 Halewood, M., Chiurugwi, T., Sackville Hamilton, R., Kurtz, B., Marden, E., Welch, E., Michiels, F., Mozafari, J., Sabran, M., Patron, N., Kersey, P., Bastow, R., Dorius, S., Dias, S., McCouch, S. & Powell, W. (2018). Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. *New Phytol*, 217(4), 1407-1419. <https://doi.org/10.1111/nph.14993>
- 5 Aubry, S. (2019). The Future of Digital Sequence Information for Plant Genetic Resources for Food and Agriculture. *Frontiers in Plant Science*, 10(1046). <https://doi.org/10.3389/fpls.2019.01046>
- 6 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(5). <https://doi.org/10.1371/journal.ppat.1006287>
- 7 Bull, S. E., Ndunguru, J., Gruissem, W., & Beeching, J. R. (2011). Cassava: Constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Reports*, 30(5), 779-87. <https://doi.org/10.1007/s00299-010-0986-6>
- 8 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* (pp. 231-248). Springer.
- 9 Wang, W., Hostettler, C. E., Damberger, F. F., Kossmann, J., Lloyd, J. R., & Zeeman, S. C. (2018). Modification of Cassava Root Starch Phosphorylation Enhances Starch Functional Properties. *Frontiers in Plant Science*, 9(1562). <https://doi.org/10.3389/fpls.2018.01562>
- 10 Welsch, R., Arango, J., Bär, C., Salazar, B., Al-Babili, S., Beltrán, J., . . . Beyer, P. (2010). Provitamin A Accumulation in Cassava Roots Driven by a Single Nucleotide Polymorphism in a Phytoene Synthase Gene. *The Plant Cell*, 22(10), 3348-3356. <https://doi.org/10.1105/tpc.110.077560>
- 11 Narayanan, N., Beyene, G., Chauhan, R. D., Gaitán-Solís, E., Gehan, J., Butts, P., Sirtunga, 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. *Nature biotechnology*, 37(2), 144-151. <https://doi.org/10.1038/s41587-018-0002-1>
- 12 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 biotechnology journal*, 17(2), 421-434. <https://doi.org/10.1111/pbi.12987>
- 13 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., Sasaki, 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 Genomics*, 20(1), 420. <https://doi.org/10.1186/s12864-019-5734-x>
- 14 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. *The Plant Journal*, 93(2), 338-354. <https://doi.org/10.1111/tpj.13781>
- 15 Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., . . . Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, 457(7229), 551-556. <https://doi.org/10.1038/nature07723>
- 16 Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., . . . Wang, J. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature communications*, 4, 2320. <https://doi.org/10.1038/ncomms3320>
- 17 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 Biology*, 12(11), R114. <https://doi.org/10.1186/gb-2011-12-11-r114>
- 18 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). Erratum to: SorGSD: a sorghum genome SNP database. *Biotechnology for biofuels*, 9(30). <https://doi.org/10.1186/s13068-016-0450-0>

- <sup>19</sup> 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. *TAG. Theoretical and applied genetics*, 132(3), 751-766. <https://doi.org/10.1007/s00122-018-3212-5>
- <sup>20</sup> 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. *Frontiers in plant science*, 8(1237). <https://doi.org/10.3389/fpls.2017.01237>
- <sup>21</sup> 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. *Plant Genome*, 10(2). <https://doi.org/10.3835/plantgenome2016.09.0091>
- <sup>22</sup> Gilding, E. K., Frère, C. H., Cruickshank, A., Rada, A. K., Prentis, P. J., Mudge, A. M., Mace, E. S., Jordan, D. R., & Godwin, I. D. (2013). Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. *Nature Communications*, 4(1483). <https://doi.org/10.1038/ncomms2450>
- <sup>23</sup> Campbell, B. C., Gilding, E. K., Mace, E. S., Tai, S., Tao, Y., Prentis, P. J., . . . Godwin, I. D. (2016). Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnology Journal*, 14(12), 2240-2253. <https://doi.org/10.1111/pbi.12578>
- <sup>24</sup> Diatloff, E., Mace, E.S., Jordan, D.R. et al. (2017). The vegetative nitrogen response of sorghum lines containing different alleles for nitrate reductase and glutamate synthase. *Mol Breeding* 37(138). <https://doi.org/10.1007/s11032-017-0738-1>
- <sup>25</sup> Massel, K., Campbell, B. C., Mace, E. S., Tai, S., Tao, Y., Worland, B. G., . . . Godwin, I. D. (2016). Whole Genome Sequencing Reveals Potential New Targets for Improving Nitrogen Uptake and Utilization in Sorghum bicolor. *Frontiers in Plant Science*, 7(1544). <https://doi.org/10.3389/fpls.2016.01544>
- <sup>26</sup> Cuevas, H.E., Prom, L.K., Cooper, E.A., Knoll, J.E. and Ni, X. (2018). Genome Wide Association Mapping of Anthracnose (*Colletotrichum sublineolum*) Resistance in the U.S. Sorghum Association Panel. *The Plant Genome*, 11(2), 1-13. <https://doi.org/10.3835/plantgenome2017.11.0099>
- <sup>27</sup> Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., Srivastava, R. K., Chitkineni, A., Fan, G., Bajaj, P., Punnuri, S., Gupta, S. K., Wang, H., Jiang, Y., Couderc, M., Katta, M., . . . Xu, X. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology*, 35(10), 969-976. <https://doi.org/10.1038/nbt.3943>
- <sup>28</sup> Kane N.A., Berthouly-Salazar C. (2020). Population Genomics of Pearl Millet. In Rajora, O. (Ed.) *Population Genomics*. Springer. [https://doi.org/10.1007/13836\\_2020\\_80](https://doi.org/10.1007/13836_2020_80)
- <sup>29</sup> Minten, B., Taffesse, A. S., & Brown, P. (Eds.) (2018). *The Economics of Tef: Exploring Ethiopia's Biggest Cash Crop*. International Food Policy Research Institute (IFPRI). <https://doi.org/10.2499/9780896292833>
- <sup>30</sup> Ketema, S. (1997). *Tef. Eragrostis tef (Zucc.) Trotter Promoting the conservation and use of underutilized and neglected crops*. Biodiversity International Publ.
- <sup>31</sup> 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 Biotechnology Journal*, 12(5), 534-540. <https://doi.org/10.1111/pbi.12199>
- <sup>32</sup> 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. *Advances in Crop Science and Technology*, 4(6). <https://doi.org/10.4172/2329-8863.1000249>
- <sup>33</sup> Van Buren, R., Wai, C.M., Wang, X., Pardo, J., Yocca, A.E., Wang, H., Chaluvadi, S.R., Han, G., Bryants, D., Edger, P.P.... Michael, T.P. (2020). Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-14724-z>
- <sup>34</sup> Sheridan C. (2020). Coronavirus and the race to distribute reliable diagnostics. *Nature Biotechnology*, 38(4), 382-384. <https://doi.org/10.1038/d41587-020-00002-2>