



Integrative and inclusive genomics to promote the use of underutilised crops

Oluwaseyi Shorinola, Rose Marks, Peter Emmrich, Chris Jones, Damaris Odeny & Mark A. Chapman



Underutilised crops are important for diversifying food systems. The genomics of these crops should not be done in isolation but should align with their breeding and capacity building strategies while leveraging advances made in major crops.

Before the advent of agriculture, humans regularly consumed nearly 7000 plant species, yet only ~250 species have been fully domesticated, and only ~30 of these provide ~95% of global calorie intake¹. Investment in this narrow set of crops has reduced nutritional and ecological resilience. There are a variety of lesser known domesticated or semi-domesticated crops that have significant potential for making our food systems more diverse and resilient, and whilst being locally important, barriers to their wider adoption exist. We use the term “underutilised” to describe these crops, rather than other terms which imply they are abandoned (“neglected” or “orphan”), unimportant (“minor”), or simply from a certain place (“indigenous”). Each of these terms would encompass a different group of crops, and by using the term “underutilised”, we wish to focus on the unfulfilled potential of these crops to contribute to global nutrition.

Although initially lagging, the genomics of underutilised crops has gathered pace due to technological advances in genome sequencing, assembly, and annotation (Fig. 1). The availability of genomic resources is proving useful in elucidating underutilised crop histories - their polyploidization², domestication³ and artificial selection histories⁴, among others. Genomics has also shed light on the genetic architecture of important agronomic and end-use traits like tuber quality in greater yam⁵, and the unique climate resilience of crops like tepary bean that is adapted to desert habitats⁶. Recent reviews provide detailed description of progress made in underutilised crop genomics^{7,8}. There have also been several reviews on the broad use, policy and ethical considerations around underutilised crops in general^{9–11}.

In this article, we focus our discussion on how to maximise positive impact in this new era of underutilised crop genomics. As genome sequencing platforms become more and more accessible, there is a tendency to “just sequence everything”. However, we argue that the development of underutilised crop genomic resources should not be done in isolation. Instead, such efforts should be aligned with the overall breeding strategies of the target crops as well as the needs and perceptions of the communities that depend on these crops, while leveraging existing advances in major crops (Fig. 2). We describe these points in detail below and provide examples where applicable.

Underutilised crops genomics efforts should align with their seed system and breeding strategies

While improved seeds of most major crops are easily accessible through formal seed systems (involving structured public and/or private sector breeding and distribution channels), underutilised crops are predominantly distributed in informal seed systems between smallholder farmers, complicating rapid adoption of improved materials. Although genomics may not be immediately relevant across the entire seed systems, initial engagement of stakeholders particularly at the germplasm assessment and breeding stages can significantly increase the attractiveness, profitability, and uptake of underutilised crops. This is especially because underutilised crops’ farmers and end-user perceptions are sometimes based on tacit traditional knowledge that are not always captured by metric traits used in formal breeding. Incorporating this tacit knowledge in genomic screening as demonstrated by Woldeyohannes et al.¹² and Gesesse et al.¹³ can facilitate a deeper understanding of the adaptive potential and end-use value of existing genetic diversity in farmers’ fields and conserved diversity in gene banks. This understanding can be used to integrate formal and farmer-participatory models to identify, prioritise or develop adapted and farmer-preferred germplasm. This model does not necessarily require formal plant breeding, as valuable germplasm (landraces) identified from genomic-assisted participatory screening can be made widely and readily available to farmers through integrated seed systems as suggested in the “Seeds for Needs” model¹⁴.

To maximise impact and relevance of underutilised crop genomics, the choice of germplasm used in generating genomic resources should align with the demand from the breeding and end-use communities. This is particularly important in the selection of cultivars for reference genome assemblies, which should be representative of important founder parents, have rich genetic resources (markers, genetic maps) or harbour demand-led traits that are locally important. For instance, in sequencing the genome of greater yam, Bredeson et al.⁵ sequenced a breeding line that is frequently used as a parent in the greater yam breeding program and that is moderately resistant to anthracnose, an important disease in this crop.

Representation and capacity building will promote the use of underutilised crop genomes

As is the trend in plant genomics more broadly¹⁵, genomic research on underutilised crops is dominated by affluent nations and institutions. Some exceptions include the whole genome sequencing of International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) mandate crops – coordinated largely from India¹⁶, the greater yam genome from Nigeria⁵, the Lima bean genome from Colombia¹⁷, and the recent lablab genome from East Africa³. Given the overwhelming significance of underutilised crops to many indigenous communities¹⁸,

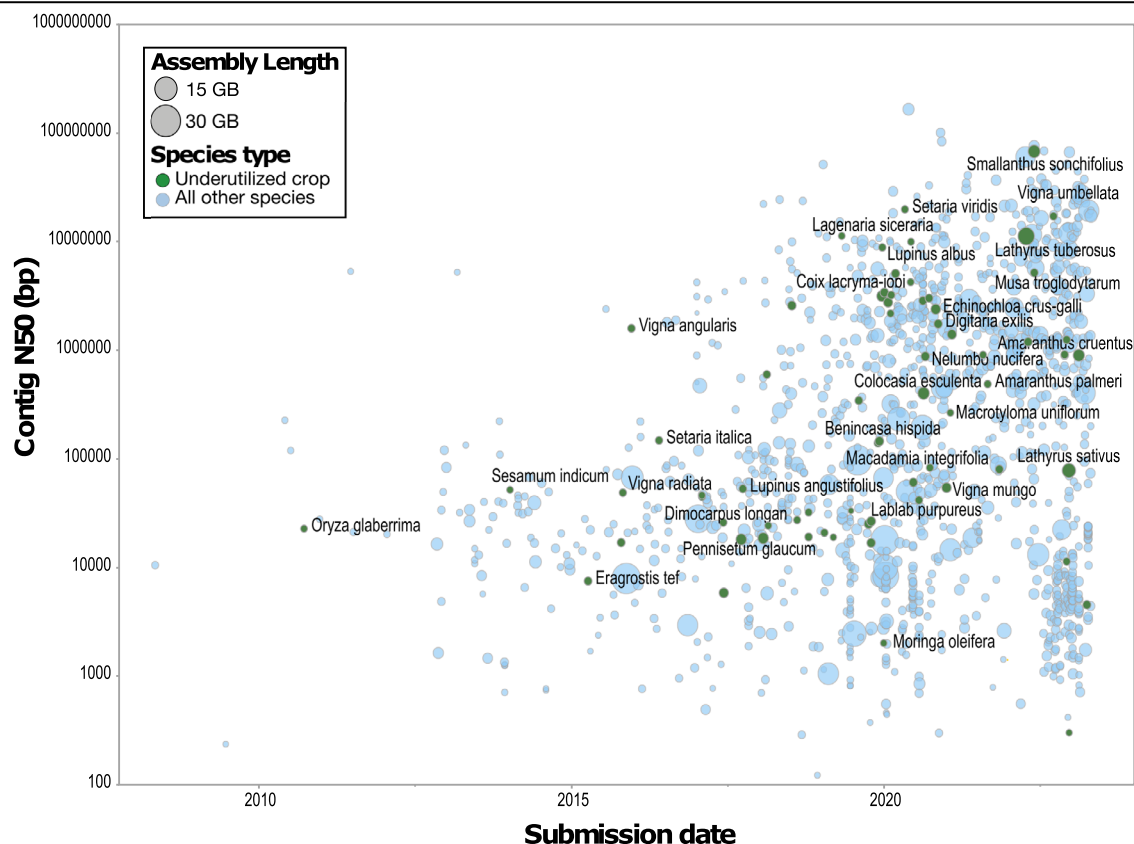


Fig. 1 | Summary of underutilised crops genomics studies. Assembly metrics by submission date for 1324 land plant species with publicly available genome assemblies from 2008 to 2023. An example (non-exhaustive) set of underutilised crops are highlighted in green while other plant species are highlighted in blue.

Genome size is represented by bubble size. Source data are provided as a Source Data file, which consists of Supplementary Table 1 from Marks et al.¹⁵ updated with data for assemblies published till July 2023.

genomics projects on underutilised crops should be used to develop research capacities in these regions. The ability to utilise a genome is just as important as having access to the genome. While it is not essential for everyone to become proficient with genomic data, systems with overlapping expertise and effective communication chains across interest groups are needed to fully realise the potential of genomic data and tools in underutilised crops. While short-term (days) capacity building efforts are useful, we argue that longer-term, practical and in-country trainings are more impactful, especially when customised to fill specific gaps within the local context. These efforts should not be afterthoughts but should be considered at the inception of genomics projects on underutilised crops. Researchers from communities with a good understanding of farmer and consumer preferences should co-design and co-lead genomic research on the target crops.

A growing number of international collaborations and consortia (e.g Africa BioGenome Project¹⁹, African Orphan Crops Consortium²⁰, and CABANA/Capacity Building for Bioinformatics in Latin America) are now promoting inclusion and capacity building in genomics research on under-represented crops. This inclusivity and diversity help to promote curiosity-driven research and the integration of local knowledge enriches research questions leading to more relevant, trusted and translatable research outputs, better adoption of

recommendations, and increased ownership and stewardship of the resulting resources.

Advances in major crops should be used to fast-track underutilised crops genomics

There is a need to accelerate the rate of genetic improvement in underutilised crops. We have a broad knowledge of the genes and molecular pathways underlying many agriculturally relevant traits in major crops, and a suite of tools (transgenics, gene editing, genomic selection, speed breeding) have been developed to improve such crops. These tools should be leveraged to transfer knowledge and fast-track the de novo domestication and improvement of underutilised crops.

Genomic selection and speed breeding have emerged as some of the most prominent breeding innovations that significantly reduce breeding time and cost. These tools could be especially valuable for helping small breeding programmes (as is typical of underutilised crops) accelerate genetic gains using limited resources^{21,22}. Importantly, genomic selection and speed breeding protocols that have been established for many major crops could be easily transferred to closely related underutilised crops to achieve immediate impact. Furthermore, genomic advances in major crops have highlighted the importance of coalescing multiple de novo assemblies into a crop

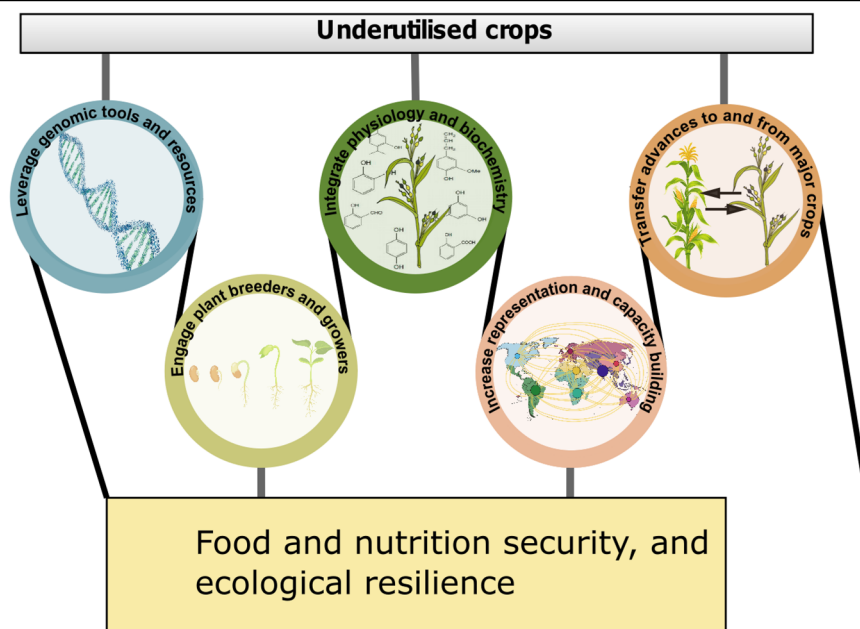


Fig. 2 | Integrated and inclusive genomics for underutilised crops. To achieve a diversified and resilient food system through the utilisation of underutilised crops, the genomics of these crops must align with their breeding programmes and end-users' needs, promote participation and capacity building of the communities that

depend on them, and leverage information from other fields and crops. The map present in this Figure was adapted from Marks et al.¹⁵. The other images in the Figure were purchased from Shutterstock.

pangenome to present a wider picture of the gene content of a species. Core genome regions are found in all accessions, but may make up only half of the pangenome⁷, while dispensable regions could contain adaptive variation, for example in foxtail millet²³. If applied to underutilised crops, these approaches would give a much more complete picture of genomic variation.

Gene editing provides an unprecedented opportunity to accelerate the improvement and domestication of underutilised crops. Knowledge of gene function from related major crops could inform the selection of candidate gene targets in underutilised crops, while taking into account that gene function is not always conserved across species. The simplicity, versatility and cost-effectiveness of the CRISPR/Cas systems has democratised gene editing, and is already showing potential for domesticating underutilised crops, as was demonstrated for groundcherry²⁴ and African rice²⁵ by targeting genes that were firstly studied in major crops.

Despite this potential, the use of gene editing particularly in underutilised crops should be carefully considered in light of the traditional and cultural value of these crops to indigenous communities. The lack of efficient and genotype-independent transformation and regeneration protocols stalls the broad adoption of gene editing in many underutilised crops²⁶. Also, licensing issues still present a major hurdle to the use of the CRISPR/Cas systems in varietal development²⁷. As part of the capacity building effort suggested above, we advocate for dedicated initiatives and funding to promote the transfer of technical knowledge (in transformation and regeneration) and establishment of well-resourced, and appropriately licensed tissue culture labs for the use of transformation-enhancing systems (including morphogenic genes like *WUSCHEL*, *BABY BOOM* and *GRF-GIF*) and gene editing in communities working on underutilised crops.

Linking genomics with physiology and biochemistry to unlock the resilience and nutraceutical potential of underutilised crops


Many underutilised crops are adapted to degraded, marginal or otherwise poor soils, and tolerate drought, cold or heat stresses. They are often the only crops standing, when most staple crops have failed, giving them the reputation of famine or insurance crops. Many indigenous communities use underutilised crops to promote healthy living or treat ailments due to their rich nutrient and nutraceutical composition²⁸. Underutilised crops are therefore a reservoir of novel adaptive genetic, physiological and biochemical variation.

There is a need to integrate underutilised crop genomics with plant physiology and biochemistry to build a more cohesive picture of their adaptation and metabolic phenotypes at the cell, organ, plant and species levels. As genomics is becoming more accessible, detailed physiological and metabolic studies on underutilised crops are needed to achieve this integration. Studies like those of Edwards et al.²⁹ and Moghaddam et al.⁶ demonstrate the value of combining genomics with physiology and biochemistry to uncover adaptation and metabolic phenotypes in underutilised crops. Several adaptive genes or alleles from underutilised crops have already been shown to provide climate resilience in model plants or staple crops²⁸.

Conclusion. Underutilised crop genomics has come of age but should not stand alone. Translating genomic advances into viable pathways towards food and nutritional security will require integrating different systems (basic research, breeding, and seed systems), participants (e.g., academics, breeders and indigenous communities), tools (e.g., gene editing, genomic selection, speed breeding) and disciplines (e.g., genomics, physiology, biochemistry). Research into both underutilised and major crops would benefit each other: underutilised

crops could inform the improvement of resilience and nutritional traits in major crops while domestication, yield and agronomy improvement in underutilised crops would benefit from knowledge acquired in major crops. It is also important to note that sharing of germplasm and digital sequence information (DSI) is a part of this process, and a review of international policies on germplasm access and benefit sharing is needed to promote research into underutilised crops^{11,30}.

Oluwaseyi Shorinola ^{1,2} , **Rose Marks**^{3,4}, **Peter Emmrich**⁵, **Chris Jones** ¹, **Damaris Odeny** ⁶ & **Mark A. Chapman** ⁷

¹International Livestock Research Institute, Naivasha Road, Nairobi, Kenya. ²School of Bioscience, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ³Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA. ⁴Plant Resilience Institute, Michigan State University, East Lansing, MI 48824, USA. ⁵Norwich Institute for Sustainable Development, School of Global Development, University of East Anglia, England NR4 7TJ, UK. ⁶Center of Excellence in Genomics and Systems Biology, ICRISAT, Patancheru 502324 Telangana, India. ⁷Biological Sciences, University of Southampton, Life Sciences Building 85, Highfield Campus, Southampton SO17 1BJ, UK.  e-mail: o.shorinola@bham.ac.uk

Received: 14 September 2023; Accepted: 14 December 2023;
Published online: 08 January 2024

References

- von Wettberg, E., Davis, T. M. & Smykal, P. Wild plants as source of new crops. *Front. Plant Sci.* **11**, 591554 (2020).
- VanBuren, R. et al. Exceptional subgenome stability and functional divergence in the allo-tetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
- Njaci, I. et al. Chromosome-level genome assembly and population genomic resource to accelerate orphan crop lablab breeding. *Nat. Commun.* **14**, 1915 (2023).
- Lin, J. et al. Signatures of selection in recently domesticated macadamia. *Nat. Commun.* **13**, 242 (2022).
- Bredeson, J. V. et al. Chromosome evolution and the genetic basis of agronomically important traits in greater yam. *Nat. Commun.* **13**, 2001 (2022).
- Moghaddam, S. M. et al. The tepary bean genome provides insight into evolution and domestication under heat stress. *Nat. Commun.* **12**, 2638 (2021).
- Chapman, M. A., He, Y. & Zhou, M. Beyond a reference genome: pangenomes and population genomics of underutilized and orphan crops for future food and nutrition security. *New Phytol.* **234**, 1583–1597 (2022).
- Ye, C.-Y. & Fan, L. Orphan crops and their wild relatives in the genomic era. *Mol. Plant* **14**, 27–39 (2021).
- Mabhaudhi, T. et al. Diversity and diversification: ecosystem services derived from underutilized crops and their co-benefits for sustainable agricultural landscapes and resilient food systems in Africa. *Front. Agron.* **4**, 859223 (2022).
- Talabi, A. O. et al. Orphan crops: a best fit for dietary enrichment and diversification in highly deteriorated marginal environments. *Front. Plant Sci.* **13**, 839704 (2022).
- Smith, S., Nickson, T. E. & Challender, M. Germplasm exchange is critical to conservation of biodiversity and global food security. *Agron. J.* **113**, 2969–2979 (2021).
- Woldeyohannes, A. B. et al. Data-driven, participatory characterization of farmer varieties discloses teff breeding potential under current and future climates. *eLife* <https://doi.org/10.7554/eLife.80009> (2022).
- Gesesse, C. A. et al. Genomics-driven breeding for local adaptation of durum wheat is enhanced by farmers' traditional knowledge. *Proc. Natl. Acad. Sci. USA.* **120**, e2205774119 (2023).
- Fadda, C. et al. Integrating conventional and participatory crop improvement for smallholder agriculture using the seeds for needs approach: a review. *Front. Plant Sci.* **11**, 559515 (2020).
- Marks, R. A., Hotaling, S., Frandsen, P. B. & VanBuren, R. Representation and participation across 20 years of plant genome sequencing. *Nat. Plants* **7**, 1571–1578 (2021).
- Varshney, R. K. Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* **242**, 98–107 (2016).

- Garcia, T. et al. Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nat. Commun.* **12**, 702 (2021).
- Dwyer, W., Ibe, C. N. & Rhee, S. Y. Renaming Indigenous crops and addressing colonial bias in scientific language. *Trends Plant Sci.* **27**, 1189–1192 (2022).
- Ebenezer, T. E. et al. Africa: sequence 100,000 species to safeguard biodiversity. *Nature* **603**, 388–392 (2022).
- Jamnadass, R. et al. Enhancing African orphan crops with genomics. *Nat. Genet.* **52**, 356–360 (2020).
- Yabe, S. & Iwata, H. Genomics-assisted breeding in minor and pseudo-cereals. *Breed. Sci.* **70**, 19–31 (2020).
- Chiurugwi, T., Kemp, S., Powell, W. & Hickey, L. T. Speed breeding orphan crops. *Theor. Appl. Genet.* **132**, 607–616 (2019).
- He, Q. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat. Genet.* **55**, 1232–1242 (2023).
- Lemmon, Z. H. et al. Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* **4**, 766–770 (2018).
- Lacchini, E. et al. CRISPR-mediated accelerated domestication of African rice landraces. *PLoS One* **15**, e0229782 (2020).
- Venezia, M. & Creasey Krainer, K. M. Current advancements and limitations of gene editing in orphan crops. *Front. Plant Sci.* **12**, 742932 (2021).
- Kim, D., Hilty, R., Hofmeister, E., Slowinski, P. R. & Steinhart, M. CRISPR/Cas technology and innovation: *Mapping Patent Law Issues*. (University of Munich, 2022).
- Kamenya, S. N., Mikwa, E. O., Song, B. & Odeny, D. A. Genetics and breeding for climate change in orphan crops. *Theor. Appl. Genet.* **134**, 1787–1815 (2021).
- Edwards, A. et al. Genomics and biochemical analyses reveal a metabolon key to β -L-ODAP biosynthesis in *Lathyrus sativus*. *Nat. Commun.* **14**, 876 (2023).
- Halewood, M., Bagley, M. A., Wyss, M. & Scholz, A. H. New benefit-sharing principles for digital sequence information. *Science* **382**, 520–522 (2023).

Acknowledgements

O.S. was supported by the Royal Society FLAIR award (FLR_R1_191850). R.M. was supported by NSF DBI grant no. 2213983 to the Water and Life Interface Institute. M.A.C. was supported by grant BB/X512035/1 from the BBSRC.

Author contributions

O.S., R.M., P.E., and M.A.C. conceptualised the paper. All the authors wrote the manuscript. R.M. and P.E. designed the figures.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44535-x>.

Correspondence and requests for materials should be addressed to Oluwaseyi Shorinola.

Peer review information *Nature Communications* thanks Bradley Campbell, Carlo Fadda and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024