Beyond a reference genome: pangenomes and population genomics of underutilised and orphan crops for future food and nutrition security

**Mark A. Chapman1,\*****, Yuqi He2 and Meiliang Zhou2,\***

1 Biological Sciences, University of Southampton, Life Sciences Building 85, Highfield Campus, Southampton, SO17 1BJ, UK;

2 Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Room 405, National Crop Gene Bank Building, Zhongguancun South Street No. 12, Haidian District, Beijing, 100081, China; ORCID:.

\* For correspondence: [m.chapman@soton.ac.uk](mailto:m.chapman@soton.ac.uk) (MAC) and [zhoumeiliang@caas.cn](mailto:zhoumeiliang@caas.cn) (MZ).

ORCID:

MAC: 0000-0002-7151-723X; YH: 0000-0001-9920-9653; ML: 0000-0001-5175-3684

Brief heading and social media: MAC at the Centre for Underutilised Crops can be followed on Twitter @CUCsoton

**Summary**

Underutilised crops are, by definition, under-researched compared to staple crops yet come with traits that may be especially important given climate change and feeding a globally increasing population. These crops are often stress tolerant, combined with unique and beneficial nutritional profiles. Whilst progress is being made by generating reference genome sequences, in this Tansley Review, we point out how this is only the very first step. We advocate that going ‘beyond a reference genome’ should be a priority, as it is only at this stage one can identify the specific genes and the adaptive alleles that underpin the valuable traits. We sum up how population genomic and pangenomic approaches have led to the identification of stress and disease tolerant alleles in staple crops and compare this to the small number of examples from underutilised crops. We also demonstrate how previously underutilised crops have benefitted from genomic advances and that many breeding targets in underutilised crops are often well-studied in staple crops. This cross-crop population level resequencing could lead to an understanding of the genetic basis of adaptive traits in underutilised crops. This level of investment may be crucial for fully understanding the value of these crops before they are lost.

**Keywords:** Crop Improvement,Food security, Genomics, Pangenome, Population genomics, Reference genome, Underutilised crops

Word count (summary): 200

Word count (main text): 6472

Number of figures: 2 (one in colour)

Number of tables: 1

Number of supplementary files: 2 (2 tables)

**1. Introduction**

**1.1 Genomics and its contribution to agricultural improvements**

*1.1.1 Genome-wide Association Study (GWAS)*

*1.1.2 Nested-Association Mapping (NAM)*

*1.1.3 Pangenomes*

**1.2 Genome-assisted advances in staple crops**

*1.2.1 Genomic advances in rice*

*1.2.2 Genomic advances in maize*

**2. Genomics of underutilised crops to-date**

**2.1 Reference genome sequences for underutilised crops and cross-crop comparisons**

**2.2 Population genomics in underutilised crops**

*2.2.1 Identifying crop wild relatives and untapped resources*

*2.2.2 GWAS and selection analyses*

**2.3 Genetic transformation and gene editing in underutilised crops**

**3. The way forward**

**3.1 The successful transition from underutilised and orphan to mainstream**

**3.2 What do we need and why? The advantages of going beyond a reference genome**

*3.2.1 Quality trait-marker linkage*

*3.2.2 Cross-crop analyses*

*3.2.3 Next steps*

**Acknowledgements**

**Literature Cited**

**1. Introduction**

In 2020, at least 720 million people (≥9.9% of the world’s population) faced hunger; an increase on previous years, and the greatest percentage of the total population since 2010 (FAO et al. 2021). With ongoing climate change, the increasing global population, and the COVID19 pandemic, the number of people facing hunger is expected to rise significantly. It is increasingly recognised that achieving Sustainable Development Goal (SDG) 2 (“Zero Hunger”) by 2030 is in doubt (Global Panel on Agriculture and Food Systems for Nutrition 2020). To overcome this challenge, one of the most favourable approaches is to breed new germplasm to face these stresses. Crop breeders have identified germplasm (either local varieties or crop wild relatives [CWRs]) with beneficial traits, and introgressed the adaptive alleles into elite varieties. Although substantial efforts and many successes have been made to breed climate resilient crops, yield has started to plateau because of negative effects of extreme weather events in at least some years and in some parts of the world (Zhao et al. 2017). Homogenization of diets and the increasing consumption of calorie-rich yet nutrition- deficient crops has also lead to an increase in hidden hunger, a significant factor threatening human health (Khoury *et al.* 2014; [Dawson *et al.*, 2019](#_ENREF_6)). A greater understanding of the effect of environmental stresses on crop quality, for example nutrient content and composition is also needed.

Staple crops are limited in their tolerance of a changing climate, forcing researchers and breeders to start to investigate new pathways to ensure future food security. Underutilised crops (sometimes known as semi-domesticated, orphan and/or forgotten crops) are locally important crops grown in limited regions with low-input conditions. They are currently impossible or inefficient to be produced at large-scale due to low yield, anti-nutrient content, consumer preferences, harvestability, or other properties. Because of the diverse nutrient and local adaptations, these crops are often locally important or vital, and represent a broad gene pool for future crop improvement. Examples include high iron, zinc and calcium content in tepary bean and bambara groundnut (Chibarabada et al. 2017), tolerance of drought stress in lablab, horsegram and cowpea (Ewansiha and Singh 2006; Muchow 1985), and tolerance of heat in amaranth and cassava (Kuo et al. 1992). Thus, underutilised crops have significant potential to aid food security (Massawe et al. 2016; Mabhaudhi et al. 2017; Li et al. 2020b; Siddique et al. 2021). However, some of underutilised crops are becoming abandoned due to globalisation, migration and other economic forces and therefore are at risk of loss of diversity, or complete loss of the crop, or the loss of associated indigenous knowledge (Mabhaudhi et al. 2017).

This review will summarise advances made in major crops, and how genome sequencing has brought about a step-change in agricultural research. We also discuss some crops which could have been described as underutilised only *c*. 30 years ago but have become more mainstream, and the genomics-assisted research that has taken place. Beyond this, we discuss the benefits of venturing ‘beyond a reference genome’, i.e., resequencing, population genomics and pangenomics, and what advantages this can bring. We encourage the research community to incorporate underutilised crops into wider research and to envisage the benefits that come from this increased investment.

**1.1 Genomics and its contribution to agricultural improvements**

For staple crops, notable successes have been made using genomic-related approaches to uncover the genes responsible for important phenotypes.

*1.1.1 Genome-wide Association Study (GWAS)*

GWAS relies on genome-scale data (dense single nucleotide polymorphisms [SNPs] or whole genome sequences) to associate genomic variation with trait variation (Liu and Yan 2019). Typically, the analysis of hundreds of varieties/lines is necessary to generate sufficient power to resolve quantitative trait loci (QTL), as well as to ensure rare causative variants are included (e.g., Xing et al. 2015). This breadth of variation encompasses orders of magnitude more recombination than QTL mapping therefore can resolve even small-effect QTL and with typically greater resolution. Thousands of GWAS have been reported and do not only cover the staple cereals, for example cucumber, peach, sesame and lettuce (Liu and Yan 2019), however GWAS of crops that fit the definition of underutilised are in their infancy (see section 2.2).

*1.1.2 Nested-Association Mapping (NAM)*

NAM involves the development of multiple mapping populations sharing one parent, circumventing in part the issues related to traditional QTL mapping approaches only covering the variation present in two parents and the low power to tightly define QTL due to limited recombination events (McMullen et al. 2009). With this increased power (Bouchet et al. 2017), resources are being developed for a range of crops, and have provided candidate genes for a range of agronomic traits, for example flowering time in maize (Buckler et al. 2009), barley (Maurer et al. 2015) and sorghum (Bouchet et al. 2017), and plant architectural traits in maize (Brown et al. 2011) and durum wheat (Kidane et al. 2019). Many of these studies have confirmed previous work as well as identified novel QTL for follow-up, however, are currently restricted to only the main crops.

*1.1.3 Pangenomes*

It has become increasingly clear that across varieties of a crop, presence/absence variation is common; this means that any one reference genome contains only a subset of the species’ genome, potentially lacking causative loci (Della Coletta et al. 2021). Sequencing and assembling multiple varieties and adding novel genomic regions to the reference results in the pan-genome. Genes present in all accessions as ‘core’ and others ‘dispensable’. In some cases, half or less of some crop genomes are core (Gordon et al. 2017; Wang et al. 2018; Haberer et al. 2020), although these values are highly dependent on the number of accessions sequenced, and whether wild relatives are examined. This rare variation can underlie loci of agronomic importance, for example a flavour-related gene in tomato (Gao et al. 2019). Pangenomics has also revealed some important evolutionary genomic insight; core maize genes exhibit much greater expression than dispensable genes (Haberer et al. 2020), and dispensable regions tend to have higher transposable element (TE) content than core regions (e.g., Gao et al. 2019).

Most of the above studies have focussed on agronomic traits such as yield and its associated components (flowering time, plant stature and size), however because of ensuing climate change, a concerted effort to increase the study of climate change resilience phenotypes, for example root traits and stress response would be prudent. Whilst some studies have investigated this, for example using NAM to explore drought adaptive phenotypes in barley (Pham et al. 2019) and maize (Li et al. 2016), these are comparatively rare.

**1.2 Genome-assisted advances in staple crops**

Here we focus on rice and maize as two of the most widely grown staple crops, collectively grown on 360 M ha in 2019 (<http://faostat.fao.org>; accessed August 2021), to highlight how genomic advances have led to significant crop improvement for hundreds of millions of consumers.

*1.2.1 Genomic advances in rice*

Rice is one of the most important crops providing staple food for more than half of the world population. Due to the relatively small genome size (430 Mb), co-linearity with other cereals, a high efficiency genetic transformation system, and abundant genetic resource, rice has become a model grass species equivalent to the eudicot *Arabidopsis*. Significant efforts have been made to assemble and annotate the rice genome, for the *japonica* subspecies (cv. Nipponbare; Goff et al. 2002) and the *indica* subspecies (93-11; Yu et al. 2002). Compared to *Arabidopsis*, rice exhibits a gradient in GC content, which means a large proportion of rice genes have no obvious homolog in *Arabidopsis* (Yu et al., 2002).

Rice is an excellent candidate for population genomic analysis because of its extremely strong population structure and the large extent of LD owing to self-pollination. Resequencing of 527 rice landraces clearly separate the two cultivated subspecies and further divided these by latitudinal photoperiod and temperature clines (Huang et al. 2010). Genome analysis of 446 wild rice (*Oryza rufipogon*) accessions and 1083 cultivated varieties has identified a population of *O. rufipogon* in southern China where it appears that the domestication into the *japonica* subspecies took place; from there, hybridisation with local wild rice formed *indica* rice which spread into Southeast and South Asia (Huang et al. 2012). Combining genomic and phenotypic data through GWAS, a variety of loci associated with agronomic traits in rice have been identified (Supplementary Table 1).

Whilst mapping short sequence reads of other varieties onto one of the reference genomes can allow these types of analyses, this means that genomic regions absent from the reference will be ignored. Therefore, more recently the genomes of other varieties with beneficial agronomic traits (Zhao et al. 2018; Choi et al. 2020), weedy rice (Sun et al. 2019) and other species of *Oryza* (Li et al. 2020a) have been sequenced and assembled, forming a potential rice pangenome resource. A rice pangenome initiative, involving the comparison of 16 *de novo* assembled genomes from the main population genomic groups (sequenced and assembled using long reads and optical mapping, hence are described as ‘platinum’ genomes), showed that an average of 33.7 Mb of genome was absent among all pairwise comparisons (Zhou et al. 2020). This highlights how using one reference will only ever describe variation in a subset of the entire crop’s (pan)genome. It is worth noting that the majority (nearly 90%) of the presence/absence fraction of the rice pangenome is made up of TEs indicating these are more evolutionary labile, however the remaining 10% contains potentially protein-coding loci. These genome resources will promote evolutionary studies and the identification of adaptive genetic variation in rice.

*1.2.2 Genomic advances in maize*

Maize was domesticated from teosinte in southwestern Mexico *c*. 10,000 years ago. From its wild progenitor, maize has evolved a strikingly different morphology, forming an unbranched plant with large cobs and sweet naked kernels (Doebley 1990; Matsuoka et al. 2002). Since then, maize has been continuously improved, and an array of hybrid lines suitable for modern agricultural practice have been developed. In the past century, maize yield has increased eight-fold due to the increase of yield per plant and plant density adapted by harnessing heterosis. The ancestor genome of maize experienced a tetraploid intermediate stage (n=20), and then a series of chromosome fusion led to diploidisation and the recovery of the chromosome number n=10 (Schnable 2015). Thus, the maize genome is very large (Gaut and Doebley 1997) and is especially known for its array of TEs, indeed it is from maize that McClintock first hypothesised that some genetic elements could be mobile (McClintock 1950).

Using bacterial artificial chromosome and fosmid clones, the genome of maize variety B73 was assembled, and revealed that long-terminal repeat retrotransposons (LTRs) account for 74% of genome. Proliferation of the LTRs was the primary reason why the maize genome is so expanded relative to other grasses (Schnable et al. 2009). Resequencing of 17 wild relatives, 23 traditional landraces and 35 improved maize lines and mapping back to the B73 reference genome suggested that introgression from wild relatives could be responsible for diversity recovery in maize post-domestication and identified genes with diverse biological functions having been under selection during domestication (Hufford et al. 2012).

Although significant advances have been made in maize genome sequencing and population genomics, GWAS in maize is a challenge because the linkage disequilibrium (LD) decays within 2 kb. The development of the large NAM panel (McMullen et al. 2009) has significantly increased the efficiency of GWAS in maize, and loci associated with multiple agronomic traits have been identified (Supplementary Table 2).

Using single-molecule real-time sequencing and high-resolution optical mapping, an improved B73 genome was more recently assembled (Jiao et al. 2017). Due to significant structural variation among inbred lines, the genome of B73 alone is not sufficient to fully explain the variation among other inbred lines. Thus, the genome of other inbred lines including PH207 (Hirsch et al. 2016), W22 (Springer et al. 2018), Mo17 (Sun et al. 2018), HuangZaoSi (Li et al. 2019), small-kernel (Yang et al. 2019) and B73-Ab10, a variant of B73 containing Abnormal chromosome 10 (Liu et al. 2020b), have been assembled. This and other ongoing work will facilitate the understanding of maize genome diversity, as well as the breeding and improvement of maize.

**2. Genomics of underutilised crops to-date**

The genetic improvement of underutilised crops is, in part, constrained by limited genome resources. Recent developments in genome technology and the reduction of sequencing costs means genome-scale research is no longer limited to major food crops (Table 1).

**2.1 Reference genome sequences for underutilised crops and cross-crop comparisons**

The Poaceae (grasses) is the second largest plant family, with *c*. 12,000 species. Besides the staple crops rice, maize, wheat, and sugarcane, and some previously common cereals such as barley, oats and rye, the Poaceae also contains many underutilised crops, including sorghum, foxtail millet, finger millet and broomcorn millet that all use C4 photosynthesis, in which photorespiratory losses induced by hot and arid environments are reduced. The conversion of C3 rice and wheat towards C4 photosynthesis is a long-standing biotechnological goal. Comparative genomics has revealed that genes involved in C4 carbon fixation are all present in C3 plants (Zhang et al. 2012), and therefore the C4 pathway might have evolved from ancestral C3 isoforms. The panicoid grasses maize and sorghum show greater conservation of these genes compared to the pooid grasses rice and *Brachypodium* (Bennetzen et al. 2012). Further, a tandem duplication of the carbonic anhydrase Caβ subfamily, which hydrates atmospheric CO2 to bicarbonate in the mesophyll, was found in C4 plants, potentially vital for C4 evolution (Zhang et al. 2012). Genome comparisons between underutilised C4 crops and the staple C3 crops in the Poaceae will provide new suggestions for the evolution C4 photosynthesis, with the potential to improve the photorespiration efficiency and subsequent drought tolerance of other underutilised and staple crops.

The Fabaceae is the third-largest plant family, including many agronomically important grain and forage species. Legumes can improve soil fertility through the fixing of atmospheric nitrogen via root nodule-specific bacteria. The discovery of many genes involved in nitrogen fixing has been aided through the study of underutilised legume genomes (Jain et al. 2013; Lu et al. 2018; Zhuang et al. 2019). In addition, the legumes also contain species with unique nutritional features; for example, adzuki bean, widely cultivated in Asia, is referred to as the “weight loss bean” due to its sweet taste but low caloric and fat content. Genomic comparison to other legumes found that adzuki bean has fewer starch and fatty acid biosynthesis genes, which could play a role in its unique nutritional profile (Yang et al. 2015).

Several underutilised crops are advocated as worthy of investment because of their extreme stress resilience, often greater than staple crops (Cullis and Kunert 2017; Massawe et al. 2015). Further comparative genomics in the Poaceae has identified numerous gene family expansions associated with stress tolerance in underutilised crops, and these might explain the high stress resistance in underutilised crops. Drought tolerant foxtail millet and sorghum (compared to drought susceptible rice and maize) contain expansions of stress response genes, including those encoding cytochrome P450 proteins, expansins, lipid transfer proteins, and several others, as well as miRNA169 targeting drought stress-associated transcription factor nuclear factor-YB (Paterson et al. 2009; The International Brachypodium Initiative 2010). Tef, a drought tolerant cereal mainly distributed in Ethiopia, contains a tandem duplication of the nucleotidase/phosphatase *SAL1*, a gene family involved in drought tolerance, relative to other grasses investigated (Cannarozzi et al. 2014). Pearl millet possesses more members of cutin, suberin, wax biosynthetic and metabolite transporter genes, which might be responsible for the heat and drought tolerance in this underutilised crop (Varshney et al. 2017b). The number of BTB ubiquitin E3 ligases is greater in grasses than *Arabidopsis* and one subgroup, the BTB-BACK subgroup, was only expanded in the underutilised cereal broomcorn millet (Zou et al. 2019) which may contribute to the excellent stress tolerance. Clearly, genome comparison of these underutilised crops will provide a new pool of stress targeted genes for well-studied main crops. Similar results were found in the genus *Dioscorea*, in which draft genomes have been assembled for two yam species (Tamiru et al. 2017; Siadjeu et al. 2020), and phylogenetic analyses show that *Dioscorea* has more bulb-type lectin genes than the Poaceae and *Arabidopsis*, with potential roles in the insecticidal properties of Guinea yam (Tamiru et al. 2017).

In sum, reference genomes of underutilised crops can help resolve the genetic basis of agronomic traits, especially as a (sometimes unique) resource for improving the photorespiratory efficiency, nutritional value and stress tolerance of related major food crops currently challenged by climate change.

**2.2 Population genomics** **in underutilised crops**

Population genomics of underutilised crops can help researchers understand population structure and domestication history, as well as aid in identifying candidate genes modulating key agronomic traits through GWAS and to develop molecular markers for marker-assisted breeding (MAB; Figure 1).

*2.2.1 Identifying crop wild relatives and untapped resources*

The combination of genome resequencing and phylogenetic analysis enables us to better understand the population structure and identify wild relatives. For example, resequencing of 994 pearl millet lines identified four main clusters and pinpointed a West African origin of the crop (Varshney et al. 2017b). Resequencing of 166 fonio millet accessions and 17 proposed progenitors found a significant separation between cultivated and wild accessions, and the cultivated accessions were genetically closest to wild accessions from Southern Togo and West Guinea (Abrouk et al. 2020). Further, diversity of the grain size gene *GS5* was significantly reduced in fonio millet cultivars (with larger grains than the wild progenitor) suggesting human selection on this locus during domestication. Resequencing of 510 accessions of Tartary buckwheat from the entire global distribution identified three major clades and indicate two geographically distinct domestication events (Zhang et al. 2021). Similar investigations in other underutilised crops have identified wild relatives and genetic sub-populations of the domesticates, for example adzuki bean (Yang et al. 2015), lupin (Wang et al. 2021a; Hufnagel et al. 2020), and coix (Liu et al. 2020a). Resequencing of Guinea yam and its potential wild relatives resolved a hybrid origin from a cross between the wild rainforest species *D. praehensilis* and the savanna species *D. abyssinica* (Sugihara et al. 2020).

*2.2.2 GWAS and selection analyses*

The high-density single nucleotide polymorphism (SNP) data obtained from population-scale genome resequencing can be used to identify the genetic basis of adaptive traits through GWAS and selection analyses. From this, genetic markers can be designed and used for MAB. Beyond breeding, genetic modification-type approaches can be used to insert the candidate genes from a stress tolerant or otherwise novel underutilised crop into a susceptible or trait-lacking staple crop.

Seed size and weight are some of the most important traits of many crops, affecting price and milling qualities. GWAS of 368 cowpea accessions found candidate genes involved in endosperm development, embryo development and cell elongation associated with seed size, some of which also play a role in common bean (Lo et al. 2019). Similar analyses in castor bean identified candidate genes for seed traits that differentiate the wild and cultivated types (Xu et al. 2021). Recently, GWAS of Tartary buckwheat identified a mutation in the GCC cis-element of an *AP2* transcription factor associated with grain weight (Zhang et al. 2021).

Underutilised crops can possess multiple agronomic traits which are not present in staple crops, therefore present untapped resources for traits such as abiotic tolerances and novel nutrients. GWAS has been used to identify genes involved in tuber quality in cassava (Zhang et al. 2018b) and anthocyanin content in mungbean (Noble et al. 2018). Regarding stress tolerance, genes involved in lateral root development, stress tolerance and phosphorus use efficiency of mungbean have been identified (Reddy et al. 2020).

Comparing wild and domesticated cassava genomes has identified selective sweeps in genes involved in photosynthesis, starch accumulation and stress response (Wang et al. 2014). Extending this to other crops will facilitate the development of markers associated with domestication related traits. Candidate genes identified in underutilised crops through GWAS can help improve the quality and stress tolerance of not only the underutilised crops, but related staple crops, to better suit the needs of human beings in a changing climate.

Extended population genomics, for example using reduced representation technologies can be used to link genetic markers to agronomic traits, without knowing the precise genetic basis of the trait. Markers linked to several diverse agronomic traits in foxtail millet have been identified using these approaches, for example coloration, leaf size and shape, grain yield and weight, and flowering time (Upadhyaya et al., 2015; Jaiswal et al., 2019). Similarly, QTL mapping approaches at sufficient density can provide marker-trait associations, for example markers associated with yield and flowering in pea (Annicchiarico et al. 2017) and dormancy in groundnut (Kumar et al. 2020).

Genomics-assisted breeding of underutilised crops has been limited due to the lack of molecular markers linked to traits of interest, however the recent recognition of the importance of underutilised crops and the development of genome technology has clearly started to remedy this. The use of these markers to accelerate breeding (i.e., genomic selection) has been shown for a handful of underutilised crops (Ye and Fan 2020).

**2.3 Genetic transformation and gene editing in underutilised crops**

In recent years, genetic engineering has been widely used to elucidate gene function and for crop improvement. Compared with traditional hybridisation and crossing of varieties, genetic engineering could deliver agronomically useful traits into plant faster and in a more targeted manner. *Agrobacterium*-based transformation systems are widely used for genetic transformation in plants, facilitating the integration of foreign gene copies into the host plant’s genome. Although *Agrobacterium* transformation has been successfully used for transformation in several major crops, the inherent limitations associated with resistance to *Agrobacterium* infection and their recalcitrance to *in vitro* regeneration limit the transformation of many orphan crops. At present, *Agrobacterium*-mediated transformation has only been successfully transformation in shoot apex explants of finger millet ([Ceasar & Ignacimuthu, 2011](#_ENREF_4)) and foxtail millet ([Ceasar *et al.*, 2017](#_ENREF_3)), callus derived from mature seeds of finger millet ([Hema *et al.*, 2014](#_ENREF_9)), green millet ([Martins *et al.*, 2015a](#_ENREF_15); [Nguyen *et al.*, 2020](#_ENREF_19)), sorghum ([Zhao *et al.*, 2000](#_ENREF_24); [Belide *et al.*, 2017](#_ENREF_2)) and foxtail millet ([Santos *et al.*, 2020](#_ENREF_20)), embryonic axis explants of pigeonpea ([Ghosh *et al.*, 2017](#_ENREF_8)), germinated seedlings of chickpea ([Senthil *et al.*, 2004](#_ENREF_21)), and hairy roots of chickpea ([Aggarwal *et al.*, 2018](#_ENREF_1)) and buckwheat ([Mi *et al.*, 2020](#_ENREF_18)). Still, the recalcitrant tissue culture efficiency and occasional and unpredictable chimerism hinder the efficiency of these tissue culture-based methods.

Recently, using the floral-dip *Agrobacterium*-mediated transformation method, the wild ancestor of foxtail millet, green millet, was successfully transformed ([Martins *et al.*, 2015b](#_ENREF_16)). This is a significant advance because millets are model C4 grasses, and green millet it diploid, with a rapid life cycle, small genome size, simple growth requirements and high transformation efficiency.

Despite ongoing challenges of carrying out gene editing in even the best studied crops (Yang, 2020), CRISPR/Cas9-based gene editing has been conducted in underutilised crops with relatively high tissue culture efficiency, including green millet ([Weiss *et al.*, 2020](#_ENREF_23)) and sorghum ([Jiang *et al.*, 2013](#_ENREF_11); [Che *et al.*, 2018](#_ENREF_5)). These approaches will provide the necessary technical support for improving the efficiency of confirming the function of unique genes and the development of advantageous varieties of underutilised crops.

**3. The way forward**

**3.1 The successful transition from underutilised to mainstream**

In the past 20 years, a few previously underutilised crops, such as quinoa, chickpea and pigeonpea have seen a significant boost in research and recognition. For these crops we have seen a parallel 20% to 500% increase in the area grown worldwide between the 1960s and 2010s (<http://faostat.fao.org>; accessed August 2021). Chickpea and pigeonpea were among the first underutilised crops to have their genomes sequenced (Varshney et al. 2012; Varshney et al. 2013) with the (tetraploid) quinoa genome being made available more recently (Jarvis et al. 2017). Clearly the availability of genome sequence was a major stepping-stone in resolving the genetic basis of adaptive and agronomic phenotypes in these crops.

Using pigeonpea as an exemplar, this crop was recognised as worthy of significant investment in the 1970s, with the Pigeonpea Genomics Initiative (PGI) established in 2006 (Varshney et al. 2010). After genome sequencing (Varshney et al. 2012), significant advances have been made identifying genomic regions underlying adaptive traits which could be crossed between varieties using MAB (Varshney et al. 2017a); for example markers associated with sterility mosaic disease (Saxena et al. 2017a) and fusarium wilt (Saxena et al. 2017b). Genomic analysis has revealed less genes involved in lipid biosynthesis in pigeonpea than in soybean, and more cellulose synthesis genes, which together might underlie the biochemical and morphological differences between pigeonpea and other legumes (Singh et al. 2012). In addition, a pigeonpea gene involved in disease resistance was cloned and transferred to soybean, conferring resistance to Asian soybean rust (Kawashima et al. 2016), which would have been impossible without using the pigeonpea genome sequence.

More recent GWAS of nearly 300 pigeonpea accessions (Varshney et al. 2017a) identified dozens of associations and provides significant resources for MAB (Bohra et al. 2020). Pigeonpea is probably the only underutilised crop for which a pangenome has been sequenced (Zhao et al. 2020); this has 55,512 genes, compared to the reference genome (Varshney et al. 2012), which has only 53,612 (when annotated in exactly the same way as the pangenome). Using this pangenome, novel GWAS associations have been identified (Zhao et al. 2020) which were absent using the single reference genome (Varshney et al. 2017a). This further highlights the additional insights that can be made when a pangenome is made available.

Chickpea is grown and consumed worldwide, but 30 years ago could have been considered underutilised. Although productivity has steadily increased, the development of accessions with greater yield, improved nutrition and stress resistance are essential to meet the increasing demands. Comparative genomics of legumes has identified a lack of some resistance and nodulation genes, potential reasons for the low stress resistance (Jain et al. 2013; Varshney et al. 2013). Resequencing panels have identified genetic groups of cultivars (primarily the desi and kabuli types), identified the origin of the crop, and uncovered genes involved in drought tolerance and heat stress response through GWAS (Varshney et al. 2013; Varshney et al. 2019).

Several other previously underutilised crops are seeing a revolution in their investigation, suggesting they are on the path to escaping some of the reasons they were previously underutilised. The following examples are case studies of crop species early on this trajectory and provide ideas to circumvent issues such as large genomes and examples of crops with unique attributes which have received investment.

For species with polyploid genomes, investigations of related diploids can shed light on agronomic traits. Oat is nutritional crop containing abundant calcium, dietary fibre (especially β-glucan), and unsaturated fatty acids ([Joyce *et al.*, 2019](#_ENREF_12)). Due to the cholesterol-lowering properties and the antidiabetic effect of β-glucan, oat has been widely used to adjuvant treatment of diabetes and cardiovascular disorders. The rotation of oat with other crops can improve soil structure and reduce diseases in other crops. This disease resistance has been attributed to the production of avenacins, specialised antifungal metabolites. Oat is allohexaploid, with a relatively large, highly repetitive, and rearranged genome, thus brings challenges for genome assembly. The current sequencing has mainly focussed on wild diploid oats. For example, through genome assembly of the diploid extant progenitors, candidate genes regulating flowering time and disease resistance were identified ([Maughan *et al.*, 2019](#_ENREF_17)). Genome assembly of other diploid accessions identified a 12-gene cluster responsible for avenacin biosynthesis, and this cluster was located in a subtelomeric region which may have formed since oat diverged from other crops ([Li *et al.*, 2021](#_ENREF_14)). These results shed light on the evolution of oat and will help breeding oat varieties with modified and improved health benefits.

For other underutilised crops, they bring qualities and traits which are lacking in mainstream staple crops, and as such significant investment has begun to begin their escape from being underutilised. Quinoa is one example, a crop of the Chenopodiaceae, which has been cultivated for approximately 7000 years. Its diverse environmental adaptability means it is grown from the sea level of Chile to altitudes above 4500m in Bolivia ([Suárez-Estrella *et al.*, 2018](#_ENREF_22)). Due to its extraordinary balance of essential amino acids, and abundant vitamins, minerals, dietary fibre and unsaturated fatty acids, it was recognized as a complete food and has attracted the attention of the scientific community ([Filho *et al.*, 2017](#_ENREF_7)). However, quinoa contains bitter and astringent anti-nutritional factors such as saponins. Although these substances are health-promoting, their bitter taste still limit the utilization of quinoa (Suárez-Estrella et al., 2018). Thus, selection of genotypes with low saponin content is one of the most important quinoa breeding objectives for the future. In addition, due to the outcrossing nature, genome assembly of quinoa was not trivial, requiring repeated self-pollination to reduce heterozygosity (Yasui et al., 2016a). Analysis of the subsequently assembled quinoa genome (Yasui et al. 2016a) identified expansions of gene families involved in lysine, vitamins, polyphenol and betalain synthesis, as well as ABA signalling, which together may relate to the unique nutritional and anti-nutritional factors profile and abiotic stress tolerance in quinoa (Yasui et al. 2016a; Zou et al. 2017).

Another example of an underutilised crop with novel attributes is buckwheat, a pseudocereal originating from and domesticated in China more than 4000 years ago (Zhang et al., 2021). This crop possesses outstanding nutritional profile (especially flavonoids) and an excellent ability to grow under adverse climatic and soil conditions. The main cultivated species are common buckwheat and Tartary buckwheat. Similar to quinoa, being outcrossing makes genome sequencing of common buckwheat more challenging and required repeated self-pollination to reduce heterozygosity (Yasui et al. 2016b). In contrast to common buckwheat, the sequencing of Tartary buckwheat was relatively simple because of a smaller genome and being predominantly a selfer. Comparative genomics using the chromosome-scale Tartary buckwheat genome, revealed a whole-genome duplication event after buckwheat divergence from sugar beet with some evidence this might play a role in buckwheat tolerance of extremely harsh environments (Zhang et al. 2017). Genome resequencing of Tartary buckwheat identified two independent domestication events, in southwestern and in northern China, which has resulted in the diversity of modern Tartary buckwheat varieties (Zhang et al., 2021). Candidate genes responsible for flavonoid biosynthesis were also identified and will help buckwheat breeding with improved health and medical benefits.

Given these findings from the significant investment and the ongoing work in representative orphan crops (Roorkiwal et al. 2020), we feel encouraged that the resources and investment needed for these crops to be elevated to the national stage are in place. However these are only the tip of the underutilised crop iceberg; dozens of underutilised crops have a single reference genome, and in some cases small resequencing panels (Table 1) but significant population and GWAS resources or pangenomes are absent for the vast majority.

**3.2 What do we need and why?****The advantages of going beyond a reference genome**

To efficiently breed improved varieties of underutilised crops we need to have reliable linkage between genetic markers and traits of interest. Markers identified in a single QTL mapping experiment may not be reliable given many QTL are only expressed in some environments (genotype x environment interaction) and do not always tightly define the genomic region (therefore the QTL spans dozens or hundreds of genes). More precision can be gleaned from LD mapping approaches (Thornsberry et al. 2001), including GWAS, which requires extensive panels of germplasm and high marker density. A reference genome is an asset to begin to understand important and adaptive phenotypes in underutilised crops, yet it is becoming clear that significant advances in breeding improved varieties are only possible when the genomic variants are identified, thus requiring a population of genomes and potentially a pangenome.

*3.2.1 Quality trait-marker linkage*

One main advantage to having population level sequencing is to tie this to trait data using GWAS-type approaches. This is an efficient way to start to narrow down the genetic basis of quantitative traits such as yield, seed and organ size, plant stature and so on, all traits which need to be optimised to ensure a crop is cost-effective to be grown at scale (Figure 2). A reference genome is an asset, but without the resequencing (or high-density SNP genotyping), GWAS cannot be done. Examples of GWAS-style analyses in the underutilised crops cowpea, castor bean, Tartary buckwheat, cassava and mungbean are given above.

Using reduced-representation technology (e.g., genotyping-by-sequencing, or restriction site-associated DNA sequencing) is an efficient way to screen large numbers of genetic loci throughout the genome at lower cost than resequencing. The downside being that unless LD extends for very long distances, most markers will be so widely spaced that they will be too far from the underlying causative loci to be associated. Nevertheless, these approaches have yielded marker-trait associations (MTAs) in some underutilised crops and pave the way for more in-depth analyses. For example, in Proso millet (*Panicum milaceum* L.) 13 MTAs for seven traits were resolved, however no MTAs were found for another nine traits (Boukail et al. 2021). In foxtail millet, 81 MTAs for ten traits were resolved, however most did not pass false discovery rate correction (Jaiswal et al. 2019), and in Kersting’s groundnut, ten MTAs for five traits were resolved (Akohoue et al. 2020). It is important to bear in mind that because of the marker spacing, partly due to the location of cut sites in the genome, and partly because of uneven sequencing coverage across loci (Beissinger et al. 2013), these studies are likely to only report a subset of genomic loci involved in the traits of interest.

*3.2.2 Cross-crop analyses*

Whilst underutilised crops may lack some desirable traits (fast growth, ease of harvest, high harvest index), or have additional phenotypes that are unwanted (antinutrients, perenniality), significant advances have been made in understanding the genetic basis of many of these traits in other crops (Figure 2). Thus, where these traits have been studied in detail, there may be candidate genes that can be followed up if population resequencing or pangenomic data are available in the underutilised crop.

Several underutilised crops would benefit from having more predictable flowering/fruiting or being adapted to novel environments where the daylength is different. The legume lablab is typically a short-day plant therefore expansion outside its native tropical latitudes is unlikely to be successful (Sennhenn et al. 2017). Bambara groundnut, another tropical underutilised legume, is typically short-day although a few semi-improved varieties can be grown further from the equator, however it is acknowledged that this is still a barrier to more widespread adoption (Mayes et al. 2019). Daylength response (and therefore flowering time) is relatively well-studied in staple crops, including rice, maize and the typically long-day legumes, lentil and pea (Weller et al. 2012; Itoh and Izawa 2013; Hung et al. 2012), offering candidate genes for the development of underutilised cereal and legume varieties for adaptation to non-native latitudes. Candidate genes, or genomic regions, underlying annual vs perennial growth have been identified in Brassicaceae species (Kiefer et al. 2017; Heidel et al. 2016); this is another trait which might help the adoption of underutilised crops.

An often-cited reason for the poor adoption, or decline in use, of underutilised crops is their antinutrient content. Antinutrient factors (ANFs) inhibit the uptake of beneficial minerals and vitamins, therefore a high nutrient crop with high ANF content will have low nutrient bioavailability. This is especially the case in legumes where several ANFs have been identified that affect iron, zinc and protein uptake. Whilst cooking and fermentation can reduce the presence of these compounds (e.g., (Samtiya et al. 2020)) these take time or energy (e.g., fuel for cooking). However, ANFs are usually vital for crop disease resistance, therefore breeding for high ANF during the growth period coupled with low ANF in the maturation period would clearly be advantageous. Progress has been made understanding the genetic basis of these traits (Sparvoli and Cominelli 2015; Campion et al. 2013) therefore opening the door for understanding the genetic basis of these traits in underutilised crops.

There are other traits which make underutilised crops less attractive as a choice for a farmer or the consumer, for example poor shelf-life, unpleasant taste, or lengthy cooking times (and an increase cost for fuel). Genes involved in shelf life in tomato have been elucidated (Casals et al. 2012; Zhang et al. 2018a), along with QTL for alkaloid content in lupin (Rychel and Książkiewicz 2019), and for seed hardness, and therefore cooking time, in legumes (Sandhu et al. 2018; Diaz et al. 2021).

Relatedly, many underutilised crops are known for their extreme resilience phenotypes. Any analysis of the genetic basis of drought or heat tolerance in any underutilised crops, likely requiring population sequencing for e.g., GWAS, will be of significant value to other more mainstream crops. This could resolve novel alleles or even undercover novel genes and pathways involved in these climate-change relevant tolerances. The sequencing of one reference genome of an underutilised crop cannot offer this.

Population level resequencing mapped to one reference will not be able to examine the fraction of the genome that is only present in some accessions (presence-absence variation lacking from the sequenced reference which would only be identified in a pangenome). This problem could be underestimated for underutilised crops where variation in genome size might not be recognised; for example, the underutilised legume lablab was likely domesticated twice (Maass et al. 2017; Robotham and Chapman 2015), and the two gene pools differ in genome size by ~20% (MAC, unpublished).

*3.1.3 Next steps*

We propose that efforts should be made to not only generate a reference genome but carry out population level sequencing and pangenomics. In parallel we encourage the continued collection and long-term archiving of seed resources, and addressing the challenges associated with archiving the required indigenous knowledge associated with these under-investigated species (Kamenya et al. 2021; Mabhaudhi et al. 2017). Researchers should make data free to use, and collaborations between institutes worldwide encouraged to expedite the production of results and limiting unnecessary overlap and wasted resources.

Whilst the cost and time implications of multiple reference genomes, resequencing, and collecting global germplasm are not trivial, we believe that, given the climate crisis and the need to fast-track the development of mainstream and novel crops, this is the most reliable way to ensure that underutilised crops are investigated in the depth at which reliable and meaningful data can be used. It is likely that some underutilised crops hold vital genetic variants to help the human population combat food insecurity in the next few decades; this genetic erosion is under-investigated even in staple crops (Khoury et al. 2022). Without fully investigating underutilised cropgenomes, we do not know where these variants lie, and if we delay too long, we may lose alleles, varieties and crops entirely.

**Acknowledgements**

This research was supported by grants from the National Key R&D Program of China (2019YFD1001300/2019YFD1001302) and National Natural Science Foundation of China (31871536 and 32161143005) to MZ and from the Natural Environment Research Council and Biotechnology and Biological Sciences Research Council (NE/S002022/1) to MAC. We thank our lab groups for ongoing discussions into these crops as well as three anonymous reviewers who gave insightful comments and suggestions on a previous version of this review,

**Author contributions**

MAC, YH and MZ researched the topic and wrote the manuscript.

**Literature Cited:**

**Abrouk M, Ahmed HI, Cubry P, Šimoníková D, Cauet S, Pailles Y, Bettgenhaeuser J, Gapa L, Scarcelli N, Couderc M, et al. 2020.** Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. *Nature Communications* **11**(1): 4488.

**Aggarwal PR, Nag P, Choudhary P, Chakraborty N, Chakraborty S. 2018.** Genotype-independent Agrobacterium rhizogenes-mediated root transformation of chickpea: a rapid and efficient method for reverse genetics studies. *Plant Methods* **14**: 55.

**Akohoue F, Achigan-Dako EG, Sneller C, Van Deynze A, Sibiya J. 2020.** Genetic diversity, SNP-trait associations and genomic selection accuracy in a west African collection of Kersting’s groundnut [*Macrotyloma geocarpum* (Harms) Maréchal & Baudet]. *PLOS ONE* **15**(6): e0234769.

**Annicchiarico P, Nazzicari N, Pecetti L, Romani M, Ferrari B, Wei Y, Brummer EC. 2017.** GBS-Based Genomic Selection for Pea Grain Yield under Severe Terminal Drought. *Plant Genome* **10**(2): 1-13.

**Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, Twardziok SO, Hackauf B, Gordillo A, Wilde P, et al. 2017.** Towards a whole-genome sequence for rye (*Secale cereale* L.). *The Plant Journal* **89**(5): 853-869.

**Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N. 2013.** Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* **193**(4): 1073-1081.

**Belide S, Vanhercke T, Petrie JR, Singh SP. 2017.** Robust genetic transformation of sorghum (*Sorghum bicolor* L.) using differentiating embryogenic callus induced from immature embryos. *Plant Methods* **13**: 109.

**Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. 2012.** Reference genome sequence of the model plant Setaria. *Nature Biotechnology* **30**(6): 555-561.

**Bohra A, Saxena KB, Varshney RK, Saxena RK. 2020.** Genomics-assisted breeding for pigeonpea improvement. *Theoretical and Applied Genetics* **133**(5): 1721-1737.

**Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Tuinstra M, Morris GP. 2017.** Increased Power To Dissect Adaptive Traits in Global Sorghum Diversity Using a Nested Association Mapping Population. *Genetics* **206**(2): 573-585.

**Boukail S, Macharia M, Miculan M, Masoni A, Calamai A, Palchetti E, Dell’Acqua M. 2021.** Genome wide association study of agronomic and seed traits in a world collection of proso millet (*Panicum miliaceum* L.). *BMC Plant Biology* **21**(1): 330.

**Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES, et al. 2011.** Distinct Genetic Architectures for Male and Female Inflorescence Traits of Maize. *PLoS Genetics* **7**(11): e1002383.

**Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. 2009.** The Genetic Architecture of Maize Flowering Time. *Science* **325**(5941): 714-718.

**Cai Z, Liu H, He Q, Pu M, Chen J, Lai J, Li X, Jin W. 2014.** Differential genome evolution and speciation of *Coix lacryma-jobi* L. and *Coix aquatica* Roxb. hybrid guangxi revealed by repetitive sequence analysis and fine karyotyping. *BMC Genomics* **15**(1): 1025.

**Campion B, Glahn RP, Tava A, Perrone D, Doria E, Sparvoli F, Cecotti R, Dani V, Nielsen E. 2013.** Genetic reduction of antinutrients in common bean (*Phaseolus vulgaris* L.) seed, increases nutrients and *in vitro* iron bioavailability without depressing main agronomic traits. *Field Crops Research* **141**: 27-37.

**Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson YS, Girma D, de Castro E, Chanyalew S, Blösch R, Farinelli L, et al. 2014.** Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics* **15**(1): 581.

**Casals J, Pascual L, Cañizares J, Cebolla-Cornejo J, Casañas F, Nuez F. 2012.** Genetic basis of long shelf life and variability into Penjar tomato. *Genetic Resources and Crop Evolution* **59**(2): 219-229.

**Ceasar SA, Baker A, Ignacimuthu S. 2017.** Functional characterization of the PHT1 family transporters of foxtail millet with development of a novel Agrobacterium-mediated transformation procedure. *Sci Rep* **7**(1): 14064.

**Ceasar SA, Ignacimuthu S. 2011.** Agrobacterium-mediated transformation of finger millet (*Eleusine coracana* (L.) Gaertn.) using shoot apex explants. *Plant Cell Rep* **30**(9): 1759-1770.

**Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, et al. 2010.** Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology* **28**(9): 951-956.

**Che P, Anand A, Wu E, Sander JD, Simon MK, Zhu W, Sigmund AL, Zastrow-Hayes G, Miller M, Liu D, et al. 2018.** Developing a flexible, high-efficiency *Agrobacterium*-mediated sorghum transformation system with broad application. *Plant Biotechnol J* **16**(7): 1388-1395.

**Chibarabada TP, Modi AT, Mabhaudhi T. 2017.** Expounding the Value of Grain Legumes in the Semi- and Arid Tropics. *Sustainability* **9**(1): 60.

**Choi JY, Lye ZN, Groen SC, Dai X, Rughani P, Zaaijer S, Harrington ED, Juul S, Purugganan MD. 2020.** Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. *Genome Biology* **21**(1): 21.

**Cullis C, Kunert KJ. 2017.** Unlocking the potential of orphan legumes. *Journal of Experimental Botany* **68**(8): 1895-1903.

**Dawson IK, Powell W, Hendre P, Bančič J, Hickey JM, Kindt R, Hoad S, Hale I, Jamnadass R. 2019.** The role of genetics in mainstreaming the production of new and orphan crops to diversify food systems and support human nutrition. *New Phytol* **224**(1): 37-54.

**Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. 2021.** How the pan-genome is changing crop genomics and improvement. *Genome Biology* **22**(1): 3.

**Diaz S, Ariza-Suarez D, Ramdeen R, Aparicio J, Arunachalam N, Hernandez C, Diaz H, Ruiz H, Piepho H-P, Raatz B. 2021.** Genetic Architecture and Genomic Prediction of Cooking Time in Common Bean (*Phaseolus vulgaris* L.). *Frontiers in Plant Science* **11**(2257): 622213.

**Doebley J. 1990.** Molecular evidence and the origin of maize. *Economic Botany* **44 (Supplement 3)**: 6-27.

**Ewansiha SU, Singh BB. 2006.** Relative drought tolerance of important herbaceous legumes and cereals in the moist and semi-arid regions of West Africa. *Journal of Food Agriculture and Environment* **4**(2): 188-190.

**FAO, IFAD, UNICEF, WFP, WHO. 2021.** The State of Food Security and Nutrition in the World 2021: Transforming food systems for food security, improved nutrition and affordable healthy diets for all. Rome, Italy.

**Filho AM, Pirozi MR, Borges JT, Pinheiro Sant'Ana HM, Chaves JB, Coimbra JS. 2017.** Quinoa: Nutritional, functional, and antinutritional aspects. *Crit Rev Food Sci Nutr* **57**(8): 1618-1630.

**Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, et al. 2019.** The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics* **51**(6): 1044-1051.

**Gaut BS, Doebley JF. 1997.** DNA sequence evidence for the segmental allopolyploid origin of maize. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 6809-6814.

**Ghosh G, Ganguly S, Purohit A, Chaudhuri RK, Das S, Chakraborti D. 2017.** Transgenic pigeonpea events expressing Cry1Ac and Cry2Aa exhibit resistance to Helicoverpa armigera. *Plant Cell Rep* **36**(7): 1037-1051.

**Global Panel on Agriculture and Food Systems for Nutrition. 2020.** Future Food Systems: For people, our planet, and prosperity. Global Panel on Agriculture and Food Systems for Nutrition, London, UK.

**Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002.** A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**(5565): 92-100.

**Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L, et al. 2017.** Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications* **8**(1): 2184.

**Guo C, Wang Y, Yang A, He J, Xiao C, Lv S, Han F, Yuan Y, Yuan Y, Dong X, et al. 2020.** The Coix Genome Provides Insights into Panicoideae Evolution and Papery Hull Domestication. *Molecular Plant* **13**(2): 309-320.

**Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, Yang X, Peng Q, Wang Y, Jia L, et al. 2017.** *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nature Communications* **8**(1): 1031.

**Haberer G, Kamal N, Bauer E, Gundlach H, Fischer I, Seidel MA, Spannagl M, Marcon C, Ruban A, Urbany C, et al. 2020.** European maize genomes highlight intraspecies variation in repeat and gene content. *Nature Genetics* **52**(9): 950-957.

**Hane JK, Ming Y, Kamphuis LG, Nelson MN, Garg G, Atkins CA, Bayer PE, Bravo A, Bringans S, Cannon S, et al. 2017.** A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant-microbe interactions and legume evolution. *Plant Biotechnology Journal* **15**(3): 318-330.

**Hatakeyama M, Aluri S, Balachadran MT, Sivarajan SR, Patrignani A, Grüter S, Poveda L, Shimizu-Inatsugi R, Baeten J, Francoijs KJ, et al. 2018.** Multiple hybrid *de novo* genome assembly of finger millet, an orphan allotetraploid crop. *DNA Research* **25**(1): 39-47.

**Heidel AJ, Kiefer C, Coupland G, Rose LE. 2016.** Pinpointing genes underlying annual/perennial transitions with comparative genomics. *BMC Genomics* **17**(1): 921-921.

**Hema R, Vemanna RS, Sreeramulu S, Reddy CP, Senthil-Kumar M, Udayakumar M. 2014.** Stable expression of mtlD gene imparts multiple stress tolerance in finger millet. *Plos One* **9**(6): e99110.

**Hirakawa H, Kaur P, Shirasawa K, Nichols P, Nagano S, Appels R, Erskine W, Isobe SN. 2016.** Draft genome sequence of subterranean clover, a reference for genus *Trifolium*. *Sci Rep* **6**: 30358.

**Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F, Hernandez AG, et al. 2016.** Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize. *The Plant Cell* **28**(11): 2700-2714.

**Hittalmani S, Mahesh HB, Shirke MD, Biradar H, Uday G, Aruna YR, Lohithaswa HC, Mohanrao A. 2017.** Genome and Transcriptome sequence of Finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics* **18**(1): 465.

**Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, et al. 2012.** A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**(7421): 497-501.

**Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, et al. 2010.** Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**(11): 961-U976.

**Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012.** Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**(7): 808-811.

**Hufnagel B, Marques A, Soriano A, Marquès L, Divol F, Doumas P, Sallet E, Mancinotti D, Carrere S, Marande W, et al. 2020.** High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nature Communications* **11**(1): 492.

**Hung H-Y, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, et al. 2012.** *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences* **109**(28): E1913-E1921.

**The International Brachypodium Initiative. 2010.** Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**(7282): 763-768.

**Itoh H, Izawa T. 2013.** The coincidence of critical day length recognition for florigen gene expression and floral transition under long-day conditions in rice. *Molecular Plant* **6**(3): 635-649.

**Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, Shah N, Singh VK, Garg R, Jeena G, et al. 2013.** A draft genome sequence of the pulse crop chickpea (*Cicer arietinum*L.). *The Plant Journal* **74**(5): 715-729.

**Jaiswal V, Gupta S, Gahlaut V, Muthamilarasan M, Bandyopadhyay T, Ramchiary N, Prasad M. 2019.** Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (*Setaria italica* L.) Using ddRAD Sequencing. *Sci Rep* **9**(1): 5020.

**Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, et al. 2017.** The genome of *Chenopodium quinoa*. *Nature* **542**(7641): 307-312.

**Jegadeesan S, Raizada A, Dhanasekar P, Suprasanna P. 2021.** Draft genome sequence of the pulse crop blackgram [*Vigna mungo* (L.) Hepper] reveals potential R-genes. *Sci Rep* **11**(1): 11247.

**Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, et al. 2013.** *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**(7443): 91-95.

**Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. 2013.** Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res* **41**(20): e188.

**Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, et al. 2017.** Improved maize reference genome with single-molecule technologies. *Nature* **546**(7659): 524-527.

**Joyce SA, Kamil A, Fleige L, Gahan CGM. 2019.** The Cholesterol-Lowering Effect of Oats and Oat Beta Glucan: Modes of Action and Potential Role of Bile Acids and the Microbiome. *Front Nutr* **6**: 171.

**Kamenya SN, Mikwa EO, Song B, Odeny DA. 2021.** Genetics and breeding for climate change in Orphan crops. *Theoretical and Applied Genetics* **134**(6): 1787-1815.

**Kang SH, Kim B, Choi BS, Lee HO, Kim NH, Lee SJ, Kim HS, Shin MJ, Kim HW, Nam K, et al. 2020.** Genome Assembly and Annotation of Soft-Shelled Adlay (*Coix lacryma-jobi* Variety ma-yuen), a Cereal and Medicinal Crop in the Poaceae Family. *Frontiers in Plant Science* **11**: 630.

**Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha BK, Jun TH, Hwang WJ, Lee T, Lee J, et al. 2014.** Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun* **5**: 5443.

**Kang YJ, Satyawan D, Shim S, Lee T, Lee J, Hwang WJ, Kim SK, Lestari P, Laosatit K, Kim KH, et al. 2015.** Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci Rep* **5**: 8069.

**Kaur P, Bayer PE, Milec Z, Vrána J, Yuan Y, Appels R, Edwards D, Batley J, Nichols P, Erskine W, et al. 2017.** An advanced reference genome of *Trifolium subterraneum* L. reveals genes related to agronomic performance. *Plant Biotechnology Journal* **15**(8): 1034-1046.

**Kawashima CG, Guimarães GA, Nogueira SR, MacLean D, Cook DR, Steuernagel B, Baek J, Bouyioukos C, Melo Bdo V, Tristão G, et al. 2016.** A pigeonpea gene confers resistance to Asian soybean rust in soybean. *Nature Biotechnology* **34**(6): 661-665.

**Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-Villegas J, Guarino L, Jarvis A, Rieseberg LH, Struik PC. 2014.** Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences of the United States of America* **111**(11): 4001-4006.

**Khoury CK, Brush S, Costich DE, Curry HA, de Haan S, Engels JMM, Guarino L, Hoban S, Mercer KL, Miller AJ, et al. 2022.** Crop genetic erosion: understanding and responding to loss of crop diversity. *New Phytologist* **233**(1): 84-118.

**Kidane YG, Gesesse CA, Hailemariam BN, Desta EA, Mengistu DK, Fadda C, Pè ME, Dell'Acqua M. 2019.** A large nested association mapping population for breeding and quantitative trait locus mapping in Ethiopian durum wheat. *Plant Biotechnology Journal* **17**(7): 1380-1393.

**Kiefer C, Severing E, Karl R, Bergonzi S, Koch M, Tresch A, Coupland G. 2017.** Divergence of annual and perennial species in the Brassicaceae and the contribution of cis-acting variation at FLC orthologues. *Molecular Ecology* **26**(13): 3437-3457.

**Kumar R, Janila P, Vishwakarma MK, Khan AW, Manohar SS, Gangurde SS, Variath MT, Shasidhar Y, Pandey MK, Varshney RK. 2020.** Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. *Plant Biotechnology Journal* **18**(4): 992-1003.

**Kuo CG, Chen HM, Sun HC 1992.** Membrane Thermostability and Heat Tolerance of Vegetable Leaves. In: Kuo CG ed. *Adaptation of Food Crops to Temperature and Water Stress;*. Tainan, Taiwan: AVRDC - The World Vegetable Center,160-168.

**Li C, Song W, Luo Y, Gao S, Zhang R, Shi Z, Wang X, Wang R, Wang F, Wang J, et al. 2019.** The HuangZaoSi Maize Genome Provides Insights into Genomic Variation and Improvement History of Maize. *Molecular Plant* **12**(3): 402-409.

**Li C, Sun B, Li Y, Liu C, Wu X, Zhang D, Shi Y, Song Y, Buckler ES, Zhang Z, et al. 2016.** Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations. *BMC Genomics* **17**(1): 894.

**Li W, Li K, Zhang Q-j, Zhu T, Zhang Y, Shi C, Liu Y-l, Xia E-h, Jiang J-j, Shi C, et al. 2020.** Improved hybrid de novo genome assembly and annotation of African wild rice, *Oryza longistaminata*, from Illumina and PacBio sequencing reads. *The Plant Genome* **13**(1): e20001.

**Li X, Yadav R, Siddique KHM. 2020.** Neglected and Underutilized Crop Species: The Key to Improving Dietary Diversity and Fighting Hunger and Malnutrition in Asia and the Pacific. *Frontiers in nutrition* **7**: 593711-593711.

**Li Y, Leveau A, Zhao Q, Feng Q, Lu H, Miao J, Xue Z, Martin AC, Wegel E, Wang J, et al. 2021.** Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat Commun* **12**(1): 2563.

**Liu H, Shi J, Cai Z, Huang Y, Lv M, Du H, Gao Q, Zuo Y, Dong Z, Huang W, et al. 2020a.** Evolution and Domestication Footprints Uncovered from the Genomes of *Coix*. *Molecular Plant* **13**(2): 295-308.

**Liu HJ, Yan J. 2019.** Crop genome-wide association study: a harvest of biological relevance. *The Plant Journal* **97**(1): 8-18.

**Liu J, Seetharam AS, Chougule K, Ou S, Swentowsky KW, Gent JI, Llaca V, Woodhouse MR, Manchanda N, Presting GG, et al. 2020b.** Gapless assembly of maize chromosomes using long-read technologies. *Genome Biology* **21**(1): 121.

**Lo S, Muñoz-Amatriaín M, Hokin SA, Cisse N, Roberts PA, Farmer AD, Xu S, Close TJ. 2019.** A genome-wide association and meta-analysis reveal regions associated with seed size in cowpea [*Vigna unguiculata* (L.) Walp]. *Theoretical and Applied Genetics* **132**(11): 3079-3087.

**Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo MC, et al. 2019.** The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *The Plant Journal* **98**(5): 767-782.

**Lu Q, Li H, Hong Y, Zhang G, Wen S, Li X, Zhou G, Li S, Liu H, Liu H, et al. 2018.** Genome Sequencing and Analysis of the Peanut B-Genome Progenitor (*Arachis ipaensis*). *Frontiers in Plant Science* **9**: 604.

**Maass BL, Robotham O, Chapman MA. 2017.** Evidence for two domestication events of hyacinth bean (*Lablab purpureus* (L.) Sweet): a comparative analysis of population genetic data. *Genetic Resources and Crop Evolution* **64**(6): 1221-1230.

**Mabhaudhi T, Chimonyo VGP, Chibarabada TP, Modi AT. 2017.** Developing a Roadmap for Improving Neglected and Underutilized Crops: A Case Study of South Africa. *Frontiers in Plant Science* **8**(2143): 02143.

**Mamidi S, Healey A, Huang P, Grimwood J, Jenkins J, Barry K, Sreedasyam A, Shu S, Lovell JT, Feldman M, et al. 2020.** A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. *Nature Biotechnology* **38**(10): 1203-1210.

**Martins PK, Nakayama TJ, Ribeiro AP, Cunha B, Nepomuceno AL, Harmon FG, Kobayashi AK, Molinari HBC. 2015a.** Setaria viridis floral-dip: A simple and rapid Agrobacterium-mediated transformation method. *Biotechnol Rep (Amst)* **6**: 61-63.

**Martins PK, Ribeiro AP, Cunha B, Kobayashi AK, Molinari HBC. 2015b.** A simple and highly efficient Agrobacterium-mediated transformation protocol for Setaria viridis. *Biotechnol Rep (Amst)* **6**: 41-44.

**Massawe F, Mayes S, Cheng A. 2016.** Crop Diversity: An Unexploited Treasure Trove for Food Security. *Trends in Plant Science* **21**(5): 365-368.

**Massawe FJ, Mayes S, Cheng A, Chai HH, Cleasby P, Symonds R, Ho WK, Siise A, Wong QN, Kendabie P, et al. 2015.** The Potential for Underutilised Crops to Improve Food Security in the Face of Climate Change. *Procedia Environmental Sciences* **29**: 140-141.

**Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G. J, Buckler E, Doebley J. 2002.** A single domestication for maize is shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 6080-6084.

**Maughan PJ, Lee R, Walstead R, Vickerstaff RJ, Fogarty MC, Brouwer CR, Reid RR, Jay JJ, Bekele WA, Jackson EW, et al. 2019.** Genomic insights from the first chromosome-scale assemblies of oat (Avena spp.) diploid species. *BMC Biol* **17**(1): 92.

**Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K. 2015.** Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* **16**(1): 290.

**Mayes S, Ho WK, Chai HH, Gao X, Kundy AC, Mateva KI, Zahrulakmal M, Hahiree MKIM, Kendabie P, Licea LCS, et al. 2019.** Bambara groundnut: an exemplar underutilised legume for resilience under climate change. *Planta* **250**(3): 803-820.

**McClintock B. 1950.** The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* **36**(6): 344-355.

**McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, et al. 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.

**McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al. 2009.** Genetic Properties of the Maize Nested Association Mapping Population. *Science* **325**(5941): 737-740.

**Mi Y, Zhu Z, Qian G, Li Y, Meng X, Xue J, Chen Q, Sun W, Shi Y. 2020.** Inducing Hairy Roots by Agrobacterium rhizogenes-Mediated Transformation in Tartary Buckwheat (*Fagopyrum tataricum*). *J Vis Exp*(157): e60828.

**Muchow RC. 1985.** Phenology, seed yield and water use of grain legumes grown under different soil water regimes in a semi-arid tropical environment. *Field Crops Research* **11**: 81-97.

**Nguyen DQ, Van Eck J, Eamens AL, Grof CPL. 2020.** Robust and Reproducible *Agrobacterium*-Mediated Transformation System of the C(4) Genetic Model Species Setaria viridis. *Front Plant Sci* **11**: 281.

**Noble TJ, Tao Y, Mace ES, Williams B, Jordan DR, Douglas CA, Mundree SG. 2018.** Characterization of Linkage Disequilibrium and Population Structure in a Mungbean Diversity Panel. *Frontiers in Plant Science* **8**(2102): 02102.

**Ott A, Schnable JC, Yeh CT, Wu L, Liu C, Hu HC, Dalgard CL, Sarkar S, Schnable PS. 2018.** Linked read technology for assembling large complex and polyploid genomes. *BMC Genomics* **19**(1): 651.

**Parween S, Nawaz K, Roy R, Pole AK, Venkata Suresh B, Misra G, Jain M, Yadav G, Parida SK, Tyagi AK, et al. 2015.** An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci Rep* **5**: 12806.

**Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. 2009.** The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**(7229): 551-556.

**Pham A-T, Maurer A, Pillen K, Brien C, Dowling K, Berger B, Eglinton JK, March TJ. 2019.** Genome-wide association of barley plant growth under drought stress using a nested association mapping population. *BMC Plant Biology* **19**(1): 134.

**Reddy VRP, Das S, Dikshit HK, Mishra GP, Aski M, Meena SK, Singh A, Pandey R, Singh MP, Tripathi K, et al. 2020.** Genome-Wide Association Analysis for Phosphorus Use Efficiency Traits in Mungbean (*Vigna radiata* L. Wilczek) Using Genotyping by Sequencing Approach. *Frontiers in Plant Science* **11**(1546): 537766.

**Robotham O, Chapman MA. 2015.** Population genetic analysis of hyacinth bean (*Lablab purpureus* (L.) Sweet, Leguminosae) indicates an East African origin and variation in drought tolerance. *Genetic Resources and Crop Evolution* **64**: 139-148.

**Roorkiwal M, Bharadwaj C, Barmukh R, Dixit GP, Thudi M, Gaur PM, Chaturvedi SK, Fikre A, Hamwieh A, Kumar S, et al. 2020.** Integrating genomics for chickpea improvement: achievements and opportunities. *Theoretical and Applied Genetics* **133**(5): 1703-1720.

**Rychel S, Książkiewicz M. 2019.** Development of gene-based molecular markers tagging low alkaloid pauper locus in white lupin (*Lupinus albus* L.). *Journal of Applied Genetics* **60**(3): 269-281.

**Samtiya M, Aluko RE, Dhewa T. 2020.** Plant food anti-nutritional factors and their reduction strategies: an overview. *Food Production, Processing and Nutrition* **2**(1): 6.

**Sandhu KS, You FM, Conner RL, Balasubramanian PM, Hou A. 2018.** Genetic analysis and QTL mapping of the seed hardness trait in a black common bean (*Phaseolus vulgaris*) recombinant inbred line (RIL) population. *Molecular Breeding* **38**(3): 34.

**Santos CM, Romeiro D, Silva JP, Basso MF, Molinari HBC, Centeno DC. 2020.** An improved protocol for efficient transformation and regeneration of *Setaria italica*. *Plant Cell Rep* **39**(4): 501-510.

**Saxena RK, Kale SM, Kumar V, Parupali S, Joshi S, Singh V, Garg V, Das RR, Sharma M, Yamini KN, et al. 2017a.** Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. *Sci Rep* **7**(1): 1813.

**Saxena RK, Singh VK, Kale SM, Tathineni R, Parupalli S, Kumar V, Garg V, Das RR, Sharma M, Yamini KN, et al. 2017b.** Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. *Sci Rep* **7**(1): 1911.

**Schnable JC. 2015.** Genome Evolution in Maize: From Genomes Back to Genes. *Annual Review of Plant Biology* **66**(1): 329-343.

**Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009.** The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**(5956): 1112-1115.

**Sennhenn A, Odhiambo JJO, Maass BL, Whitbread AM. 2017.** Considering effects of temperature and photoperiod on growth and development of *Lablab purpureus* (L.) Sweet in the search of short-season accessions for smallholder farming systems. *Experimental Agriculture* **53**(3): 375-395.

**Senthil G, Williamson B, Dinkins RD, Ramsay G. 2004.** An efficient transformation system for chickpea (*Cicer arietinum* L.). *Plant Cell Rep* **23**(5): 297-303.

**Shi J, Ma X, Zhang J, Zhou Y, Liu M, Huang L, Sun S, Zhang X, Gao X, Zhan W, et al. 2019.** Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat Commun* **10**(1): 464.

**Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B. 2020.** High Contiguity de novo Genome Sequence Assembly of Trifoliate Yam (*Dioscorea dumetorum*) Using Long Read Sequencing. *Genes* **11**(3): 274.

**Siddique KHM, Li X, Gruber K. 2021.** Rediscovering Asia's forgotten crops to fight chronic and hidden hunger. *Nature Plants* **7**(2): 116-122.

**Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, et al. 2012.** The first draft of the pigeonpea genome sequence. *J Plant Biochem Biotechnol* **21**(1): 98-112.

**Sparvoli F, Cominelli E. 2015.** Seed Biofortification and Phytic Acid Reduction: A Conflict of Interest for the Plant? *Plants* **4**(4): 728-755.

**Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F, Barad O, Barbazuk WB, Bass HW, Baruch K, Ben-Zvi G, et al. 2018.** The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics* **50**(9): 1282-1288.

**Suárez-Estrella D, Torri L, Pagani MA, Marti A. 2018.** Quinoa bitterness: causes and solutions for improving product acceptability. *J Sci Food Agric* **98**(11): 4033-4041.

**Sugihara Y, Darkwa K, Yaegashi H, Natsume S, Shimizu M, Abe A, Hirabuchi A, Ito K, Oikawa K, Tamiru-Oli M, et al. 2020.** Genome analyses reveal the hybrid origin of the staple crop white Guinea yam (*Dioscorea rotundata*). *Proceedings of the National Academy of Sciences* **117**(50): 31987-31992.

**Sun J, Ma D, Tang L, Zhao M, Zhang G, Wang W, Song J, Li X, Liu Z, Zhang W, et al. 2019.** Population Genomic Analysis and De Novo Assembly Reveal the Origin of Weedy Rice as an Evolutionary Game. *Molecular Plant* **12**(5): 632-647.

**Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, et al. 2018.** Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature Genetics* **50**(9): 1289-1295.

**Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu M, Yoshida K, Uemura A, Oikawa K, Abe A, et al. 2017.** Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biology* **15**(1): 86.

**Thielen PM, Pendleton AL, Player RA, Bowden KV, Lawton TJ, Wisecaver JH. 2020.** Reference Genome for the Highly Transformable *Setaria viridis* ME034V. *G3 (Bethesda)* **10**(10): 3467-3478.

**Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES, IV. 2001.** *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* **28**(pdf): 286-289.

**Upadhyaya HD, Vetriventhan M, Deshpande SP, Sivasubramani S, Wallace JG, Buckler ES, Hash CT, Ramu P. 2015.** Population Genetics and Structure of a Global Foxtail Millet Germplasm Collection*. Plant Genome* **8**(3):eplantgenome2015.07.0054.

**VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, et al. 2020.** Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat Commun* **11**(1): 884.

**Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, et al. 2012.** Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* **30**(1): 83-89.

**Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla-Garcia N, Farmer AD, et al. 2010.** Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Molecular Breeding* **26**(3): 393-408.

**Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, et al. 2017a.** Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics* **49**(7): 1082-1088.

**Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A, et al. 2017b.** Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology* **35**(10): 969-976.

**Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, et al. 2013.** Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* **31**(3): 240-246.

**Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, Bajaj P, Cubry P, Rathore A, Jian J, et al. 2019.** Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics* **51**(5): 857-864.

**Wang P, Zhou G, Jian J, Yang H, Renshaw D, Aubert MK, Clements J, He T, Sweetingham M, Li C. 2021a.** Whole-genome assembly and resequencing reveal genomic imprint and key genes of rapid domestication in narrow-leafed lupin. *The Plant Journal* **105**(5): 1192-1210.

**Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, et al. 2014.** Cassava genome from a wild ancestor to cultivated varieties. *Nature Communications* **5**(1): 5110.

**Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, et al. 2018.** Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**(7703): 43-49.

**Wang X, Chen S, Ma X, Yssel AEJ, Chaluvadi SR, Johnson MS, Gangashetty P, Hamidou F, Sanogo MD, Zwaenepoel A, et al. 2021b.** Genome sequence and genetic diversity analysis of an under-domesticated orphan crop, white fonio (*Digitaria exilis*). *GigaScience* **10**(3): 1–12.

**Weiss T, Wang C, Kang X, Zhao H, Elena Gamo M, Starker CG, Crisp PA, Zhou P, Springer NM, Voytas DF, et al. 2020.** Optimization of multiplexed CRISPR/Cas9 system for highly efficient genome editing in *Setaria viridis*. *Plant J* **104**(3): 828-838.

**Weller JL, Liew LC, Hecht VFG, Rajandran V, Laurie RE, Ridge S, Wenden B, Vander Schoor JK, Jaminon O, Blassiau C, et al. 2012.** A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences* **109**(51): 21158-21163.

**Xing A, Gao Y, Ye L, Zhang W, Cai L, Ching A, Llaca V, Johnson B, Liu L, Yang X, et al. 2015.** A rare SNP mutation in Brachytic2 moderately reduces plant height and increases yield potential in maize. *Journal of Experimental Botany* **66**(13): 3791-3802.

**Xu W, Wu D, Yang T, Sun C, Wang Z, Han B, Wu S, Yu A, Chapman MA, Muraguri S, et al. 2021.** Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. *Genome Biology* **22**(1): 113.

**Xu W, Zhang Q, Yuan W, Xu F, Muhammad Aslam M, Miao R, Li Y, Wang Q, Li X, Zhang X, et al. 2020.** The genome evolution and low-phosphorus adaptation in white lupin. *Nat Commun* **11**(1): 1069.

**Yang B. 2020.** Grand Challenges in Genome Editing in Plants. *Frontiers in Genome Editing* **2:** 2.

**Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, et al. 2017.** Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants* **3**(9): 696-703.

**Yang K, Tian Z, Chen C, Luo L, Zhao B, Wang Z, Yu L, Li Y, Sun Y, Li W, et al. 2015.** Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch and low fat accumulation and domestication. *Proc Natl Acad Sci U S A* **112**(43): 13213-13218.

**Yang N, Liu J, Gao Q, Gui S, Chen L, Yang L, Huang J, Deng T, Luo J, He L, et al. 2019.** Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nature Genetics* **51**(6): 1052-1059.

**Yasui Y, Hirakawa H, Oikawa T, Toyoshima M, Matsuzaki C, Ueno M, Mizuno N, Nagatoshi Y, Imamura T, Miyago M, et al. 2016.** Draft genome sequence of an inbred line of *Chenopodium quinoa*, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Research* **23**(6): 535-546.

**Yasui Y, Hirakawa H, Ueno M, Matsui K, Katsube-Tanaka T, Yang SJ, Aii J, Sato S, Mori M. 2016.** Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. *DNA Research* **23**(3): 215-224.

**Ye C, Fan L. 2020.** Orphan crops and their wild relatives in the genomic era. *Molecular Plant* **14**: 27-39.

**Ye CY, Wu D, Mao L, Jia L, Qiu J, Lao S, Chen M, Jiang B, Tang W, Peng Q, et al. 2020.** The Genomes of the Allohexaploid *Echinochloa crus-galli* and Its Progenitors Provide Insights into Polyploidization-Driven Adaptation. *Molecular Plant* **13**(9): 1298-1310.

**Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002.** A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**(5565): 79-92.

**Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, et al. 2012.** Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnology* **30**(6): 549-554.

**Zhang K, He M, Fan Y, Zhao H, Gao B, Yang K, Li F, Tang Y, Gao Q, Lin T, et al. 2021.** Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. *Genome Biology* **22**(1): 23.

**Zhang L, Li X, Ma B, Gao Q, Du H, Han Y, Li Y, Cao Y, Qi M, Zhu Y, et al. 2017.** The Tartary Buckwheat Genome Provides Insights into Rutin Biosynthesis and Abiotic Stress Tolerance. *Molecular Plant* **10**(9): 1224-1237.

**Zhang L, Zhu M, Ren L, Li A, Chen G, Hu Z. 2018a.** The *SlFSR* gene controls fruit shelf-life in tomato. *Journal of Experimental Botany* **69**(12): 2897-2909.

**Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X, et al. 2018b.** Genome-Wide Association Studies of 11 Agronomic Traits in Cassava (*Manihot esculenta* Crantz). *Frontiers in Plant Science* **9**: 503.

**Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciais P, et al. 2017.** Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences of the United States of America* **114**(35): 9326-9331.

**Zhao J, Bayer PE, Ruperao P, Saxena RK, Khan AW, Golicz AA, Nguyen HT, Batley J, Edwards D, Varshney RK. 2020.** Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal* **18**(9): 1946-1954.

**Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, et al. 2018.** Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics* **50**(2): 278-284.

**Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, et al. 2000.** *Agrobacterium*-mediated sorghum transformation. *Plant Mol Biol* **44**(6): 789-798.

**Zhou Y, Chebotarov D, Kudrna D, Llaca V, Lee S, Rajasekar S, Mohammed N, Al-Bader N, Sobel-Sorenson C, Parakkal P, et al. 2020.** A platinum standard pan-genome resource that represents the population structure of Asian rice. *Scientific Data* **7**(1): 113.

**Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W-C, Zhang L, Zhang X, Tang R, et al. 2019.** The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics* **51**(5): 865-876.

**Zou C, Chen A, Xiao L, Muller HM, Ache P, Haberer G, Zhang M, Jia W, Deng P, Huang R, et al. 2017.** A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Research* **27**(11): 1327-1340.

**Zou C, Li L, Miki D, Li D, Tang Q, Xiao L, Rajput S, Deng P, Peng L, Jia W, et al. 2019.** The genome of broomcorn millet. *Nature Communications* **10**(1): 436.

**Table S1. List of important traits dissected by GWAS in rice**

**Table S2. List of important traits dissected by GWAS in maize**

**Table 1 – Underutilised crops for which reference genomes are available. Crops are organized into families (subfamilies) and the genome statistics are given.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Species** | **Common name** | **Chromosome No.** | **Estimated genome size** | **Sequencing method1** | **Scaffold N50** | **Number of contigs** | **Assembled scaffolds size** | **References** |
| Poaceae  (Pooideae) | *Secale cereale* | Rye | 2n=2x=14 | 7.92 Gb | H2000 | 9.4 Kb | 1,581,707 | 2.80 Gb | Bauer et al. 2017 |
| *Aegilops tauschii* | Wheat (D-genome) | 2n=2x=14 | 4.36 Gb | H2000 | 58.0 Kb | 179,145 | 4.23 Gb | Jia et al. 2013 |
| Poaceae  (Chloridoideae) | *Eleusine coracana* | Finger millet | 2n=4x=36 | 1.45 Gb | H4000, N500 | 23.7 Kb | 525,759 | 1.20 Gb | Hittalmani et al. 2017 |
| 1.5 Gb | NS500, PB | 905.3 Kb | 2,812,919 | 1.20 Gb | Hatakeyama et al. 2018 |
|  | *Eragrostis tef* | Tef | 2n=4x=40 | 772 Mb | H2000, 454 | 85.0 Kb |  | 672 Mb | Cannarozzi et al. 2014 |
| 622 Mb | PB, H4000, Hi-C | 1.6 Mb | 1344 | 576 Mb | VanBuren et al. 2020 |
| Poaceae  (Panicoideae) | *Setaria italica* | Foxtail millet | 2n=2x=18 | 510 Mb | GAII | 12.3 Mb |  | 397 Mb | Bennetzen et al. 2012 |
| 490 Mb | GAII, H2000 | 1.0 Mb | 16,903 | 423 Mb | Zhang et al. 2012 |
|  | *Setaria viridis* | Green millet | 2n=2x=18 | 500 Mb | PB, H2000 | 11.2 Mb | 75 | 395 Mb | Mamidi et al. 2020 |
| 401 Mb | ONT, NS500 | 19.5 Mb | 44 | 397 Mb | Thielen et al. 2020 |
|  | *Cenchrus americanus* | Pearl millet | 2n=2x=14 | 1.79Gb | H2000, PB | 885.0 Kb | 175,708 | 1.79 Gb | Varshney et al. 2017b |
|  | *Panicum miliaceum* | Broomcorn/  Proso millet | 2n=4x=36 | 887 Mb | PB, X-ten, BN, Hi-C | 8.24 Mb | 1308 | 848 Mb | Shi et al. 2019 |
| 1 Gb | X-ten | 89 Kb | 171,982 | 823 Mb | Ott et al. 2018 |
| 923 Mb | PB, H2500 | 46.7 Mb | 5,541 | 855 Mb | Zou et al. 2019 |
|  | *Echinochloa crus-galli* | Barnyard millet | 2n=6×=54 | 1.38 Gb | H2000, PB | 1.8Mb |  | 1.27 Gb | Guo et al. 2017 |
| PB, H4000 | 4.1 Mb |  | 1.34 Gb | Ye et al. 2020 |
|  | *Digitaria exilis* | Fonio millet | 2n=4×=36 | 893Mb | H2500, No6000, Hi-C, BN | 10.7 Mb | 29,155 | 716 Mb | Abrouk et al. 2020 |
| PB, H3000, H4000 | 1.73 Mb | 3329 | 761Mb | Wang et al. 2021b |
|  | *Coix lacryma-jobi* | Adlay/Job's tears | 2n=2×=20 | 1.68 Gb | H2000 |  |  |  | Cai et al. 2014 |
| PB, H2500 | 2.24 Mb | 4519 | 1.62 Gb | Guo et al. 2020 |
| PB, H2500 | 594.3 Kb | 13,691 | 1.28 Gb | Kang et al. 2020 |
| PB, H2000, BN, Hi-C | 14.0 Mb | 3321 | 1.73 Gb | Liu et al. 2020a |
|  | *Sorghum bicolor* | Sorghum | 2n=2x=20 | 730Mb | BAC | 35 Mb | 12,873 | 679 Mb | Paterson et al. 2009 |
| BAC, H2500 | 68.7 Mb | 2688 | 655 Mb | McCormick et al. 2018 |
| Fabaceae | *Cajanus cajan* | Pigeonpea | 2n=2x=22 | 833 Mb | BAC, GAII, H2000 | 516.1 Kb | 173,708 | 606 Mb | Varshney et al. 2012 |
| BAC, 454 | 14.0 Kb | 332,766 | 511 Mb | Singh et al. 2012 |
|  | *Cicer arietinum* | Chickpea | 2n=2x=16 | 738 Mb | BAC, GAII, 454 | 77.3 Kb |  | 520 Mb | Jain et al. 2013 |
| BAC, H2000, BAC | 40.0 Mb | 62,619 | 532 Mb | Varshney et al. 2013 |
| 454, GAII | 39.9 Mb | 182,734 | 510 Mb | Parween et al. 2015 |
|  | *Vigna radiata* | Mungbean/ Green gram | 2n=2×=22 | 543Mb | H2000, 454 | 1.52 Mb | 180,372 | 431 Mb | Kang et al. 2014 |
|  | *Vigna. angularis* | Adzuki bean | 2n=2×=22 | 542 Mb | H2000 | 1.3 Mb | 25,426 | 467 Mb | Yang et al. 2015 |
| 612 Mb | H2000, 454 | 703 kb | 36,516 | 443 Mb | Kang et al. 2015 |
|  | *Vigna unguiculata* | Cowpea | 2n=2×=22 | 640 Mb | H2000, BAC, PB | 16.4 Mb | 765 | 519 Mb | Lonardi et al. 2019 |
|  | *Lupins albus* | White lupin | 2n=2x=50 | 584 Mb | PB, H2000, Hi-C | 18.7 Mb | 3171 | 474 Mb | Xu et al. 2020 |
| PB, H3000, BN | 17.4 Mb |  | 451 Mb | Hufnagel et al. 2020 |
|  | *Lupinus angustifolius* | Narrow leafed lupin | 2n=2x=40 |  | H2500 | 703.2 KB | 14,379 | 609 Mb | Hane et al. 2017 |
|  |  |  |  | PB, H2000 | 30.8 Mb | 123 | 616 Mb | Wang et al. 2021a |
|  | *Vigna mungo* | Blackgram | 2n=2x=22 | 574 Mb | X-ten and ONT | 1.4 Mb |  | 475 Mb | Jegadeesan et al. 2021 |
|  | *Trifolium subterraneum* | Subterranean clover | 2n=2x=16 | 540 Mb | H2000, M, 454 | 287.6 Kb | 101,010 | 472 Mb | Hirakawa et al. 2016 |
| BN | 1.4 Mb | 264 | 512 Mb | Kaur et al. 2017 |
| Polygonaceae | *Fagopyrum esculentum* | Common buckwheat | 2n=2x=16 | 1.2 Gb | H2000 | 25.1 Mb |  | 1.18 Gb | Yasui et al. 2016b |
|  | *Fagopyrum tataricum* | Tartary buckwheat | 2n=2x=16 | 540 Mb | H2000, H2500, M, PB, GBS | 550.7 Kb | 8778 | 489.3 Mb | Zhang et al. 2017 |
| Amaranthaceae | *Chenopodium quinoa* | Quinoa | 2n=4x=36 | 1.5Gb | H2500, PB | 86.9 Kb | 110,092 | 1.1 Gb | Yasui et al. 2016a |
| PB, BN | 3.8 Mb | 4232 | 1.18 Gb | Jarvis et al. 2017 |
| H2500, PB | 1.2 Mb | 10,795 | 1.3 Gb | Zou et al. 2017 |
| Euphorbiaceae | *Manihot esculentum* | Cassava | 2n=2x=36 | 746 Mb | H2000, 454, BAC | 67 Kb |  | 432 Mb | Wang et al. 2014 |
| *Ricinus communis* | Castor bean | 2n=2x=20 | 320 Mb | Plasmid, Sanger | 561.4 Kb | 25,800 | 325 Mb | Chan et al. 2010 |
| Convolvulaceae | *Ipomoea batatas* | Sweet potato | 2n=6x=90 | 4.4 Gb | H2500, N500, H4000, 454 | 200.7 Kb | 35,919 | 836 Mb | Yang et al. 2017 |
| Dioscoreaceae | *Dioscorea dumetorum* | Trifoliate yam | 2n=2x=36 | 322 Mb | H1500, ONT | 3.2 Mb | 1172 | 485 Mb | Siadjeu et al. 2020 |
| *Dioscorea rotundata* | White Guinea yam | 2n=2x=40 | 579 Mb | H2500, BAC | 2.1 Mb | 4723 | 594 Mb | Tamiru et al. 2017 |

1 Sequencing approaches are abbreviated as follows: 454, Roche 454; BAC, Bacterial Artificial Chromosome sequencing; BN, BioNano; GAII, Illumina Genome Analyzer II; H, Illumina Hiseq; Hi-C, high-throughput chromosome conformation capture; M, Illumina Miseq; N500, NextSeq 500; No6000, Novaseq6000; ONT, Oxford Nanopore Technologies; PB, PacBio; X-ten, Illumina X-ten.

**Figure 1 – Schematic diagram highlighting the main sequencing and breeding approaches discussed in this article.** Firstly (left), by comparing the reference genomes of underutilised crops and staple/main crops, genome variants responsible for superior agronomic traits (such as specific resistance and nutrient quality or quantity) of underutilised crops could be explored. In addition (right), the combination of genome resequencing and phenotyping through genome wide association study could help identifying candidate genes responsible for agronomic traits of underutilised crops. Through genetic transformation or genome editing, the function of these candidate genes could be verified. Finally, by associating genomic data to phenotypic information of different accessions, germplasm resources can be effectively screened and bred by means of molecular marker-assisted breeding and crossing, to improve the resistance and nutritional value of underutilised crops in addition to staple crops.

**Figure 2 – Schematic diagram highlighting common breeding targets for any crop, and whether the genetic basis has been investigated in underutilised crops (UCs) and/or staple crops.** ANF, Anti-nutrient factor.

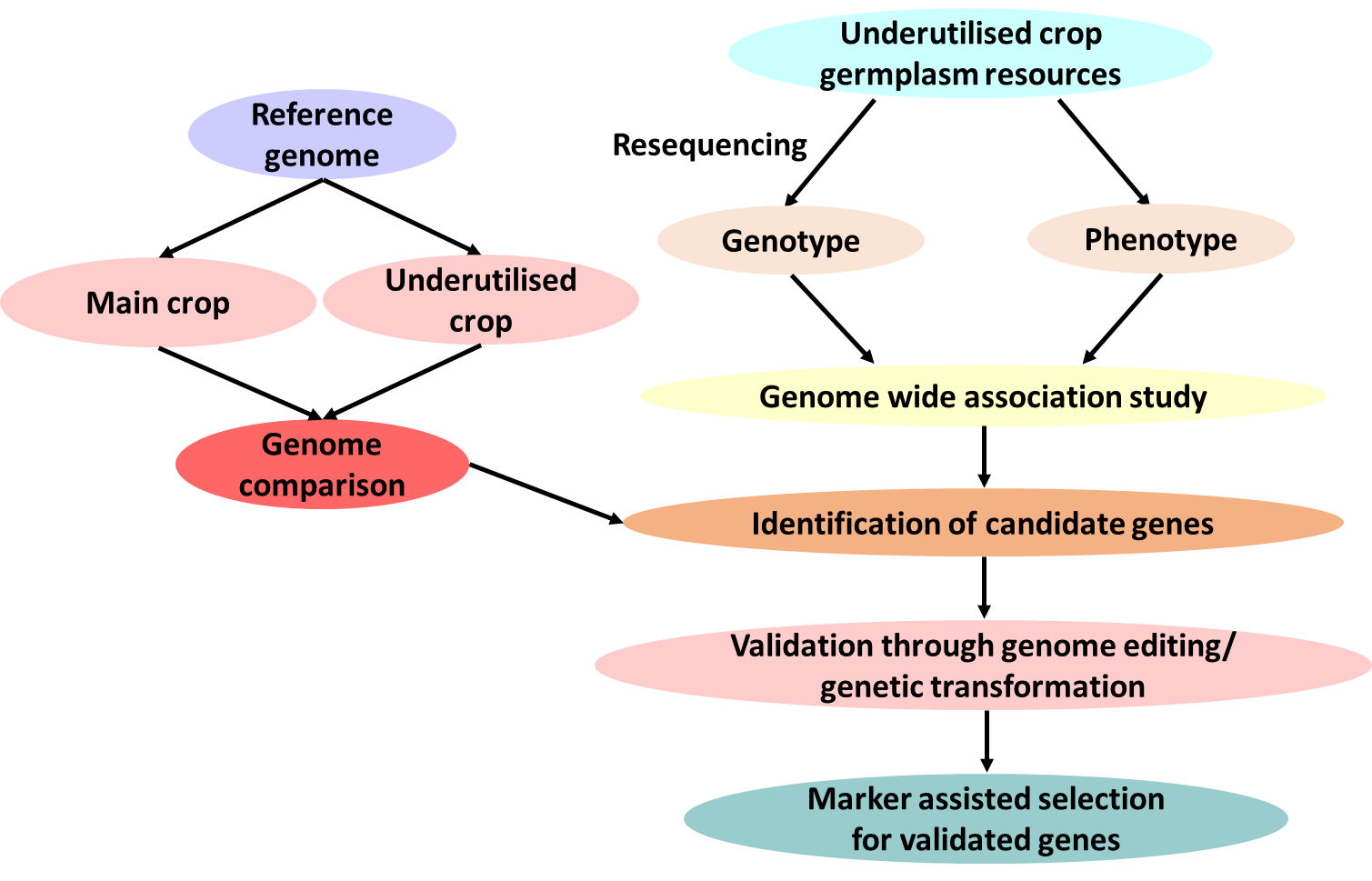
Fig 1

Fig 2

