Genome-wide Association Studies of Agronomic Traits Consisting of Field- and Molecular-based Phenotypes

Yuki Nakano1 and Yuriko Kobayashi1*

1 The United Graduate School of Agricultural Science, Gifu University, 1-1 Yanagido, 501-1193, Gifu, Japan.

ABSTRACT

Genome-wide association study (GWAS) is a powerful approach to identify the genetic factors underlying the intraspecific phenotypic variations. Recent advances in DNA sequencing technology, including next generation sequencing has enabled us to easily genotype high density genome-wide SNPs. In addition, many accessions of various plant species have been widely collected in recent years. These genetic resources have made GWAS a markedly more popular approach for investigation of natural variations occurring in various traits using large populations. In addition to genotyping technology, advances in high-throughput phenotyping technologies have enabled us to acquire variation data on a large number of accessions characterized for various traits, including not only the field traits (e.g., yield and disease resistance) but also molecular traits (e.g., gene expression level and metabolite content). Thus, it is possible to expand the range of application of GWAS and enhance the detection power of genomic association. In this review, we summarize recent GWAS of various agronomic traits at field and molecular scale, following which we highlight the integration approach involving GWAS and high-throughput phenotyping technologies including transcriptome, ionome and metabolome.

Keywords
crop breeding, genome breeding, genome-wide association study, high-throughput phenotyping, natural variation, omics analysis

1. Introduction

Sustainable food production is important for enhancing human health and food security. During the Green Revolution, crop yields were markedly improved through the introduction of dwarf genotypes of rice and wheat, which contributed toward preventing the occurrence of food crisis in the developing countries (Khush, 2001). However, global crop production faces various threats, such as climate change and population growth (Wheeler and von Braun, 2013; Lesk et al., 2016; Deutsch et al., 2018). Although various solutions have been proposed for sustainable crop production, enhancing plant robustness against environmental stress is one of the important factors that can enable stable food production. In addition, in consideration of the implications for human health, it is globally desirable to enrich the edible parts of the crops with beneficial phytonutrients, while decreasing the toxic component during food production. For example, increasing the content of essential minerals, such as iron (Fe) and zinc (Zn) and reducing the content of toxic elements, such as cadmium (Cd) and arsenic (As) in the crops are important targets for crop production (White and Broadley, 2005; Zhao and Shewry, 2011). Various rice breeding methods, including the transgenic approach, have been employed to increase the Fe/Zn content and to reduce the Cd content in the rice grains (Ishikawa et al., 2012; Slamat-Loedin et al., 2015). For example, ‘Koshihikari Kan No. 1’ has been developed as a new cultivar of low-Cd rice in Japan, which also has reduced As content (Ishikawa et al., 2016). Thus, crop improvement surely plays a key role in ensuring global food security.
Interspecies accessions of plants show phenotypic variation based on their genotypes in various traits. Hence, the study of genetic variation in the natural accessions adapting to the habitats is a powerful and effective approach to investigate the genetic architecture of complex traits using the wide genetic diversity compared to traditional approaches (e.g., mutant screening). Subsequently, natural allelic variations associated with the target traits can be directly exploited in breeding in general. One popular approach is quantitative trait loci (QTLs) mapping using biparental population, such as recombinant inbred lines (RIL), which is a method to identify the genomic region determining phenotypic differences between the parents (Borevitz and Chory, 2004). Several studies have used QTL mapping of various traits to successfully identify the key genes contributing to the variations in model plants and crops (Takeda and Matsuoka, 2008; Roy et al., 2011; Blümel et al., 2015). These data indicated that the wild and cultivated accessions of plant genetic resources are powerful resources that can be instrumental in identifying the genetic factors determining intraspecific variations in various agronomic traits, such as yield, stress tolerance, and metabolite content, which may provide resources and information for plant breeding. Most agricultural traits are complex traits controlled by polygenes, indicating that several molecular mechanisms are involved in the phenotypic variation among the accessions. To fully elucidate the mechanisms responsible for the genetic diversity of resources, large number of accessions and crop varieties have been collected by plant resource centers for various species in the world and have been released to the research community (Jackson, 1997; Garcia-Hernandez et al., 2002; Massawe et al., 2016; Portwood et al., 2019).

Recently, the cost of single nucleotide polymorphism (SNP) typing and DNA sequencing has markedly decreased due to advances in next generation sequencing (NGS), which has enabled us to obtain genetic information, such as whole genome sequences and genome-wide SNPs, across many accessions to be collected for each project involving various plant species (rice: The 3000 rice genomes project, 2014; Arabidopsis: The 1001 Genomes Consortium, 2016; maize: Portwood et al., 2019; soybean: Song et al., 2013, 2015; tomato: Bauchet et al., 2017). Consequently, genome-wide association study (GWAS), using a large number of accessions with genome-wide SNPs, has emerged as an alternative approach for bi-parental QTL mapping in various plants. GWAS is a population statistics approach to detect the genomic region involved in target trait based on association analysis between the phenotypic variation and genome-wide SNPs in accessions with higher resolution as compared to the biparental populations (Rafalski, 2010; Zhang et al., 2010). GWAS can identify the genetic factors determining the phenotypic variations from the genetic pool of multiple accessions, indicating that multiple genetic factors relating to various molecular mechanisms involved in phenotype variations can be identified.

In contrast to the genetic mapping using bi-parental progeny, there is no need to develop mapping population for GWAS because it detects associated genomic region using the historic recombination events among the accessions. Furthermore, once the accessions are genotyped, they can be used for GWAS for various traits without further genotyping, leading to accelerated dissection of genetic architecture of various agronomic traits. While advances in sequencing technology to generate genome-wide SNP markers has made it easy to implement GWAS, phenotyping large number of accessions is still a bottleneck for GWAS in plants. Increasing the population size generally improves the detection power of GWAS (Korte and Farlow, 2013). However, measuring agronomical traits in a large mapping population is time consuming and labor intensive, especially the cultivation traits in field experiments. Therefore, measuring the traits of large number of accessions with high accuracy and less effort is important for large scale phenotyping through high-throughput technologies. Further, advances in high-throughput phenotyping technologies for molecular traits (e.g., metabolite and mineral content) have enabled comprehensive
Table 1: Representative report of GWAS for agronomic traits

<table>
<thead>
<tr>
<th>Traits</th>
<th>Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering time</td>
<td>Arabidopsis</td>
<td>Shindo et al., (2005); Balasubramanian et al., (2006); Atwell et al., (2010); Wollenberg and Amasino, (2012); Li et al., (2014); Sasaki et al., (2015); Zan and Carlborg, (2018)</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>Yang et al., (2013); Bouchet et al., (2013)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Huang et al., (2012)</td>
</tr>
<tr>
<td></td>
<td>Barley</td>
<td>Muñoz-Amatriain et al., (2014)</td>
</tr>
<tr>
<td></td>
<td>Soybean</td>
<td>Zhang et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>Rapeseed</td>
<td>Xu et al., 2015</td>
</tr>
<tr>
<td></td>
<td>Common bean</td>
<td>Moghaddam et al., (2016)</td>
</tr>
<tr>
<td>Morphological phenotype</td>
<td>Arabidopsis</td>
<td>Kobayashi et al., (2016)</td>
</tr>
<tr>
<td></td>
<td>Root growth under salt condition</td>
<td>Satbhai et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Famoso et al., (2011)</td>
</tr>
<tr>
<td></td>
<td>Root growth under zinc deficiency condition</td>
<td>Bouain et al., (2018)</td>
</tr>
<tr>
<td></td>
<td>Root growth under water stress condition</td>
<td>Ayalew et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>RSA under control condition</td>
<td>Pace et al., (2014); Pace et al., (2015); Sanchez et al., (2018)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Biscarini et al., (2016)</td>
</tr>
<tr>
<td></td>
<td>Cowpea</td>
<td>Burridge et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis</td>
<td>Slovak et al., (2014)</td>
</tr>
<tr>
<td></td>
<td>RSA under salt condition</td>
<td>Julkowska et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>RSA under drought condition</td>
<td>Li et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Leaf shape</td>
<td>Yang et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>Yield assorted traits</td>
<td>Yang et al., (2014)</td>
</tr>
<tr>
<td>Metabolite contents</td>
<td>Arabidopsis</td>
<td>Verslues et al., (2014)</td>
</tr>
<tr>
<td></td>
<td>Glucosinolates contents</td>
<td>Brachi et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>Compounds relating quality of tomato</td>
<td>Sauvage et al., 2014</td>
</tr>
<tr>
<td></td>
<td>Metabolite content in fruit</td>
<td>Ye et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Metabolite content in leave</td>
<td>Chen et al., (2014)</td>
</tr>
<tr>
<td></td>
<td>Metabolite content in grain</td>
<td>Chen et al., (2016)</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis</td>
<td>Chao et al., (2012)</td>
</tr>
<tr>
<td></td>
<td>Aegilops tauschii</td>
<td>Qin et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>Barley</td>
<td>Wu et al., (2015)</td>
</tr>
<tr>
<td></td>
<td>Rapeseed</td>
<td>Chen et al., (2018)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Zhao et al., (2018a)</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>Zhao et al., (2018b)</td>
</tr>
<tr>
<td>Mineral contents</td>
<td>Arabidopsis</td>
<td>Chao et al., (2014)</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>Zhao et al., (2018c)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Frouin et al., (2019)</td>
</tr>
<tr>
<td></td>
<td>Chickpea</td>
<td>Upadhyaya et al., (2016)</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>Hindu et al., (2018)</td>
</tr>
</tbody>
</table>
Table 1: (Continued)

<table>
<thead>
<tr>
<th>Traits</th>
<th>Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mineral contents</td>
<td>Chickpea</td>
<td>Ozkuru et al., (2019)</td>
</tr>
<tr>
<td>Mineral content</td>
<td>Barley</td>
<td>Gyawali et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Foxtail millet</td>
<td>Jaiswal et al., (2019)</td>
</tr>
<tr>
<td></td>
<td>Rice</td>
<td>Yang et al., (2018)</td>
</tr>
<tr>
<td></td>
<td>Soybean</td>
<td>Ziegler et al., (2018)</td>
</tr>
<tr>
<td>Mineral and metabolite content</td>
<td>Wheat</td>
<td>Alomari et al., (2017); Kumar et al., (2018); Alomari et al., (2019); Cu et al., (2020).</td>
</tr>
<tr>
<td>Transcription level</td>
<td>Sorghum</td>
<td>Melo et al., (2018)</td>
</tr>
<tr>
<td></td>
<td>Arabidopsis</td>
<td>Sadhukhan et al., (2017)</td>
</tr>
<tr>
<td></td>
<td>Maize</td>
<td>Li et al., (2013)</td>
</tr>
</tbody>
</table>

profiling of various quantitative molecular phenotypes simultaneously. Integrative approach of GWAS and the high-throughput techniques could lead to bridge the gap between the agronomic traits regulated by complex mechanisms and their responsive genes.

Thus, GWAS in plants has been applied for the study of various complex traits at molecular level and in the field (Table 1). These studies identified associated loci, and in some cases, causal genes determining variation of traits among the accessions in various plant species. In this review, we summarize the application of GWAS to various agronomic traits at molecular level and in the fields in various plants. In addition, we highlight the works that GWAS for agronomic traits has generated through high-throughput phenotyping technologies for molecular phenotypes

2. GWAS for morphological phenotypes

In the beginning of a GWAS, visible agricultural phenotypes of traditional agronomic traits that are associated with high yield and resistance to environmental stress are analyzed. The flowering time in response to photoperiod and stress is an important breeding trait as a determinant of yield of crop cultivation; its molecular mechanism and responsible genes have been studied using traditional mutant analysis in Arabidopsis (Jung and Müller, 2009). In addition, the natural variation and the genetic architecture of flowering time adaptations to environments have been well studied. Candidate gene-based association analysis revealed that major genes controlling the flowering time, FLOWERING LOCUS C (FLC), FRIGIDA (FRI), VERNALIZATION-INSENSITIVE 3 (VIN3) and CRYPTOCHROME 2 (CRY2), are associated with natural variation (Shindo et al., 2005; Balasubramanian et al., 2006; Wollenberg and Amasino, 2012). After developing a number of accessions with the SNP information, the genome-wide association of genes regulating flowering time was identified by GWAS (Atwell et al., 2010; Li et al., 2014; Sasaki et al., 2015; The 1001 Genomes Consortium, 2016; Zan and Carlborg, 2018). GWAS on flowering time have been conducted in various crops other than Arabidopsis, leading to the discovery of new loci associated with flowering time (maize: Bouchet et al., 2013, rice: Huang et al., 2012, barley: Muñoz-Amatriain et al., 2014, soybean: Zhang et al., 2015b, rapeseed: Xu et al., 2015, common bean: Moghaddam et al., 2016). The polymorphism of these genes associated with flowering time may have driven the spread of plants into a different photoperiod environment.

On the contrary, root growth is an underground phenotype that has been used as an index for tolerance to soil stresses, such as mineral deficiency and excess water stress inhibiting root growth. Tolerance to mineral stress was
evaluated by measuring the root length under control and stress conditions. Kobayashi et al. (2016) evaluated the salt tolerance of Arabidopsis accessions under varying stress intensities using relative root length (RRL: ratio of the primary root length in the presence of salt to that in the absence of salt), and demonstrated by GWAS that distinct loci were involved in the salt tolerance variations between mild to severe stress conditions. Similarly, GWAS using root length phenotype was performed for various other stress conditions, such as iron deficiency (Arabidopsis: Satbhai et al., 2017), aluminum tolerance (rice: Famoso et al., 2011), zinc deficiency (Arabidopsis: Bouain et al., 2018), and water stress (wheat: Ayalew et al., 2015). However, rather than simple root length, a more complex root system architecture (RSA), which comprises of primary root and lateral roots, is an important trait for adaptation against various environments (Gruber et al., 2013). Julkowska et al., (2017) acquired 17 RSA phenotypes including lateral root density and root angle for 347 Arabidopsis accessions under salt stress conditions. They found that the reported salt tolerance gene \textit{HIGH AFFINITY K$^+$ TRANSPORTER1} and salt responsive gene \textit{CYTOCHROME P450 FAMILY 79 SUBFAMILY B2} were involved in the variation of RSA under salt condition by GWAS using both the RSA phenotypes and principle components (PCs) derived from the multiple RSA phenotypes. Especially in crops, RSA is known as an important trait involved in increasing the productivity not only under normal but also under stress conditions (de Dorlodot et al., 2007). Many loci associated with RAS that were estimated by several root traits under normal conditions were identified in various crops by GWAS (Pace et al., 2015; Biscarini et al., 2016; Burridge et al., 2017). Li et al., (2017) identified loci associated with RAS under drought stress in rice; however, there are only a few reports of GWAS on RAS under stress conditions in crops.

Figure 1: The example of analysis of root system architecture (RSA) by EZ-Rhizo software. (A) Root detection by EZ-Rhizo software from the original root photo of Arabidopsis. Roots are automatically detected from image, and primary root and lateral root of each plant are distinguished by the software as shown in different color. (B) Parameters of constructing RSA. The software automatically measures various parameters constructing RSA, such as length and number of main root as shown in orange and lateral root as shown in red.
Instead of traditional phenotyping technology, image analysis is a key technology for largescale phenotyping in large number of accessions with high accuracy and high throughput in GWAS. It was developed as root analysis tools for efficient evaluation of RSA including 3D root scan imaging (WinRHIZO: Pro 2014; EZ-Rhizo: Armengaud et al., 2009; SmartRoot: Lobet et al., 2011; ARIA: Pace et al., 2014) (Figure 1). Slovak et al., (2014) developed a pipeline using a photo scanner for evaluation of the root phenotypes of Arabidopsis grown on agar plates (BRAT; Busch-lab Root Analysis Toolchain). This pipeline could be used for non-destructive phenotyping of various root phenotypes (e.g., total root length and root angle); the authors acquired 16 root phenotypes for 163 accessions during a time course of 5 days using this pipeline and identified several loci that were significantly associated with each phenotype by GWAS. In maize seedlings, GWAS for RSA evaluated by ARIA software, which can extract 27 different root traits in one analysis including 2D and 3D traits, identified reported QTLs and new loci controlling root development including root network area (Pace et al., 2014; Sanchez et al., 2018). In addition to root traits, leaf traits involved in photosynthesis, yield, and stress resistance were also analyzed by high-throughput leaf phenotyping using scan camera, which was applied for GWAS. It was found that a large number of new loci associated with 29 leaf traits expressed during different growth stages in rice (Yang et al., 2015). Furthermore, GWAS for various traditional agronomic traits, including green leaf area and tiller number in rice, which were analyzed as image traits by color-imaging device and linear X-ray computed tomography, enabled detection of loci associated with yield assorted traits (Yang et al., 2014) and stress tolerance (Guo et al., 2018).

3. GWAS for metabolite content

Plant metabolites play an important role in not only plant growth and stress tolerance (Pastori et al., 2003; Ye et al., 2017), but also as beneficial nutrients for human health (De Luca et al., 2012). Even though there have not been many reports of metabolic-GWAS (mGWAS) as compared to phenotypic-GWAS (pGWAS), certain studies involving non-targeted mGWAS and targeted mGWAS in Arabidopsis and crops have been reported during the last decade. Verslues et al., (2014) measured the proline accumulation, which is a compatible solute, in 180 Arabidopsis accessions under low water potential conditions. They identified a series of genes underlying the variations in proline accumulation using combination approach involving GWAS and reverse genetics. Additionally, Brachi et al., (2015) profiled the content of 22 methionine-derived glucosinolates (GSLs), which are involved in the defense against herbivory and pathogens, of 595 Arabidopsis accessions, and identified three loci that could explain approximately 40% of the variation in the GSL profile among the accessions by GWAS. These studies revealed the relationship between the genes regulating metabolite content and stress tolerance, at least in part. In addition to stress tolerance, metabolite content affects the fruit quality of the crops. GWAS for compounds related to the quality of tomato, such as sugar and amino acid content, and metabolites including GABA, malate, citrate, and ascorbate, found 44 significant associations and enabled the identification of useful genetic variants (Sauvage et al., 2014). Later, Al-Activated Malate Transporter 9 (SIALMT9) was identified as a causal gene determining natural variation of fruit malate content, which influences the flavor and palatability of fruits among the tomato accessions by GWAS (Ye et al., 2017). In addition, they demonstrated that the polymorphism of SIALMT9 can contribute to not only malate content, but also Al tolerance in tomato by Al-chelating ability.

Metabolomics approach using LC-MS/ GC-MS has made it possible to profile large number of metabolites simultaneously, and mGWAS based on a non- and widely-targeted profiling can reveal the genetic architecture for natural variation in secondary metabolism (Matsuda et al., 2015), while revealed unexpected relationships between metabolite-gene and metabolite-phenotype as compared to the mGWAS targeting specific metabolite. Chen et al., (2014) profiled 840 metabolites in the leaves of more than 500 rice accessions. They found a series of loci associated
with each metabolite, although some metabolites were controlled by major single loci providing an explanation for over 50% variation observed for each metabolite. Furthermore, they successfully identified the function of the genes for major associated loci by transgenically increasing the content of target metabolites. Moreover, although it is suggested that metabolite content may influence the plant phenotype, there are relatively few reports that have demonstrated the association between the metabolite and phenotype. Parallel analysis of mGWAS for grain metabolites and pGWAS for six grain-related phenotypes (e.g., grain color and grain width) identified 17 loci that were co-detected in both GWAS (Chen et al., 2016). Among them, six loci associated with trigonelline content that was also associated with grain width/grain thickness. Validation studies on one of the genes from these associated loci showed over-accumulation of trigonelline caused wider grains with cell expansion in the overexpressing transgenic plant. These results bridged the gap between the metabolite-phenotype involved in certain biological processes, and could improve breeding strategies by genetic and biochemical regulation and selection.

4. GWAS for mineral content

Similar to the metabolites, the mineral content of plants is an important trait not only for plant growth and stress tolerance but also for human health. Plants require both macro nutrients (e.g., nitrogen (N), potassium (K), and phosphorus (P)) and micronutrients (e.g., copper (Cu) and Zn) for adequate growth. In modern agriculture practices, large amounts of fertilizers, in particular macro nutrients, are used to increase the yield, raising concerns regarding their disturbance on the environment due to outflow of the fertilizers (Withers and Lord, 2002). Therefore, improving the nutrient uptake and use-efficiency in crops is important to reduce the fertilizer input, which can mitigate the impact on environment and save the ingredients of fertilizer (e.g., phosphorus ore). Moreover, plants uptake and accumulate minerals that are beneficial (e.g., Zn and Fe) and deleterious (e.g., Cd and As) to human health, and these plants form a major source of these minerals for human intake (White and Broadley, 2009; Uraguchi and Fujiwara, 2013). Thus, it is important to understand the mechanisms regulating the mineral content, that is, the mineral homeostasis in plants for compatibility of sustainable agriculture and enhancing the benefits for human health.

Heavy metal content in plant tissue is involved not only in stress tolerance but also human health because plant-derived food is a major source of human intake of toxic metals. To identify the genes determining heavy metal content in crops, GWAS for especially Cd and As content was conducted. Molecular mechanism of regulation of Cd has been identified and exploited for plant modification, especially in rice which faces severe problems of Cd in soil in the rice fields; many QTL and mutant studies found accumulation of various transporters that regulate Cd tolerance and accumulation in rice grain (Gao et al., 2016). Cd uptake transporter NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 5 (OsNRAMP5) and Cd translocator LOW AFFINITY CATION TRANSPORTER 1 (OsLCT1) in rice were mutated for low-Cd rice breeding using ion beam mutagenesis (Ishikawa et al., 2016) and genome editing (Tang et al., 2017; Songmei et al., 2019). However, GWAS further identified loci, including new genes associated with Cd accumulation and tolerance (Arabidopsis: Chao et al., 2012; Aegilops tauschii: Qin et al., 2015; barley: Wu et al., 2015; rapeseed: Chen et al., 2018; rice: Zhao et al., 2018a, maize: Zhao et al., 2018b). Similarly, GWAS for As accumulation was conducted, by which some associated genes were identified that were functionally distinct from the previously characterized genes (Chao et al., 2014; Zhao et al., 2018c; Frouin et al., 2019). The information of loci revealed by GWAS can be exploited for improving the efficiency of breeding of not only significantly low accumulation in the edible parts but also for enhancing the tolerance for toxic minerals, in addition to using major responsible genes and the orthologue genes in various plants.

Plant-derived food provides essential micronutrients for human health. An estimated two billion people worldwide suffer from Zn and Fe deficiency, which is the leading health risk. In addition, other minerals deficiencies,
including that of calcium (Ca) and trace elements are worldwide health problems (White and Broadley, 2009; Stein, 2010). Because of the importance of Zn and Fe for human health, recently, GWAS of Zn and Fe contents in edible parts of the major cereals and crops was actively conducted to identify the loci controlling the mineral content (chickpea: Upadhyaya et al., 2016; barley: Gyawali et al., 2017; maize: Hindu et al., 2018; rice: Yang et al., 2018; wheat: Kumar et al., 2018, Alomari et al., 2019, Cu et al., 2020; foxtail millet: Jaiswal et al., 2019). Moreover, many loci associated with Ca concentration (Alomari et al., 2017), Cu and Manganese (Mg) concentration in wheat grains (Cu et al., 2020), and Se and molybdenum (Mo) in chickpea (Ozkuru et al., 2019) were identified by GWAS. Interestingly, Kumar et al., (2018) and Cu et al., (2020) reported loci controlling multiple grain micronutrients, phosphorus content, and yield-related traits.

Similar to metabolomic approach, advances in the ionome technology, such as ICP-MS, has enabled us to profile the content of multiple minerals simultaneously. Yang et al., (2018) profiled 17 mineral nutrients, such as macro (e.g., N, P, and K), micro (e.g., Mo and Zn), and nonessential nutrients (e.g., As and Cd) for 529 rice accessions. Next, they identified several candidate genes associated with the natural variation of each mineral content by GWAS, that included previously reported mineral transporters involving the mineral uptake, such as HKT1;5 for K content and MOLYDATE TRANSPORTER 1;1 (MOT1;1) for Mo content. Furthermore, they demonstrated that the ion profile is influenced by the supply of N and P fertilizer. In soybean population, ionomic GWAS for the concentration 19 elements and weights of the seed identified 21 SNPs associated with candidate and already characterized genes (Ziegler et al., 2018). Additionally, some associated genes were co-detected in several traits. For example, COPPER CHAPERONE ANTIOXIDANT-1 (ATOX1) for Cu, Fe, Zn, and P content, ALUMINUM-ACTIVATED MALATE TRANSPORTER (ALMT) for Al, Mo, K, cobalt, and rubidium content, and each metal transporter were included in the gene list. It is known that there is cross-talk in mineral uptake and accumulation among the ions, such as between Zn-Fe and As-P (Pineau et al., 2012; LeBlanc et al., 2013). Especially in transport system, molecular mechanism of Cu, Fe, and Zn homeostasis and their interactions, which often include interactions with toxic metals, were relevance in plants (Puig et al., 2007; Bernal et al., 2012; Sinclair and Krämer, 2012; Bashir et al., 2016). These reports suggested that integration of multiple GWAS results targeting various minerals could identify the loci controlling the cross-talk of uptake and accumulation of multiple minerals. This information of loci and causal genes controlling multiple mineral contents could lead to the development of crops that contain higher content of beneficial micronutrients and lower content of toxic metals.

5. GWAS for gene expression

Expression level polymorphism (ELP) among accessions is an important determinant underlying natural variation of various traits, such as morphological phenotype, metabolite content, and mineral content as described above. Although there are relatively limited number of reports compared to other traits, GWAS targeting expression level (eGWAS) has been conducted as an alternative approach for bi-parental QTL analysis to identify the genetic factors underlying ELP.

The genes encoding organic acid (OA) transporter have been identified as important Al tolerance genes in various plants, and it has been demonstrated that the ELP of OA transporter caused by cis-element polymorphism is associated with the variation of Al tolerance among the accessions (Hoekenga et al., 2006; Sasaki et al., 2006; Magalhaes et al., 2007; Chen et al., 2013; Yokosho et al., 2016). The higher expression of the OA transporter gene has notably contributed toward high Al tolerance in various crops in the ELP. However, it was reported that introgression of the tolerant allele of Sorghum bicolor MULTIDRUG AND TOXIC COMPOUND EXTRUSION (SbMATE), which is an important citrate transporter gene in Al tolerance of sorghum, into the sensitive accession
not always resulted in enhanced expression level of \( SbMATE \) and elevated Al tolerance (Melo et al., 2013). This result suggested that trans-factors may play a role in enhancing the expression level of OA transporter in addition to the cis-factors. In fact, SbWRKY1 and SbZNF1 are the trans-factors involved in variation of \( SbMATE \) expression levels by eGWAS and bi-parental eQTL mapping, in addition to previously reported cis-element polymorphism (Melo et al., 2019). Contrarily, parallel approach of pGWAS and eGWAS of the pGWAS-detected genes was used to demonstrate the direct link between the phenotype and gene based on co-detection of same locus in both GWAS (Figure 2), which explained the phenotype variation at the gene expression level (Li et al., 2013; Sadhukhan et al., 2017). eGWAS has high potential to identify the direct and indirect regulating factors involved in the expression of target genes. This suggests that integration of multiple eGWAS results targeting key genes would provide important information to understand the gene regulatory networks for the molecular mechanism controlling the trait.

Recent decrease in sequencing costs has allowed the acquisition of large transcriptome data from large number of accessions by RNA-seq. Li et al., (2013) profiled kernel oil-related traits (i.e., oil concentration and fatty acid compositions) of maize accessions, and identified several candidate genes associated with the oil-related traits by GWAS. In addition, the eGWAS of the candidate genes using transcriptome data of 368 maize accessions acquired by RNA-seq identified their own locus as eQTL, suggesting that the genes affect the phenotypic variation of the oil-related traits via ELP. This result demonstrated that transcriptome data of large number of accessions, which is followed by the integration of pGWAS and eGWAS, is useful to dissect the mechanisms responsible for the generation of trait variation through transcriptional regulation.

Figure 2: Parallel approach of pGWAS and eGWAS of the pGWAS-detected gene. Example of Manhattan plots for pGWAS of an agronomic trait and eGWAS of the pGWAS-detected gene are shown with SNPs indicated by each dot in different color at each chromosome. Solid rectangles indicate same gene identified by pGWAS and eGWAS. The direct link between expression level polymorphism (ELP) of the pGWAS-detected gene and phenotype variation can be estimated by co-detection of same locus in pGWAS and eGWAS. The gene regulates the variation of agronomic trait via the expression difference.
6. Conclusion and perspective

Crop breeding has contributed immensely to food production to date and is becoming an important tool for increasing world food security. Traditional crop breeding based on crossing and selection requires long time to establish new cultivars. To overcome the issue, molecular breeding based on functional genomics information (e.g., genomic editing, marker assist selection; MAS) is expected to increase the efficiency and speed of breeding. In combination with recent advances in genome sequencing technology, which can be easily used to produce genome-wide markers in a lot of accessions, GWAS has become a popular approach to dissect the complex genetic architecture of agronomic traits. This approach can provide useful genomic information for molecular breeding (Figure 3). GWAS has been applied for various agronomic traits, such as field phenotypes, metabolite content, mineral content, and transcription level, and identified the genetic determinants underlying natural variations of the traits. Furthermore, application of combined approaches using GWAS and high-throughput phenotyping technologies (e.g., metabolome and ionome) has enabled the discovery of unexpected molecular mechanisms that has bridged the gaps between the genome and the traits. These genetic architectures including causal genes for target phenotype would be useful information that can be exploited for molecular breeding in crops targeting various kinds of trait.

High-throughput and accurate phenotyping technologies has allowed us to conduct GWAS using large number of accessions. Recent advances in hyperspectral technology has enables to the non-destructive measuring of the biochemical traits (e.g., water content, protein content, and chlorophyll content) (Yi et al., 2013; Sun et al., 2017; Sun et al., 2019), in addition to morphological traits. Recently, remote sensing using unmanned aerial vehicle (UAV) combined with image analyses and hyperspectral technologies has been highlighted (Condorelli et al., 2018; Cen et al., 2019; Koh et al., 2019). Furthermore, an attempt to apply machine learning technology in plant phenotyping has been made (Ghosal et al., 2018). These new techniques are expected to allow non-destructive phenotyping of multiple agronomic traits for large number of plants planted in large fields with high accuracy and within short time to accelerate the investigation of genetic architecture of agronomic phenotypes.

The large scale GWAS lead to detection of a single locus that showed strongly significant associations as compared to GWAS using limited accessions, and assisted in achieving an excellent germplasm and DNA marker for the key gene, which can be used for the breeding. However, most agronomic traits are quantitative traits controlled by polygenes, in which the effect of each single locus is relatively small. Therefore, it is important to identify a series of loci that can cumulatively explain the variation of trait not only loci collocated with high-ranking SNPs. Identification of these polygenic loci is required for breeding with high performance, as well as loci involved in pleiotropic effects and environmental interactions. Therefore, to estimate the cumulative contribution of GWAS detected loci, combined approach with GWAS and genomic prediction (GP) model have been employed (Kobayashi et al., 2016; Kooke et al., 2016). This approach could detect a series of SNPs including associated loci with weak effect, which can be target-SNP markers for crop breeding by genomic selection. In addition to additive effects, an attempt to estimate epistatic effects was conducted using the GP model (Wang et al., 2012; Maurer et al., 2015). Furthermore, epistasis analysis was conducted in combination with GWAS (Lachowiec et al., 2015; Zhang et al., 2015a; Moellers et al., 2017).

Integration of GWAS results targeted for different traits (parallel GWAS) is a reasonable approach to bridge the gap between the traits, such as gene, metabolite, and phenotype. In addition to this approach, multi-trait GWAS has been implemented to identify the genetic factors underlying the variation in multiple traits (Thoen et al., 2017). This approach has led not only to identify the gene that is commonly involved in multiple traits but also in gaining an insight into the molecular processes underlying gene to phenotype expression. Similarly, integration of GWAS with
Figure 3: Overview of a GWAS workflow for crop breeding. GWAS exploits genetic diversity of natural accessions. Large number of accessions and crop varieties of various plant species have been collected by plant resource centers and released for the research community. The genotype information of the accessions and varieties have been determined and accumulated in the database or used in own research. These resources are enabled us to conduct GWAS with high resolution and accuracy. GWAS can identify the loci associated with phenotypic variations of agronomic traits among the accessions, which possibly plays a role in the regulation of the traits and can be useful genetic factor for breeding. The advances in high-throughput phenotyping technologies has enables the comprehensive profiling of various quantitative traits across large mapping populations simultaneously. It could not only enhance the accuracy of GWAS but also reveal the relationship between different traits. GWAS can provides useful information for breeding strategies, such as QTLs, key genes, and excellent germplasm and genetic architecture of the traits. These information may lead to increased efficiency of crop breeding, and accelerate the development of new cultivars for sustainable food production.
gene co-expression network analysis is a powerful approach toward understanding the molecular processes controlling the phenotype (Kobayashi et al., 2016; Li et al., 2018; Schaefer et al., 2018), in which large expression data would be acquired from accessions under GWAS than public database to explain the GWAS phenotype. Recently, public databases for plant phenotype have been established (Arabidopsis: AraPheno [Seren et al., 2017], rice: RiceVarMap [Zhao et al., 2015], Maize: Panzea [Zhao et al., 2006]). These databases collect both phenotypic and molecular traits (e.g., metabolite and mineral contents). These accumulated data is a useful information to conduct multi-trait GWAS or parallel GWAS through combined with own data.

GWAS have been conducted for various traits and have identified a series of genes and molecular mechanisms underlying phenotypic variation. Advanced GWAS has been conducted in mainly model plants and major crops; however, GWAS has also been reported in other plants with considerably less dense SNPs and population, such as apple (Amyotte et al., 2017; Farneti et al., 2017; Lee et al., 2017), beans (Sallam et al., 2016; Zuiderveen et al., 2016), and other vegetables (Nimmakayala et al., 2016; Han et al., 2018; Okada et al., 2019). Such kinds of studies in various plants will be conducted to promote NGS platform and resource collection. In contrast, GWAS has successfully identified the genetic factors that have actually driven phenotypic variation in nature, suggesting that these genetic factors may function when integrated into different lines in crop breeding. The information on these genetic factors are useful for crop breeding based on genome editing and MAS as well as genomic selection (Desta and Ortiz, 2014).

Acknowledgments
This work was supported by JSPS KAKENHI Grant-in-Aid for JSPS Fellows (to YN, Grant Number 18J11757).

REFERENCES


Chen ZC, Yokosho K, Kashino M, Zhao F-J, Yamaji N, Ma JF (2013) Adaptation to acidic soil is achieved by increased numbers of cis-acting elements regulating ALMT1 expression in Holcus lanatus. Plant J 76: 10–23.


