DNA Research, 2020, 27(4), 1–8
doi: 10.1093/dnares/dsaa023
Advance Access Publication Date: 26 September 2020
Invited Review

Invited Review

History and future perspectives of barley genomics

Kazuhiro Sato *

Institute of Plant Science and Resources, Okayama University, Kurashiki 710-0046, Japan

*To whom correspondence should be addressed. Tel. +81 86 434 1249. Email: kazsato@okayama-u.ac.jp

Received 22 August 2020; Editorial decision 10 September 2020; Accepted 12 September 2020

Abstract

Barley (*Hordeum vulgare*), one of the most widely cultivated cereal crops, possesses a large genome of 5.1 Gbp. Through various international collaborations, the genome has recently been sequenced and assembled at the chromosome-scale by exploiting available genetic and genomic resources. Many wild and cultivated barley accessions have been collected and preserved around the world. These accessions are crucial to obtain diverse natural and induced barley variants. The barley bioresource project aims to investigate the diversity of this crop based on purified seed and DNA samples of a large number of collected accessions. The long-term goal of this project is to analyse the genome sequences of major barley accessions worldwide. In view of technical limitations, a strategy has been employed to establish the exome structure of a selected number of accessions and to perform high-quality chromosome-scale assembly of the genomes of several major representative accessions. For the future project, an efficient annotation pipeline is essential for establishing the function of genomes and genes as well as for using this information for sequence-based digital barley breeding. In this article, the author reviews the existing barley resources along with their applications and discuss possible future directions of research in barley genomics.

Key words: *Hordeum vulgare*, genome sequencing, genetic resources

1. Introduction

1.1. Origin of genomic diversity in barley

Before barley (*Hordeum vulgare* ssp. *vulgare*) was domesticated, hunter–gatherers used the ancestral wild form (*H. vulgare* ssp. *spontaneum*) of domesticated barley as a human food source. Both wild and domesticated barley were found in archaeological sites in the Fertile Crescent dating back about 10,000 years,[1] which is believed to be the origin of barley domestication. Wild barley kernels can be distinguished from those of domesticated barley by their brittle and smooth rachis.[2] The ancestral wild form of barley shares the same genome as domesticated barley and is classified as a barley subspecies.[3] Millions of generations of this wild barley provide a source of diversity to the present-day cultivated form, although domestication partially narrowed the diversity. Soon after domestication, mutations responsible for agronomically valuable traits, such as the jump from two to six row spike,[4] spring growth habit,[5] and hull-less caryopsis,[6] were selected for and spread quickly to all cultivated barley within a few thousand years. Allelic diversity at these loci corresponds with ecological conditions and with different uses of the cereal (human food, animal feed, and malt production) and made barley landraces well suited for cultivation throughout the world in most conditions, except in the tropics. Naturally occurring diversity in locally adapted landraces was the only source of available diversity until barley cross-breeding and mutation induction started in the early 20th century.[7]

2. Relationships of plant species within the Poaceae family

Barley ranks fourth among grain cereals (Poaceae species) after maize (*Zea mays*), wheat (*Triticum aestivum*), and rice (*Oryza*

sativa) in terms of global production. Barley is self-pollinating with a diploid genome consisting of seven chromosomes ($2n = 2x = 14$). Barley, common wheat, rye (*Secale cereale*), and their wild relatives (e.g. *Aegilops* spp.) are closely related and are included in the Triticeae tribe, which evolved some 12 million years ago within the Pooideae subfamily of the Poaceae (grasses).[8] The estimated barley genome size is 5.1 Gbp[9] with >80% of repetitive elements. Each subgenome of hexaploid wheat is characterized by the same genetic components as barley in terms of genome size, gene content, and repetitive elements.[10] Although it has a large genome, barley can be considered a good genomic model for cultivated hexaploid wheat due to its simple diploid genome. Large differences in genome sizes exist among Poaceae species. Both *Brachypodium distachyon* (272 Mbp) and rice (430 Mbp) have small genomes and share a common ancestor with the Triticeae tribe, with divergence times of 32–39 and 40–53 million years ago, respectively.[11] Their evolutionary relationships reflect the sequence similarity within Poaceae species and facilitate the identification of orthologous genes of importance across crop genomes. Among Poaceae species, barley has been well studied genetically. Several mutant traits have practical importance and can be used as model traits in cereal crops. In particular, many barley and wheat genes exhibit similar functions: information about a gene in barley can therefore be readily applicable to estimating the genes responsible for similar traits in wheat.

3. Genetic and genomic resources

3.1. Seed collection for barley

Barley is the only cultivated species of the genus *Hordeum*, which includes about 32 species and about 45 taxa.[3] More than 485,000 accessions for the genus *Hordeum* are preserved at more than 200 different institutions worldwide.[12] These collections include 299,165 accessions of *H. vulgare* ssp. *vulgare* (primarily new and old cultivars and landraces), 32,385 accessions of *H. vulgare* ssp. *spontaneum*, 4,681 accessions of wild species, and a substantial representation of genetic stocks, breeding lines, and mapping populations.[13] Many accessions are duplicated between gene banks for safety or to avoid quarantine problems. The world's largest seed storage facility is the Svalvard Seed Vault, which preserves over 1 million crop-related accessions, including 92,075 *Hordeum* accessions as of July 2020 (https://www.nordgen.org/sgsv/).

As an East Asian Center of barley genetic resources, Okayama University maintains 10,980 cultivars and landraces, 2,498 genetic stocks, and 628 wild barleys. Their collection, preservation, and distribution are partly supported by the National Bioresource Project (nbrp.jp). These materials have been collected since the 1940s for crop evolutionary studies at Okayama University. About 5,300 cultivated barley accessions have been intensively characterized both by genomic markers and for some agronomically relevant traits, such as resistance to powdery mildew. The data sets are available online from Barley DB (http://www.shigen.nig.ac.jp/barley/).

These collections are part of the South and East Asian subset of the international barley core collection, which constitutes the entire *ex situ* (stored) barley genetic diversity. Okayama University is responsible for the distribution of the South and East Asian subset, comprising 380 accessions. Other materials can be requested from the United States Department of Agriculture (USDA) small grain collection (Americas), the International Center for Agricultural Research in the Dry Areas (ICARDA, West Asia and North Africa and ssp. *spontaneum*), Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK, Germany), and the Nordic Genetic Resource Center in Sweden (for wild *Hordeum*).

Barley has been the subject of much mutation research and breeding. These mutants have been mainly collected at the USDA small grain collection, located in Aberdeen, Idaho, USA, and at the Nordic Genetic Resource Center, in Alnarp, Sweden.[14] To enhance the utility and accessibility of these mutants, >400 mutant alleles have been introgressed into the cultivar Bowman.[15] Information on the Bowman introgression lines is available in the Barley Genetics Newsletter, Vol. 26 (http://wheat.pw.usda.gov/ggpages/bgn/26/bgn26tc.html). These historical mutants provide a rich resource for functional studies and gene cloning.

Okayama University also maintains mutants, tetraploid lines, linkage testers, and near isogenic lines for barley. These lines were largely developed through the research activities at Okayama University and used for the development of linkage maps and genetic analysis of mutant traits.[15] Tetraploids helped to develop a series of trisomic lines that can be used to identify the chromosome location of a given locus of interest. Currently, three mapping populations[16–18] have been deposited from the North American barley genome mapping project. Another mapping population (Haruna Nijo × H602) was developed at Okayama University.[19] These populations were instrumental in developing molecular genetic maps for barley and are being exploited for high-throughput mapping by single-nucleotide polymorphism (SNP) arrays. Recombinant chromosome substitution lines (backcross introgression lines) are also useful resources to identify genes in specific genomic regions to precisely study and map a locus of interest. Several sets of populations are available for distribution.[20,21]

3.2. Genomic resources

Multiple genomic resources have been developed to analyse partial or total genomic sequences and their associated functions in barley.

Since 2000, several large barley expressed sequence tag (EST) projects have generated large numbers of sequences from expressed genes (http://harvest.ucr.edu/). Eight different genotypes provided the material for these projects. The polymorphisms identified between genotypes contributed to promoting high-throughput SNP genotyping, genetic mapping, and marker generation. The first comprehensive barley full-length cDNA (FL-cDNA) sequences were collected by Sato et al.[22] mRNA samples were isolated from 15 organs and treatments from the Haruna Nijo cultivar and later pooled to develop a FL-cDNA library using the Cap-trapper method.[23] A total of 5,006 clones were sequenced (http://www.shigen.nig.ac.jp/barley/). Another set of about 25,000 clones was generated and sequenced from the same cultivar, from 40 different organs and treatments.[24] These sequences have allowed the annotation of genes on contigs[25] or chromosome-scale genome sequences.[26,27]

Barley geneticists have developed high-quality genetic maps, based on mutant phenotypes, using classical three-point linkage tests.[15] These efforts are being updated and complemented with genome-wide genetic maps generated from molecular markers. Stein et al.[28] developed a consensus barley map with 1,032 EST-based loci assayed using a combination of marker assays. Sato et al.[19] developed a high-resolution barley EST map with 2,890 loci using a single-mapping population. High-throughput and high-quality multiplex PCR-based genotyping assays based on Golden Gate technology (Illumina Inc., CA, USA) involve the allele-specific detection of SNPs. A consensus genetic linkage map containing >2,900 gene-based SNP markers has been developed (http://harvest.ucr.edu/).[29,30]

Several bacterial artificial chromosome (BAC) libraries have been constructed. The first library was developed from the Morex cultivar[31]; it has been joined by other five Morex BAC libraries that collectively cover >25 genome equivalents.[32] A library from the cultivar Haruna Nijo consists of 294,912 clones with an average insert size of 115.2 kbp and a coverage of about 6.6 genome equivalents.[33] Another BAC library from wild barley accession H602 was developed at Okayama University (unpublished). Libraries from Haruna Nijo (coded as HNB) and H602 are available from the National Bioresource Project (nbrp.jp).

4. Genome sequencing

4.1. The nuclear genome

The International Barley Genome Sequencing Consortium (IBSC) was established in 2006 to generate a high-quality barley genome sequence.[34] Two general approaches have been undertaken for the analysis of genome structure in barley: (i) identifying a minimum tiling path of 87,075 genetically mapped BAC clones and sequencing these clones (BAC-by-BAC strategy) and (ii) performing whole-genome shotgun sequencing.

The development of a BAC-based physical map for genome sequencing is a large-scale and worldwide effort. Madishetty et al.[35] developed a high-throughput approach to use overgo probes to identify gene-containing BAC clones. They used >10,000 overgo probes derived from EST sequences to identify 83,381 gene-containing clones from the initial Morex BAC library.[31] Fingerprinting of these clones resulted in contigs that comprise roughly two-thirds of all barley genes. Four new Morex BAC libraries have been prepared that cover an estimated 25 haploid genome equivalents.[34] From these libraries, approximately 550,000 clones (covering about 14 genome equivalents) have been fingerprinted and assembled in contigs.[32] To complement these efforts, BAC-end sequencing will be conducted on 350,000 BACs. In addition, there are efforts to integrate the resulting BAC contigs with the SNP-based genetic maps. Thus, a robust BAC-based physical map was integrated with the genetic map.

Of the 5.10 Gbp of the barley genome, IBSC developed a physical map of 4.98 Gbp, with more than 3.90 Gbp anchored to a high-resolution genetic map.[9] Projecting a deep whole-genome shotgun assembly, EST, FL-cDNA, and newly developed RNA-seq data onto this framework supports 79,379 transcript clusters, including 26,159 high-confidence genes with homology support from other plant genomes.[9] More than 80% of the genome is occupied by repeat sequences.

For chromosome-scale, high-quality assembly, each BAC clone from the minimum tiling path was barcoded and sequenced by Illumina short-read sequencing. Then, a high-resolution genetic map created by population sequencing methodology[36] and a highly contiguous optical map were combined to construct super-scaffolds composed of merged assemblies from individual BACs. Finally, chromosome conformation capture sequencing (Hi-C) was used to order and orient BAC-based super-scaffolds. The final chromosome-scale assembly represents 4.79 Gbp (~95%) of the genome.[37] Mapping of transcriptome data identified 39,734 high-confidence loci and 41,949 low-confidence loci on the basis of sequence homology to related species.

4.2. The organellar genomes

Middleton et al.[38] determined that the chloroplast sequences from cultivated and wild barley were closely related (sequence identity 99.98%). The divergence time of these haplotypes is estimated to be 80,000 ± 20,000 years using semi-penalized likelihood. A comparison of the chloroplast genome from cultivated barley and common wheat identified four insertions and five deletions >50 bp relative to the common wheat chloroplast genome. The extent of chloroplast sequence similarity indicates that cultivated and wild barleys are more closely related to each other than they are to cultivated wheat.

Hisano et al.[39] assembled the complete nucleotide sequences of the mitochondrial genomes from wild and cultivated barley. Two independent circular maps of the 525,599-bp barley mitochondrial genome were constructed by *de novo* assembly of high-throughput sequencing reads from the wild accession H602 and from the cultivar Haruna Nijo. These two maps detected only three SNPs between the two haplotypes. Both mitochondrial genomes contained 33 protein-coding genes, three ribosomal RNAs, 16 transfer RNAs, 188 new ORFs, six major repeat sequences, and several types of transposable elements. Mitochondrial genome sequencing is essential for annotating the barley nuclear genome; indeed, these mitochondrial sequences identified a significant number of fragmented mitochondrial sequences in the reported nuclear genome sequences.[37]

4.3. Transcriptomes

Deep sequencing of the transcriptome (RNA-seq) from the cultivar Morex and FL-cDNAs from the cultivar Haruna Nijo helped to annotate the reference genome of the cultivar Morex.[9,37] The recent development of a single-molecule sequencing technique may also support the sequencing of long transcripts. To this end, an international collaborative project on full transcript sequencing for major haplotypes is underway (Waugh et al., unpublished). These long transcript sequences will also contribute to the annotation of pan-genome assemblies (Stein et al., unpublished). As it is essential to isolate intact mature transcripts (mRNA) to obtain FL-cDNA sequences, techniques are being developed to maintain transcript integrity.

A *de novo* RNA-seq-based genotyping procedure for barley strains used in breeding programs has been implemented.[40] Using RNA samples from several tissues, reads were mapped onto transcribed regions, which correspond to ~590 Mbp out of the ~4.8 Gbp reference genome. Using 150 samples from 108 strains, this approach detected 181,567 SNPs and 45,135 indels, located in 28,939 transcribed regions distributed throughout the Morex genome.[37] The quality of this polymorphism detection method was validated by analysing 387 RNA-seq-derived SNPs by amplicon sequencing. These results demonstrated that this RNA-seq-based *de novo* polymorphism detection system can generate genome-wide markers, even in the closely related barley genotypes used in breeding programs.

4.4. Genomic information and available databases

The barley research community maintains a diverse array of databases that house information pertaining to barley genetics and genomics that can be easily accessed. Table 1 lists such databases and their respective information type.

For genome assembly and annotation, EnsemblPlants provides easy access to the most updated barley genome assembly, including chromosome sequences, genes, transcripts, and predicted proteins. The same website also supports Basic Local Alignment Search Tool (BLAST) searches against the barley genome. Similarly, IPK allows the user to conduct BLAST searches against all sequence resources published by the worldwide barley community, including activities related to the International Barley Sequencing Consortium.

Table 1. Barley genome databases

Name	URL	Function
EnsemblPlants	http://plants.ensembl.org/Hordeum_vulgare	Browser, BLAST
IPK (IBSC) barley BLAST server	https://webblast.ipk-gatersleben.de/barley_ibsc/	BLAST
PLEXdb	http://www.plantgdb.org/prj/PLEXdb/	Gene expression analysis
HarvEST	http://harvest.ucr.edu/	cDNA sequence
barleyGenes	https://ics.hutton.ac.uk/barleyGenes/	RNA-seq data
bex-db	https://barleyflc.dna.affrc.go.jp/bexdb/	cDNA, gene expression
GrainGenes	http://www.graingenes.org	Markers, maps, mutants, etc.
Barley DB	http://www.shigen.nig.ac.jp/barley/	Seed collection, cDNA sequence

Several databases are specifically related to transcripts. PLEXdb is a public resource for gene expression analysis of plants and plant pathogens. This website currently hosts microarray data sets from a range of species, including barley and wheat. HarvEST is principally an EST database viewing platform that emphasizes gene function and is geared towards comparative genomics and oligonucleotide design, in support of activities such as microarray content design, functional annotation, and physical and genetic mapping. The site also allows the display of consolidated maps of approximately 3,000 SNP markers from four barley mapping populations[29] and offers a rice synteny viewer. A subset of these mapped SNPs is integrated with the minimum BAC clone tiling path. barleyGenes provides access to predicted genes from an assembly of whole-genome shotgun sequences from barley (cultivar Morex). These genes were predicted from the mapping of RNA-seq data to the genome assembly. Gene expression levels were also calculated from the RNA-seq data and are available in the form of FPKM values associated with the predicted genes. bex-db provides 5′- and 3′-end sequences for 175,000 FL-cDNA clones as well as their expression data in a searchable database. The database also provides a genome browser, showing the locations of cDNA sequences on the barley genome.[9]

Several databases contain barley genomic resource information. GrainGenes, the database of choice for legacy and classical genetics data,[41] gathers information such as genetic maps, genes, alleles, genetic markers, phenotypic data, quantitative trait loci studies, experimental protocols, and publications about Triticeae species. Barley DB includes information on barley germplasm and barley genome resources from Okayama University. The database contains 5,006 FL-cDNAs and 134,928 EST entries. The barley germplasm collection from Okayama University and associated information are also available on the site.

5. Applied use of the genomic information of barley

5.1. Towards the digital bioresource project of barley

The ultimate goal of barley genomics is to *de novo* sequence all natural and induced germplasm. The collection of seed samples with unchanged sequences can become a digital bioresource.[42,43] However, even the current state of techniques and cost reduction are not sufficient to make sequencing more than a few thousand barley haplotypes possible. After the analysis of sequence data sets, the TRITEX pipeline[27] can assemble chromosome-scale sequences for one haplotype in 3–4 weeks for barley. A parallel sequencing and computational analysis may save time, but it is not feasible to sequence thousands of accessions by high-quality chromosome-scale assembly with the current technical standards. For these reasons, partial sequencing of genomes may give useful preliminary information for digital bioresource development in barley. Here, The author summarizes some of the activities and tools used to estimate diversity in natural and induced barley accessions.

5.2. Sequence accessions to summarize natural sequence variation

The IPK crop gene bank has a major world collection of barley seed samples. To estimate the diversity in genome sequences in wild and cultivated barleys, a genotyping-by-sequencing platform was applied to single plants from 21,405 accessions in the IPK barley collection.[44] All haplotype reads were aligned to the reference genome from the barley cultivar Morex,[37] which allowed the detection of 171,263 bi-allelic SNPs. A principal component (PC) analysis of cultivated barleys indicated that PC2 separated Eastern and Western barleys, whereas PC1 set Ethiopian barleys apart. Figure 1 shows that the geographical origin of collection agreed well with the SNP analysis. Representative accessions based on PC analysis are being used to analyse the barley pan-genome (Stein et al., unpublished), which may present a full complement of sequence variation within the barley genome.

5.3. Exome sequencing

Plant biologists and breeders concentrate most of their efforts on deciphering the functions of genes. Exome sequencing specifically targets coding sequences from the genome, and hence alleviating the need to sequence entire genomes and their repeat content, while bringing the computational needs and cost down to a reasonable range. The barley exome capture system selectively enriches for 61.6 Mbp of coding sequences based on the Morex cultivar HC (high-confidence gene model) sequences, FL-cDNA sequences, and *de novo* assembled RNA-seq consensus sequence contigs by hybridizing the gene-related fragments (or exons) from genomic libraries.[46]

The platform provides a highly specific and targeted capture of barley exons and closely related species. Using this platform, Russel et al.[47] sequenced the exomes of a collection of 267 geo-referenced landraces and wild accessions of barley. A combination of genome-wide analyses demonstrated that the patterns of variation in barley have been strongly shaped by geography and that variant-by-environment associations for individual genes are prominent. A high-density 50 K SNP array with the Illumina Infinium whole-genotyping array was also developed, based on SNP design on exome capture data from 170 lines from a barley diversity panel.[48]

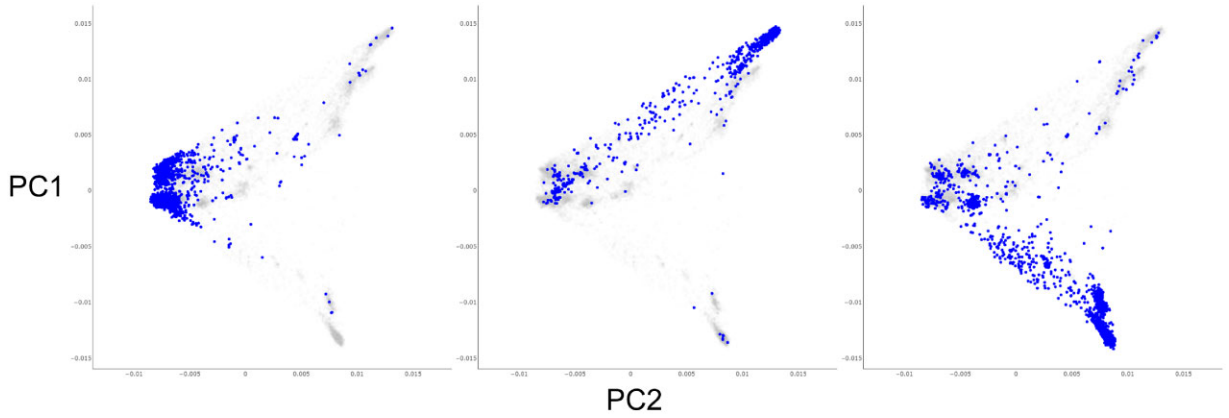

Figure 1. Principal component analysis of 19,778 domesticated barleys based on 76,102 genotyping-by-sequencing markers from the IPK Bridge Web Portal (https://bridge.ipk-gatersleben.de/bridge/).[45] Accessions from Germany (left), Japan (middle), and Ethiopia (right) are plotted on PC1 (*y* axis) and PC2 (*x* axis). The proportion of variance explained by the PCs is indicated on the axes.

5.4. Mapping of mutations for gene annotation

The use of pooled sequencing approaches may accelerate genetic mapping and identification of causal mutations. Mascher et al.[49] applied exome sequencing to pooled samples from a barley mapping population segregating for the phenotype caused by the *mnd* (*many noded dwarf*) mutant, which increases the rate of leaf initiation. The pool of mutant plants should display a local and specific peak in read depth corresponding to the mutant genomic background around the candidate locus, which can be confirmed by the analysis of independent mutant alleles exhibiting the same phenotype.

Another example of mapping by exome sequencing borrowed from an existing pipeline for QTL-seq,[50] which was initially developed for genome sequencing in rice. This 'exome QTL-seq' approach allowed the mapping of the causal locus underlying the black lemma and pericarp (*Blp*) phenotype from a segregating population derived from doubled haploid barley lines.[51] Exome sequences were assembled into pseudo-contigs by first-ordering exomes based on the genomic coordinates of their respective genes and then limiting each locus to 200 bp in the pseudo-map (but including all relevant SNPs). Short reads generated by the sequencing of the exome capture library are then analysed through this QTL-seq pipeline. The causal loci responsible for the trait of interest are identified based on the relative enrichment in SNP allele frequencies from their original genomic background, as described above for QTL-seq.

5.5. Identification of useful genes for the genetic and genomic resources established

Natural diversity remains a major source of agronomically relevant traits for barley breeding programs. Genetic and genomic resources have been capitalized on to isolate genes of interest that control agronomic and industrial traits. Some such genes were isolated through homology-based cloning, whereas others were isolated via positional cloning strategies. In other cases, several approaches were combined to map genes: positional cloning, synteny with related grasses, and homology-based approaches. Here, we provide two examples of genes underlying important phenotypes that were isolated by our group.

The most important step that allowed barley domestication is linked to mutations in the two adjacent, dominant, and complementary genes *Brittle rachis* (*Btr*) *Btr1* and *Btr2*. Their loss of function caused barley grains to remain on the inflorescence at maturity, enabling easier and effective harvesting.[2] To identify the *btr1* and *btr2* genes, we crossed the cultivars Kanto Nakate Gold (carrying a *btr1*-type allele) and Azumamugi (bearing a *btr2*-type allele) to produce a mapping population segregating at both *btr* loci. We mapped two candidate genes genetically from >10,000 segregating individuals. We then identified BAC clones using some of the genetic markers used for mapping and sequenced positive clones. We confirmed the identification of both genes via complementation tests by transforming functional *Btr* alleles in the respective haplotypes.

Dormancy allows wild barley grains to survive dry summers in the Near East. After domestication, barley was selected for shorter dormancy periods. Sato et al.[52] isolated the major seed dormancy gene *QTL for Seed Dormancy 1* (*Qsd1*) from wild barley, which encodes an alanine aminotransferase (AlaAT). We first built a high-resolution genetic map between the cultivar Haruna Nijo and the wild barley accession H602, narrowing the mapping interval down to two Haruna Nijo BAC clones, which we then annotated, utilizing information from barley EST and FL-cDNA sequences. The candidate gene was knocked down by RNA interference and subjected to complementation tests to determine the phenotypic effects on seed dormancy. The seed dormancy gene is expressed specifically in the embryo. The two *Qsd1* alleles responsible for long and short dormancy periods encode proteins that differ by a single amino acid.

5.6. Variations in the proteomes useful for trait analysis

A unique example of an application of transcript/protein sequence information in barley is related to its industrial product: malt and beer. Iimure et al.[53] conducted a two-dimensional gel-based proteome analysis to identify proteins associated with quality traits related to malt and beer production. Several protein species were identified in malt, wort (the first extraction step after malt maceration and mashing), and beer by gel electrophoresis, followed by trypsin digestion and mass spectrometry analyses and/or liquid chromatography tandem mass spectrometry. In addition, low-molecular-weight polypeptides were isolated from beer by the combination of non-enzymatic digestion and mass analysis. Collectively, these data sets of polypeptides from barley proteomes provide a platform for analysing protein functions in beer. Several novel proteins related to beer quality traits such as foam stability and haze formation have been identified through analysing these proteomes. Some of the proteins

have also been turned into efficient protein or DNA markers for trait selection in malting barley breeding.[54]

5.7. Transformation, genome editing, and functional validation of identified genes

As described above (see description of the *brt1*/*brt2* and *Qsd1* loci), introducing a gene of interest by stable transformation is the standard technique for validating the gene responsible for the target trait. Targeted genome modification technology (so-called genome-editing) offers tremendous promise as a technology that can efficiently produce mutations in desired genes, and several cases of barley genome-editing by clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated nuclease Cas9 have already been reported.[55] However, transformation and genome-editing experiments may suffer from some limitations resulting from the low transformation potential of some accessions. Indeed, the old Scottish malting cultivar Golden Promise is one of the few reliable haplotypes for Agrobacterium (*Agrobacterium tumefaciens*)-mediated transformation. Hisano and Sato[56] identified loci controlling transformation amenability in the regions of chromosomes 2H and 3H in an F_2 population derived from a cross between the cultivars Golden Promise and Haruna Nijo. Introducing these genomic regions in target haplotypes may increase their transformation efficiency and genome-editing capabilities.

5.8. The digital bioresource project for digital breeding

Natural and induced variation provides opportunities to analyse traits of interest. However, how to combine sequence and trait information remains a challenging question that needs to be addressed. High-quality genome sequences are essential to establish a digital bioresource centre for world barley *ex situ* collections but are unfortunately also insufficient. To better understand the contribution of genomic sequence variation to various traits, automatic sequence annotation must become faster and more efficient, since this process currently relies on a slow and manual data curation step. Automated phenotyping is also emerging as an essential goal to diminish the bottleneck associated with the characterization of increasing numbers of accessions. Systematic gene inactivation or modification by genome editing may provide a functional picture of candidate genes of interest in the target haplotype. However, agronomically important traits are often controlled by multiple interacting genes, which demands a deep knowledge of trait-based genetics. Finally, the combined information collected from genomic sequences and the systematic functional analysis of genes may provide novel strategies for trait improvement, for example, the sequence-based digital breeding of barley.

6. Conclusion and future perspectives

The genomes of various organisms have been sequenced and analysed for some 40 years. During this period, development and application of new technologies including computer software frequently changed sequencing strategies. Initially, most plant genome sequencing projects had to rely on a single haplotype to establish a reference genome. However, it has become clear that to sequence multiple haplotypes in parallel would be essential for our understanding of the genetic and genomic features underlying the natural and induced variations. Newly developed technologies of various sorts have allowed us to make it quite efficient to analyse multiple accessions and lines simultaneously, even for species possessing a large genome size such as barley.

However, we realize at the same time that we need to establish genetically stable reference accessions to obtain consistent results for genomic and proteomic sequencings. In these regards, a digital bioresource including sequence and related information of all accessions appears to be quite challenging. In the meantime, however, sequencing of a limited number of accessions that cover most of the major sequence variations in a crop species may provide an alternative and efficient strategy to establish the 'pan-genome'.

High-quality sequence assembly will certainly be an indispensable component of the pan-genome infrastructure. The ongoing attempt of a barley pan-genome project aims to construct chromosome-scale sequence assembly for 20 genotypes, consisting of landraces, cultivars, as well as a wild barley accession selected to represent the global barley diversity. The details of the project will be published in the near future which is expected to present an advanced view of barley genomics.

Acknowledgments

The author thanks the National Bioresource Project, Japan, for collection, preservation and distribution of barley resources stored at Okayama University described in this review.

Conflict of interest

The author declare that there is no competing interest.

References

1. Zohary, D., Hoph, M. and Weiss, E. *Domestication of Plants in the Old World*, 4th edition. New York: Oxford University Press, 2012, pp. 20–74.
2. Pourkheirandish, M., Hensel, G., Kilian, B., et al. 2015, Evolution of the seed dispersal system in barley, *Cell*., **162**, 527–39.
3. Bothmer, R. V., Sato, K., Komatsuda, T., Yasuda, S. and Fischbeck, G. The domestication of cultivated barley. In: von Bothmer, R., van Hintum, Th., Knüpffer, H. and Sato, K., eds. *Diversity in Barley (Hordeum vulgare)*. The Netherlands: Elsevier Science B.V., 2003, pp. 9–27.
4. Komatsuda, T., Pourkheirandish, M., He, C., et al. 2007, Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene, *Proc. Natl. Acad. Sci. USA*., **104**, 1424–9.
5. Yan, L., Fu D., Li C., Blechl A., et al. 2006, The wheat and barley vernalization gene VRN3 is an orthologue of FT, *Proc. Natl. Acad. Sci. USA*., **103**, 19581–6.
6. Taketa, S., Amano, S., Tsujino, Y., et al. 2008, Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway, *Proc. Natl. Acad. Sci. USA*., **105**, 4062–7.
7. Fischbeck, G. Diversification through breeding. In: von Bothmer, R., van Hintum, Th., Knüpffer, H. and Sato, K., eds. *Diversity in Barley (Hordeum Vulgare)*, The Netherlands: Elsevier Science B.V., 2003, pp. 28–50.
8. Gaut, B.S. 2002, Evolutionary dynamics of grass genomes, *New Phytol*., **154**, 15–28.
9. International Barley Sequencing Consortium (IBSC) 2012, A physical, genetic and functional sequence assembly of the barley genome, *Nature*., **491**, 711–6.
10. Middleton, C.P., Stein, N., Keller, B., Kilian, B. and Wicker, T. 2013, Comparative analysis of genome composition in Triticeae reveals strong variation in transposable element dynamics and nucleotide diversity, *Plant J*., **73**, 347–56.

11. International Brachypodium Initiative 2010, Genome sequencing and analysis of the model grass *Brachypodium distachyon*, *Nature*., **463**, 763–8.
12. FAO. The second report on the state of the world's plant genetic resources for food and agriculture. In: *Commission on Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations, 2010, p. 370. (http://www.fao.org/docrep/013/i1500e/i1500e00.htm)
13. Knüpffer, H. Triticeae genetic resources in ex situ Genebank collections. In: Feuillet, C. and Muehlbauer, G.J., eds. *Genetics and Genomics of the Triticeae*. New York: Springer, 2009, pp. 31–79.
14. Lundqvist, U. and Franckowiak, J.D. Diversity of barley mutants. In: von Bothmer, R., van Hintum, Th., Knüpffer, H. and Sato K., eds. *Diversity in Barley (Hordeum Vulgare)*. The Netherlands: Elsevier Science B.V., 2003, pp. 75–94.
15. Franckowiak, J.D., Lundqvist, U. and Konishi, T. 1997, New and revised names for barley genes, *Barley Genet. Newsl*., **26**, 4–8.
16. Kleinhofs, A., Kilian, A., Saghai Maroof, M., et al. 1993, A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome, *Theoret. Appl. Genetics*, **86**, 705–12.
17. Marquez-Cedillo, L.A., Hayes, P.M., Jones, B.L., et al. 2000, QTL analysis of malting quality in barley based on the doubled-haploid progeny of two elite North American varieties representing different germplasm groups, *Theor. Appl. Genet*., **101**, 173–84.
18. Szűcs, P., Blake, V.C., Bhat, P.R., et al. 2009, An integrated resource for barley linkage map and malting quality QTL alignment, *The Plant Genome*., **2**, 134–40.
19. Sato, K., Nankaku, N. and Takeda, K. 2009, A high density transcript linkage map of barley derived from a single population, *Heredity*., **103**, 110–7.
20. Sato, K. and Takeda, K. 2009, An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines, *Theor. Appl. Genet*., **119**, 613–9.
21. Sato, K., Close, T.J., Bhat, P.R., Muñoz-Amatriaín, M. and Muehlbauer, G.J. 2011, Development of genetic map and alignment of recombinant chromosome substitution lines from a cross of EST donors by high accuracy SNP typing in barley, *Plant Cell Physiol*., **52**, 728–37.
22. Sato, K., Shin-I, T., Seki, M., et al. 2009, Development of 5006 full-length cDNAs in barley: a tool for accessing cereal genomics resources, *DNA Res*., **16**, 81–9.
23. Carninci, P., Kvam, C., Kitamura, A., et al. 1996, High-efficiency full-length cDNA cloning by biotinylated CAP trapper, *Genomics*., **37**, 327–36.
24. Matsumoto, T., Tanaka, T., Sakai, H., et al. 2011, Construction and sequence analysis of barley 24,783 full-length cDNAs, *Plant Physiol*., **156**, 20–8.
25. Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J. and Itoh, T. 2016, Improvement of barley genome annotations by deciphering the Haruna Nijo genome, *DNA Res*., **23**, 21–8.
26. Schreiber, M., Mascher, M., Wright, J., et al. 2020, A genome assembly of the barley 'transformation reference' cultivar golden promise, *G3-Genes Genom. Genet*., **10**, 1823–7.
27. Monat, C., Padmarasu, S., Lux, T., et al. 2019, TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools, *Genome Biol*., **20**, 284.
28. Stein, N., Prasad, M., Scholz, U., et al. 2007, A 1000 loci transcript map of the barley genome—new anchoring points for integrative grass genomics, *Theor. Appl. Genet*., **114**, 823–39.
29. Close, T.J., Bhat, P.R., Lonardi, S., et al. 2009, Development and implementation of high-throughput SNP genotyping in barley, *BMC Genomics*., **10**, 582–94.
30. Muñoz-Amatriaín, M., Moscou, M.J., Bhat, P.R., et al. 2011, An improved consensus linkage map of barley based on flow-sorted chromosomes and SNP markers, *Plant Genome*., **4**:238–49.
31. Yu, Y., Tomkins, J.P., Waugh, R., et al. 2000, A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes, *Theor. Appl. Genet*., **101**, 1093–9.
32. Schulte, D., Ariyadasa, R., Shi, B., et al. 2011, BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.), *BMC Genomics*., **12**, 247.
33. Saisho, D., Myoraku, E., Kawasaki, S., Sato, K. and Takeda, K. 2007, Construction and characterization of a bacterial artificial chromosome (BAC) library for Japanese malting barley 'Haruna Nijo', *Breed. Sci*., **57**, 29–38.
34. Schulte, D., Schulman, A., Graner, A., et al. 2009, Plant physiology focus issue on grasses—barley, *Plant Physiol*., **149**, 142–7.
35. Madishetty, K., Condamine, P., Svensson, J.T., Rodriguez, E. and Close, T.J. 2007, An improved method to identify BAC clones using pooled Overgos, *Nucleic Acids Res*., **35**, e5.
36. Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., et al. 2013, Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ), *Plant J*., **76**, 718–27.
37. Mascher, M., Gundlach, H., Himmelbach, A., et al. 2017, A chromosome conformation capture ordered sequence of the barley genome, *Nature*., **544**, 427–33.
38. Middleton, C.P., Senerchia, N., Stein, N., et al. 2014, Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe, *PLoS One*., **9**, e85761.
39. Hisano, H., Tsujimura, M., Yoshida, H., Terachi, T. and Sato, K. 2016, Mitochondrial genome sequences from wild and cultivated barley (*Hordeum vulgare*), *BMC Genomics*., **17**, 824.
40. Tanaka, T., Ishikawa, G., Ogiso-Tanaka, E., Yanagisawa, T. and Sato, K. 2019, Development of genome-wide SNP markers for barley via reference-based RNA-Seq analysis, *Frontiers Plant Sci*., **10**, 577.
41. O'Sullivan, H. 2007, GrainGenes: a genomic database for Triticeae and Avena, *Methods Mol. Biol*., **406**, 301–14.
42. Mascher, M., Schreiber, M., Scholz, U., et al. 2019, Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding, *Nat. Genet*., **51**, 1076–81.
43. Khan, A.W., Garg, V., Roorkiwal, M., Golicz, A.A., Edwards, D. and Varshney, R.K. 2020, Super-pangenome by integrating the wild side of a species for accelerated crop improvement, *Trends Plant Sci*., **25**, 148–58.
44. Milner, S.G., Jost, M., Taketa, S., et al. 2019, Genebank genomics highlights the diversity of a global barley collection, *Nat. Genet*., **51**, 319–26.
45. König, P., Beier, S., Basterrechea, M., et al. 2020, BRIDGE—a visual analytics web tool for barley genebank genomics, *Front. Plant Sci*., **11**, 701.
46. Mascher, M., Richmond, T.A., Gerhardt, D.J., et al. 2013, Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond, *Plant J*. **76**, 494–505.
47. Russell, J., Mascher, M., Dawson, I., et al. 2016, Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation, *Nat. Genet*., **48**, 1024–30.
48. Bayer, M.M., Rapazote-Flores, P., Ganal, M., et al. 2017, Development and evaluation of a barley 50k iSelect SNP array, *Frontiers Plant Sci*., **8**, 1792.
49. Mascher, M., Jost, M., Kuon, J., et al. 2014, Mapping-by-sequencing accelerates forward genetics in barley, *Genome Biol*., **15**, R78.
50. Takagi, H., Abe, A., Yoshida, K., et al. 2013, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, *Plant J*, **74**, 174–83.
51. Hisano, H., Sakamoto, K., Takagi, H., Terauchi, R. and Sato, K. 2017, Exome QTL-seq maps monogenic locus and QTLs in barley, *BMC Genomics*., **18**, 125.
52. Sato, K., Yamane, M., Yamaji, N., Kanamori, H., et al. 2016, Alanine aminotransferase controls seed dormancy in barley, *Nat. Commun*., **7**, 11625.
53. Iimure, T., Kihara, M., Sato, K. and Ogushi, K. 2015, Purification of barley dimeric α-amylase inhibitor-1 (BDAI-1) and avenin-like protein-a (ALP) from beer and their impact on beer foam stability, *Food Chem*., **172**, 257–64.

54. Iimure, T., Kihara, M., Ichikawa, S., Ito, K., Takeda, K. and Sato, K. 2011, Development of DNA markers associated with beer foam stability for barley breeding, *Theor. Appl. Genet*, **122**, 199–210.
55. Kumar, N., Galli, M., Ordon, J., Stuttmann, J., Kogel, K.H. and Imani, J. 2018, Further analysis of barley MORC 1 using a highly efficient RNA-guided Cas9 gene-editing system, *Plant Biotechnol. J.*, **16**, 1892–903.
56. Hisano, H. and Sato, K. 2016, Genomic regions responsible for amenability to Agrobacterium-mediated transformation in barley, *Sci. Rep.*, **6**, 37505.