

Foxtail Millet Genome Sequencing, Assembly, Annotation, and Application

2

Mehanathan Muthamilarasan, Shweta Shweta
and Manoj Prasad

Abstract

Among millets, foxtail millet (*Setaria italica* L.) is a well-studied crop, which is known for its genetic close-relatedness to biofuel grasses and cereals, C₄ photosynthesis, and appreciable tolerance to broad-spectrum abiotic stresses. Foxtail millet, along with its wild ancestor, green foxtail (*S. viridis* L.), are accentuated as model crops for studying the aforementioned traits. In view of their importance, the genomes have been sequenced and released. The present chapter summarizes the sequencing efforts, the outcomes, and the application of sequence information in expediting genomics studies. In addition, the chapter also provides a snapshot of how the genome sequence information has been exploited to develop different genomic resources useful for crop improvement.

2.1 Introduction

The term ‘millets’ denotes the small-seeded, polyphyletic, annual C₄ grasses belonging to the Poaceae family. This includes several species, including foxtail millet (*Setaria italica*), green foxtail (*Setaria viridis*), pearl millet (*Pennisetum glaucum*), finger millet (*Eleusine coracana*), proso millet (*Panicum miliaceum*), and other minor

millets which are farmed as food and fodder crops on deprived and trivial areas in drylands of temperate, subtropical, and tropical sections across the world (Dwivedi et al. 2011; Lata et al. 2013). Millets stand fifth in global grain production after rice, wheat, maize and sorghum (Dwivedi et al. 2011; Lata et al. 2013); however, they are not as popular as chief staple crops such as rice, maize, and wheat. Of note, millet species are a prime resource of energy for many civilizations, especially in developing countries, and it is still being consumed as a primary source of energy by considerable global population inhabiting the arid and semi-arid regions of the world (Muthamilarasan and Prasad 2015). Millets are referred to as species from not less than four distinct tribes of

M. Muthamilarasan (✉) · S. Shweta (✉) ·
M. Prasad (✉)

National Institute of Plant Genome Research
(NIPGR), Aruna Asaf Ali Marg, New Delhi 110067,
India
e-mail: manoj_prasad@nipgr.ac.in

PACMAD grasses, namely Paniceae, Paspaleae, Cynodonteae, and Eragrostideae. Pearl millet, proso millet, foxtail millet, and green foxtail millet belong to tribe Paniceae. Irrespective of their distinct domestication history, C_4 photosynthesis is a common attribute shared by all the millets. In C_4 photosynthesis, unique bundle sheaths and mesophyll cells (Kranz anatomy) are utilized to concentrate CO_2 in the neighborhood of ribulose biphosphate carboxylase/oxygenase. This mechanism decreases photorespiration and enhances water utilization efficiency in millets, especially during drought and heat stress. C_4 plants also have superior nitrogen utilization efficiency, and therefore need a lower nitrogen supply to attain photosynthetic rates equivalent to C_3 plants (Huang et al. 2016). These attributes of C_4 photosynthesis accentuate millets as climate-resilient crops as they can flourish in soils of low nutritional content, typically acclimatize to diverse environmental stresses, require minimal resources, and have significant nutritional values (Muthamilarasan and Prasad 2015; Muthamilarasan et al. 2016a).

Genomics has offered several tools and techniques to exploit an interesting genome to study the genetic factors that contribute to a trait-of-interest and transfer the genes/alleles/QTLs (quantitative trait locuses) to target crops for their improvement. In this context, millet genomes could be a repertoire of novel genes and alleles, although the polyploid and complex nature of their genomes obstructs large-scale genomic studies (Goron and Raizada 2015; Saha et al. 2016). As a first step, sequencing a small and simple genome among millets could facilitate the implementation of comparative genomics tools to unravel the genomic complexities of other millets. In this regard, foxtail millet (*S. italica* L.) has been chosen as a tractable model because of its small genome size (~ 512 Mb), inbreeding nature, low repetitive DNA content, and short life cycle (Doust et al. 2009; Li and Brutnell 2011; Lata et al. 2013; Diao et al. 2014; Muthamilarasan and Prasad 2015). In this context, two international efforts have been made to sequence the genome of foxtail millet along with its wild relative, green

foxtail (*S. viridis*), and the draft genome sequence is made available to the global research community through public databases (Zhang et al. 2012; Bennetzen et al. 2012). The present chapter summarizes the factors which have accentuated foxtail millet and green foxtail (collectively called *Setaria*) as models, their genome sequencing efforts, the outcome of the projects, annotation data, comparative genome mapping against grass genomes, and the application of the genome sequence information.

2.2 Foxtail Millet and Green Foxtail as Prospective Models

Foxtail millet and its wild ancestor green foxtail belong to the genus *Setaria* under the subfamily Panicoideae, which comprises nearly 125 different cultivated and wild species. Both the grasses share significant genome-level synteny with several bioenergy grasses, including switchgrass (*Panicum virgatum*), napiergrass (*Pennisetum purpureum*), and pearl millet (*Pennisetum glaucum*) (Doust et al. 2009; Li and Brutnell 2011). The *Setaria* genomes are also evolutionarily close to major cereals including sorghum (*Sorghum bicolor*), maize (*Zea mays*), and rice (*Oryza sativa*) (Doust et al. 2009). Although maize and sorghum were initially considered as the model for biofuel research, the large genome size of maize along with paleopolyploid evolutionary history, the bigger stature of plants, and the longer life cycle have limited the applicability of these crops. The small genome size with comparatively less repetitive DNA along with small plant size have given *Setaria* the advantage to serve as excellent surrogate genomes for studying the genetics and genomics of bioenergy grasses (Diao et al. 2014; Muthamilarasan and Prasad 2015).

In addition, *Setaria* utilizes a C_4 mode of photosynthesis, where phosphoenolpyruvate carboxylase (PEPC) supports the instantaneous uptake of carbon dioxide and transports it to RuBisCO for photosynthesis. Therefore, being a C_4 plant, *Setaria* possesses an elevated photosynthesis rate as compared to other C_3 plants,

even in the presence of high light intensity and high temperature (Way et al. 2014). Instant quenching and release of carbon dioxide by PEPC do not need a continuous opening of stomata for a longer duration, and thus a lesser amount of water is lost by transpiration, consequently contributing to the efficient use of water (Lata et al. 2013). In this context, *Setaria* could be an ideal model for studying C_4 photosynthesis. Recently, the abiotic stress tolerance of foxtail millet has gained importance, and several functional genomics studies aimed at characterizing several stress-responsive genes including *NAC* (Puranik et al. 2011, 2013), *WD40* (Mishra et al. 2012, 2014), *AP2/ERF* (Lata et al. 2011, 2014), *C₂H₂ zinc finger* (Muthamilarasan et al. 2014a), *MYB* (Muthamilarasan et al. 2014b), *DCL*, *AGO*, and *RDR* (Yadav et al. 2015a), *WRKY* (Muthamilarasan et al. 2015a), and *ADP-ribosylation factors* (Muthamilarasan et al. 2016b) suggested their putative involvement in stress-responsive molecular machinery. Thus, *Setaria* could be a prospective model system to study C_4 photosynthesis, abiotic stress tolerance, and biofuel traits.

2.3 Genome Sequencing Projects

The release of the whole genome sequence of foxtail millet and green foxtail is considered as a milestone in *Setaria* genomics. The genome was

sequenced by two independent groups led by Beijing Genome Initiative (BGI), China and the United States Department of Energy Joint Genome Institute (USDOE-JGI), USA (Zhang et al. 2012; Bennetzen et al. 2012). The summary of these two sequencing efforts is given in Table 2.1.

2.3.1 Beijing Genome Initiative

2.3.1.1 Sequencing Overview

The foxtail millet strain ‘Zhang gu’ was sequenced using whole genome shotgun combined with next-generation sequencing technology. DNA libraries of various sizes were constructed and sequenced using Illumina second-generation sequencers, which generated a raw data of 63.5 Gb (127× sequencing depth). Post filtering of raw data, ~40 Gb (80×) clean reads were used for genome assembly by SOAPdenovo. The contig N50 and scaffold N50 were 25.4 kb and 1.0 Mb, respectively, which showed that 90% of contigs and scaffolds were present in 16,903 contigs and 439 longest scaffolds, respectively. The total length of all scaffolds was estimated to be 423 Mb, with 28 Mb (6.6%) gaps, and the genome size was determined as ~490 Mb (Zhang et al. 2012). Comparing the telomeric repeats identified in sorghum with foxtail millet showed the abundant presence of 155 bp repeat elements (Zhang et al.

Table 2.1 Summary of the outcomes of foxtail millet genome sequencing efforts

Information	Beijing Genome Initiative, China (Zhang et al. 2012)	United States Department of Energy Joint Genome Institute (Bennetzen et al. 2012)
Accessions sequenced	Foxtail millet cv. ‘Zhang gu’ and ‘A10’	Foxtail millet cv. ‘Yugu1’ and green foxtail cv. ‘A10’
Platforms used	Illumina second-generation sequencer	ABI3730xl capillary sequencer; 454 FLX platform; Illumina Genome Analyzer II platform
Sequence data generated	~40 Gb	~4 Gb for ‘Yugu1’ and ~3.5 Gb for ‘A10’
Genome coverage	~86% (81% excluding gaps)	~80%
Genome size	~423 Mb	~400 Mb
Total number of genes	38801 (~82% expressed)	24,000–29,000 expressed genes

2012). The repeat elements were identified in the foxtail millet genome, which showed that 46% of the total element is composed of different transposable elements (TEs), namely DNA transposons (39.7 Mb–9.4%; CACTA, *hAT*, Tourist, Helitron, Stowaway and other tandem repeats) and retroelements (133.6 Mb–31.6%; LTR/*Copia*, LTR/*Gypsy*, LTR/other, LINEs and SINEs).

2.3.1.2 Genetic Map

A widely cultivated photo-thermosensitive male-sterile line ‘A2’ was also sequenced using Illumina GA II platform (~10× depth) and the reads were mapped onto the reference genome assembly. This identified a total of 542,322 single nucleotide polymorphisms (SNPs), 33,587 small insertion and deletions (InDels), and 10,839 structural variants (SVs) between the two foxtail millet strains. From this, 118 SNPs and 641 SVs were used to genotype an F₂ population comprising 480 individuals developed by crossing ‘Zhang gu’ and ‘A2’. Of these, 751 markers were used to construct the genetic map covering 1,865 centimorgan (cM) in total. These markers clustered into nine linkage groups and the comparison of genetic and physical distances revealed that 33% of the genome was located in low-recombination regions.

2.3.1.3 Foxtail Millet Genes

An integrated annotation pipeline was used to identify 38,801 genes in foxtail millet. The genes were mapped against several protein databases including SwissProt, TrEMBL InterPro, KEGG, and GO to retrieve the homologs with known function. This information was used to assign the function of foxtail millet genes, and a total of 30,579 genes (78.81%) were annotated whereas 8,220 (21.19%) were unannotated. The transcriptome of four tissues, namely root, leaf, spica, and stem, was sequenced and aligned to the genome assembly. The analysis showed that 81.7% of the predicted genes were expressed in these tissues. Average lengths of intron and exon of these genes were 442 and 256 bp, respectively, and the average exon per gene is 4.3 (Zhang et al. 2012). Gene ontology annotation of

identified genes revealed that 78.8% of the genes have homologs with defined functions in the public database. In addition, 1,367 pseudogenes were predicted, which could be duplicated (both parent genes and pseudogenes are multi-exon genes), retrotransposed (parent genes are multi-exon genes whereas pseudogenes are single-exon genes), or unclassified (both parent genes and pseudogenes are single-exon genes or pseudogenes are multi-exon genes). Prediction of non-coding RNA genes showed the presence of 159 miRNA genes, 382 small nuclear RNA genes (CD-box, HACA-box, and splicing), 704 tRNA genes, and 99 rRNA genes (18S, 28S, 5.8S, and 5S) in foxtail millet genome. Large clusters of rRNA genes were found on chromosomes 1, 7, 8, and 9, whereas other non-coding RNA genes showed biased chromosomal distribution (Zhang et al. 2012).

2.3.1.4 Data Availability

The sequence reads of both genome and transcriptome are available in NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra?term=SRA048234>) and DDBJ (<https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=SRA048234>). The genome sequence and annotation data set are available in NCBI (Project ID: PRJNA77795 (chromosomes) and PRJNA73995 (scaffolds); the accession number is GSM892310). The BGI team has also established a dedicated database for foxtail millet genome (<http://foxtailmillet.genomics.org.cn>), and the genome assembly, annotation, and other relevant data are available at ftp://ftp.genomics.org.cn/pub/Foxtail_millet.

2.3.2 United States Department of Energy Joint Genome Institute

2.3.2.1 Sequencing Overview

The USDOE-JGI has sequenced foxtail millet inbred line ‘Yugu1’ and green foxtail accession ‘A10’ using Sanger sequence analysis on ABI3730xl capillary sequencer and Illumina Genome Analyzer II platform, respectively. Also, 50688 BAC library clones (~12× genome

coverage) were also sequenced. The final genomic sequence of ‘Yugu1’ contained 396.7 Mb in nine chromosomes and 4.2 Mb in 327 scaffolds, which provided a genome coverage of $\sim 80\%$. In the case of ‘A10’, $\sim 3,500$ Mb of generated data provided $\sim 7\times$ coverage (Bennetzen et al. 2012).

2.3.2.2 Genetic Map

A total of 247 progenies were generated by crossing foxtail millet inbred line ‘B100’ and green foxtail accession ‘A10’, and using this progeny, a recombinant inbred line (RIL) population was constructed (Bennetzen et al. 2012). This RIL population was sequenced using paired-end Illumina technology, and 3,149,093 SNPs were identified. Among these, 992 SNPs distributed at ~ 400 -kb intervals were used to construct the genetic map comprising of nine linkage groups with a coverage of 1,416 cM. The genomes of foxtail millet and green foxtail were compared with sorghum and rice genomes, which revealed the presence of 188 and 163 syntenic blocks. These blocks were analyzed to identify the number of species-specific rearrangements in *Setaria* and sorghum relative to rice, which showed the presence of seven and four species-specific inversions in *Setaria* chromosomes 1, 2, 4, 6, 8, and 9 and the orthologous sorghum chromosomes relative to rice (Bennetzen et al. 2012).

2.3.2.3 Genome Annotation and Analysis

Annotation of fully assembled whole genome sequence predicted the presence of 35,472 protein-coding genes as primary transcripts and 5,128 are alternate transcripts. The analysis also showed that $\sim 11\%$ of these genes could be novel to foxtail millet and are interesting candidates to study. The average length of introns and exons of annotated genes were 163 and 135 bp, respectively, and the average length of proteins was 329 amino acids. These were in accordance with the findings reported in other grasses and Arabidopsis, which suggests a high degree of gene conservation among angiosperms (Bennetzen et al. 2012). In the case of transposable elements, 40% of ‘Yugu1’ genome is predicted to contain

TEs which is relatively equal to rice ($\sim 40\%$) but lower than other grass genomes, including sorghum ($\sim 62\%$), maize, and wheat (more than 80%). The relatively lower content of TEs favors *Setaria* as a model. In addition, transcriptomes of different tissues and treatments were sequenced using different sequencing platforms, which showed 34,088 RNA-seq loci distributed on 9 chromosomes of foxtail millet, and 48 miRNA encoding genes were predicted. Similar to Zhang et al. (2012), the study also analyzed the *PEPC*, *PPDK*, and *MDH* genes among the sequenced grass genomes, which showed a high degree of conservation between maize and sorghum orthologs. The malic enzyme isoform of *Setaria* is different from those in maize and sorghum, which could be interesting to study for delineating the novel signatures of C_4 evolution. In addition, analysis of *Setaria* genomes identified six drought-associated gene clusters which possess more gene members in drought-tolerant species (*Setaria* and sorghum) as compared with drought-susceptible species (maize and rice) (Bennetzen et al. 2012; Muthamilarasan and Prasad 2017). These include plant lipid transfer protein, NADH oxidase, multi-antimicrobial extrusion protein, aldo/keto reductase, glutathione S-transferase, and AMP-dependent synthetase/ligase. The genes encoding for corresponding proteins could be potential candidates for further functional studies in foxtail millet and green foxtail to elucidate their genetic basis of stress adaptation.

2.3.2.4 Data Availability

The whole genome shotgun sequence reads of foxtail millet are available in the NCBI Nucleotide database under the accession number AGNK00000000 (<https://www.ncbi.nlm.nih.gov/nucleotide/AGNK00000000>) and Phytozome (https://gold-dev.jgi.doe.gov/analysis_projects?id=Ga0039986). The updated version of the sequence reads are available under the accession number AGNK02000000, and consists of sequences AGNK02000001–AGNK02006778 (<https://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AGNK02>). The scaffolds are available under the accession numbers CM003528–CM003536 and

KQ475381–KQ475707. EST and RNA-Seq data are available at NCBI EST database (accession numbers JK546897–JK608602) and Short Read Archive (SRX116346–SRX116357).

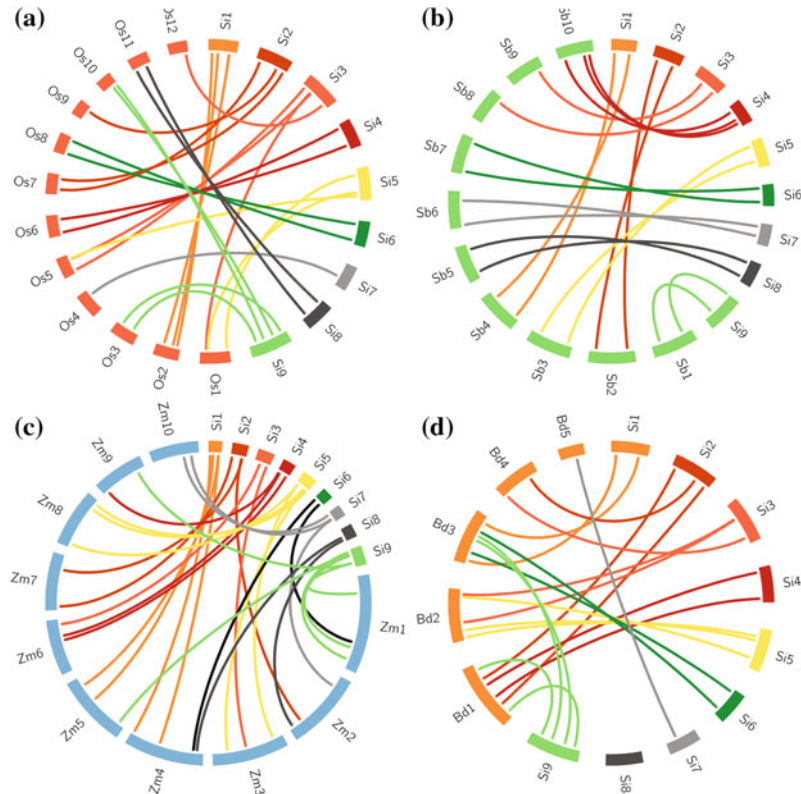
2.4 Comparative Mapping Between Foxtail Millet and Other Sequenced Grass Genomes

Foxtail millet genome was compared with the genome of other grasses, namely sorghum, maize, rice, and *Brachypodium*. A total of 24, 19, 20, and 29 large collinear blocks were identified between foxtail millet–rice, –sorghum, –*Brachypodium*, and –maize, respectively (Fig. 2.1). This indicated that foxtail millet genome is 71.8, 72.1, 61.5, and 86.7% collinear with rice, sorghum, *Brachypodium*, and maize genomes, respectively. This supports the close evolutionary relationships among the genomes, although the chromosomes have undergone extensive chromosomal shuffling

(Zhang et al. 2012). The arrangement of conserved collinear blocks suggested the nested chromosomal fusions frequently observed in grass genomes. For example, foxtail millet chromosomes 2, 3, and 9 were collinear with rice chromosomes 7 and 9, 5 and 12, and 3 and 10, respectively, suggesting the divergence of these species from a common ancestor (Zhang et al. 2012). These three pairs of rice chromosomes were individually fused to form three chromosomes of foxtail millet and, among these three fusions, two also occurred in sorghum during evolution (Zhang et al. 2012). This suggests that the two fusions occurred before the divergence of foxtail millet and sorghum, around 27 million years ago.

In addition to genome-level synteny maps, several comparative maps have been constructed between the grass genomes at the level of markers and gene families (Muthamilarasan and Prasad 2015). The first comparative map using genomic SSRs (simple sequence repeats) was

Fig. 2.1 Comparative mapping of conserved collinear blocks across the genomes of foxtail millet and **a** rice (*Os*; *Oryza sativa*), **b** sorghum (*Sb*; *Sorghum bicolor*), **c** maize (*Zm*; *Zea mays*), and **d** *Brachypodium* (*Bd*; *B. distachyon*)



performed by Pandey et al. (2013), and later by Kumari et al. (2013) using EST-derived SSRs. Later, Muthamilarasan et al. (2014a) developed comparative maps between foxtail millet and sequenced grass genomes using genome-wide intron-length polymorphic markers. Similarly, Yadav et al. (2014, 2015b) developed synteny maps using miRNA-based and transposable elements-based molecular markers. Comparative analysis of these maps revealed the percentage of sequence-based orthology and distribution of syntenic genomic regions between the genomes. Many of the physically mapped molecular markers in foxtail millet showed orthologous regions in more than one chromosome of other grass genomes. However, a higher percentage of synteny was observed between foxtail millet and sorghum, followed by foxtail millet and maize, in the case of all the molecular markers. Interestingly, it is unlike the comparative map constructed using collinear blocks, where maximum synteny was observed between foxtail millet and maize (86.7%; Zhang et al. 2012). These comparative maps also suggested several unique syntenic relationships between the genomes which were not reported by Devos (2005), Zhang et al. (2012), and Bennetzen et al. (2012) at genome level. The synteny maps also suggested the occurrence of nested chromosomal fusions frequently observed in grass genomes (Muthamilarasan and Prasad 2015). These marker-based comparative maps are very useful in the map-based isolation of genes of agronomic significance from foxtail millet using marker-based genotyping data of other related grass members.

Comparative maps were also constructed using the gene families responsible for stress response identified in foxtail millet. Members of gene families including *NAC* (Puranik et al. 2011, 2013), *WD40* (Mishra et al. 2012, 2014), *AP2/ERF* (Lata et al. 2011, 2014), *C₂H₂ zinc finger* (Muthamilarasan et al. 2014a), *MYB* (Muthamilarasan et al. 2014b), *DCL*, *AGO*, and *RDR* (Yadav et al. 2015a), *WRKY* (Muthamilarasan et al. 2015a), and *ADP-ribosylation factors* (Muthamilarasan et al. 2016b) were identified in foxtail millet to delineate their molecular and physiological roles in stress response, and

comparative maps were constructed between foxtail millet and other grass genomes. Similar to marker-based maps, the gene-based maps also revealed the close evolutionary relationship between foxtail millet and sorghum, followed by maize and rice. The decrease in synteny between foxtail millet and rice as well as *Brachypodium* was expected because of the difference in the sub-families, where rice belongs to Ehrhartoideae and *Brachypodium* is a member of Pooideae. Muthamilarasan et al. (2015b) have identified and characterized the genes encoding proteins and enzymes which participate in cellulose (*CesA/CsI*), callose (*Gsl*), and monolignol biosynthesis, (*PAL*, *C4H*, *4CL*, *HCT*, *C3H*, *CCoAOMT*, *F5H*, *COMT*, *CCR*, and *CAD*) in foxtail millet. Comparative mapping of these genes on switchgrass genome revealed maximum homology between these two crops. This demonstrates the application of comparative maps in selecting candidate orthologous genes from related genomes for functional characterization.

2.5 Applications of Genome Sequence Information

The release of foxtail millet and green foxtail genomes have expedited the structural as well as functional genomics of both crops along with other millets, thus proving the efficient application of the sequence information in dissecting the agronomic traits and also facilitating crop improvement programs. This is evident from the increasing number of publications that arise on several aspects of *Setaria* (Fig. 2.2).

2.5.1 Molecular Marker Discovery

The release of whole genome sequences has facilitated the discovery of high-throughput molecular markers at genome-scale for large-scale genotyping applications. Zhang et al. (2012) have aligned the whole genome sequence of foxtail millet varieties ‘Zhang gu’ and ‘A2’ and identified 542,322 SNPs, 33,587 InDels, and 10,839 SVs. Bennetzen et al. (2012) have

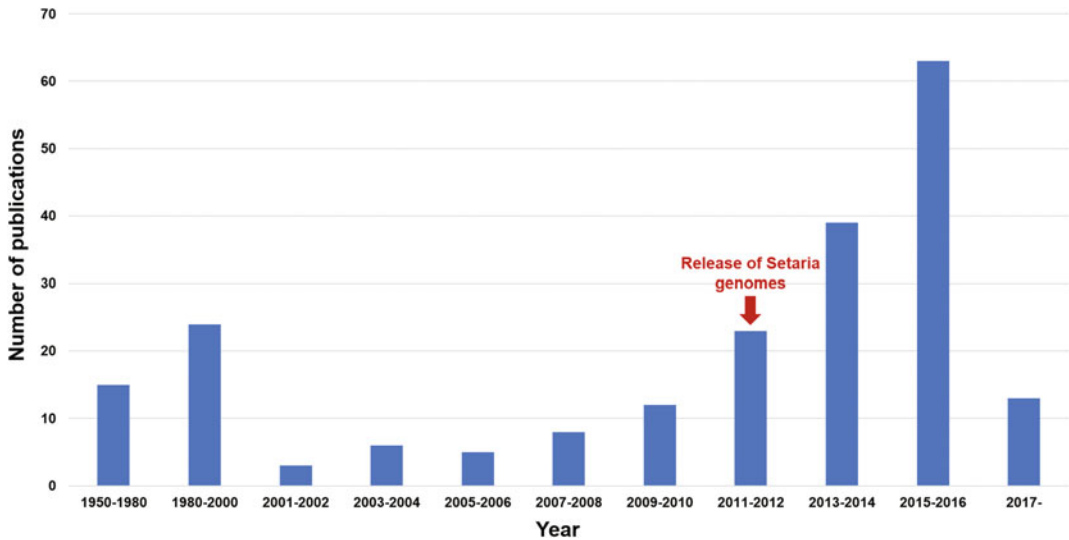


Fig. 2.2 Graphical representation of number of articles published on foxtail millet and green foxtail and indexed in NCBI PubMed

sequenced the RIL population from foxtail millet ‘B100’ X green foxtail ‘A10’ and identified 3,149,093 SNPs and used them for the construction of the genetic map. Pandey et al. (2013) have scanned the whole genome to identify 28,342 microsatellites and constructed a high-density physical map using 15,573 markers. Khan et al. (2014) have identified microRNAs from whole genome sequence information, which was then exploited to develop 176 miRNA-based markers (Yadav et al. 2014). Similarly, the genome sequence data was scanned to identify transposable elements, from which 20,278 repeat junction-based markers were developed (Yadav et al. 2015a, b). To build a haplotype map of foxtail millet, Jia et al. (2013) sequenced 916 foxtail millet varieties and identified more than 2.5 million SNPs. Through genome-wide association studies (GWAS), phenotyping of 916 varieties was carried out to identify 512 loci associated with 47 agronomic traits under five different environments. This large data set of loci with minor allele frequency >0.05 are linked with several essential agronomic traits in diverse environments. Similarly, Bai et al. (2013) performed resequencing of a foxtail

millet landrace ‘Shi-Li-Xiang’ (SLX) and aligned the reads to both reference genomes to identify the patterns of genetic variations. The study demonstrated approximately 700,000 SNPs, 26,000 InDels 1–5 bp long, and 10,000 sequence variants between SLX and Yugu1 (Bennetzen et al. 2012); over 900,000 SNPs, 20,000 InDels, and 12,000 SVs between SLX and Zhang gu (Zhang et al. 2012; Bai et al. 2013). Recently, Upadhyaya et al. (2015) have resequenced 190 foxtail millet accessions including a core collection and identified 17,714 SNPs, which were then used for genetic structure analysis and GWAS. Altogether, the release of genome sequence data has facilitated the development of large-scale genome-wide molecular markers useful for high-throughput genotyping purposes.

2.5.2 Construction of High-Density Genetic and Physical Maps

Construction of genetic and physical maps followed by saturation approaches to pinpoint the genomic regions regulating agronomic traits is

indispensable for molecular as well as genomics-assisted breeding. In the case of foxtail millet, genetic maps were constructed by Zhang et al. (2012), Bennetzen et al. (2012), Jia et al. (2013), Bai et al. (2013), and Upadhyaya et al. (2015), which were shown to be useful in mapping important traits. High-density physical maps were constructed by Pandey et al. (2013) using SSRs, Kumari et al. (2013) using EST-SSRs, Muthamilarasan et al. (Muthamilarasan et al. 2014c) using intron-length polymorphic markers, and Yadav et al. (2014, 2015a, b) using miRNA- and TE-based markers. In addition, gene family members including *NAC* (Puranik et al. 2011, 2013), *WD40* (Mishra et al. 2012, 2014), *AP2/ERF* (Lata et al. 2011, 2014), *C₂H₂ zinc finger* (Muthamilarasan et al. 2014a), *MYB* (Muthamilarasan et al. 2014b), *DCL*, *AGO*, and *RDR* (Yadav et al. 2015a), *WRKY* (Muthamilarasan et al. 2015a), and *ADP-ribosylation factors* (Muthamilarasan et al. 2016b) were also physically mapped onto foxtail millet chromosomes to construct physical maps. These maps are certainly useful in map-based selection of genes and/or markers for breeding and characterization purposes.

2.5.3 Gene Discovery

Whole genome sequence information provides access to the total number of genes that contribute to the growth, development and stress response. Increasing numbers of studies of individual genes and their corresponding gene families have suggested their involvement in diverse molecular, physiological and biological processes, and have provided novel clues on their regulation per se. In foxtail millet, *NAC* (Puranik et al. 2011, 2013), *WD40* (Mishra et al. 2012, 2014), *AP2/ERF* (Lata et al. 2011, 2014), *C₂H₂ zinc finger* (Muthamilarasan et al. 2014a), *MYB* (Muthamilarasan et al. 2014b), *DCL*, *AGO*, and *RDR* (Yadav et al. 2015a), *SET* (Yadav et al. 2016), and *Heat shock protein* (Singh et al. 2016)

have been characterized at genome-wide level and interesting candidate genes showing differential expression in response to different abiotic stresses have been identified. In addition, a transcriptome-wide analysis of *WRKY* transcription factors was performed in green foxtail (Muthamilarasan et al. 2015a), and Muthamilarasan et al. (2016b) have also performed a comparative study on structure, organization, and evolution of *ADP-ribosylation factors* in rice and foxtail millet. Zhang et al. (2012) have untapped 586 genes in the genome which was annotated as 'response to water'. These genes could be candidates for studying the molecular machinery of drought and dehydration stress adaptation. Similarly, the foxtail millet-specific genes identified by Zhang et al. (2012) and Bennetzen et al. (2012) could also be studied for their specific roles in this climate-resilient crop.

2.5.4 Comparative Genomics

The availability of high-density genetic linkage maps as well as physical maps facilitates comparative mapping of genes and genomic regions of agronomic importance in other related crop species. The close evolutionary relationship between *Setaria* and several biofuel grasses as well as cereals and millets enables the transfer of genes, alleles, and QTLs from *Setaria*. It also facilitates the identification of orthologous genes in the related crop genome for characterizing their functional roles. The higher percentage of cross-genera transferability of molecular markers, namely SSRs, eSSRs, and ILPs, also paves the way for the use of these markers in other related genera whose genome sequence information is not available. In view of these, comparative physical maps were constructed in all the studies related to marker discovery and gene family identification in foxtail millet. In addition, comparative genomics also permits the evolutionary studies including rates of evolution of genes and their corresponding families, differential gene

loss or retention following duplications and chromosomal rearrangements, and divergence patterns which collectively contribute to taxonomic, morphological, and physiological variations.

2.5.5 Sequencing and Resequencing of Genomes

The existence of predominant genome-level synteny between *Setaria* and other related species facilitates the sequencing and assembly of those complex genomes. Bennetzen et al. (2012) had sequenced the tetraploid switchgrass and demonstrated the applicability of foxtail millet genome in assembling the sequence reads. Availability of reference genomes also enables the resequencing of different cultivars, varieties, accessions, and landraces and alignment of sequence reads to the reference genome for identifying the SNPs and InDels which are useful for large-scale genotyping applications. In addition, reference genomes simplify the sequencing and analysis of transcriptomes. Unlike genomes, transcriptomes are highly dynamic, and assembly of RNA sequence reads using de novo approaches are often error prone. However, reference-based assembly allows an efficient assembly and annotation of transcripts, and thus the genome sequence plays an important role in these aspects.

2.6 Conclusions

The initiation of *Setaria* as model crops for dissecting the physiological, evolutionary and architectural traits of C_4 panicoid grasses was accentuated by the release of whole genome sequence information. This has accelerated the research in this neglected yet model crop, which promoted the crop from 'orphan' to 'crop with rich genetic and genomic resources'. It should be noted that foxtail millet research has now obtained numerous scientific leads to proceed further toward crop improvement. Attempts to develop genomic resources at large-scale and

providing unrestricted access to the research community via web-based databases would certainly accelerate functional genomics studies and molecular breeding for crop improvement. Furthermore, the crop's potential abiotic stress tolerance has encouraged the plant research community to explore the respective molecular mechanism which would enable the generation of crops with improved stress tolerance, thus ensuring food security in the scenario of global climate change.

Acknowledgements Studies on millet genomics in Dr. Manoj Prasad's laboratory are supported by Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Govt. of India [Grant No. EMR/2015/000464], by Department of Biotechnology, Govt. of India [Grant No. BT/HRD/NBA/37/01/2014], and by Core Grant of National Institute of Plant Genome Research (NIPGR), New Delhi, India. Shweta S. acknowledges the National Post-Doctoral Fellowship received from DST-SERB, Govt. of India [File No. PDF/2016/001430].

References

- Bai H, Cao Y, Quan J, Dong L, Li Z, Zhu Y, Zhu L, Dong Z, Li D (2013) Identifying the genome-wide sequence variations and developing new molecular markers for genetics research by resequencing a Landrace cultivar of foxtail millet. *PLoS ONE* 8: e73514
- Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye CY, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell TP, Doust AN, Tuskan GA, Rokhsar D, Devos KM (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol* 30:555–561
- Devos KM (2005) Updating the crop circle. *Curr Opin Plant Biol* 8:155–162
- Diao X, Schnable J, Bennetzen JL, Li J (2014) Initiation of *setaria* as a model plant. *Front Agri Sci Eng* 1:16–20
- Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. *Plant Physiol* 149:137–141
- Dwivedi S, Upadhyaya H, Senthilvel S, Hash C, Fukunaga K, Diao X, Santra D, Baltensperger D, Prasad M (2011) Millets: genetic and genomic resources. *Plant Breed Rev* 35:247–375

- Goron TL, Raizada MN (2015) Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front Plant Sci* 6:157
- Huang P, Pu H, Studer AJ, Schnable JC, Kellogg EA, Brutnell TP (2016) Cross species selections can identify components of C4 photosynthesis in the grasses. *J Exp Bot* 68:127–135
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H, Zhu C, Lu Y, Zhou C, Fan D, Weng Q, Guo Y, Huang T, Zhang L, Lu T, Feng Q, Hao H, Liu H, Lu P, Zhang N, Li Y, Guo E, Wang S, Wang S, Liu J, Zhang W, Chen G, Zhang B, Li W, Wang Y, Li H, Zhao B, Li J, Diao X, Han B (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat Genet* 45:957–961
- Khan Y, Yadav A, Suresh BV, Muthamilarasan M, Yadav CB, Prasad M (2014) Comprehensive genome-wide identification and expression profiling of foxtail millet [*Setaria italica* (L.)] miRNAs in response to abiotic stress and development of miRNA database. *Plant Cell Tissue Organ Cult* 118:279–292
- Kumari K, Muthamilarasan M, Misra G, Gupta S, Subramanian A, Parida SK, Chattopadhyay D, Prasad M (2013) Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PLoS ONE* 8:e67742
- Lata C, Bhutty S, Bahadur RP, Majee M, Prasad M (2011) Association of an SNP in a novel DREB2-like gene SiDREB2 with stress tolerance in foxtail millet [*Setaria italica* (L.)]. *J Exp Bot* 62:3387–3401
- Lata C, Gupta S, Prasad M (2013) Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Crit Rev Biotechnol* 33:328–343
- Lata C, Mishra AK, Muthamilarasan M, Bonthala VS, Khan Y, Prasad M (2014) Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e113092
- Li P, Brutnell TP (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J Exp Bot* 62:3031–3037
- Mishra AK, Puranik S, Bahadur RP, Prasad M (2012) The DNA-binding activity of an AP2 protein is involved in transcriptional regulation of a stress-responsive gene, SiWD40, in foxtail millet. *Genomics* 100:252–263
- Mishra AK, Muthamilarasan M, Khan Y, Parida SK, Prasad M (2014) Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e86852
- Muthamilarasan M, Prasad M (2015) Advances in *Setaria italica* genomics for genetic improvement of cereals and bioenergy grasses. *Theor Appl Genet* 128:1–14
- Muthamilarasan M, Prasad M (2017) Genetic determinants of drought stress tolerance in *Setaria*. In: Doust A, Diao X (eds) *Genetics and genomics of Setaria*. Springer, pp 267–289
- Muthamilarasan M, Suresh BV, Pandey G, Kumari K, Parida SK, Prasad M (2014a) Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. *DNA Res* 21:41–52
- Muthamilarasan M, Bonthala VS, Mishra AK, Khandelwal R, Khan Y, Roy R, Prasad M (2014b) C₂H₂-type of zinc finger transcription factors in foxtail millet define response to abiotic stresses. *Funct Integr Genomics* 14:531–554
- Muthamilarasan M, Khandelwal R, Yadav CB, Bonthala VS, Khan Y, Prasad M (2014c) Identification and molecular characterization of MYB transcription factor superfamily in C₄ model plant foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e109920
- Muthamilarasan M, Bonthala VS, Khandelwal R, Jaishakar J, Shweta S, Nawaz K, Prasad M (2015a) Global analysis of WRKY transcription factor superfamily in *Setaria italica* identifies potential candidates involved in abiotic stress signaling. *Front Plant Sci* 6:910
- Muthamilarasan M, Khan Y, Jaishankar J, Shweta S, Lata C, Prasad M (2015b) Integrative analysis and expression profiling of secondary cell wall genes in C₄ biofuel model *Setaria italica* reveals targets for lignocellulose bioengineering. *Front Plant Sci* 6:965
- Muthamilarasan M, Dhaka A, Yadav R, Prasad M (2016a) Exploration of millet models for developing nutrient rich graminaceous crops. *Plant Sci* 242:89–97
- Muthamilarasan M, Mangu VR, Zandkarimi H, Prasad M, Baisakh N (2016b) Structure, organization and evolution of ADP-ribosylation factors in rice and foxtail millet, and their expression in rice. *Sci Rep* 6:24008
- Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D, Prasad M (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. *DNA Res* 20:197–207
- Puranik S, Bahadur RP, Srivastava PS, Prasad M (2011) Molecular cloning and characterization of a membrane associated NAC family gene, SiNAC from foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Mol Biotechnol* 49:138–150
- Puranik S, Sahu PP, Mandal SN, Venkata Suresh B, Parida SK, Prasad M (2013) Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *PLoS One* 8:e64594
- Saha D, Dipnarayan S, Gowda MVC, Lalit A, ManjushaV Bansal KC (2016) Genetic and genomic resources of small millets. *Crit Rev Plant Sci* 35:56–79
- Singh RK, Jaishankar J, Muthamilarasan M, Shweta S, Dangi A, Prasad M (2016) Genome-wide analysis of heat shock proteins in C₄ model, foxtail millet identifies potential candidates for crop improvement under abiotic stress. *Sci Rep* 6:32641

- 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
- Way DA, Katul GG, Manzoni S, Vico G (2014) Increasing water use efficiency along the C₃ to C₄ evolutionary pathway: a stomatal optimization perspective. *J Exp Bot* 65:3683–3693
- Yadav CB, Muthamilarasan M, Pandey G, Prasad M (2014) Development of novel microRNA-based genetic markers in foxtail millet for genotyping applications in related grass species. *Mol Breed* 34:2219–2224
- Yadav CB, Muthamilarasan M, Pandey G, Prasad M (2015a) Identification, characterization and expression profiling of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in foxtail millet. *Plant Mol Biol Rep* 33:43–55
- Yadav CB, Bonthala VS, Muthamilarasan M, Pandey G, Khan Y, Prasad M (2015b) Genome-wide development of transposable elements-based markers in foxtail millet and construction of an integrated database. *DNA Res* 22:79–90
- Yadav CB, Muthamilarasan M, Dangi A, Shweta S, Prasad M (2016) Comprehensive analysis of SET domain gene family in foxtail millet identifies the putative role of *SiSET14* in abiotic stress tolerance. *Sci Rep* 6:32621
- Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, Tao Y, Bian C, Han C, Xia Q, Peng X, Cao R, Yang X, Zhan D, Hu J, Zhang Y, Li H, Li H, Li N, Wang J, Wang C, Wang R, Guo T, Cai Y, Liu C, Xiang H, Shi Q, Huang P, Chen Q, Li Y, Wang J, Zhao Z, Wang J (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30:549–554

The Foxtail Millet Genome

Prasad, M. (Ed.)

2017, XVI, 152 p. 17 illus., 16 illus. in color., Hardcover

ISBN: 978-3-319-65616-8