차세대 염기서열분석을 통한 밀 기능유전체 연구의 현황과 전망

최창현¹ · 윤영미¹ · 손재한¹ · 조성우^{2*} · 강천식^{1*}

¹국립식량과학원, ²전북대학교

Current Status and Prospect of Wheat Functional Genomics using Next Generation Sequencing

Changhyun Choi¹, Young-Mi Yoon¹, Jae-Han Son¹, Seong-Woo Cho^{2*}, and Chon-Sik Kang^{1*}

¹National Institute of Crop Science, Wanju, 55365, Republic of Korea

²Institute of Agricultural Sciences and Technology, Chonbuk National University, Jeonju, 54896, Republic of Korea

Abstract Hexaploid wheat (common wheat/bread wheat) is one of the most important cereal crops in the world and a model for research of an allopolyploid plant with a large, highly repetitive genome. In the heritability of agronomic traits, variation in gene presence/absence plays an important role. However, there have been relatively few studies on the variation in gene presence/absence in crop species, including common wheat. Recently, a reference genome sequence of common wheat has been fully annotated and published. In addition, advanced next-generation sequencing (NGS) technology provides high quality genome sequences with continually decreasing NGS prices, thereby dawning full-scale wheat functional genomic studies in other crops as well as common wheat, in spite of their large and complex genomes. In this review, we provide information about the available tools and methodologies for wheat functional genomics research supported by NGS technology. The use of the NGS and functional genomics technology is expected to be a powerful strategy to select elite lines for a number of germplasms.

Keywords Wheat, Ngs, Functional genomics, MAS, SNP, Genome editing, Marker

Received on October 25, 2018. Revised on October 26, 2018. Accepted on October 31, 2018.

* Co-corresponding Author (Chon-Sik Kang, E-mail: kcs1209@korea.kr, Tel: +82-63-238-5453, Fax: +82-63-238-5463)

* Corresponding Author (Seong-Woo Cho, E-mail: tottoriu2009@naver.com, Tel: +82-63-270-2533, Fax: +82-63-270-2640)

서 언

세계적으로 밀은 쌀, 옥수수 등과 함께 가장 중요한 식량작물 중 하나로서 세계 인구의 40% 이상이 식량으로 사용하고 있으며, 연간 600만톤 이상 생산되고 있다(Gupta et al. 2008). 밀은 인류가 작물을 재배하고 가축을 사육하여 농업의 변형과 정착생활을 이루는데 일조하였으며, 그 중 보통밀(*Triticum aestivum* L.)은 높은 수량과 가공 품질의 우수성으로 인해 가장 널리 재배되는 품종이 되었다. 보통밀은 계놈의 구조적 비교분석 결과 대략 8천년전~1만년전에 A와 B 계놈을 가지고 있는 4배체인 emmer (*T. dicoccum*, AABB)와 D계놈을 가지고 있는 2배체인 goat grass (*Aegilops tauschii*, DD)간의 자연 교배에 의해 유래되었다는 가설이 지배적이다(Tang et al. 2008). 밀의 조상인 grass 종들은 계놈의 배가(whole genome duplication, WGD)에 의해 진화적으로 보통밀이 탄생되었다. 전세계에서 널리 재배되고 있는 밀은 다양한 표현형과 다양한 환경에서 적응하는 능력을

갖추었다. 비록 밀의 지역 적응성에 관한 유전적 근거는 완전히 이해되지 않지만 WGD에 의한 것이라고 추정된다(Akhunova et al. 2010). 많은 연구자들이 그 유전요인들을 밝히기 위하여 유전정보를 탐색하고자 계놈 염기서열을 분석하였다. 식물 최초 계놈 염기서열은 대표적 모델식물인 애기장대에서 완전히 밝혀졌다(*Arabidopsis* Genome Initiative, 2000). 5년 후 단자엽의 모델식물인 벼의 계놈분석이 완성되었다(International Rice Genome Sequencing Project, 2005). 이 두 식물의 염기서열 분석법은 고전 염기서열 분석방법을 통해서 전체 계놈 염기서열을 분석하였다. 이 방법은 계놈을 적당한 크기로 잘라 bacterial artificial chromosomes (BAC) 벡터에 삽입하여 클로닝 함으로써 계놈 라이브러리(genomic library)를 만든 후, 각각의 클론을 Sanger 방법으로 염기서열을 분석하는 것이다. BAC 라이브러리의 벡터 하나 하나를 물리적 지도와 비교해 가며 전체 염기서열을 조합한다. 이 방법보다 더 빠르고 효율적인 염기서열 분석 방법인 ‘계놈 샷건’(genome shotgun, WGS) 방법은 수수(*Sorghum bicolor*)

및 포도(*Vitis vinifera*) 등의 많은 작물의 전체 게놈 염기서열 분석에 이용되었다(Paterson et al. 2009). 이 방법은 잘게 자른 DNA 조각들을 동시에 염기서열분석하여 많은 염기서열 정보를 얻은 후, 그 정보들 중 서로 겹치는 부분을 이용하여 ‘염색체 보행’(chromosome walking)을 통해 전체 염기서열을 정렬(assembly)하는 방식이다. 이 염기서열 분석방법은 시간을 어느 정도 단축시켰지만, 염색체 크기가 크고 복잡한 구조를 지닌 작물에는 비용이 너무 비싸고 염기서열의 assembly 역시 어렵다. 특히 보통밀은 WGD에 의해 유사한 염색체 6개가 존재함에 따라 게놈 크기(15×10^9)가 거대할뿐만 아니라 반복염기서열로 인해 전장유전체 염기서열 분석이 어렵다. 이러한 상황에서 다른 작물에 비해 뒤쳐진 밀 기능유전체학 연구에 큰 변화를 줄 수 있는 ‘차세대염기서열분석’(next generation sequencing, NGS) 기술 발전은 밀 기능유전체학 연구자들에게 기대가 크다(Ling et al. 2013). 앞선 기술과 NGS 기술을 통해 이배체의 원조인 A와 D 게놈뿐만 아니라 6배체 밀 게놈에 대한 유전자 서열도 거의 완성되었고, *Ae. tauschii*와 Chinese Spring의 최신 게놈 버전의 이용이 가능하게 되었다(Chapman et al. 2015, Clavijo et al. 2017). 이러한 역사적 과정을 통해 International Wheat Genome Sequencing Consortium (IWGSC)는 2005년부터 전게놈 염기서열 분석을 시작하여 13년 만에 2018년 8월에 완성된 6배체 밀 21개 염색체 염기서열 정보(Chinese spring, 약 98%, 13.8 Gb, 107,891개 유전자) 모두를 완전히 제공하였다(IWGSC 2018). 이렇게 공개된 다량의 유전정보들은 reference sequence (RefSeq)으로 이용되어 차세대염기서열 분석을 통해 계통간 resequencing 분석이 용이하게 되었으며, 분자생물학 또는 생화학적 접근방법을 응용한 ‘역유전학’(reverse genetics)에 의해 유전자의 기능이 구체적으로 밝혀지고, 이러한 기능유전자들은 육종 목표에 부합하는 ‘유전자 기반 선발 마커’(functional molecular marker, FM marker) 개발뿐만 아니라 새로운 육종 신소재 개발 등으로 다양한 밀 품종 육성에 일조할 것으로 예측된다.

본 논문에서는 차세대염기서열 분석을 통해 이루어낸 밀의 기능유전체 연구의 최근 진행과정을 정리하고 앞으로의 전망을 분석하였다. 앞으로 밀 기능유전체 연구의 발전은 유전자 지도 작성을 위한 새로운 전략과 유전적 변이 발견을 위한 새로운 소재 개발의 융화 속에서 가속화 될 것이다. 아울러 다양한 밀 특성에 부합하는 유전자가 동정 되어 다양한 소비자의 요구에 걸맞는 고품질 밀 육종 연구가 지속적으로 가능하게 될 것이다.

NGS 기반 밀 염기서열 분석

1. 전게놈 염기서열 분석

최근 가장 두드러진 발전을 보인 기술 분야는 적은 비용으로 전게놈 및 전사체 염기서열분석의 시대를 이끈 NGS 기술이라고 해도 과언이 아니다. NGS는 유전체의 고속염기서열 분석 기술로서 high-throughput sequencing, massive parallel sequencing, second-generation sequencing라고도 불리우며, 기존의 Sanger 염기서열 분석법과는 달리 많은 수의 DNA 서열을 동시에 병렬로 처리하는 특징이 있다. 대표적 NGS 기술의 예로는 ‘일루미나 시퀀싱’이다. 그 원리를 간단히 요약하자면, 임의로 자른 게놈 DNA 조각을 유리판에 부착된 탐침자(프라이머)에 결합시키고, 중합효소에 의해 결합된 DNA의 염기서열에 따라 상보적으로 프라이머가 신장될 때 서로 다른 형광물질이 표시된 염기의 배열을 읽어내는 방식이다. 일루미나 시퀀싱법은 다량의 DNA 조각의 염기서열을 동시에 읽어낼 수 있을뿐만 아니라 오류가 적은 장점이 있다(Ahmadian et al. 2006). NGS 기술은 또한 전통적인 기술에 비해 상대적으로 저렴하여 큰 게놈의 염기서열을 분석이 가능하다(Table 1). 이를 통해 중요한 형질에 관련된 유전 변이를 빠르고 체계적인 방법으로 분석할 수 있다. 이렇듯 NGS 기술의 발전은 크고 복잡한 6배체 밀과 같은 작물의 유전분석에 용이하다. 하지만, 2.5~6만년 이상 다른 종과 분리되고 유지된 보통밀의 유전자 서열은 80~90%의 높은 상동성 비율로 존재하는 반복 염기서열이 존재하기 때문에, 염기서열 분석을 통해 얻은 데이터를 구분하기가 여전히 어렵다(Safar et al. 2010). 따라서, 보통밀 기원의 이배체(D와 A 게놈의 기원인 *Ae. tauschii*와 *T. urartu*) 염기서열 분석을 통해 제공된 염기서열 정보는 보통밀에서 반복서열을 구분하는데 도움이 된다(Ling et al. 2013).

보통밀 게놈 염기서열 분석의 시작은 대표 품종인 Chinese Spring 염색체 분석으로 시작되었다. 21개의 염색체쌍 중 유전정보가 7D 단완 부분은 기존 염기서열 분석 기술을 통해 염기서열

Table 1. Comparison of genome sizes from several species.

Species	Genome sizes
<i>Saccharomyces cerevisiae</i>	12 Mb
<i>Arabidopsis thaliana</i>	135 Mb
<i>Drosophila melanogaster</i>	180 Mb
<i>Oryza sativa</i>	440 Mb
<i>Glycine max</i>	1.1 Gb
<i>Zea mays</i>	2.5 Gb
<i>Homo sapiens</i>	3.2 Gb
<i>Hordeum vulgare</i>	4.9 Gb
<i>Triticum aestivum</i>	17 Gb

조합(*de novo assembly*)을 새롭게 완성하였다(Berkman et al. 2011). 이후 7B 염색체의 단완과 4A 염색체의 장완에서 진화 초기에 전좌 현상이 일어났음이 확인되었다(Berkman et al. 2012). 보통밀의 3B 염색체는 다른 염색체와 비교하였을 때 상대적으로 온전하게 존재하여 염기서열 조합이 완성되었다. Opata M85와 W7984 품종의 계놈은 Illumina Solexa 분석을 통한 shotgun 방법으로 염기서열이 분석되었는데, 염색체 3B 염기서열은 가장 길고 정확한 염기서열 조합을 이루어 냈으며, 물리적 지도 제작의 기반이 되었다(Chapman et al. 2015). 최근에 완성된 보통밀의 RefSeq 정보와 NGS 분석 기술은 밀의 계통들 간 유전적 차이를 밝히기 위한 재염기서열분석(*re-sequencing*)을 통해 얻은 염기서열 정보의 서열화를 용이하게 한다. 이로써 전세계 밀 자생 유전자원 간의 유전적 차이를 분석하는데 있어서 용이해졌다(Hufford et al. 2012, IWGSC 2018). 2017년에 야생 *emmer* (*T. turgidum* ssp. *dicoccoides*)의 14개 염색체의 10.1 Gb 조합은 NRGene 계놈 조합 알고리즘을 통해 생성되었고, 보통밀의 AB 계놈 기원과 관련된 유전자 구성, 계놈 구조 및 유전적 다양성에 대한 자세한 유전정보를 제공함으로써 RefSeq의 완성에 일조하였다(Avni et al. 2017).

밀 계놈의 염기서열분석은 기능 유전체 연구를 위해 필요한 정보를 제공하고 있지만, 현재 계놈의 크기와 복잡성으로 인해 세부적인 염기서열 분석, 특히 단백질을 암호화하는 염기서열(*coding sequence*, *cds*)의 분석은 지연되고 있었다. 그럼에도 불구하고 일부 밝혀진(*annotation*이 이루어진) 밀 계놈 염기서열을 이용해 유전자들의 배수성 관련 연구뿐만 아니라 *transposon*의 삽입 여부, 다양한 원인으로 인한 유전자의 미성숙 종결 코돈(*premature stop codon*) 및 SNP에 의한 아미노산서열 변화 연구 등이 진행되었다(Luo et al. 2009, Choulet et al. 2010).

2. 특정 DNA 염기서열 분석

새로운 특성이 도입된 작물의 육종을 위하여 ‘분자표지 활용선발’(marker-assisted selection, MAS)의 활용이 날로 증가되고 있으며, 그 중 대표적인 기술이 염기서열정보를 활용한 ‘유전체 선발’(genomic selection, GS) 방법이다(Sallam et al. 2015). 유전정보를 활용한 GS 중, 현대 분자유종에서 가장 많이, 그리고 가장 효율적으로 사용되는 GS 기술은 ‘단일염기다형성’(single nucleotide polymorphism, SNP) 마커를 이용한 것이며, 상당히 많은 SNP 마커들이 개발되고 있다(Lai et al. 2015, Winfield et al. 2015). 식물의 특정 시기 및 조직에서 발현한 모든 전사체(*transcriptome*) 염기서열 분석은 상대적으로 크기가 작고 적은

수를 다루기 때문에 경제적이며 매우 효과적으로 SNP를 탐색할 수 있다(Cavanagh et al. 2013). 뿐만 아니라 *cds*의 염기서열 정보는 작물의 계통별 재염기서열 분석을 통해 얻어진 정보에서 유전적 다형성을 얻는데 유익하다(Jordan et al. 2015). 소위 ‘엑솜 캡처’(exome capture)라 일컬어지는 유전적 다형성을 확인하는 방법은, 특히 밀이나 배추와 같은 복잡한 계놈 구조를 가진 작물에서, 전체 염기서열 중 적은 비율로 존재하는 특정한 종류의 유전자들을 목표로 하여 유전적 다양성을 빠르고, 간단하며, 적은 비용과 DNA (1~3 μ g)량으로 분석할 수 있는 장점이 있다(Jupe et al. 2013, Wulff et al. 2014). 전계놈 염기서열 중 아주 적은 비율만이 단백질로 번역되는 암호화된 염기서열을 지니고 있다. 20,000~21,000 유전자로 구성되어 있는 인간의 경우는 200,000개의 엑손(*exon*)이 계놈상에 펼쳐져 있기 때문에 계놈당 엑손의 비율이 약 1~2%에 해당되며, 그 평균 길이는 145 bp로 매우 짧다. 밀의 경우 4배체 밀(10 Gb)과 6배체(17 Gb) 밀의 전체 엑손(*Exome*, 엑솜) 크기는 170~340 Mb 이다(Paux et al. 2006). 따라서 엑솜 염기서열분석은 엑손 부위의 유전적 다형성 발견 가능성이 전장유전체 염기서열 분석에 비해 100배 높을뿐만 아니라 반복염기서열의 출현 비율도 현저히 낮춤으로서 더욱 정확한 정보를 얻을 수 있다(Pennisi et al. 2012). 엑솜 캡처의 기술적 방법에는 *array/chip*과 *In solution* 방법으로 나눌 수 있는데, 두 가지 방법 모두 *input DNA* 샘플에서 목표하는 유전자들을 동정(*capture*) 하기 위한 탐침자(*probe/bait*)의 제작이 가장 중요하며, 두 방법 간의 큰 차이점은 *input DNA*의 양과 탐침자량의 비율이 다르다는 점이다(Grover et al. 2012). Fig. 1은 엑솜 캡처의 여러 방법 중에서 한 과정을 대표로 설명하였다. 물리적(초음파파쇄기) 또는 화학적 방법(제한효소)에 의해 잘게 자른 계놈 DNA와 biotin과 결합시킨 cDNA library의 혼성화 반응을 통해 상보적인 염기서열끼리 결합시키고, biotin 수용체가 결합된 비즈(*beads*)와 반응시킴으로써 RNA와 결합된 DNA 조각을 분리하여 최종적으로 NGS 분석을 통해 엑솜의 염기서열을 분석하는 방법이다(Fig. 1). 밀 유전체 연구에서 엑솜 염기서열 분석의 대표적 성과로, Chinese Spring의 초기 계놈 염기서열 분석을 통해 얻어진 정보와 주요 밀 재배 지역 자생 및 품종이 포함된 62종의 계통을 엑솜 캡처를 통해 얻어진 정보를 이용하여 1.57 Mb의 SNP 및 161,719의 삽입/결손(*indel*)의 다형성을 발견한 사례가 있다(Jordan et al. 2015). 이러한 연구 결과는 앞으로 GWAS (*genome-wide association study*) 수행을 통해 물리적 지도(*physical map*)를 구축함으로써 밀의 복잡한 표현형을 결정하는 유전적 요인을 밝히는데 큰 도움이 된다.

Fig. 1. Schematic diagram showing a process of exome-capture. It consists of two steps. The first step is to select only DNA fragment encodes proteins (exon regions) by probes like complementary DNA (for example, cDNA library). The second step is to sequence these exon fragment (exome) using high-throughput DNA sequencing technology (NGS).

Montenegro et al. (2017)은 기존의 IWGSC 팀(2014)에 의해 구성된 참조유전체서열정보(Reference Sequence, RefSeq)를 개선하였고, 그 향상된 Chinese Spring의 RefSeq를 바탕으로 하여, 18개 밀 품종의 유전적 다양성을 Chinese Spring의 RefSeq에 지도화 함으로써 범유전체(pangenome)를 구성하였다. 범유전체는 140,500 (± 102)개의 유전자로 구성되었으며, 각 밀 품종 당 81,070 ($\pm 1,631$)개의 핵심유전체(core genome)와 평균 128,656 개의 유전자로 구성되어 있음이 확인되었다. 아울러 범유전체 정보를 통해서 18개 밀 품종에서 발견된 12,150개의 유전자가 Chinese Spring에서는 발견되지 않았으며, Chinese Spring에서는 존재하는 245개 유전자가 18개의 품종에서는 발견되지 않음을 확인하였다. 범유전체 정보의 활용을 통해 유전자의 존재 여부뿐만 아니라, 36만개 이상의 SNP가 18개 품종에서 존재하고, 이중 2.8만개의 SNP가 Chinese Spring의 RefSeq에서는 존재하지 않음을 확인할 수 있었다(Montenegro et al. 2017). 최근 IWGSC 연구팀은 밀 염색체 21개의 염기서열 모두를 해독하였으며, 107,891개의 유전자 및 4,729,004개의 마커 정보 {504개

single-stranded repeats (SSRs) 마커, 205,807개 single nucleotide polymorphisms (SNP) 마커, 3,025개 diversity array technologies (DArTs) 마커, 6,689개의 expressed sequence tags (ESTs) 마커, 4,512,979개의 ISBPs 마커}를 공개했다. 이 성과는 73개 대학과 기업, 연구소 등이 참여하여 13년만에 이루어졌으며 개화조절, 농업특성 등 관련 유전자들을 구분하였다(IWGSC 2018).

작물의 농업 형질을 결정하는 유전적 요인은, 해당 유전자의 존재 여부, 그리고 cds 또는 regulatory sequence (promoter, untranslation region, intron 등) 내에서 SNP 또는 Indel로 인해 정상적인 단백질이 만들어지지 않거나 유전자 발현에 이상이 발생한 경우를 들 수 있다. 앞으로 이들 정보를 이용한 GS를 통해 새로운 농업 특성이 도입된 새로운 작물 생산이 활발히 진행될 것이다(Montenegro et al. 2017).

밀 계통의 유전적 다양성을 경제적으로 분석할 수 있는 또 다른 접근법으로 ‘제한효소 관련 DNA 염기서열분석’(restriction site-associated DNA sequencing: RAD-Seq) 방법을 들 수 있다. 제한효소 부위에 가까운 짧은 영역만 서열화 함으로써 데이터를 생성하는 방법으로, 적당한 제한효소를 이용하여 전체 DNA를 200~500 bp 크기로 자른 DNA의 일부 염기서열만을 분석하는 방식이다. 이 방법에 의하면 전체놈의 20~30%만을 분석하게 됨에 따라 적은 비용으로 신속히 계통간 유전적 다양성을 확인할 수 있다는 장점이 있다. 이 방법을 ‘염기서열분석을 통한 유전자 타이핑’(genotype by sequencing, GBS)이라는 용어로 자주 불리우며, 생물종에 대한 RefSeq을 요구하지 않고도 집단 특이 변이들을 찾아낼 수 있는 장점이 있다. 따라서 RAD-Seq은 RefSeq 정보가 완성되지 않은 비모델 생물체에 대한 생태 및 진화 연구에서 다량의 SNP 탐색 및 유전자형 분석에 가장 널리 사용되는 계놈 접근법이 되었다(Andrews et al. 2016). 물리적 지도 및 샷건(shotgun) 염기서열분석으로부터 발생한 정보를 이용하지 않는 GBS의 기술은, 전체놈 중 일부 정보만을 이용함에도 불구하고, 육종가들에게 매우 매력적이다. 최근 Chinese Spring의 RefSeq이 완성이 되었기 때문에 NGS 기술에 의한 GBS 정보의 정확성 및 이용성은 더욱 높아질 것으로 예상된다.

3. 전사체 염기서열 분석(RNA-seq)

밀 계놈 염기서열 분석은 다양한 환경에서 발현되는 유전자를 지도화 하기 위한 기초자료를 제공하고, 계놈의 구성요소들의 물리적 구조를 이해하는 첫번째 단계이다. 전사체 프로파일링은 발현된 모든 유전자의 동정과 발현 양상을 확인하는 것이다. 이러한 정보는 유전자 발현 조절과 그 유전자의 기능뿐만 아니라

복잡한 상호작용에 대해 깊은 분석이 가능하게 한다. 오랫동안 cDNA 미세배열(cDNA micro-array)은 밀 전사체의 연구를 위한 주된 도구가 되어왔다(Gao et al. 2012, Zhang et al. 2014). 그러나 미세배열을 이용한 전사체 정보는 미세배열에 심겨져 있는 탐침자의 수와 유전자 염기서열 정보의 정확성에 달려있기 때문에 그 이용이 제한적이다. 예를 들어, cDNA 미세배열의 경우, 유전자 발현 정도를 측정하고자 특정 조직이나 세포에서 추출한 mRNA를 cDNA로 합성하여 탐침자와 결합을 통해 나타난 신호의 세기 정도를 이용하여 유전자의 발현양을 간접적으로 확인하게 되지만, 이 과정에서 mRNA 추출을 위한 샘플의 정확성, 탐침자 염기서열의 종류 및 길이의 한계 등으로 인해 얻어진 정보의 정확성 여부가 문제점으로 대두된다(Wang et al. 2009).

‘전체 전사체 염기서열화 분석’(RNA-seq)은 NGS를 이용하여 유전자의 발현 정도를 양적으로 비교하는 방식 외에도 한 유전자로부터 ‘선택적 스플라이싱’(alternative splicing) 과정을 통해 발생하는 다양한 전사체까지도 세밀히 분석할 수 있다(Li et al. 2014, Feng et al. 2017). 최근 연구에서 Pfeifer et al. (2014)은 더 신뢰할 수 있고 유용한 전사 발현 정보를 제공하는 효과적인 RNA-seq 분석의 좋은 예를 제시하였다. 그 보고에 의하면, RNA-seq을 통해 밀 종자 생성 중에 같은 양상의 발현을 보이는 공발현(co-expression) 유전자 집단을 확인했다. 이 발현 양상들은 밀 배유가 발달하는 동안에 특정 유전자가 조직별, 시기별로 어떻게 발현하고 기능적으로 유사한 유전자들과 어떻게 상호작용하는지 여부를 유추할 수 있도록 했다. 이 연구를 통해서 조직 및 시기별 발현이 우세하는 유전자들이 발견되었고, 그 유전자들의 상호작용과 하위 유전자들의 발현 조절에 대한 정보를 제공되었다(Pfeifer et al. 2014). 이러한 분석은 전체 밀 유전자의 발현 양상을 제시하여 공발현 집단을 확인함으로써 유전자 간의 발현 조절 메커니즘을 규명할 수 있게 되고 유전자의 기능까지 유추할 수 있게 되었다. 최근에 Feng et al. (2017)은 어린 이삭부터 꽃 조직의 분화까지의 초기 밀 화기 발달 과정에 대한 전사체 연구를 통해 밀 개화 발달과정 유전자 발현 조절 네트워크를 도식화 하였으며, 밀 수확량 향상을 위한 주요 타겟 유전자들을 제시함으로써 RNA-Seq을 통한 밀 기능유전체학을 한층 발전시켰다.

밀 돌연변이체군을 이용한 유전적 다형성 분석

대부분의 밀 유전자들의 경우, 3개의 동일한 유전자(동조체, homoeologs)가 존재하기 때문에 쌀과 옥수수나 같은 이배체성 종에 사용된 기존의 유전자 식별 방법은 밀에 사용되기에는

문제가 있다. 6배체 밀의 유전체 분석 시 어떤 한 유전자에서 변이가 생기더라도 하더라도 정상인 나머지 유전자들이 존재하기 때문에 3개의 동조체 모두를 확인해야 하는 어려움이 있다. TILLING (targeted induced local lesions in genomes)은 작물의 유전자들의 기능을 검정할 수 있는 중요한 접근법임에도 불구하고 그런 이유로 인해 배수체 밀에서 TILLING 집단을 만드는 것이 필요하였다. 밀의 대표적 모델식물인 A계능 이배체 밀(*T. monococcum*)을 이용하여 EMS처리에 의한 돌연변이 집단이 만들어졌다(Rawat et al. 2012). 1532개 M₂ 돌연변이 집단은 전분합성을 수반하는 WAXY유전자와 세개의 리그닌 생합성 관련 유전자(COMT1, HCT2, 4CL1)들을 위한 연구에 사용되어, 그 유전자들이 92 kb마다 한 개씩 존재함이 확인되었다(Wang et al. 2012). 듀럼밀과 빵용밀의 TILLING 자원을 생성하기 위한 노력은 Uauy et al. (2009)에 의해 수행되었는데, 1368개의 4배체와 1536개의 6배체 M₂ 식물이 생성되었다. 돌연변이 발생 정도는 6배체에서 38 kb 당 1개로 발생된 반면, 4배체에서는 6배체보다 낮은 51 kb 당 1개로 돌연변이 비율로 발생하였다. 이 TILLING 집단 연구를 통해 SBEIIa 효소에서 돌연변이가 발생 시 아밀로스 함량이 높아지며 저항성 전분(소화가 잘 되지 않는 전분) 등 새로운 전분 기능이 생성됨을 확인하였다(Botticella et al. 2011, Hazard et al. 2012). Chinese Spring의 밀 계능 정보가 공개됨에 따라 다양한 밀 자원에서도 유전정보가 더욱 활발히 밝혀질 것이며 이러한 정보들을 통해 애기장대와 같이 언젠가는 모든 밀 유전자에서 돌연변이가 존재하는 집단을 구축하게 될 것이다.

염기서열 데이터베이스를 통한 밀 유전자형 분석

1. SNP-chip을 이용한 밀 유전적 다형성 분석

지속적인 NGS 가격 하락에도 불구하고 밀 계능 연구는 비용 및 데이터 처리에 있어서 여전히 어렵다. 그래서 SNP 미세배열(SNP-arrays/chip)은 육종가에게 있어 특정 밀 계통에서 전유전체 정보의 이해를 돕는 가장 좋은 수단이다. SNP-chip은 수 많은 SNP들을 지니고 있기 때문에 유전적 다양성을 밝히기 위한 강력한 수단이다. 그 정보들은 유전자 마커의 특징을 연구하여 더욱 세밀한 유전자지도를 제작함으로써 집단에서 각 개체 간의 유전적 근연 관계를 추론하는데 이용할 수 있다. 지난 몇 년 동안 다양한 밀도를 가진 SNP-chip이 만들어져 왔다(Jia et al. 2017).

밀 SNP 미세배열은 8개 밀 계통의 재염기서열 분석을 통해 얻은 특정 엑솜 염기서열을 바탕으로 제작되어 발전되었으며, 보통밀의 재래종과 품종의 향상을 위한 다중 표적을 위해 거의

3000개의 전 세계적 계통의 유전자형 분석에 사용되었다(Allen et al. 2011). 더욱 최근에는 9K iSelect SNP-chip을 이용하여 유전자형 분석을 하였으며, 262 계통의 중국 밀 집단을 이용하여 표현형을 분석한 결과 총 2420과 2396 SNP가 A와 B 게놈 염색체에서 존재하였다(D 게놈에서는 340개로 미비함). 아울러 이 연구결과를 통해 878개의 ‘일배체형 블록’(haplotype block, 한 염색체 내에 서로 연관되어 모여 있는 SNP 지역)이 발견되었으며, 전체 SNP의 92.4%가 그 일배체형 블록에 포함되어 있음을 확인하였다. 또한 A 게놈으로만 구성된 일배체형 블록이 B 게놈으로만 존재하는 일배체형 블록보다 더 크며, 이는 B 게놈 염색체보다 A 게놈 염색체가 육종 시 선발에 더 큰 영향을 미칠 수 있다는 것을 알 수 있었다(Hao et al. 2017). 밀 분자 육종의 한 전략으로서 일배체형 블록 기반 SNP 마커는 밀 MAS를 위해 유용하게 이용 될 것이다(Hao et al. 2017, Jia et al. 2017).

90K SNP-chip 역시 다양한 지역으로부터 19개 빵밀 계통을 바탕으로 개발되어 많이 사용되고 있다. 이 미세배열은 보통밀 계통 전반에 걸친 수 많은 SNP를 포함하고 있으며 52,607개 마커(51,655 SNPs, 667 SSR, 266 DarT 등)를 지도화 하였다(Wang et al. 2014, Wen et al. 2017). ‘밀도 기반 군집화 알고리즘’(density-based spatial clustering algorithm) 개발의 발전으로 복잡한 데이터를 대량신속처리 하여 유전자형을 분석할 수 있을 뿐만 아니라 저감도 군집들도 탐지 할 수 있게 되었다.

가장 최근의 SNP 미세배열은 192 보통밀과 60 세계적인 현대 밀 품종, 72 재래종 밀, 30 야생종 *emmer* 계통, 그리고 *Ae. tauschii*의 30 계통을 포함한 모든 계통에 적합하도록 낮은 커버리지 서열의 630,517 SNPs (660K array)를 포함한다(Cui et al. 2017). 이 SNP 미세배열을 통해 얻어진 유전자지도는 기존의 90K와 820K (unpublished)를 활용한 유전자지도와 동일한 결과 값을 얻었으며, 이 SNP들은 Chinese Spring 게놈 분석에 할당되었다. 아울러 이 SNP 미세배열은 품종화된 밀 뿐만 아니라 밀 원조 등 모든 밀 계통을 분석할 수 있는 장점이 있다.

SNP 배열은 수 많은 밀 계통의 유전자형에 적합한 플랫폼을 나타내며, 육종 적용에 이상적이다(Allen et al. 2011, Wang et al. 2014). 작지만 대표적이고 유익한 SNPs는 CerealsDB에 이용 가능한 Breeders 35K Axiom Array와 같이 실험적 비용을 더욱 줄일 수 있다. 게다가 배열에 기반을 둔 분석은 비용과 정확도 면에서 훨씬 좋은 STARP 마커 또는 게놈 특정 KASP 마커로 변환 될 수 있다(Long et al. 2017). 이 SNP들은 고품질 밀 품종, 재래종, 원형종을 포함한 다양한 유전자원에 이르기까지 폭 넓게 적용을 할 수 있을 뿐만 아니라 많은 부분이 유전적으로 지도화

되어 육종과 유전체 연구 모두에 크게 활용되고 있다.

2. Genome-wide association studies를 통한 기능유전체학

전장유전체연관분석연구(genome-wide association studies, GWAS)는 자연 집단의 유전적 변이를 확인하기 위한 또 다른 접근 방법으로 한 지역에서 적응한 재래종이 어떻게 다른 농업적 조건에서도 적응하는지에 대한 분자적 메커니즘을 연구함으로써 더 나은 작물 개량에 적용 될 수 있는 중요한 정보를 제공한다(Yu et al. 2006). 보통밀의 염기서열 기반의 GWAS는 아직 미약하지만 다양성배열기술(diversity arrays technology, DArT)과 같이 합리적인 기술들이 주로 수행되고 있다(Wisniewski et al. 2001). GWAS를 통한 일부 밀 유전체 연구의 사례를 살펴보면, 밀에서 중요한 가뭄 저항성 기작 중 하나인 수용성 탄수화물 축적 관련 유전자가 염색체 1A, 1B, 1D, 2D, 4A에서 GWAS에 의해 확인되어 이러한 분자 마커들은 내건성 밀 선발에 이용할 수 있게 되었다. 최근 중국의 10개 지역에서 확보한 723점의 재래종들이 6개의 환경에서 어떻게 생육하는지를 23개의 농업적 특성으로 평가되었다(Liu et al. 2017). 52,303개의 DArT 마커를 이용하여 GWAS를 수행하였고, 이 결과를 통해서 29개의 양적 형질 유전자 좌(quantitative trait loci, QTL)에서 149개 주요 마커가 개발되었다. 또 다른 GWAS는 9740개 DArT-seq와 178,803개 SNP 마커를 사용하여 717 중국 재래종 밀 중 194 점에서 수발아 저항성을 확인하였다.

GBS는 GWAS를 위한 효과적인 비용절감 방법 중 하나이다. 다중 샘플들은 바코드로 표지 할 수 있고 전장분자 마커 발견은 유전자형과 연결될 수 있다. 177개 *Ae. tauschii* 계통의 집단을 GBS 분석을 수행하여 유전적 다양성 분석에 이용 가능한 11,489 SNP 마커를 확인할 수 있었다(Arora et al. 2017). 또한 5249개 SNP 마커를 가지고 수행한 GWAS에서 114개 계통의 밀 종실 크기와 연관된 17개의 SNP 마커가 확인되었다. 이들 마커에 의해 세포 분열과 분화와 관계된 유전자가 염색체 6D, 5D, 2D에 위치하고 있음이 밝혀졌다. *Ae. tauschii*에서 카드름 과다 및 인 결핍 스트레스와 같은 경제적으로 중요한 특성들을 연구하기 위해 사용되어 왔다(Qin et al. 2015, Liu et al. 2015). 이렇게 NGS 기술 발전을 통한 GBS 분석에 의해 SNP를 발견하고 이들을 활용한 SNP-Chip은 밀 GWAS를 위해 더 빈번하게 사용되고 있다. 그 예로서 Sun et al. (2017)은 밀 90K 미세배열을 사용하여 163개 빵밀 품종에서 13개의 수량 특성 관련 1769개의 유전자 좌를 확인한 사례를 들 수 있다.

NGS의 도움을 받은 GWAS는 유전자원 집단에서의 유전적

변이 확인이 가능할뿐만 아니라 병원균에 대한 반응이나 환경변화 적응에 대한 염색체내 관련 유전자의 위치를 확인하는 등 다양한 유전체분석에 활용될 수 있는 기술이다(Maccaferri et al. 2015, Ovenden et al. 2017). 앞으로 GWAS는 더 정확한 표현형 조사 및 향상된 샘플 채취 방법의 모색으로 밀 기능 유전체 연구에서 중심 역할을 하게 될 것이다.

Reverse genetics를 통한 밀 기능유전체학

1. 밀 기능 유전자 클로닝

유전자 지도 기반 클로닝(map-based 또는 positional cloning)은 진보된 고전적 유전자 클로닝 방법으로서, 애기장대 또는 벼와는 달리, 유전 정보가 부족한 곡류 작물에서 농업적 중요성의 유전자를 확인하기 위한 효율적인 방법이다. 완전한 게놈 서열(IWGSC 2018)과 고품질 유전 모델(IWGSC R EFSEQ v.10, IWGSC 2014, Clavijo et al. 2017), 전사체학 정보은행, 고밀도의 SNP-chip이 포함된 유전자원을 가지고 연구하는 밀 유전체 연구는 밀 유전자의 map-based 클로닝을 훨씬 쉽게 수행하도록 한다. 결과적으로, 정밀 QTL 지도화 작업이 그와 같은 밀 게놈의 자원의 활용으로 더욱 활성화되었다. 예를 들어, 종자 무게 QTL *Qtgw-cb.5A*는 염색체 5A에 위치함을 확인하였으며, 이후 발견된 밀 게놈 서열과 유전자 모델을 이용하여 정밀한 물리적 지도를 작성했뿐만 아니라, 과피 세포의 확장을 통해 최종 종자 크기를 결정한다는 직접적인 타겟 유전자 정보를 제공했다(Brinton et al. 2017).

복잡한 게놈 6배체 밀과는 대조적으로, 이배체 밀은 map-based 클로닝과 유전 연구에 더 용이하다. 예를 들어, 광범위하게 치명적으로 줄기에 퍼지는 *Ug99*녹병에 대한 저항성 유전자를 갖고 있는 *Sr35*는 *T. monococcum* (AA)로부터 복제되었고, *T. urartu* (A-게놈 원형)와 배수체 밀에서는 그 인자가 없는 것으로 밝혀졌다. *Sr35*는 원 교배에 의해 이배체 밀로부터 이동된 후에 일반 밀에서 *Us99*에 대한 강한 저항성을 보였다(Saintenac et al. 2013).

또 다른 최근에 복제된 유전자인 *Male Sterile 2* (*Ms2*)는 map-based 클로닝에 의해 확인되었다(Ni et al. 2017, Xia et al. 2017). 밀 형질전환체 및 돌연변이를 이용하여, *Ms2* 유전자의 프로모터 영역에서 TRIM (terminal-repeat retrotransposons in miniature)을 확인하였고, 이 TRIM이 *Ms2* 유전자 활성화하여 꽃밥 특이적 발현과 웅성불임의 결과함을 확인하였다. *Ms2*는 *triticeae* 집단에서만 존재하고 다른 어떤 동족체에서도 발견되지 않아 “고아”라고 불리는 유전자다. *Ms2* 유전자의 클로닝은 밀 잡종 종자 생산 및 육종 선발을 위한 상당한 잠재력을 제공한

다. 돌연변이 스크리닝과 차세대 염기 서열 분석은 이 유전자의 클로닝 과정에서 중요한 역할을 한다. 복잡한 구조의 게놈을 가지고 있는 작물로부터 유전자 클로닝 하기 위한 방법으로 ‘긴 범위 조립’(targeted chromosome based cloning via long range assembly, TACCA)이라고 불리는 새로운 전략이 대두되었다. 새로운 유전자를 가진 품종을 개발하기 위해서는 새롭게 형성된 고품질 게놈 염기서열에 달려있다. 분류된 염색체 정보와 Chicago long-range에 연결된 프로그램을 사용하여 분석한 염기서열 조합을 이용하여 생성된 고품질 유전자 서열을 바탕으로, 이전에 사용되던 분자 마커와 EMS 처리에 의한 돌연변이를 이용하여 4개월 만에 밀 잎녹병 저항성 유전자인 *Lr22a*를 클로닝 하는데 성공하였다(Thind et al. 2017). 이러한 연구 사례들은 클로닝이 어려운 유전자를 다루는 연구자에게 큰 희망을 줄 수 있다.

2. 형질전환 기술을 활용한 밀 기능유전체

1990년대 초반까지 단자엽 식물체는 아그로박테리아(*Agrobacterium tumefaciens*)에 의해 성공한 예가 없었기 때문에 밀 형질전환체 생산은 particle bombardment 장비와 밀 미성숙배를 이용하여 뒤늦게 성공하였으며(Vasil et al. 1992), 2000년 초반에 비로소 아그로박테리아를 이용하여 안정적인 밀 형질전환체를 생산하게 되었다(Ye et al. 2002). 이후 다양한 유전자를 바탕으로 밀 형질전환체 관련 연구가 활발히 시작되었다. 최근 연구결과를 소개하면, R2R3 Myb transcription factor인 *TaPIMP1* 유전자의 과발현체를 통해 잎집무늬병균(*Bipolaris sorokiniana*)에 저항성을 보였을뿐만 아니라 스트레스 관련 유전자의 발현을 조절함으로써 가뭄저항성 증진 효과가 확인되었고(Zhang et al. 2012), 병저항성 유전자로 알려진 CC-NBS-LRR 유전자군 중 하나인 *Yr10*을 도입한 밀 형질전환체에서 줄기녹병저항성이 확인되었으며, pathogenesis-related protein 1 (*TcLr19PR1*)의 유전자 도입 밀 개체에서 잎녹병 저항성을 확인하였다(Liu et al. 2014, Gao et al. 2015). 밀 calcium-dependent protein kinase *TaCPK7-D* 유전자 도입 형질전환체 밀에서 밀 잎집눈무늬병(eyespot disease)에 저항성이 증진됨을 확인되었고, NAC transcription factor 중 하나인 *TaNAC2-5*가 밀 수확량과 관계 있음을 또한 밀 형질전환체를 통해 확인하였다(He et al. 2015, Wei et al. 2016).

밀 게놈 염기서열의 공개로 많은 유용 유전자의 동정이 수월해졌으며 발전된 형질전환 기술을 통해 많은 유전자의 기능이 밝혀질 것이고, 이러한 정보들의 축적은 다양한 밀 우수 계통을 선발하는데 큰 도움이 될 것이다.

3. 유전자편집기술을 통한 밀 기능유전체학

유전자편집기술은 새로운 작물 생산에 있어 중요한 기술로 주목받았다. 특히 GMO (genetically modified organism)의 사회 문제로 인해 식물분자생물학 연구자들에게 특정 유전자 변이체 생산 기술인 유전자편집기술이 각광받고 있다.

미국 농무성에서는 외부 유전자가 옮겨지지 않는 유전자편집 작물은 GMO로 판단하지 않는다고 판단하였고, 유럽의 경우는 현재 법정 심사 중에 있고 2018년 내로 결론이 날 것으로 예상된다. 그에 따라 현재 polyphenol oxidase (PPO) 유전자편집 벼가 2016년에 상용화 되었으며 동일한 해에 waxy gene 편집을 통해 찰성이 높고 광택이 우수한 옥수수가 판매되었다(Waltz 2016a, 2016b).

이러한 CRISPR/Cas9 시스템은 초기 유전자 편집기술인 ZFN (zinc finger nuclease)와 TALEN (transcription activator-like effector nuclease)의 후 버전으로써 목표하는 유전자의 특정부위를 자르는 특이성 및 효율성 면에서 월등히 우수하다. CRISPR/Cas9 시스템은 DNA 이중나선을 자르는 Cas9 단백질과, 유전자의 특이성을 결정하는 single guide RNA (sgRNA)으로 구성되어 있다(Fig. 2). Guide RNA와 상보성을 갖는 단일 염기서열의 결합은 Cas9 단백질의 결합을 유도하여 PAM 자리(NGG 염기서열) 앞 3번째와 4번째 염기 결합을 끊게 되고, 이렇게 잘린 부분에는 1개에서 많게는 수십개의 임의의 염기가 추가되거나 제거가 되는 DNA 회복과정을 통해 다양한 돌연변이가 발생하게 된다 (Jiang & Doudna 2017). 밀에서 실제 연구 사례를 살펴보면, CRISPR/Cas9에 의해 밀 원형질체에서 병 저항성 유전자, TaMLO 유전자를 돌연변이 시킨 사례가 있으며(Shan et al. 2014), TALEN

에 의해 흰가루병 저항성을 증진시킨 사례도 있다(Wang et al. 2017). 또한 밀 수확량과 병 저항성 관련 유전자인 TaGW2와 TaGASR7 유전자를 CRISPR/Cas9에 의해 유전자편집에 성공하였다(Liang et al. 2017).

최근 유전자 편집기술은 ‘염기편집(base editing)’ 형태로 발전하였다. CRISPR/Cas9이 획기적인 기술임에는 틀림이 없다. 하지만 인간의 질병 등 상당수의 유전적 변이는 SNP로 존재하기 때문에 하나의 염기서열을 바꾸어야 하는 상황에서는 CRISPR/Cas9의 한계가 드러난다. 그런 면에서 염기편집 기술은 많은 인간의 유전병뿐만 아니라 개선된 밀 육종에서도 주목받고 있다. 이 염기편집은 Cas9의 nuclease 활성을 억제하도록 돌연변이 시키고 시토신(C)을 유리딘(U)로 치환하는 cytidine deaminase 효소를 융합한 dCas9 (dead Cas9)을 이용하여 CRISPR/Cas9과 동일한 방법으로 해당 유전자를 편집한다(Fig. 2). 단, 변형된 dCas9으로 인해 DNA를 절단할 수 없고, 타겟 염기서열의 C를 cytidine deaminase에 의해 U로 치환한다. 이후 DNA 복제 또는 회복 과정을 통해 치환된 U는 T로 바뀔으로써 최종적으로 염기 하나만을 치환할 수 있게 되었다(Komor et al. 2016). 밀 연구 사례로서, 밀가루 색깔에 중요한 역할을 담당하며 지질 가수분해 효소 중 하나인 TaLOX2 유전자를 dCas9을 통해 밀가루 색을 개선한 사례가 있다(Zong et al. 2017).

앞으로 NGS 기술의 발달로 밀 유전정보의 질적, 양적 성장과 아울러 유전자 또는 염기편집기술의 발전으로 인해 더욱 풍성한 밀 기능유전체학의 발전을 기대할 수 있다.

Fig. 2. The schematic diagram shows the two processes of genome editing with CRISPR/Cas9 and CRISPR/dCas9. CRISPR-Cas9 system relies on creating cut nucleotide bond of target DNA with Cas9 has endonuclease activity, while dCas9 activation systems for base editing technique rely on similar components like Cas9 to intended targets. dCas9 is one of Cas9 mutants without endonuclease activity. Base editing technique to irreversible converting one target DNA base into another have been used dCas9 fused to a cytidine deaminase enzyme.

밀 분자육종을 위한 유전자 특이 마커

FM 마커(Functional molecular marker) 또는 유전자 특이 마커(gene-specific marker)는 기능을 담당하는 유전자 염기서열 내에서 유전적 다형성을 확인할 수 있는 마커로써, PCR (polymerase chain reaction)을 이용하는 면에서는 기존의 마커(SSR, RFLP, AFLP, RAPD, DArT 마커 등)들과 유사하지만, 표현형과 직접적으로 관련 있는 유전자 자체의 존재 여부 또는 다형성을 판단하기 때문에 가장 효율적인 선발 마커라 할 수 있다(Bagge et al. 2007). 밀에서 2012년까지 30여개의 유전자좌 또는 유전자가 밀에서 마커 유전자로 밝혀졌고, 97개의 FM 마커가 제작되어 93개의 대립유전자를 판독할 수 있게 되었으며, 농업형질과 병저항성 우수 밀 계통 선발을 위해 쓰이고 있다(Liu et al. 2012).

밀은 다른 작목에 비해 다양한 용도로 가공되기 때문에 용도별 가공 적성이 품종 육성에서 중요한 요인이다. 밀 종자의 질을 결정짓는 주요 요인으로는 빵용 밀의 경우는 글루텐의 종류

(HMW-GS, LMW-GS)이 가장 중요한 유전적 요인이며, 밀 가공 제품의 색깔을 결정짓는 유전적 요인으로는 Polyphenol oxidase (PPO) 활성, Lipoxynase (LOX) 활성, Yellow pigment content (YPC)이 발견되어 마커로 이용되고 있고, 그 밖에 이삭의 경도 (Puroindoline protein, Pin), 전분 특성(Granule-bound starch synthase/Waxy protein, GBSS I/Wx)이 대표적으로 밝혀져 마커로 이용되고 있다. 농업적 형질 특성 마커는 밀의 초장 관련 인자로써 Semi-dwarf (Rht) 유전자에서에서 개발되었으며, 출수에 관련하여 마커로 개발된 유전자는 Photoperiod response (Ppd) 와 Vernalization (vm) 유전자가 있다. 수확량(천립중)과 관계되는 유전자에는 Sucrose synthase 2 (Sus2), Grain width (GW2), Cell-wall invertase (CWI)에서 마커가 개발되어 이용되고 있다. 밀 주요 병 관련 마커로, 흰가루병에는 Pm3/Pm38 유전자좌, 잎녹병/줄무늬녹병 저항성에 관련된 Lr24/Yr18 유전자가 클로닝 되고 기능이 밝혀져 병 저항성 마커로 제작되어 이용되고

Locus names		
Abiotic stress	Drought	<i>Dreb-B1</i>
	Powdery mildew	<i>Pm3</i> <i>Pm38</i>
Disease	Leaf & Strip rust	<i>Lr34</i> <i>Yr18</i> <i>Pm38</i>
Agricultural traits	Semi-dwarfing genes	Rht-B1b/D1b (<i>Rht1/Rht2</i>)
	Vernalization	Photoperiod response genes (<i>Ppd</i>) Vernalization genes (<i>Vrn</i>)
	Grain weight	Sucrose synthase 2(<i>Sus2</i>)
		Grain width (<i>GW2</i>)
		Cell-wall invertase (<i>CWI</i>)
		Dehydration-responsive element binding (<i>DREB</i>)
End use quality	Noodle color	Polyphenol oxidase (<i>PPO</i>) Lipoxynase (<i>LOX</i>) Phytoene synthase (<i>PSY</i>) Zeta-carotene desaturase (<i>ZDS</i>)
	Storage protein	HMW-GS
		LMW-GS
		Gliadine
	Grain hardness	Puroindoline (<i>Pin</i>)
	Starch property	Waxy (<i>Wx</i>)

Fig. 3. Functional markers (FMs) known to date in wheat were listed. FMs for agricultural traits and end use quality have been designed from several functional genes on genome of common wheat. However, the FMs for abiotic stress and disease are rare until recently.

Fig. 4. Forward genetics seeks to find the genetic basis of a phenotype or trait from germplasm such as TILLING and genetic resources by techniques such as whole genome sequencing or RNA-Sequencing technology, while reverse genetics seeks to find what phenotypes arise as result of particular genetic sequences by using several techniques such as over-expressing, knock-down, knock-out and genome editing.

있다(review by Liu et al. 2012, Fig. 3). 환경스트레스 중 가뭄에 저항성을 갖는 유전자 마커는 Dhydration-responsive element binding (DREB) 전사조절 단백질을 타겟으로 개발되었다(Wei et al. 2009). 하지만, 더 많은 병 및 환경스트레스 관련 FM 마커 개발이 밀 육종에 필요하다.

NGS 기술의 도입을 통해 이미 밝혀진, 그리고 앞으로 밝혀질 유전정보를 통해, 유전적 다양성을 보이는 유전자원을 발굴하고, reverse genetics 기술의 도입으로 유전자 기능이 더욱 정밀하게 밝혀짐에 따라, 다양한 유전적 특성을 간단하고 정확하게 판단할 수 있는 FM 마커들이 더욱 무수히 제작될 것이다. Fig. 4에서 요약한 대로 표현형 분석과 유전형 분석을 통해 얻은 정보를 GWAS 분석을 이용하여 다양한 분자 마커가 완성되면 앞으로 다양한 요구에 맞는 새로운 밀 품종 개발이 이루어질 것으로 기대한다.

적 요

차세대 염기 서열 분석 기술의 적용은 빠르게 식물 유전체학의 지식을 확장시킴으로 기능유전자 연구의 발전을 도모하고 있다. 특히, 밀의 기능유전체학의 발전은 기존의 염기서열 분석 기술로는 가능성이 없어 보였다. 하지만 NGS의 발전은 고품질 보통밀의 RefSeq를 완성뿐만 아니라 다양한 밀 계통들의 재염기서열 분석을 가능하게 한다. 현재 이렇게 얻어진 고품질 유전정보와

유전적 다양성이 밝혀진 유전자원의 이용으로 밀 기능유전체 연구가 새로운 단계로 접어들고 있다. NGS 기술 및 reverse genetics의 발전은 앞으로 전세계에 펼쳐져 있는 야생형 밀과 재배종 밀 계통들의 유전적인 다양성 분석을 가능케 하고 밀의 유전과 진화 과정을 깊게 이해하는데 큰 도움이 될 것이다. NGS 기술의 사용과 생물정보학의 결합은 타 작물에 비해 뒤쳐진 밀의 기능유전체 연구 속도를 가속화할 것이다. 기능유전체 연구를 활용한 밀 육종의 시대, 애기장대 및 벼 분야와 같이, 다가오고 있다.

사 사

본 연구는 차세대바이오그린21 농생명기술활용연구사업(협동과제명: 밀 핵심집단구축을 위한 유전자원 특성 분석, 협동과제번호: PJ01315903)에 의해 이루어진 것임.

REFERENCES

- Ahmadian A, Ehn M, Hober S. 2006. Pyrosequencing: History, biochemistry and future. Clin Chim Acta 363: 83-94.
- Akhunova AR, Matniyazov RT, Liang HQ, Akhunov ED. 2010. Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC Genomics 11: 505.
- Allen AM, Barker GL, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brechley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). Plant Biotechnol J 9: 1086-1099.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17: 81-92.
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the rice genome. Nature 436: 793-800.
- Arora S, Singh N, Kaur S, Bains NS, Uauy C, Poland J, Chhuneja P. 2017. Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. Front Plant Sci 8: 886.
- Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H et al. 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357: 93-97.

8. Bagge M, Xia XC, Lubberstedt T. 2007. Functional markers in wheat. *Curr Opin Plant Biol* 10: 211-216.
9. Berkman PJ, Skarszewski A, Lorenc T, Lai K, Duran C, Ling EY et al. 2011. Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant Biotechnol J* 9: 768-775.
10. Berkman PJ, Skarszewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubalakova M, Simkova H, Batley J, Dolezel J, Hernandez P, Edwards D. 2012. Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theor Appl Genet* 124: 423-432.
11. Botticella E, Sestili F, Hernandez-Lopez A, Phillips A, Lafiandra D. 2011. High resolution melting analysis for the detection of EMS induced mutations in wheat SBEIIa genes. *BMC Plant Biol* 11: 156.
12. Brinton J, Simmonds J, Minter F, Leverington-Waite M, Snape J, Uauy C. 2017. Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytol* 215: 1026-1038.
13. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S et al. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci USA* 110: 8057-8062.
14. Chapman JA, Mascher M, Buluc A, Barry K, Georganas E et al. 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. 2015. *Genome Biol* 16: 26.
15. Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P et al. 2010. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* 22: 1686-1701.
16. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H et al. 2017. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res* 27: 885-896.
17. Cui F, Zhang N, Fan XL, Zhang W, Zhao CH, Yang LJ, Pan RQ et al. 2017. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. *Sci Rep* 7: 3788.
18. Feng N, Song G, Guan J, Chen K, Jia M, Huang D, Wu J, Zhang L, Kong X, Geng S, Liu J, Li A, Mao L. 2017. Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. *Plant Physiol* 174: 1779-1794.
19. Gao F, Jordan MC, Ayele BT. 2012. Transcriptional programs regulating seed dormancy and its release by after-ripening in common wheat (*Triticum aestivum* L.). *Plant Biotechnol J* 10: 465-476.
20. Gao L, Wang S, Li XY, Wei XJ, Zhang YJ, Wang HY, Liu DQ. 2015. Expression and functional analysis of a pathogenesis-related protein 1 gene, TcLr19PR1, involved in wheat resistance against leaf rust fungus. *Plant Mol Biol Rep* 33: 797-805.
21. Grover CE, Salmon A, Wendel JF. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *Am J Bot* 99: 312-319.
22. Gupta PK, Mir RR, Mohan A, Kumar J. 2008. Wheat genomics: present status and future prospects. *Intl J Plant Genomics* 2008: 896-451.
23. Hao C, Wang Y, Chao S, Li T, Liu H, Wang L, Zhang X. 2017. The iSelect 9 K SNP analysis revealed polyploidization induced revolutionary changes and intense human selection causing strong haplotype blocks in wheat. *Sci Rep* 7: 41247.
24. Hazard B, Zhang X, Colasuonno P, Uauy C, Beckles DM, Dubcovsky J. 2012. Induced mutations in the starch branching enzyme II (SBEII) genes increase amylose and resistant starch content in durum wheat. *Crop Sci* 52: 1754-1766.
25. He X, Qu B, Li W, Zhao X, Teng W, Ma W, Ren Y, Li B, Li Z, Tong Y. 2015. The nitrate-inducible NAC transcription factor TaNAC2-5A controls nitrate response and increases wheat yield. *Plant Physiol* 169: 1991-2005.
26. Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppeler SM, Lai J et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat Genet* 44: 808-811.
27. International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* 436: 793-800.
28. International Wheat Genome Sequencing consortium (IWGSC). 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345: 1251788.
29. International wheat genome sequencing consortium (IWGSC). 2018. Shifting the limits in wheat research and breeding

- using a fully annotated reference genome. *Science* 361: 661.
30. Jia M, Guan J, Zhai Z, Geng S, Zhang X, Mao L, Li A. 2017. Wheat functional genomics in the era of next generation sequencing: An update. *The Crop J* 6: 7-14.
  31. Jiang FG, Doudna JA. 2017. CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* 46: 505-529.
  32. Jordan KW, Wang S, Lun Y, Gardiner LJ, MacLachlan R, Hucl P et al. 2015. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol* 16: 48.
  33. Jupe F, Witek K, Verweij W, Sliwka J, Pritchard L, Etherington GJ, Maclean D, Cock PJ, Leggett RM, Bryan GJ, Cardle L, Hein I, Jones JD. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J* 76: 530-544.
  34. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533: 420-424.
  35. Lai K, Lorenc MT, Lee HC, Berkman PJ, Bayer PE, Visendi P et al. 2015. Identification and characterization of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat. *Plant Biotechnol J* 13: 97-104.
  36. Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J, Yin L, Zhang R, Wu L, Zheng Y, Mao L. 2014. mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26: 1878-1900.
  37. Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, Liu J, Zhang H, Liu C, Ran Y, Gao C. 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat Commun* 8: 14261.
  38. Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C et al. 2013. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496: 87-90.
  39. Liu Y, He Z, Appes R, Xia X. 2012. Functional markers in wheat: current status and future prospects. *Theor Appl Genet* 125: 1-10.
  40. Liu W, Frick M, Huel R, Nykiforuk CL, Wang X, Gaudet DA, Eudes F, Conner RL, Kuzyk A, Chen Q, Kang Z, Laroche A. 2014. The stripe rust resistance gene Yr10 encodes an evolutionary-conserved and unique CC-NBS-LRR sequence in wheat. *Mol Plant* 7: 1740-1755.
  41. Liu Y, Wang L, Deng M, Li Z, Lu Y, Wang J, Wei Y, Zheng Y. 2015. Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. *Theor Appl Genet* 128: 2203-2212.
  42. Liu Y, Lin Y, Gao S, Li Z, Ma J, Deng M, Chen G, Wei Y, Zheng Y. 2017. A genome-wide association study of 23 agronomic traits in Chinese wheat landraces. *Plant J* 91: 861-873.
  43. Long YM, Chao WS, Ma GJ, Xu, Qi LL. 2017. An innovative SNP genotyping method adapting to multiple platforms and throughputs. *Theor Appl Genet* 130: 597-607.
  44. Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ et al. 2009. Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc Natl Acad Sci USA* 106: 15780-15785.
  45. Ni F, Qi J, Hao QQ, Lyu B, Luo MC, Wang Y, Chen FJ, Wang SY, Zhang CZ, Epstein L, Zhao XY, Wang HG, Zhang XS, Chen CX, Sun LZ, Fu DL. 2017. Wheat Ms2 encodes for an orphan protein that confers male sterility in grass species. *Nat Commun* 8: 15121.
  46. Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D, Bossolini E, Chen X, Pumphrey M, Dubcovsky J. 2015. A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3. Genes Genomes Genet* 5: 449-465.
  47. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CKK, Visendi P, Lai K, Dolezel J, Batley J, Edwards D. 2017. The pangenome of hexaploid bread wheat. *Plant J* 90: 1007-1013.
  48. Ovenden B, Milgate A, Wade LJ, Rebetzke GJ, Holland JB. 2017. Genome-wide associations for water-soluble carbohydrate concentration and relative maturity in wheat using SNP and DArT marker arrays, G3. *Genes Genomes Genet* 7: 2821-2830.
  49. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* 457: 551-556.
  50. Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P et al. 2006. Characterizing the composition and evolution

- of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J* 48: 463-474.
51. Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing Consortium, Mayer KF, Olsen OA. 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345: 1250091.
52. Pennisi E. 2012. ENCODE project writes eulogy for junk DNA. *Science* 337: 1159-1161.
53. Qin Q, Li X, Zhuang J, Weng L, Liu W, Tai P. 2015. Long-distance transport of cadmium from roots to leaves of *Solanum melongena*. *Ecotoxicology* 24: 2224-2232.
54. Rawat N, Sehgal SK, Joshi A, Rothe N, Wilson DL, McGraw N, Vadlani PV, Li WL, Gill BS. 2012. A diploid wheat TILLING resource for wheat functional genomics. *BMC Plant Biol* 12: 205.
55. Safar J, Simkova H, Kubalakova M, Cihalikova J, Suchankova P, Bartos J, Dolezel J. 2010. Development of chromosome-specific bac resources for genomics of bread wheat. *Genome Res* 129: 211-223.
56. Saintenac C, Zhang W, Salcedo A, Rouse MN, Trick HN, Akhunov E, Dubcovsky J. 2013. Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* 341: 783-786.
57. Sallam AH, Endelman JB, Jannink JL, Smith KP. 2015. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *Plant Genome* 8: 1-15.
58. Shan Q, Wang Y, Li J, Gao C. 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc* 9: 2395-2410.
59. Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, Chen F. 2017. Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnol J* 15: 953-969.
60. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperms gene maps. *Genome Research* 18: 1944-1954.
61. Thind AK, Wicker T, Simkova H, Fossati D, Moullet O, Brabant C, Vrana J, Dolezel J, Krattinger SG. 2017. Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nat Biotechnol* 35: 793-796.
62. Uauy C, Paraiso F, Colasuonno P, Tran RK, Tsai H, Berardi S, Comai L, Dubcovsky J. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biol* 9: 115.
63. Vasil V, Castillo AM, Fromm ME, Vasil IK. 1992. Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Nat Biotechnol* 10: 667-674.
64. Waltz E. 2016a. Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532: 293.
65. Waltz E. 2016b. CRISPR-edited crop free to enter market, skip regulation. *Nature Biotech* 34: 582.
66. Wang TL, Uauy C, Robson F, Till B. 2012. TILLING in extremis. *Plant Biotechnol J* 10: 761-772.
67. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63.
68. Wang GT, Peng B, Leal SM. 2014. Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. *Am J Hum Genet* 94: 770-783.
69. Wang Y, Zong Y, Cao C. 2017. Targeted mutagenesis in hexaploid bread wheat using the TALEN and CRISPR/Cas systems. *Methods Mol Biol* 1679: 169-185.
70. Wei B, Jing RL, Wang CS, Chen JB, Mao XG, Chang XP, Jia JZ. 2009. *Dre1* genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs. *Mol Breed* 23: 13-22.
71. Wei XN, Shen FD, Hong YT, Rong W, Du LP, Liu X, Xu HJ, Ma LJ, Zhang ZY. 2016. The wheat calcium-dependent protein kinase TaCPK7-D positively regulates host resistance to sharp eyespot disease. *Mol Plant Pathol* 17: 1252-1264.
72. Wen W, He Z, Gao F, Liu J, Jin H, Zhai S, Qu Y, Xia X. 2017. A high-density consensus map of common wheat integrating four mapping populations scanned by the 90K SNP array. *Front Plant Sci* 8: 1389.
73. Winfield MO, Allen AM, Burrige AJ et al. 2015. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J* 14: 1195-1206.
74. Wisniewski K, Zagdanska B. 2001. Genotype-dependent proteolytic response of spring wheat to water deficiency. *J Exp Bot* 52: 1455-1463.
75. Wulff BBH, Moscou MJ. 2014. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Front Plant Sci* 5: 1-11.

76. Xia C, Zhang LC, Zou C, Gu YQ, Duan JL, Zhao GY, Wu JJ, Liu, Fang XH, Gao LF, Jiao YN, Sun JQ, Pan YH, Liu X, Jia JZ, Kong XY. 2017. A TRIM insertion in the promoter of *Ms2* causes male sterility in wheat. *Nat Commun* 8: 15407.
77. Ye XG, Shirley S, Xu HJ, Du LP, Clement T. 2002. Regular production of transgenic wheat mediated by *Agrobacterium tumefaciens*. *Agric Sci China* 1: 239-244.
78. Yu J, Buckler ES. 2006. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17: 155-160.
79. Zhang Z, Liu X, Wang X, Zhou M, Zhou X, Ye X, Wei X. 2012. An R2R3 MYB transcription factor in wheat, TaPIMP1, mediates host resistance to *Bipolaris sorokiniana* and drought stresses through regulation of defense- and stress-related genes. *New Phytol* 196: 1155-1170.
80. Zhang H, Zhu B, Qi B, Gou X, Dong Y, Xu C, Zhang B, Huang W, Liu C, Wang X, Yang C, Zhou H, Kashkush K, Feldman M, Wendel JF, Liu B. 2014. Evolution of the BBAA component of bread wheat during its history at the allohexaploid level. *Plant Cell* 26: 2761-2776.
81. Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, Qiu JL, Wang D, Gao C. 2017. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol* 35: 438-440.