

## 배추 유전체 염기서열 해독 전략과 현황

문정환 · 권수진 · 박범석

### The strategy and current status of *Brassica rapa* genome project

Jeong-Hwan Mun · Soo-Jin Kwon · Beom-Seok Park

Received: 8 April 2010 / Accepted: 13 May 2010  
© Korean Society for Plant Biotechnology

**Abstract** *Brassica rapa* is considered an ideal candidate to act as a reference species for *Brassica* genomic studies. Among the three basic *Brassica* species, *B. rapa* (AA genome) has the smallest genome (529 Mbp), compared to *B. nigra* (BB genome, 632 Mbp) and *B. oleracea* (CC genome, 696 Mbp). There is also a large collection of available cultivars of *B. rapa*, as well as a broad array of *B. rapa* genomic resources available. Under international consensus, various genomic studies on *B. rapa* have been conducted, including the construction of a physical map based on 22.5X genome coverage, end sequencing of 146,000 BACs, sequencing of >150,000 expressed sequence tags, and successful phase 2 shotgun sequencing of 589 euchromatic region-tiling BACs based on comparative positioning with the *Arabidopsis* genome. These sequenced BACs mapped onto the *B. rapa* genome provide beginning points for genome sequencing of each chromosome. Applying this strategy, all of the 10 chromosomes of *B. rapa* have been assigned to the sequencing centers in seven countries, Korea, UK, China, India, Canada, Australia, and Japan. The two longest chromosomes, A3 and A9, have been sequenced except for several gaps, by NAAS in Korea. Meanwhile a China group, including IVF and BGI, performed whole genome sequencing with Illumina system. These Sanger and NGS sequence data will be integrated to assemble a draft sequence of *B. rapa*. The imminent *B. rapa* genome sequence offers novel insights into the organization

and evolution of the *Brassica* genome. In parallel, the transfer of knowledge from *B. rapa* to other *Brassica* crops would be expected.

#### 서론

배추과 (Brassicaceae family)는 350 속 (genera)에 속하는 약 3,700 종 (species)을 포함하며, 다양한 배추과 식물이 농업적으로 중요한 채소, 양념, 사료 및 유지 작물로 이용되고 있다 (Beilstein et al. 2006). 경제적으로 볼 때 배추과 식물은 세계 채소 생산량의 약 10%와 식용유 생산량의 약 12%를 담당하고 있다 (Economic Research Service USDA 2008). 배추과는 13-19개의 족 (族)으로 구분할 수 있는데, 배추족 (Brassiceae tribe)은 약 240개 종을 포함한다. 작물로서 중요한 종들은 유지작물인 유채 (*B. napus*)와 갓 (*B. juncea*), 채소작물인 배추 (*B. rapa*)와 양배추 (*B. oleracea*), 그리고 양념작물인 흑겨자 (*B. nigra*) 등이 있다. 배추류 식물은 현재 이용되는 것 외에도 종간 교잡 등을 통한 새로운 채소작물의 개발은 물론 산업용 오일 등 물질 생산을 위한 분자농업 (molecular farming) 기주식물로의 개발 가능성도 크다. 또한 야생종들은 핵 및 세포질 융성불임, 병충 저항성, 기상 스트레스 저항성 등 농업적으로 유용한 형질을 보유하고 있는 중요한 자원이다. 따라서 배추류 식물을 지속적으로 육종 프로그램에 도입하고 활용하기 위하여 배추의 유전체에 대한 연구와 이해가 필요하다.

배추과 작물은 크고 다양한 잎과 꽃 기관의 형태적 특징을 나타낸다. 이러한 형태적 다양성은 배추과 작물의 염색체 배수성 (polyploidy)에 기인하는 유전체 수준의 변화에 근거한다 (Lukens et al. 2004). 배추과 식물들 간의 비교유전학적 연구는 염색체 배열의 유사성을 보여주며

J.-H. Mun · S.-J. Kwon · B.-S. Park (✉)  
농촌진흥청 국립농업과학원 농업생명자원부 유전자분석개발과  
(Genomics Division, Department of Agricultural Biotechnology,  
National Academy of Agricultural Science, Rural Development  
Administration)  
e-mail: pbeom@korea.kr

(Schmidt et al. 2001), 애기장대 (*Arabidopsis*)와 배추과 식물의 비교 결과 이들이 약 1천4백만 년-2천만 년 전에 공통조상으로부터 분화되었음을 추정케 해준다 (Bowers et al. 2003). 배추류의 유전체는 애기장대의 유전체가 3배수화 (triplication) 된 구조를 보이는데, 애기장대와 분화된 직후인 약 1천3백만 년-1천7백만 년 전에 전체 유전체가 3배로 증가되었다고 사료된다 (O'Neill and Bancroft 2000; Town et al. 2006; Yang et al. 2006). 이후 배수성 배추류는 염색체 재배열을 통하여 염색체 수와 크기에 변화를 겪었으며, 한편으로는 1만년 전 정도에 자연적인 이질배수체화 (allopolyploidization)를 통하여 다양하게 진화하였다 (그림 1) (U 1935; Rana et al. 2004; Johnston et al. 2005). 현재 작물로 재배되는 배추 (*B. rapa*, AA, 2n=20), 흑겨자 (*B. nigra*, BB, 2n=16), 양배추 (*B. oleracea*, CC, 2n=18)는 동질 이배체 (monogenomic diploids)이고 이들 사이의 자연적인 종간교잡에 의해서 탄생한 갓 (*B. juncea*, AABB, 2n=36), 유채 (*B. napus*, AACC, 2n=38), 에티오피아겨자 (*B. carinata*, BBCC, 2n=34)은 이질사배체 (allotetraploid)이다. 따라서 배추 유전체 연구는 배수체에서의 유전체 구조의 진화, 유전자 기능의 분화와 발현 조절 메카니즘 등을 연구하는데 필요한 중요한 정보를 제공할 것이다.

배추과 식물과 모델식물인 애기장대 사이의 밀접한 계통적 유연관계는 모델식물에 대한 연구 결과들을 직접적으로 작물의 개선에 응용할 수 있음을 의미한다. 그러나 배추과 식물에서 일어난 복잡한 염색체 재배열과 배수화 과정으로 인해 모델식물의 유전자와 직접 대응되는 배추의 유전자 (orthologous gene)가 어느 것인가를 동정하는 일이 쉽지 않다. 실제로 배추과 작물에서는 유전체의 3배

수화 이후에 유전자의 소멸과 재배열이 동반되는 염색체 단편의 소실과 복제 등 대규모의 염색체 재배열이 1천 5백만 년 이상 동안 이루어졌다 (Town et al. 2006; Yang et al. 2006). 따라서 배추는 물론 다른 배추과 작물에 대한 유전체 분석 및 비교유전체 연구가 향후 지속적으로 수행되어야 할 것이다.

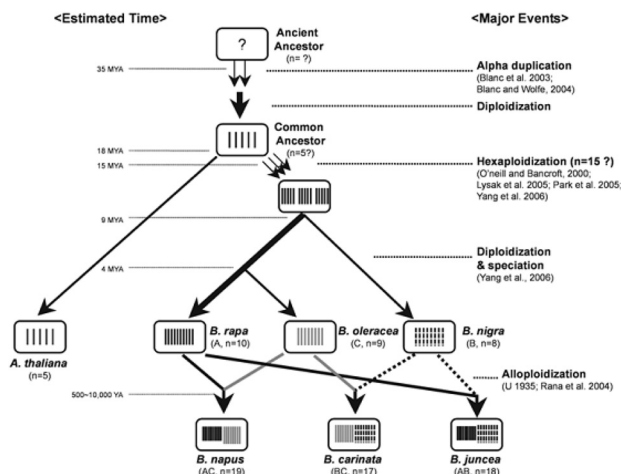
배추 (*B. rapa*)는 동지중해 유역과 동중국에서 유래한 것으로 추정되는데, 엽형 (leafy type)인 배추와 청경채, 무형 (turnip type)인 순무, 유지형 (oil type)인 옐로우사순 (yellow sarson) 등 재배 목적에 따라 다양한 작물이 지역의 특성에 맞게 전 세계에서 재배되고 있다. 배추는 이배체이면서 비교적 유전체 크기가 작고 (약 529 Mbp), 형질 전환이 가능하며, 특히 한국을 비롯한 동북아시아에서 경제적으로 중요한 채소로서 A genome을 포함한 배추과 식물의 유전 및 유전체 연구의 중심 대상으로 여겨져 왔으며, 8개국이 참여하는 다국적 유전체 프로젝트 (Multinational Brassica Genome Project, MBGP)에서 2003년 *B. rapa* ssp. *pekinensis* cv. *Chiifu-401-42*를 유전체 연구 소재로 선택하였다. 프로젝트 초기에 한국 주도로 BAC (bacterial artificial chromosome) 유전자은행 제작, 핵형 분석, 분자마커 개발 및 유전지도 작성 등을 시작으로 clone-by-clone 염기서열 분석전략에 따라 BAC 말단 염기서열 및 물리지도 정보를 기반으로 참여국들이 할당된 염색체별 염기서열 분석을 수행하였다 (<http://www.brassica.info>). 본고는 농촌진흥청 국립농업과학원의 배추 유전체 해독 연구를 통해 수행된 연구 결과를 중심으로 작성하였다.

## 배추 유전체의 특성

### 염색체 구조

배추 유전체의 세포유전학적 분석 결과 배추의 10개 염색체는 비교적 작고 조밀하며, 대부분의 유전자는 진정염색질 영역 (euchromatic space)에 집중되어 있고, 동원체는 다양한 반복서열로 구성되어 있으며 rDNA는 일부 염색체에 특징적으로 존재한다 (Lim et al. 2005, 2007). 중기 (metaphase) 염색체의 길이는 2.1-4.5  $\mu\text{m}$  크기이며 그 총 길이는 32.6  $\mu\text{m}$ 로 측정되었다. 태사기 (pachytene) 염색체에 대한 DAPI (4',6-diamidino-2-phenylindole dihydrochloride) 및 FISH (fluorescent *in situ* hybridization)에 의한 핵형 분석은 염색체의 길이가 23.7-51.3  $\mu\text{m}$  정도이고 총 길이는 385.3  $\mu\text{m}$ 인 것으로 측정되었다. 따라서 염색체의 길이는 감수분열 전기 (prophase) 시기에서는 체세포분열 중기 (metaphase) 시기보다 약 12배 길게 관찰되었다 (Koo et al. 2004).

배추 세포유전지도는 FISH 방법으로 분석한 33개 STS



**Fig. 1** Schematic representation of the evolutionary process and genetic relationships between *Arabidopsis* and *Brassica* species with the chromosome numbers, based on literature review and our research. Estimated time of major events represented as million years ago (Mya) on the left side; major events of genome divergence were summarized on the right side

(sequence tagged site) 마커와 반복서열 5종의 염색체상의 위치를 유전지도와 비교함으로써 각각의 염색체와 연관 군지도를 일치시켜 작성하였다 (Lim et al. 2005; Kim et al. 2006). 현재까지 유전자를 포함하는 BAC 클론 100 여개와 중요 반복서열의 염색체상의 위치를 FISH 분석을 통하여 결정하였고, 이를 통해 10개 염색체의 동원체 위치, 이질염색질 (heterochromatin)의 크기와 패턴, rDNA 등 다양한 반복서열의 종류와 분포가 파악되었다. 동원체를 구성하는 반복서열이 중기 염색체 분석에서 약 30%로 측정되었고, 특히 태사기 FISH에서 동원체 주변의 이질염색질 영역 (pericentromeric heterochromatic region)이 38.2 um로 전체의 약 10%로 측정되어 배추의 전체 이질염색질은 약 40%인 것으로 추정되었다.

한편 유전자 지역에 위치하는 BAC 클론들을 사용한 태사기 FISH 분석에 의하면, 태사기 염색체 1 um는 약 400-500 kbp인 것으로 계산되었다. 이상의 결과는 염기서열, 분자마커, 세포유전학적 결과를 통합하는 근거가 되었고, 유전자 지역에 대한 clone-by-clone 염기서열 분석을 수행하여 대다수의 유전자 구조를 밝히는데 이정표 역할을 할 것으로 기대된다.

#### 반복염기서열과 이질염색질

진화적으로 빠르게 구조가 변하고 있는 동원체는 연속된 반복염기서열 (tandem satellite repeats)과 동원체 특이 전이인자로 구성되어 있다. 식물의 동원체 지역을 구성하는 반복서열은 150-180 bp 단위로 종종 수백만 bp 이상 반복배열 지역을 형성하는데, 애기장대의 pAL1 (Copenhaver et al. 1999), 벼의 CentO (Zhang et al. 2004), 옥수수의 CentC (Ananiev et al. 1998), 그리고 메디카고 (*Medicago truncatula*)의 MtRs (Kulikova et al. 2004)가 알려져 있다. 배추의 BAC 클론 10,204개의 말단 염기서열을 분석한 결과 약 30%의 BAC 말단 서열에서 양쪽 끝에 제한효소 *HindIII* 자리가 있는 176 bp의 동원체 반복서열을 발견하였다. 배추의 동원체 반복서열은 82% 염기서열 유사성을 보이는 두 종류로 구성되어 있는데, 각각 CentBr1과 CentBr2로 명명하였다. 배추의 동원체 반복서열을 이용한 FISH 분석에 의하면 CentBr1은 염색체 A1, A2, A4와 A6-A10에 주로 분포하고 CentBr2는 염색체 A3와 A5에서만 분포하는 것으로 나타났다. 한편 양배추의 whole-genome shotgun (WGS) 염기서열에서도 CentBr이 발견되었는데, 이는 CentBr이 배추류 식물의 공통적인 동원체 염기서열임을 의미한다.

한편 이질염색질을 함유한 BAC 클론의 염기서열 분석을 통하여 동원체 특이 Ty1-copia-like retrotransposon (CRB), 238 bp 길이의 단순연속 반복서열 (TR238), 변형된 rDNAs, 그리고 동원체 주변 특이 (pericentromere-specific) Ty3-gypsy-like retrotransposons (PCRBr) 등을 발견할 수 있었다 (Lim

et al. 2007). 전이인자 CRB는 배추류의 이배체 3종과 이질사배체 3종의 동원체에서 모두 관찰되는 주요 인자였으나, TR238과 PCRBr은 A genome인 배추에서만 특이적으로 발견되었다. 이러한 동원체 및 그 주변 특이 반복서열은 진정염색질과 이질염색질의 경계를 탐색하는데 중요한 단서로 이용될 수 있다. 이 외에도 배추 BAC 말단 염기서열을 조사한 결과 약 50%에 달하는 BAC 말단 염기서열은 애기장대의 유전체 염기서열과는 유사성이 없는 것으로 분석되었다. 이러한 사실들은 배추 특이적인 서열의 증폭이 배추 유전체의 팽창에 기여하고 있으며, 그들 대부분은 이질배수체 영역의 전이인자 혹은 반복서열에 의해 일어남을 의미한다 (Lim et al. 2005; Yang et al. 2006).

전이인자 중 불완전한 구조로 non-autonomous한 독특한 LTR retrotransposon의 일종인 TRIM (terminal-repeat retrotransposons in miniature)이 발견되었다 (Yang et al. 2007). TRIM은 길이가 100-250 bp 정도인 말단 반복서열 (TR, terminal repeats)을 특징으로 하는데, 안쪽 염기서열은 ~300 bp 정도이고, 전이인자 양끝에는 5 bp TSD (target site duplication)를 갖고 있다. 안쪽의 염기서열은 methionine-tRNA 상보적인 서열로 시작하여 전형적인 polypurine tract motifs로 끝나고 있다. 배추의 paralogous BAC 클론 4종의 염기서열로부터 4종의 배추 TRIM (Br1-Br4)을 분리할 수 있었는데, 그 길이는 364~1,311 bp로 다양했다 (Yang et al. 2007). 약 96 Mbp의 배추 BAC 말단 염기서열의 분석을 통해 추정되는 배추 TRIMs의 카피 수는 애기장대 보다 무려 6배나 많았다. 이는 다양한 TRIM elements가 배추와 애기장대의 종 분화 이후에 배추 유전체에서 증폭되었음을 의미한다. 아직 배추 유전체에서 전이인자 TRIMs의 정확한 위치는 파악하지 못했지만, 애기장대의 경우 유전자 지역인 진정염색질에서의 TRIMs 분포 양상과 배추의 ESTs (expressed sequence tags) 염기서열에서 키메라 (chimeric) 형태의 TRIM이 발견됨을 볼 때, 배추에서도 TRIM elements가 유전자 지역에 주로 분포할 것으로 추정된다. TRIMs의 이러한 분포 양상은 진화적으로 중요한 의미를 갖는데, 이들의 이동과 증폭은 배추 유전체의 재구성은 물론 프로모터나 터미네이터의 재배치를 통하여 유전자의 기능이나 발현 양상에 영향을 주었을 것으로 유추된다. 이러한 유전체 서열의 재구성은 배추 유전체의 3배체화 이후 중복된 유전자들의 기능이 분화하거나 새로운 능력을 갖게 되는 계기가 되었을 것으로 추정된다. 한편 배추 TRIMs의 유전체 분포 특성으로 볼 때 다양한 배추류 작물의 종내 또는 종속간 구분을 위한 분자마커로서 활용 가능하다. 따라서 우리는 Br1과 Br2의 특정 염기서열을 이용하여 전이인자 전시(transposon display)함으로써 사용되는 배추의 F<sub>1</sub> 종자 16종은 물론 다른 배추류의 종과 속을 구분할 수 있었다 (Kwon et al. 2007).

## 배추 유전체의 3배수화 특성

모델 식물들의 전체 유전체 구조가 해독됨으로써 얻어진 지식과 정보들은 현재의 생물학적 지식을 수정하거나 미래의 중요한 농업적 문제들에 해답을 제공할 것으로 기대된다. 식물의 유전체 정보가 생산, 축적되는 양이 증가하면서 벼과와 콩과의 몇몇 식물간의 비교유전체 연구가 이루어졌다 (Paterson et al. 2003; Choi et al. 2004; Cannon et al. 2006). 애기장대는 배추과에 속하며 유전적, 진화적으로 배추와 매우 밀접한 관계이므로 이미 보고된 애기장대의 유전체 정보는 배추의 유전체 구조를 유추하거나 진화적 비교를 하는데 매우 유용하다. 분자유전학적인 방법과 일부 상동성이 있는 유전체 일부분의 염기서열 비교로 우선 밝혀진 배추류 유전체 구조의 특성은 3배수성 (triplicated nature)으로 나타났는데, 이러한 결과는 배추류와 애기장대의 종 분화 직후 배추류 유전체가 3배로 증가하였거나, 이런 유전체의 증가가 배추류와 애기장대의 종 분화의 원인이었을 것으로 해석된다 (Lysak et al. 2005; Town et al. 2006; Yang et al. 2006). 즉, 애기장대의 유전체는 2배수성 구조로 반수체 (n) 중에 유사한 염기서열이 두 지역에 존재하는데 (The *Arabidopsis* Genome Initiative 2000), 배추 반수체에는 애기장대의 대응 염기서열이 일반적으로 각각 세 지역에서 존재한다.

배추 유전체의 3배수성은 유전지도와 물리지도의 비교 또는 염기서열 수준에서의 synteny 비교에 의하여 알 수 있었다. 장원 배추의 EST 염기서열에 근거한 RFLP (restriction fragment length polymorphism) 마커 545개의 유전지도와 애기장대의 유전자 염기서열을 비교한 결과 단지 62개의 마커가 애기장대와 단일 유전자좌위 (single locus)에서 일치하였고 나머지는 모두 2개 이상의 유사 염기서열을 보였다 (Kim et al. 2006). 또한 유사한 synteny를 보이는 유전체 단편에서의 분자마커의 분포를 조사한 결과는 배추 유전체의 3배수성을 보여주었다 (Park et al. 2005). 한편 *FLC* (*FLOWERING LOCUS C*) 유전자 지역을 포함하는 배추의 BAC 클론 4개와 애기장대 5번 염색체의 대응 부분의 염기서열 수준에서의 비교는 6배체 (hexaploid) 모양의 조상 배추가 유전자 또는 일부 염기서열의 소실과 재배치를 통하여 이배체로 진화하는 패턴을 자세히 보여주고 있다 (Yang et al. 2006). 염기서열의 변화로 추정된 애기장대와 배추류의 종분화는 약 1천7백만 년-1천8백만 년 전에 있었고, 배추류의 유전체 3배수화는 약 1천3백만 년-1천7백만 년 전에 일어난 것으로 유추된다. (이 결과는 그림 1에 요약되어 있다.) 또한 배추 유전체에서는 약 8십만 년 전에 국부적인 복제 (segmental duplication)가 이루어진 것으로 추정된다. 배추 유전체는 이배체화 (diploidization) 과정을 거치면서 유전자 또는 염기서열이 심각하게 소실되고 일부 복제와 TRIM과 같은 전이인자의 삽입이 있었

는데, 전체적으로는 3배수화로 증가된 유전체의 약 40%를 잃어버렸으며, 그 결과 배추가 갖는 전체 유전자 수는 애기장대의 약 1.7배인 것으로 추정된다 (Mun et al. 2009).

배추의 진화과정에서 유전체 반복성 증가와 이배체화는 잉여의 유전자가 소실되는 것 외에도 다양한 돌연변이를 수용하여 새로운 기능의 유전자의 탄생, 유전자 기능의 감소 또는 융합 등을 가능하게 하여 배추 유전자의 기능 분화에 기여했을 것으로 사료되는데, 배추류 식물의 진화 중 배수체 과정의 영향에 대한 심도 있는 연구가 진행되어야 할 것이다.

## 배추 유전체 연구 소재

### BAC 유전자은행과 BAC 말단 염기서열

식물 유전체 연구에는 다양한 원천소재가 반드시 필요한데, BrGSP (The *B. rapa* Genome Sequencing Project)에서 생산 사용된 소재들을 표 1에 정리하였다. 특히 중요한 BAC 유전자은행은 2008년까지 총 5세트를 제작하였으며, 이들의 전체 길이는 배추 유전체의 53 배에 해당한다. BAC 유전자은행 제작에는 제한효소 *EcoRI*, *BamHI*, *HindIII*, 그리고 *Sau3AI*이 사용되었다.

BrGSP 참여국은 이들 BAC 유전자은행의 클론 146,688 개를 사용하여 총 260,637개의 BAC 말단 염기서열을 생산하였는데, 총 염기서열 양은 약 203 Mbp에 달한다. BES 서열의 분석을 통하여 배추 유전체 구조의 특징을 파악할 수 있다. BES의 BLAST 분석 결과, BES의 약 25%가 동원체 또는 동원체 주변의 반복서열인 것으로 나타났다. 또한 약 15%의 BES는 다른 전이인자 혹은 반복서열로 나타났기 때문에 염색체의 FISH 분석 결과와 종합해보면 배추의 이질염색질 영역은 약 40%이고 유전체 염기서열 분석의 실제 대상인 유전자를 함유하는 진정염색질 영역은 약 60%인 것으로 추정되었다 (Yang et al. 2005; Lim et al. 2007). BES를 배추의 ESTs 및 애기장대의 CDS (coding sequences)와 비교한 결과 BES의 약 11%가 단백질을 암호화하는 유전자 (protein-coding genes)였다. BAC 염기서열을 조사했을 때 단백질 암호화 유전자가 평균 1.1-1.3 kbp 크기였으므로 배추 유전체에는 약 45,000-53,000개의 유전자가 있을 것으로 추산되었는데, 이는 Yang 등 (2006)의 추산과 비슷하다.

배추와 애기장대 유전체 구조의 유사성으로 볼 때 배추의 BES와 애기장대 유전체 서열을 비교하여 배추 유전체 해독을 위한 seed BAC을 선발하는 것은 매우 효과적인 방법이다. 따라서 91,000개의 배추 BES를 애기장대 유전체 염기서열에 BLAST 분석하여 단일 BAC 클론의 양말단 염기서열이 애기장대에서 동일선상에 500 kbp 이

**Table 1** Summary of the genomic resources available for the *B. rapa* Genome Sequencing Project

Resources	Material	Characteristics	Number <sup>a</sup>	Reference <sup>b</sup>
BAC library	5 libraries		234,544 clones (53.4X)	
KBrB	cv. <i>Chiifu/BamHI</i>	avr. insert size 120 kbp	55,296 clones (12.5X)	KBGP
KBrE	cv. <i>Chiifu/EcoRI</i>	avr. insert size 139 kbp	23,040 clones (6X)	
KBrH	cv. <i>Chiifu/HindIII</i>	avr. insert size 120 kbp	56,448 clones (12.8X)	
KBrS1	cv. <i>Chiifu/Sau3AI</i>	avr. insert size 100 kbp	55,296 clones (10.5X)	
KBrS2	cv. <i>Chiifu/Sau3AI</i>	avr. insert size 132 kbp	46,464 clones (11.6X)	
BAC-end sequence	146,688 clones		260,637 reads (203 Mbp)	KBGP, NCBI
KBrB	55,296 clones	both end single-pass sequence	97,912 reads (75 Mbp)	KBGP, NCBI
KBrE	23,040 clones		43,168 reads (33 Mbp)	KBGP
KBrH	50,688 clones		88,951 reads (73 Mbp)	KBGP, NCBI
KBrS1	6,144 clones		8,117 reads (5 Mbp)	KBGP, NCBI
KBrS2	11,520 clones		22,489 reads (17 Mbp)	KBGP
Genetic map	3 reference maps			
JWF3p	cv. <i>Jangwon</i>	F <sub>2:3</sub> population 134 lines	1,287 cM, 545 markers	Kim et al. 2006
CK	cv. <i>Chiifu</i> × cv. <i>Kenshin</i>	DH F <sub>2</sub> population 78 lines	1,182 cM, 556 markers	Choi et al. 2007
VCS	cv. <i>VCS3</i> × cv. <i>SR5</i>	DH F <sub>2</sub> population 91 lines	837 cM, 355 markers	KBGP
Physical map				
Build 2	67,468 BAC fingerprints	HICF map	1,428 contigs (717 Mbp)	Mun et al. 2008
Expressed sequence tag	33 cDNA libraries of cv. <i>Chiifu</i>	cDNA single-pass sequence	152,253 ESTs (91 Mbp)	KBGP
Microarray	2 NimbleGen chips			
KBGP-24K	24,000 unigenes	60mer, 6 probes/gene		Lee et al. 2008
KBGP-50K	32,000 unigenes, 17,000 CDS	60mer, 7 probes/gene		KBGP
BAC sequence	KBrB, KBrE, KBrH, KBrS clones	BAC shotgun sequence	Ongoing <sup>c</sup>	KBGP, NCBI

<sup>a</sup> Total sequence length or genome coverage is represented in parenthesis. Genome coverage was estimated based on the haploid genome equivalent of *B. rapa* as 529 Mbp

<sup>b</sup> KBGP, The Korea *B. rapa* Genome Project (<http://www.brassica-rapa.org>). NCBI, the National Center for Biotechnology Information

<sup>c</sup> As of December 2008, 886 BAC sequences have been deposited in NCBI

내의 간격에 올바른 방향으로 염기서열 유사성을 보일 경우, 이들 BAC 클론들을 선발하고 이들로부터 애기장대 서열 기준으로 최소 중첩되면서 애기장대의 유전자 지역을 대부분 커버할 수 있는 600 여개의 배추 BAC 클론을 seed BAC (염색체별 염기서열의 기준점)으로 선정하였다 (Mun et al. 2009). 비교유전체 방법으로 선정된 seed BAC은 모두 “phase 2” - (염기서열 결정 후 조립한 contig의 방향과 contig 사이의 gap을 알며 contig의 길이

합이 원래 BAC 클론의 90% 이상) - 이상 수준으로 염기서열을 분석하였고, 이 염기서열에 근거하여 SSR (simple sequence repeat) 및 IBP (intron based polymorphism) 마커들을 개발하고 유전지도에 seed BAC의 위치를 정하였다. 염기서열이 분석된 seed BAC은 모두 핵산지문분석 (fingerprinting)하여 물리지도의 contig에 포함시켜 clone-by-clone 염기서열을 진행할 때 다음 염기서열을 결정할 BAC 클론 선발에 이용하였다.

## 유전지도

고밀도 유전지도 작성은 식물 유전체 연구에 매우 중요한 기본 요소인데, 유전체 염기서열의 기준점을 제공하거나 염기서열 진행을 점검하는데 반드시 필요하다. 또한 최근의 차세대 염기서열 (Next Generation Sequencing, NGS) 방법에 의한 유전체 해독에서도 NGS 서열의 미비함을 보완하는 중요한 요소로 필수적이다. 배추에서는 다양한 유전집단으로부터 작성된 수종의 유전지도가 보고되었다. 초기에는 주로 혼성화 (hybridization) 마커들이 이용되었지만, 최근에는 사용하기 간편하고 비용이 저렴한 PCR (polymerase chain reaction) 마커들이 주로 개발되어 사용된다.

최근에 3종의 배추 (*B. rapa* ssp. *pekinensis*) 유전지도가 배추 유전체 프로젝트와 관련하여 보고되었다. 그중 하나는 장원 (cv. *Jangwon*) 배추의 F<sub>2:3</sub> 유전지도로서 초기에 520개 RFLP 마커와 25개 PCR 마커를 갖는 총 길이 1,287 cM의 지도였다 (Kim et al. 2006). 비록 이 장원 유전지도는 염기서열 분석 대상인 지부 (cv. *Chiifu*)와 계통은 다르지만 유전체 해독 초기의 완성도 높은 배추 유전지도였기 때문에 핵형분석을 위한 마커 제공, seed BAC의 염색체 상의 위치 결정, 새로 개발된 분자마커의 추가 등 배추 유전체 연구에 중요하게 사용되었다. 따라서 우리는 장원 유전지도에 유전자 표적 (gene-targeting) 또는 SSR 마커 등 500개 이상의 PCR 기반의 STS 마커를 추가하여 장원 유전지도를 유전체 해독에 이용하기 충분하도록 보강하였다. 2007년 말까지 총 513개의 염기서열이 분석된 seed BAC 클론이 장원 유전지도 작성에 사용되었는데, 이는 1,311 cM에 905개 유전좌위 (genetic loci)를 포함한다. 두 번째 유전지도는 MBGP에서 배추의 표준 유전집단으로 사용하고 있는 지부 × 권심 (cv. *Chiifu* × cv. *Kenshin*) 교배집단의 유전지도인데, 이 CK 집단은 F<sub>1</sub> 소포자를 배양 고정된 doubled haploid 78 계통으로 구성되었다. CK 유전지도는 주로 AFLP 마커를 포함하여 총 556개의 다양한 마커로 이루어졌고 총 길이는 1,182 cM 이다 (Choi et al. 2007). 세 번째 VCS 유전지도는 병저항성, 속잎색 등의 중요 형질을 포함하는 cv. *VCS3*와 cv. *SR5*의 doubled haploid 집단으로 작성되었는데, 장원 유전지도에는 포함되지 않은 BAC 클론 225개의 위치를 포함하고 있다 (<http://www.brassica-rapa.org>). 따라서 VCS 유전지도는 장원 유전지도에서 나타날 수 있는 유전적 재조합 (recombination)이나 분리 (segregation)의 불균형을 보완하는 역할을 할 것이다. 현재 3종의 유전지도에서 공통적으로 사용된 마커들을 이용하여 통합된 유전지도를 작성하는 연구가 진행 중이다.

이러한 유전지도들의 고밀도 분자마커와 유전체 염기서열은 작물의 중요한 양적/질적 형질, 예를 들어 병저항

성, 개화시기, 또는 유용한 이차대사산물 생산 등과 연관된 마커의 개발은 물론 관련 유전자의 탐색과 기능연구에 크게 활용될 것으로 기대된다. 또한 배추와 애기장대를 비롯한 추가적인 배추류 유전체 염기서열의 생산과 축적은 비교유전체적 방법을 통해 관심있는 orthologous 유전자들을 쉽게 동정 분리할 수 있는 기반을 제공하여 유전자의 기능과 발현을 연구하고 농업 및 관련 산업적으로 중요한 분자마커와 형질전환체를 개발하는데 크게 기여할 것이다.

## 물리지도

유전체 전체 길이를 포괄하는 물리지도의 제작은 성공적인 clone-by-clone 염기서열 분석을 위하여 반드시 필요하다. 잘 구축된 물리지도는 염기서열 분석을 위한 효율적인 BAC 클론 선발과 분석된 클론들을 순서에 맞게 연속적인 염기서열로 연결하는데 중요한 정보를 제공한다. 유전체 해독에서 물리지도의 유용성은 인간 (Consortium 2001), 애기장대 (Marra et al. 1999), 벼 (Chen et al. 2002), 그리고 *M. truncatula* (Mun et al. 2006) 유전체 프로젝트에서 충분히 입증되었다. 이러한 물리지도는 과거에는 BAC 클론들의 제한효소 절단 패턴을 아가로스 젤에서 읽고 FPC software package를 이용하여 contig로 조립함으로써 제작하였다 (Soderlund et al. 2000). 그러나 이 방법은 전기영동 및 제한효소 밴드 읽기가 수작업으로 이루어지기 때문에 속도와 정확성에서 한계가 있다. 한편, 최근에는 제한효소 밴드를 형광표지하여 자동염기서열장치 (automatic capillary sequencer)로 분석하는 방법이 고안되었다 (Gregory et al. 1997; Ding et al. 2001; Luo et al. 2003; Xu et al. 2004). 이 새로운 형광표지 자동분석 (Fluorescence-labeled capillary electrophoresis) 방법에는 3-enzyme 방법과 high-information content fingerprinting (HICF) 방법이 있는데, 각각 type IIS restriction enzymes 또는 SNaPshot labeling techniques을 사용한다. 이 방법은 종전의 아가로스 전기영동 보다 수행능력과 정확도, 분석속도 면에서 훨씬 우수한데, 최초의 HICF 물리지도는 옥수수에서 보고되었다 (Nelson et al. 2005).

우리는 배추 유전체의 물리지도를 SNaPshot 방법으로 제작하였다 (Mun et al. 2008). BAC 유전자은행 3종으로부터 총 99,456개 BAC 클론을 선정하고 (배추 유전체 크기의 약 22.5 배 길이), 5종의 제한효소 (*EcoRI*, *BamHI*, *XbaI*, *XhoI*, and *HaeIII*)를 동시에 사용하여 절단한 후 4가지 형광색의 SNaPshot 시약으로 3'-end를 표지하여 ABI 3730xl 자동염기서열 분석장치로 제한효소별 단편의 크기를 분석하였다. 분석 결과 93%에 해당하는 총 93,689 BAC 클론의 핵산시퀀스가 성공적으로 분석되었는데, 이중에서 이질염색질의 반복서열을 포함하는 26,221 BAC 클론의 데

이터를 제외하고 contig를 제작하였다. Contig 제작에 사용된 진정염색질을 포함하는 고품질의 67,468 BAC 클론은 배추 유전체의 약 15.2 배 길이인데, 그 결과 1,417개 contig를 얻었다. 이 결과를 바탕으로 contig 구조의 신뢰성을 검증하여 최종적으로 1,428개 contig를 제작하였는데, contig의 평균 길이는 512 kbp로서 전체 길이는 배추 유전체의 1.3배에 해당하는 717 Mbp 였다. 이 contig들에는 핵산지문분석한 전체 BAC 클론의 약 15%에 달하는 다수의 Q 클론 (밴드의 50% 이상이 consensus map과 일치하지 않는 클론)이 포함되어 있었으나, 3개의 특별히 큰 contig에 전체 Q 클론의 48.3% 정도가 포함되었다. 따라서 이들 3개 contig를 중복오류 (false-positive overlaps)로 제외하면 나머지 contig에 포함되는 Q 클론은 약 7.7%가 되며, 이 정도의 수치는 catfish (7.3%) (Quiniou et al. 2007)와 옥수수 (11%) (Nelson et al. 2005)의 HICF 물리지도에서와 비슷한 수준이다.

물리지도 contig 내의 BAC 클론 242개는 유전지도의 분자마커 315개와 연관되어 있었기 때문에 이들 BAC 클론은 배추의 염색체별 염기서열 분석에 직접 사용될 수 있었다. 이렇게 유전지도와 연관된 contig의 총 길이는 배추 유전체의 약 30% 길이에 해당하는 약 160.7 Mbp에 달했다. 그러나 유전지도와 연관된 물리지도 contig가 전체 유전체를 충분한 수준으로 커버하지 못하고, 각 contig의 클론 중첩도 (depth) 역시 아직 낮기 때문에 물리지도의 보강이 필요했다.

#### 발현 유전자와 전사체 분석

유전체 염기서열 후 주석달기 (annotation)와 유전자 기능 연구를 목적으로 배추의 발현 유전자 (EST) 152,253개의 염기서열을 분석하였다. EST 분석을 위하여 배추의 다양한 조직과 발달 단계에서 그리고 병 처리 및 수분 스트레스 처리 후 등 총 33종의 cDNA 유전자은행을 제작하고 각 cDNA 유전자은행별로 분석되는 EST의 중복성을 고려하여 EST 데이터를 생산하였다. 2006년까지 생산된 EST 127,144개의 자료는 NCBI에 등록공개 하였다. 분석된 EST는 Cap3 및 TGICL assembly program 사용, 97% 일치도를 기준으로 조립했을 때 39,095개의 단일 조합 유전자서열 (unigene)로 정리되었다. 이는 추정되는 배추 유전자 수의 약 74-87% 정도였는데, 이 중에 16,898개는 조립된 (tentative consensus sequence) 서열이었고 22,199개는 단일서열 (singleton) 이었다. 배추 unigene의 유전자 ontology 분석이나 길이 분포 분석을 애기장대와 비교했을 때 특정 영역으로의 편중은 발견되지 않았다. 또한 배추의 unigene과 애기장대의 유전체 정보를 비교했을 때, 85%의 unigene이 애기장대와 일치하였는데, 이는 애기장대 유전자 전체의 약 75%에 해당하였다. 일치하는 배추 unigene 들은 편중되지 않게 애기장대 염색체에 골고루 매핑되는

양상을 보였다. 한편 애기장대와 일치하지 않는 15%에 해당하는 5,747개의 unigene은 배추에 특이적인 유전자를 반영하고 있을 것으로 생각된다. 분석된 EST의 염기서열을 보완하여 총 21,409개의 완전장 (full-length) 유전자를 확보하였는데, 이러한 자료와 소재는 배추 특이의 유전자 발현과 기능 연구는 물론 다른 배추류 작물의 유전자 기능분석 연구에 기여할 것으로 전망된다.

분석된 cDNA의 염기서열 정보를 이용한 유전자칩 (microarray)은 조직과 발달, 생물적/비생물적 자극에 따른 유전자의 발현 패턴을 조사하는데 널리 사용되어 왔다. 배추에서 2 종, 즉 KBGP-24K와 KBGP-50K, 유전자칩이 각 유전자 당 6개 (KBGP-24K) 혹은 7개 (KBGP-24K) 표지(probe)를 갖도록 NimbleGen 방식으로 제작되었다. 24K 칩은 20개 cDNA 유전자은행의 127,144 ESTs로 조립된 23,932 unigene을 반영하고, 50K 칩은 약 8,500개의 추가적인 unigene과 seed BAC 염기서열에서 분석된 17,454개 CDS를 합쳐 약 50,000개의 유전자 그룹을 반영하여 제작되었다.

24K 칩을 이용하여 조직 및 발달단계별, 생물적/비생물적 스트레스 처리 등 다양한 10 종의 전사체 분석이 이루어졌다 (Lee et al. 2008). 비생물적 스트레스의 경우 저온 (4°C), 염 (250 mM NaCl), 그리고 건조 (air-dry) 처리하고 48시간 동안 정상과 비교하여 5배 이상 전사량이 증가하거나 감소하는 반응을 보인 유전자는 각각 저온 417개 (1.7%), 염 202개 (0.8%), 건조 738개 (3.1%)였다. 이들 유전자의 발현 패턴의 변화는 각각의 조건에서 RT-PCR (reverse transcriptase PCR) 방법으로 확인하였다. 이러한 비생물적 스트레스에 반응하는 유전자 중에는 56개의 전사조절인자 (transcription factors)와 60 종의 스트레스 공통 반응 유전자가 포함되어 있었다.

#### 기능유전체 연구와 생물정보

농촌진흥청 국립농업과학원의 배추 유전체 웹 사이트 (<http://www.brassica-rapa.org>)는 BAC 클론 염기서열, 유전자 예측을 위한 임시적인 주석달기, 유전지도와 마커 정보, WebFPC로 보이는 물리지도, cDNA 염기서열, 그리고 다양한 전사체 분석 정보 등을 제공하고 있다. BAC과 cDNA 유전자은행의 클론들도 이 웹 사이트를 통하여 주문할 수 있다. 충남대학교의 배추과소재은행 (KBGRB, <http://www.brassica-resource.org>)은 배추 자원의 데이터베이스를 운용하며 다양한 배추 계통의 종자, RI (recombinant inbred) 계통을 포함하는 유전집단, 그리고 DNA 클론들을 보존 분양하는 역할을 수행하고 있다.

경희대학교 식물분자유전학 실험실은 유전자 기능 연구를 위하여 T-DNA 삽입 돌연변이 배추를 다수 생산하고 있으며, MBGP의 TILLING (targeting induced local lesions in genomes) Consortium (<http://www.brassica.info>)은 reverse

genetics의 소재로 다수의 배추 돌연변이 집단을 확보하고 서비스하고 있다. TILLING 돌연변이 집단은 주로 EMS (ethylmethane sulfonate) 처리에 의하여 제작되었는데, *Chiifu* (한국), *rapid-cycle* (미국), 그리고 *R-O-18* (영국) 계통이 사용되었다. 배추의 유전 및 유전체 자료는 MBGP의 TILLING을 이용한 배추 유전자 기능 연구에 귀중한 정보를 제공할 것이다.

## 배추 유전체 염기서열

### 유전자 지역 염기서열 분석 전략

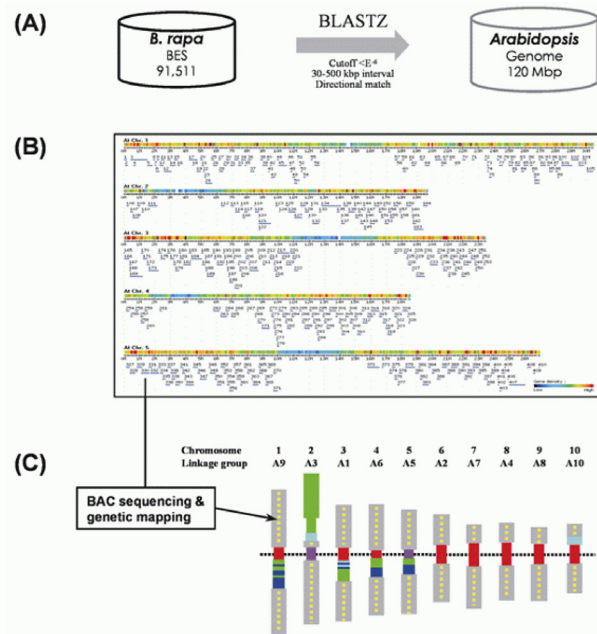
배추과 작물은 동서양을 막론하고 경제적, 사회적으로 중요한 작물일 뿐만 아니라 식물과학의 관점에서도 복잡한 배수체 진화와 분화 연구를 위하여 애기장대 이후의 모델로서 차지하는 비중이 크기 때문에 배추의 유전자 지역의 완성도 높은 염기서열 및 구조 분석은 의미가 크다. 유전체 해독에는 clone-by-clone 방법과 WGS 방법이 각각의 장단점을 가지고 병행되어 왔는데, 우리는 (엄밀히 말해 BrGSP는 협의에 의하여) 배추 유전체의 3배수체 구조의 복잡성 때문에 시간과 비용면에서 다소 불리하지만 분석 결과의 정밀성이 훨씬 높은 clone-by-clone 유전체 해독 전략을 선택하였다. Clone-by-clone 전략은 통상적으로 평균 120 kbp 이상의 BAC 클론으로 이루어진 완성도 높은 최소중첩경로 (minimum tiling path)를 기반으로 진행되는데, 이는 유전지도와 물리지도 및 세포유전학적 분석을 종합하여 만들 수 있다.

한편, Yang 등 (2006)은 배추에는 특별히 유사한 (homologous) 염색체 지역이 많고, 이들 지역에서 선택된 BAC 클론들의 염기서열을 분석한 결과, 매 3-4 kbp 마다 유전자가 존재하는 높은 밀도의 유전자 분포를 보일 뿐만 아니라 각 지역 간의 대응 유전자의 염기서열 유사성이 82% 이상으로 매우 높다는 보고를 하였다. 따라서 중기와 태사기 염색체를 사용한 다양하고 정밀한 FISH 분석을 통하여 일부 염기서열 만으로는 확정할 수 없는 BAC 클론들의 물리적 위치를 실험적으로 검증하고 결정하는 과정이 반드시 필요하다. 또한 양배추 유전체에 대한 WGS 염기서열 0.44X를 분석한 결과 전이인자 염기서열이 예상 밖으로 많이 생성되기도 하였다 (Zhang and Wessler 2004; Ayele et al. 2005). 따라서 clone-by-clone 전략을 통해 배추 유전체를 해독할 경우 효율성과 정확성이 높을 것으로 사료된다.

### 비교유전체 분석에 근거한 seed BAC 선발

Clone-by-clone 전략은 각 염색체별 seed BAC을 기점으로

로 서열을 신장하여 연결하는 전략이므로 우선 seed BAC을 다수 확보하는 것이 중요하다. 애기장대의 유전자 지역이 배추 유전체에 기본적으로 3배수화되어 분포한다는 사실에 근거하여, 배추의 BES와 애기장대의 유전체 서열을 비교하여 애기장대 진정염색질 영역을 최대한 커버하면서 최소 중첩하도록 배추 BAC 클론들을 선발하고, 이들을 분자마커를 이용하여 배추 염색체에 위치시켜서 seed BAC으로 선발하였다 (그림 2). 이러한 전략을 바탕



**Fig. 2** Schematic representation of the comparative tiling sequencing method to select seed BAC clones from the euchromatic regions of *B. rapa*, using in silico matching of BESs with known *Arabidopsis* sequences. (A) Based on the knowledge of the sequence level collinearity between *Arabidopsis* and *Brassica*, BAC-end sequences of *B. rapa* ssp. *pekinensis* cv. *Chiifu* were BLAST analyzed against *Arabidopsis* genome sequence and the corresponding BACs were selected based on the pair-end best hits within 30-500 kbp interval. (B) Selection of minimally overlapping 589 BAC clones based on *in silico* comparative allocation on the *Arabidopsis* genome. A total of 4,317 *B. rapa* BAC clones with an average insert size of 120 kbp were mapped in silico onto the euchromatic regions of the *Arabidopsis* genome. On *in silico* comparative allocation, 589 minimally overlapping BAC clones were mapped onto the comparative positions in the corresponding *Arabidopsis* genome sequence, then were selected and sequenced as phase 2 and scattered on the euchromatic regions of the 10 *B. rapa* chromosomes as seed BACs by a combination of genetic and cytogenetic mapping. Mixed-color bar represents each chromosome of *Arabidopsis* and blue bars depict *B. rapa* BAC contigs tiled on the counterparts of *Arabidopsis* genome. (C) Based on the sequence of each BAC clone, we developed STS markers and anchored the clones on the genetic maps. Overall, 513 sequenced BAC clones were mapped onto the 10 linkage groups and provided to each sequencing center of the BrGSP for completion of the chromosome sequencing. Detailed numbers of BAC clones on each chromosome are denoted in Table 3



으로 약 46,000개 배추 BAC 클론의 BES를 애기장대와 BLASTZ 분석하였다. 이 때 각 BES가  $< E^{-6}$  조건에서 가장 유사성이 높은 애기장대 염기서열 한 곳만 선정하도록 한 후, BAC 클론의 양쪽 BES가 애기장대의 동일 염색체에서 30-500 kbp 이내에 올바른 방향을 갖는 클론 4,647개를 선발 정리하였다 (Yang et al. 2005). 이들 BAC 클론은 애기장대 유전자 지역 약 93 Mbp를, 즉 전체 유전체의 약 78% 또는 유전자 지역의 약 92%에 해당하는 부분을 커버하였다. 선발된 BAC 클론을 대상으로 BAC-FISH 및 BES를 이용한 STS mapping을 통하여 배추 염색체 상의 위치를 결정하면 염기서열 시작을 위한 seed BAC을 확보하게 된다. 배추 유전체는 애기장대에 비하여 약 70% 정도로 위축되었기 때문에 평균 120 kbp 크기의 배추 BAC 클론은 애기장대 약 165 kbp에 해당하므로 이론적으로 약 600개의 배추 BAC 클론은 애기장대 유전자 지역 유전체 약 100 Mbp를 커버할 수 있다. 2008년 중반까지 우리는 589개의 최소 중첩되는 BAC 클론을 선발하고 염기서열 분석하여 배추 염색체에 mapping 하였다 (Mun et al. 2009). 한편 분석된 BAC의 mapping에는 추가적인 FISH 분석과 물리지도 contig 정보가 이용되기도 하였다 (그림 2). 이들 seed BAC 염기서열의 총 길이는 65.8 Mbp 였는데, 실제로는 애기장대의 유전자 영역 75.3 Mbp를 커버하였다. Seed BAC 및 물리지도와 BES 정보를 통합 제공하는 배추 유전자 영역의 contig 160 Mbp는 각 염색체별로 seed BAC 으로부터 염기서열을 신장하여 완성하는데 사용되었다. 분석된 seed BAC의 염기서열 자료는 국립농업과학원 배추 유전체 웹사이트에서 열람할 수 있다.

유전자 특성

염기서열분석이 완료된 seed BAC 589개의 염기서열은 NCBI의 HTGS (High Throughput Genome Sequence) 데이터베이스에 등록하였다. Seed BAC 염기서열을 배추의 cDNA 염기서열로 최적화시킨 FGENESH (www.softberry.com) 프로그램을 사용하여 단백질을 암호화하는 유전자 구조를 예측하였다 (표 2). 배추 유전자는 구조적으로 양배추의 것들과 매우 유사하였다 (Town et al. 2006). 그러나 배추 유전자의 평균 길이는 1.6 kbp로 애기장대의 2.2 kbp보다 짧았는데, 그 이유는 유전자 당 평균 엑손 (exon)의 개수가 애기장대가 배추보다 1개 정도 많고, 엑손과 인트론 (intron)의 평균길이도 애기장대가 배추보다 10-20 bp 정도 길기 때문이라고 판단된다. 한편 비교된 유전체 염기서열에서의 유전자 밀도는 배추와 애기장대가 각각 4.2 kbp와 4.5 kbp 마다 한 개의 유전자가 존재하는 것으로 나타나 배추의 유전자 분포가 애기장대 만큼 조밀하였다 (Mun et al. 2009). 추정된 배추의 단백질 암호화 유전자들로 공개된 데이터베이스 자료와 유사성 분석을 하

였을 때 약 18%의 배추 유전자들은 기 보고된 자료와 전혀 유사성이 없었다. 이상의 결과들은 EST unigene이 보여준 결과 및 FLC 주변의 염기서열 분석 결과 (Yang et al. 2006)와 일치한다. 배추 유전체 프로젝트의 선도 그룹인 국립농업과학원은 현재 EST와 완전장 유전자 자료를 이용하여 유전체 염기서열에 자동으로 주석을 달고 유전자의 구조와 기능을 예측할 수 있는 파이프라인을 개발 중인데, 완성 후에는 배추의 염기서열과 유전자에 대한 고품질의 정보를 연구자와 개발자들에게 제공할 것이다.

배추 유전체 염기서열

국제컨소시엄 BrGSP는 그림 3과 같이 염색체를 국가별로 할당하고 각 염색체의 유전자 지역의 염기서열을 clone-by-clone 방법으로 해독하기로 합의하였다. 또한 BrGSP는 각 BAC 클론의 염기서열 분석은 phase 2 수준으로 수행할 것을 결정하였다. 이와 더불어 BrGSP 참가국

Table 2 Comparison of overall composition of annotated protein coding genes in the *B. rapa* BAC sequences and *A. thaliana* genome

Feature	<i>B. rapa</i>	<i>A. thaliana</i> <sup>a</sup>
Average gene size (kbp)	1.6	2.2
Number of exons per gene	4.7	5.8
Intron size (bp)	141	165
Exon size (bp)	225	268
Average gene density (kbp/gene)	4.2	4.5
Overall G/C content (%)	35.2	35.9
Exons	46.3	44.1
Introns	32.6	32.6
Intergenic regions	31.3	32.9

<sup>a</sup> Statistics for *A. thaliana* is based on the annotation of TAIR7 from the *Arabidopsis* Information Resource website ([http://www.arabidopsis.org/portals/genAnnotation/genome\\_snapshot.jsp](http://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp))

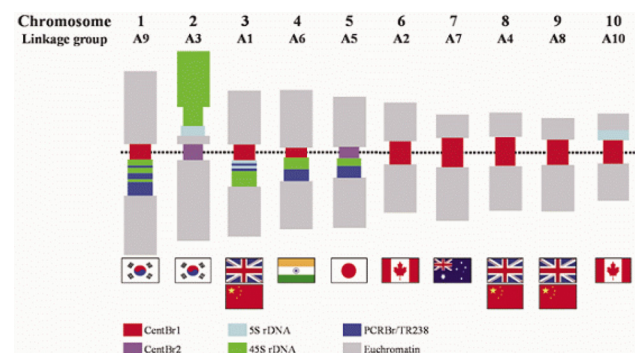


Fig. 3 Idiogram of *B. rapa* chromosomes represented by euchromatic arms and heterochromatic repetitive sequence blocks. Each chromosome was assigned to a sequencing center in one of the seven countries of the BrGSP consortium

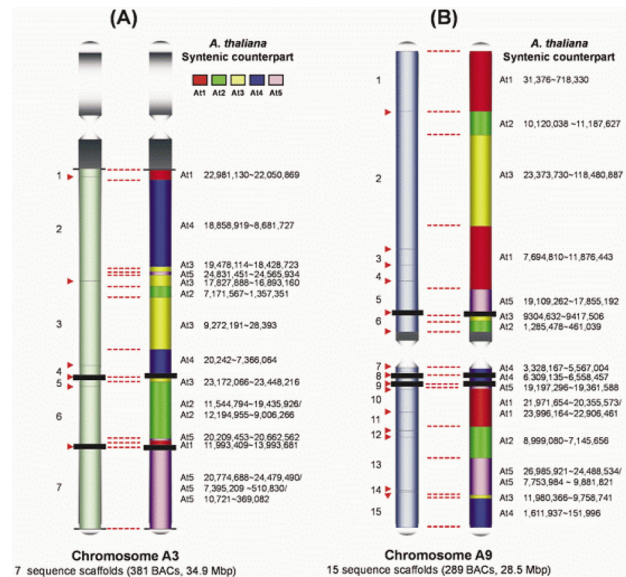
**Table 3** Current status of *B. rapa* BAC sequencing by NAAS, as of March 2010

Chromosome	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Unanchored	Total
Anchored seed BACs	56	62	94	14	53	58	50	47	58	34		526
Extended BACs	1	5	441	6	1	14	7	7	291	12		785
Total	57	67	535	20	54	72	57	54	349	46	543	1,854

의 필요에 따라서 gap이 없이 완벽한 염기서열 해독 (phase 3)을 할 수도 있다. BAC 염기서열 해독은 해당 클론을 삽입단편의 길이가 5 kbp 이내인 짧은 sublibrary로 만든 후 원래 BAC 클론 크기의 약 7-8X 양을 ABI 3730xl 을 사용하여 shotgun 방식으로 분석하여 수행하였고, Phred/Phrap/Consed software package를 이용하여 염기서열 contig를 제작하였다.

2009년 말까지 배추 BAC 클론 886개 (한국 681, 영국 200, 호주 5)의 염기서열 총 107 Mbp가 NCBI의 HTGS 데이터베이스에 등록되었다. 한국은 배추 염색체 중에서 가장 길이가 긴 염색체 두 개 A3와 A9의 염기서열 해독을 BAC 클론 884개를 분석하여 완성하였다 (표 3). 이들 BAC 클론을 최소 중첩되도록 연결한 결과 염색체 A3는 BAC 클론 381개와 스캐폴드 (scaffold) 7개의 34.9 Mbp, 염색체 A9는 BAC 클론 289개와 스캐폴드 15개의 28.5 Mbp로 마무리 되었다 (그림 4). 이는 추정되는 염색체의 유전자 지역의 길이에 비하여 A3는 90%, A9는 85% 완성된 것으로 추정된다. 따라서 스캐폴드 사이에 남아있는 장단 길이의 gap, 동원체 주변 및 염색체 말단의 telomere 부분과 유전자 지역의 경계 부분에 대한 세밀한 분석이 추가된다면 두 염색체의 염기서열 완성도는 더 높아질 것이다.

염기서열 분석 기술과 기기의 개발은 발전 속도가 매우 빠른 분야로서 프로젝트가 진행되는 중에 분석 능력이 크게 향상되고 분석 비용도 크게 감소한 차세대 염기서열 분석기기가 상용화 되었다. 따라서 BrGSP 참여국에서는 GS-FLX 염기서열 분석기를 사용하여 10개 이상의 BAC 클론을 함께 분석하여 clone-by-clone 분석의 단점을 보완하고자 시도를 하였다. 이에 한국도 GS-FLX와 ABI 3730xl을 병행한 BAC 클론 분석을 수행하여 약 20% 정도의 비용을 절감할 수 있었지만, 서열 통합 작업이 복잡하고 GS-FLX의 pyrosequencing에서 나타나는 poly-mononucleotide의 해독 불능 등의 문제점도 노출되었다. 한편, 중국의 소재화웨이연구소 (Institute of Vegetables and Flowers, IVF)와 북경유전체연구소 (Beijing Genome Institute, BGI)는 Illumina 방식의 WGS를 통하여 배추 유전체의 70X 이상의 염기서열을 생산하고 성공적으로 조립하였다. 현재 한국 국립농업과학원과 중국의 IVF는 양측의 염기서열, 유전지도, 물리지도 등의 정보를 일괄하여 배추 전체 유전체 서열에 대한 초안의 공동완성을 협약하고 진행 중이다.



**Fig. 4** Current sequencing status of the two longest chromosomes, A3 and A9. BAC clones were sequenced as phase 2 and assembled into sequence scaffolds. (A) A total of 381 minimally tiled BAC clones generated seven sequence scaffolds spanning 34.9 Mbp for chromosome A3. (B) 15 scaffolds spanning 28.5 Mbp were assembled from 289 minimum tiled BAC clones for chromosome A9. Syntenic counterpart regions of the *A. thaliana* genome identified by sequence comparison were also represented in the colored boxes, along with position information. Red arrow heads indicate major gaps between scaffolds

## 결론

우리는 배추의 완성된 표준 유전체 염기서열과 배추류 작물들의 유전체계를 이해하는데 필요한 정보를 조만간 제공할 예정이다. 배추 유전체 염기서열의 가장 중요한 수혜자는 배추류를 포함하는 쌍떡잎 작물의 유전학, 육종학, 분자생물학은 물론 유전체학, 생물정보학을 연구하는 과학자와 개발자들이 될 것이다. 많은 수의 배추 발현유전자 수집, 완전장 cDNA 세트, 그리고 완성도 높은 유전체 염기서열은 농산업적으로 중요한 유전자들을 동정/분리하고 기능을 연구하는데 활용될 수 있다. 다수의 분자마커가 잘 배열된 고밀도 유전지도, STS 마커와 관련 염기서열 정보는 이미 중요 형질과 연관된 분자마커의 개발 및 유전지도 기반 유전자 클로닝 (map-based cloning)에 사용되고 있다. 배추 유전체 염기서열 정보는 양배추,

유채, 브로콜리 등 경제적으로 중요하며 다양한 배추류 작물의 비교유전체 연구 및 유전자 개발에 확대 이용될 전망이다. 예를 들어, 배추의 유전체 정보와 소재를 이용하여 양적형질 (quantitative trait loci, QTL) 분석 또는 연관 매핑 (association mapping)으로 유채에서 가치 있는 식물유래 화합물 (phytochemical) 관련 유전자를 찾았다 (Meng, Huazhung, pers. comm.). 또한 현재의 배추 유전체 정보는 양배추나 유채 등의 배추류 작물의 유전체 연구에서 염기서열이나 스캐폴드 조립을 위하여 우리가 애기장대와 비교유전체 기술을 이용한 것보다 훨씬 유용하게 이용될 수 있다. NGS 기술 등 최근의 기술적 진보는 GS-FLX Titanium (Roche), Genome Analyzer (Illumina), 그리고 SOLiD (Life Technologies) 기기를 이용한 유전체 연구를 가능케 하였는데, 배추의 표준 유전체 염기서열은 배추과 작물의 유전체 연구는 물론 유전과 육종, 분자생물학 연구에 NGS 기술의 적용을 가속화시킬 전망이다.

## 사 사

우리는 한국 배추 유전체 프로젝트에 참여했던 모든 분들에게 감사드립니다. 이 글의 내용은 국립농업과학원에서 수행한 연구과제 (05-1-12-2-1, PJ006759, PJ006691)와 농촌진흥청의 바이오그린21 사업에서 지원한 연구과제 (2005 0301034438)의 결과 중 일부를 포함하고 있습니다.

## 인용문헌

- Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc Natl Acad Sci USA* 95:13073-13078
- Ayele M, Haas BJ, Kumar N, Wu H, Xiao Y, Van Aken S, Utterback TR, Wortman JR, White OR, Town CD (2005) Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*. *Genome Res* 15:487-495
- Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. *Am J Bot* 93: 607-619
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433-438
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KF, Rogers J, Quétiér F, Oldroyd GE, Debelle F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van de Peer Y, Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc Natl Acad Sci USA* 103:14959-14964
- Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higinbottom S, Phimpilai J, Phimpilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. *Plant Cell* 14:537-545
- Choi HK, Mun J-H, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. *Proc Natl Acad Sci USA* 101:15289-15294
- Choi SR, Teakle GR, Plaha P, Kim JH, Allender CJ, Beynon E, Piao ZY, Soengas P, Han TH, King GJ, Barker GC, Hand P, Lydiat DJ, Batley J, Edwards D, Koo DH, Bang JW, Park BS, Lim YP (2007) The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project. *Theor Appl Genet* 115:777-792
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286:2468-2474
- Ding Y, Johnson MD, Chen WQ, Wong D, Chen YJ, Benson SC, Lam JY, Kim YM, Shizuya H (2001) Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using type IIS restriction endonucleases. *Genomics* 74:142-154
- Economic Research Service USDA. (2008) Vegetables and melons outlook: <http://www.ers.usda.gov/Publications/VGS/Tables/World.pdf>
- Gregory SG, Howell GR, Bentley DR (1997) Genome mapping by fluorescent fingerprinting. *Genome Res* 7:1162-1168
- Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. *Theor Appl Genet* 90:157-165
- International Human Genome Sequencing Consortium (2001) A physical map of the human genome. *Nature* 409: 934-941
- Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in Brassicaceae. *Ann Bot* 95:229-235
- Kim JS, Chung TY, King GJ, Jin M, Yang TJ, Jin YM, Kim HI, Park BS (2006) A sequence-tagged linkage map of *Brassica rapa*. *Genetics* 174:29-39
- Koo DH, Plaha P, Lim YP, Hur Y, Bang JW (2004) A high-resolution karyotype of *Brassica rapa* ssp. *pekinensis* revealed by pachytene analysis and multicolor fluorescence in situ hybridization. *Theor Appl Genet* 109:1346-1352
- Kulikova O, Geurts R, Lamine M, Kim DJ, Cook DR, Leunissen J, de Jong H, Roe BA, Bisseling T (2004) Satellite

- repeats in the functional centromere and pericentromeric heterochromatin of *Medicago truncatula*. *Chromosoma* 113:276-283
- Kwon SJ, Kim DH, Lim MH, Long Y, Meng JL, Lim KB, Kim JA, Kim JS, Jin M, Kim HI, Ahn SN, Wessler SR, Yang TJ, Park BS (2007) Terminal repeat retrotransposon in miniature (TRIM) as DNA markers in *Brassica* relatives. *Mol Genet Genom* 278:361-370
- Lee SC, Lim MH, Kim JA, Lee SI, Kim JS, Jin M, Kwon SJ, Mun JH, Kim YK, Kim HU, Hur Y, Park BS (2008) Transcriptome analysis in *Brassica rapa* under the abiotic stresses using *Brassica* 24K oligo microarray. *Mol Cells* 26:595-606
- Lim KB, de Jong H, Yang TJ, Park JY, Kwon SJ, Kim JS, Lim MH, Kim JA, Jin M, Jin YM, Kim SH, Lim YP, Bang JW, Kim HI, Park BS (2005) Characterization of rDNAs and tandem repeats in the heterochromatin of *Brassica rapa*. *Mol Cells* 19:436-444
- Lim KB, Yang TJ, Hwang YJ, Kim JS, Park JY, Kwon SJ, Kim J, Choi BS, Lim MH, Jin M, Kim HI, de Jong H, Bancroft I, Lim YP, Park BS (2007) Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species. *Plant J* 49:173-183
- Lukens LN, Quijada PA, Udall J, Pires JC, Schranz ME, Osborn TC (2004) Genome redundancy and plasticity within ancient and recent *Brassica* crop species. *Biol J Linn Soc Lond* 82:665-674
- Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82:378-389
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe *Brassicaceae*. *Genome Res* 15:516-525
- Marra M, Kucaba T, Sekhon M, Hillier L, Martienssen R, Chinwalla A, Crockett J, Fedele J, Grover H, Gund C, McCombie WR, McDonald K, McPherson J, Mudd N, Parnell L, Schein J, Seim R, Shelby P, Waterston R, Wilson R (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nat Genet* 22:265-270
- Mun J-H, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook DR (2006) Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* 172:2541-2555
- Mun J-H, Kwon SJ, Yang TJ, Kim HS, Choi BS, Baek S, Kim JS, Jin M, Kim JA, Lim MH, Lee SI, Kim HI, Kim H, Lim YP, Park BS (2008) The first generation of a BAC-based physical map of *Brassica rapa*. *BMC Genom* 9:280
- Mun J-H, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol* 10:R111
- Nelson WM, Bharti AK, Butler E, Wei F, Fuks G, Kim H, Wing RA, Messing J, Soderlund C (2005) Whole-genome validation of high-information-content fingerprinting. *Plant Physiol* 139:27-38
- O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J* 23:233-243
- Park JY, Koo DH, Hong CP, Lee SJ, Jeon JW, Lee SH, Yun PY, Park BS, Kim HR, Bang JW, Plaha P, Bancroft I, Lim YP (2005) Physical mapping and microsynteny of *Brassica rapa* ssp. *pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5. *Mol Genet Genom* 274:579-588
- Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA (2003) Structure and evolution of cereal genomes. *Curr Opin Genet Dev* 13:644-650
- Quiniou SMA, Waldbieser GC, Duke MV (2007) A first generation BAC-based physical map of the channel catfish. *BMC Genom* 8:40
- Rana D, van den Boogaart T, O'Neill CM, Hynes L, Bent E, Macpherson L, Park JY, Lim YP, Bancroft I. (2004) Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J* 40:725-33
- Schmidt R, Acarkan A, Boivin K (2001) Comparative structural genomics in the Brassicaceae family. *Plant Physiol Biochem* 39:253-262
- Soderlund C, Humphray S, Dunham I, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* 11:934-941
- The *Arabidopsis* Genome Initiative (2004) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815
- Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick M, Bancroft I (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18:1348-1359
- U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* 7:389-452
- Xu Z, Sun S, Covalada L, Ding K, Zhang A, Wu C, Scheuring C, Zhang HB (2004) Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality. *Genomics* 84:941-951
- Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M,

- Park JY, Lim MH, Kim HI, Lim YP, Kang JJ, Hong JH, Kim CB, Bhak J, Bancroft I, Park BS (2006) Sequence-level analysis of the diploidization process in the triplicated *FLOWERING LOCUS C* region of *Brassica rapa*. *Plant Cell* 18:1339-1347
- Yang TJ, Kim JS, Lim KB, Kwon SJ, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Kim SH, Lim YP, Park BS (2005) The Korea *Brassica* Genome Projects: a glimpse of the *Brassica* genome based on comparative genome analysis with *Arabidopsis*. *Compar Funct Genom* 6:138-146
- Yang TJ, Kwon SJ, Choi BS, Kim JS, Jin M, Lim KB, Park JY, Kim JA, Lim MH, Kim HI, Lee HJ, Lim YP, Paterson AH, Park BS (2007) Characterization of terminal-repeat retrotransposon in miniature (TRIM) in *Brassica* relatives. *Theor Appl Genet* 114:627-636
- Zhang X, Wessler SR (2004) Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc Natl Acad Sci USA* 101:5589-5594
- Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. *Nucl Acid Res* 32:2023-2030