

## 배추 유전체연구의 현황과 전망

최수련<sup>1</sup>, 박지영<sup>2</sup>, 박범석<sup>2</sup>, 김호일<sup>2</sup>, 임용표<sup>1\*</sup>

<sup>1</sup>충남대학교 농업생명과학대학 식물자원학부, <sup>2</sup>농업생명공학연구원 배추제넨팀

## Korea Brassica Genome Project: Current Status and Prospective

Su Ryun Choi<sup>1</sup>, Jee Yong Park<sup>2</sup>, Beom-Seok Park<sup>2</sup>, Ho-Il Kim<sup>2</sup>, and Yong Pyo Lim<sup>1\*</sup>

<sup>1</sup>Department of Horticulture, College of Natural Science, Chungnam National University, Daejeon 305-764, Korea

<sup>2</sup>National Institute of Agricultural Biotechnology, RDA, Suwon 441-707, Korea

**ABSTRACT** *Brassica rapa* is an important species used as a vegetable, oil, and fodder worldwide. It is related phylogenetically to *Arabidopsis thaliana*, which has already been fully sequenced as a model plant. The 'Multinational Brassica Genome Project (MBGP)' was launched by the international *Brassica* community with the aim of sequencing the whole genome of *B. rapa* in 2003 on account of its value and the fact that it has the smallest genome among the diploid Brassica. The genome study was carried out not only to know the structure of genome but also to understand the function and the evolution of the genes comprehensively. There are two mapping populations, over 1,000 molecular markers and a genetic map, 2 BAC libraries, physical map, a 22 cDNA libraries as suitable genomic materials for examining the genome of *B. rapa* ssp. *pekinensis* Chinese cabbage. As the first step for whole genome analysis, 220,000 BAC-end sequences of the KBrH and KBrB BAC library are achieved by cooperation of six countries. The results of BAC-end sequence analysis will provide a clue in understanding the structure of the genome of *Brassica rapa* by analyzing the gene sequence, annotation and abundant repetitive DNA. The second stage involves sequencing of the genetically mapped seed BACs and identifying the overlapping BACs for complete genome sequencing. Currently, the second stage is comprises of process genetic anchoring using communal populations and maps to identify more than 1,000 seed BACs based on a BAC-to-BAC strategy. For the initial sequencing, 629 seed BACs corresponding to the minimum tiling path onto *Arabidopsis* genome were selected and fully sequenced. These BACs are now anchoring to the genetic map using the development of SSR markers. This information will be useful for identifying near BAC clones with the seed BAC on a genome map. From the BAC sequences, it is revealed that the *Brassica rapa* genome has extensive triplication of the DNA segment coupled with variable gene losses and rearrangements within the segments. This article introduces the current status and prospective of Korea Brassica Genome Project and the bioinformatics tools possessed in each national team. In the near future, data of the genome will contribute to improving *Brassic*as for their economic use as well as in understanding the evolutionary process.

### 서 론

생명공학관련 연구와 산업화에 가장 기본적이고 필수적인

정보를 제공하는 유전체 연구는 21세기 생명공학연구의 발전에 있어서 핵심 연구분야로 국가 및 민간 차원에서 체계적이고 지속적인 연구가 이루어 지고 있다. 현재까지 전세계적으로 인간을 비롯한 고등동물 (*Arabidopsis thaliana*, rice, *Drosophila melanogaster*)종과 미생물종 (*Mycoplasma genitalium*,

\*Corresponding author Tel 042-821-5739 Fax 042-822-6712

E-mail: yplim@cnu.ac.kr

*Treponema pallidum*, *Borrelia burgdorferi*)이 분석완료 되는 등 그 속도가 가속화되고 있다.

우리 나라에서도 유전체 연구에 대한 집중적 연구가 이루어 지고 있으며 이중에서도 특히 고등동식물로서 국내에서 유전체 분석연구를 독창적으로 수행할 수 있는 대상작물로서 배추를 중심으로 한 배추과작물이 그 대표적 연구대상으로 2003년부터 본격적으로 전체염기서열분석이 진행되고 있다.

배추과 (십자화과)는 A, B, C genome으로 대표되는 *B. rapa* (2n=20 AA genome), *B. nigra* (2n=16, BB genome), *B. oleracea* (2n=18, CC genome) 의 이배체종과 이들의 이형 배수체인 *B. carinata* (BBCC), *B. juncea* (AABB), *B. napus* (AACC)로 구성되어 있다. 배추과내에는 약 3,350여종이상의 다양한 종으로 분포하고 있으며, 쌍자엽식물 유전체연구의 모델 식물로서 이미 전체 염기서열 분석이 완료된 아기장대 (*Arabidopsis*)를 위시하여, 배추 (*Brassica rapa*), 무 (*Raphanus sativus*), 양배추 (*B. oleracea*), 유채 (*B. napus*), 갯 (*B. juncea*) 등 전세계적으로 주요 채소류, 식물성 유지, 향신료, 사료 등 유용한 경제작물들이 포함되어 있다.

농업적으로 중요하게 이용되고 있는 배추는 *Brassica rapa* (syn. *campestris*)에 속하며, 한국배추는 그 중에서도 *pekinensis* 아종에 속하며 유전체 크기는 약 529 Mb로 아기장대플과는 약 4배의 크기를 가지고 있어 유전체 연구를 하는데 있어서 기존의 주요 경제작물들과 비교하였을 때, 그 중에서는 유전체 크기가 상대적으로 작아 비교적 경제적으로 연구가 가능할 뿐 만 아니라, 다양한 유용한 작물이 같은 배추과에 속함으로써 이를 통한 비교유전체연구 효과도 기대할 수 있겠다.

또한 배추는 우리 나라 4대 채소작물 중의 하나이며 한국 고유 식품인 김치의 주원료로서 국민의 식생활에 없어서는 안 되는 주요 채소이다. 우리 나라의 배추생산은 약 5만 ha에서 약 25만 톤을 연간 생산한다. 또한 배추는 국내 시장만 해도 약 6000억원에 달하고 있는 주요 경제작물이다. 그리고 최근에는 김치산업이 발전하여 1억불 이상의 대외 수출을 기록하고 있으며 급속한 증가추세에 있다.

역사적으로도 우리 나라는 우장춘 박사의 U's triangle (종의 합성이론)에 의거 배추 및 관련 십자화과 (배추과) 식물의 재능 진화이론의 개발 등을 포함하여 전통적으로 세계 최고 수준의 고전 유전학적, 육종학적 연구가 진행되어 있어 유전체연구 진행 및 결과 활용에 세계적으로도 유리하며 독창적인 위치에 있다.

또한 배추는 식물유전체 연구의 모델인 아기장대와 매우 밀접한 관계를 가지고 있어 이미 전체 재능분석이 완료된 아기장대의 유전정보를 이용하는데 매우 큰 장점을 가지고 이와 더불어 진화의 연구에도 좋은 결과를 가져올 수 있는 작물이다.

이러한 여러 가지 장점을 인정받아 2003년 1월 미국 San

Diego에서 개최된 International Brassica Genome Project Consortium에서 한국의 배추를 우선 배추과 식물 중 최초로 전 염기서열을 분석하기로 하였으며, 오늘의 배추재능연구에 이르게 되었다.

본 논문에서는 1990년대부터 준비된 배추의 유전분석집단 양성 및 BAC library의 작성, 유전자지도 및 물리지도의 작성, BAC-end 염기서열분석, 전 염색체 염기서열분석에 이르는 그 동안의 국제 배추과 재능콘소시엄의 활동과 국내 배추재능연구의 진행을 검토하고, 앞으로의 전망에 대하여 서술하고자 한다.

### 유전체 크기와 핵형분석

배추속에서 6가지 종이 세계 농업시장에서 재배되고 있다. 3가지 종, *B. rapa* (배추), *B. oleracea* (양배추), *B. nigra* (흑겨자) 는 이배체이며 각기 AA (2n=20), BB (2n=16), CC (2n=18) 유전체로 분류된다. 나머지 3가지 종인 *B. napus* (유채), *B. carinata* (겨자), *B. juncea* (잎겨자)는 이형 배수체로 AACC (2n=38), BBCC (2n=34), AABB (2n=36) 유전체로 구성된다. 이형배수체는 자연 상태에서 종간교잡이 일어나 새로이 합성된 종임을 1935년 우장춘 박사 (U, 1935) 가 염색체 관찰을 통해 밝혔다 (Figure 1). Johnston et al. (2005)은 이들 유전체의 실제적인 크기를 보고하였으며 이중 *B. rapa* (배추)는 1.1 pg/2C (529 Mbp/1C)으로 가장 작았다.

생물 종, 개체, 세포 고유의 염색체 구성을 핵형이라 하는데, 그 특성은 세포의 핵분열 중의 중기 또는 후기에 나타나는 염색체의 형태, 크기, 수로 나타낸다. 다양한 핵형 분석 방법 중에서 FISH는 *B. rapa* (배추종) 처럼 metaphase에서 염색체의 형태 및 크기가 비슷한 식물에서 효율적인 핵형분석을 가능하게 한다. Fukui et al. (1998)는 pro-metaphase 단계에서

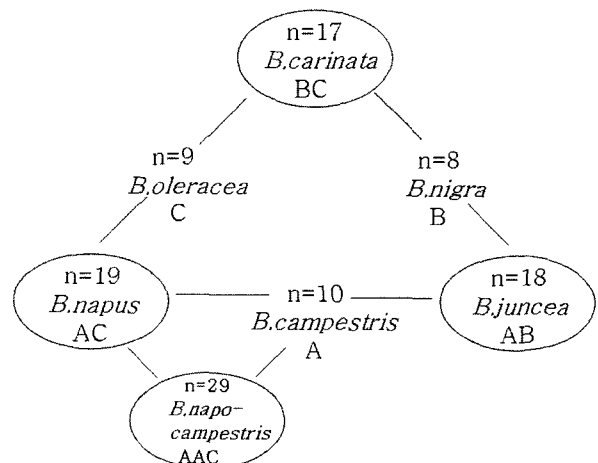


Figure 1. The Triangle of U; Diagrammatic representation of the genomic relations among the species in Brassica (1935)

*B. rapa*, *B. oleracea*, *B. nigra* 간에 45S rDNA의 위치와 개수가 다르다는 것을 보였고, 이로써 염색체 1, 2, 4, 5, 7을 구분한 바 있다. Snowdon et al. (2002)은 선행 연구의 염색체 순서를 기본으로 하여 5S rDNA를 probe로 실험하여 1, 4, 5, 7번 염색체의 동원체 근방에 25S rDNA가 특이적으로 존재하는 것을 확인하고 2, 7번 염색체에 25S rDNA와 5S rDNA가 공존하는 것을 보였으며 5S rDNA의 위치로 염색체 10번을 구분한 바 있다. Koo et al. (2004)는 현 유전체 연구에서 기본재료로 사용되고 있는 지부계 배추를 재료로 하여 mitotic metaphase 및 pachytene 상태의 염색체에서 5S rDNA, 45S rDNA, 및 C11-350H (반복 DNA 서열)을 probe로 하여 분자세포유전학 연구를 한 결과 mitotic metaphase 염색체의 길이는 1.46  $\mu\text{m}$ 에서 3.30  $\mu\text{m}$ 로 보고 하였다. 반복 DNA 서열을 probe로 이용하여 2, 5번 염색체를 제외한 나머지 8개의 염색체에 모두 존재하는 것을 보고 하였으며, 45S rDNA를 이용하여 5개 염색체 (1, 2, 4, 5, 7), 5S rDNA를 이용하여 3개의 염색체 (2, 7, 10)를 구분 하였다. Lim et al. (2005, 2006)은 centromere의 주요 구성을 차지하는 176bp tandem repeat과 CRB, PCRBr 등을 밝히고 이들의 분포를 통해 10개의 염색체를 명확히 구분할 수 있음을 보여주었다.

#### 유전체의 배수성, 전 유전체의 중복과 배추 특이적 중복

3,500여종으로 구성된 배추과는 애기장대와 분화된 후 monophyloitic group으로 진화했다. 이들의 유전체 크기는 529~696Mb로 그 형태만큼이나 다양하다 (Rich 1991, Johnston et al. 2005). 이러한 다양성은 원시 조상에서 진화하면서 유전체의 배수성, 유전자 혹은 염색체 단편의 중복, 결실, 반복염기서열에 기인한다.

*Brassica* 유전체 구조가 어떻게 진화하였는가에 대한 가설은 분분하다. 원시 조상이 4배체였고 연속적인 segmental 중복이 일어났을 것이라는 가설이 있으나 (Lukens et al. 2004) 가장 주목 받는 설은 애기장대와 분화된 후 유전체 양이 3배수가 된 6배체가 되었고 이후 결실과 삽입 등의 과정을 거쳐 유전적으로 안정된 이배체 형태로 진화했을 것이라는 것이다 (Schmidit et al. 2001, Parkin et al. 2003). 초기에 RFLP 방법이 이러한 가설을 뒷받침했다. RFLP 마커를 이용한 분석으로 유전체내에 중복유전자가 많이 있음을 볼 수 있었는데 이는 진화과정 동안 대단위의 중복 현상이 있었으며 염색체 구조의 재변환이 있었음을 시사했다. 최근에는 애기장대 유전체의 8.7 Mb BAC contig를 이용해 염색체의 상동부위를 배추과에 속하는 21개의 종에서 살펴본 결과 (Lysak et al. 2005)와 *B. napus*에서 1,000개 이상의 RFLP 마커를 이용하여 염기서열 수준에서 애기장대와 상동적인 위치를 비교한 결과 (애기

장대 거의 모든 유전체가 중복되고 재배열 하여 *B. napus* 유전체를 구성하고 있었다)가 보고 됨으로써 더 확고해 졌다. 물리지도기술을 이용하여 *B. oleracea*, *B. rapa*, 그리고 *B. napus* 유전체의 특정 유전자군 부위에서 microstructure를 비교연구한 결과에서도 이배체 배추 유전체가 삼중복 되어 있다는 것을 확인했다 (O'Neill and Bancroft 2000, Rana et al. 2004, Park et al. 2005). 또한 Yang et al. (2006)은 배추 유전체에 존재하는 삼중복 된 유전자를 이용하여 진행중인 역동적인 진화과정을 보여 주었다. 삼중복된 개화조절에 관련된 유전자 부위를 이용하여 염기서열 수준에서 원시 6배체 배추가 현재의 이배체 배추로 진화되어 오는 기본 메커니즘을 분석 보고한 것이다. 5개의 *B. rapa* BAC 클론과 애기장대에서 새롭게 중복된 두 부위 (5번 염색체의 3.0-3.3 Mb 부위와 25.8-26.2 Mb)를 비교해 본 결과 삼중복된 배추의 각각 DNA 단편은 애기장대의 비교부위 보다 0.6-0.8 배 짧았는데 이는 다른 염색체 부위에서도 동일한 결과로 나타났다 (Yang et al. 2005). 전체 유전자 수를 비교해 보면 삼중복으로 인해 *B. rapa* 에서 애기장대보다 1.7배 증가한 것을 볼 수 있었고, *B. rapa*의 전체 유전자수는 애기장대의 유전자 수를 기본으로 하여 추정해 볼 때 49,000개정도로 예상되며 *B. rapa* 유전체의 15%는 애기장대와 분화된 후 배추에 새로이 삽입된 유전자임을 제시했다. 삼중복 되는 부위에서 약 40%의 유전자가 단일 카피로 남아 있고 상당수의 것이 아직 그 기능이 밝혀지지 않았다는 점에서 그 해답을 찾을 수 있으리라 기대한다.

#### *Brassica rapa*의 유전체 연구 재료

재배되고 있는 다양한 배추 종들 중에서, *B. rapa* ssp. *pekinensis* 는 A-genome의 전형적인 대표로 고려되고 있다. 이배체 배추간에 유전체 크기가 529 Mb로 가장 작고 (Johnston et al. 2005) 다음과 같이 유전체 연구에 필요한 기본 재료를 갖추고 있었기 때문이다.

#### 유전자 지도 작성을 위한 집단 및 유전자 지도

유전자 지도 작성을 위한 집단을 개발하고 양친간의 다형성을 나타내는 마커를 개발하여 연관군 그룹을 할당하는 과정인 유전자 지도 작성은 유전체를 분석하기 위한 기본 단계이다. 우리는 내혼계인 '지부'와 '권심' 배추 간의 교배로 얻은 F<sub>1</sub> 식물을 약배양 하여 만든 78계통의 doubled haploid (CKDH)집단과 F<sub>8</sub> 세대까지 진전한 recombinant inbred (CKRI) 집단 251계통이 확보되었으며, 이 집단은 현재 MBGP의 전 염기서열 작성을 위한 표준 유전자지도작성용 집단으로 이용되고 있다.

현재 DH 집단을 이용하여 10개의 연관그룹으로 구성된 표준 유전자 지도가 있는데, 총 556 마커로 구성되었으며 전체 길이는 1,181 cM이다. 10개의 연관그룹은 *B. napus* 표준연관군지도에 사용된 SSR 마커를 사용함으로써 *B. napus* 유전체에서 *B. rapa* 유전체를 근원으로 하는 R1-R10으로 명명하였다 (Parkin et al. 1995, Suwabe et al. 2002, Lowe et al. 2004). 보유한 모든 마커는 보다 많은 계통으로 구성된 RI 집단에 적용될 것이며 보다 정밀한 유전자 지도 작성을 위해 지속적으로 분자 마커가 개발 중이다. 국내에서 장원배추 모,부분을 사용한 F<sub>3</sub>집단을 이용한 유전자지도도 개발된 바 있다. 이 집단에서는 520개의 RFLP 마커가 개발되어 있으며, *B. napus*의 21개의 SSR 마커를 사용하여 R1-R10으로 명명한 바 있다 (Kim et al. 2006)

이렇게 개발된 유전자지도는 경제적으로 중요한 형질을 tagging 할 수 있고 육종에 있어 마커를 이용한 선발 (MAS) 및 유전자집적 (gene pyramiding)과 유용한 유전자의 클로닝에 사용할 수 있다. 추가로 유전체의 구성, 작물 종의 진화 및 관련된 종간의 비교분석을 가능하게 한다. 또한 전 유전체 염기서열 분석을 위한 초석으로 사용할 수 있다.

#### BAC (Bacterial Artificial Chromosome) Library

세균인공염색체 (BAC) library는 거대한 genomic 단편을 운반할 수 있다는 점에서 구조 및 기능 유전체 연구에서 빠질 수 없는 재료이다. 내흔계인 ‘지부’ 배추의 잎을 재료로 하여 거대한 genomic 단편을 운반하고자 2가지 제한효소 (*Hind*III, *Bam*HI)를 이용해 세균인공염색체 (BAC) library (KBrH, KBrB)를 만들었고, 이들은 각각 56,592개와 50,688개의 클론으로 구성되어 있다 (Park et al. 2005). 배추의 유전체 크기를 529Mb로 추정할 적에 이 두 library는 배추 유전체의 약 24배를 포함하고 있다.

#### EST 및 Microarray

농업생명공학연구원의 ‘배추제넘탐’과 ‘BioGreen 21 program (BG21)’ 연구단의 공동연구를 통해 잎, 뿌리, 배축, 자방 및 꼬투리와 같은 배추의 각기 다른 조직과 잎과 뿌리에 염류 및 저온 처리를 한 것을 재료로 하여 22가지의 cDNA library를 만들었다. 이들을 이용하여 EST 분석을 실시하였으며 평균 길이가 575 bp으로 된 104,914개의 클론으로 구성되어 있다. 이들을 이용한 DNA chip 제작 목적과 현재까지 분석된 EST contig에 대한 결과는 (<http://www.brassica-rapa.org>)에서 볼 수 있다.

현재 ‘Korea Brassica Genome Resource Bank (KBGRB)’가 한국과학재단의 지원으로 설립되었고 이는 MBGP의 목적과

부합되는 구조유전체연구의 활성화를 위해 국제공동체에 *B. rapa*의 유전자 및 유전체 자원공급을 지원하고 있다. 이는 BAC 및 cDNA library외에도 BAC-end 염기서열과 EST 염기서열에서 개발된 microsatellite 및 EST 마커를 보유하고 있다. KBGRB에 대한 상세한 정보는 <http://www.brassica-resource.org>에서 볼 수 있다. 또한 그동안 수집된 배추과 작물의 유전자원의 가치를 인정받아 농촌진흥청으로부터 유전자원 관리기관으로 선정되어 활동 중이다.

#### BAC Fingerprinting

두 BAC 라이브러리의 지문분석 (DNA Fingerprinting)을 하고 있다. 2006년 말까지 모두 110,000 BAC 클론을 지문분석한다는 목표를 가지고 추진하고 있으며 결과들은 <http://www.brassica-rapa.org>에서 볼 수 있다. 현재 KBrH 라이브러리의 56,592 BAC 클론이 *Hind*III 제한효소 절단방법으로 지문분석 (DNA Fingerprinting) 되었고, 그 결과 3,854의 FPC contigs가 만들어졌으며 그 크기는 전체 유전체의 약 90%에 해당한다. 이러한 물리지도는 각 염색체 별로 BAC-to-BAC 염기서열 분석을 위한 초석으로 사용될 종자BAC을 선별하는데 도움이 될 것이다. 보다 상세한 물리지도 결과를 얻기 위해 KBrH와 KBrB 라이브러리를 SNaPShot 방법으로 분석 중 이다.

#### 유전자 지도와 세포유전학 지도간의 통합 및 유전체 염기서열 분석

*B. rapa*에서 여러 아종을 이용하여 많은 유전자 지도가 작성되어 있다 (McGrath and Quiros 1991, Song et al. 1991, Chyi et al. 1992, Teutonico and Osborn 1994, Nozaki et al. 1997, Kim et al. 2006, Choi et al. unpublished data). 그러나, 이러한 유전자 지도의 연관그룹은 세포유전학 지도의 염색체와 연관성에 대해 발표한바 없다. 그러므로, 유전자 지도에서 연관그룹과 실제 염색체 번호와 연결시키는 것이 구조 유전체 연구에 있어서 기동력이 될 것이다. 유전자 지도와 세포유전학 지도의 통합된 결과가 이미 밀, 보리 및 애기장대에서 보고된 바 있다. 이들은 염색체 결실, 전좌 및 삼염색체성을 이용하였다 (Koornneff and Vanderveen 1983, Kunzel et al. 2000, Sandhu et al. 2001). *B. rapa*는 위와 같은 현상으로 식물체를 유지하기가 어려울 뿐더러 염색체의 크기가 매우 비슷하여 크기나 모양과 같은 방법으로는 구분하기가 어렵기에 다른 방법에 요구된다. 이런 문제점에 Fluorescence in situ hybridization (FISH)가 해결안이 될 수 있으며 이미 감자 (Dong et al. 2000), 콩 (Kulikova et al. 2001)에서 이용되었다. 양배추에서 Howell 등 (2002)은 유전자 지도의 9개의 연관그룹을 FISH 방법으로

BAC클론을 probe으로 사용하여 염색체와 연관지어 보였다. PCR 방법으로 유전자좌 특이적 증폭을 하여 BAC 클론과 유전자 지도를 연결시켰다. 우리는 연관 그룹에 있는 EST, STS, SSR 과 같은 유전자 마커에 해당하는 BAC 클론을 KBrH library에서 선별해 냈고 이들은 유전자지도와 Koo et al. (2004)가 보고한 세포유전학 지도와 통합하기 위해 사용하고 있다.

#### 생물정보학 방법과 배추 유전체 연구를 위한 Database

최근 GenBank에 배추 종의 DNA 염기서열의 축적은 폭발적으로 증가하고 있다. Genoplante Consortium에서 31,011 개의 배추 EST를 제공했고, TIGR에서 300,000개의 양배추 유전체 shotgun sequence를 제공했다. 이러한 염기서열 정보가 증가됨에 따라 보다 효율적인 관리를 위해 database의 개발을 필요했고 현재 각 국에서 운영되고 있는 database의 종류는 다음과 같다.

영국은 Brassica 'A' 와 'C' 유전체의 기본적인 물리지도를 만들기 위해 BAC 라이브러리의 DNA fingerprinting (지문분석)을 하였다. 또한 유전자 probe을 선별하여 애기장대 유전체 염기서열과 앞서 언급한 물리지도를 통합하고자 하였다. Brassica와 애기장대에 관심있는 사용자들에게 현재까지 출간된 자료, 애기장대 유전체 염기서열과 A, C 유전체간에 연결을 위해 사용된 probe와 BAC클론 등의 정보를 제공하기 위해 BBSRC Brassica IGF Project (<http://brassica.bbsrc.ac.uk/IGF/>)를 운영하고 있다. 유럽공동체와 중국은 공동으로 IMSORB: oilseed rape programme (<http://brassica.bbsrc.ac.uk/IMSORB/>)를 운영하면서 다양한 환경 조건에서 경쟁력 있는 oilseed rape 자원 품종을 개발하고 지속적인 신품종개발을 위한 지식을 축적하고 있다. 미국은 애기장대 유전체의 annotation을 높이고, Brassica와 애기장대간에 염기서열 보존 (conservation) 정도를 탐색하기 위해 *B. oleracea* GSS database (<http://www.tigr.org/tdb/e2k1/bog1/TIGR>)을 운영하고 있다. Cold Spring Harbor 그룹에서는 Brassica 유전체의 0.5 - 1배를 목표로 하여 전체 유전체 shotgun sequencing을 하고 있고, 이렇게 확보된 염기서열을 이미 분석이 완료된 애기장대와 비교, 위치 (anchoring)시킴으로써 아직 기능을 알 수 없는 것으로 분류되었던 유전자를 새로 해석 할 수 있을 뿐만 아니라 유전자간의 공통된 혹은 차별화된 부위를 찾아낼 수 있으리라 기대한다. 캐나다는 Brassica와 애기장대 유전체간의 비교를 할 수 있는 browser, AAFC Comparative Genome Viewer (<http://brassica.agr.gc.ca/navigation/viewere.shtml>, SRC)를 운영하고 있다. TIGR에서 기능분석(annotation)한 애기장대 유전체와 관련된 *B. napus* EST (60,000 cDNA의 3'과 5' sequences)를 보여주고 있다. 애기장대 유전체 정보는 'The Institute for Genomic Research'의 website에서 볼 수 있다 (<http://www.tigr.org>).

한국은 [www.brassica-rapa.org](http://www.brassica-rapa.org) (농업생명공학연구원/ 충남대학교/KOBIC)를 운영하면서 통합지도 작성, 1번과 2번 염색체 염기서열 분석, BAC-ends 염기서열 분석, EST와 DNA Chip, 세포유전학, T-DNA와 Transposon Tagging Mutation 연구과제에 대한 결과를 관리하고 있다. JIC (영국)에서는 BAC 라이브러리 screening과 배급을 위해 (<http://jicgenomelab.co.uk/libraries.html>)을 운영하고 있다. 호주에서는 Brassica ASTRA database (Love et al. 2004, 2005), SSR 및 SNP 개발, 유전체 지형도, 유전자 ontology 연구를 위해 PBC Bioinformatics (<http://hornbill.csp.la.trobo.edu.au/>)를 운영하고 있다. Brassica 유전체에 관련된 모든 정보를 집적하여 공유하고자 모든 자료가 <http://www.brassica.info> 에 공개 및 연결 되어 있다.

#### 배추 전 유전체 염기서열 분석 현황

국제 배추과 제넘콘소시엄을 대표해서 구성된 MBGP는 배추 유전체 연구분야에 있어서 연구 공동체의 전략적인 목표를 규정하고 서로간의 협동을 촉진하는 역할을 하고 있다. MBGP 내에 BrGSP (*Brassica rapa* Genome Sequencing Project)가 구성되어 529Mb 크기의 *B. rapa* ssp. *pekinensis* 가 A-genome의 대표 유전체로 전 염기서열을 분석하기로 국제회의에서 결정되었다 (<http://www.brassicagenome.org>). 첫 번째 단계로 KBrH 및 KBrB library의 모든 BAC 클론 107,285개의 BAC-end 염기서열을 분석하기로 하였다. 두 번째 단계는 공동으로 사용하는 유전자 지도에 대략 1000개의 BAC 클론을 마커로 연결하고 이를 염기서열 분석의 초석으로 삼는다. 유전자 지도에 연결된 종자 BAC 클론을 기반으로 하여 연결될 수 있는 BAC 클론을 BAC-end 염기서열 정보를 통해 찾아가면서 전체 유전체의 염기서열을 분석하고자 한다. 이를 위해 배추 유전자 지도의 R3과 R9 (염색체로는 각각 2번과 1번에 해당) 은 대한민국, R1과 R8은 영국과 중국, R2는 호주, R4는 미국, R5, R6, R7은 캐나다, R10은 프랑스와 독일이 염기서열 분석을하기로 결정 하였다.

#### BAC-end Sequence

BAC-end 염기서열은 BAC library의 클론을 한쪽 말단에서 염기서열을 분석한 것이다. BAC 클론의 DNA fingerprinting 과 병행하여 이와 같은 BAC 클론의 말단 염기서열은 유전체 염기서열 분석에 있어 기본적인 골격으로 이용될 수 있다 (Vender et al. 1996). 더욱이 이러한 염기서열은 단백질 코딩 부위, CpG islands 및 전이인자와 microsatellite과 같은 반복되는 DNA 부위들을 포함함으로써 유전체의 구성을 조망하는데 이용 될 수 있다 (Mahairas et al. 1999, Hong et al. 2004)

**Table 1.** Current Status of BAC-end sequencing in Chinese cabbage by MBGP activity

Plates	No. clones	Group	Country
KBrH001 - KBrH015	5,760	NIAB	S. Korea
KBrH016 - KBrH050	13,440	AAFC + NRC	Canada
KBrH051 - KBrH062	4,608	JIC	UK
KBrH063 - KBrH087	9,600	DPI	Australia
KBrH088 - KBrH117	11,520	JIC, Bath, TIGR	UK/USA
KBrH118 - KBrH136	7,296	U. Bielefeld	Germany
KBrH137 - KBrH144	3,072	CNU	S. Korea
KBrB001 - KBrB096	36,864	NIAB, CNU	S. Korea
KBrB097 - KBrB132	13,824	AAFC	Canada

BGP 위원회는 국제공동으로 KBrH, KBrB library의 전체 BAC 클론 107,280개의 양쪽 말단 염기서열 분석을 하기로 결정하였다. 이들 클론의 분석은 대한민국, 호주, 독일, 영국, 캐나다, 미국에서 수행되었으며 2006년 9월 말 말단 염기서열 분석이 완료되었다 (Table 1). 현재까지 분석된 BAC-end 염기서열 분석의 현황 및 염기서열 정보, 이들의 관리 규약은 (<http://www.brassica.info>)에서 볼 수 있다. 이러한 BAC-end 염기서열 정보는 BAC-to-BAC 방법으로 전 유전체 염기서열 분석하고자 하는데 있어 종자 BAC 을 효율적으로 선별하는데 유용하게 이용 될 것이다. In silico 물리지도를 통한 비교분석은 *B. rapa* 염색체에서 chromosome walking 및 종자 BAC 선별과정에 도움을 줄 수 있을 것이다. BAC-FISH와 in silico 물리지도에서 비교분석으로 골라낸 BAC클론의 말단 염기서열을 이용하여 sequence tagged connectors (STC)를 개발, mapping 하는 것은 마커가 없는 부위에서 BAC 클론을 선별하는데 도움이 될 것이다. 염기서열 분석이 끝난 모든 BAC 클론들은 앞으로 STC를 개발하고 최소로 중첩되는 인접한 종자 BAC 클론 선별을 위해 기점으로 작용할 것이다.

#### 각 염색체 염기서열 분석

전 유전체 염기서열 분석을 위한 방법으로 BAC-by-BAC 방법이 선정되었다. 이를 위한 시발점으로 농업생명공학연구원의 ‘배추제농팀’에서는 BAC-end 염기서열을 이용하여 애기장대 유전체와 비교하여 Minimum tiling path로 애기장대 염색체에 상동을 나타내는 BAC 클론들을 선별하고 (Yang et al. 2005) 선별한 629개 모두 염기서열 분석 하였다. 이 모든 정보는 GeneBank 에 공개되어 있다.

이 BAC 클론은 유전체 전 염기서열 분석을 위해 기본 골격이 될 종자 BAC으로서 역할 할 것이다. 또한 배추 유전체 분석의 두 번째 단계 (1,000개의 BAC 클론을 마커로 유전자 지도에 연결하기)를 위해 사용되었다. Minimum tiling path로 선별한 모든 BAC 클론을 유전자 지도에 올리고자 전체 BAC

클론 염기서열에서 반복 염기서열 부위를 이용해 SSR 마커로 만들고자 했다. 현재 331개의 BAC 에서 SSR 마커를 개발 하였으며 이는 331개의 BAC 클론을 유전자 지도와 연결 할 수 있음을 의미한다. 331개 중에 101개의 마커가 표준 유전자 지도에 연결 되었고 나머지 마커들에서 분석이 수행 중 이다.

#### 결 론

배추 종은 채소나 유지 및 사료작물로 이용됨으로써 중요한 작물 중 하나이다. 이러한 작물에서 유전체 연구는 같은 과에 속하는 애기장대 유전체가 전부 염기서열 분석됨에 따라 탄력을 받았다. 유전자지도를 위한 집단, 공동으로 사용될 유전자 지도 및 BAC library와 같이 유전체 연구에 필수적인 재료를 갖추고 있다는 점에서 *B. rapa* ssp. *pekinensis*는 이들 종의 유전체에 관심 있는 연구자들의 주목을 받게 되었다. 현재 국제 배추과 제농 콘소시엄이 *B. rapa*의 전 염기서열 분석이라는 목표를 가지고 추진 중이다. *B. rapa*의 구조 유전체 연구 및 전 염기서열 정보는 다른 배추속에 속하는 경제적 가치를 가진 종들의 유전체 분석의 기본 자료를 제공할 수 있고 이들과 비교 유전체를 함으로써 식물학의 다양한 형태적 변이에 대한 정보도 제공할 수 있으리라 기대한다. 또한 배추 유전체의 특성상 진화과정에서 대부분 작물 유전체의 특징인 다배체화, 재이배체화에 대한 해석을 가능하게 할 것이다. 이런 모든 정보는 궁극적으로 *B. rapa* 뿐만 아니라 상당한 경제적 가치를 지닌 배추 속에 속하는 다양한 작물의 개선에 유용하게 이용 될 수 있으리라 기대한다.

#### 적 요

유전체 연구란 목적하는 유전체의 구조를 밝히고 가지고 있는 모든 유전자의 기능 및 진화과정을 망라하여 이해하고자 하는 것이다. 계통발생학상 애기장대와 연관되어 있는 *Brassica rapa*는 채소, 유지 및 사료로 이용되는 중요한 작물의 하나이다.

*Brassica rapa*의 유전체 연구를 착수하는 데는 적합한 유전학적 재료 및 유전체 재료가 있어야 한다. 우리는 배추 (*Brassica rapa* spp. *pekinensis*)를 재료로 하여 표준 mapping 집단을 개발하여, 78계통으로 구성된 DH집단과 약 250 계통으로 구성된 RI집단을 개발하였다. 2가지 제한효소 (*HindIII*, *BamHI*)를 이용해 세균인공염색체 (BAC) library (KBrH, KBrB)를 만들었고, 이들은 각각 56,592개와 50,688개의 클론으로 구성되어 있다. 또한 배추의 각기 다른 부위를 이용하여 만든 22가지의 cDNA library를 이용하여 평균 575bp의 길이를 가지는 104,914개의 EST 분석을 실시 하였다.

세계적으로 'Multinational Brassica Genome Project (MBGP)' 조직이 구성되었고 배추의 전 염기서열 분석을 하기로 2003년 결정되었다. 그 첫 단계로 104,914개의 BAC 클론의 BAC-end 염기서열분석이 제안되어 2006년 9월 5개국 공동 프로젝트로 추진하여 완성하게 되었다. 이러한 BAC-end 염기서열 분석의 결과는 유전자의 염기서열 해석, 및 풍부하게 존재하는 반복염기서열 DNA를 분석함으로써 배추의 유전체 구조를 이해할 수 있는 실마리를 주었다. BAC 클론의 전체 염기서열분석은, 비록 단편 내에 유전자의 결실이 변화무쌍하게 일어나지만 배추 DNA 단편이 유전체에서 광범위하게 삼중복으로 존재함을 나타냈다. 이러한 BAC-end 염기서열을 아기장대 염기서열에 비교하여 629개의 종자 BAC을 선정하게 되었고, 이들의 염기서열 분석을 완성하였다.

MBGP에서는 2단계로서 배추의 전 유전체 염기서열 분석을 추진하게 되었고, 유전자지도에 위치한 종자 BAC을 이용하여 인접한 BAC 클론을 찾아 염기서열 분석하는 BAC-to-BAC 방법을 추진하고 있으며 8개국에서 참여하여 현재 염기서열 분석을 추진 중 이다. 최근에 각 국에서는 생물정보학기법을 활용한 염기서열 분석 기반에 대하여 많은 토론이 진행되고 있다. 앞으로 다양한 유전체 정보가 축적됨에 따라 배추의 유전체 구조를 이해하고 농업적으로 적용하고자 하는데 기여를 할 것이다.

## 사 사

본 연구는 농촌진흥청 바이오그린21사업과 과학재단 특성화장려사업 (R21-2004-000-10010-0)의 연구 결과의 일부이며 연구비 지원에 감사드립니다.

## 인용문헌

Chyi YS, Hoenecke ME, Sernyk JL (1992) A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). *Genome* 35: 746-757  
Dong F, Song J, Naess SK, Helgeson JP, Gebhardt C,

Jiang J (2000) Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor Appl Genet* 101: 1001-1007  
Fukui K, Nakayama S, Ohmido N, Yoshiaki H, Yamabe M (1998) Quantitative karyotyping of three diploid *Brassica* species by imaging methods and localization of 45s rDNA loci on the identified chromosomes. *Theor Appl Genet* 96: 325-330  
Hong CP, Lee SJ, Park JY, Plaha P, Park YS, Lee YK, Choi JE, Kim KY, Lee JH, Lee J, Jin H, Choi SR, Lim YP (2004) Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. *Mol Gen Genomics* 271: 709-716  
Howell EC, Baker GC, Jones GH, Kearsey MJ, King GJ, Kop EP, Ryder CD, Teakle GR, Vicente JG, Armstrong J (2002) Integration of cytogenetic and genetic linkage maps of *Brassica oleracea*. *Genetics* 161: 1225-1234  
Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in *Brassicaceae*. *Annals of botany* 95: 229-235  
Kim JS, Chung TY, King JG, Jin[lyp3] M, Yang TJ, Jin YM, Kim HI, Park BS (2006) A Sequence-Tagged Linkage Map of *Brassica rapa*. *Genetics* 174: 29-39  
Koo DH, Lim YP, Hurr YK, 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  
Koomneff M, Vanderveen JH (1983) Trisomics of *Arabidopsis thaliana* and the location of linkage groups. *Genetica* 61: 41-46  
Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J* 27: 49-58  
Kunzel G, Korzun L and Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics* 154: 397-412  
Lim KB, De Jong H, Yang TJ, Park JY, Kwon SJ, Kim JS, Lim MH, Kim JA, Jin M, Jin YM (2005) Characterization of rDNAs and tandem repeats in heterochromatin of *Brassica rapa*. *Mol Cells* 19: 436-444  
Lim KB, Yang TJ, Hwang YJ, Kwon SJ, Kim JA, Choi BS, Lim MH, Kim JS, Jin M, Park JY et al. (2006) Characterization of the major centromere and pericentromere retrotransposons in *Brassica rapa* and distribution in the related *Brassica* species. *Plant J*. In press  
Love CG, Batley J, Lim G, Robinson AJ, Savage D, Singh D, Spangenberg GC, Edwards D (2004) New computational tools for *Brassica* genome research. *Comp Funct Genom* 5: 276-280  
Love CG, Robinson AJ, Lim GAC, Hopkins CJ, Batley J, Baker G, Spangenberg GC, Edwards D (2005) *Brassica* ASTRA: an integrated database for *Brassica* genomic research. *Nucleic Acids Res* 33: 656-659

- Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in Brassica crop species. *Theor Appl Genet* 108: 1103-12
- Lukens L, Zou F, Lydiate D, Parkin I, Osborn T (2003) Comparison of a Brassica oleracea genetic map with the genome of *Arabidopsis thaliana*. *Genetics* 164(1): 359-72
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. *Genome Res* 15: 516-25
- Mahairas GG, Wallace JC, Smith K, Swartzell S, Holzman T, Keller A, Shaker R, Furlong J, Young J, Zhao S, Adams MD, Hood L. (1999) Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. *Proc Natl Acad Sci USA* 96: 9739-44
- McGrath JM and Quiros CF (1991) Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B. oleracea*. *Theor Appl Genet* 82: 668-673
- Nozaki T, Kumazaki A, Koba T, Ishikawa K, Ikehashi H (1997) Linkage analysis among loci for RAPDs, isozymes and some agronomic traits in *Brassica campestris* L. *Euphytica* 95: 115-123
- O'Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. alboglabra that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J* 23: 233-43
- 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 kb Gene-Rich Region of *Arabidopsis* Chromosome 4 and Partially Duplicated on Chromosome 5. *Mol Gen Genomics* 274: 579-588
- Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122-1131
- Parkin IA, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the Brassica napus genome. *Genome* 46: 291-303
- Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171: 765-81
- 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
- Rich TCG (1991) Crucifers of Great Britain and Ireland. Botanical Society of the British Isles, London. 336
- Sandhu D, Champoux JA, Bondareva SN, Gill KS (2001) Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. *Genetics* 157: 1735-1747
- Snowdon RJ, Friedrich T, Friedt W, Köhler W (2002) Identifying the chromosomes of the A- and C-genome diploid *Brassica* species *B. rapa* (syn. *campestris*) and *B. oleracea* in their amphidiploid *B. napus*. *Theor Appl Genet* 104: 533-538
- Song KM, Suzuki JY, Slocum MK, Williams PH and Osborn TC (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length loci. *Theor Appl Genet* 82: 296-304
- Suwabe K, Iketani H, Nunome T, Kage T, Hirai M (2002) Isolation and characterization of microsatellites in *Brassica rapa* L. *Theor Appl Genet* 104: 1092-1098
- Teutonico RA, Osborn TC (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa*, and comparison to linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. *Theor Appl Genet* 89: 885-894
- 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
- Vender JC, Smith HO, Hood L (1996) A new strategy for genome sequencing. *Nature* 381: 364-366
- Yang TJ, Kim JS, Lim KB, Kwon SJ, Kim JI, Jin M, Park JY, Lim MH, Kim HI, Kim SH et al. (2005) The Korea Brassica Genome Project: A glimpse of the *Brassica* genome based on comparative genome analysis with *Arabidopsis*. *Comp Funct Genomics* 6: 138-146
- 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

(접수일자 2006년 10월 11일, 수리일자 2006년 10월 21일)