

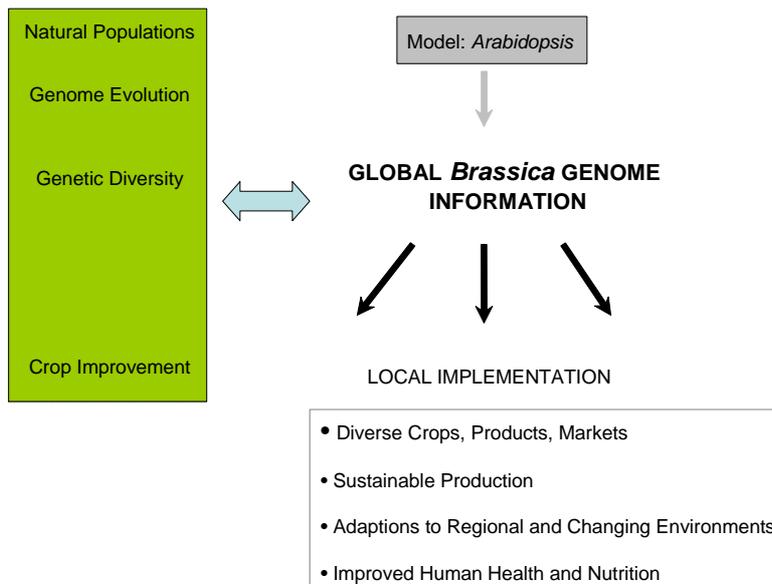
# *Brassica*: Harvesting the Genome, Diversity and Products

A (draft) "White Paper"<sup>1</sup> for

## THE MULTINATIONAL *BRASSICA* GENOME PROJECT

v 1.10

ed. Graham J King, Rothamsted Research, Mar 2005



<sup>1</sup>White Paper: "An educational report, made available to the public, which expounds on a particular (industry) issue."

## **EXECUTIVE SUMMARY**

*Brassica* species play an important role in global agriculture and horticulture, as well as being the closest crop relatives to Arabidopsis. The species are characterised by a wide range of adaptations that have been domesticated into crops including oilseed rape/Canola and swede (*B. napus*); cabbage, cauliflower, broccoli, Brussels sprout (*B. oleracea*); chinese cabbage, pak choi, turnip and oil (*B. rapa*) and mustards (*B. nigra*, *B. juncea*, *B. carinata*). *Brassica* crops contribute both to the economies and health of populations around the world. Crop improvement is a key route to ensuring that continued benefits arise from these foods and plant products. A wide range of genetic and genomic resources are available from *Brassica*, as well as the easy access to information derived from Arabidopsis. Although considerable progress has been made in the genetic analysis of agronomic and related plant traits in *Brassica*, there are currently limitations in identifying and characterising the behaviour of the underlying genes, genomic regulatory networks and associated metabolism. A range of major challenges focus on harnessing genetic diversity through an information-led approach to crop improvement. There are a wide range of valuable traits for which genetic variation exists but where understanding is required. These include improving harvest index and yields in the context of changing climate and need to reduce inputs, optimising quality fit for purpose, and identifying scope for added value through nutritional, prophylactic health or non-food use. Given the unique position of brassicas with their diverse crop types and metabolites, the vision for maximising the benefits from recent advances in genomics can be summed up in terms of "global information, local implementation". Genomic information is the key to exploiting knowledge gained at the level of gene expression, biochemistry, metabolism and physiology. In particular, the *Brassica* research communities need to be able to locate genes in their relevant genomic context, in order to understand regulation and pressures on evolution. For effective crop-based studies it is also essential to be able to navigate between trait and gene, and thus integrate information from agronomy, breeding, genetics and genomics. This will provide a functional understanding of relevant systems to underpin future crop and product development within a sustainable framework.

- 1. Overviews and Challenges**
- 2. The Role of *Brassica* Genomes in Agriculture and the Economy**
- 3. Scientific Status and Unique Features of Brassica species**
- 4. *Brassica* Biodiversity and Crop Improvement**
- 5. Status and Aims of *Brassica* Genome Research**
- 6. A Unique Opportunity from Model to Crops: Harnessing *Arabidopsis***
- 7. Setting Priorities: MULTINATIONAL *BRASSICA* GENOME PROJECT Goals and Objectives**
- 8. Status of *Brassica* Research Programmes around the World**
- 9. Bibliography**
- 10. Technical paper on Sequencing the *B. rapa* Genome**

## 1. OVERVIEW AND CHALLENGES

*Brassica* species play an important role in global agriculture and horticulture, as well as being the closest crop relatives to *Arabidopsis*. The species are characterised by a wide range of adaptations that have been domesticated into crops including oilseed rape/Canola and swede (*B. napus*); cabbage, cauliflower, broccoli, Brussels sprout (*B. oleracea*); turnip, chinese cabbage and pak choi (*B. rapa*) and mustards (*B. nigra*, *B. juncea*, *B. carinata*).

*Brassica* crops contribute both to the economies and health of populations (e.g. *via* anti-oxidants, vitamins, anti-carcinogenic compounds, etc.) around the world. Crop improvement is a key route to ensuring that continued benefits arise from these foods and plant products. *Brassica* is typical of many crop species in having a larger and more complex genome than the model. The genomic relationships are well characterised, as shown in the 'triangle of U', and these have been exploited to understand the basis of chromosome evolution since divergence from a common progenitor shared with *Arabidopsis*. The *Brassica* genomes are particularly 'plastic', with several crop morphotypes within each species. This is the analogous to the situation with dogs, where the diverse morphotypes provide ideal material to study man-directed evolution (artificial selection) and the processes involved in domestication.

A wide range of genetic and genomic resources are available from *Brassica*, as well as the easy access to information derived from *Arabidopsis*. Reference linkage maps, a wide range of QTL relevant to basic processes and crop phenotypes and sequence-based information are all available. In addition, there is increasing emphasis on characterising and utilising the diverse allelic variation present in genetic resource collections.

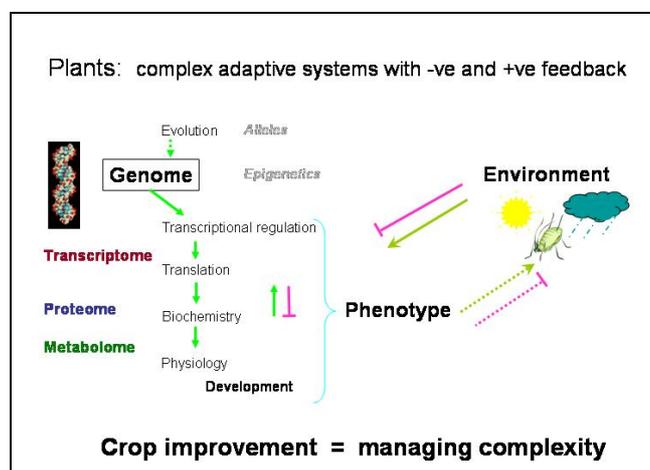
Considerable progress has been made in the genetic analysis of agronomic and related plant traits in *Brassica*. Many of these have a complex quantitative inheritance which is exacerbated by interactions between genes, plant development and the environment. Despite the advantages of using information from *Arabidopsis*, the current challenge remains the need to identify key genes and understand their regulation in crop plants. There is a pressing requirement to resolve recently identified functional loci (major genes and QTLs) in terms of locus-specific copies of candidate or novel genes located in the context of physical map contigs and the emerging complete genome sequence.

### 1.1. The Challenges include:

- Understanding, conserving, accessing and harnessing genetic diversity
- Developing an information-led approach to crop improvement, with associated cost and efficiency benefits
- Identifying traits - adapting to changing climate, markets, land use and energy requirements
  - Increasing marketable yields and harvest index
  - Improved sustainable production, including pest and disease control, nutrient and water use efficiency
  - Diversity of products affecting human health and wealth
- Understanding the genome: intra and inter-genomic polyploidy
- Navigating the genome - from traits to sequence and *vice versa*
- Identifying genetic components of variation
- Global Integration of information and resources in public domain.
  - "Global information, local implementation"

### 1.2. Managing Complexity: Crop Improvement for Yield, Quality, Sustainability and Adaptation to climate change

Higher Plants are complex adaptive systems with overlapping and interacting mechanisms for modulating their own development in response to environmental cues.



## ***Brassica* : Harvesting the Genome, Diversity and Products**

They are characterised by a highly developed secondary metabolism, and multilayered regulatory networks, reflected in complex and diverse genome organisation. Crop domestication, selection and targeted breeding has enabled human civilisation to develop. This continues into the modern age, with increasingly sophisticated technologies available for the interchange and selection of genetic material meeting the needs of different markets and growing environments. Crop improvement is based both on understanding which genes are involved in a phenotype, as well as the degree of environmental variation. Thus the ability to understand the genetic basis and heritability of traits (phenotype) selected in breeding programmes provides the opportunity to deploy novel allelic combinations.

### **1.3. Understanding and Exploring Diversity**

*Brassica* species are characterised by extensive morphological diversity and the ability to adapt to a wide range of habitats and growing environments. There appear to be several primary and secondary centres of diversity, with crop domestication involving further selection of adaptations from northern temperate (sub-arctic?) to tropical regions. Understanding the basis of particular crop morphotypes in the context of the wider range of genetic, biochemical and metabolic variation present within the species will make it possible for breeders of a given crop to utilise beneficial alleles found in other crop types or wild germplasm in an informed and efficient manner.

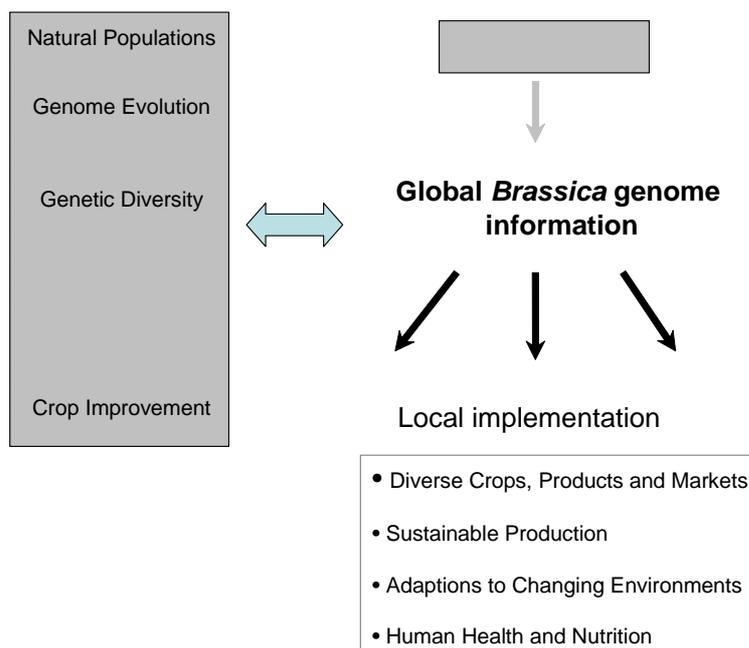


### **1.4. Benefits from International Co-operation**

*Brassica* is a unique genus with global impact and diversity in terms of crop types. There is considerable scope to benefit and add value to national programmes from sharing public-domain pre-competitive knowledge.

### **1.5. A Key Role for Genomic Information**

Interpretation of genomic information in the context of trait characterisation is an essential component of future targeted crop improvement. It is expected that increasing use will be able to be made of knowledge gained from an understanding of gene function, allelic variation and interaction between genes and the environment. Genomic information is the key to exploiting knowledge gained at the level of gene expression, biochemistry, metabolism and physiology.



The size of the *Brassica* genomes and their close relationship to *Arabidopsis*, make them tractable for whole-genome sequencing. In particular, there are considerable advantages to be gained from comparative sequencing between and within the *Brassica* genomes, especially in the context of the known segmental genome duplications that have characterised genome evolution in this genus. In particular, the *Brassica* research communities need to be able to locate genes in the relevant genomic context in order to understand regulation and pressures on evolution. For effective crop-based studies it is essential to be able to navigate between trait and gene, and to distinguish different locus-specific copies.

#### 1.6. **Initiation and Development of the MBGP**

The international Brassica research community has agreed on the need to focus on common aims, and generating and exploiting **public-domain** *Brassica* genomic information. The **Multinational Brassica Genome Project** (MBGP) was initiated in 2002, and a Steering Group established to help crystallize objectives and promote national and international research programmes. The long term aim is to capitalise on the immense opportunities and benefits that can arise from a combined and integrated effort to understand and utilise genomic information.

## 2. THE ROLE OF *BRASSICA* GENOMES IN AGRICULTURE AND THE ECONOMY

### 2.1. **Climate and Regional Adaptations**

2.1.1. Brassicas are primarily adapted to temperate climates, and due to the inherent phenotypic plasticity, are well adapted to a range of intensive or low input cultivation techniques. The genetic diversity within the primary and secondary gene pools includes gene variants (alleles) that encompass adaptations to abiotic (eg water, nutrient) and biotic (pests, pathogens) stresses, as well as considerable variation in secondary metabolites (oils, glucosinolates, vitamins). There has been regional specialisation of crop types such as *B. rapa* vegetables in S.E. Asia, *B. juncea* oil crops in India. In general domestication has followed a typical path of selection of specialised forms followed by adaptive radiation and adoption amongst wider human populations.

**2.1.2. Crops, Cultivated Area, Economy and markets**

**2.1.2.1. Oils**

The total harvested area for Rapeseed (Canola) is **23,694,731 Ha** (FAO, 2003), producing 35,931,452 Mt. This represents % of vegetable oil production. Of this, % is for human consumption.

Yield per hectare varies by ten-fold in different countries, reflecting different levels of inputs and efficiency of production. This highlights the scope for crop improvement, and yield will continue to be the primary focus of many crop improvement programmes. The relatively low harvest index (harvested : total biomass) of *Brassica* oil crops compared to other arable crops also provides considerable impetus for crop improvement.

- o Global market value
- o Value to farmers

	<b>Area Harvested Ha</b>	<b>Yield (Hg/Ha)</b>	<b>Production (Mt)</b>
China	7,200,004	15,972	11,500,005
Canada	4,689,200	14,222	6,669,200
India	4,800,000	7,604	3,650,000
Germany	1,270,000	28,661	3,640,000
France	1,077,000	30,836	3,321,000
United Kingdom	460,000	39,935	1,837,000
Australia	1,500,000	9,467	1,420,000
Poland	426,300	17,677	753,556
USA	432,700	15,865	686,470
Czech Republic	250,959	15,979	401,000
Pakistan	290,000	8,621	250,000
Bangladesh	297,447	7,329	218,000
Denmark	85,000	25,647	218,000
Sweden	67,469	26,679	180,000
Austria	56,000	24,821	139,000
Lithuania	60,000	22,333	134,000
Russian Federation	125,000	10,000	125,000
Finland	66,200	15,529	102,800
Hungary	58,000	12,069	70,000
Belarus	140,000	4,643	65,000
<b>World Total</b>	<b>23,694,731</b>		<b>35,931,452</b>

- **Processing**  
Oil from *B. napus*, *B. rapa* and *B. juncea* is primarily used as an edible feed for humans or livestock. Oils are either used directly in cooking or processed into a variety of food products as a replacement for dairy-based fats in margarine. Residual meal contains valuable protein (~20%) which is used as livestock feed.
- **Added-value and non-food uses**  
Due to increasing pressures to reduce reliance on fossil-fuel carbon sources, Brassica derived is being adopted in some regions as a biodiesel engine fuel. Dependent upon fatty acid composition, other uses include lubricant oils and as a raw material for biocomposites and plastics.
- **Changes in market requirements**  
There is increased interest in diversifying the Brassica-based oils available in the market place. This is primarily dependent upon modifying the fatty acid profiles to develop crops such as HERO (High erucic acid, for ...) and HILO (High linoleic acid).

**2.1.2.2. Vegetables**

The total harvested area for Cabbage and Cauliflower is 3,160,792Ha. FAO statistics do not include or distinguish between many *Brassica* vegetable crops (Broccoli, Chinese cabbages et al.)

<b>Cabbage</b>	<b>Area Harvested Ha</b>	<b>Yield (Hg/Ha)</b>	<b>Production (Mt)</b>
China	1,621,050	170,741	27,678,001
India	280,000	217,857	6,100,000
Russian Fed'tion	180,000	222,222	4,000,000

## **Brassica : Harvesting the Genome, Diversity and Products**

Korea, Rep	44,409	580,029	2,575,849
Japan	56,000	427,143	2,392,000
USA	105,000	219,048	2,300,000
Indonesia	90,405	159,083	1,438,194
Poland	32,000	390,625	1,250,000
Ukraine	85,000	141,176	1,200,000
Uzbekistan	16,000	562,500	900,000
Romania	42,000	202,381	850,000
Turkey	32,000	229,688	735,000
Korea, Dem Peoples Rep	34,000	200,000	680,000
Germany	12,899	489,185	631,000
Egypt	20,000	285,000	570,000
<b>World Total</b>	<b>3,160,792</b>		<b>62,001,881</b>

Korea: Chinese Cabbage 4,451,135; Krw 546 billion (€ 266 million)

<b>Cauliflower</b>	<b>Area Harvested Ha</b>	<b>Yield (Hg/Ha)</b>	<b>Production (Mt)</b>
China	353,616	183,405	6,485,502
India	260,000	184,615	4,800,000
Italy	24,053	209,415	503,705
France	27,000	145,185	392,000
USA	17,500	171,429	300,000
Spain	12,500	225,280	281,600
Pakistan	11,000	182,727	201,000
Mexico	18,000	111,111	200,000
Poland	10,000	200,000	200,000
UK	12,000	109,333	131,200
Germany	5,123	249,854	128,000
Australia	11,300	106,195	120,000
Japan	10,000	118,000	118,000
Egypt	4,500	244,444	110,000
<b>World total</b>	<b>857,692</b>		<b>15,084,200</b>

### **2.1.2.3. Condiments**

The *Brassica* species *B. nigra*, *B. juncea* and *B. carinata* are important mustard crops in sub-tropical regions of the world, including Northern Africa and the Indian sub-continent.

### **2.1.2.4. Fodder**

*Brassica* kales and turnips have traditionally been used as fodder and forage crops.

### **2.1.3. Diversity and specific adaptations for diverse environments and climate change**

The area of the world's arable land is decreasing, whilst the human population continues to grow. There is increasing pressure to grow arable and other crops on land with marginal nutrition or water availability. The innate variation within the *Brassica* gene pool should drive a concerted effort to understand the basis of adaptability. Access to the relevant genomic resources and information will provide breeders with the necessary tools to select the optimal combinations of alleles adapted to local and changing growing environments.

## **2.2. Agricultural practices**

### **2.2.1. Land use and crop nutrition**

In some systems, oilseed crops are grown in rotation with cereals, with added beneficial effects as a soil fumigant. Sustainable yields are dependent upon efficient use of macro- and micro-nutrients. Nitrogen and phosphate limit yield potential, but there is considerable genetic variation for their efficient uptake and utilisation within *Brassica* crops.

### **2.2.2. Labour**

Where *Brassica* oil crops are mechanically sown and harvested in many regions of the world, labour inputs are relatively low. *Brassica* vegetable crops are subject to more intensive cultivation, and mechanisation varies region to region and crop to crop. There is increasing pressure on reducing labour inputs associated with crop establishment, crop protection and harvest. In addition, there are requirements to attain sustainable production through

## ***Brassica* : Harvesting the Genome, Diversity and Products**

reduction in mechanised post-processing and waste. In many parts of the world the majority of losses occurs post production, often associated with poorly developed supply chain systems. There are thus important drivers for improved post-harvest keeping quality (especially of vegetables).

### 2.2.3. Environmental impact

There is an increasing amount of quantitative information becoming available on the environmental impact of different crops. There is variation in the amounts of inputs in terms of water and chemical use, as well as in amount of run off pollution in terms of fertiliser and pesticide use. For intensive cultivation there are pressures to reduce inputs and pollutants.

### 3.2.4 Crop protection

There are continuing threats to the production of Brassica crops worldwide, with different pests and diseases having global or regional significance. Major diseases include *Leptosphaeria* blackleg (esp. on oilseeds), *Xanthomonas* black rot (on vegetables, clubroot and downy and powdery mildews. Major pests include diamond back moth and cabbage aphid. In some regions (UK, elsewhere) bird and slug damage have major effects on crop establishment, particularly of oilseeds.

### 3.2.5 Effects of climate change

Changes in climate patterns and overall increases in average global temperatures are likely to have profound effects on Brassica production worldwide. In particular, the availability of usable land, water, incidence and changes in pests and diseases and seasonality may have unpredictable knock-on effects. Understanding plant adaptations to changes in the environment, and the availability of robust genetic variation will be essential to underpinning relevant crop improvement.

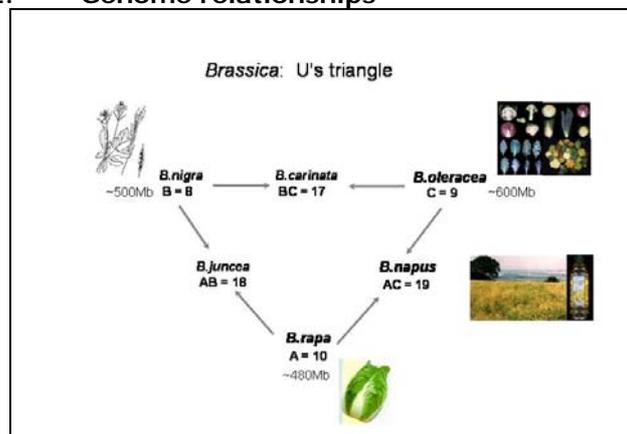
## **3. SCIENTIFIC STATUS AND UNIQUE FEATURES OF *BRASSICA* SPECIES**

### 3.1. One genome, many crops

3.1.1. The relationships between the different genomes characterising the different species in the Brassica genus are outlined in the well-known "Triangle of U". There are three distinct diploid 'genomes' or 'cytodemes' recognised, each represented by a type species:

Genome	Species	n=	~ genome size
A	<i>B. rapa</i>	10	500-550 Mbp
B	<i>B. nigra</i>	8	470
C	<i>B. oleracea</i>	9	600-650
AB	<i>B. juncea</i>	18	1100-1500
AC	<i>B. napus</i>	19	1130-1240
BC	<i>B. carinata</i>	17	1540

### 3.1.2. Genome relationships



## ***Brassica* : Harvesting the Genome, Diversity and Products**

Within each species, a number of distinct crop types have been domesticated. A non-exhaustive list is shown below:

<i>B. rapa</i>	Vegetable	Turnip, turnip tops, Pak choi, turnip rape, sarson
	Oil	
<i>B. nigra</i>	Condiment	Black mustard
<i>B. oleracea</i>	Vegetable	Broccoli, Brussels sprout, cabbage, cauliflower, kohlrabi, collard, kale, Chinese white kale (Kailaan)
	Fodder	Kale
<i>B. juncea</i>		Brown mustard, oil
<i>B. napus</i>	Oil	Oilseed rape, rapeseed, Canola
	Vegetable	Swede
	Fodder	
<i>B. carinata</i>	Condiment	Ethiopian mustard

### 3.1.3. A model for chromosome evolution

Due to the distribution of a series of closely related genomes and combined amphidiploids, the *Brassica* genus provides a useful system to study chromosome evolution over different time-spans. In particular, it has been possible to start to understand some of the dynamics of duplication and divergence associated with speciation and polyploidisation. In addition, the close relationship with *Arabidopsis* allows comparisons over longer time spans at the level of chromosome segment down to sequence.

### 3.1.4. Polyploidy

Genome duplication or polyploidy is a widespread phenomenon throughout the plant kingdom, and provides a mechanism and context for genetic and evolutionary adaptations. *Brassica* species provide excellent experimental systems to understand the consequences of genome duplication. There are well studied examples of whole genome duplication that result in allopolyploidy (combining two or more distinct genomes) or autopolyploidy (duplication of a diploid genome). New allopolyploid species such as *Brassica napus* often combine parental phenotypes with novel traits and adaptations and the fixation of hybrid vigour. The diploid *Brassica* species are also comprised of ancestral chromosomal segmental duplications that result in blocks of paralogous genes that may then provide a wider range of developmental or adaptive variation in terms of expression or specificity.

## 3.2. Genetics

Molecular marker based genetic maps have been established for all *Brassica* species. In recent years, considerable effort has been placed in developing high-density maps based on DNA sequence-tagged genetic markers (RFLPS, SSRs, CAPS, SNPs etc). The development of genomics tools is dependent on the ability of researchers to navigate amongst different genetic maps that are anchored with common sequence-tagged markers. An increasing amount of information is available that provides links between *Brassica* genetic maps and the *Arabidopsis* genome – primarily through the use of sequence-tagged markers highly orthologous to defined *Arabidopsis* genes and loci.

### 3.2.1. Quantitative Genetics

Quantitative Trait Loci (QTL) have been identified for a wide range of morphological, physiological and crop traits in the different *Brassica* crop types. The ability to resolve QTL is dependent upon access to a larger number of recombinant individuals (segregating populations, substitution or near-isogenic lines), as well as to high density genetic and physical maps. There is a continuing requirement to provide the tools that allow researchers to resolve QTL to the level of physical maps (BAC contigs) and candidate genes. This can be achieved through an integration of physical BAC contigs anchored with genetic markers onto common reference genetic linkage maps, as well as access to comparative genomic data between *Brassica* and *Arabidopsis*.

## 3.3. Breeding systems and technologies

### 3.3.1. Self-incompatibility

*Brassica* species have been established as the model sporophytic pollen self-incompatibility system for research, with a well-characterised and diverse allelic series of 'S-' alleles in both *B. oleracea* and *B. rapa* having strong to weak incompatibility reactions. The S locus has been well characterised at the genomic and sequence level. Alleles of genes at the S-locus display considerable variability, with a very high level of non-synonymous substitution mutation. Comparative studies indicate that this variation appears to have occurred prior to the speciation of *B. rapa* and *B. oleracea*. Prior to adoption of cytoplasmic male sterility systems, self-incompatibility was widely used in development of F1 hybrid brassica crops.

## ***Brassica* : Harvesting the Genome, Diversity and Products**

### 3.3.2. Male sterility

Cytoplasmic male sterility (CMS) occurs infrequently in natural populations of *Brassica*, and provides an excellent tool to study genetic interactions between mitochondria and nucleus during flower development. CMS plants can be selected either following sexual crosses between different species of the same family or by somatic hybridisation between unrelated species. Several systems and sources of CMS have been characterised and used in *Brassica*, and CMS has increasingly replaced self-incompatibility in hybrid *Brassica* production. A single restorer locus appears to be capable of restoring different forms of *Brassica* CMS.

### 3.3.3. Genetic modification

*Brassica* crops were amongst the first to be targeted for commercial transgenic genetic modification, for traits such as herbicide resistance and modification of male sterility. There have been many research programmes and patents associated with modification of fatty acid pathways to yield novel oils.

### 3.3.4. Doubled haploids

Due to the strong self-incompatibility system, most brassicas are outbreeders with a high degree of heterozygosity in natural populations and open-pollinated crops. Doubled haploid technology is widely applied in brassica crop improvement and research programmes, either via anther culture or microspore culture.

## 3.4. ***Brassica* consumption: Health, Medical and Clinical benefits**

Cardiovascular disease and cancer are ranked as the leading causes of death in most industrialised countries. Epidemiological studies demonstrate that increased fruit and vegetable intake decrease the risk of cancer and heart disease, with *Brassica* vegetables appearing to be especially protective against cancer and heart disease. Although the mechanisms are not fully understood, they are likely due to secondary metabolites. Understanding the basis of variation in such metabolites and the detailed regulation of the relevant biochemical pathways will provide a sound basis for more targeted clinical experimentation in the context of human allelic variation. It is important in dietary studies to quantify the 'metabolic profile' of the food, as opposed to just the metabolite of particular interest. *Brassica* vegetables are also known to be beneficial in the prevention of other major illnesses such as Alzheimers, cataracts, and some of the functional declines associated with ageing. Plant secondary products have complementary and overlapping actions, including antioxidant effects from flavanoids and other phenolic compounds, modulation of detoxification enzymes, stimulation of the immune system, reduction of inflammation, modulation of steroid metabolism, and antibacterial and antiviral effects. Most studies show that phytochemicals in *Brassica* plants up-regulate many detoxification enzyme systems in the animal that consumes them.

3.4.1. Considerable genetic variation underlies products derived from plant cell biochemistry, which in turn affects the nutrients humans ingest. Underlying genetic variation has led to *Brassica* plants to produce higher amounts of particular compounds than do other plants in our diet.

3.4.2. The genes in the Brassicaceae uniquely enable the production of a particular class of biochemicals (glucosinolates) which break down to compounds (isothiocyanates such as sulphoraphane) known to provide some protection against a range of human cancers. Glucosinolates and other sulphur containing metabolites act as anti-cancer agents due to their ability to induce detoxification enzymes in mammalian cells and to reduce the rate of tumour development. Isothiocyanates are modulators of Phase 1 and 2 enzyme activity and neutralise cancer-causing chemicals that damage cells by interfering with tumour growth. The protective effect of *Brassica* consumption and the associated decrease in risk of cancer is related to the Human *GST* genotype. Significant effects are identified in cabbage and broccoli that have antiproliferative activities.

3.4.3. In vegetable brassicas, levels of vitamins A, C and E, folate and potassium are notably high. *Brassica* contains folates ranging from Glu<sub>4</sub> to Glu<sub>8</sub>. Nutritional demands for folates are particularly high during pregnancy. Folate supplementation prior to conception can significantly reduce the incidence of neural tube defects. Folate deficiencies have also been implicated in the aetiology of megaloblastic anaemia, *Spina bifida*, neuropsychiatric disorders and various forms of cancer. There is also considerable scope for increasing levels of tocopherol (Vit E) in brassicas.

3.4.4. There is genetic variation in *Brassica* species for the uptake and availability of trace elements such as selenium (Se). Supplementation studies on humans have demonstrated the efficacy of Se for prevention of colon cancer. The metabolism of Se

## **Brassica : Harvesting the Genome, Diversity and Products**

depends upon its chemical form, and the form that occurs in broccoli appears to be particularly effective at protecting laboratory animals from cancer

3.4.5. The fatty acid composition of *Brassica* seeds has been a target for selective breeding to meet different end uses and market demands. It has been shown that lipid profiles may have significant effects on human obesity, with emerging evidence to show that lipid profiles improve on a controlled-carbohydrate diet.

3.4.5.1. Since nervonic acid is an important constituent of brain and membrane phospholipids, it is vital that the body has an adequate provision of this fatty acid. If, for any reason, the body cannot make this lipid, then a dietary source is indicated. Two such situations are under active investigation; multiple sclerosis (MS), where there are indications that there is inadequate provision of nervonic acid; and premature birth, where the infant no longer benefits from maternal nutrition

3.4.5.2. A new iodised oil, Brassiodol utilises rapeseed oil as vehicle of iodination. Brassiodol This is proposed to prevent or eradicate 127I-deficiency disorders including goitre.

### **3.5. Agronomic adaptations and the genetic basis of trait variation**

Within different growing regions and crop types there are priorities vary for research focused on agronomic traits. Dependent upon availability of natural or managed resource, the genetic variation within *Brassica* can be utilised to provide information and solutions to improve nutrient use efficiency, water use efficiency, germination, seedling establishment and vigour, yield components and the basis of genotype x environment interactions. As with all crops there is a continuing requirement to provide protection against pests and diseases.

3.5.1. Potential for increasing harvest index of a major oil crop  
The harvest index (harvestable yield as a proportion of total biomass) of *Brassica* rapeseed crops is low compared with other arable crops such as cereals. There is considerable scope for improving the harvest index to increase overall yields and lower the level of fossil-fuel and other inputs for a more sustainable production. Developing a systems approach to understanding the genetic and genomic basis of yield components requires a multidisciplinary international research environment with access to relevant genomic and trait information.

3.5.2. Heterosis  
The strong self-incompatibility within *Brassica* species leads to a natural outcrossing breeding system. Recurrent selfing tends to lead to inbreeding depression. The fixation of genetic variation in F1 varieties provides considerable advantages in terms of hybrid vigour, also known as heterosis. The natural variation, breeding material and experimental systems for *Brassica* provide an excellent opportunity to determine the genomic basis of heterosis, which is not yet fully understood.

### **3.6. Non-food uses**

There is considerable scope for increasing the added value from Brassica-based production. For oil-based products these include high-value lubricants, bio-fuels, plastics and composites, as well as fibre-based products.

## **4. BRASSICA BIODIVERSITY AND CROP IMPROVEMENT**

### **4.1. The genepool and genetic resources**

*Brassica* crops are derived from wild species which still exist in nature, and which they are able to be inter-crossed. There are concerns that in some crop types there has been progressive genetic erosion and reduction in genetic diversity within the crop gene-pool. The wild relatives thus have the potential to provide access to a much wider range of allelic variation.

#### **4.1.1. Natural variation**

In situ genetic variation exists for the diploid *Brassica* species, in the centres of diversity. For *B. oleracea* (and C genome species) this is centred on the Mediterranean and coastal regions of Europe. For *B. rapa*, natural populations exist near waterways in Europe.

#### **4.1.2. Domesticated**

*Brassica* crops appear to have been under different selection pressures reflecting the period over which they have been domesticated. For leafy vegetable crops such as kale, there is a wide genetic diversity and eco-geographic distribution. More specialised leafy, root or floral organ based vegetables have been progressively selected and refined as mutations have accumulated through domestication, and hence may have more restricted eco-geographic

## ***Brassica* : Harvesting the Genome, Diversity and Products**

centres of diversity. The *B. napus* oil crops appear to be relatively recent crop types that have been selected since stable hybridisation between domesticated diploid progenitors. A further restriction of the genepool is likely to have taken place following development of oil-extraction technologies and more recent selection of varieties possessing low glucosinolate and low erucic acid content (double lows).

### **4.1.3. Genetic resources**

*Ex situ* genetic resources are available in the public and private sectors. The European *Brassica* database (Bras-EDB) compiles information on genetic resource accessions held in public collections. Similar information is held in the USDA/ARS Grin system. The pressure for more intensive agriculture has resulted in an overall reduction of the natural genetic diversity of crop plants, a process called 'genetic erosion' and in some cases leading to a significant loss in the adaptability and sustainability of crops, particularly in combating natural pests and diseases. Traditionally, most crops were based on local selections ('landraces') developed over generations in particular farming communities and locations.

4.1.4. Characterising and using crop biodiversity is essential to provide the range of alternative crop varieties needed to meet the changing needs of the world. A large international effort over the past few decades has ensured (*ex situ*) conservation of many crop genetic resources. This has involved collecting traditional varieties and landraces from around the world, and in particular from centres of genetic diversity for specific crops. Modern crops contain a surprisingly narrow genetic base, despite the existence of extensive allelic variation within germplasm collections. Although breeders exploit the existing natural variation for particular target traits within crop species, there has been considerable reduction in biodiversity through canalisation of alleles.

4.1.5. Core Collections have been developed for *Brassica* spp in the past, within EU GenRes and other programmes. These were based on accessions held within public genetic resource collections. Seed were multiplied and accessions screened for a number of fungal and other resistance traits. Whilst valuable in identifying 'hotspots' of variation within the relevant gene-pools, these collections consisted of heterogeneous and heterozygous material. This limits their long-term use for correlating detailed genetic studies

4.1.6. Diversity Fixed Foundation Sets (DFFS) are defined as: "*an informative set of genetically fixed lines representing a structured sampling of diversity across a genepool*". Recent molecular analysis has shown that there is considerable genotypic variation within individual extant genetic resource accessions. This is present both as within-line heterogeneity and within-plant heterozygosity. As such, any co-ordinated trait and genetic studies will be very difficult to interpret unless material with acceptable levels of homozygosity is produced and made publicly available. To address this issue and provide researchers with relevant resources, 'Diversity Fixed Foundation Sets' (DFFS) are being constructed. These are designed to represent the majority of allelic variation within respective gene pools in a form suitable for long term exploitation by researchers and end users.

*Information to be added in next edition:*

4.2. **Harnessing Genomics:** The vision for *Brassica* Crop improvement

4.2.1. Conventional- current breeding programs

4.2.2. Potential for allele mining and marker assisted breeding

4.2.3. GMO

4.3. Specific examples of agronomic trait targets and expected impact of the MGBSP on germplasm enhancement.

## **5. STATUS AND AIMS OF *BRASSICA* GENOME RESEARCH**

### **5.1. Sequence information**

The latest release of EMBL/Genbank includes 404Mb of *Brassica oleracea* DNA sequence, with additional large datasets of *B. napus* and other ESTs.

#### **5.1.1. ESTs**

A number of EST programmes have been initiated in different countries. A *B. napus* EST programme funded by Genoplante in France released sequences into the public databases in

## **Brassica : Harvesting the Genome, Diversity and Products**

summer 2003. *B. rapa* ESTs are being sequenced as part of the S. Korean *B. rapa* sequencing effort. A UK-Canadian *B. oleracea* EST programme is releasing data during 2005.

### **5.2. Genetic maps**

Genetic maps have been generated for all *Brassica* species. Most effort has been focused on *B. napus*, *B. oleracea* and *B. rapa*. Until recently there has been relatively little widespread integration of maps with public-domain sequence-tagged common markers. There is an ongoing effort to generate such integrated information and to standardise nomenclature and orientation of linkage groups. A number of maps exist that have been constructed with EST or other sequence tags from Arabidopsis.

### **5.3. Physical maps and BAC libraries**

BAC libraries are available for many *Brassica* species in the public domain. The UK genome effort has initially focused on developing fingerprinted contigs of the A and C genomes, anchored to the Arabidopsis genome through hybridisation with 1200+ gene-specific probes. An overgo-hybridisation programme with additional Arabidopsis probes is underway in the USA.

Reference BAC libraries have been generated in Korea for the *B. rapa* sequencing programme, and distributed to laboratories around the world.

### **5.4. Karyotype analysis**

For *B. oleracea* chromosome FISH analysis has allowed orientation and assignment of all chromosomes to the linkage groups. For *B. rapa* a similar exercise has been carried out in Korea.

### **5.5. Mutants, TILLING**

Mutant lines of *Brassica* have been compiled and characterised in the past. Sets of rapid cycling lines were generated from the Wisconsin Crucifer Genetics Co-operative, and are used as a successful teaching tool worldwide. The first plant to complete its life cycle in space was a rapid cycling line of *B. rapa* (Astro-plant) on the MIR space station. The Astro-plant lines have been used in subsequent zero-gravity experiments.

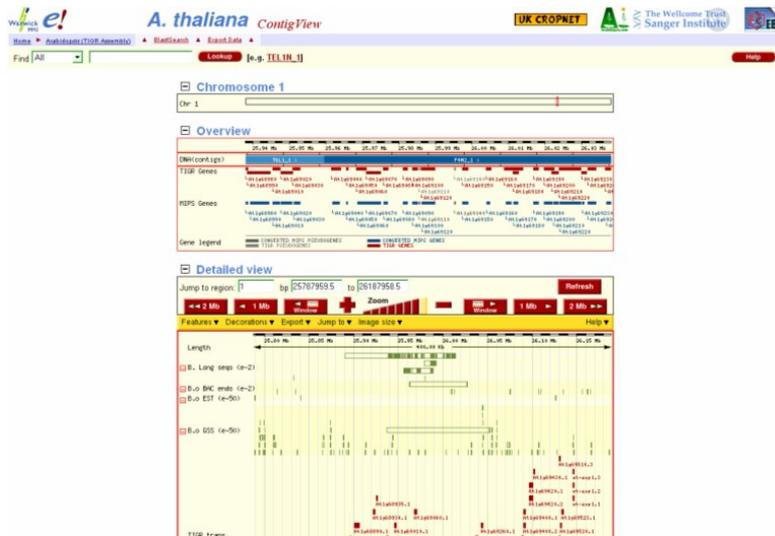
5.5.1. The Targeted Induced Local Lesions In Genomes (TILLING) platform provides a means to screen (usually EMS) mutagenised populations for mutations in specific candidate gene families. The presence of duplicated chromosomal segments and several paralogous copies of each gene suggest that *Brassica* species should be able to carry a relatively high mutational load (due to gene redundancy). Initial exploratory TILLING populations have been established for *B. napus* and *B. oleracea*, and are underway for *B. rapa*. The labour-intensive tasks of maintaining tens of thousands of mutagenised lines indicates that future development of comprehensive *Brassica* TILLING resources should be managed within an international network.

### **5.6. Other experimental resources**

A range of functional genomic resources are required to fully utilise genomic information arising from Brassica and Arabidopsis. Genotypes of some species of *Brassica* are relatively easy to transform, and thus study up and down regulation of individual genes or gene families.

### **5.7. Information, knowledge and bioinformatics**

A considerable amount of information is available world-wide relating to Brassica traits, genetics and genomics. There is a requirement for the international community to develop methodologies for establishing conventions for nomenclature and information exchange, in order to harness the full benefits from genome-based resources. Some progress has been made in this area, with specific examples of comparative genomic data that enable navigation between the Arabidopsis and *Brassica* genomes.



Alignment of *Brassica* genomic sequence information and Arabidopsis in an Ensemble viewer (<http://ensembl.warwick.ac.uk>)

**6. A UNIQUE OPPORTUNITY FROM MODEL TO CROPS: HARNESSING ARABIDOPSIS** (INFORMATION TO BE ADDED IN NEXT EDITION)

- 6.1. How good a model is Arabidopsis?
  - 6.1.1.1. Commonalities
  - 6.1.1.2. Crop-specific issues
  - 6.1.1.3. Field based studies
  - 6.1.1.4. Harvestable product
  - 6.1.1.5. Deployment of resistance genes
- 6.2. Sequence relationships
- 6.3. Chromosome evolution and organisation
- 6.4. Mutants and genomic regulatory networks
- 6.5. Common developmental pathways
- 6.6. Impact of duplicated loci (paralogues) in *Brassica*
- 6.7. Use of common resources (eg microarrays, complementation)

Relationship between the international Brassica and Arabidopsis communities

## **7. SETTING PRIORITIES: MBGP GOALS AND OBJECTIVES**

### **7.1. Prioritisation of objectives**

History of meetings of the MBGP

- Jan 2002 Initial discussion at PAGX
- Mar 2002 Discussion and tabulated list of MBGP resources from Crucifer Genetics Workshop
- Jan 2003 1st Steering Group Meeting, San Diego
- Jan 2004 2nd Steering Group Meeting, San Diego
- Oct 2004 3rd Steering Group Meeting, Daejeon, S. Korea
- Jan 2005 4th Steering Group Meeting, San Diego

### **7.2. Navigating between trait and genome**

### **7.3. Integration of genetic and physical maps**

### **7.4. DETERMINING THE GENOME SEQUENCE**

**7.4.1. *B. rapa* Genome sequence** A Concept note for the Brassica genome sequencing project was prepared 11th June, 2003 by Ian Bancroft (ian.bancroft@bbsrc.ac.uk) and Yong Pyo Lim (yplim@cuvic.cnu.ac.kr) on behalf of the Steering Group for the Multinational *Brassica* Genome Project.

**7.4.2. Deliverables:** The project aims initially to produce, from BAC clones, "Phase 2" sequence (i.e. fully oriented and ordered sequence but some small sequence gaps and low quality sequences) for the ca. 500 Mb genome of *Brassica rapa* subspecies *pekinensis*. The genome sequence is to be anchored to a reference genetic map by ca. 1000 molecular markers. Note that sequencing to Phase 2 halves the cost of the systematic project compared to the production of fully finished sequence. Scientists requiring finished sequence from a specific region can finish it themselves, accessing trace files that will be archived in an agreed format at TIGR, MIPS and NIAB (National Institute of Agricultural Biotechnology, Suweon, S. Korea). The systematic annotation of the genome sequence, although not to fully finished standard, will be feasible due to the availability of the annotated genome sequence of the related species *Arabidopsis thaliana*.

**7.4.3. Work plan:** An international consortium, working to agreed standards and using common resources, will conduct the sequencing of the Brassica genome. TIGR, MIPS and NIAB will conduct integration, annotation and public serving of the data. The work plan comprises three partially overlapping stages:

#### **7.4.4. STAGE 1: PRELIMINARY WORK**

- Establish a general *Brassica* information site (<http://www.brassica.info/> and gateway site for the Multinational *Brassica* Genome Project (<http://brassica.bbsrc.ac.uk> & <http://www.niab.go.kr>). Establish communal databases for genetic mapping data (<http://brassica.bbsrc.ac.uk/BrassicaDB> and <http://www.niab.go.kr>) and physical mapping data (<http://brassica.bbsrc.ac.uk/IGF/> and <http://www.brassicagenome.org>).
- Define **reference BAC libraries** for genome sequencing: *B. rapa pekinensis* (Chinese cabbage) variety Chiifu-401 libraries KBrH and KBrB. Each library consists of 144 x 384-well plates; made using *Hind*III (KBrH) or *Bam*HI (KBrB) digested genomic DNA. A total of 110,592 clones, providing 20-fold redundant representation of the genome are available. The libraries were constructed at Chungnam National University (CNU), Daejeon, S. Korea.
- Establish **international distribution centres** for KBrH and KBrB **BAC libraries**: at NIAB (for eastern hemisphere) and JIC (John Innes Centre, U.K.) (for western hemisphere).
- Define **communal reference populations** for genetic mapping in Chiifu: CKDH and CKRI. CKDH consists of 86 DH lines available now; use for low-resolution mapping. CKRI will consist of at least 200 RI lines when complete (2004); use for high-resolution mapping. Construct high quality genetic maps for each population using communal markers and make publicly available. Includes input from CNU, AAFC Agriculture and Agri-Food Canada), WHRI (Warwick HRI, U.K.).

## ***Brassica* : Harvesting the Genome, Diversity and Products**

- Establish international **distribution centres** for CKDH and CKRI **plant populations**: at CNU (for eastern hemisphere) and HRI (for western hemisphere).
  - End sequence all 110,592 BAC clones in reference libraries and collate data in searchable community databases at TIGR, MIPS and NIAB.
- 7.4.5. STAGE 2: GENETIC ANCHORING OF SEED BACS** Anchor ca. 1000 seed BACs unambiguously to the Arabidopsis genome sequence (via both end sequences) and to *B. rapa* genetic map via single amplification product molecular markers (SSRs, SNPs, InDels) on CKDH or CKRI population.
- 7.4.6. STAGE 3: GENOME SEQUENCING**
- Divide sequencing on chromosome-by-chromosome basis.
  - Participants to start by sequencing all seed BACs on their chromosome, or defined sub-region of a chromosome, for which funding to complete the sequencing is in hand.
  - Sequence to Phase 2.
  - Deposit all sequence data at TIGR, MIPS, and NIAB as soon as Phase 2 achieved, using a data format to be agreed.
  - Check collinearity of clones: a) by reference to the Arabidopsis genome. If discontinuity in conserved microsynteny, proceed to b) look for independent end-sequenced BACs that confirm authenticity of discontinuity. If confirmation cannot be found, consider terminating sequencing walk. If walk must be continued, c) develop a genetic marker from unique sequence beyond discontinuity and map in CKDH/CKRI population to confirm map location. Proceed to step 6 only if have confirmation that clone is likely to be collinear with the genome.
  - Select next BAC for sequencing using end sequences; select to have minimum overlap. (If other end of clone overlaps existing sequenced BACs, find a clone with minimum aggregate overlaps.)
  - Sequence to Phase 2; deposit data; check collinearity; continue iteratively.
- 7.4.7. Timetable:** Most of the work involved in this project involves well-established high throughput DNA sequencing technology. There is ample capacity worldwide to conduct such work, therefore the rate of progress will be limited primarily by the availability of the BAC libraries, reference mapping population and, in particular, funding.
- 7.4.7.1.** The distribution of the libraries and mapping population is under way. It is expected that the remainder of the preliminary (Stage 1) work, i.e. end sequencing of the BAC libraries, can progress rapidly.
- 7.4.7.2.** Everyone interested in participation in the main sequencing work (Stage 3) is encouraged to participate in the BAC end sequencing, with the aim of completing the end sequencing of all 110,592 clones by mid 2004.
- 7.4.7.3.** The selection and genetic anchoring of seed BACs (Stage 2) can commence as soon as the reference mapping populations and substantial numbers of BAC end sequences are available. The mapping of BACs via SNP markers (as will often be required) is a specialised task, but several members of the international Brassica genomics community have such capabilities. The commencement of these mapping activities should be encouraged as soon as the end sequences start to become available, i.e. from late 2003, with the aim of completion of seed BAC anchoring by the end of 2004.
- 7.4.8.** The completion of the anchoring of all 1000 seed BACs (Stage 2) should not be viewed as a prerequisite for the commencement of large-scale genome sequencing (Stage 3). The early initiation of sequencing, building upon the experience of the pilot Brassica BAC sequencing project already conducted in the USA by TIGR, should be encouraged. However, the timing of progression beyond sequencing of the initial seed BACs should be such that all of the BAC end sequences are available for the selection of minimally overlapping clones. The agreed quality control standards (in particular those aimed at ensuring that the sequencing of chimaeric clones does not lead to erroneous sequencing of incorrect regions of the genome) must be adhered to strictly.

## ***Brassica* : Harvesting the Genome, Diversity and Products**

7.4.9. The main sequencing work (Stage 3) is structured such that many clones can be sequenced in parallel. This provides the opportunity for very rapid progress if sufficient funds are available. Allowing for the effect of the timing of national funding cycles, we should aim for completion of systematic genome sequencing (to Phase 2) by the end of 2007. In order to achieve this ambitious goal, it will be necessary to engage all nations with a significant interest in Brassica as either a crop or model system, and work in a highly cooperative manner.

7.5. *B. oleracea* :- adding value to existing *B. oleracea* sequence

7.6. Comparative sequencing. B genome, *B. napus*

7.7. Maximising access and use of common resources

7.8. Information management and sharing

## **8. STATUS OF *BRASSICA* RESEARCH PROGRAMMES AROUND THE WORLD**

### **OVERVIEW:**

*Brassica* research is characterised by its diversity of experimental materials, trait biology and end-uses. The common thread is access to well defined genomic resources and close relationship with the model plant *Arabidopsis*. Throughout the world the research communities comprise public and privately funded programmes, with a long history of interaction and technology transfer between public sector researchers and breeding companies.

### **COMMUNICATION:**

- The [www.brassica.info](http://www.brassica.info) website has been established as the common portal for compiling information about Brassica species, crop and genome
- The genome gateway sites have also been established at <http://brassica.bbsrc.ac.uk/> and <http://www.niab.go.kr>
- The [brassica.info](http://www.brassica.info) **e-mail list** has been established to encourage communication and discussion amongst the international Brassica research community, and currently has over X members world-wide.
- International *Brassica* meetings (ISHS Symposia on Brassicas) have been held as follows:
  - 1994: Lisbon, Portugal 1994
  - 1997: Rennes, France, 1997
  - 2000: Wellesbourne, UK, 2000
  - 2004: Daejeon, Korea, 2004
  - 2008: next meeting planned in Norway, 2006

International Crucifer Genetics Workshops have been held concurrent with these meetings. In addition, Crucifer Genetics Workshops have been held regularly since ?

1998: Quebec, Canada  
2002: Davis, USA  
2006: Wageningen, The Netherlands

1999: 10th Rapeseed Congress, Canberra, Australia  
2003: 11th Rapeseed Congress, Copenhagen, Denmark  
2007: 12th Rapeseed Congress, Wuhan, China

### **COUNTRY BY COUNTRY:**

Australia  
Canada  
China  
Czech Republic  
Denmark  
European Union  
France  
Germany  
India  
Italy  
Japan

***Brassica* : Harvesting the Genome, Diversity and Products**

Korea  
New Zealand  
Pakistan  
Poland  
Portugal  
Spain  
Sweden  
The Netherlands  
UK  
USA  
+