The Molecular Genetics Of Curd Morphology And The Domestication Of Cauliflower (*Brassica oleracea* L. var. *botrytis* L.)

Lee Brett Smith B.Sc. (Hon's.)

Ph.D. thesis

University of Warwick

Department of Plant Genetics and Biotechnology
Horticulture Research International

December 1999

2000

TABLES AND FIGURES

floral induction in *A. thaliana*.

Figure 8.2: Theoretical model of role of *BoAP1* threshold requirement to induce floral induction in the *Brassica* cauliflower.

Figure 8.3: Theoretical model of how the functional status of *BoCAL-a* and *BoAP1-a*, allied to the status of inflorescence fusion may control the generic morphologies covering the entire range of crop morphotypes within broccolis and cauliflower.

ACKNOWLEDGEMENTS

I am indebted to many people who have supported me during this research.

I thank Neil Periam for his advice and (sometimes too) incisive comments regarding my chosen approaches, and specifically for advice on SSR analysis. I thank Carol Ryder for her hard work in generating a *Brassica oleracea* BAC library which has been of such benefit to my research. I am also very grateful to Rosemary McClenaghan for help and advice on all plant matters, although I still don't think I'll ever make it as a gardener....

I thank Angela Pinnegar and Dave Astley at the HRIGRU and Doug Blazey at Yates' Seed Company for provision of *B. oleracea* seed, and the ABRC for *Arabidopsis* seed. I also thank Rachel Sebastian for providing me with access to her genetic map prior to publication. I am very grateful to Sue Armstrong for training in Fluorescent *In Situ* Hybridisation. I am also indebted to Marty Yanofsky and Sherry Kempin for access to pre-published sequence information and the SK162 *BobCAL* clone. I would also like to thank Michael Purugannan for providing a pre-published manuscript of his *BoAP1* paper.

Finally I am grateful to my supervisor Graham King. I have valued his support, advice, and the wealth of resources he has made available for my research purposes, I hope I have fulfilled the trust he has placed in me.

Julie, Duncan and Sarah, thank you, it's finished!

DEDICATION

This thesis is dedicated to my son Daniel, whose arrival on the 16th November 1999 provided an amazing culmination to a hectic year. Hopefully he'll let me get some sleep one day.....

DECLARATION

The author declares that the work presented in this thesis has not been published previously, or submitted in part or whole towards the award of another degree.

Lee Smith
December 1999

SUMMARY

The characteristic curd phenotype of the *Brassica* cauliflower consists of proliferating, arrested inflorescence and floral meristems. Recent analysis of the similar phenotype in the *ap1-1/cal-1* mutant of the related crucifer *Arabidopsis thaliana* has led to speculation that the orthologous genes from *Brassica oleracea* L. may be responsible for this characteristic trait.

Application of molecular genetic analysis to this hypothesis allows the presentation of a genetic model based on specific, mapped loci of *BoCAL* and *BoAP1*. This model accounts for differences in the stage of arrest between the heading phenotypes of cauliflower (*B. oleracea* var. *botrytis* L.) and Calabrese broccoli (*B. oleracea* var. *italica* Plenck), and is also predictive in accounting for intermediate stages of arrest similar to those observed in Sicilian Purple types.

Further molecular genetic analysis characterised three independent loci of the floral meristem identity gene *BoAP1*. Integration of this data into the genetic model proposed for curd development, suggests a combination of point mutations and expression thresholds of several copies of the key meristem identity genes *BoCAL* and *BoAP1* respectively may account for the development of curd tissue in the *Brassica* cauliflower.

The association of alleles of the *BoCAL-a* gene with the curding phenotypes of *B. oleracea* was also demonstrated through a survey of over 200 crop accessions. This reveals strong correlations between specific *BoCAL-a* alleles and discrete inflorescence morphologies, and allows the presentation of a possible scenario for the domestication of cauliflowers.

Molecular genetic analysis of *BoCAL-a* utilising monosomic addition lines has also demonstrated the potential for integration of the genetic and cytogenetic maps of *B. oleracea.* Such analysis may have significant utility for physical characterisation of replicated loci in *B. oleracea,* prior to the development of a strong physical map.

Further examination of inflorescence morphologies amongst the heading brassicas revealed a shared trait, termed Fused Inflorescence. Preliminary investigation suggests this trait may be under the control of multiple loci, providing a possible indication of the delineation between heading and sprouting *B. oleracea* crops.

ABBREVIATIONS

BAC	Bacterial Artificial Chromosome
bp	Base pair
CAPS	Cleaved Amplified Polymorphic Sequence
cm	Centimetre
cm^2	Centimetre squared
cM	Centimorgan
°C	Degrees Celsius
DNA	Deoxyribonucleic Acid
DH	Doubled Haploid
F1	First filial generation
F2	Second filial generation
GOF	Gain Of Function
g	Gram
ha	Hectare
hr	Hour
Kbp	Kilobase pair
LOF	Loss Of Function
Mbp	Megabase pair
m	Metre
m^2	Metre squared
m^{-2} s^{-1}	Per metre squared per second
µg	Microgram
µg ml^{-1}	Microgram per millilitre
µl	Microlitre
mm	Millimetre
mM	Millimolar
M	Molar
ng	Nanogram
nm	Nanometre
pmol	Picomoles
PCR	Polymerase Chain Reaction
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase PCR
RNA	Ribonucleic Acid
SSR	Simple Sequence Repeat
V cm^{-1}	Volts per centimetre
W	Watts
w/v	Concentration, weight by volume

CHAPTER 1: INTRODUCTION

Crop morphotypes of *Brassica* species have existed for many centuries although analysis of domesticated traits has, until recently, relied solely on morphological and physiological examination. Disproportionate emphasis of these characters during cladistic analysis has resulted in confusion regarding the genetic relationships between crop types (Crisp, 1983). The dogma arising from this confusion has promoted continued debate in the literature regarding the classification, genetics and process of domestication of crop brassicas.

Resources provided by the recent extensive genetic and physiological investigations into the related crucifer *Arabidopsis thaliana,* suggests there now exists a good opportunity to illuminate the genetics underlying one of the most recognisable *Brassica* crop traits, the curd of cauliflower. Application of knowledge gained from *A. thaliana* to the molecular genetic investigation of the *Brassica* cauliflower may help dispel some of the uncertainties regarding the development and domestication of this crop trait.

1.1 The taxonomy of *Brassica* spp.

The brassicas make up 35 species of the tribe *Brassicaceae* within the family *Cruciferae*. Members of the *Cruciferae* are most notably identified by the cruciform shape of the four petals of the open flower. Over 3000 species have been described in the *Cruciferae* (Warwick and Black, 1993), including the model species *Arabidopsis thaliana* which has provided the medium for much of the initial detailed genetic and physiological analysis amongst dicotyledonous plants.

Domestication of several *Brassica* species has led to the development of many distinct crops for which differing specific traits have been selectively bred (table 1.1).

Table 1.1: The diversity of major crops developed from *Brassica* species.

Brassica species	*Crop*	*Tissue harvested*
B. oleracea	Cabbage	leaf
	Kale	leaf/stem
	Broccoli	inflorescence
	Cauliflower	inflorescence
	Kohl rabi	stem
	Brussels sprouts	axillary buds
	Tronchuda cabbage	leaf
	Chinese white kale	leaf
B. rapa	Turnip	root
B. napus	Swede	root
	Oilseed rape (Canola)	seed
B. juncea	Yellow mustard	seed
B. carinata	Ethiopian mustard	seed
B. nigra	Black mustard	seed

1.2 The taxonomy of cauliflower and broccoli

Cauliflower (*B. oleracea* L. var. *botrytis* L.) and Broccoli (var. *italica* Plenck) differ significantly from other brassicas in inflorescence morphology, which has been developed and cultivated for domestic use. The taxonomy of crop types covered by the varietal names of *botrytis* and *italica* is confused (see Crisp 1982, for review) (Gray, 1989). Although in English the terms 'cauliflower' and 'heading broccoli' are used in a generic sense to describe distinct retailed products, these names are actually interchangeable to some extent depending on geographic origin. For example, a specific morphotype may be described as either broccoli or cauliflower in different regions of Italy, the proposed centre of diversity (Massie, 1998).

Taxonomic distinction has classically been defined largely on the basis of 'sprouting'

verses 'heading' habit (Synge, 1956), with cauliflower and heading broccoli (such as Calabrese) classified in one group. More recent classification has utilised the distinction in inflorescence morphology between cauliflower and broccoli to delineate varietas status (Gray, 1982). This classification places importance on the presence or absence of buds at the marketable maturity stage, with var. *italica* referring to all accessions possessing an inflorescence of floral buds and var. *botrytis* to accessions possessing an inflorescence of aborted proliferating apical meristems (Gray, 1989). The major impact of this classification was to move the Sicilian Purple cauliflower from var. *botrytis* to var. *italica,* as this crop develops precocious floral bud primordia across the surface of an otherwise classic cauliflower curd. Conversely, White Sprouting Broccoli moved in the opposite direction, on the basis of curd like structures on an otherwise classic broccoli inflorescence (Gray, 1989). (This more recent taxonomic classification probably best describes the distinction made between cauliflower and broccoli by the general public).

1.3 The ontogeny of cauliflower and morphology of curd

Initial examination of curds by botanists was concerned for the most part with identifying whether curds were formed from vegetative or florally induced tissue. A summary of these investigations (Sadik, 1962) suggested a consensus could not at that time readily been reached. The classic description of cauliflower curd, which has now persisted for over 30 years and is derived from further morphological analysis and physiological measurements, can be summarised as:

"At marketable maturity, the cauliflower head is a curd or dome of tissue made up of a mass of proliferated floral meristems, of which 90% or more abort prior to flowering" (Gray, 1989).

Recent molecular genetic analysis in *B. oleracea* has failed to resolve whether cauliflower curd consists of floral or inflorescence tissue. Carr and Irish (1997) have suggested curds are a mixture of floral and inflorescence meristems based on expression patterns of floral meristem identity genes, although there is no mention of the quality of the retailed material (see 'riceyness' 1.11.3). A similar examination by

Anthony *et al.* (1996) identified the cauliflower curd as consisting of inflorescence meristems, with floral meristems developing across the curd after a variable period. The confusion between the two definitions appears to arise from the problem of relying on gene expression as a determinate of ontological state (Carr and Irish, 1997), which is of course not diagnostic of gene function (thus curd could be florally induced from a genetic viewpoint, yet still consist solely of inflorescence tissue). For this study therefore, the definition of curding tissue will be relatively general in terminology, but extremely specific in ontological stage. The curding stage of the *Brassica* cauliflower will be defined as:

'A proliferation of inflorescence meristems from the primary inflorescence meristem, prior to the development of flower initials'.

1.4 Morphological and physiological analysis of curd development

Although, as has been discussed, the term 'cauliflower' is used in a generic sense to describe the widely grown classic white cauliflower, the curd phenotype is seen in a myriad of forms across the crop types defined as *B. oleracea* vars. *botrytis* and *italica,* including the regional morphotypes from the Italian centre of diversity, Romanesco, Di Jesi and Macerata (Fig. 1.1) (Massie *et al.*, 1996).

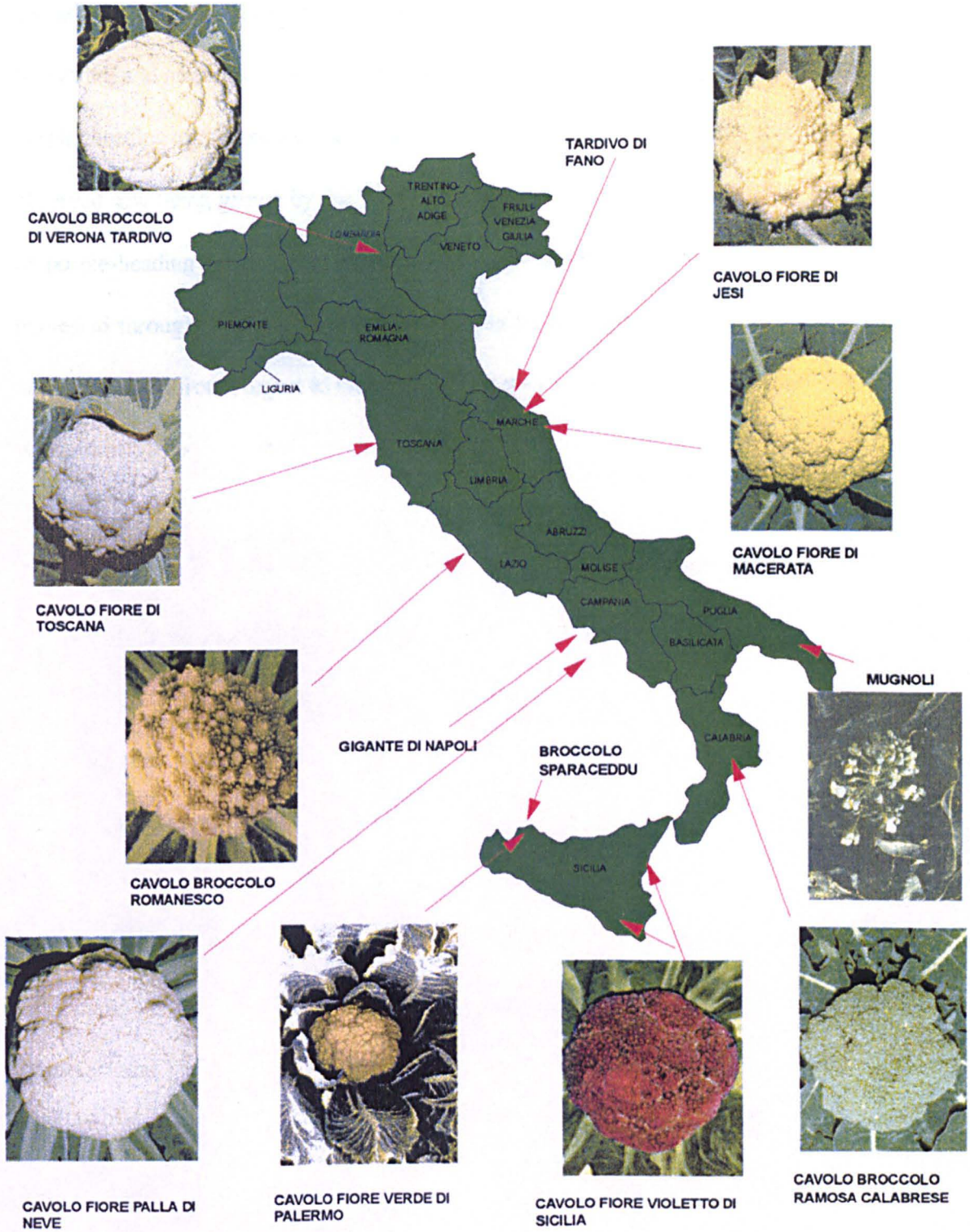


Figure 1.1: Perceived eco-geographic variation of cauliflower and broccoli regional morphotypes across Italy based on taxonometric analysis of 40 inflorescence and leaf characters. Reproduced from Massie (1998).

In all cauliflowers, there is a period of vegetative growth prior to curd initiation, the length of which varies according to genotype and environmental conditions (Hand and Atherton, 1987; Salter, 1969). This period comprises a juvenile phase, followed by an inductive phase during which curd initiation is possible (Salter and James, 1974).

Following initiation, the apical inflorescence meristem, which in wild-type would begin to produce flowers from the peripheral zone, instead proliferates secondary meristems, which in turn proliferate tertiary meristems, quaternary meristems and so forth to form the curd phenotype (Fujime and Okuda, 1996). The gross conformation of the curd tissue, which differs significantly across the range of morphotypes, is dependent to a large extent on three quantitative characters (Kieffer *et al.*, 1998). These characters are, i) the rate of production of branch primordia on the flanks of the apical meristems, ii) the number of branch primordia produced before the first formed produces branch primordia of its own, and iii) the duration of the pre-inflorescence stage (before the production of floral primordia). Geometrical analysis suggested that small changes in each of these three variables are able to explain the differences between gross inflorescence conformation of white cauliflowers, the exaggerated pyramidal shape of the Romanesco cauliflowers, and the Calabrese broccoli (Kieffer *et al.*, 1998).

Following ontogenic arrest at this stage (marketable maturity), for a period again partially defined by temperature (Fujime and Okuda, 1996), approximately 10 percent of primary peduncles elongate into inflorescence spikes (Crisp and Walkey, 1974) simultaneously producing wild-type cruciform flowers which are fully fertile (Sadik, 1962).

1.5 Classical breeding investigations of curding

Investigations into the genetics of the curding phenotype of cauliflowers have been carried out in the past by breeders, with limited success (Crisp, 1982; Pease, 1926; Yeager, 1943). Classical genetic evidence suggested curding may be under polygenic control, given the lack of any easily scorable Mendelian ratio for crop phenotype

between broccoli and cauliflower crosses (Crisp, 1982). At the F8 and F9 generation however, there is a clear indication of a single recessive gene determining the difference between the formation of a curd or a broccoli head (186 plants scored) (Crisp, 1982).

Similar unquantifiable segregation to that reported in Crisp's 1982 paper has been observed in separate experiments (Gray, 1982). Gray reported the presence of a clear intermediate phenotype in the F2 of crosses between cauliflower and Calabrese (Gray, 1982). These intermediates possess rough curds with immature flower buds developing across the surface of the curd, reminiscent of both the 'ricey' phenotype (Fujime, 1983). (1.11.3) and the Sicilian Purple cauliflower discussed earlier (1.2). The development of bud primordia suggests there has been a relaxation of control over floral induction. Both pale green and purple intermediates have been observed (Crisp and Gray in Gray, 1982), and at least one distinct cultivar, cv. Green Ball (Honma and Heecht, 1971), has been developed directly from this material, suggesting it is a stable phenotype.

It is surprising that throughout many decades of cauliflower breeding, no genetic model has ever been proposed to account for curd development. It may be that this lack of evidence, indicative of an underlying level of genetic complexity, is not readily tractable by classical breeding studies alone.

1.6 Exploitation of *Brassica* - *Arabidopsis* synteny

Arabidopsis thaliana, the mouse-eared, or thale cress is a dicotyledonous weedy annual and a close relative of *Brassica*. This species has been used in plant biology for many decades due to its small stature, rapid generation time (seed to seed in 5 weeks), high seed set, ease of growth and ability to self and cross fertilise (for reviews see (Meyerowitz, 1987; Pang and Meyerowitz, 1987)). *A. thaliana* has a small, relatively simple genome, with a DNA content of the haploid genome of approximately 130 Mbp (AGI, 1999) distributed across 5 chromosomes. The genome contains very little repetitive DNA (approximately 10-15%) when compared with other dicots (Meyerowitz, 1987), and this makes *A. thaliana* very amenable to molecular genetic

analysis. For example, genomic libraries can be manufactured and scored with greater ease than for plants with higher DNA contents and more repetitive DNA. *A. thaliana* is also the subject of a genome project which aims to sequence the entire *A. thaliana* genome by the year 2004 (AGI, 1999). The release of information from this project has provided researchers with the ability to conduct *in silico* experiments. Thus DNA sequences isolated from other plant species can now be tested for homology to *A. thaliana* genes through computer analysis rather than in the laboratory.

Molecular genetic analysis of *Arabidopsis* and *Brassica* has demonstrated the close phylogenetic relatedness of these two genera, with homologous protein sequence identity reported around 86% (Lydiate *et al.*, 1993). In contrast to *A. thaliana* however, cultivated *B. oleracea* has a relatively large haploid genome size which, estimated at around 630 Mbp (Arumuganathan and Earle, 1991), is at least four times the size of the *A. thaliana* genome. This is thought to be the result of both replication, and the presence of repetitive DNA (Slocum *et al.*, 1990).

Comparisons of specific regions of genome between *A. thaliana* and *Brassica* demonstrate that at some levels genome structure is also largely conserved. Regions of colinearity in marker order have repeatedly been reported between replicated loci of *Brassica* and single loci of *A. thaliana* (Cavell *et al.*, 1998; Conner *et al.*, 1998; Lagercrantz *et al.*, 1996).

Genome structure and coding sequence conservation at such a high level suggests genetic models of molecular and developmental processes in *A. thaliana* development may be applicable to similar processes in *B. oleracea.*

1.7 Floral induction in *A. thaliana*

Analysis of *A. thaliana* has greatly increased the understanding of genetic and physiological events governing the process of floral initiation (for reviews see (Levy and Dean, 1998; Ma, 1998; Pidkowich *et al.*, 1999) and flower development (Bowman *et al.*, 1991; Coen and Meyerowitz, 1991). As discussed earlier (1.4), recent experiments have suggested that cauliflower curd develops at the specific ontological

stage of floral induction (Anthony *et al.*, 1996; Carr and Irish, 1997). Elucidation of the genetics controlling this stage of development in *A. thaliana* may therefore provide insights into possible genetic interactions governing curd development in the related *Brassica* cauliflower.

In *A. thaliana*, the floral initiation process (FLIP) is controlled by the interaction of several master homeotic genes. Genetic models of the floral initiation process have been constructed following characterisation of mutagenised *A. thaliana* plants (for reviews see (Ma, 1998; Pidkowich *et al.*, 1999)). Combinations of mutants were used in crossing experiments to identify interactions between loci, the majority of which represented Loss Of Function (LOF) mutations. Once cloned, ectopic expression of genes from ubiquitous promoters was also employed to test the effects of Gain Of Function (GOF) on phenotype (table 1.2). The current model constructed around the discrete switch from inflorescence to floral fate in wild-type *A. thaliana* is summarised in figure 1.2.

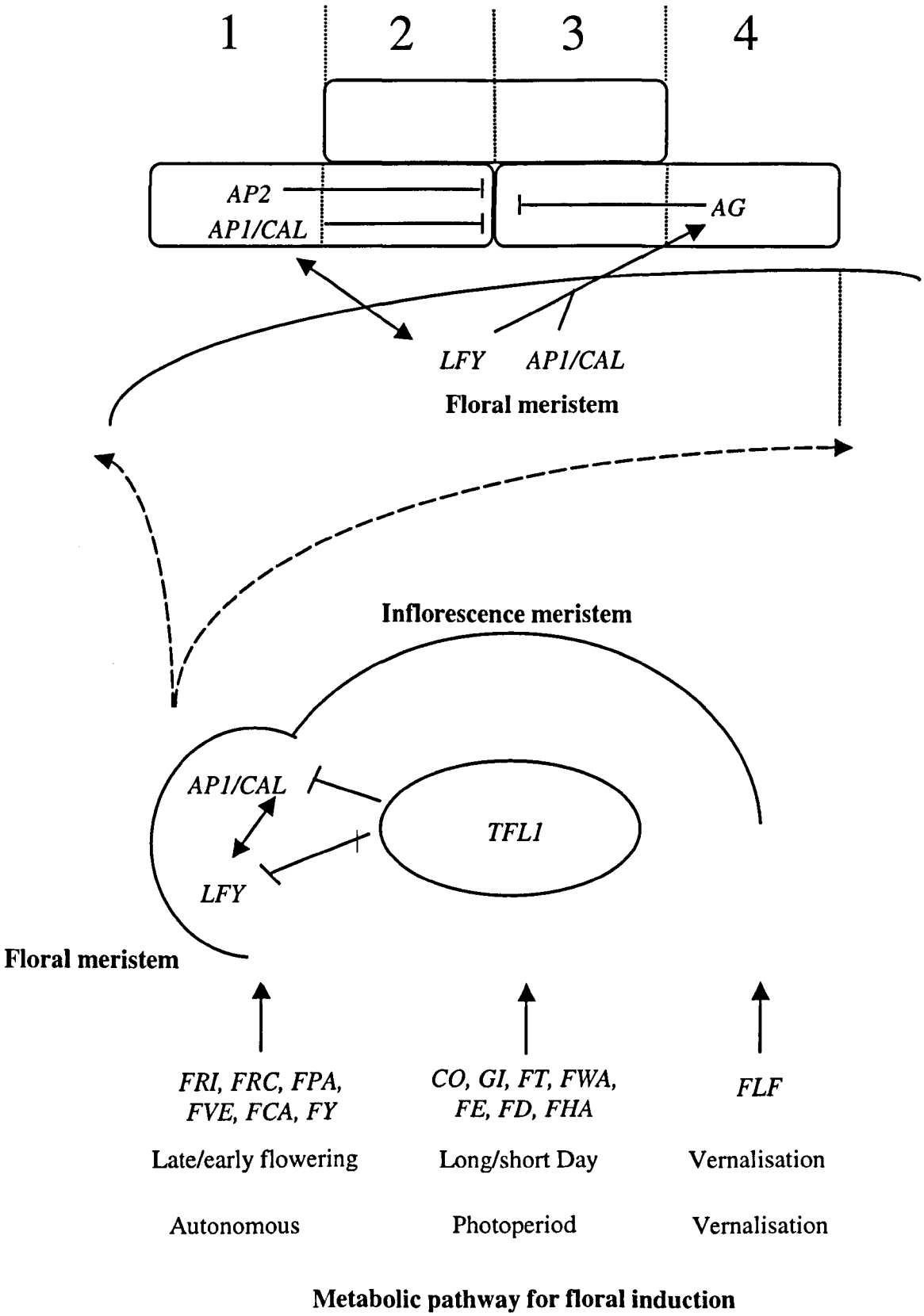


Figure 1.2: Schematic showing expression pattern and interactions between floral meristem identity gene products in *Arabidopsis thaliana*. Following inductive signalling from one of three identified metabolic

pathways (for review see (Levy and Dean, 1998), the peripheral zone of the inflorescence meristem begins to express the transcription factor *LFY*. This promotes floral initial development and results in the activation of the two closely related MADS-Box transcription factors *APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*) (Liljegren *et al.*, 1999). Once activated, *LFY*, *AP1* and *CAL* act to upregulate each others expression, whilst *TERMINAL FLOWER 1* (*TFL1*) acts indirectly to exclude expression of these three genes from the inflorescence meristem, thus maintaining its indeterminacy. *AP1* and *CAL* also act to prevent lateral meristem formation in the flower primordia ensuring the flower remains laterally determinate. Expression of *LFY CAL* and *AP1* activate *AGAMOUS* (*AG*) expression which acts to maintain terminal determinacy of the flower, as well as delineating the 'C' domain of the flower (table 1.2, fig. 1.2).

As the flower develops *AG* expression limits *CAL* and *AP1* to whorls 1 and 2 (sepal and petal). *AP1*, *CAL* and newly expressed *APETALA2* (*AP2*) act antagonistically to limit *AG* expression to whorls 3 and 4 (stamen and carpel). *LFY* acting with *UNUSUAL FLORAL ORGANS* (*UFO*) is thought to be involved in determining B function (petal and stamen) (Parcy *et al.*, 1998; Weigel and Meyerowitz, 1993), characterised by *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) expression (Goto and Meyerowitz, 1994; Jack *et al.*, 1994).

Table 1.2: Genetic and molecular experiments used to develop the current model of floral induction in wild-type *A. thaliana.* Descriptions given refer to specific functions at this ontological stage. Genes may also function elsewhere.

Gene	LOF	GOF	Deduced function in floral meristem induction
LFY	Partial floral → inflorescence conversion (Weigel *et al.*, 1992)	Early flowering. transforms inflorescence meristem to floral meristem (Weigel and Nilsson, 1995)	Controls the switch from inflorescence to flowers
AP1	Partial floral → inflorescence conversion (Bowman *et al.*, 1993)	Early flowering. transforms inflorescence meristem to floral meristem (Mandel and Yanofsky, 1995)	Controls the switch from inflorescence to flowers
CAL	No effect (Bowman *et al.*, 1993)	Early flowering. transforms inflorescence meristem to floral meristem (Liljegren *et al.*, 1999)	Controls the switch from inflorescence to flowers
LFY/AP1	Severe floral → inflorescence conversion (Weigel *et al.*, 1992)	35S:*LFY/ap1-1* (weaker than *35S:LFY*) *35S:AP1/lfy* (no change from *35S:AP1*) (Liljegren *et al.*, 1999)	Additive nature, suggests only partial functional overlap. GOF suggests *AP1* is downstream and activated by *LFY* (concurs with temporal expression pattern)
CAL/LFY	Partial floral → inflorescence conversion (Bowman *et al.*, 1993)	untested	*LFY* is epistatic to *CAL*
CAL/AP1	Severe floral → inflorescence conversion Lateral meristem formation in the flower primordia (Bowman *et al.*, 1993)	untested	*CAL* also controls switch from inflorescence to flowers, but LOF can be completely compensated for by *AP1* function
CAL/LFY/ AP1	Phenotype identical to *LFY/AP1* (Bowman *et al.*, 1993)	untested	
TFL1	Inflorescence → floral conversion (indeterminate → determinate) early flowering (Alvarez *et al.*, 1992)		In situ suggests *TFL* acts to exclude *LFY* and *AP1* from inflorescence meristem thus maintaining indeterminacy (Bowman *et al.*, 1993)
AG	Floral → inflorescence conversion under short days (Mizukami and Ma, 1997)	Early flowering. determinacy of inflorescence meristem	Key for switch from indeterminate to determinate meristem activity
UFO	Resembles a weak *lfy* mutant (Levin and Meyerowitz, 1995)		Acts to promote floral induction overlapping function with *LFY*
AP2	in short days similar to *ap1-1* LOF (Shannon and Meeks-Wagner, 1993)	N/A	Acts under certain conditions to promote floral induction

1.8 Identification of a curding *A. thaliana* mutant

The elucidation of the floral initiation process in *A. thaliana* involved analysis of combinations of many *A. thaliana* mutants. The role of *CAL* and *AP1* in preventing the initiation of secondary meristems from the floral meristem was identified through just such an analysis, and provided an interesting insight into cauliflower curd development.

In 1993 an *A. thaliana* mutant was reported that enhanced the effect of the *apetala1-1* (*ap1-1*) mutant phenotype (Bowman *et al.*, 1993). The gene was named *CAULIFLOWER (CAL)* and the mutant allele designated *cal-1*. A plant homozygous for *ap1-1/cal-1* shows more severe loss of control over floral induction than *ap1-1* plants alone. Each meristem that in a wild-type plant would give rise to a flower instead acts like an inflorescence meristem. These meristems themselves produce higher order meristems that behave in the same phylotactic spiral manner (Bowman *et al.*, 1993). This results in structures that closely resemble the phenotype of a typical *Brassica* cauliflower curd (figure 1.3).

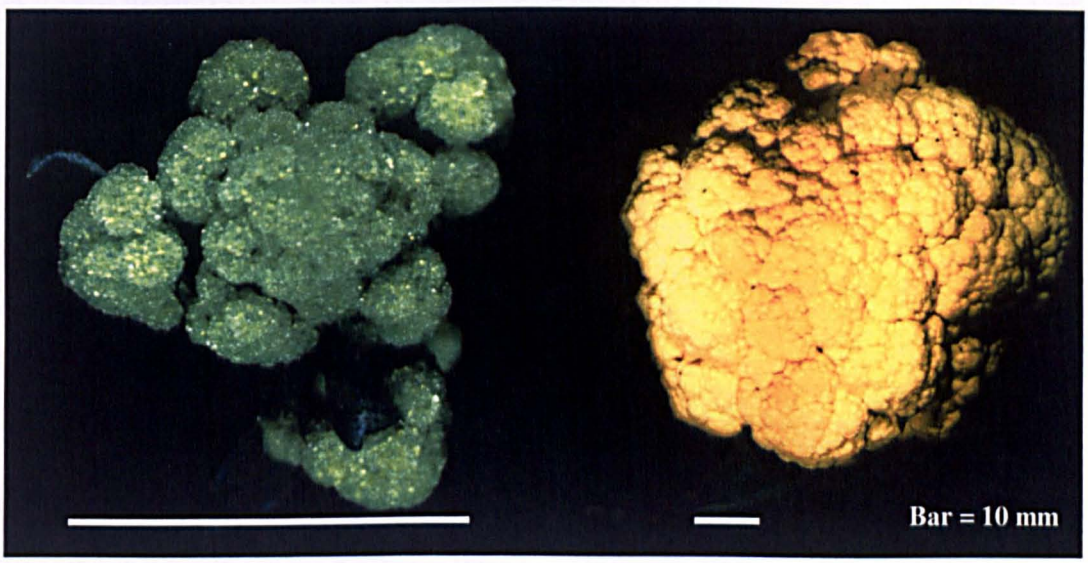


Figure 1.3: Visual comparison of the *A. thaliana ap1-1/cal-1* mutant inflorescence to the *Brassica* cauliflower curd.

1.9 *CAULIFLOWER*: an *A. thaliana* gene promoting floral induction

The *CAULIFLOWER* (*CAL*) gene from *A. thaliana* was cloned in 1994 (Kempin *et al.*, 1995). *CAL* was found to be a MADS-Box transcription factor. Phylogenetic analysis demonstrated that *CAL* groups with *AP1* in a subclade of the generic MADS-Box phenogram (Munster *et al.*, 1997) as the most closely related MADS-Box gene to *AP1* yet identified in *A. thaliana.* Amino-acid sequence similarity of these two proteins is 76%. This rises to 88% taking into account conservative substitutions (Kempin *et al.*, 1995).

1.10 The structure and function of MADS-Box genes

The MADS box family of transcription factors has been defined on the basis of primary sequence similarity amongst numerous proteins from a diverse range of Eukaryotic organisms including yeasts, plants, insects, amphibians and mammals (Theissen *et al.*, 1996). The MADS box motif is a highly conserved 56-amino acid region found within the DNA binding domain of these transcription factors. The name is derived from the initials of the first four members isolated, namely: *MCM1, AG, DEFA* and *SRF* (Scwarz-Sommer *et al.*, 1990).

The MADS box genes have significant biological roles. For example, the human *Serum Response Factor* (*SRF*) is involved in co-ordinating the proto-oncogene *c-fos* (Norman *et al.*, 1988), whilst *MCM1* plays a key role in the regulation of mating type specific genes in *Saccharomyces cerevisiae* (Acton *et al.*, 1997). In *A. thaliana*, MADS box proteins play fundamental roles in the control of plant architecture, including floral initiation (*AP1, CAL, AG,*), and flower development (*AP3, PI, AP1, CAL, AG,*). (for reviews see: (Reichmann and Meyerowitz, 1997; Shore and Sharrocks, 1995).

The structure of MADS-Box polypeptides can be separated into several distinct modules (Fig. 1.4). There is a variable size N-terminal region. In the human MADS-Box protein *SRF*, the presence of an N-terminal extension to the MADS-box has a profound effect on the specificity of DNA-binding (Pellegrini *et al.*, 1995). Although in the *A. thaliana* genes *CAL* (Kempin *et al.*, 1995) and *AP1* (Mandel *et al.*, 1992) an N-terminal

extension is completely absent and the protein begins with the MADS-Box. The 56-amino acid MADS domain is required for DNA binding and dimerisation (Pellegrini *et al.*, 1995). The region following the MADS-Box is designated I (intervening) or L (linker), and in plants, is involved in dimerisation and functional specificity (Reichmann *et al.*, 1996). The K domain follows this and is named for its sequence similarity to the coiled-coil of keratin (Ma *et al.*, 1991). The K-box is thought to stabilise dimers (Reichmann and Meyerowitz, 1997). Finally there is a variable carboxy-terminal sequence. This region is not required for functional specificity, as shown by experiments involving the swapping of C-terminal regions between *API*, *AG*, *AP3* and *PI*, which had no effect on the specific function of the proteins (Krizek and Meyerowitz, 1996). Variable sequence of the C-terminus region, between different MADS-Box genes along with high numbers of glutamine codons in this region suggests the C-terminus may act in regulating transcriptional activation (Reichmann and Meyerowitz, 1997). Indeed evidence for the activation of transcription by the C-terminus of *API* has been reported recently (Cho *et al.*, 1999).

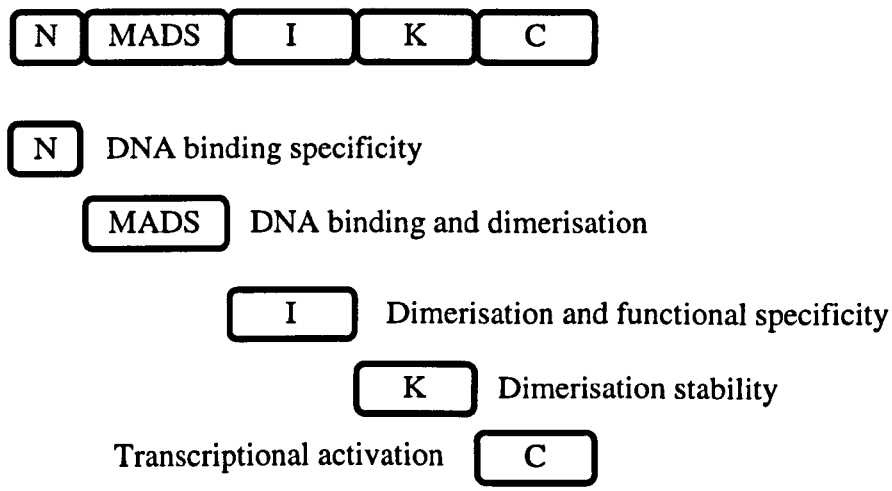


Figure 1.4: Modular structure and function of MADS-Box transcription factors

MADS-Box polypeptides act as transcription factors, binding as a dimer to specific regions of promoters. MADS-Box polypeptides have been shown to form both homo- and hetero-dimers prior to DNA binding (Reichmann *et al.*, 1996). It has been shown for

several genes that MADS-Box proteins act in these regions to bend the DNA of the target promoter (Acton *et al.*, 1997; West and Sharrocks, 1999; West and Sharrocks, 1997; West *et al.*, 1997). It has been suggested that this change in secondary structure of the DNA may allow further transcription factors to bind to the protein-DNA complex, or else allow access of RNA polymerase II to the DNA, thereby directing transcription (Pellegrini *et al.*, 1995).

1.11 Comparative floral induction in *A. thaliana* and the *Brassica* cauliflower

1.11.1 *BoCAL*, the *B. oleracea CAL* orthologue

The *A. thaliana* MADS-Box gene *CAL* was used as a probe to isolate the orthologous gene from *Brassica oleracea* var. *botrytis* (Snowball Y improved cauliflower), designated *BobCAL*, and from a rapid cycling line *(BoCAL)* (Kempin *et al.*, 1995). The allele isolated from the *Brassica* cauliflower was found to have a single base change in exon five that was not present in *BoCAL*. This resulted in an in-frame stop codon and theoretically a truncated product missing 105 of the total 255 amino acids. Similar stop codons in this region of the *A. thaliana* *AP1* gene result in plants having a severe *ap1* phenotype, suggesting the *BobCAL* allele is likely to be non-functional (Kempin *et al.*, 1995).

1.11.2 Comparative expression of meristem identity genes during curd development in *A. thaliana* and cauliflower

In both *A. thaliana* and the *Brassica* cauliflower the stability of the curd arrest stage is affected by temperature fluctuations (Bowman *et al.*, 1993; Fujime, 1983). In *A. thaliana* the arrest prior to floral initiation becomes more complete as the temperature drops. Conversely in cauliflower, arrest is successfully maintained at higher temperatures, with complete vegetative reversion above a cultivar-specific threshold (fig 1.5) (Anthony *et al.*, 1996). At these non-permissive temperatures, both *BoAP1* and the *Brassica LFY* homologue *BoFH*, and *AP1* and *LFY* expression is virtually undetectable (Anthony *et al.*, 1996; Bowman *et al.*, 1993).

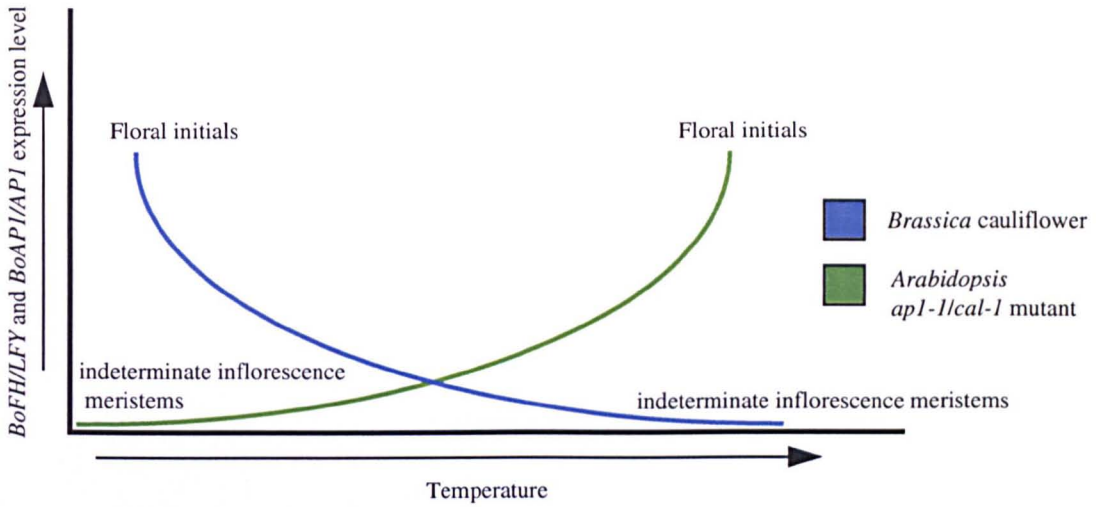


Figure 1.5: The effect of temperature on meristem identity gene expression and floral induction in the *Brassica* cauliflower and the *Arabidopsis ap1-1/cal-1* mutant.

In both *A. thaliana* and cauliflower, growth at non-permissive temperature induces bract development (Bowman *et al.*, 1993; Fujime, 1983). In *B. oleracea*, the presence of these axillary leaves protruding through the curd detract from the marketability of the cauliflower as they can number up to several thousand, forming a papillate covering across the curd (Crisp and Tapsell, 1993). Although temperature fluctuation promotes development of bracting, in cauliflower this trait has a heritability of 73%, based on regression analysis of progeny on parent in the field (Crisp *et al.*, 1975) This suggests there is a strong genetic component. In the *A. thaliana ap1-1/cal-1* mutant, bract development is known to be associated with low levels of *LFY* expression (Bowman *et al.*, 1993), suggesting that the similar bracting trait of cauliflowers could be the result of low levels of *BoFH* product.

At the permissive higher temperature of 25 °C in the *A. thaliana ap1-1/cal-1* mutant, expression of the meristem identity genes *AP1* and *LFY* is much reduced when compared to that in either *ap1* or *lfy* single mutants. This suggests that in wild-type, *CAL* may act to upregulate expression of both *LFY* and *AP1* (Bowman *et al.*, 1993).

Although much reduced in the *ap1-1/cal-1* mutant, at the permissive higher temperature, *LFY* expression is still at a level sufficient to inhibit bract development

(Bowman *et al.*, 1993). *LFY* activity gradually accumulates, eventually resulting in the production of *ap1-1* mutant flowers. *ap1* product also increases over time, although interestingly in the higher order meristems first, i.e. those that will eventually form flowers. Analysis of the orthologous genes *BoFH* (*B. oleracea Floricaula Homologue*) (Anthony *et al.*, 1993), and *BoAP1* (Anthony *et al.*, 1995; Carr and Irish, 1997) demonstrated exactly the same distinct pattern of increase in expression in cauliflower curd at permissive lower temperatures.

Expression studies at permissive temperatures demonstrate that *CAL* and *BoCAL* have similar expression patterns in the *A. thaliana ap1-1/cal-1* mutant and the *Brassica* cauliflower respectively (Kempin *et al.*, 1995). In both, *BoCAL* and *CAL* mRNA accumulates in young flower primordia but unlike *AP1* is found only at low levels in sepals and petals. Expression patterns of *CAL* and *BoCAL* have not been reported at restrictive temperatures.

Comparative analysis of meristem identity gene expression suggests there is a strong correlation between the processes governing curd development in both the *Brassica* cauliflower and the *A. thaliana ap1-1/cal-1* mutant. This suggests that the application of models of floral induction developed in *A. thaliana* to the *Brassica* cauliflower is not unreasonable, and may provide insights into theoretical models that could be tested by experimentation in *Brassica.*

1.11.3 Theoretical models

The similarity between the *ap1-1/cal-1* double mutant *A. thaliana* and the cauliflower curd, along with the discovery of a non-functional copy of *BoCAL* in var. *botrytis,* suggested that the curding phenotype of cauliflower may be at least in part due to a mutation in the *BoCAL* locus (Kempin *et al.*, 1995).

If *BoCAL* is involved, models proposed both by Anthony *et al.* (1996) for cauliflower, and Bowman *et al.* (1993) for the *ap1-1/cal-1 A. thaliana* mutant, require a threshold level of *AP1* (*BoAP1*) and *LFY* (*BoFH*) meristem identity gene transcript level to be reached for floral induction. Under this scenario, cauliflower curd could be considered

a transient stage, until *BoAP1* and *BoFH* expression reach sufficient levels to induce downstream factors. From analogy to *A. thaliana* (Bowman *et al.*, 1993), this process would take longer when *BoCAL* is non-functional, as the co-operative upregulation of expression would be inhibited, presumably resulting in development of the curd phenotype.

The time taken to reach this threshold is also temperature sensitive (Anthony *et al.*, 1996; Bowman *et al.*, 1993). As the temperature moves away from the restrictive threshold (increases in *A. thaliana*, decreases in cauliflower) the control of arrest falters. This results in development of floral primordia from the surface of the curd.

In *Brassica*, this faltering of arrest is seen to a greater or lesser extent in many cauliflower cultivars, and is well characterised as 'riceyness' (Fujime, 1983). Riceyness in cauliflowers has a large environmental component and is responsive to temperature fluctuations, with loss of arrest at the curd stage occurring when plants are grown below a cultivar-specific threshold (Fujime, 1983). However, several studies report a good response to selection in progeny from crosses between ricey and non-ricey cauliflowers, suggesting strong heritability, with current genetic models predicting only a few genes underlying this trait (Crisp and Tapsell, 1993; Dickson and Lee, 1980; Watts, 1966).

The model of *Brassica* cauliflower curd development constructed from comparative meristem identity gene analysis, follows closely the model constructed for *A. thaliana* curd development. However, there is an inconsistency between the two models regarding final flower structure. In *A. thaliana*, *AP1* or *LFY* expression is insufficient to produce fully developed flowers (Weigel *et al.*, 1992; Bowman *et al.*, 1993). Both gene products are required to direct wild-type floral organogenesis as the *ap1-1/cal-1* mutant produces *ap1-1* type flowers (Bowman *et al.*, 1993). In the *Brassica* cauliflower however, all flowers are wild-type. If the models of curd ontogeny are generically identical in the *A. thaliana ap1-1/cal-1* mutant and the *Brassica* cauliflower, *BoAP1* should have function at later ontological stages in the *Brassica* cauliflower.

An alternative genetic model would incorporate the overall replication of the *Brassica* genome when compared to *A. thaliana* (Slocum *et al.*, 1990). The possibility that more than one copy of each meristem identity gene may be present in the *Brassica* genome raises the question of functional redundancy or even divergence between copies, with control of meristem identity and floral organ identity split between loci.

The stepwise inflorescence development of other heading brassicas is very similar to that of cauliflower, although the stage at which arrest occurs varies. The Sicilian Purple cauliflower develops a cauliflower-type curd, but this is covered with immature flower buds (Gray, 1989). The Calabrese broccoli also has a small curding stage (Kieffer *et al.*, 1998) followed by arrest after full bud development (Carr and Irish, 1997). The length of this arrest may also be affected by temperature (Bjorkman and Pearson, 1998). Comparative analysis of the expression patterns of floral meristem identity genes in cauliflower and *A. thaliana* suggests they may also contribute to the arrest seen at later ontological stages in these other morphotypes.

1.12 Introgression of desirable traits into cauliflowers

The development of horticultural forms in well defined cultivars by modern breeders, and previously by local farmers developing landraces, has involved the introgression of desirable traits into the crop genepool (for review see Crisp, 1983). The major limiting factor for introgression such as pathogen resistance into crops, is the haplotype dilution and subsequent difficulty in resynthesising the edible portion to a quality that is satisfactory to the consumer (reviewed in (Crisp, 1983). In *Brassica,* as a general rule, introgression of desirable traits has been more successful where the donor is more closely related to the recipient genepool (Crisp, 1983). Wide crosses, such as attempts to introgress clubroot resistance from kale to cauliflower, has resulted in material of poor curding quality that has failed to reach market (Weisaeth, 1974). The same is true of introgression from cabbage sources to cauliflower and broccoli (Gallegly, 1956; Vriesenga and Honma, 1971). It has been estimated that satisfactory transfer of a variable trait such as clubroot resistance between a wide cross such as kale to cauliflower would require a five hectare field, a glasshouse and vernalisation facility, and considerable labour input over a period of some ten years

before breeding of finished cultivars could begin (Toxopeus, 1977).

The development of molecular genetic technologies has promised to revolutionise the approaches taken during crop improvement programs. Molecular markers for traits of interest are being developed for many distinct crops (Kaneko *et al.*, 1999; Mazur *et al.*, 1999; Quarrie *et al.*, 1999; Seyfarth *et al.*, 1999) including *Brassica* (Jourdren *et al.*, 1996; Song *et al.*, 1995). If successful, the development of molecular markers for the curding habit of cauliflowers may allow rapid resynthesis of phenotype during introgression programs.

1.13 The domestication of cauliflowers

The documented history of the domestication of cauliflowers from its wild progenitors is far from complete, relying for the most part on ancient text and drawings to reconstruct the sequence of events (Crisp, 1982; Crisp and Tapsell, 1993).

The earliest extant records that refer to brassicas containing identifiable floral differences come from early Greek and Roman authors who referred to sprouting forms of cabbages (Hooper and Ash, 1935). The earliest reference to a heading type was by Pliny (23-79 AD, quoted in Crisp, 1982) and the first distinction between the two types was in the 12th century by Ibn-al-Awan, a Spanish Moor (Crisp, 1982). European 16th century herbalists drew pictures and descriptions of heading forms, but even these picture sources are scarce and little of what could be positively identified as a cauliflower exists before the seventeenth century (Zeven, 1996). As with all historical sources however, meaning can be lost during translation and interpretation and it is wise to be cautious regarding the reliability of this evidence.

The most prevalent cauliflower domestication theory is of an Eastern Mediterranean origin, but is based on poorly defined 'heading' inflorescence structures (Gates, 1953; Hyams, 1971; Snogerup, 1980). Under this theory, a cauliflower precursor was introduced into Italy, around 1490 (Thompson, 1976), where a secondary centre of diversity arose with development of distinct local forms, isolated by geography and strong trade barriers between the many regional states that today form modern Italy

(Massie, 1998).

The geographical distribution of cauliflower and broccoli crop morphotypes in the regions of modern Italy has been examined on several occasions (Branca and Iapichino, 1997; Crisp and Gray, 1984). Recent taxonometric study of 40 curd and leaf traits in Italian cauliflowers and broccolis confirmed that in spite of the strong outbreeding strategy, the historical eco-geographic groupings and relationships amongst the local Italian varieties still exist (Figure 1.1) (Massie *et al.*, 1996).

The conservation of these groupings suggests that analysis of the genetics associated with the development of curd may allow a reconstruction of the events leading to the domestication of the *Brassica* cauliflower phenotype from a wild progenitor, independent of evidence supplied by questionable historical sources.

1.14 Hypothesis

The proposed hypothesis is that the genetics underlying the curd trait of the *Brassica* cauliflower are generically identical to that described for the development of a similar phenotype in the related *A. thaliana ap1-1/cal-1* mutant (Bowman *et al.*, 1993). Thus that single loci of *BoCAL* and *BoAP1* will prove to be the contributing genes.

By inference, this hypothesis also suggests that any further replicated loci discovered will not be redundant at this stage of development, but have become non-functional or else have diverged to ensure wild-type flower development.

1.15 Objectives

1. Develop an understanding of the genetic interactions responsible for the curding habit of cauliflowers and construct a genetic model.
2. Utilise molecular genetic techniques to address the conflicting debate regarding the classification of broccoli and cauliflower types within the *B. oleracea* genepool.
3. Compare the genomic distribution and the functional status of each copy of the meristem identity genes *BoCAL* and *BoAP1* in *Brassica oleracea*
4. Utilise knowledge gained from molecular genetic analysis to examine the genetic relationship between broccoli and cauliflower and trace the domestication of the *Brassica* cauliflower curd trait.

1.16 Rationale

The primary approach taken was to exploit the genetic relatedness between *A. thaliana* and *Brassica* to investigate the genetics underlying the curd phenotype of cauliflower. Previous examination of the *A. thaliana ap1-1/cal-1* mutant, together with analysis of meristem identity genes in both *A. thaliana* and the *Brassica* cauliflower suggested these genes may be involved in the development of curd. This analysis would also clarify the taxonomic debate, by directly testing the level of genetic distinction between the

inflorescence morphotypes of cauliflower and broccoli.

Evidence from genetic mapping projects in *B. oleracea* has suggested that many loci that are single copy in *A. thaliana* are multicopy in *Brassica*. This may have serious implications for the control of floral induction in the *Brassica* cauliflower. Comparative analysis between *A. thaliana* and *B. oleracea* was carried out to test the possibility of functional diversification or redundancy between several copies of the genes involved in the development of cauliflower curd.

Confirmation that the historical eco-geographic groupings and relationships amongst the local Italian cauliflower and broccoli varieties still exist (Massie *et al.*, 1996) suggested that identification of the genes specifically associated with the curding trait may allow tracing of the domestication of cauliflowers, in a manner similar to that reported recently for maize (Rong-Lin *et al.*, 1999).

1.17 Experimental approach

Molecular and classical genetics were employed, initially to develop a genetic model of curd development in a doubled haploid (DH) population arising from a cross between a DH cauliflower and a recurrent inbred Calabrese broccoli. Segregation of candidate genes was then tested against phenotypic segregation to identify the likely contributing loci.

Once identified, these genes were utilised during screening of a *B. oleracea* BAC library to identify further copies of these candidate loci. These loci were physically and genetically mapped where possible, and compared to the homologous region from *A. thaliana* to ascertain genome conservation. Expression studies were used to examine the possibility of functional redundancy of multiple loci; and DNA sequence comparison was employed to determine possible contributing mutations.

Diagnostic markers developed from one of the putative causal loci were applied to analysis of over 200 *B. oleracea* accessions to examine their utility as molecular markers, the strength of the proposed genetic model, and the possible domestication of

the curd trait within Italy, the centre of diversity for cauliflower development.

Further genetic analysis was employed to preliminarily characterise a distinct trait that contributes to the overall inflorescence morphologies of the heading brassicas.

CHAPTER2: MATERIALS AND METHODS

2.1 Plant material

2.1.1 NxB doubled haploid (DH) population

2.1.1.1 First planting

Five seeds from each of the Cauliflower parent (DJ7032), the Broccoli parent (BI88908), the F1 of the cross and 61 DH progeny lines derived from anther culture of the F1, were sown in 2.5 cm^2 hassie tray modules containing Levington F2 compost. The plants were placed in the glasshouse to allow germination with an average night temperature of 10 °C and a day venting temperature of 20 °C. Supplementary lighting was provided at 300 µmol photons m^{-2} s^{-1} (400-700 nm, supplied by thorn 85 W white fluorescent tubes) up to 14 hour daily light periods. At the six leaf stage two of each of the plants were repotted into 13 cm diameter round pots containing Levington M2 and 5 mm pea shingle in a ratio of 3:1 and placed back under glass. Plants were watered daily until scoring was complete.

2.1.1.2 Second planting

Plants were grown exactly as the first trial with the exception of being moved from the glasshouse to the cold-frame following repotting at the six leaf stage. No supplementary lighting was given after this point. Plants were watered daily until scoring was complete.

2.1.2 A12DHxGD33 DH mapping population

Five seeds of each of the Chinese White Kale parent (A12DH), the Calabrese parent (GD33) and 169 DH lines derived from anther culture of the F1 were sown in 2.5 cm^2 hassie tray modules containing John Innes F2 compost. The population was grown and manipulated exactly as for the NxB first planting until scoring was complete.

2.1.3 NxG mapping population

Five seeds of each of the Cauliflower parent (DJ7032), the Brussels Sprout parent (DJ3753) and 97 DH lines derived from anther culture of the F1 were sown in 2.5 cm^2

hassie tray modules containing John Innes F2 compost. The population was grown and manipulated exactly as for the NxB first planting until scoring was complete.

2.1.4 A12DHxN F2

One hundred and twenty F2 plants derived from selfing of the F1 from the cross between the Cauliflower (DJ7032) and the Chinese White Kale (A12DH), were sown in 2.5 cm^2 hassie tray modules containing John Innes F2 compost. The population was grown and manipulated exactly as for the NxB first planting until scoring was complete.

2.1.5 *Arabidopsis thaliana ap1-1/cal-1* mutant

Arabidopsis plants were grown as recommended by the *Arabidopsis* Biological Resource Centre (ABRC) in a controlled environment room at a constant temperature of 25 °C. Supplementary lighting was provided at 120 µmol photons m^{-2} s^{-1} (400-700 nm, supplied by Growlux and Warmwhite fluorescent tubes in a 3:2 ratio) with an artificial day length of 16 hours.

2.1.6 Diversity study

Seed of all of the *Brassica* accessions used were obtained from the HRI Genetic Resources Unit (HRIGRU, Wellesbourne, Warwick, CV35 9EF, UK), with the exception of a pedigree line from Yates Vegetable Seeds Pty. Ltd (Narromine, N.S.W. 2821, Australia). Ten seeds of each accession were sown in 2.5 cm^2 hassie tray modules containing John Innes F2 compost. The population was grown and manipulated exactly as for the NxB second planting, with the exception of transplantation to the field following three weeks in the cold-frame. Two plants from each accession were planted 0.5 m apart in rows of four accessions, resulting in a total field plot of 250 m^2. Guard plants surrounded the entire plot at 0.5 m distance to reduce edge effects. Plants were watered three times a week until scoring was complete.

2.2 Molecular analyses

2.2.1 Chemicals and reagents

Unless specifically stated, all chemicals and reagents were obtained from Sigma-Aldrich Ltd. Addresses of all reagent providers are given in appendix 3.

2.2.2 Restriction endonucleases

PCR products were digested in reactions containing 20 µl PCR reaction, 2.5 µl 10X buffer, 1 unit of enzyme and H_2O to 25 µl. (any recommendations by manufacturers, e.g. addition of Bovine Serum Albumin (BSA) during specific digests, were followed). Samples were incubated at enzyme specific temperatures for 4 hours. BAC digests were conducted with similar ratios of reagents in volumes adjusted to account for lower template yield.

Table 2.1: Restriction enzymes used during this research and respective suppliers.

Supplier	*Restriction endonuclease*
Roche diagnostics Ltd	*Spe*I, *Mbo*II,
Promega UK Ltd	*Dra*I, *Eco*RI, *Bam*HI, *Kpn*I, *Xba*I, *Hin*dIII *Not*I
New England Biolabs UK Ltd	*Tsp*RI

2.2.3 DNA extraction

Leaf tissue was taken from each plant at the two leaf stage and stored at –80 °C until required. Small amounts of genomic DNA required for PCR analysis were extracted from plant tissue employing a Cetyltrimethyl-ammonium bromide (CTAB) miniprep method (Doyle and Doyle, 1990).

2.2.3.1 DNA extraction for RFLP analysis

Large quantities of high quality genomic DNA required for RFLP analysis were extracted from leaf tissue utilising the phenol/chloroform based method described in (Bohuon *et al.*, 1996).

2.2.3.2 Estimation of concentration

DNA concentration was estimated using a bisbenzimide assay in a Hoeffer 200 DNA fluorimeter and adjusted with dH_2O to a final concentration of 10 μg ml^{-1} for PCR.

2.2.4 Polymerase Chain Reaction (PCR) Amplification

All PCR reactions were set up thus: final free magnesium (Life Technologies) concentration of 1.5 mM-3.5 mM depending on specific primers used (appendix 1), one unit of *Taq* DNA polymerase (Life Technologies), 4 pmol of each primer (Life Technologies), 2 μl 10x PCR buffer (Life Technologies) 500 pmol of each of the four dNTP's, 20 ng of template DNA, and sterile H_2O to a total volume of 20 μl. PCR cycling was conducted on omnigene thermocyclers (Hybaid) which required mineral oil overlay of samples to prevent evaporation. Temperature was regulated under 'block' control as this was considered to be the most reproducible setting. Amplification itself consisted of 94 °C 90 seconds, and then 35 cycles of X °C 60 seconds, 72 °C Y seconds, 94 °C 60 seconds, with a final 10 minute extension step at 72 °C (X varied from 51-63 °C and Y vary from 60-120 seconds due to primer design and expected amplicon length respectively).

2.2.4.1 Agarose gel electrophoresis

Following cycling, PCR reactions were mixed 4:1 with loading buffer, (1x TBE, 1 M sucrose, bromophenol blue), samples were subjected to electrophoresis on 0.5x TBE, 1% agarose gels (Seakem GTG, FMC Bioproducts) at 7.5 V cm^{-1} for 90 minutes, with the exception of the amplicon from AP15PF-AP15PR, which was run on a 3% low melting temperature (Nusieve GTG, FMC Bioproducts) agarose gel at 7.5 V cm^{-1} for 300 minutes. For Southern blotting from agarose gels, samples were subjected to electrophoresis through a 0.7% agarose gel (Seakem GTG, FMC) at 1.5 V cm^{-1} for 18 hours. A size standard was loaded on each gel to aid resolution (100 bp, 1 Kbp ladder, Life Technologies). Agarose gels were washed in H_2O containing ethidium bromide at a final concentration of 270 ng ml^{-1} for 20 minutes, washed in dH_2O for 20 minutes and visualised on an ultraviolet light box at 302 nm wavelength (UVP). Gel images were captured using a camera linked to a UVP image store 5000 (UVP).

2.2.5 Cloning

All amplicons cloned prior to sequencing or for hybridisation purposes were ligated into the pCRII vector supplied with the TA cloning kit (Invitrogen). Conditions followed manufacturer's recommendations, with a 3:1 insert:vector ratio used during ligation. INV-αF' cells were transformed with plasmids and transformants selected by the introduction of ampicillin (50 μg ml^{-1}). Blue-White selection was carried out and suspected transformants were analysed by both restriction digest with *Eco*RI (manufacturers conditions, Roche Diagnostics) and DNA sequencing from the M13 consensus primers.

2.2.5.1 Plasmid purification

Plasmids were purified from 10 ml liquid cultures (LB containing ampicillin at 50 μg ml^{-1}) using a plasmid miniprep extraction kit (Qiagen), following manufacturers conditions. Plasmids were resuspended in 20 μl dH_2O and stored at –20°C until required.

2.2.5.2 Plasmid quantification

Plasmid concentration was estimated by comparison to known concentrations of calf thymus DNA, following electrophoresis and staining as 2.2.4.1.

2.2.6 DNA sequencing

Cycle sequencing was employed according to manufacturer's instructions (Perkin Elmer-ABI). Bigdye fluorescently labelled di-deoxyNTPs were chosen in place of dRhodamine to give extended sequence reads. Following cycle sequencing, samples were purified according to the ethanol/NaOAc precipitation method (ABI sequencing handbook) with a second 70% ethanol wash added to reduce salt levels in the sample. Samples were resuspended and subjected to electrophoresis on an ABI377 DNA sequencer according to manufacturer's instructions.

2.2.7 SSR analysis

CALSSRF was labelled with the fluorescent tag 6-Carboxyfluorescin (6-FAM) (PE-ABI). PCR amplifications between CALSSRF-CALSSRR involved 35 cycles of 94 °C for 30 seconds, 63 °C for 30 seconds, with a final 10 minute extension step at 72 °C.

Analysis of the length polymorphism within the *BoCAL-a* intron seven simple sequence repeat (SSR) was carried out on an ABI 377 DNA sequencer (PE-ABI) and analysed using the Genescan software package, in accordance with manufacturers recommendations. A Rox350 fragment size standard was added to each sample prior to electrophoresis. Analysis of allele lengths was based on this standard, ignoring the 140 bp band due to interactions between this and the 140 bp SSR allele. All allele sizes were adjusted by one base to account for the incorporation of a template independent dATP by *Taq* DNA polymerase (Clark, 1988). This was confirmed by sequencing alleles from several accessions.

2.2.8 Denaturing Polyacrylamide Gel Electrophoresis (P.A.G.E.)

Oligonucleotide primers were end labelled with ^{33}P-dATP (Amersham), catalysed by *Polynucleotide Kinase* (Life Technologies) according to manufacturers recommendations.

Acrylamide and bisacrylamide was dissolved in a solution of 1x TBE and 7.5 M urea to a final percentage of 6:1.39 respectively. 450 µl of 10% ammonium persulphate and 45 µl Temed were added to 65 ml of the acrylamide solution and mixed. The gel was poured between two 30x20 cm gel plates (sealed with electrical tape around three sides) spaced to give a gel thickness of 0.4 mm. Shark tooth combs were inserted and the gel allowed to set for 60 minutes. 5 µl of each ^{33}P labelled sample was mixed with 5 µl load dye (100% formamide plus xylene cyanol and bromophenol blue markers). Samples were denatured at 94 °C for 180 seconds and snap cooled on ice prior to loading. The gel was pre-warmed to 50 °C prior to sample loading (Model S2 gel apparatus, Life Technologies). Electrophoresis was carried out at 7.5 V cm^{-1} for 180 minutes. Following electrophoresis the top siliconised plate was removed and the gel transferred to 3MM paper. Gels were then dried for 120 minutes in a vacuum gel drier (Model 583, Biorad). Dried gels were moved to autoradiograph cassettes (GRI) and exposed to film (Fuji RX) at room temperature for 4-7 days.

2.2.8.1 Film developing

Autoradiograph films were processed on a RPN1700 hyperprocessor (Amersham) according to manufacturer's instructions.

2.2.9 ^{32}P labelled hybridisations

Hybridisation and analysis of all ^{32}P-labelled probes during this research was conducted as outlined below.

2.2.9.1 Southern blotting

Agarose gels (0.7%, 2.2.4.1) were washed in depurination solutions (0.4 M HCl) for 15 minutes. They were then moved to a bath of neutralising solution (0.2 M NaOH) for twenty minutes prior to membrane transfer. Gels were capillary blotted overnight from a solution of 0.2 M NaOH onto a charged nylon membrane (Hybond N+, Amersham) in accordance with standard protocol (Sambrook *et al.*, 1989). Following transfer, membranes were washed in 2x SSC (Sambrook *et al.*, 1989), baked at 80°C for 2hours and stored in 2x SSC until required.

2.2.9.2 Prehybridisation

Prehybridisation solution contained final concentrations of 1 M sodium phosphate ($NaHPO_4$) pH 7.2, 20% w/v Sodium Dodecyl Sulphate (SDS), 10% w/v BSA, and 0.5 M Ethylenediaminetetra-acetic acid (EDTA) pH 8.0. Sheared salmon sperm DNA (50 $\mu g\ ml^{-1}$) was heated to 100 °C for five minutes and then added to pre-hybridisation solution pre-heated to 65 °C. Nylon membranes to be probed were then added.

2.2.9.3 Probe labelling

Probes for radioactive hybridisation were labelled with ^{32}P-dCTP (Amersham). This was incorporated during a nick translation reaction conducted in accordance with manufacturers instructions (Promega, cat No. U1001). Following labelling, the entire probe mix was heated to 94 °C for 180 seconds and added to the pre-hyb/membrane solution and incubated for at least 18 hours in a shaking incubator set at 65 °C.

2.2.9.4 Washes

Following hybridisation, membranes were washed twice in solutions of 2x Saline Sodium Citrate (SSC), 0.1% SDS (Sambrook *et al.*, 1989) at 65 °C for twenty minutes. Higher stringency washes were then carried out with final washes ranging from 1x SSC to 0.1x SSC dependent on specific hybridisation strength. Membranes

were then exposed to X-ray film (Fuji RX) and stored at –80 °C for 2-5 days to allow visualisation of hybridisation results.

2.2.10 BAC extraction from bacterial culture

Bacteria (*E. coli*, strain DH5α, Life Technologies) were cultured overnight at 37 °C in LB containing chloramphenicol at 12.5 μg ml^{-1} in an incubator shaking at 200 rpm. The samples were centrifuged at 3200 rpm (2200 x*g*) for 5 minutes and the supernatant removed. Samples were resuspended in residual LB and transferred to 2x1.5 ml eppendorf tubes. The cultures were subjected to centrifugation briefly in a bench-top microfuge, and the LB removed completely. The bacteria were resuspended in 300 μl of solution I (50 mM Tris.Cl pH 8.0, 50 mM glucose and 10 mM EDTA). 600 μl of solution II (0.2 M NaOH, 1% Triton X-100) was then added and the sample mixed gently. The lysis step was allowed to proceed for 5 minutes before addition of 450 μl of Solution III (3 M NaOAc). The samples were mixed gently and placed on ice for 30 minutes. The samples were then subjected to centrifugation at maximum speed (13000 rpm) in a bench top microfuge for 30 minutes at 4 °C. The supernatant was poured to fresh tubes and mixed with an equal volume of phenol/chloroform/iso-amyl alcohol (25:24:1). The samples were then subjected to centrifugation at maximum speed in a bench top microfuge for 30 minutes. The top aqueous phase was carefully pipetted into fresh tubes. Nucleic acids were precipitated in 1:0.1:2.5, sample/NaOAc (pH 5.3)/100% ethanol. The samples were placed on ice for 30 minutes then subjected to centrifugation at maximum speed in a bench top microfuge for 30 minutes at 4 °C. The supernatant was removed and 2x 70% ethanol washes were carried out. The samples were dried and resuspended in 30 μl H_2O. 0.2 μl RN*ase* A (100 μg ml^{-1}) was added to each sample and incubated at 37°C for 30 minutes. Samples were then placed at –20 °C until required.

2.2.11 Fluorescent *In Situ* Hybridisation (FISH)

Tissue for FISH analysis was harvested from A12DH immature stamens, these were fixed as entire buds in Carnoys solution (6:3:1 ethanol:chloroform:acetic acid) Stamens were dissected out and a sample tapped out in aceto-orcein and examined to determine developmental stage. The remaining anthers were washed six times with

ethanol:acetic acid (3:1) to remove the fix and three times in 0.01 M citrate buffer pH 4.5 (44.5:55.5, citric acid 0.1 M:sodium citrate 0.1 M, 1:10 dilution). The tissue was placed at 37 °C in citrate buffer containing 0.3% w/v *pectolyase*, 0.3% w/v *cytohelicase* and 0.3% w/v *cellulase* for two hours. Reactions were stopped by transferring to ice-cold H_20. Anthers were tapped out on clean slides in 20 µl 60% acetic acid to make the cells transparent. Slides were placed on a hot block (45 °C) and stirred gently for 60 seconds with a second aliquot of 60% acetic acid added after 30 seconds. Slides were then washed with 200 µl of Carnoy's fix and air dried.

2.2.11.1 Probe labelling

Probes were labelled with Digoxigenin-dUTP (DIG) during a nick translation reaction according to manufacturer's recommendations (Roche Diagnostics cat No. 976776).

2.2.11.2 *In Situ* Hybridisation

Slides were washed in 2x SSC for 10 minutes, then covered with 20 µl of *RNase* (100 µg ml^{-1}) and incubated at 37 °C for 30 minutes. Slides were then washed 3 times in 2x SSC for 5 minutes, then in a solution containing *pepsin* (0.01% w/v) at 37 °C for 90 seconds. Samples were then post-fixed in a solution of 4% paraformaldehyde pH 8.0, rinsed in dH_2O and dehydrated in 70%, 85% and 100% ethanol at room temperature for 2 minutes each. Slides were dried and placed at 37 °C until application of the probe.

Twenty microlitres of a solution containing 14 µl of master mix (5 ml deionised formamide, 1 ml 20x SSC and 1 g dextran sulphate made up to 7 ml with H_2O, pH 7.0) and 6 µl labelled probe, was warmed to 37 °C and applied to the slides. The slides were sealed and placed at 75 °C for 5 minutes, then moved to 37 °C and left overnight.

The slides were unsealed and washed 3 times in a solution of 50% formamide, 2x SSC, once in 2x SSC and once in 4x SSC, 0.05% Tween 20, at 45 °C for 5 minutes each. The slides were then washed with 4x SSC, 0.05% Tween 20 at room temperature for 5 minutes.

Anti-DIG (Roche diagnostics) was mixed in a 1:20 ratio with DIG block. 100 µl was applied to each slide, covered with parafilm and incubated for 30 minutes. Three washes in 4x SSC, 0.05% Tween 20 at room temperature for 5 minutes were carried out. Seven microlitres of DAPI/vectorshield solution was added and the slides covered with a coverslip prior to analysis.

2.2.12 *Reverse Transcriptase*-Polymerase Chain Reaction (RT-PCR)

2.2.12.1 Harvesting of tissue

DJ7032 Cauliflower curd was shaved to give 4g of meristematic tissue for RNA extraction. This tissue was snap frozen in liquid nitrogen and stored at –80 °C Harvesting was conducted at both the arrested curd stage and following the visual identification of floral bud primordia across the curd surface.

2.2.12.2 RNA extraction

Curd tissue was ground to a fine powder in liquid nitrogen and mixed with extraction buffer (50 mM Tris pH 7.6, 150 mM LiCl, 5 mM EDTA, 1% SDS, in DEPC treated water) in a 1:3 w/v ratio respectively. After vigorous shaking, phenol:chloroform:isoamyl alcohol (PCA), 25:24:1 was added to the samples in a 1:1 ratio respectively. Samples were subjected to centrifugation for 10 minutes at 3200 rpm (2200 x*g*). The top phase was carefully drawn into a fresh tube and two further PCA extractions were conducted. A further extraction of sample:chloroform in a 1:1 ratio was conducted, before LiCL was added to a final concentration of 2 M. The samples were left overnight at 4 °C to precipitate the RNA. Samples were then centrifuged for 60 minutes at 3200 rpm (2200 x*g*). The supernatant was removed and the pellet resuspended in 4 ml Diethylpyrocarbonate (DEPC) treated water. An ethanol precipitation was conducted in a ratio of 1:0.1:2.5, sample/NaOAc (pH 5.3)/100% ethanol. Samples were placed at –20 °C for 60 minutes, then subjected to centrifugation for 10 minutes at 3200 rpm (2200 x*g*) at a temperature of 4 °C. Pellets were resuspended in 5 ml 70% ethanol, and subjected to centrifugation as before. The supernatant was pipetted off and the samples allowed to dry. The samples were then redissolved in 500 µl DEPC treated water. Ten units of the enzyme *DNA polymerase I* were added to each sample and incubated at 37 °C for 120 minutes. Samples were

subjected to a phenol chloroform extraction and ethanol precipitation as described above, and resuspended in 500 µl DEPC treated water. Concentration of samples was ascertained by measuring the absorbence at 260 nm in a Genequant spectrophotometer (Pharmacia). Samples were then stored at –20 °C until required.

2.2.12.3 RT-PCR

First strand complementary-DNA (cDNA) was generated from the RNA samples using the 1st strand cDNA synthesis kit for RT-PCR (AMV) (Boehringer Mannheim) according to manufacturers instructions. Both Oligo-$p(dT)_{15}$ and random $p(dN)_6$ primers were used for each sample, in separate reactions.

One microlitre of each 1st strand cDNA synthesis reaction was used as template for PCR. Where possible, PCR primers were selected to distinguish spliced from unspliced mRNA and possible genomic DNA contamination. PCR amplification was conducted as 2.2.4, with an annealing temperature of 57 °C.

2.3 Computer software

2.3.1 Genetic mapping

Genes were mapped independently on two *B. oleracea* collinear maps (Bohuon *et al.*, 1996) (Sebastian *et al.*, 1999). Segregation data were scored on 169 (A12DH x GD33) and 97 (N x G) lines respectively and checked for double recombinants against existing data sets. Linkage analysis was carried out for both mapping populations using JoinMap 2.0 (Stam and Van Ooijen, 1995) with the parameters set for DH derived progeny. Map distances were calculated using the Kosambi mapping function.

2.3.2 Molecular biology software

DNA sequence alignment, contig construction, PCR primer design and restriction site analysis were all conducted using programmes within the software package DNASTAR (DNASTAR Inc.) according to manufacturers instructions. The default settings were consistently used.

2.4 Nomenclature

There is strong evidence that *BoCAL* and *BoAP1* are present at multiple loci in *B. oleracea*, (Carr and Irish, 1997; Lowman and Purugganan, 1999) for *BoAP1;* (This study, *BoCAL* and *BoAP1*). To prevent possible confusion the precedent set by Lowman and Purugganan (Lowman and Purugganan, 1999), using a letter following the gene name to distinguish between loci has been followed.

CHAPTER 3: THE MOLECULAR GENETICS OF CAULIFLOWER CURD

3.1 Introduction

The ontogeny of the heading brassicas as a whole can be viewed as increasing stringency on control of the switch from vegetative growth to flowering, with cauliflower curds representing the most extreme phenotype of early arrest. An intermediate stage of arrest is characteristic of Sicilian Purple types, where a cauliflower curd form arrests following development of precocious floral bud primordia across its surface (Gray, 1989). The most ontologically advanced of the heading brassicas is characterised by the inflorescence of the Calabrese broccoli. This structure consists of several fused, arrested inflorescence spikes with proliferation of almost fully developed floral buds, together forming a compact wide-domed structure (Carr and Irish, 1997) (fig. 3.4). All of these arrest stages are to some extent temperature sensitive (Fujime, 1983), with eventual bolting and the production of wild-type flowers.

Classical breeding studies have attempted to characterise the genetics underlying the curding phenotype with limited success (Crisp, 1982; Gray and Crisp, 1979; Pease, 1926; Yeager, 1943). Crosses between cauliflower and Calabrese often produce intermediate phenotypes (Branca and Iapichino, 1997; Gray, 1989), suggesting curding is a multifactorial trait. However, it has been suggested that the number of factors controlling gross curd morphology are more likely to be a few key loci rather than an accumulation of many 'less significant' mutations (Crisp, 1982).

The limited understanding of the genetic interactions underlying curding has meant attempts to introgress desirable traits into cauliflowers have often proved difficult. This success has been limited largely by the inability to resynthesise a satisfactory curd of marketable quality (Crisp, 1982). Knowledge of the genetic interactions governing curd development would allow early identification of plants that are likely to be of interest. This would reduce the effort required during introgression programmes by screening out non-informative plants, potentially as early as the

cotyledon stage.

The similarity between the *cal-1/ap1-1* double mutant *Arabidopsis* and the *Brassica* cauliflower curd (fig. 1.3), along with the discovery of a 'mutated' copy of *BoCAL-a* in var. *botrytis* (*BobCAL*), suggested that the curding phenotype of cauliflower may be at least in part due to a mutation at the *BoCAL-a* locus. The *BobCAL* allele therefore presents itself as potentially a very useful marker for the selection of curding habit in cauliflowers.

3.2 Objectives

From this starting point experiments were designed to investigate the genetics of curding in cauliflowers with three specific aims:

- Develop an understanding of genetic interactions responsible for the curding habit of cauliflowers.
- Construct a genetic model of curding and identify loci involved.
- Develop molecular markers to aid rapid resynthesis of curd during introgression programmes involving cauliflowers.

3.3 Strategy

Genetic segregation analysis provides a rapid method of excluding candidate loci as contributing to specific phenotypes. Examination of recombination frequency between a specific allele of a candidate gene and the character in question provides a good indication of whether or not that allele is likely to be contributing to the phenotype. However, ascertaining the interactions between a locus and a phenotype relies completely upon the ease and reliability of phenotypic scoring. To investigate the genetic interactions underlying the curd trait of cauliflower a parallel approach utilising both classical and molecular genetics was employed. A *B. oleracea* population segregating for curding was examined. Having identified phenotypic groupings and ascertained the stability of scorable variation within and between these groups, it was proposed to construct the most probable genetic model that could

account for such phenotypic variation. Tests of linkage of candidate loci to the segregating curd phenotype could then be conducted, and molecular markers developed to trace putative causal alleles.

3.4 Results

3.4.1 Classical genetic analysis

A previously constructed Doubled Haploid (DH) sibling population was employed to examine the segregation of candidate loci against phenotype. The DH population of 61 unselected genetically distinct lines was generated from anther culture of the F1 from a cross between a DH cauliflower, DJ7032 (derived from cv. Nedcha) and a recurrent inbred Calabrese, BI87053 (Ockendon, King and McClenaghan unpublished). This resulted in a segregating population in which, due to the complete homozygosity of DH lines, masking of the phenotypic effects of recessive alleles was removed. The cross was designated NxB.

3.4.1.1 Scoring phenotypic variation

To allow a valid investigation of the genetics underlying this trait it was first necessary to define the specific phenotypic characters that were to be scored. Preliminary examination during field trials in 1992 and 1993 (Ockendon, King and McClenaghan unpublished) and a coldframe experiment in 1995 (King pers. comm.) suggested plants were distinguishable on the basis of gross morphology of the inflorescence. Scoring was conducted in the period following inflorescence development but prior to bolting of the inflorescence spikes. Thus all lines were scored at the 'heading stage'. Close inspection of this stage during 1996 resulted in delineation of three discrete classes of phenotype. Plants were scored as 'parental-cauliflower' if expressing "a complete curd of proliferating apical meristems". Alternatively, plants were scored 'parental-Calabrese' if expressing "a heading inflorescence consisting entirely of fully developed floral buds possessing pedicels". Plants were also observed expressing an inflorescence phenotype similar to that expressed by Sicilian Purples and the curd quality-reducing trait of 'riceyness' (Fujime 1983). This phenotype was defined as "a cauliflower type curd, which is in

part or completely covered by precocious floral bud primordia". Lines expressing such phenotypes appeared intermediate between cauliflower and Calabrese inflorescence morphotypes and were therefore scored as 'intermediate'.

3.4.1.2 Stability of phenotypic groupings

For genetic analysis to be valid, stability of the trait to be scored is of paramount importance. To address this question, phenotypic comparisons were made between individuals of the same DH line grown under different environmental conditions and in different years. In 1996 and 1998 two individuals of each of 61 DH lines were grown in the glasshouse and the coldframe respectively and scored for inflorescence morphology against the criteria outlined above. Examination of these two trials allowed phenotypic stability to be compared across four individuals from each DH line, grown under the crudely differing environmental conditions of glasshouse and cold-frame.

During 1996, both individuals from each of the 61 DH lines always grouped in the same discrete class on the basis of phenotype, suggesting the phenotype was relatively stable across identical individuals. The same observation was made from the 61 DH lines grown during the cold-frame trial of 1998. Comparison of the same lines across years showed trait scores were also highly conserved with just two out of sixty one accessions grown in 1998 not grouping with the 1996 scores. This suggested phenotypic variation spanning the defined groupings *is* possible within individual DH lines, and that the artificial groupings imposed upon the population were invalid. However, in hindsight, molecular analysis of *BoCAL-a* and *BoAP1-a* showed these two accessions were wrongly labelled during sowing of the cold frame trial in 1998. The allelic state of these two loci differed in individuals from the 1996 and 1998 trials, although phenotypic predictions of specific individuals based on the allelic state of these loci still held true.

Tentative comparisons of phenotypes, from photographs of the DH lines taken during the field trials of 1992 and 1993, also showed conservation of phenotype was largely maintained within and across years, with replicates from just a single line falling into two groups during 1992 (NxB 182a, one curd, three intermediate).

Overall, the phenotypic stability of the DH lines across the four trials was strong. Excluding known labelling mistakes, only one line falls within more than one of the phenotypic groupings. The comparison of results from trials conducted across several years under differing environmental conditions suggests inflorescence phenotypes are defined more by genetic components than external variable environmental effects at the specific scoring stage. The identification of discrete groupings of phenotypes that can be reliably scored under different conditions provides a strong base from which to conduct a genetic investigation.

3.4.2 Genetic analysis of phenotypic variation

3.4.2.1 Hypothesis testing: single causal locus

Phenotypic scoring of the DH population revealed 18 parental-Calabrese, 17 parental-cauliflower and 26 intermediate inflorescence phenotypes (Fig 3.4). This segregation of phenotypes was clearly not consistent with the simple 1:1 ratio expected for segregation of a single causal locus in a DH population, where all plants are homozygous at every locus. Because the plants *are* homozygous, it was also possible to discount a theory of co-dominant heterozygosity at a single locus being responsible for the intermediate phenotypes. The most plausible conclusion arising from this evidence is that the genetic determinants responsible for the phenotypic differences between Calabrese and cauliflower inflorescence morphology are located at more than one position in the genome. Statistical analysis of the phenotypic segregation pattern was therefore employed to identify the most likely number of independent loci that could explain the observed ratios.

3.4.2.2 Chi-squared (χ^2) test of segregation ratios

Chi-squared (χ^2) is a statistical method of ascertaining goodness of fit of observed data to specific discrete groupings (Bailey, 1981). This test is ideal for examining theoretical genetic models proposed to explain phenotypic variation.

Examination of the observed segregation ratios in discrete tests against those expected for increasing numbers of interacting loci was used to establish the most likely

number of loci responsible for the differences in inflorescence phenotype. Chi-squared tests were conducted against theoretical models of two and three interacting loci.

3.4.2.3 Hypothesis testing: Two loci

The null hypothesis states: The segregation of phenotypes amongst the NxB population is inconsistent with a model of two unlinked loci.

Table 3.1: Genotype-phenotype interactions under a theoretical two locus hypothesis for curd development in the NxB population.

DH progeny genotype		*Phenotype*
AABB	Parental	Calabrese
aaBB	Recombinant	Intermediate
AAbb	Recombinant	Intermediate
aabb	Parental	Cauliflower

Expected ratio = 1:2:1 Calabrese:Intermediate:Cauliflower. Plants expressing an intermediate phenotype were grouped as no distinction could be made by visual inspection.

Table 3.2: Chi-squared test of observed phenotypic segregation ratios in the NxB population against those expected for a two locus hypothesis.

	O	*E*	*O-E*	$O-E^2$	$O-E^2/E$
Calabrese	18	15.25	3.25	10.56	0.69
Intermediate	26	30.50	4.50	20.25	0.66
Cauliflower	17	15.25	2.25	5.06	0.33

$\chi^2 = \sum O\text{-}E^2/E = 1.69$, with 2 degrees of freedom.

The probability of observing a χ^2 with 2 degrees of freedom greater in value than 0.10 is 95%. The probability of observing a χ^2 with 2 degrees of freedom smaller in value than 5.99 is 95%. The calculated χ^2 value of 1.69 lies within these limits thus suggesting that the null hypothesis is likely to be incorrect. The observed data has a 'good fit' with ratios expected for a two gene hypothesis.

3.4.2.4 Hypothesis testing: Three loci

Although the chi-squared test demonstrates the two gene model is consistent with observations, the action of further loci must be ruled out. The null hypothesis states: The segregation of phenotypes amongst the NxB population is inconsistent with a model of three unlinked loci.

Table 3.3: Genotype-phenotype interactions under a theoretical three locus hypothesis for curd development in the NxB population.

DH progeny genotype		*Phenotype*
AABBCC	Parental	Calabrese
aaBBCC	Recombinant	Intermediate
AAbbCC	Recombinant	Intermediate
aabbCC	Recombinant	Intermediate
AABBcc	Recombinant	Intermediate
aaBBcc	Recombinant	Intermediate
AAbbcc	Recombinant	Intermediate
aabbcc	Parental	Cauliflower

Expected ratio = 1:6:1 Calabrese:Intermediate:Cauliflower. Again plants expressing an intermediate phenotype were grouped as no distinction could be made by visual inspection.

Table 3.4: Chi-squared test of observed phenotypic segregation ratios in the NxB population against those expected for a three locus hypothesis.

	O	*E*	*O-E*	$O\text{-}E^2$	$O\text{-}E^2/E$
Calabrese	18	7.63	10.38	107.64	14.12
Intermediate	26	45.75	-19.75	390.06	8.53
Cauliflower	17	7.63	9.38	87.89	11.53

$\chi^2 = \sum O\text{-}E^2/E = 34.17$, with 2 degrees of freedom.

The probability of observing a χ^2 with 2 degrees of freedom greater in value than 13.82 is 0.1%. The calculated χ^2 value of 34.17 lies outside of this limit thus demonstrating that the null hypothesis is likely to be correct. Observed and expected values are significantly different to the extent that the likelihood of the observed data arising from a three gene model is less than one in a thousand.

3.4.2.5 Summary

From these analyses we are able to conclude that the most likely explanation for the observed result is a model of two, unlinked loci. A similar model of two interacting loci (although linked) was also established following characterisation of the curd phenotype of the *Arabidopsis ap1-1/cal-1* mutant (Bowman *et al.*, 1993). The phenotypic similarity of the cauliflower curd to the inflorescence of the *Arabidopsis ap1-1/cal-1* mutant, together with the identification of a similar genetic model in both species, suggests the *B. oleracea* orthologues of *AP1* and *CAL*, *BoAP1-a* and *BoCAL-a* respectively, are strong candidates for the loci controlling the curding trait of cauliflowers.

3.4.3 Molecular genetic analyses

3.4.3.1 Molecular genetic analysis of *BoCAL-a*

3.4.3.1.1 Marker design

A molecular assay was designed and employed to exploit the presence of the stop codon in the *BobCAL* allele, initially isolated from a single shop-bought, cauliflower (Kempin *et al.*1995). This assay is able to distinguish biallelic differences associated with the base pair mutation in the *BobCAL* allele of the *BoCAL-a* locus.

To analyse the allelic state of the *BoCAL-a* locus a Cleaved Amplified Polymorphic Sequence (CAPS) assay was designed, based primarily around the similar assay described by Kempin *et al.* (1995). The DNA oligonucleotide primers CAL4F and CALI6R (appendix 1) were designed from genomic sequence of *BoCAL-a* (M. Yanofsky, Pers. Comm.). These primers amplify a region of genomic DNA spanning exon 4 to intron 6, approximately 700 bp in cauliflower (figure 3.1).

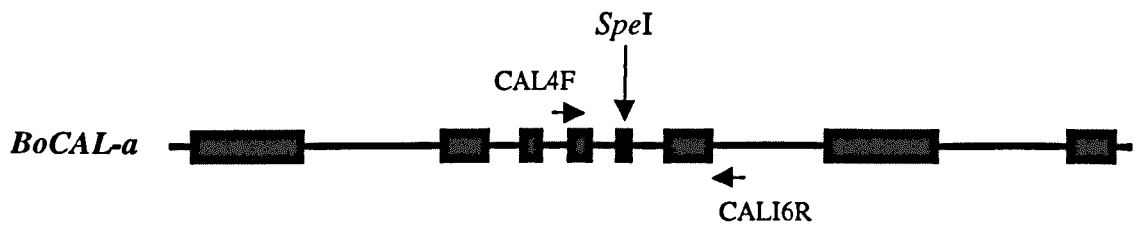


Figure 3.1: Schematic representation of the *BobCAL* CAPS assay designed from *BoCAL-a*. Blocks denote exons, lines denote introns. Amplification between CAL4F and CALI6R generates an amplicon that possesses a *Spe*I restriction site in the cauliflower parent that is absent from the Calabrese broccoli parent.

Following amplification of this region, a restriction digest was conducted exploiting the *Spe*I recognition site, created by the introduction of the base pair mutation into exon 5 of *BobCAL*. Presence of the non-sense mutation in the amplicon resulted in fragments of approximately 500 bp and 200 bp upon digestion. Absence resulted in an uncleaved product. Following optimisation of the reaction all subsequent digestions contained both parents as positive controls of enzymatic activity. Amplicons of both parents were also sequenced to ensure absence of the restriction site was not due to a secondary mutation in the recognition sequence. This confirmed BI87053 did not contain the stop codon.

DNA was extracted from the parents, the F1 and the DH segregant lines, and analysis of the allelic state of *BoCAL-a* was determined for each line in the population.

3.4.3.1.2 Segregation distortion

The Chi-squared test was utilised to examine whether observations of the segregation of alleles of the *BoCAL-a* locus are in agreement with the 1:1 segregation ratio expected under classical Mendelian rules. This was employed to identify whether segregation distortion is present at this locus which can occur through linkage drag during the natural selection pressure of anther culture.

The null hypothesis states: Segregation of *BoCAL-a* alleles in the NxB population is inconsistent with classical Mendelian ratios.

Table 3.5: Chi-squared test of observed segregation ratios of *BoCAL-a* alleles amongst the NxB population against those expected under classical Mendelian segregation.

Observed frequency (O)	*Expected frequency (E)*	*O-E*	*$O-E^2$*	*$O-E^2/E$*
35	30.50	4.50	20.25	0.66
26	30.50	4.50	20.25	0.66

$\chi^2 = \sum O\text{-}E^2/E = 1.33$, with 1 degree of freedom.

The probability of observing a χ^2 with 1 degree of freedom greater in value than 0.004 is 95%. The probability of observing a χ^2 with 1 degree of freedom smaller in value than 3.84 is 95%. The calculated χ^2 value of 1.33 lies within these limits thus suggesting that the null hypothesis is incorrect. The observed and expected ratios are not significantly independent, demonstrating that the observed data has a 'good fit' with ratios expected for classical Mendelian segregation. This demonstrates that segregation distortion has not significantly altered the segregation ratio of alleles at

this locus.

3.4.3.1.3 Comparison of *BoCAL-a* segregation against phenotype

Table 3.6: Segregation data of *BoCAL-a* against phenotype in the NxB population.

	Calabrese	*Intermediate*	*Curd*	*total*
BoCAL-a SpeI+	0	18	17	35
BoCAL-a SpeI-	18	8	0	26
Total	18	26	17	61

Analysis of the segregation pattern of the *BoCAL-a* alleles in the DH progeny strongly suggests this locus is linked to phenotype. The *BoCAL-a* allele containing the *Spe*I recognition site and consequently the premature termination codon, is always present in offspring expressing the parental curd phenotype. This suggests that the presence of this allele, or the haplotype it represents, is required for development of full curd.

Segregation of the presumptive causal allele of *BoCAL-a* amongst the intermediate and curding lines of the NxB population suggests that although the presence of this allele is required, it is not diagnostic of curd development. This provides molecular genetic confirmation that the development of full curd is dependent on segregation and allelic state of more than one locus.

3.4.3.2 Molecular genetic analysis of *BoAP1-a*

In light of the genetic interaction underlying the phenotypically similar *ap1-1/cal-1* *Arabidopsis* mutant, evidence pointing to the action of a second gene in *B. oleracea* immediately suggests the *B. oleracea AP1* orthologue, *BoAP1-a* (Anthony *et al.*, 1995) would be a good candidate for investigation.

3.4.3.2.1 Marker design

A polymorphism in *BoAP1-a* was identified by examining the locus for possible amplicon length variation between the parents of the NxB population. PCR primers

were constructed at regular intervals along the locus which allowed rapid examination of the region, whilst also being close enough together to resolve possible regions of interest upon agarose gel electrophoresis of the product (Figure 3.2) (appendix 1).

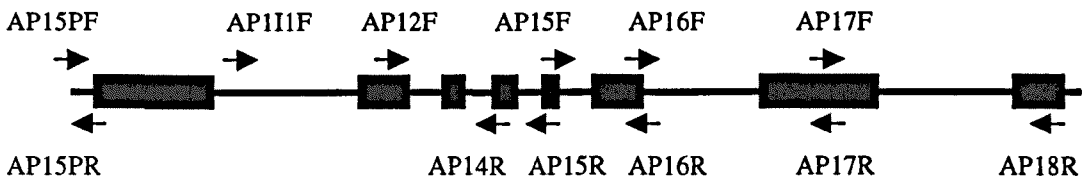


Figure 3.2: Schematic showing the relative distribution of PCR primers along the *BoAP1-a* locus

Amplification between AP15PF and AP15PR, situated 100 bp five-prime of the presumptive start codon revealed a length polymorphism on electrophoresis. The amplicon length was approximately 120 bp with allele size differing between the parents of the NxB population by approximately 10 bp.

This polymorphism was scored on the F1 of the population and each DH progeny line.

3.4.3.2.2 Segregation distortion

The Chi-squared test was once again employed to examine whether the segregation ratio of *BoAP1-a* alleles is in agreement with the 1:1 ratio expected under classical Mendelian rules. Calculations were conducted as given in 3.4.2.3.

$\chi^2 = \Sigma\ O\text{-}E^2/E = 1.98$, with 1 degree of freedom. Thus there is no significant difference between the observed and expected values, given that the 5% point is at 3.84. *BoAP1-a* is segregating within expectations based on classical Mendelian inheritance.

3.4.3.2.3 Comparison of *BoAP1-a* segregation against phenotype

Table 3.7: Segregation data of *BoAP1-a* against phenotype in the NxB population

	Calabrese	*Intermediate*	*Curd*	*total*
DJ7032 *BoAP1-a*	0	8	17	25
BI88908 *BoAP1-a*	18	18	0	36
Total	18	26	17	61

Analysis of the segregation pattern of the *BoAP1-a* alleles in the DH progeny strongly suggests that the allelic state at this locus, or the haplotype it represents, is also linked to the curding phenotype. DH progeny expressing full curds always possess the *BoAP1-a* allele from the cauliflower parent. Again presence of this allele in lines expressing intermediate phenotypes demonstrates that variation at this locus is not diagnostic of curd development.

3.4.4 Is curding a recessive trait?

Although extremely useful in simplifying scoring of the genotype/phenotype interaction of specific loci, the sole use of a DH population prevents any conclusions being drawn regarding dominance status of specific alleles. To address this question phenotypic and molecular genetic examination of the F1 from the NxB cross was undertaken. The F1 was confirmed to be heterozygous for both *BoCAL-a* and *BoAP1-a* utilising the assays discussed above. Phenotypic examination by independent scorers of plants grown under glass, in a cold frame, and in the field confirmed that the F1 expresses a complete 'Parental-Calabrese' inflorescence type (fig. 3.3). There is no evidence in the F1 of the intermediate curd qualities found in the NxB intermediates, which would support the hypothesis of dominant alleles. The results of this analysis strongly suggest that, alleles at both contributing loci arising from the cauliflower parent are recessive, and thus that the curd of cauliflower is likely to be a recessive trait.

Figure 3.3: Inflorescence morphology of the NxB F1 resembles a 'Parental-Calabrese' phenotype.

3.5 Discussion

3.5.1 A Genetic model of curding in cauliflower

Molecular genetic analysis of the NxB population allows the construction of a relatively simple genetic model that is able to explain fully the observed phenotypic differences between cauliflower and Calabrese inflorescence morphology. The results strongly suggest that these differences can be explained by the allelic state at two unlinked loci. The data also demonstrate that the orthologues of the genes responsible for the similar cauliflower phenotype in the related crucifer *Arabidopsis thaliana* are present at these two loci. Segregation of these loci amongst the DH NxB population results in three distinct inflorescence phenotypes, namely Calabrese, intermediate and cauliflower. Although indistinguishable on the basis of phenotype, DH progeny expressing an intermediate phenotype can be divided into two groups on the basis of presence of alleles at one locus, either *BoCAL-a* or *BoAP1-a*, arising from the cauliflower parent. Evidence from this study suggests these lines can be distinguished entirely on that basis. The distinction is therefore quite clear and can be summarised thus: i) Presence of no alleles from the cauliflower parent at either locus results in expression of Calabrese inflorescence type. ii) Presence in a homozygous state of

alleles of *either BoCAL-a* or *BoAP1-a* from the cauliflower parent results in an earlier stage of arrest, characterised by the expression of the intermediate phenotype. iii) Presence in a homozygous state of alleles from *both BoCAL-a and BoAP1-a* from the cauliflower parent results in arrest at an earlier stage and development of full curd (figure 3.4).

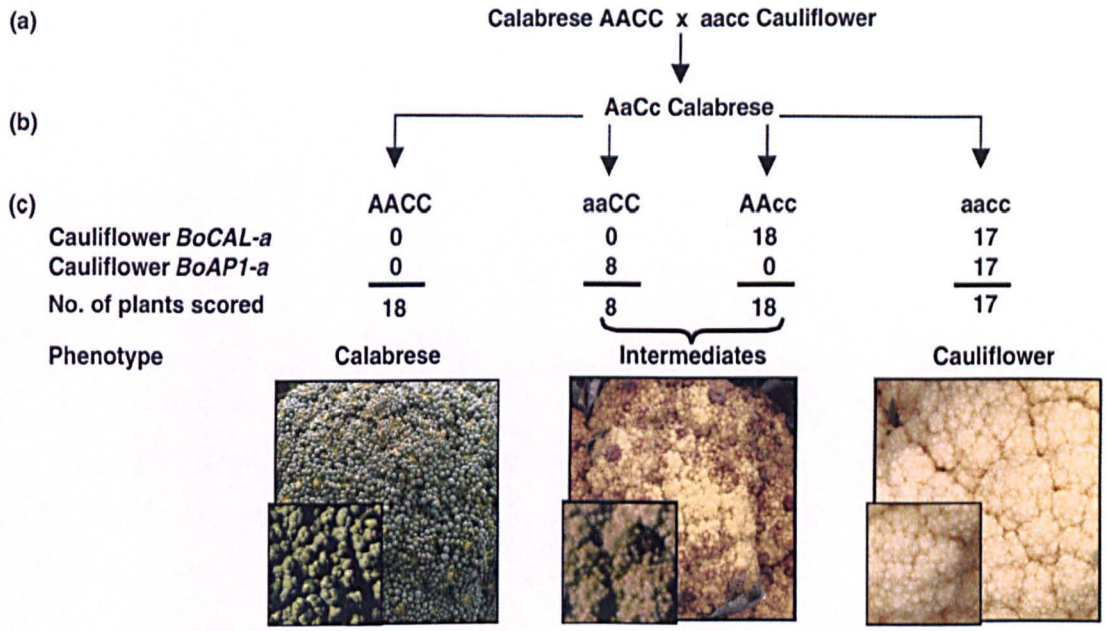


Figure 3.4: A genetic model of the interactions between the *BoCAL-a* and *BoAP1-a* loci required to explain the curding phenotype of *Brassica oleracea* var. *botrytis*. AA = *BoAP1-a*, CC = *BoCAL-a*. Lower case letters denote presumptive mutant alleles. Calabrese is taken as the 'wild-type' phenotype as cauliflower arrests at an earlier developmental stage. (a) DH Parental genotypes and phenotypes. (b) F1 of the cross, full Calabrese phenotype suggests mutant alleles at both loci are recessive. (c) DH offspring generated from the F1, the two loci segregate such that double recessive plants exhibit full curding whilst double dominant plants display the Calabrese phenotype, recombinants between the parental genotypes results in intermediate type heads (insets show magnified view).

3.5.2 Utility for plant breeders

The difficulties encountered during previous attempts to characterise the genetics underlying this trait through conventional breeding techniques (Crisp, 1982; Crisp *et al.*, 1975; Crisp and Tapsell, 1993) could be explained by the synergistic interaction of

these two loci and the recessive nature of the causal alleles. These genetic interactions are compounded by the effect on phenotype of factors such as temperature fluctuation. Together the genotype/environment interactions have probably inhibited previous identification of a consensus model of curd synthesis.

The molecular assays tested during this study may be of significant interest to cauliflower breeders as they represent powerful selectable markers for curding habit in *B. oleracea.* They would have utility both in characterising *ex situ* genetic resources and in reducing the costs of introgression trials. By employing just the molecular markers we have developed for *BoCAL-a,* seventy five percent of the seedlings grown could be removed as 'non-curding', predictions that can be made as early as the cotyledon stage. Development of a more robust assay for *BoAP1-a* segregation could increase this specificity to almost ninety four percent.

3.5.3 Conclusions

1. Two unlinked genetic loci are responsible for differences in inflorescence morphology between Calabrese and cauliflower.
2. Orthologues of the genes responsible for the similar curd trait of *Arabidopsis ap1-1/cal-1* mutant are present at these loci.
3. The two loci characterised by *BoCAL-a* and *BoAP1-a* segregation, suggest that alleles arising from the cauliflower parent are recessive, and thus that cauliflower curd is a recessive trait.
4. The two loci characterised by *BoCAL-a* and *BoAP1-a* appear to have additive effect on control of floral induction.

CHAPTER 4: MOLECULAR GENETIC CHARACTERISATION OF THREE INDEPENDENT *BoAP1* LOCI IDENTIFIED IN *BRASSICA OLERACEA*

4.1 Introduction

Comparative mapping between *Brassica* and the related crucifer *A. thaliana* has suggested that diploid *Brassica* genomes are of paleopolyploid origin, derived from a hexaploid ancestor (Lagercrantz and Lydiate, 1996). Regions of genetic marker colinearity between *Brassica* and *A. thaliana* have been reported following several studies (Cavell *et al.*, 1998; Grant *et al.*, 1998; Lagercrantz *et al.*, 1996; Scheffler *et al.*, 1997). This suggests that diploid *Brassica* genomes may have evolved from the combination of more than one *A. thaliana*-like genome (Lagercrantz, 1998). Alternatively, *A. thaliana* could have undergone genomic reduction since diverging from the *A. thaliana/Brassica* common ancestor.

The suggestion of a hexaploid origin for diploid brassicas has been further substantiated by recent evidence derived from genetic mapping projects in *B. oleracea.* These analyses suggest that many loci that are single copy in *A. thaliana* are multicopy in *Brassica* (Bohuon *et al.*, 1996; Lagercrantz, 1998; Sebastian *et al.*, 1999; Slocum *et al.*, 1990). The implications of this for the genetic control of important agronomic traits in *Brassica* are intriguing, as it raises the possibility of redundancy or functional divergence between gene copies. An example of such redundancy has already been reported (Lagercrantz *et al.*, 1996). Two copies of the flowering time gene *CONSTANS* (*CO*) have been identified in the diploid *Brassica nigra.* These duplicate copies appear to exert differing levels of control over flowering time, based on evidence from QTL analysis of a polymorphic population (Lagercrantz *et al.*, 1996).

RFLP banding patterns suggest that the *Brassica* orthologues of the *A. thaliana* genes *AP1*, *AG* and *AP3* are all replicated in *B. napus* (Isobel Parkin Pers. Comm.). Replication of such loci, and in particular *AP1,* may be of interest regarding the control of floral induction in the *B. oleracea* cauliflower.

Initial examination of a cauliflower curd cDNA library identified what appeared to be a single *BoAP1* mRNA species (Anthony *et al.*, 1995). The first indication that there may be multiple copies of *BoAP1* within *B. oleracea* was made following the cloning of two distinct mRNA species (*Boi1AP1* and *Boi2AP1*) from broccoli (Carr and Irish, 1997). However, no effort had been made to rule out the possibility of heterozygosity at a single locus which could also account for the observations (V. Irish, Pers. Comm.). A duplication of the *BoAP1* locus in the *Brassica* cauliflower was recently confirmed and the novel copy designated *BoAP1-b* (Lowman and Purugganan, 1999). This copy contains a premature termination codon in exon 4 in single accessions of five distinct *B. oleracea* crops, and was not found to be expressed in *Brassica* cauliflower curd. DNA sequence analysis suggests that *BoAP1-b* is distinct from either mRNA species identified by Carr and Irish (1997) (Lowman and Purugganan, 1999).

Lowman and Purugganan (1999) proposed a reduction in the level of *BoAP1* redundant function resulting from the mutation in *BoAP1-b*, in conjunction with the LOF of *BoCAL-a* as a model to account for the curd phenotype of the *Brassica* cauliflower (table 4.1).

Table 4.1: Summary of genetic model for curd development proposed by Lowman and Purugganan (1999). The reduction in *BoAP1* redundant function in combination with LOF of *BoCAL* was proposed to account for the development of the *Brassica* cauliflower curd.

	BoCAL function	*BoAP1-a function*	*BoAP1-b function*
Wild-type	+	+	-
Cauliflower	-	+	-

Perhaps the most significant conclusion that could be drawn from the identification of multiple loci of *BoAP1* was that previous *in situ* and northern blot analysis would have been unable to distinguish between expressed copies (Anthony *et al.*, 1996; Carr and Irish, 1997; M Yanofsky, Pers. Comm.). There therefore exists the possibility that previously published expression data may include expression profiles from more than one locus.

4.2 Objectives

Evidence of multiple loci of *BoAP1* raised several important issues relating both to this study and to the investigation of *Brassica* genes in general. These form the basis of the objectives:

1. Determine the likely number of *BoAP1* loci in *B. oleracea.*
2. Examine the possibility that the previous approach of PCR screening (Chapter 3) has resulted in the production of chimeric gene sequences, constructed from overlapping PCR amplicons derived from *B. oleracea* genomic DNA.
3. Identify whether *Boi1AP1* and *Boi2AP1* (Carr and Irish, 1997) represent loci distinct from *BoAP1* (Anthony *et al.*, 1995). If so, does the existence of these expressed sequence in broccoli have functional significance in terms of cauliflower curd development?
4. Assign a linkage position to each locus and re-examine the cohesiveness of the genetic model proposed for curding (chapter 3).
5. Examine the role of *BoAP1* during floral induction in the *Brassica* cauliflower. In particular, whether it is likely to involve multiple *BoAP1* loci.

4.3 Strategy

RFLP analysis with cross-hybridising probes from *A. thaliana* was employed to give an indication of likely copy number. These probes were then utilised during screening of a *B. oleracea* BAC library to identify further copies of candidate loci. Locus specific assays were designed to allow DNA sequencing, genetic mapping and analysis of regions collinear with *A. thaliana,* to ascertain the extent of genome conservation. RT-PCR was used to examine the possibility of functional redundancy or divergence of multiple loci or gene silencing during floral induction in the *Brassica* cauliflower.

4.4 Results

4.4.1 SECTION 1: ASCERTAINING GENE COPY NUMBER OF *BoAP1*

4.4.1.1 Probe construction

Specific sections of cDNA clones of the *A. thaliana* genes *CAL* and *AP1* were amplified by PCR to exclude the highly conserved MADS-Box region from cross-hybridisation experiments. A third probe was generated from *B. oleracea* genomic DNA (A12DH) which represented the *BoAP1-a* MADS-Box and a small region of intron 1 (table 4.2).

Table 4.2: Probes for cross-hybridisation experiments between *B. oleracea* and *A. thaliana*. Primer sequences are given in appendix 1.

Probe name	*Primer 1*	*Primer 2*	*Probe size (bp)*	*Generated from:*
CAL	CALcDNAF	M13R	800	*CAL*-cDNA
AP1	AP1cDNAF	M13F	1050	*AP1*-cDNA
MADS-Box	AP1MADSF	MADINTR	432	*BoAP1*-genomic PCR amplicon

4.4.1.2 RFLP analysis

*Eco*RI genomic digests of the NxB (Smith and King unpublished) and NxG (Sebastian *et al.*, 1999) populations, alongside *A. thaliana* genomic DNA, were Southern blotted and probed with the AP1 and CAL amplicons to provide segregation data based on RFLP marker loci.

These probes hybridised distinctly to single bands of different size on *A. thaliana* genomic DNA, demonstrating that there was no cross-hybridisation between the CAL and AP1 probes. This was confirmed in a control experiment where the CAL and AP1 probes were hybridised to the IGF Arabidopsis BAC library (Mozo *et al.*, 1998). This demonstrated that both probes identified BACs at the correct genomic positions for *CAL* and *AP1* respectively, and that no other loci were hybridised. This suggested that

there was a high probability that any multiple hybridising bands in *Brassica* were likely to represent closely homologous sequences to *CAL* and *AP1*.

Although the *B. oleracea* genes *BoAP1* and *BoCAL* are more closely related to their respective *A. thaliana* orthologues than to each other (based on exonic sequence), cross-hybridisation *was* observed between CAL and AP1 cDNA probes when hybridised to *Brassica* genomic DNA, although the intensity of specific loci varied between probes.

Analysis of restriction fragment lengths suggested there are at least four *BoCAL* and *BoAP1* like loci in *B. oleracea*. Two of these loci were polymorphic in the NxG population and were designated AP1/CAL-like1 and AP1/CAL-like2. The detection of polymorphism allowed linkage mapping in the context of the NxG genetic linkage map (4.4.2.2) (Sebastian *et al.*, 1999).

4.4.1.3 Bacterial Artificial Chromosome (BAC) library analysis

In light of the cross-hybridisation problems, a *B. oleracea* genomic BAC library derived from A12DH was employed in an attempt to isolate distinct loci of *BoCAL* and *BoAP1* for molecular characterisation.

Probing of the *B. oleracea* BAC library with the *A. thaliana* cDNA amplicons from *CAL* and *AP1*, repeatedly revealed approximately 400 positively hybridising colonies, even at high stringency (0.5x SSC, 0.1% SDS at 65 °C).

The estimated genome coverage of the BAC library was five haploid genomes (C. Ryder Pers. Comm.). This suggested either that the probes contained a conserved widespread motif, or else *AP1* and *CAL* like sequences are extremely common in *B. oleracea*. Alternatively, the technique used to generate the BAC library (*Hin*dIII restriction enzyme partial digest rather than sonication) may have resulted in bias towards specific genomic regions, resulting in over representation of certain loci. Secondary screening of 190 positively hybridising colonies (selected as the 30 most strongly hybridising, together with all hybridsations from half of the filter sets, A-E) revealed two colonies (BoB28L01 and BoB48N13) hybridising for *AP1* at a

significantly stronger level than other preliminarily selected BACs. The resolution of these colonies above background was aided by growing to a diameter of 7 mm (as opposed to the initial 2 mm colony diameter on the library membranes). Similar screening with the *CAL* probe failed to identify any *BoCAL* positive colonies, in spite of further screening with an exonic PCR primer set (CALcDNAF to LSCALR1). This suggested that *BoCAL* loci may be under-represented in the library, as PCR amplification with these primers *is* successful from A12DH genomic DNA.

To address this possibility, further screenings were conducted with a labelled *BoCAL* amplicon (LSCALF1 to Bob2; exon 7 to exon 8, 450 bp). This screening also failed to identify any *BoCAL* positive colonies, suggesting that the most likely explanation is under representation of *BoCAL* loci in the BAC library.

Further hybridisation with the generic MADS-Box probe identified many of the colonies previously identified in the *CAL* cDNA and *AP1* cDNA probings. This suggested that the most likely explanation for the high number of false positives was cross-hybridisation to other, over represented members of the MADS-Box gene family, of which at least 50 members have been identified in the related *A. thaliana* (M. Yanofsky, pers. Comm.).

4.4.1.4 Molecular Analysis of *BoAP1* positive BACs

In light of the cross-hybridisation problems encountered, and the limited time available, it was decided to concentrate efforts on analysis of the positively identified *AP1* hybridising BACs.

PCR amplification with the exonic *BoAP1* primers *AP1*cDNAF and GK96 (appendix 1) confirmed the presence of a copy of *BoAP1* on both of the BACs (BoB28L01 and BoB48N13). Probing of these BACs with the generic MADS-Box probe following restriction digest with a range of enzymes implied the presence of just a single MADS-Box motif on each of the BACs (fig. 4.1).

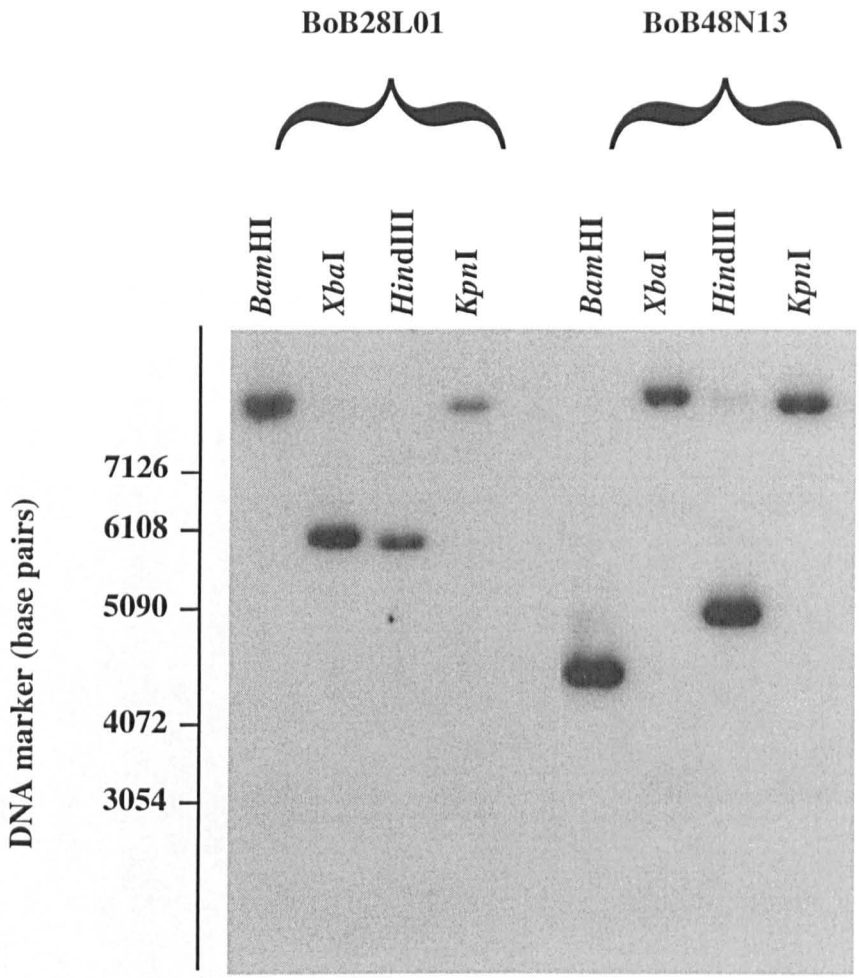


Figure 4.1: Autoradiograph showing hybridisation of *BoAP1*-MADS-Box probes to BAC DNA digested with a range of enzymes. Banding patterns suggest that a single copy of *BoAP1* is present on each BAC. Fragment size polymorphisms suggest that these copies are distinct.

This strongly suggested that the hybridising *BoAP1* loci were the only MADS-Box genes on their respective BACs, and that the use of *BoAP1* specific primers would be unlikely to generate a chimeric sequence during DNA sequencing from the BACs. Whilst it was recognised that in theory further *BoAP1* loci may be present and simply lack the probe target (the MADS-Box within exon 1), DNA sequencing from a second distinct locus would be likely to generate heterozygote peaks on fluorescent sequence trace files. No such heterozygote peaks were observed, thus supporting the single

distinct locus hypothesis.

Further analysis of the MADS-Box probing showed that not only was there just a single band, but the size of the hybridising band was polymorphic between the two *BoAP1* positive BACs (fig. 4.1). This suggested that the two copies of *BoAP1* identified were distinct.

4.4.1.5 Two distinct *BoAP1* loci

PCR amplifications conducted using intronic primers (appendix 1) designed from an unpublished *BoAP1* genomic sequence (M. Yanofsky, Pers. Comm) demonstrated that amplification was from only one of the two loci. The combination of hybridisation analysis and specific PCR amplification with conserved exonic but not intronic primer sets confirmed that the second copy of *BoAP1* (BAC BoB28L01) was distinct from *BoAP1*-a (BAC BoB48N13). In light of the nomenclature proposed by Lowman and Purugganan (1999) this novel locus was designated *BoAP1-c.*

4.4.1.6 DNA sequencing

The genomic sequences of *BoAP1-a* and *BoAP1-c* were characterised from the BACs following PCR amplification from conserved primers (fig. 4.2)

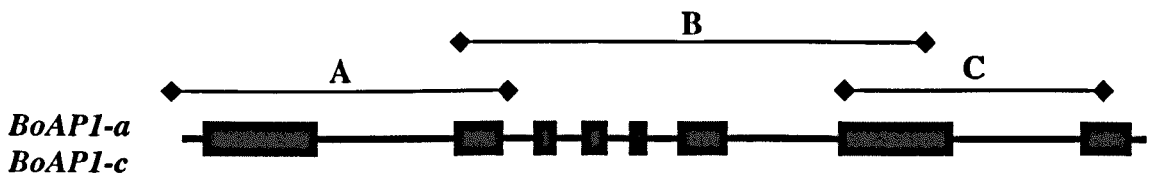


Figure 4.2: Schematic of overlapping amplicons derived from *BoAP1-a* and *BoAP1-c* genomic BAC clones. (A) AP15PF-AP1cDNAR. (B) AP1cDNAF-GK96. (C) AP17F-AP1B8R. Primer sequences are given in appendix 1.

DNA sequencing from exonic primers was then employed to obtain the genomic sequence of both of the copies of *BoAP1* from A12DH present in the different BACs (Fig. 4.3).

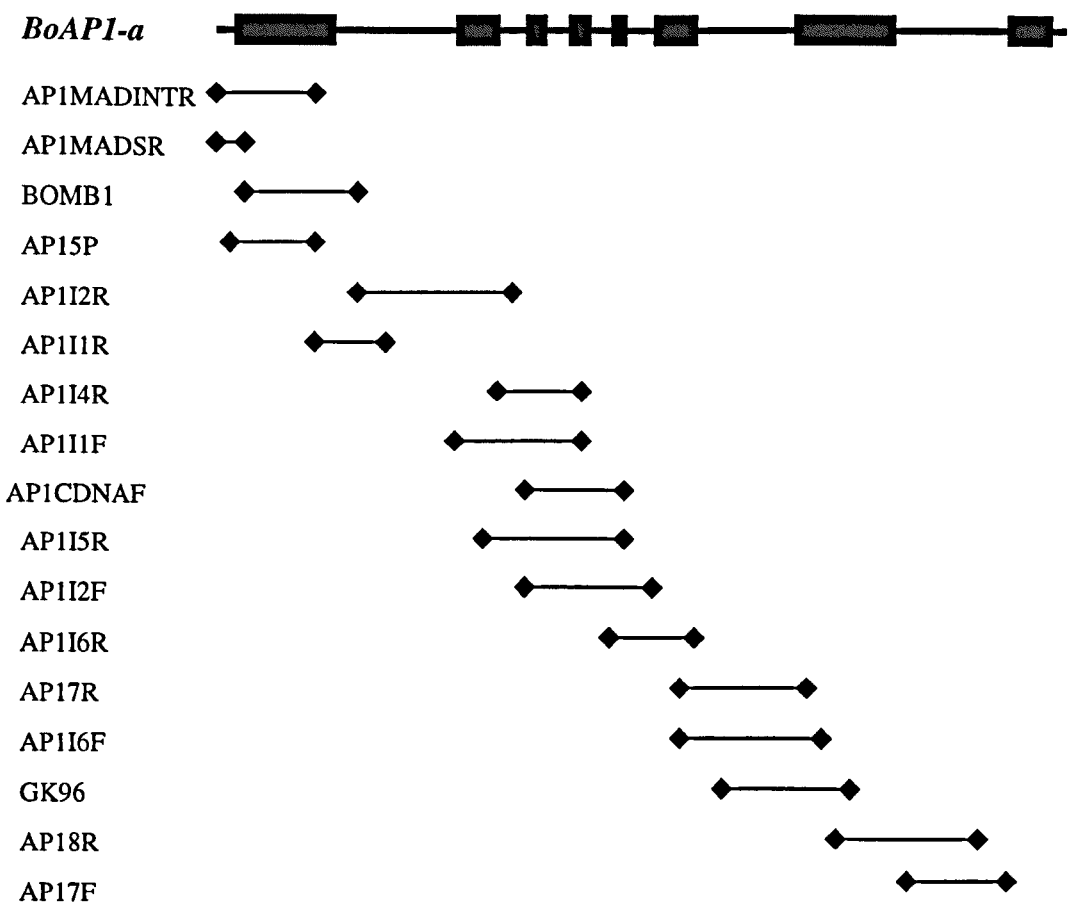


Figure 4.3: DNA sequence contig construction of *BoAP1-a* derived from BAC BoB48N13 (A12DH). Sequence primer names are given in the left column, genomic position of resultant sequence is shown against a schematic.

4.4.1.7 Locus specific markers for *BoAP1-a* and *BoAP1-c*

Locus specific PCR primers were designed following DNA sequencing of amplicons between the conserved primers. These novel primers then acted as locus specific anchors during PCR amplifications from stock genomic DNA of each of four distinct accessions (DJ7032, GD33, BI87053 and DJ3753).

As a control, all locus-specific primer sets were tested in PCR amplifications from both BACs. This confirmed results from the *in silico* experiments which suggested that these primers would be locus specific. These analyses were done in the

knowledge that this approach could not guarantee locus specificity verses as yet undiscovered loci in amplifications from genomic DNA. However, it was decided to proceed with this approach as it was found to be the most pragmatic, and would allow rapid re-analysis of locus specificity upon discovery of further loci. In addition, it was decided to limit sequence analysis to DH plants thus providing the theoretical safeguard of heterozygote peak analysis. Due to the nature of DH plants any such peaks should represent distinct loci.

4.4.1.8 A third copy of *BoAP1*

Relatively recently a novel copy of *BoAP1* has been identified by Lowman and Purugganan (1999). As this was the second *BoAP1* locus reported in the literature it was designated *BoAP1-b*. Comparative sequence alignment of exon 4 with that of *BoAP1-a* and *BoAP1-c* supports suggestions that this is a novel locus, as *BoAP1-b* contains a distinct insertion which results in an in-frame termination codon.

```
A12DH-BoAP1-a  GCA ATG AGC --- --- --- CCT AAG GAA
A12DH-BoAP1-b  GCA ATG AGC TAA ATA GGC TCT AAG GAG
A12DH-BoAP1-c  GCA ATG AGC --- --- --- CCT AAA GAG
```

Figure 4.4: Comparative sequence alignment of a section of exon 4 from *BoAP1-a, boAP1-b* and *BoAP1-c*. The sequence from *BoAP1-b* contains an insertion which results in the generation of an in-frame termination codon. This suggests *BoAP1-b* is a novel locus when compared to *BoAP1-a* and *BoAP1-c*. Note: the sequences from all other accessions scored also contained this mutation.

PCR primers were designed spanning this novel region (AP1INDEL4F and AP1INDEL4R), to allow anchored locus-specific PCR both 5-prime and 3-prime from exon 4 (Fig. 4.5). The absence of this sequence from either *BoAP1-a* or *BoAP1-c* ensured these loci were not amplified.

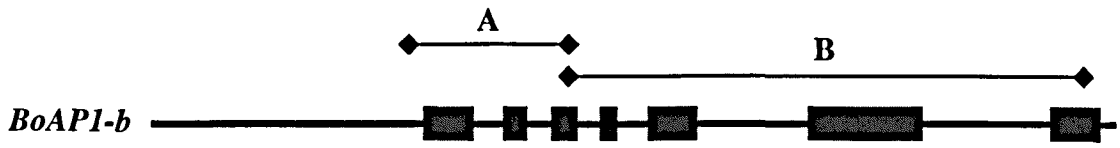


Figure 4.5: Schematic of anchored locus specific PCR amplicons against *BoAP1-b* genomic sequence. (A) AP1cDNAF-AP1INDEL4R. (B) AP1INDELF-AP1B8R. No amplification 5-prime of exon 2 was forthcoming, despite attempts with four distinct MADS-Box specific forward primers, which based on both *in silico* analysis and empirical experiments should amplify from both *BoCAL* and *BoAP1* genomic template. This suggested that either exon 1 is not present, has significantly diverged from other *BoAP1* loci, or is too far away for successful PCR amplification.

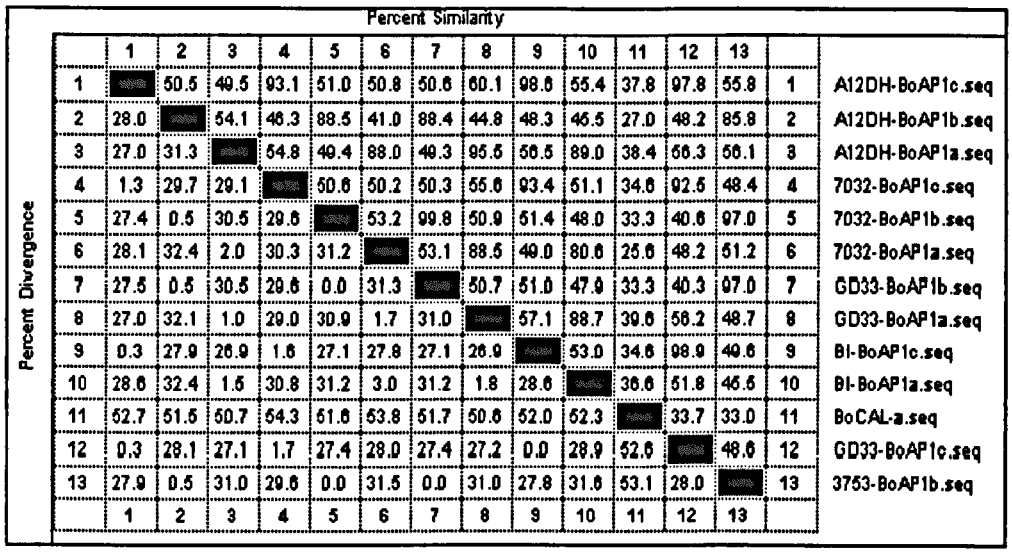
The late identification of this locus during the period of this research led to a decision to characterise this locus solely through anchored-PCR means, rather than returning to the BAC library. The problems encountered with previous BAC library hybridisations, and in particular the failure to identify any *BoCAL* or *BoAP1-b* positive BACs during the initial screen, suggested that a large amount of effort in the limited time available would still not guarantee success.

In silico comparisons of DNA sequences derived from all three *BoAP1* loci confirmed that previously designed intronic primer sets would not have amplified from *BoAP1*-b, and as a consequence, it was surmised that no chimeric sequences had been constructed from these three loci.

4.4.1.9 Analysis of DNA sequence

DNA sequence was generated for each locus in each of five distinct lines, A12DH (Chinese White Kale, i.e. BAC derived sequences), GD33 (Calabrese), DJ7032 (N) (Cauliflower), DJ3753 (G) (Brussels Sprout) and BI87053 (B) (Calabrese). Time limitations prevented the complete genomic sequence being characterised for each accession. However, the volume of sequencing was sufficient to allow examination of locus distinction (fig. 4.6), the assignment of loci to previous published cDNA sequences (fig. 4.7), and identification of mutations that could possibly contribute to the development of the curd phenotype of cauliflower (4.4.1.11).

(a)



(b)

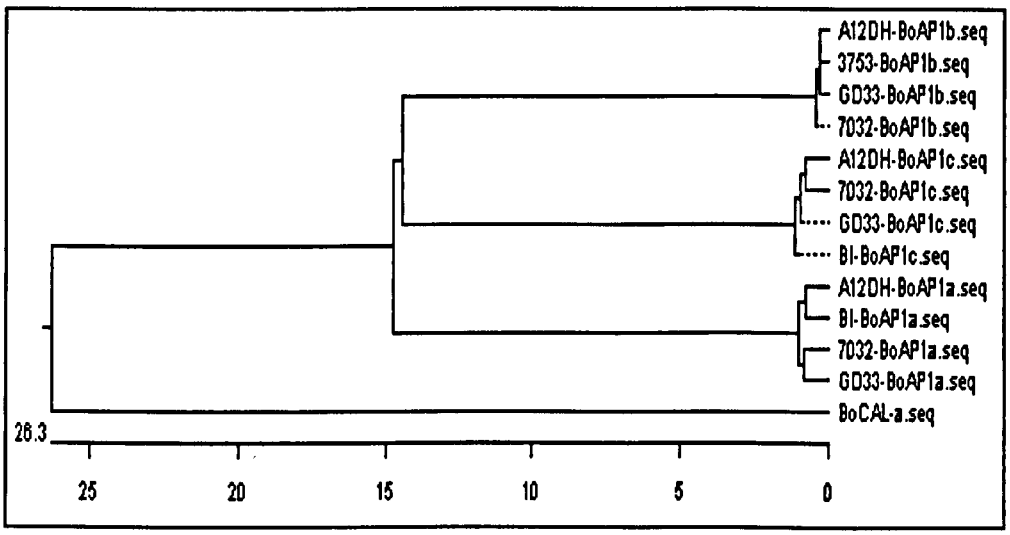


Figure 4.6: (a) Comparative alignment matrix of DNA sequences from exon 6 and intron 6 showing the percent DNA sequence identity and divergence between sequences from each of the three *BoAP1* loci and *BoCAL-a*. (b) A phenogram constructed from this analysis and shown here purely as a graphical representation of the alignment (See 4.4.1.9 for key). These analyses demonstrate that the three *BoAP1* loci could be grouped into three distinct clades based on sequence similarity across different crop types. *BoCAL-a* was added to the alignment to provide an outgroup. Exon and intron six were chosen as the most widely sequenced region of the three loci across the crop types. Note: DNA sequence comparison utilised the MEGALIGN sequence alignment program within the DNASTAR software package and are given in appendix 2. Alignments were made using the clustal method, with all parameters set to default.

DNA sequence alignment strongly supports evidence suggesting that the three identified loci of *BoAP1* are independent. The availability of sequence data also raised the possibility of assigning previously published cDNA sequences to specific loci.

4.4.1.10 Assignment of loci

DNA sequencing of exons one to seven (minus exon 4) of *BoAP1*-a and *BoAP1*-c allowed comparative alignment with previously characterised *BoAP1* sequences (fig. 4.7). The alignment strongly suggested that *BoAP1*-a is the same locus as *BobAP1* (M. Yanofsky, Pers. Comm.) and *Boi2AP1* (Carr and Irish, 1997). This analysis further suggested that *BoAP1*-c corresponded to the mRNA species identified as *Boi1AP1* from broccoli (Carr and Irish, 1997). As a result, the second mRNA species discovered by Carr and Irish (1997) was confirmed as representing a distinct genetic locus of *BoAP1*, and thus that *BoAP1*-c is expressed in broccoli.

(a)

Percent Similarity

Percent Divergence

	1	2	3	4	5	6	7	8		
1		93.6	96.0	96.0	96.2	96.0	96.0	96.0	1	7032-BoAP1a (ant)
2	4.2		97.4	97.4	97.6	91.1	91.1	91.1	2	7032-BoAP1a
3	3.9	0.5		99.7	99.8	93.1	93.1	93.1	3	Bobap1a(yan)
4	3.9	0.5	0.3		99.8	92.9	92.9	92.9	4	A12DH-BoAP1a
5	3.7	0.4	0.2	0.2		93.2	93.2	93.2	5	Boi2AP-1
6	4.1	6.9	7.1	7.1	6.9		100.0	100.0	6	A12DH-BoAP1c
7	4.1	6.9	7.1	7.1	6.9	0.0		100.0	7	7032-BoAP1c.
8	4.1	6.9	7.1	7.1	6.9	0.0	0.0		8	Boi1AP-1
	1	2	3	4	5	6	7	8		

(b)

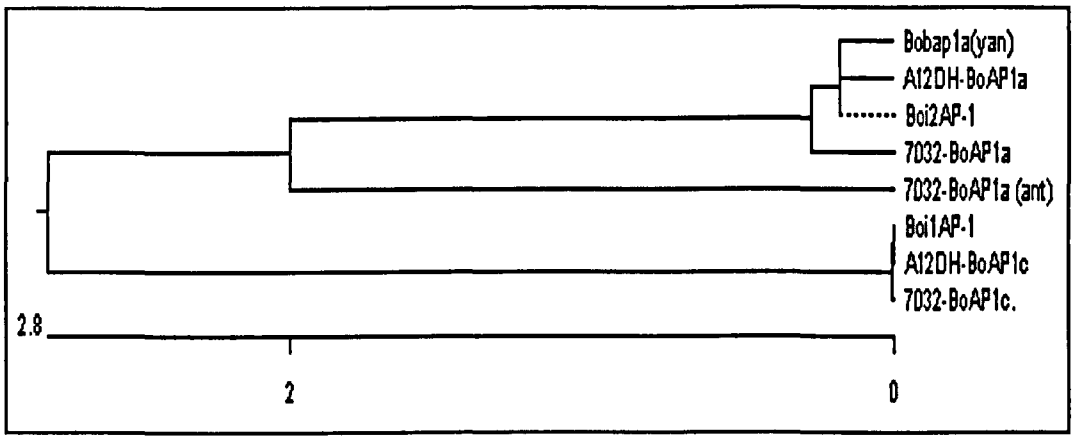


Figure 4.7: (a) Comparative alignment matrix of DNA sequences from exon 1 to exon 7 (minus exon 4) showing the percent DNA sequence identity and divergence between sequences from each of the published *BoAP1* cDNA sequences and putative exonic sequences from *BoAP1-a* and *BoAP1-c*. (b) A phenogram constructed from this analysis and shown here purely as a graphical representation of the alignment. These analyses suggest that the published *BoAP1* cDNAs could be grouped into two distinct clades based on the distinct *BoAP1-a* and *BoAP1-c* loci, suggesting that these cDNAs can now be assigned to a genomic locus, and thus that these loci can retrospectively be assigned expression information. The *BoAP1* cDNA sequence (Anthony *et al.*, 1995) appears to be distinct from each of the major clades. This is surprising given that *BoAP1-a* and *BoAP1-c* sequences from the same line used by Anthony *et al.*, (1995) (DJ7032) are included in the analysis. Further examination suggests that *BoAP1* represents a chimeric sequence constructed from overlapping cDNA clones from distinct loci. KEY: Bobap1a(yan) = M. Yanofsky, unpublished sequence derived from a cauliflower; A12DH-BoAP1a and

A12DH-BoAP1c = derived from Chinese White Kale; Boi1AP-1 and Boi2AP-1 = derived from Calabrese broccoli (Carr and Irish, 1997); 7032-BoAP1a and 7032-BoAP1c = derived from cauliflower DJ7032; 7032-BoAP1a (ant) = published cauliflower *BoAP1* sequence (Anthony *et al.*, 1995). Note: DNA sequence comparison utilised the MEGALIGN sequence alignment program within the DNASTAR software package and are given in appendix 2. Alignments were made using the clustal method, with all parameters set to default. Exons 4 and 8 were excluded from the analysis as sequence was unavailable for comparison from all accessions.

Further examination of published cDNA sequences in comparison with genomic exon sequences of *BoAP1-a* and *BoAP1-c* suggested an error had been made during previous analysis. DNA sequence alignment strongly suggests that the published *BoAP1* sequence derived from cauliflower cDNA (Anthony *et al.*, 1995) is in fact a chimera, constructed from both *BoAP1-a* and *BoAP1-c* expressed sequences (appendix 2). Sequence comparison shows that bases 1-70 of the coding region correspond to *BoAP1-a*, bases 71-290 are from *BoAP1-c*, and bases 291-800 from *BoAP1-a*. This could be the result of expression from a novel *BoAP1* locus. However, three pieces of evidence suggest that this is unlikely to be the case. Firstly, examination of the methods by which the *BoAP1* cDNA sequence was obtained demonstrates that it was constructed from overlapping 5-prime and 3-prime clones *in silico,* rather than from a single full length clone (Anthony, 1994). Secondly, genomic sequence analysis utilising locus specific primers demonstrates that the sequence of *BoAP1-a* and *BoAP1-c* from DJ7032 are both distinct from the published DJ7032-*BoAP1* sequence along their entire length, yet share specific contiguous regions which roughly correspond to the distinct clones initially used. Finally, both *BoAP1-a* and *BoAP1-c* are expressed during curd development (4.4.3), and give the same size fragment on electrophoresis. *In silico* analysis of the PCR primers (RAAPET2-RAAPET3) used by Anthony *et al.* (1995) demonstrates that both *BoAP1-a* and *BoAP1-c* would have been present in the sample used for cloning. If proven, this finding may have significant implications, providing both a good example of the problems encountered during molecular genetic analysis of replicated genomes, and possibly affecting previous analysis of the role of *BoAP1-a* in cauliflower curding.

4.4.1.11 Putative contributing mutations

4.4.1.11.1 Distinctions between *BoAP1-a* and *BoAP1-c*

Comparative sequence analysis of the exonic sequences of *BoAP1-a* and *BoAP1-c* from both the Chinese White Kale A12DH and the cauliflower DJ7032 revealed several amino acid substitutions (fig. 4.8, appendix 2). *BoAP1-c* was found to be identical in both the kale and the cauliflower yet possessed 12 amino acid substitutions in comparison with the *BoAP1-a* sequences (table 4.3). Examination of the sequences from each locus of the Brussels Sprout (DJ3753), and the two Calabrese (BI87053, GD33) sequenced suggested that these substitutions were conserved at each locus and may therefore prove to possess functional significance.

```
A12DH   BoAP1-c     MGRGRVQLKRIENKINRQVTFSKRRAGLFKKAHEISVLCDAEVALVVFSHKGKLFEYS
A12DH   BoAP1-a     ---------------N-----------M---------------------------
DJ7032  BoAP1-a     ---------------S?----?--?---M-?-------------?--------?---

A12DH   BoAP1-c     TDSCMEKILERYERYSYAERQLIAPESDVNTNWSMEYNRLKAKIELLERNQRHYLGED
A12DH   BoAP1-a     ----------------------------S----------------------------
DJ7032  BoAP1-a     ---???????-----------------S----------------------------

A12DH   BoAP1-c     LQAMSPKELQNLEQQLDTALKHIRSRKNQLMYDSVNELQRKEKAIQEQNSMLSKQIKE
A12DH   BoAP1-a     ----------------------------------I----------D?-----K----
DJ7032  BoAP1-a     ------?--------------------------I-------???????----N----

A12DH   BoAP1-c     REKVLMAQQEQWDQQNHGQNMPSPPPPQQHQIQHPYMLSHQPSPFLNMGGLYQEEDQM
A12DH   BoAP1-a     --N--R-------E----H---P-S-------------------------------?
DJ7032  BoAP1-a     --N--R-------E----H---P-P-----?--------------------------

A12DH   BoAP1-c     AMRRNDLDLSLEPVYNCNLGCSAA*
A12DH   BoAP1-a     ------------------L-SF---
DJ7032  BoAP1-a     ------------------F-?????
```

Figure 4.8: Manual alignment of theoretical exonic translation of *BoAP1-a* and *BoAP1-c* from the cauliflower DJ7032 and the Chinese White Kale A12DH. Note: *BoAP1-c* sequence is identical in both accessions.

Table 4.3: Amino acid substitutions and possible effects on protein function between *BoAP1-a* and *BoAP1-c*.

Position	*BoAP1-a*	*BoAP1-c*	*Conservative/Divergent*
29	methionine	phenylalanine	hydrophobic (conservative)
87	serine	valine	hydrophilic/phobic (divergent)
151	isoleucine	valine	conservative
163	aspartic acid	glutamic acid	conservative
177	asparagine	lysine	divergent
180	arginine	methionine	hydrophilic/phobic (divergent)
188	glutamic acid	glutamine	charge change
193	histidine	glutamine	divergent
197	proline	serine	divergent (bending-compensated by 199?)
199	serine	proline	divergent (bending-compensated by 197?)
253	serine	cysteine	hydrophilic/phobic (divergent)
254	phenylalanine	serine	hydrophobic/philic (divergent)

4.4.1.11.2 Cauliflower specific mutations

Comparative analysis of incomplete *BoAP1-a* sequence between A12DH and DJ7032 revealed four amino acid substitutions (table 4.4). This raises the possibility of identification of possible contributing mutations to the curd trait of the cauliflower.

Table 4.4: Amino acid substitutions and possible effects on protein function of *BoAP1-a* between the Chinese White Kale A12DH and the cauliflower DJ7032.

Position	*Kale*	*Cauliflower*	*Conservative/Divergent*
16	asparagine	serine	divergent
170	lysine	asparagine	charge change (divergent)
199	serine	proline	divergent (bending)
251	leucine	phenylalanine	conservative

The residue defined by the asx→ser substitution in the MADS-Box, acts to bind the CARG-Box promoter sequence of target genes (Pellegrini *et al.*, 1995). The asparagine residue is highly conserved in plant MADS-Box genes with only *DEFH24* and *TM3* possessing a substitution at this residue (Pellegrini *et al.*, 1995). However, the residue *is* substituted by serine, suggesting serine may still allow the molecule to function.

A second substitution ser→pro at residue 199 may prove to be significant. Examination of *BoAP1-a* and *BoAP1-c* show identical residue substitutions extremely close to this position (table 4.3). However, the run of prolines, known to be extremely important in protein bending and transcriptional activation, may have a requirement for a serine residue at this point. Thus the inverse substitutions seen between *BoAP1-a* and *BoAP1-c* at residues 197 and 199 could act as compensating mutations. The replacement of the sole serine residue with an extra proline at this region may have an effect on the bending of the polypeptides and thus inhibit its function in some way. The position of this substitution in the C-terminal region of the gene may have further significance in light of recent work demonstrating the role of this region in directing transcriptional activation (Cho *et al.*, 1999).

4.4.1.12 Conclusions

1. Three distinct copies of *BoAP1* have been identified

2. DNA sequence comparisons between loci strongly suggest that *BoAP1* (Anthony *et al.,* 1995) has been artificially constructed as a chimeric hybrid between *BoAP1-a* and *BoAP1-c.*
3. *BoAP1*-a is equivalent to *BobAP1* (M. Yanofsky, pers. comm.) and Boi2*AP1* (Carr and Irish, 1997), and thus is expressed in broccoli.
4. *BoAP1*-c is equivalent to Boi1*AP1* (Carr and Irish, 1997) and is expressed in broccoli
5. DNA sequence analysis has revealed four cauliflower specific coding polymorphism in *BoAP1-a* that could contribute to the development of curd.

4.4.2 SECTION 2: GENETIC LINKAGE MAPPING

The discovery of three distinct loci suggested that further examination of the role of each in the development of curd was required. In view of this, genetic linkage mapping was employed in an attempt to assign genomic positions to loci, and to assess via genetic means the possible contribution of each locus.

4.4.2.1 Marker design

4.4.2.1.1 *BoAP1*-a

DNA sequencing of *BoAP1*-a between the locus specific primers AP1I6F and AP1I8R (appendix 1) identified a single nucleotide difference in intron 6 between the mapping parents A12DH and GD33. This difference could be exploited by digesting the resultant amplicons with the restriction enzyme *Mbo*II which cut the amplicon from GD33 but not A12DH (fig. 4.9). The segregating DH population constructed from these lines was screened with the generated Cleaved Amplified Polymorphic Sequence (CAPS) marker, and the resultant segregation data subjected to linkage analysis.

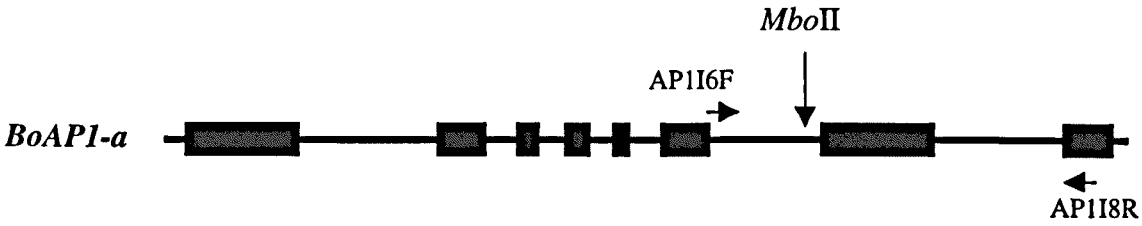


Figure 4.9: Schematic representation of CAPS-assay designed from *BoAP1-a MboII* cuts the amplicon from GD33 but not A12DH, allowing mapping in the context of the A12DH x GD33 genetic linkage map.

4.4.2.1.2 Chi-squared analysis

Chi squared analysis was used to examine the possibility of segregation distortion affecting the ratios of alleles at this locus in the A12DH x GD33 population.

The null hypothesis states: Segregation of *BoAP1-a* alleles in the A12DH x GD33 population is inconsistent with classical Mendelian ratios expected for a single locus (i.e. 1:1).

Table 4.5: Chi-squared test of observed segregation ratios of *BoAP1-a* alleles amongst the A12DH x GD33 population against those expected for a single locus under classical Mendelian segregation (1:1 ratio).

Observed frequency (O)	*Expected frequency (E)*	*O-E*	$O\text{-}E^2$	$O\text{-}E^2/E$
24	44	-20	400	9.09
64	44	20	400	9.09

$$\chi^2 = \sum O\text{-}E^2/E = 18.18, \text{ with 1 degree of freedom.}$$

The probability of observing a χ^2 with 1 degree of freedom greater in value than 0.004 is 95%. The probability of observing a χ^2 with 1 degree of freedom smaller in value than 3.84 is 95%. The calculated χ^2 value of 18.18 lies outside of these limits thus suggesting that the null hypothesis is correct. The observed and expected ratios are significantly independent, demonstrating that the observed data does not have a 'good fit' with ratios expected for classical Mendelian segregation at a single locus in a DH

population. This demonstrates that segregation distortion has significantly altered the segregation ratio of alleles at this locus, and confirms that *BoAP1-a* segregation is distorted in the A12DH x GD33 population.

4.4.2.1.3 *BoAP1*-b

DNA sequencing from all of the parents of the mapping populations did not reveal any polymorphism that would allow map localisation of *BoAP1*-b.

4.4.2.1.4 *BoAP1*-c

PCR amplification and sequencing of *BoAP1-c* between the locus specific primers AP1B5F and AP1667R (appendix 1) allowed the design of a further CAPS assay. The restriction enzyme *Tsp*RI was found to digest the amplicon from DJ3753 (G) but not DJ7032 (N) (fig. 4.10), which allowed mapping of *BoAP1*-c in the context of the NxG genetic map. Chi-squared analysis was conducted in an identical manner to *BoAP1-a.* Results are given in table 4.6.

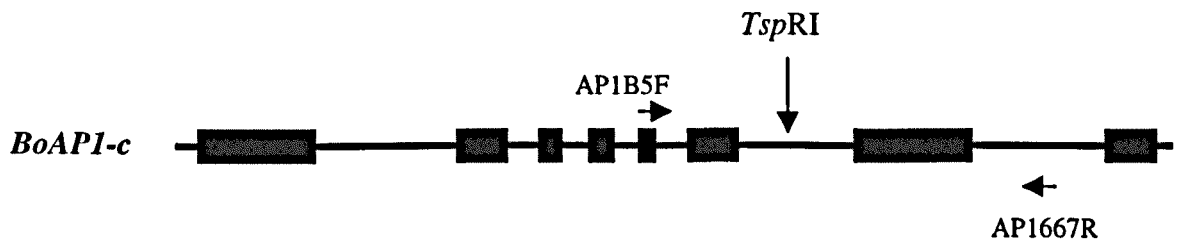


Figure 4.10: Schematic representation of CAPS-assay designed from *BoAP1-c.* The restriction enzyme *Tsp*RI was found to digest the amplicon from DJ3753 (G) but not DJ7032 (N), which allowed mapping of *BoAP1*-c in the context of the NxG genetic map.

4.4.2.1.5 *AP1*/CAL-like 1 and 2

Further Chi-squared analysis on two RFLP markers found to hybridise CAL and *AP1* in the NxG population are also given in table 4.6. Note. The CAPS assay designed for *BoAP1-c* ruled out either of these two loci as representing *BoAP1-c.*

Table 4.6: Chi-squared analysis and segregation distortion results of *BoAP1*-like loci.

	Population	*Population size*	*Chi-squared (1. d.f.)*	*Signif. Diff between observation and expected Mendelian ratios*
BoAP1-a	A12DH x GD33	88 (64:24)	18.18	Yes
BoAP1-c	NxG	91 (36:55)	3.96	Yes
AP1/CAL-like 1	NxG	91 (34:57)	5.82	Yes
AP1/CAL-like 2	NxG	94 (34:60)	7.20	Yes

Examination of the Chi-squared analysis of these markers shows that the observed and expected ratios are significantly independent. This demonstrates that the observed data does not have a 'good fit' with ratios expected for classical Mendelian segregation of a single locus, and suggests that segregation distortion has significantly altered the segregation ratio of alleles at these loci.

4.4.2.2 Linkage mapping

Identification of the map location of these distorted loci provided a possible explanation for the significant distortion of each away from classical Mendelian ratio of 1:1 expected for a single locus. Genotype scores for each locus in each line were subjected to linkage analysis using the program JOINMAP (Stam and Van Ooijen, 1995). Lowest recombination frequency and LOD scores to linked markers are given in table 4.7.

Table 4.7: Recombination frequency and LOD scores to the closest linked markers of *BoAP1*-like loci.

Locus	*Mapping population*	*Linkage group*	*Nearest markers*	*Recombination frequency*	*LOD score*
BoAP1-a	A12DH x GD33	O6	pW197E2	0.02	16.37
			mNGA111J1	0.00	7.42
BoAP1-c	NxG	O6	pW134J1	0.11	11.99
			ACCTAJ07	0.08	14.48
AP1/CAL-like1	NxG	O6	pW134J1	0.07	15.81
			ACCTAJ07	0.12	11.24
AP1/CAL-like2	NxG	O6	mNGA111J1	0.06	15.35
			ACCTAR08	0.06	14.89

As these loci were mapped in distinct, yet collinear populations, the linkage results for O6 from each population could be combined to produce an integrated map (Fig. 4.11). Further examination also identified a collinear region with *A. thaliana*, where flanking markers of *AP1*, are conserved around *BoAP1*-a.

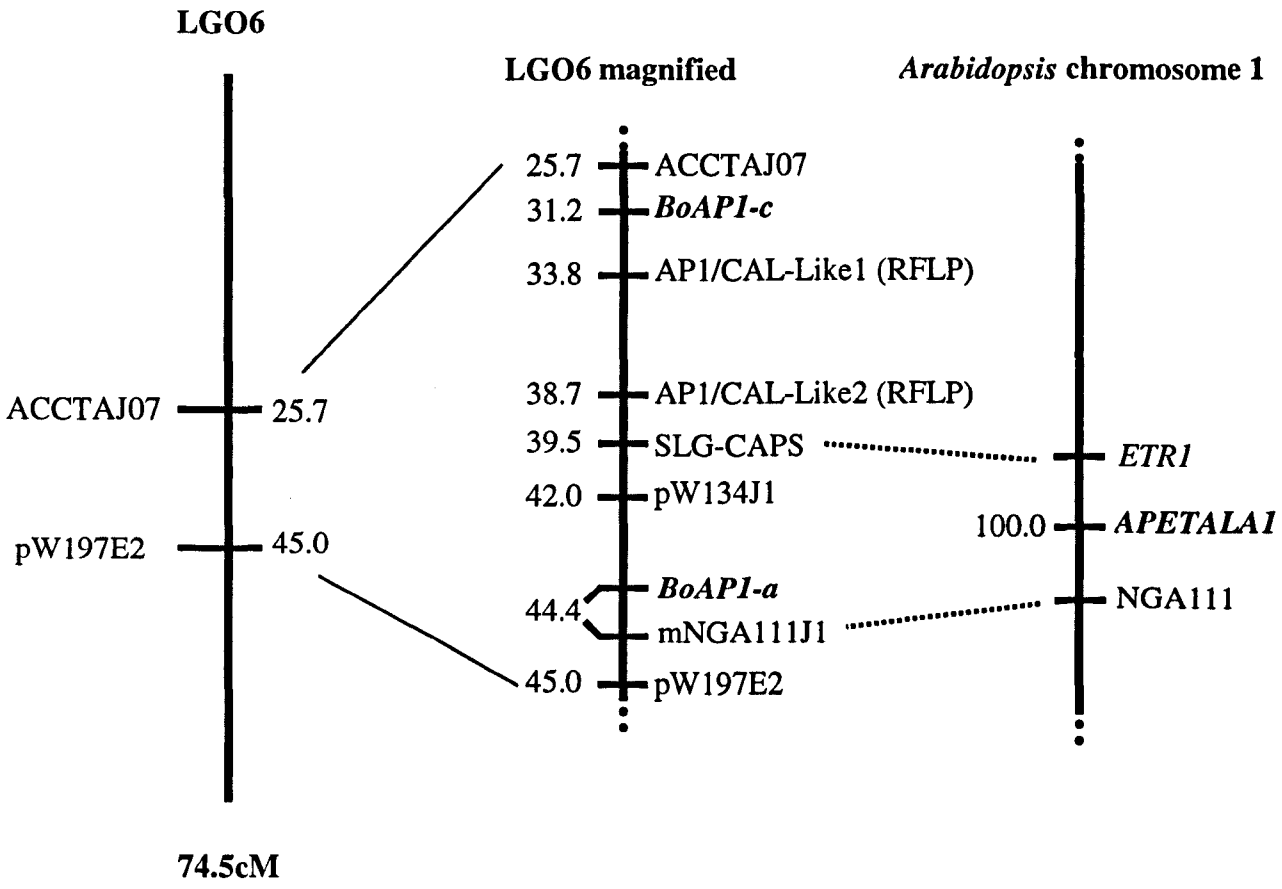
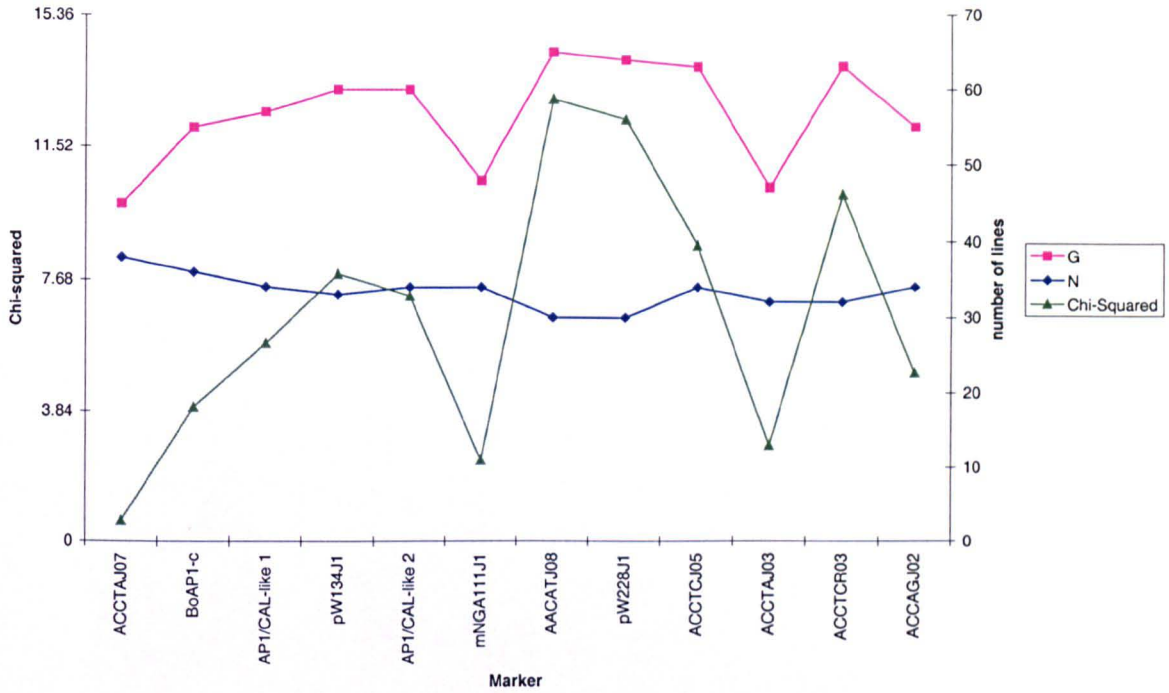


Figure 4.11: An integrated *B. oleracea* genetic map of linkage group O6 derived from the A12DH x GD33 and NxG populations. Genetic map locations of *BoAP1-a* and *BoAP1-c* are shown along with RFLP segregation scores for two loci identified during probing with the AP1 and CAL probes. The genetic map position of the Arabidopsis *AP1* linked microsatellite NGA111 is also shown, along with the Self Incompatibility locus (*SLG*). A gene flanking *SLG, ETR1,* has also been shown to map close to *AP1* in *A. thaliana* (Conner *et al.*, 1998). Distances are given in Centimorgans (cM).

4.4.2.3 Analysis of segregation distortion on LGO6

The mapping of loci showing distorted linkage to O6 suggests that this linkage group as a whole may not be segregating according to classical Mendelian genetics. To address this possibility the chi-squared values of markers located on O6 in both mapping populations were calculated (all marker scores except those relating to *BoAP1* and *AP1*/CAL-like loci were provided by R. Sebastian (Sebastian *et al.*, 1999) and (Bohuon *et al.*, 1996)). The results of this analysis are summarised in figure 4.12.

(a)



(b)

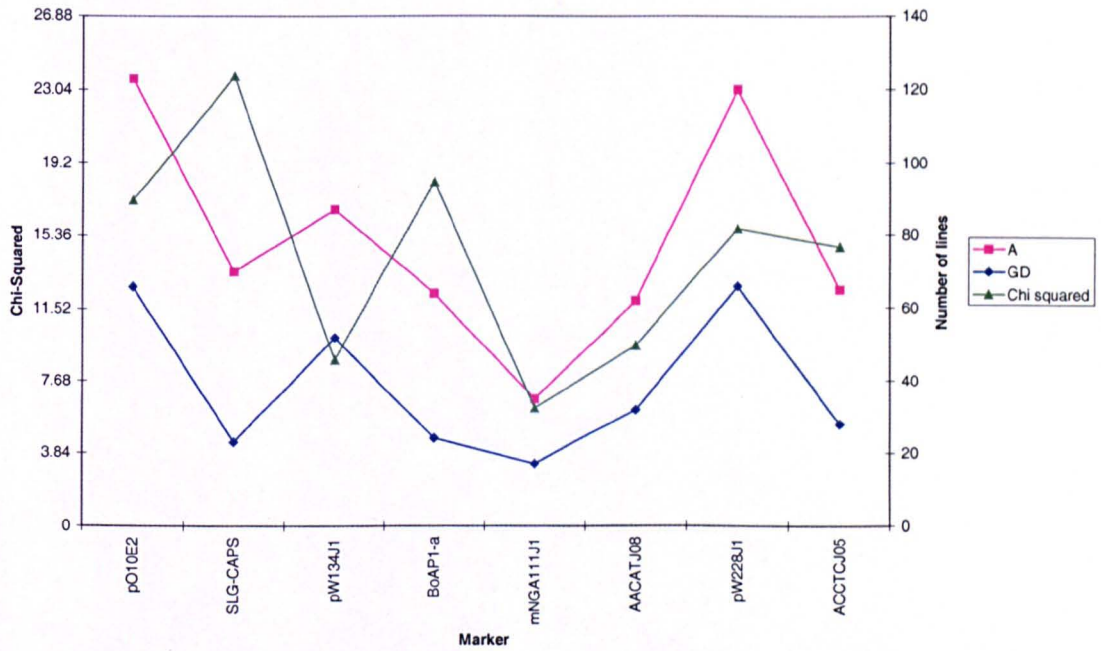


Figure 4.12: Graphs showing the chi-squared values and parental origin of markers along LGO6 of (a) the NxG and (b) the A12DHxGD33 populations. Chi-squared values greater than 3.84 denote a significant difference between observed and expected segregation ratios $p<0.05$, and thus the presence of segregation distortion. Number of lines scored is also shown.

The NxG population shows segregation has consistently been distorted in favour of the male DJ3753 (Brussels Sprout) alleles along O6 with regions defined by mNGA111J1 and ACCTAJ03 alone segregating in Mendelian fashion (1:1). The A12DH x GD33 population shows segregation has consistently been distorted in favour of the female A12DH (Chinese White Kale) across the entire chromosome.

Comparison of chi-squared values of collinear markers between the two populations suggests that although distorted across the entire chromosome, the A12DHxGD33 population shows a generally similar pattern of segregation distortion to the NxG population (correlation coefficient of 0.63) with minimal distortion around mNGA111J1 (fig. 4.13).

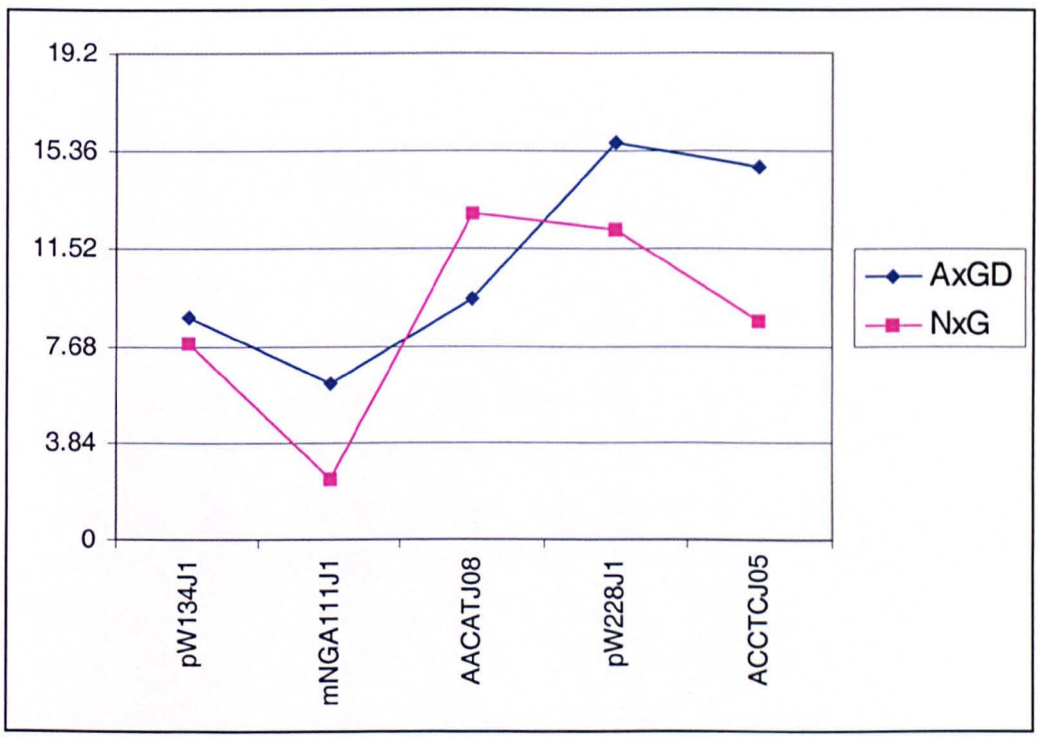


Figure 4.13 A Graph showing the chi-squared values of collinear markers along O6 of the NxG and the A12DHxGD33 populations. Values greater than 3.84 denote a significant difference between observed and expected segregation ratios ($p<0.05$), and thus the presence of segregation distortion.

4.4.2.4 Conclusions

1. Molecular locus specific markers have been designed for each locus
2. *BoAP1*-a and *BoAP1*-c and two further CAL/AP1-like RFLP markers map to O6 in significantly distorted regions of the genome.
3. O6 as a whole shows significant segregation distortion in both the A12DHxGD33 and NxG DH populations.
4. Colinearity between *B. oleracea* and *A. thaliana* has been identified around *BoAP1-a.*

4.4.3 SECTION 3: EXPRESSION PROFILING

4.4.3.1 Functional analysis of *BoAP1*

The identification of three distinct *BoAP1* loci raises the question of attributing function to each copy. In particular, is the control of floral induction in the *Brassica* cauliflower controlled by multiple copies of *BoAP1*, or has gene silencing reduced control to a single key copy.

To address this question, RNA was isolated from cauliflower curd tissue of the DH cauliflower DJ7032, both at the arrested curd stage and curd from tissue expressing floral bud primordia. The timing of this initial sampling was chosen as the maximal expression of *BoAP1* (Anthony *et al.*, 1996) and spanning the ontological stage of floral induction.

Total-RNA was converted to cDNA and *Reverse Transcriptase*-PCR (RT-PCR) was carried out between generic *BoAP1* primers (table 4.8).

Table 4.8: Primers and restriction enzymes used in the analysis of *BoAP1* cDNA

Locus	*Primer 1*	*Primer 2*	*Restriction enzyme*
BoAP1-a	AP1BPF	AP1B8R	*Dra*I
BoAP1-b	AP1INDEL4F	AP1B8R	N/A
BoAP1-c	AP1BPF	AP1B8R	*Sca*I

DNA sequence analysis revealed that use of the exonic primer AP1INDEL4F would be sufficient to distinguish *BoAP1*-b from either *BoAP1*-a or *BoAP1*-c expression. However, due to sequence similarity, locus-specific primers from *BoAP1*-a and *BoAP1*-c were located exclusively within introns. To overcome potential difficulties in distinguishing between loci, restriction digest polymorphisms between the two loci were utilised, supported by DNA sequencing of digested bands extracted from agarose gels (table 4.8).

4.4.3.2 *BoAP1-a*

Restriction digest of amplicons from the generic *BoAP1* RT-PCR amplification showed that, *BoAP1-a* is expressed during floral induction of the *Brassica* cauliflower (fig. 4.14).

4.4.3.3 *BoAP1-c*

Further analysis revealed that *BoAP1*-c is also expressed during floral induction in the *Brassica* cauliflower (fig. 4.14).

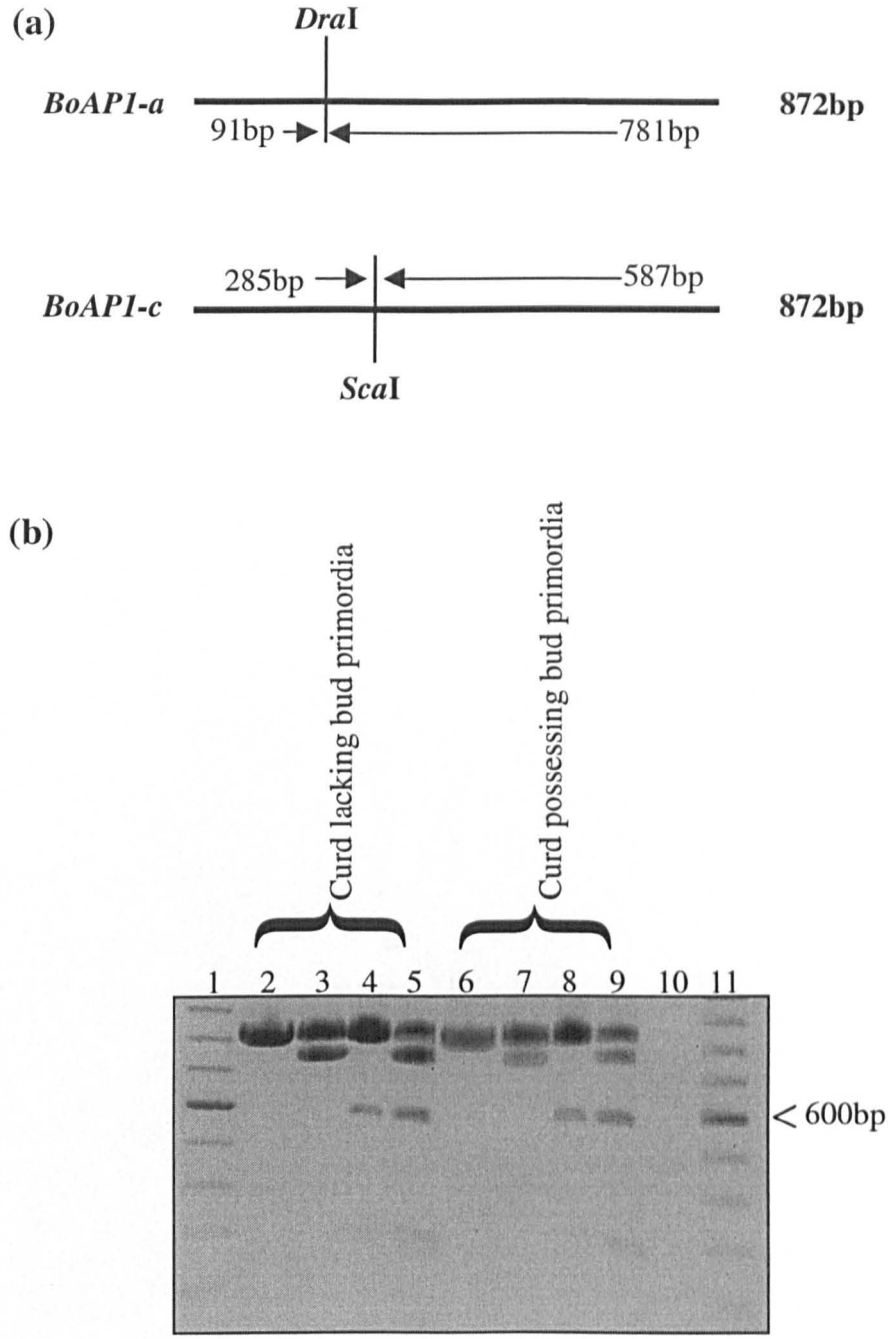


Figure 4.14: Restriction map (a) and digest (b) of amplicons between AP1BPF-AP1B8R on *Brassica* cauliflower cDNA reveals expression of both *BoAP1-a* and *BoAP1-c* at the curding stage and following floral induction. Lanes 1 and 11 = marker lane, (100 bp increments, 600 bp shown for orientation). Lanes 2 and 6 = uncut amplification product. Lanes 3 and 7 = *Dra*I digested amplicon showing *BoAP1-a* is present. Lanes 4 and 8 = *Sca*I digested amplicon showing *BoAP1*-c is present. Lanes 5 and 9 = double *DraI*/*Sca*I digest. The remaining undigested band corresponds to *BoAP1*-c (DNA sequencing analysis) suggesting low activity of *Sca*I in double digest buffer. Lane 10 = negative control. (All digested bands were DNA sequenced to confirm presence of each locus.

4.4.3.4 *BoAP1-b*

RT-PCR with the *BoAP1-b* specific primers (table 4.8) did not reveal an expected fragment size of 350bp on gel electrophoresis (fig. 4.15), confirming the previous conclusion that *BoAP1-b* is not expressed during floral induction in the *Brassica* cauliflower (Lowman and Purugannan, 1999).

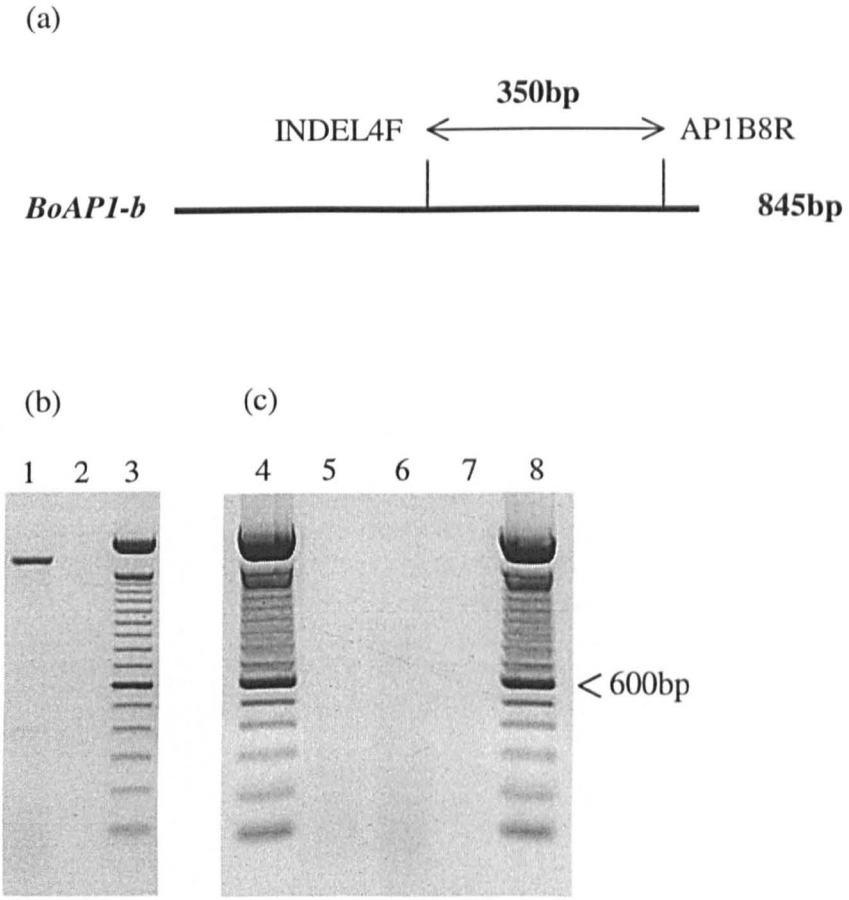


Figure 4.15: (a) cDNA amplification map and expected fragment size from *BoAP1*-b RT-PCR (b) Amplification from genomic and (c) cDNA. Lanes 3, 4 and 8 = marker lane, (100 bp increments, 600 bp shown for orientation). Lane 1 = Genomic DNA control. Lanes 2 and 7 = negative control. Lane 5 = curd lacking bud primordia. Lane 6 = curd possessing bud primordia.

4.4.3.5 Conclusions

1. *BoAP1*-a and *BoAP1*-c are expressed during the switch to floral induction in the *Brassica* cauliflower.
2. *BoAP1*-b is not expressed during floral induction.

4.5 DISCUSSION

4.5.1 Replication of *BoAP1* in *B. oleracea*

Molecular genetic analysis demonstrates that the MADS-box gene *BoAP1* it is present in at least three distinct copies in *B. oleracea.* This observation supports previous evidence suggesting a possible hexaploid ancestry for *B. oleracea* (Lagercrantz and Lydiate, 1996).

The implications of the identification of multiple loci have widespread impact, both in retrospect for previous investigations into cauliflower floral induction, and for investigation of *Brassica* genetics as a whole.

Perhaps not surprisingly the replicated nature of the *BoAP1* locus has only been addressed following the identification of *BoAP1-b* by Lowman and Purugannan (1999) and *BoAP1-c* by the author. Excluding the 'accidental' discovery of a second locus by Carr and Irish (1997) (so called as, although published, evidence of two distinct loci was not proven in this paper), previous investigation had relied largely on Northern and ISH analysis to characterise expression levels (Anthony *et al.*, 1996; M. Yanofsky, Pers. Comm). Once evidence of multiple loci came to light it was apparent that these techniques were too crude to be able to distinguish between loci. This suggested that, until further investigations were conducted utilising a more specific technique such as RT-PCR, the evidence presented for expression levels and patterns of *BoAP1-a* by (Anthony *et al.*, 1996) and (Carr and Irish, 1997) remained in question.

Such reservations proved well-founded upon identification of the probable chimeric *BoAP1* sequence constructed unwittingly by Anthony *et al.*, (1995).

The novel identification of two expressed loci of *BoAP1* at the stage of floral induction in the *Brassica* cauliflower, together with examination of the published cauliflower *BoAP1* sequence from curd suggests that the expression data presented by Anthony *et al.* (1996) is likely to refer to expression of *BoAP1* from at least two distinct loci, both *BoAP1-a* and *BoAP1-c.*

To address this possibility, examination of the expression levels of *BoAP1-a, BoAP1-b,* and *BoAP1-c,* were conducted by RT-PCR on the same DH cauliflower line used by Anthony *et al.* (1996). This demonstrated conclusively that *BoAP1-a* and *BoAP1-c* were both expressed during floral induction in the *Brassica* cauliflower.

Analysis during this study and by Lowman and Purugganan (1999) demonstrated that *BoAP1*-b is not expressed at the stage of floral induction in cauliflower. Further to this the presence of a premature termination codon in exon 4 has been observed in a wide variety of crop types. This suggests that *BoAP1-b* is unlikely to represent the genetic difference between cauliflowers and other crop types. Confirmation through genetic linkage analysis and transformation would ideally be required as proof.

4.5.2 Re-examination of genetic model for curding

The identification of two expressed loci of *BoAP1 (BoAP1-a and BoAP1-c)* at the stage of floral induction in the *Brassica* cauliflower is made all the more significant by the genetic map location of these loci.

The genetic segregation data gathered during the analysis of the curding phenotype amongst the NxB population (chapter 3) identified *BoAP1*-a as a marker linked to curding. The close proximity of the two genes *BoAP1*-a and *BoAP1*-c raises the possibility that either or indeed both of these genes may be contributing to the development of curd. Analysis of segregation of *BoAP1*-c amongst the NxB population identified a single recombinant line (NxB 616b) which disrupts the co-segregation of *BoAP1*-a and *BoAP1*-c from the cauliflower parent. Examination of eight further replicates of this line grown during a randomised field trial in 1999 (E. Kop, Pers. Comm.) confirmed the original scoring of this line as 'Parental-Curd' (see chapter 3).

This suggested that the phenotype was still completely linked to *BoAP1*-a, as line NxB 616b possessed the copy of *BoAP1*-c from the broccoli parent. Whilst this evidence is strong enough to rule out *BoAP1*-c as directly involved in curding it does not definitively state that *BoAP1*-a is the causal locus. The tentative nature of this evidence suggests significant further analysis is warranted before conclusive proof of the genetic model will be forthcoming. Further DNA sequence analysis of the correct *BoAP1-a* sequence from DJ7032 may help address this question.

In conclusion, the genetic model for curding proposed in chapter 3 remains unchanged. If this genetic model is correct, there exists a yet uncharacterised change in function of *BoAP1*-a (or a closely linked gene) that cannot be completely compensated for by overlapping expression of *BoAP1*-c.

4.5.3 Colinearity analysis

The genomic location of *BoAP1*-a and *BoAP1*-c is also interesting from the point of view of colinearity with *A. thaliana.* Flanking regions of the *A. thaliana ETR1* gene, which is localised just above *AP1* in *A. thaliana*, have been identified as being closely linked to collinear regions of the SLG-locus in *Brassica rapa.* (Conner *et al.*, 1998). This SLG locus is collinear between the A and C genomes of *B. napus* (i.e. *B. rapa* x *B. oleracea* amphidiploid), and has therefore also been mapped in *B. oleracea.* The SLG locus of *B. oleracea* is situated on O6 (C. Ryder. Pers. Comm) (Sebastian *et al.*, 1999), just above *BoAP1-a.* Analysis of a second marker, the *A. thaliana* microsatellite nga111, provides a similar result. In *A. thaliana* this marker is situated just below *AP1.* The *B. oleracea* homologue co-localises with *BoAP1-a.*

This analysis therefore raises the question of origin of these genes in the hexaploid ancestry of *B. oleracea.* The common ancestor of *A. thaliana* and *Brassica* could make up just one third of the hexaploid progenitor of *B. oleracea,* or all three loci could result from within genome replication or hybridisation. Further experiments are required to characterise flanking markers around other *BoAP1* loci to address this.

4.5.4 Segregation distortion

Segregation distortion across linkage group O6 has been identified in similar patterns in both the A12DHxGD33 and the NxG populations.

DH lines undergo selection for seed production following initial embryo growth. This suggests that the observed evidence may correlate with the presence of the S-locus on linkage group O6. The rationale for this is that the weaker the S-allele possessed from the parent the greater the seed set and the more likely the line is to be selected. This theory applies well to the A12DH x GD33 population in which A12DH is the more fertile of the two progenitors, suggesting that segregation distortion of this chromosome may be due to linkage drag of the S-locus. However, in the NxG population the Brussels Sprout parent is vastly poorer at setting seed (R. McClenaghan. Pers. Comm.), yet linkage group O6 is skewed in its favour. However, examination of tissue culture experiments suggests DJ3753 responds well to anther culture (R. McClenaghan. Pers. Comm.). This may indicate that a region on O6, independent of the S-locus, affects regeneration following anther culture.

4.6 Final conclusions

- Three independent copies of *BoAP1* have been identified and characterised in *Brassica oleracea.*
- Examination of these loci has allowed the development of a consensus for assignment of previously published *BoAP1* cDNA sequences, and a combining of independent evidence of expression patterns for specific loci.
- Examination also strongly suggests that the previously published *BoAP1* sequence from the cauliflower DJ7032 (Anthony *et al.,* 1995) is in fact an artificial chimera, constructed from overlapping partial cDNA clones.
- Genetic mapping demonstrates that at least two of the three identified *BoAP1* loci (*BoAP1-a* and *BoAP1-c*) are closely linked on LGO6 of the consensus *B. oleracea* map (Sebastian *et al.,* In press), in a region showing high segregation distortion in two distinct populations (A12DHxGD33 and NxG).
- In spite of this close linkage a single recombinant (NxB 616b) suggests that *BoAP1-c* is not directly involved in curd development in the cauliflower DJ7032,

and that the initial genetic model described in chapter 3 continues to be supported.

- RT-PCR experiments have demonstrated that both *BoAP1-a* and *BoAP1-c* are expressed during floral induction in the *Brassica* cauliflower, however, *BoAP1-b* does not appear to be expressed at this stage.
- DNA sequencing analysis confirmed the presence of a premature termination codon in exon 4 of *BoAP1-b*. This mutation is found in sequences of several distinct crops suggesting it plays no direct role in distinguishing between cauliflower and Calabrese inflorescence morphology.
- DNA sequencing though incomplete, highlighted four polymorphisms in the coding regions of *BoAP1-a* that could potentially contribute to the curding phenotype of cauliflower.

CHAPTER 5: GENETIC AND CYTOLOGICAL LOCALISATION OF *BoCAL-a* IN *BRASSICA OLERACEA*

5.1 Introduction

Several detailed genetic maps have been produced for *Brassica oleracea* which have then allowed comparative mapping with *A. thaliana*. These investigations suggest that many single copy loci in *A. thaliana* are replicated in collinear regions of *B. oleracea* (Bohuon *et al.*, 1996; Sebastian *et al.*, 1999; Slocum *et al.*, 1990). The most recently characterised example of this is *BoAP1* (chapter 4).

Comparative analysis between *Brassica* and *A. thaliana* suggests that although the colinearity of markers is conserved over small regions (Cavell *et al.*, 1998; Conner *et al.*, 1998; Lagercrantz *et al.*, 1996), the chromosomal organisation and physical distance between these markers differs between *Brassica* and *A. thaliana.*

Several of the problems associated with identification and mapping of replicated loci in *B. oleracea* have arisen from the ability to identify usable polymorphisms. Much of the initial mapping information relied on RFLP data from digests with just a few enzymes (Bohuon *et al.*, 1996; Sebastian *et al.*, 1999; Slocum *et al.*, 1990). This has resulted in the genetic maps failing to show the true level and distribution of marker replication due to the problem of monomorphic RFLP bands. With physical mapping in *Brassica* still in its infancy, a more informative mechanism of identifying and characterising replicated loci could be employed following the detailed karyotyping of *B. oleracea* chromosomes.

Fluorescent *in situ* hybridisation (FISH) is becoming an increasingly utilised tool for genome analysis, as this cytological approach allows rapid physical localisation of loci regardless of recombination events (Armstrong *et al.*, 1998). This approach also has, in theory, the advantage of providing a true estimate of physical distance of loci of interest. A valuable piece of information for future physical mapping, as this may highlight recombination hotspots and 'silent' regions of the genome which distort genetic map distances and reduce the likelihood of successful map based cloning.

The difficulties which were encountered whilst attempting to identify replicated loci of *BoCAL* through RFLP data and BAC library analysis, makes *BoCAL-a* an ideal candidate to test the possibilities provided by an approach of cytological analysis, in combination with molecular genetic characterisation.

5.2 Objectives

1. Identify the genetic map position of *BoCAL-a.*
2. Identify which cytological chromosome corresponds to the genetic linkage group.
3. Examine the cytological hybridisation of *BoCAL-a,* and determine likely copy number and distribution of *BoCAL* copies.

5.3 Strategy

The genetic map position of *BoCAL-a* was identified utilising the CAPS assay designed to exploit the premature termination codon in exon 5 of *BoCAL-a* (chapter 3). The amplification required to perform this assay was then applied to an addition line of *B. rapa* thought to contain the *B. oleracea* chromosome/linkage group in question. FISH analysis of a genomic clone containing the *BoCAL-a* locus was then applied to identify multiple loci of *BoCAL-a* in *B. oleracea* and determine their distribution.

5.4 Results

5.4.1 SECTION1: Genetic linkage mapping

5.4.1.1 Marker design

The CAPS assay designed to follow segregation of *BoCAL-a* in the NxB population (3.4.2.2) was utilised to identify the genetic map position of *BoCAL-a* in the context of the NxG map (Sebastian *et al.*, 1999). This was possible as the *Spe*I restriction site linked to curding was absent from the Brussels Sprout progenitor (G) of the

population. This map position was later confirmed with two further assays. The first was designed around a *Dra*I restriction site polymorphism in the amplicon between CALI1F-CALI2R (intron1-2) which digested the amplicon from DJ7032 (N) but not DJ3753 (G) (fig. 5.1). The second was a Simple Sequence Repeat (SSR) identified in intron 7 (see chapter 6 for details) (fig. 5.1). Amplification of these primer sets against the *BoCAL-a* genomic clone SK162 showed that all primer pairs amplified from this locus. Segregation data from each of the assays demonstrated that amplifications from each section of the *BoCAL-a* locus mapped to identical positions in the genome. These two complimentary sets of evidence suggested that these amplicons were likely to have been derived from the same copy of *BoCAL.*

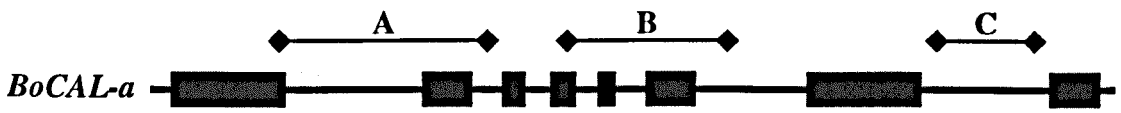


Figure 5.1: Schematic showing position of amplicons used independently for genetic linkage mapping, against the genomic sequence of *BoCAL-a.* Blocks denote exons, joining lines denote introns. Amplicon A = CALI1F to CALI2R, amplicon B = CAL4F to CALI6R, amplicon C = CALSSRF to CALSSRR (appendix 1).

5.4.1.2 Chi-squared analysis

Chi-squared analysis was employed to investigate whether *BoCAL-a* was situated in a region of the genome showing segregation distortion in the NxG population.

The null hypothesis states: Segregation of *BoCAL-a* alleles in the NxG population is inconsistent with classical Mendelian ratios expected for a single locus (i.e. 1:1).

Table 5.1: Chi-squared test of observed segregation ratios of *BoCAL-a* alleles amongst the NxG population against those expected for a single locus (i.e. 1:1) under classical Mendelian segregation.

Observed frequency (O)	*Expected frequency (E)*	*O-E*	*O-E²*	*O-E²/E*
41	46	-5	25	0.54
51	46	5	25	0.54

$\chi^2 = \sum$ O-E²/E = 1.08, with 1 degree of freedom.

The probability of observing a χ^2 with 1 degree of freedom greater in value than 0.004 is 95%. The probability of observing a χ^2 with 1 degree of freedom smaller in value than 3.84 is 95%. The calculated χ^2 value of 1.08 lies within these limits thus suggesting that the null hypothesis is incorrect. The observed and expected ratios are not significantly independent, demonstrating that the observed data has a 'good fit' with ratios expected for classical Mendelian segregation. This demonstrates that any segregation distortion has not significantly altered the segregation ratio of alleles at this locus.

5.4.1.3 Genetic map position

Segregation data for each line were submitted for linkage analysis using the genetic linkage map construction programme JOINMAP (Stam and Van Ooijen, 1995). Recombination frequency and LOD scores to the closest linked markers are given in table 5.2.

Table 5.2: Recombination frequency and LOD scores to the closest linked markers of *BoCAL-a* in the NxG population.

Locus	*Mapping population*	*Linkage group*	*Nearest markers*	*Recombination frequency*	*LOD score*
BoCAL-a	NxG	LGO3	pW146J1	0.01	25.41
			pO172J1	0.01	24.71

A graphical representation of the Genetic map position of *BoCAL-a* on the NxG map is given in figure 5.2.

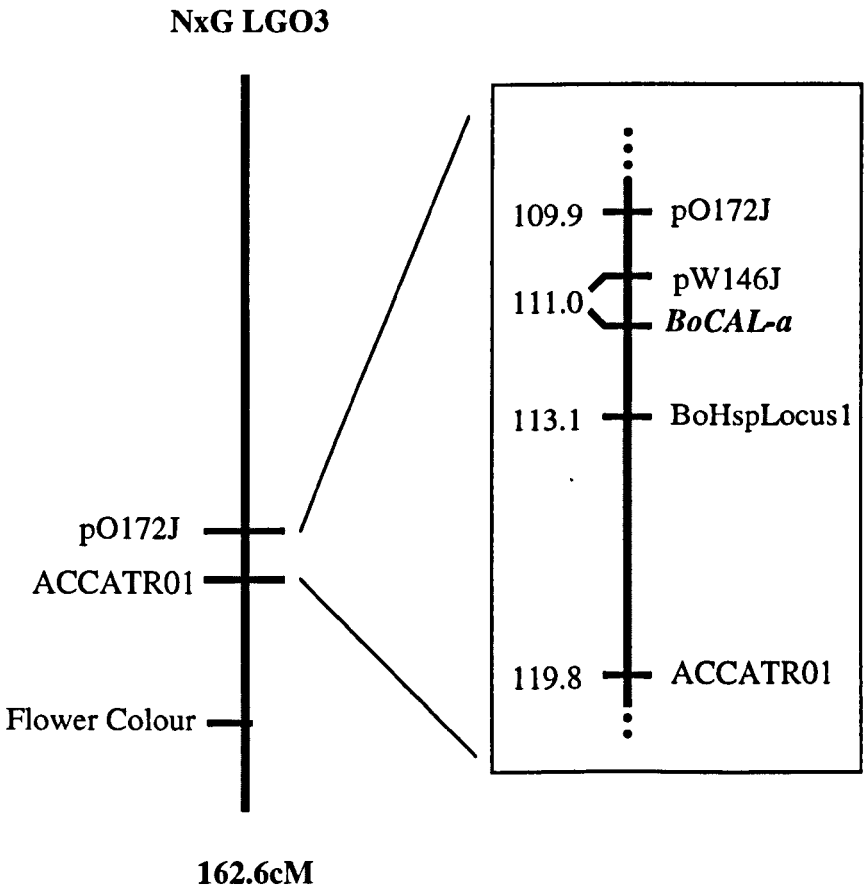


Figure 5.2: Genetic linkage map position of *BoCAL-a* on the NxG linkage map. Closest linked markers are shown. Distances are given in Centimorgans (cM). Note: *BoCAL-a* is linked to the flower colour locus of *B. oleracea.*

5.4.2 SECTION 2: Addition lines: A bridge between molecular and cytological analysis

The observation that *BoCAL-a* was linked to flower colour on LGO3 raised the possibility of assigning linkage group O3 to a cytologically characterised chromosome.

Flower colour in diploid *Brassicas* can be either white or yellow, depending upon the allelic status at a single locus. Analysis of an F2 population (NxA12DH) segregating for flower colour was scored to assess the relative dominance of flower colour. This experiment suggested white was dominant to yellow (table 5.3).

5.4.2.1 Chi-squared

The null hypothesis states: Segregation of flower colour in the NxA12DH F2 population in is inconsistent with a classical Mendelian 3:1 ratio expected for a single dominant gene.

Table 5.3: Chi-squared test of observed segregation of flower colour amongst the NxA12DH F2 population.

	Observed frequency (O)	*Expected frequency (E)*	*O-E*	*$O-E^2$*	*$O-E^2/E$*
White flower	22	22.5	0.50	0.25	0.01
Yellow flower	8	7.50	0.50	0.25	0.01

$\chi^2 = \sum O\text{-}E^2/E = 0.02$ with 1 degree of freedom.

The probability of observing a χ^2 with 1 degree of freedom greater in value than 0.004 is 95%. The probability of observing a χ^2 with 1 degree of freedom smaller in value than 3.84 is 95%. The calculated χ^2 value of 0.02 lies within these limits thus suggesting that the null hypothesis is incorrect. The observed and expected ratios are not significantly independent, demonstrating that the observed data has a 'good fit' with ratios expected for classical Mendelian segregation. This strongly suggests that white flower colour is dominant to yellow.

5.4.2.2 Monosomic addition lines

Monosomic addition lines were developed by Chen *et al.* (1997) from a yellow flowered *B. rapa* genotype (2n=20), with the addition of a single chromosome from a white flowered *B. oleracea* (2n=18), to give lines segregating with a proportion of 2n=20+1 plants.

Examination of individuals from the specific addition line containing the *B. oleracea* cytological chromosome 4 (BCA4), showed that this addition line segregated for flower colour (fig. 5.3), whereas all other addition lines possessed only yellow flowers. This suggested that chromosome 4 from *B. oleracea* may contain the flower colour locus, and thus that chromosome 4 may correspond to linkage group three of the genetic map.

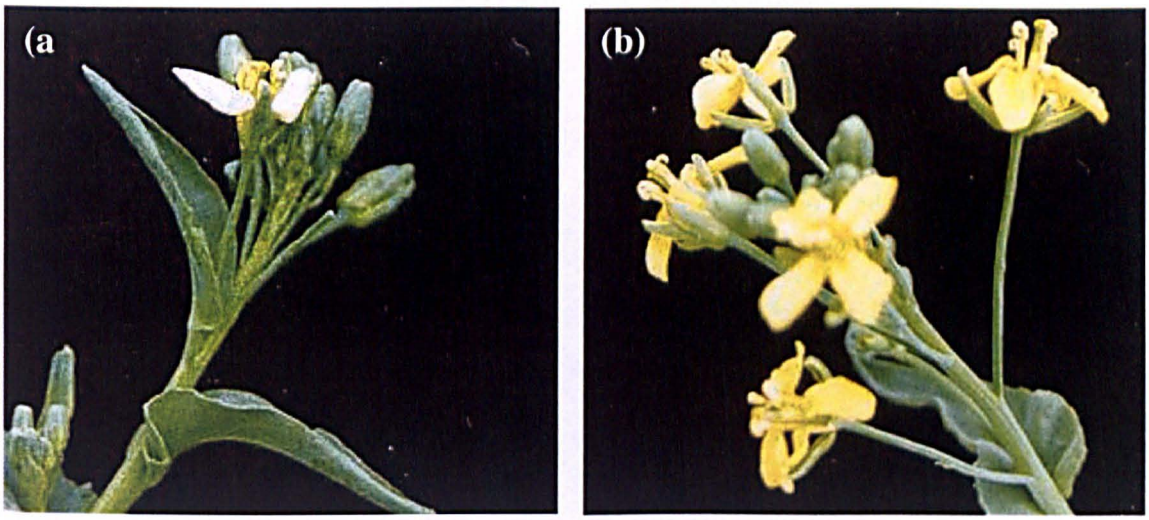


Figure 5.3: phenotype of segregating plants of the addition line BCA4. *B. rapa* 2n = 20 plants express yellow flower colour (b), whereas plants possessing the extra chromosome 4 from *B. oleracea* express white flower colour (a). This suggests that the flower colour locus is present on *B. oleracea* chromosome 4.

To examine this possibility, flow cytometry (Plant Cytometry Services, P.O. Box 299, AG Schijndel, The Netherlands) was initially employed to examine the relative amounts of DNA present in the addition lines segregating for flower colour, compared to the *B. rapa* parent. The results of this analysis demonstrated that all of the offspring of the addition line BCA4 displaying the white flowered phenotype were found to contain approximately five percent more DNA than the yellow flowered plants (table 5.4). These results suggested that some of the plants contained at least one extra chromosome and that presence of this extra DNA is linked to the white flower colour.

Table 5.4: Comparison of DNA content and flower colour in the addition line BCA4. This shows that white flower colour is linked to the addition of monosomic chromosomal material.

Individual	*Flower colour*	*Relative DNA content (C value)*
A1 (*B. rapa* parent)	Yellow	1.00
A2 (*B. oleracea* donor)	White	1.38
A3	White	1.05
A4	White	1.05
A5	Yellow	1.00
A6	Yellow	1.00
A7	Yellow	1.00
A8	White	1.05

The addition of a single chromosome to give 2n = 20+1 in line BCA4 was confirmed with cytological evidence during FISH analysis (fig. 5.4)

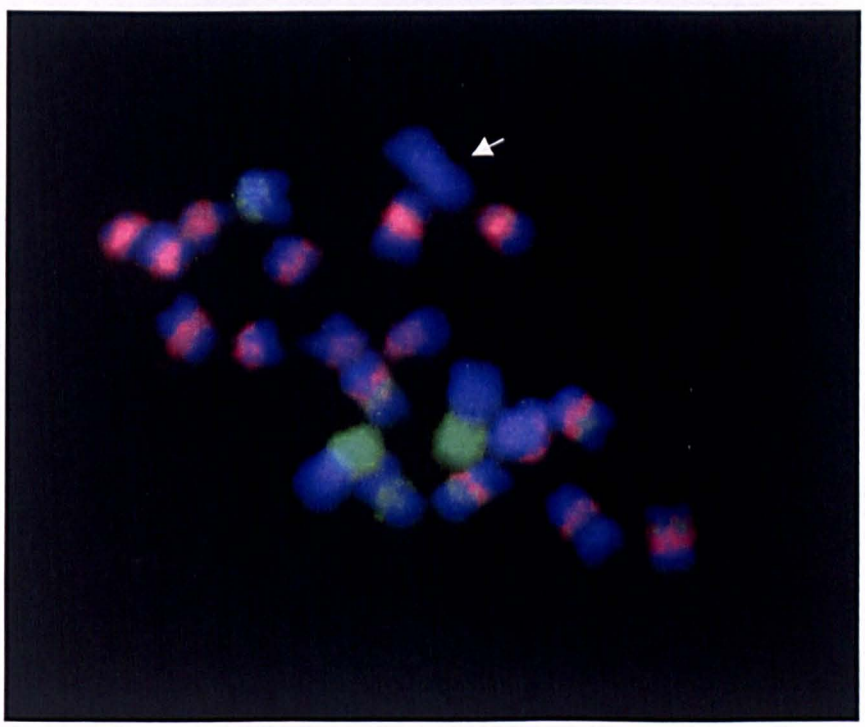


Figure 5.4: FISH Image of mitotic metaphase of addition line BCA4 showing 20+1 chromosomes. Green = rDNA probe, Orange = *B. oleracea* BAC BoB02C24 containing sub-centromeric sequences. The *B. oleracea* single chromosome is marked by an arrow.

To establish whether this extra chromosome represented linkage group O3, the PCR amplification initially utilised to map *BoCAL-a* was tested against the addition lines (CAL4F-CALI6R, fig. 5.1). Initial amplification attempts from *B. rapa* demonstrated that this assay was also *B. oleracea* specific. This suggested that presence of an amplicon of the correct size from the addition lines would demonstrate presence of *BoCAL-a* and thus of LGO3. When applied to the segregating plants of addition line BCA4, products were observed from the white flowered plants only (fig. 5.5).

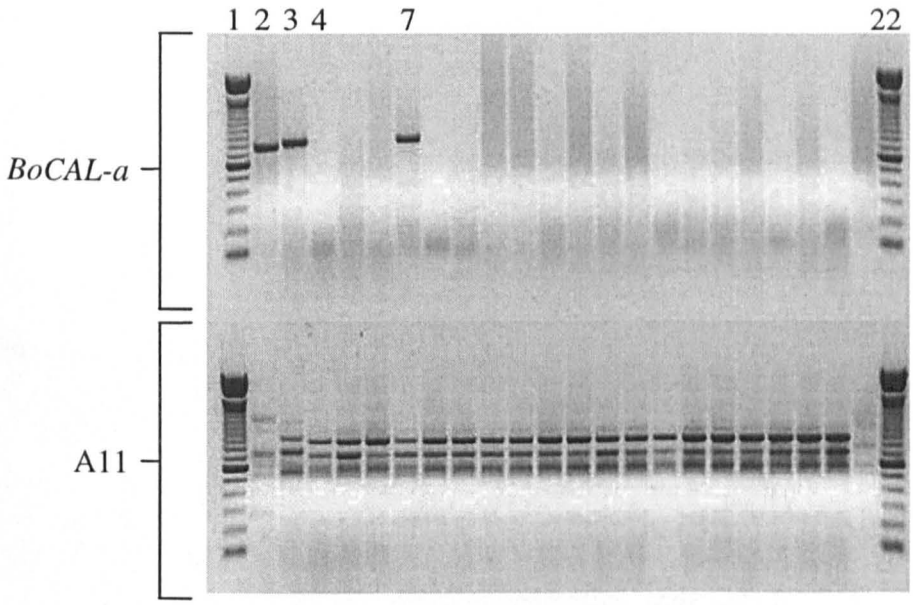


Figure 5.5: PCR amplification of *BoCAL-a* from the segregating addition line BCA4. The RAPD primer A11 was used as a positive control of DNA integrity throughout the lines. Lane 1 + 22 = marker. Lane 2 = A12 positive control (*B. oleracea*). Lane 3 = Donor *B. oleracea*. Lane 4 = *B. rapa* parent. Lane 7 = White flowered BCA4 line. All other lanes = yellow flowered BCA4 lines.

This evidence strongly suggested that the extra *B. oleracea* chromosome in the line BCA4 would correspond to linkage group 03 from the integrated A12DH x GD33 and NxG genetic map.

5.4.3 SECTION 3: Cytological examination of *BoCAL-a* copy number

The failure to identify and map further copies of *BoCAL* through time consuming RFLP analysis suggested a different approach would have to be taken. To address the question of likely copy number and distribution of *BoCAL* loci in *B. oleracea* it was decided to conduct Fluorescent *In Situ* Hybridisation (FISH) of a *BoCAL* genomic probe onto chromosome spreads from the Chinese White Kale A12DH. As well as providing an estimate of the likely copy number and distribution of *BoCAL*, this analysis would also examine the feasibility of future FISH experiments to examine the physical distance between *BoAP1-a* and *BoAP1-c*.

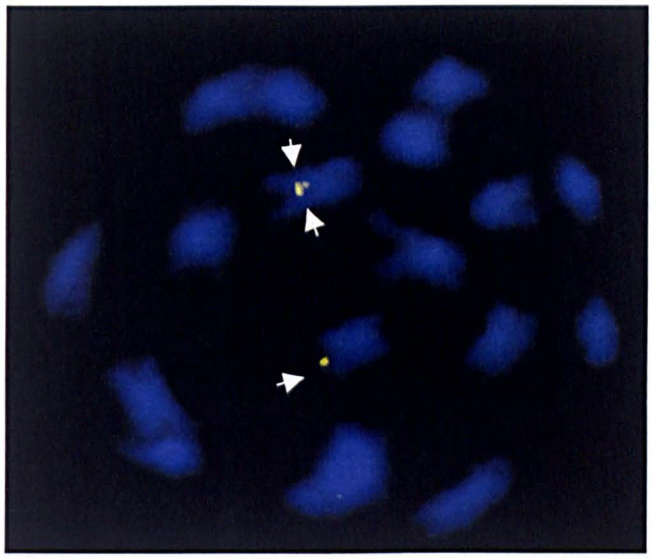
5.4.3.1 Probe construction

The 15Kb *BoCAL-a* genomic clone SK162 (S Kempin Pers. Comm.) which contains the entire genomic sequence of *BoCAL* plus 10Kb of 5-prime sequence was labelled with Dioxygenin (DIG), linked to a fluorescent antibody. It was hoped that the large tract of upstream sequence would confer a level of specificity above that provided by the gene sequence alone.

5.4.3.2 Fluorescent *In Situ* Hybridisation

Hybridisation to A12DH chromosomes at both pachytene (meiotic) and metaphase I (mitotic) highlighted two distinct regions of the genome that strongly hybridised the Genomic *BoCAL-a* clone. Further examination across ten different pachytene spreads suggested that one of these two regions appears to contain two closely linked hybridising bands (fig. 5.6).

(a)



(b)

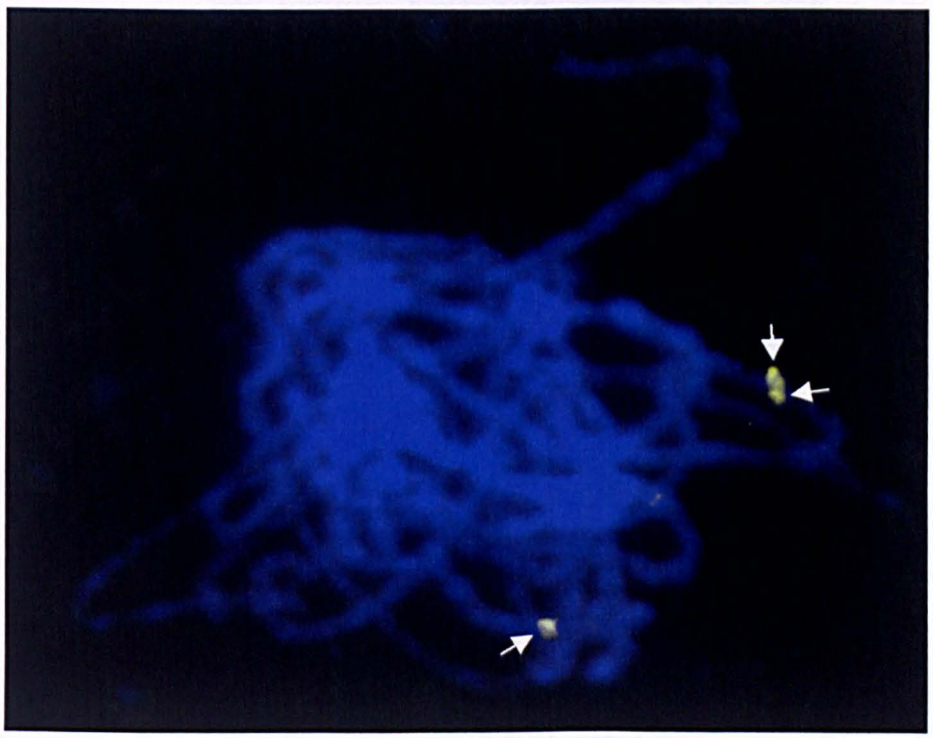


Figure 5.6: (a) Mitotic metaphase and (b) meiotic pachytene A12DH chromosomes showing hybridisation of the *BoCAL-a* genomic clone. Three discrete regions hybridise the probe strongly (arrowed) suggesting multiple copies of *BoCAL* in *B. oleracea.*

This analysis suggested that the likely copy number of *BoCAL* loci in *B. oleracea* is at least three, two of which are physically linked. Unfortunately, RFLP analysis utilising SK162 as a probe revealed no polymorphic bands in the NxB population.

5.5 Discussion

5.5.1 Genetic map localisation of *BoCAL-a*

The identification of the map position of *BoCAL-a* on LGO3 in *B. oleracea* may be of significance. In *A. thaliana* the homologue of *BoCAL*, *CAULIFLOWER*, is situated on chromosome 1, although the *A. thaliana* homologue of a marker close to *BoCAL-a*, *BoHSPLocus1* (fig. 5.1), maps to chromosome 3. This suggests that a breakpoint in synteny may be situated between *BoCAL-a* and *BoHSPLocus1* on linkage group O3. Identification of a breakpoint region suggests that characterisation of this region in both *B. oleracea* and *A. thaliana* could provide information relating to the genome evolution of *B. oleracea* from the *A. thaliana/Brassica* common ancestor.

5.5.2 Linking the genetic and cytological maps of *B. oleracea*

Preliminary analysis of the *B. oleracea* var. *alboglabra* karyotype has utilised the 45S and 5S rDNA sequences to positively identify chromosomes. So far this has allowed positive identification of three chromosomes, whilst a further two are identifiable with some confidence (Armstrong *et al.*, 1998). Analysis of the addition line BCA4 against both phenotypic (flower colour) and molecular (*BoCAL-a*) markers derived from linkage group O3 has strongly suggested that Linkage group O3 is likely to correspond to cytological chromosome 4. This represents the first example of an association between the genetic and cytological maps of *B. oleracea*, and suggests examination of further markers in other chromosome addition lines may provide a valuable resource for direct physical mapping in *B. oleracea.*

5.5.3 Determination of Locus copy number through cytological analysis.

Cytological analysis, and in particular FISH analysis has the potential to identify the genomic position of loci regardless of recombination events. This characteristic has been utilised during the analysis of *BoCAL*, and demonstrated that although only a

single copy could be identified and mapped by the molecular genetic means employed, there are likely to be at least two further copies.

As the relationship between genetic and cytological maps is further elaborated, the potential exists for identifying to which linkage group specific FISH hybridisations bind. A good example of this is the identification of two possible linked loci of *BoCAL*, which raises questions regarding the previous identification of two closely linked *AP1/CAL* hybridising RFLP bands on LGO6. Especially given that comparisons of RFLP and CAPS data suggests that neither of these RFLP loci represent *BoAP1-c*. Further analysis provided by a dual probing of *BoCAL* and *BoAP1* clones labelled with different florescence tags would immediately address whether copies of *BoCAL* and *BoAP1* are indeed linked. Furthermore, probing with *BoAP1* would help characterise genomic distribution of *BoAP1* loci, including the so far unmapped *BoAP1-b*, and would also identify the relationship between genetic and physical distance within this region of the genome.

The ability to investigate physical distance is an important requirement for any map based cloning attempt. With a *Brassica* genome sequencing project some years away, a powerful means of establishing physical distance and marker order is provided by FISH experiments using known molecular probes.

With the development of Fibre-FISH for *Brassica* chromosomes (Jackson *et al.*,1998), high resolution ordering of closely flanking markers to a region of interest is now possible. This may allow rapid mapping and increase the speed at which genes of interest may be cloned, by aiding *in situ* BAC contig construction.

5.5.4 Conclusions

1. *BoCAL-a* maps to linkage group O3 in *B. oleracea*
2. Addition line analysis suggests that linkage group O3 corresponds to cytological chromosome 4

3. FISH analysis demonstrates that a *BoCAL* genomic clone hybridises to three distinct loci, two of which are closely linked, suggesting that *BoCAL* is present in multiple copies in *B. oleracea.*
4. Although at a preliminary stage, this approach could be applied to physically map *BoAP1-a* and *BoAP1-c*, and to identify the physical and genetic map position of *BoAP1-b,* this avoids difficulties associated with time consuming conventional RFLP analysis.

CHAPTER 6: THE DOMESTICATION OF THE CAULIFLOWER

6.1 Introduction

The underlying genetic relatedness within landrace brassicas is more dependent on geographic proximity of populations rather than on the phenotypic similarity of the few characters artificially selected by man (Dias *et al.*, 1992). Examples of this have been observed in both wild (Lanner-Herrera *et al.*, 1996) and cultivated *B. oleracea* (Dias *et al.*, 1992) where RFLP data showed that the only correlation between the limits of gene flow and the development of distinct landraces was geographical isolation of populations (Dias *et al.*, 1992).

Genetic relatedness based on geographical proximity is probably aided by the breeding strategy of diploid brassicas which have evolved to become facultative outbreeders. In *B. oleracea,* a strong sporophytic self incompatibility system is defined by the allelic composition at a single locus, 'S' (De Nettancourt, 1977). This breeding strategy has been shown to result in relatively heterogeneous wild and landrace populations (Massie, 1998), which suggests that retention and development of specific crop traits with a lower selective advantage are heavily dependent on selection by man (Crisp, 1982).

Divergence due to geographical isolation and artificial selection has been proposed to account for the distinct morphologies of regional types of cauliflower within Italy (Massie, 1998) (Fig. 1.1). It appears that the spread of genetic information between populations has been inhibited for the most part by eco-geographic and political barriers, rather than by physiological means (Massie, 1998).

The identification of a mutation in the putative contributing locus *BoCAL-a* in a single population (NxB) raises the question of involvement of this allele throughout the many distinct cauliflower morphotypes (and by definition its use as a molecular marker). A robust method to address the potential involvement of *BoCAL-a* in curd development could be provided by examination of the distribution of putative causal alleles through the *B. oleracea* genepool. Heavy selection of specific alleles in spite of

so many distinct recombination events would be a good indicator of involvement.

Analysis of the distribution of such alleles may also help illuminate the processes underlying the domestication of cauliflower, independent of questionable historical sources (1.13).

6.2 Objectives

1. Determine whether the E151→Stop mutation previously detected in *BobCAL* has utility as a molecular marker for the curding habit across the entire *B. oleracea* genepool.
2. Depending on the utility of the marker, identify and develop molecular assays for further putative causal alleles.
3. Utilise the involvement of *BoCAL-a* in the propensity to curd suggested by the genetic model developed in chapter 3 to trace the possible domestication of cauliflowers through molecular genetic analysis.

6.3 Strategy

Genetic linkage of the *BobCAL* allele of *BoCAL-a* to curding (chapter 3) suggested this would be a good marker for the curding trait of cauliflowers. However, if this marker was to have widespread utility it was important to determine whether this mutation was involved in generic curd development in all regional morphotypes.

A 'diversity collection' of two hundred and nine accessions representing eight different varietas of *B. oleracea* selectively bred for crop production was grown and examined. This included nine distinct groups of *B. oleracea* var. *botrytis* and var. *italica* accessions shown to display curding features. The majority of accessions were specifically chosen to maximise eco-geographic diversity within the *B. oleracea* gene pool (Boukema *et al.*, 1997), with thirty five commercial cultivars included for comparison.

The genetic diversity within this wide-ranging survey enabled rapid screening of a

very large number of historical recombination events, which would have occurred both prior to and during domestication over the last few thousand years (fig. 6.1). This allowed testing of the hypothesis that the E151→Stop mutation described in the *BobCAL-a* allele (Kempin *et al.*, 1995) was indeed diagnostic of all cauliflowers.

Seventy six of the accessions representing *B. oleracea* vars. *botrytis* and *italica* were specifically chosen as they had previously been subject to detailed taxonometric study of 40 curd and leaf traits (Massie *et al.*, 1996), which confirmed eco-geographic groupings and relationships across the centre of diversity, in what is now Italy. It was hoped that direct comparison of the molecular data arising from analysis of the putative causal locus *BoCAL-a* to those results would provide insights into how the curd trait was domesticated.

6.4 Results

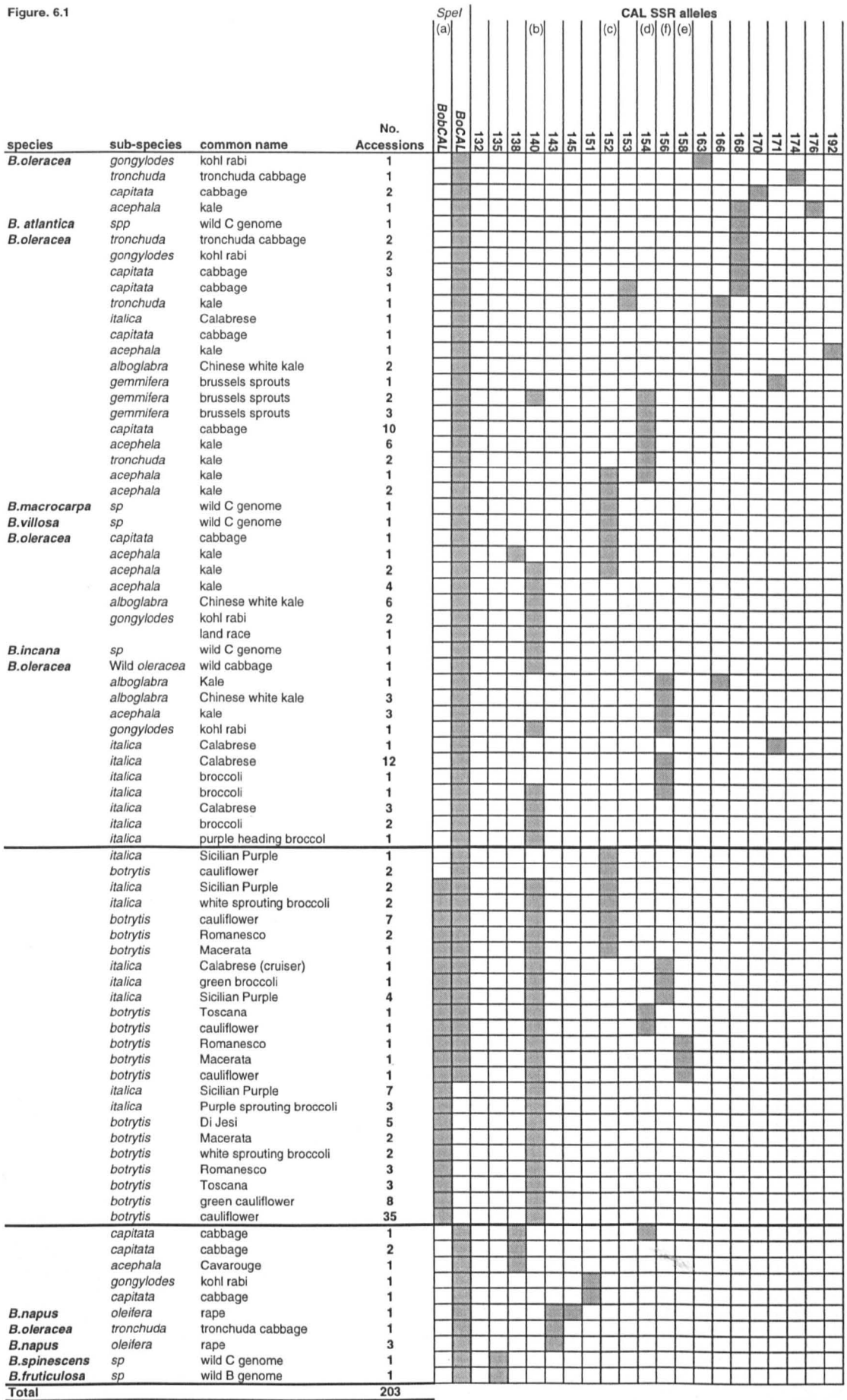
6.4.1 Distribution of the *BobCAL* allele amongst the domesticated *B. oleracea* gene pool

Molecular genetic analysis of the *BobCAL* allele utilised the molecular assay described during the analysis of the NxB population (3.4.3.1.1). Molecular results were compared to phenotypic scores taken in the field to ascertain the distribution of the premature termination mutation of *BobCAL* through the domesticated *B.oleracea* gene pool.

The distribution of this allele was striking (fig. 6.1). The *BobCAL* allele is present at extremely high frequency in cauliflowers, Sicilian Purple cauliflowers, and Purple and White Sprouting Broccoli (94 accessions), yet is completely absent in representatives of all other brassicas tested from the centre of diversity. A very high proportion of these accessions are homozygous for the *BobCAL-a* allele, supporting the original suggestion that this allele is involved in the propensity to curd. However, nineteen curding accessions were found to be heterozygous and two lacked the allele altogether. This was confirmed by extracting DNA directly from curd tissue and repeating the assays on several occasions. In one case of a white cauliflower, examination of a progenitor breeding line, originally selected in the late 1950's (Doug

Blazey Pers. comm.) was utilised to confirm inheritance of a novel allele.

Figure. 6.1



species	sub-species	common name	No. Accessions
B.oleracea	*gongylodes*	kohl rabi	1
	tronchuda	tronchuda cabbage	1
	capitata	cabbage	2
	acephala	kale	1
B. atlantica	*spp*	wild C genome	1
B.oleracea	*tronchuda*	tronchuda cabbage	2
	gongylodes	kohl rabi	2
	capitata	cabbage	3
	capitata	cabbage	1
	tronchuda	kale	1
	italica	Calabrese	1
	capitata	cabbage	1
	acephala	kale	1
	alboglabra	Chinese white kale	2
	gemmifera	brussels sprouts	1
	gemmifera	brussels sprouts	2
	gemmifera	brussels sprouts	3
	capitata	cabbage	10
	acephela	kale	6
	tronchuda	kale	2
	acephala	kale	1
	acephala	kale	2
B.macrocarpa	*sp*	wild C genome	1
B.villosa	*sp*	wild C genome	1
B.oleracea	*capitata*	cabbage	1
	acephala	kale	1
	acephala	kale	2
	acephala	kale	4
	alboglabra	Chinese white kale	6
	gongylodes	kohl rabi	2
		land race	1
B.incana	*sp*	wild C genome	1
B.oleracea	Wild *oleracea*	wild cabbage	1
	alboglabra	Kale	1
	alboglabra	Chinese white kale	3
	acephala	kale	3
	gongylodes	kohl rabi	1
	italica	Calabrese	1
	italica	Calabrese	12
	italica	broccoli	1
	italica	broccoli	1
	italica	Calabrese	3
	italica	broccoli	2
	italica	purple heading broccol	1
	italica	Sicilian Purple	1
	botrytis	cauliflower	2
	italica	Sicilian Purple	2
	italica	white sprouting broccoli	2
	botrytis	cauliflower	7
	botrytis	Romanesco	2
	botrytis	Macerata	1
	italica	Calabrese (cruiser)	1
	italica	green broccoli	1
	italica	Sicilian Purple	4
	botrytis	Toscana	1
	botrytis	cauliflower	1
	botrytis	Romanesco	1
	botrytis	Macerata	1
	botrytis	cauliflower	1
	italica	Sicilian Purple	7
	italica	Purple sprouting broccoli	3
	botrytis	Di Jesi	5
	botrytis	Macerata	2
	botrytis	white sprouting broccoli	2
	botrytis	Romanesco	3
	botrytis	Toscana	3
	botrytis	green cauliflower	8
	botrytis	cauliflower	35
	capitata	cabbage	1
	capitata	cabbage	2
	acephala	Cavarouge	1
	gongylodes	kohl rabi	1
	capitata	cabbage	1
B.napus	*oleifera*	rape	1
B.oleracea	*tronchuda*	tronchuda cabbage	1
B.napus	*oleifera*	rape	3
B.spinescens	*sp*	wild C genome	1
B.fruticulosa	*sp*	wild B genome	1
Total			**203**

Figure 6.1: Distribution of the *BobCAL* and SSR alleles (see below) in *BoCAL-a* through the *B. oleracea* and associated gene-pools. **(a)** The *BobCAL* allele is present at high frequency in accessions with curd or 'intermediate' inflorescence types in accessions from the centre of diversity, suggesting possible selection at this locus. **(b)** *BobCAL* is completely linked to allele *SSR140.* The widespread distribution of this allele suggests SSR allelic states pre-date the premature stop mutation in *BobCAL.* **(c)** *SSR152* is present in fifteen percent of curding and intermediate crop types, demonstrating the stop mutation in *BobCAL* is not diagnostic of all cauliflowers. Two other alleles, *SSR154* **(d)** and *SSR158* **(e)** are present at extremely low frequency in cauliflowers and constitute the allelic state for the remaining curding plants. The extremely low frequency of these alleles suggests they may be relatively recent introductions into the cauliflower genepool, or alternatively merely represent recent mutations of the SSR class whilst retaining the previous contributing mutations. **(f)** Allele *SSR156* is shared by Sicilian Purples and Calabrese. Reported transient arrest stages in Calabrese floral induction (Kieffer *et al.*, 1998) suggest this may represent a weak mutant allele of *BoCAL-a.*

Presence of the *BobCAL* allele in a homozygous state in Purple Sprouting broccoli (fig. 6.1) (which is not curding) suggested that either the penetrance of this mutation is complemented by a second locus, or else the *BoCAL-a* gene is not involved in curd development, and may just be linked to the contributing gene. To address this possibility, the cauliflower DJ7032 was crossed onto a Doubled Haploid Chinese White Kale (A12DH) (which is non-curding). Analysis of both the F_1 and 120 F_2 lines revealed no curding inflorescence types and just a single intermediate type (where 45 intermediates were expected; 9:3:3:1, Wild-type:intermediate:intermediate:curding) (data not shown). This suggested that non-heading brassicas may contain modifying genes that can completely complement the loci that contribute to curding in the heading brassicas. These complementing loci could conceivably represent further copies of *BoCAL* and *BoAP1.* This evidence suggests that further investigation is required if we are to widen the genetic model described for cauliflower and Calabrese to include non-heading brassicas.

6.4.2 A Simple Sequence Repeat (SSR) diagnostic of the *BoCAL-a* locus.

A SSR repeat was identified in intron seven of the *BoCAL-a* gene. This SSR was discovered during routine scanning of the *BoCAL-a* locus by PCR amplification when searching for markers to facilitate genetic mapping. Amplification between the PCR primers LSCALF1 and BoB2 resulted in amplicons ranging from approximately 430 bp in the DH cauliflower DJ7032 to 460 bp in the Chinese white kale A12DH.

Linkage of this marker to the putative causal *BoCAL-a* locus was confirmed by genetic segregation data from the NxB DH population (no recombinants were observed between the SSR allele from the cauliflower parent (DJ7032) and the *BobCAL* allele in a comparison of 49 DH progeny lines) and DNA sequencing of a cloned genomic fragment (SK162, chapter 5).

6.4.2.1 Polyacrylamide gel electrophoresis (P.A.G.E) analysis

^{33}P-dATP was used to end label the primer LSCALF1 prior to PCR amplification. P.A.G.E. analysis of all the mapping population progenitor lines and F1 hybrid parents demonstrated the potential utility of this SSR as an allele specific marker of the *BoCAL-a* locus (fig. 6.2).

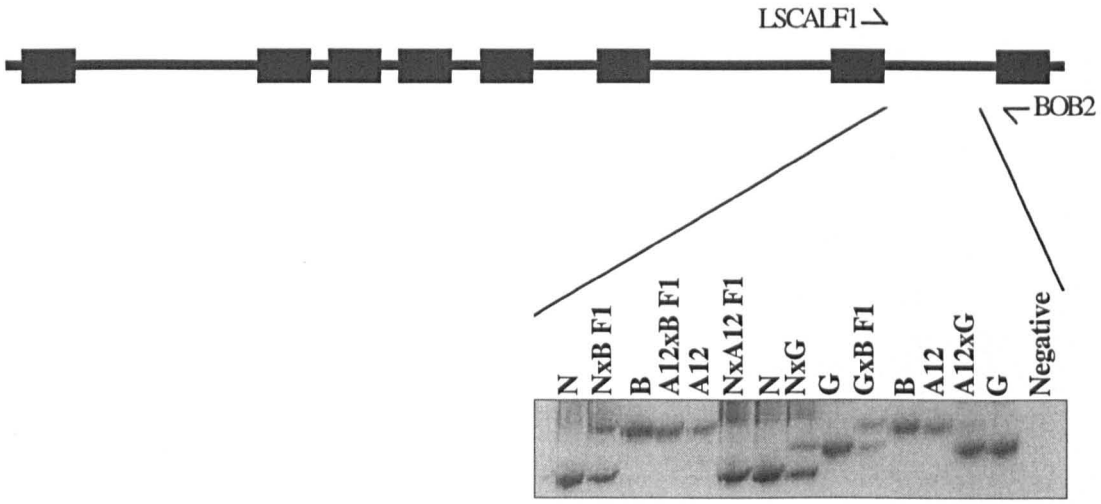


Figure 6.2: Schematic showing allele length variation in the SSR from intron seven of the *BoCAL-a* locus, amongst progenitor lines and F1 parents of the *B. oleracea* mapping populations. A12=DH Chinese white kale, GD33=DH Calabrese, N=DJ7032 DH cauliflower, B=BI88908 recurrent inbred broccoli, G=DJ3753 DH Brussels sprout.

6.4.2.2 Genescan analysis

Examination of SSR allele length polymorphisms through the *B. oleracea* genepool was conducted on an ABI377 DNA sequencer. This system has an advantage over original P.A.G.E. analysis in that an internal standard is added to each sample prior to

electrophoresis, thus increasing resolution, reproducibility and accuracy of comparisons.

6.4.2.3 Marker Design

The variable region in the LSCALF1-BOB2 amplicon was identified following DNA sequencing. This allowed the design of PCR primers (CALSSRF-CALSSRR) much closer to the informative section. PCR amplification between these primers gave a much smaller fragment, approximately 140-165 base pairs, which allowed more accurate resolution of allele length than the original 430bp fragment. (Fig. 6.3).

Figure 6.3: Diagrammatic representation of SSR allele variation and PCR primers. Consensus sequence is shown along with the variable region from a cauliflower (DJ7032) and a Chinese White Kale (A12DH), top and bottom strands respectively. CALSSRF is fluorescently labelled with 6-fam prior to PCR amplification resulting in detectable allele lengths of 140bp and 163bp respectively, upon electrophoresis on an ABI377 DNA sequencer.

6.4.2.4 Statistical analysis of accuracy

Amplification products from the cauliflower DJ7032 were included in each gel run as a positive control of allele identification. Comparison of allele length in DJ7032 over all gel runs resulted in an estimated average size of 141.33bp with a standard deviation of 0.44 (12 samples tested). This correlated well to the 141bp size expected from DNA sequencing. This demonstrated that allele length could be detected consistently within one base-pair despite possible gel to gel inconsistencies, and thus that grouping of accessions based on allele length would be likely to be consistent

across gels.

6.4.3 SSR Results

6.4.3.1 BobCAL is linked to SSR140

The marker was scored on the diversity collection of two hundred and three accessions and resolved twenty SSR alleles (fig. 6.1). The *BobCAL* allele was found to consistently cosegregate with *SSR140* such that presence of the *Spe*I site denoted presence of *SSR140*. This was also the case in the specific clone of *BobCAL* first described by Kempin *et al.* (1995). *SSR140* was found to be the most common allele detected, distributed throughout the majority of *B. oleracea* crop types analysed. This suggests that the SSR allele was established in the gene pool before the E151→Stop mutation in *BobCAL* arose. The alternative hypothesis is that the *SSR140* allele arose independently on a number of occasions, which is not inconceivable in such a diverse survey.

6.4.3.2 Detection of additional alleles of *BoCAL* associated with the curding phenotype.

The strong co-segregation of *BobCAL* with a specific SSR class, along with the relatively recent development of curding crops, (ca. 1500 onwards, (Thompson, 1976)), suggested that despite possible parallelism of SSR classes, it would not be unreasonable to utilise the SSR allelic diversity as an allele specific marker of *BoCAL*, within the relatively limited genepool of the heading brassicas. The SSR marker was scored on the diversity collection of two hundred and three accessions and resolved a further nineteen SSR/*BoCAL* alleles (fig. 6.1). Three of these alleles, *SSR152, SSR154* and *SSR158* were shown to be associated with curding within the collection (fig. 6.1).

The most common allele amongst the broccoli/Calabrese accessions, *SSR156*, was absent from the collection of cauliflower accessions (fig. 6.1). However, Sicilian Purple accessions *were* shown to contain this allele, as well as containing both the *BobCAL* and *SSR152* alleles common to curding accessions, but absent from Calabrese. Thus Sicilian Purples were found to share mutually exclusive alleles with both cauliflower and Calabrese, raising the possibility that these crops represent an

intermediate morphotype between the two extremes..

6.4.3.3 Morphotype and Geographic distribution of SSR alleles

A comparison of all heading *Brassica* accessions (Calabrese, Sicilian Purples and Cauliflower) arising specifically from Italy provides insights into the possible genetic interactions during cauliflower domestication (fig. 6.4).

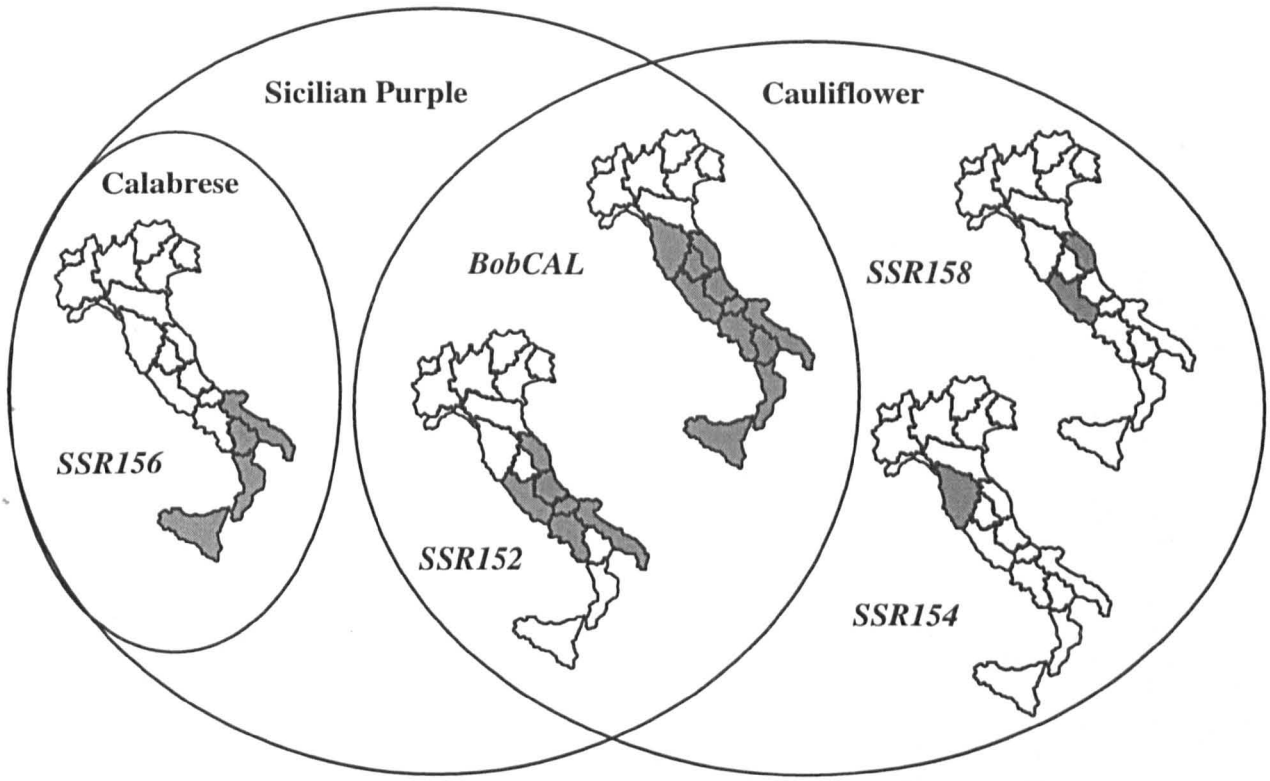


Figure 6.4: Crop type and geographical distribution of *BoCAL-a* alleles amongst the heading brassicas of Italy and Sicily. Five alleles have been identified as significant markers of specific head morphologies. Allele distribution both geographically and through the differing morphotypes supports the hypothesis that Sicilian Purples represent an intermediate stage between cauliflower and Calabrese (fig. 6.5). Sicilian Purples are always observed with at least one of either *BobCAL* or *SSR152* alleles at the *BoCAL-a* locus, suggesting plants possessing only allele *SSR156* are unable to form even crude curds, and may in fact represent purple Calabrese, in agreement with distinctions made during taxonometric studies (Massie, 1998).

The *BobCAL* allele was found in accessions collected from throughout mainland Italy

and Sicily, and is involved in the genotype of all Sicilian Purples tested from Sicily and Calabria. This allele is completely absent from all Calabrese accessions tested from the centre of diversity (which is surprising given its recessive nature). Sicilian Purples also share an allele at high frequency with Calabrese (*SSR156*) which is excluded from all curding accessions tested, suggesting Sicilian Purples represent an intermediate form between Calabrese and cauliflower (fig. 6.4, 6.5). In discrete regions of southern Italy, Sicilian Purples are found possessing the *BoCAL* allele characterised by *SSR152* which is otherwise limited to cauliflowers, and is linked in repulsion to the mutation in *BobCAL*. This characterises a possible secondary centre of hybridisation between cauliflowers and Calabrese/Sicilian Purples in this area. The limited radiation of the remaining two alleles, *SSR154* and *SSR158*, linked to curding, could suggest that mutations in these *BoCAL* alleles may be recent introductions into the cauliflower gene-pool. Alternatively the SSR allelic state may have mutated whilst retaining the previous contributing mutation.

6.4.4.4 Commercial cultivars

Thirty five accessions were omitted from this geographical analysis as they represent modern commercial cultivars of cauliflower, Calabrese and Sicilian Purples. Analysis of these accessions followed closely the distribution of specific alleles of *BoCAL-a* to specific crop types. This provides further evidence for causal involvement of these alleles in heading morphotypes, demonstrating they are heavily selected by modern commercial breeders, as well as the regional Italian growers. Interestingly, the commercial Calabrese 'Cruiser' was heterozygous for *BobCAL*, suggesting the presence of cauliflower in its recent ancestry.

6.5 Discussion

6.5.1 The utility of *BobCAL* as a molecular marker of curding

Involvement of *BoCAL-a* in the propensity to curd suggested by the genetic model developed in chapter 3 has been substantiated by an extensive survey of *ex situ* crop genetic resources (fig. 6.1). This demonstrates that the E151→Stop mutation in *BobCAL-a* (Kempin *et al.*, 1995) is present at high frequency, accounting for the genotype of ninety percent of all curding accessions tested at the *BoCAL-a* locus. This

allele is completely absent from varietas *capitata, acephala, gongylodes, alboglabra, tronchuda,* or *gemmifera.* The high frequency of the *BobCAL-a* allele in both landrace and commercial cultivars, together with the genetic segregation data presented in chapter three, presents a powerful argument in favour of causal involvement of this allele in curding.

6.5.2 Identification of further putative causal alleles

Three further alleles of *BoCAL* appear to be associated with the developmental arrest during curd formation, none of which possess the premature termination codon associated with the *BobCAL* allele (Kempin *et al.*, 1995). This evidence demonstrates that the *BobCAL* allele (Kempin *et al.*, 1995) is not diagnostic of all cauliflowers (fig. 6.1).

6.5.3 Morphotype and geographical analysis of *BoCAL-a* allele distribution

Comparison of morphotype and geographic distribution of SSR alleles of all heading *Brassica* accessions (Calabrese, Sicilian Purples and Cauliflower) collected specifically from Italy (86 accessions), provides insights into the genetic interactions taking place during cauliflower domestication (fig. 6.4). The introgression of *SSR152* into Sicilian Purples is key in demonstrating that the spread of this allele at least is not inhibited by climatic factors. This suggests that limited allele distribution between crop types, particularly allele exclusion from geographically overlapping populations, is probably largely the result of artificial selection for specific phenotypes. Thus specific phenotypes are unable to develop when certain *BoCAL* alleles are in a homozygous state.

Evidence for such a delineation is provided by examination of the mixed Sicilian Purple/Calabrese population. Segregation of *BobCAL* amongst the integrated Calabrese and Sicilian Purples appears to define discrete classes of inflorescence phenotype amongst the relatively heterogeneous populations across southern mainland Italy and Sicily (fig. 6.4). This is in agreement with distinctions made during detailed taxonometric studies (Massie *et al.*, 1996) and highlights the significant role specific alleles of *BoCAL* may play in morphotype delineation.

The high frequency of allele *SSR156* in Calabrese suggests that this allele could be under selection. Transient curding stages reported in Calabrese (Kieffer *et al.*, 1998) would be consistent with the action of a weak allele of a gene such as *BoCAL*. A brief arrest and proliferation of inflorescence meristems would result in a greater number of inflorescence and flowers developing, the key characteristics of Calabrese morphology. Alternatively this could simply be indicative of a reduced gene-pool amongst heading brassicas. The absence of the *BoCAL* allele defined by *SSR156* from cauliflowers, is probably wholly the result of its inability to direct curd development, thus cauliflowers do not develop if this allele is present.

6.5.4 The domestication of the cauliflower

The combination of genetic, phenotypic and molecular information, combined with knowledge of eco-geographic distribution of alleles allows construction and proposition of a model which would account for some of the events leading to the domestication of cauliflowers (Fig. 6.5).

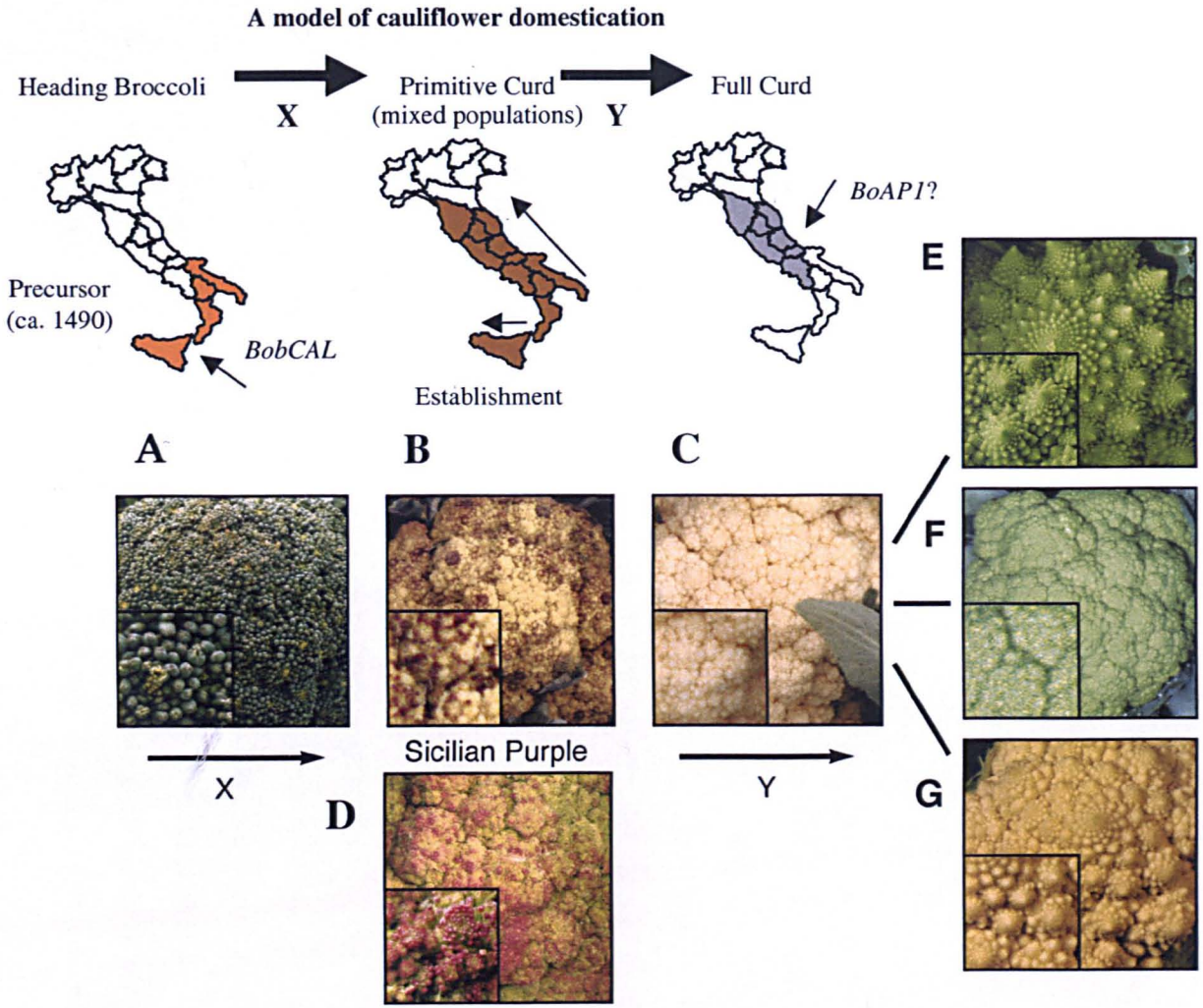


Figure 6.5: A theoretical 'two-step' model of cauliflower domestication. **(X)** The introduction of a 'mutant' allele of either the *BoCAL-a* or *BoAP1-a* loci in heading broccoli type (characterised by Calabrese in this instance (a)) results in a 'ricey' curding population phenotypically similar to both the Sicilian Purple types (d) (shown here at the onset of anthocyanin expression) and the intermediates of the NxB DH population (b). The date proposed for this introduction is ca. 1490 (Thompson, 1976). A period of establishment follows with primitive curding plants segregating amongst heterogeneous populations of heading types. **(Y)** Introduction of a further mutation into the second of the two loci results in stronger control over floral induction and development of the classic curd phenotype (c) , followed by proliferation into several distinct forms including Romanesco (e), Macerata (f), and Di Jesi (g).

The intermediate phenotypes from the DH cauliflower x Calabrese population resemble closely the weakened arrest state observed in the Sicilian Purples (chapter 3) and is in agreement with earlier breeding studies (Branca and Iapichino, 1997; Gray, 1989). An historic scenario could therefore be proposed whereby crops such as this were valued as a primitive curding type, and selected following unwitting introduction of a mutant *BoCAL-a* or *BoAP1-a* allele into the genepool of basic heading plants (fig. 6.5). A period of establishment would follow during which selection for this "new" crop encompassed local and seasonal adaptations (fig. 6.5). A more complete arrest of floral induction would arise following the introduction of a 'mutant' copy of the second locus, either *BoCAL-a* or *BoAP1-a* resulting in full curd development. This two step process leading to increased modulation in control of floral meristem induction is consistent with the model of two related genes with additive effects described for *CAL* and *AP1* in *Arabidopsis* (Bowman *et al.*, 1993), and is also supported by the strong genotype-morphotype correlation amongst the integrated populations of Southern Italy and Sicily.

A stepwise progression towards a more refined crop is consistent with classical breeding techniques, where selection for desirable traits is based on selecting the optimal progeny from any cross. The hypothesis proposed does not therefore require distinct mutational events to have arisen recently in order for development of the curding habit. Rather it relies on bringing together, either by intentional or serendipitous combination, a group of alleles already present in the *B. oleracea* genepool.

6.5.5 Conclusions

1. *BobCAL-a* allele is associated with the majority of curding accessions.
2. The E151→Stop mutation in *BobCAL-a* is not diagnostic of all cauliflowers.
3. In total, four alleles of *BoCAL-a* have been identified as associated with curding.
4. The SSR assay has utility as a molecular diagnostic tool of the *BoCAL-a* locus.
5. Geographic distribution of alleles suggests curding is probably monophyletic with multiple introgression of further causal alleles during domestication of the cauliflower.

6. Morphotype distribution of alleles suggests Sicilian Purples are intermediates in the domestication of cauliflower from Calabrese.

CHAPTER 7: FUSED INFLORESCENCE, A CHARACTER INVOLVED IN BOTH CAULIFLOWER AND CALABRESE INFLORESCENCE STRUCTURE.

7.1 Introduction

Until relatively recently the taxonomic distinction between *B. oleracea* vars. *botrytis* and *italica* was based largely on the heading verses sprouting habit (Synge, 1956). This trait divides the curding white sprouting broccolis from the remainder of curding plants in spite of similar ontological arrest stages. The importance placed upon heading both as a taxonomic character and as a defining feature of several *Brassica* crops, suggests a genetic dissection of contributing traits would be of interest to *Brassica* breeders.

Visual examination of many accessions during the diversity trial (chapter 6) suggested that the physical distinction between heading and sprouting types can largely be summarised as a combination of the level of control over ontological arrest, and inflorescence architecture. During analysis of inflorescence morphology, a trait was identified that appeared to contribute to the overall 'heading' nature of specific accessions. This trait was characterised as "a fusion of secondary inflorescences to the primary inflorescence above the subtending cauline leaf". The net result of this fusion was to draw together several inflorescence spikes into a single or few, dense 'heads' of buds or curd (fig. 7.1). This character was found to segregate amongst three populations, suggesting the possibility of genetic control.

7.2 Objectives

1. Identify genetic map positions for loci likely to be responsible for the Fused Inflorescence trait of heading *Brassica oleracea,* in preparation for future investigation.

7.3 Strategy

Three segregating DH populations were scored for presence or absence of Fused Inflorescence. The resulting data from two of these populations were used in genetic

linkage analysis to identify loci which may contain genes that contribute to the Fused Inflorescence phenotype. The third population, NxB (cauliflower x Calabrese broccoli) was examined to test whether the similar Fused Inflorescence phenotype of each of these crops has the same genetic basis.

7.4 Results

7.4.1 Scoring phenotypic variation.

The fused inflorescence character associated with heading brassicas results from the fusion of secondary inflorescence spikes to the primary inflorescence, above the subtending cauline leaf (Fig. 7.1). This helps to compact many inflorescences into a single, dense head structure (Fig. 7.2).

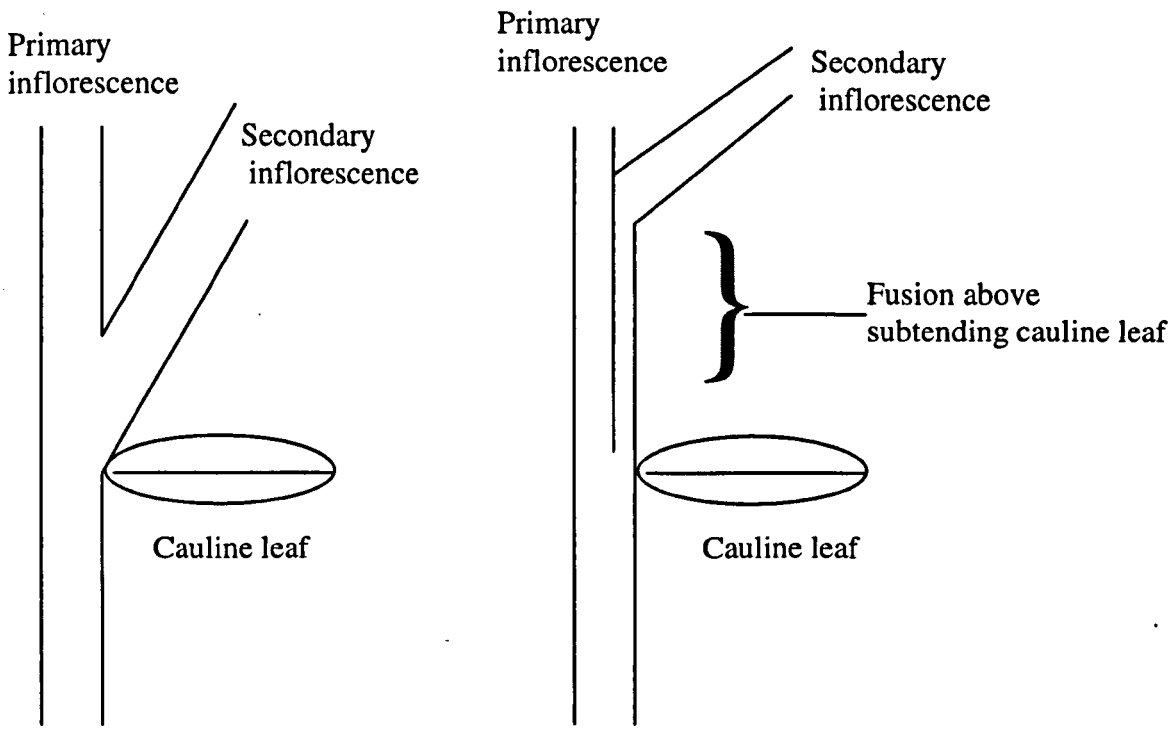


Figure 7.1: Schematic representation of the distinction between fused and unfused inflorescence scored in the A12DHxGD33, NxG and the NxB populations.

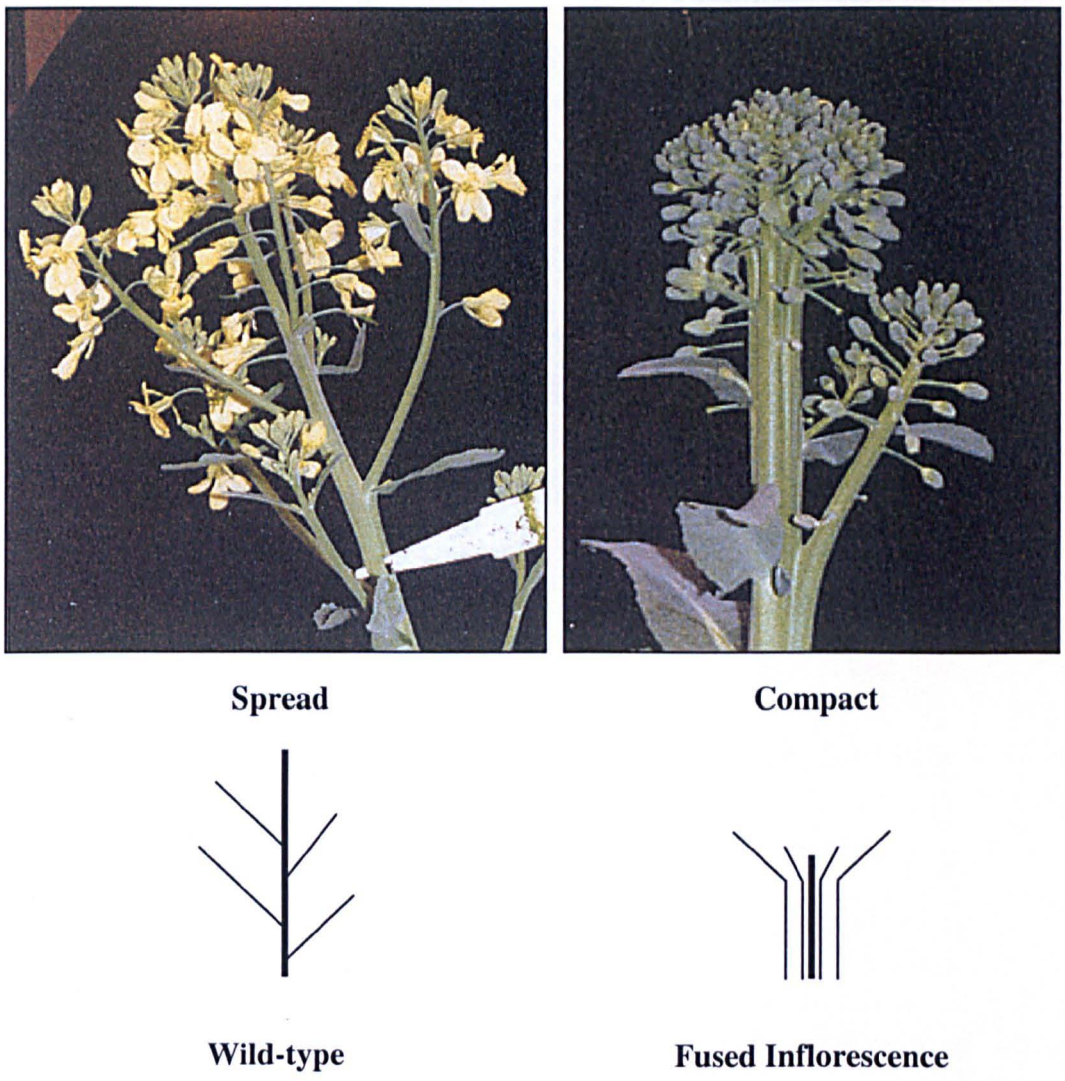


Figure 7.2: The effect on inflorescence architecture of the fused inflorescence trait segregating in the A12DH x GD33 population. Unfused plants exhibit a wild-type inflorescence spike, whereas fusion of the primary and secondary inflorescences contributes to the development of a dense head structure.

7.4.2 Phenotypic stability

Phenotypic stability was tested in a subset of the NxG DH population [cauliflower (fused) x Brussels Sprout (unfused)] and more completely in the A12xGD33 DH population [Chinese White Kale (unfused) x Calabrese (fused)] (table 7.1). All plants were scored as either 'fused' or 'unfused' following bolting of the inflorescence. Scoring was conducted following bolting to reduce early destructive scoring that would hamper later phenotypic analysis. Dissection of several replicate plants confirmed fusion was visible prior to bolting as well as post bolting.

7.4.3 Fused Inflorescence in Calabrese broccoli

Scoring in the A12xGD33 population was conducted across four replicate plants of each of 124 lines (table 7.1). Two of these lines were grown under glass and two in a coldframe. Plants were not randomised to a statistical algorithm as they were not initially grown with this analysis in mind. Thus any conclusions drawn must be treated as extremely preliminary.

Table 7.1 Relative stability of the fused inflorescence trait of heading brassicas in the A12DH x GD33 DH population. Eleven percent of lines scored resulted in one or more conflicting scores between replicates although these never exceeded more than one of the four replicates.

Consensus across 4 replicates	*Number of lines*
Fused	70
FU (conflicting scores)	14 (no more than 3:1 ratio)
Unfused	40

7.4.4 Fused Inflorescence in cauliflower

The Fused Inflorescence phenotype of cauliflower was investigated in the NxG population. Preliminary data was gathered during a field trial conducted at Birmingham University. Five replicates of each line were planted in a fully randomised field plot. Scoring of Fused Inflorescence in this population was incomplete as the population was destroyed shortly after scoring due to constraints placed on researchers at Birmingham. However, the data collected enabled linkage analysis, and thus comparison to the results from the A12DH x GD33 population.

Table 7.2 Relative stability of the fused inflorescence trait of heading Brassicas in the NxG DH population. A conflicting score was recorded in no more than one of five replicates for each of twelve lines.

Consensus across 5 replicates	*Number of lines*
Fused	5
FU (conflicting scores)	12 (no more than 4:1 ratio)
Unfused	29

The relatively high ratio between Fused and unfused plants scored in the NxG population (5:29 respectively) could potentially be attributed to the time of scoring. Only plants which had bolted were scored. The high number of unfused plants scored appears to correlate with time to bolting, suggesting the fused character may indeed contribute to the development of a head rather than an inflorescence spike.

7.4.5 Genetic linkage mapping

In an attempt to ascertain the crude genetic map position of any contributing loci, only the extreme phenotypic scores were included in the linkage analysis. Lines with conflicting scores were omitted from the analysis. Though far from ideal, this preliminary analysis was conducted with the aim of aiding future experimental design rather than to provide a definitive answer.

Independent genetic linkage mapping on the NxG and the A12DHxGD33 DH populations was hampered by 'noisy data' (interpreted from inspection of likely double recombinants). These 'mis-scores' could perhaps be attributed to environmental effects, or the effect of several independently segregating contributing loci, or a combination of both; alternatively there could have been mistakes during phenotype scoring. This resulted in relatively weak linkage to several regions of the genome rather than to a single specific locus (table 7.3).

Table 7.3: Recombination frequency and LOD score for linkage of Fused Inflorescence to fixed markers on the A12DHxGD33 and NxG genetic linkage maps

Population	*Linkage group*	*Recombination frequency/ LOD score*	*Linked marker*
A12DH x GD33	LGO4	0.40/0.54	pW139E1
	LGO5	0.37/1.91	pO92J1
NxG	LGO5	0.31/1.12	pO105J1

The behaviour algorithm employed by the JOINMAP genetic mapping programme (Stam and Van Ooijen, 1995) leads to extension of the linkage group length if the data is a particularly poor fit to the existing data set. Thus a new marker, may be positioned some distance away to accommodate several 'theoretical' recombination events.

Initial examination of the linkage results from the A12DHxGD33 population suggested this was the case on LGO5. However, analysis of the completely independent NxG population also highlighted the same region at the top of LGO5. The addition of AFLP markers to the NxG map place the Fused Inflorescence marker, albeit weakly, inside the existing data set without extending the linkage group length. Interestingly, the anchored marker pO105J1 is positioned at approximately the same genetic distance from Fused Inflorescence (fin) in both maps (fig. 7.3).

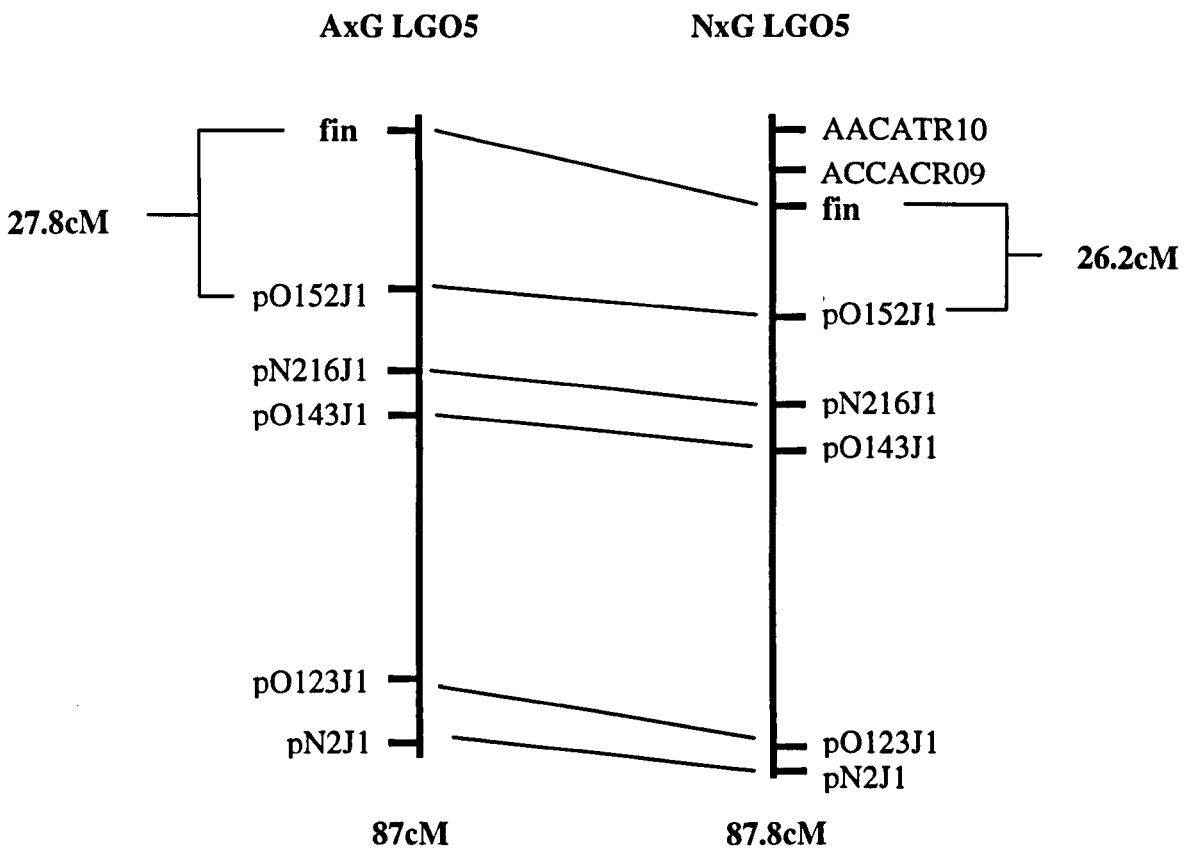


Figure 7.3: Comparative map location of Fused Inflorescence on LGO5 in the A12DHxGD33 and NxG genetic linkage maps. Note: although linkage is weak, *fin* is situated approximately the same distance from pO152J1 in both independent populations.

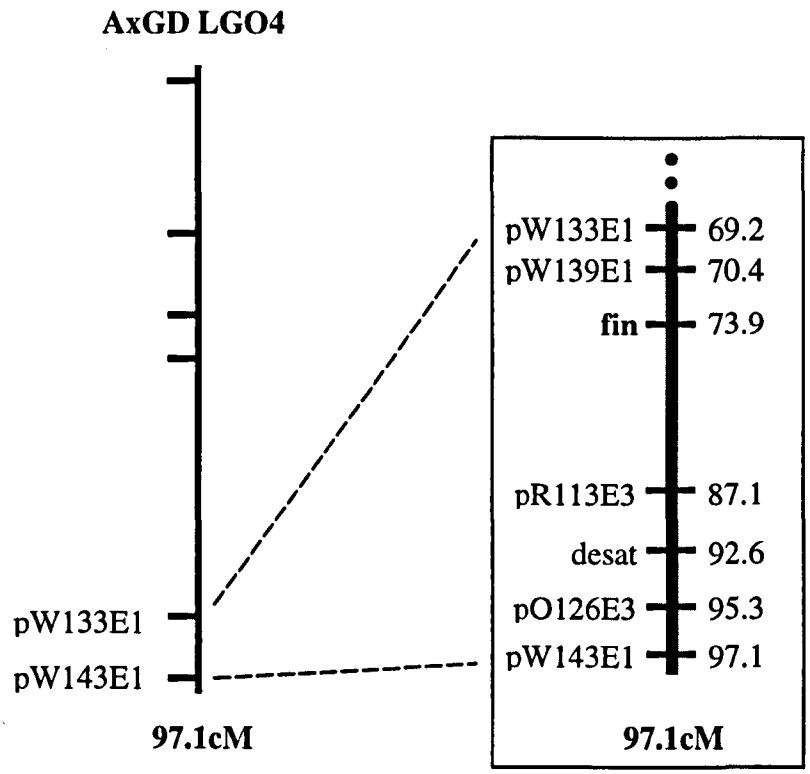


Figure 7.4: Map location of Fused Inflorescence on LGO4 in the A12DHxGD33 genetic linkage map

A second locus on LGO4 was also identified in the A12DHxGD33 population raising the possibility that Fused Inflorescence is a quantitative, multi-genic trait (Fig. 7.4).

7.4.6 Cauliflower/Calabrese comparison

Examination of Fused Inflorescence was also made in the NxB cauliflower x Calabrese-broccoli DH population. Both the cauliflower (DJ7032) and the Calabrese (BI87053) were found to express the Fused Inflorescence character, as was the F1 of the cross. However, Fused Inflorescence was found to segregate amongst the DH progeny lines (table 7.4).

Table 7.4: Segregation of the fused inflorescence trait amongst the DH progeny lines of the NxB cross.

Consensus across 2 replicates	*Number of lines*
Fused	26
FU (conflicting scores)	4
Unfused	22

The presence of unfused plants in a population derived from fused parents suggests that this phenotype is being complemented by non-allelic loci, and thus that Fused Inflorescence may be a multigenic trait.

7.5 Discussion

Visual examination has suggested that Fused Inflorescence is a character that contributes to heading propensity in the vegetable brassicas. However, examination of segregation in three DH populations suggests that Fused Inflorescence is a complex trait. The presence of conflicting scores within-genotype in all three populations raises the possibility of environmental effects impacting on any underlying genetic contribution.

Results from preliminary linkage analysis demonstrated extremely weak linkage to more than one locus, although it is interesting that the same region on LGO5 was identified in two independent populations as possibly containing a contributing locus. The identification of a second locus on LGO4 in the A12DHxGD33 population suggest that Fused Inflorescence may be a multifactorial trait, which could help to explain the poor LOD scores and the high levels of double recombinants when using both the JOINMAP pairwise linkage mapping software, and manual alignment. Perhaps a more detailed quantitative scoring system, rather than a simple 'presence' or absence' in combination with better controlled environment growth conditions and stronger statistical design and analysis, would help to narrow regions of interest.

The question of Fused inflorescence being multifactorial may also be supported by the observed segregation within the NxB population.

A. thaliana plants with mutations in several genes show similar phenotypes to specific characters of heading brassicas, although no single one displays all characters. Plants homozygous for mutations in the genes *FASCIATA1* and *FASCIATA2* show inflorescence stem enlargement and bifurcation. This bifurcation results in establishment of two or more new spiral phylotaxies. The mutant inflorescence also fails to maintain stem elongation so that the mature structure appears crowded with flowers and siliques (Ottoline Leyser and Furner, 1992). The arrested stem elongation and the crowding of flowers are reminiscent of the Calabrese broccoli phenotype.

Visual examination shows that the inflorescence of the *ap1-1/cal-1* double mutant *A. thaliana* in fact more closely resembles a White Sprouting Broccoli rather than a cauliflower. Crossing experiments between the *ap1-1/cal-1* mutant and the *fas1* and *fas2* mutants could help to determine whether the *Brassica* homologues of either of these genes contribute to the heading habit of certain *Brassica* crops.

7.5.1 Conclusions:

- Fused Inflorescence appears to contribute to head formation in crop brassicas.
- This trait is affected by environment but probably has a strong genetic component.
- Genetic mapping and population segregation data suggests the possibility of multiple contributing loci.

CHAPTER 8: DISCUSSION

This final chapter aims to pool together the discrete conclusions derived by previous chapters. and address the initial objectives in light of these results. This analysis enables suggestions to be made regarding the direction of further research into this subject area.

8.1 Objective 1: Develop an understanding of the genetic interactions responsible for the curding habit of cauliflowers and construct a genetic model.

Prior to this research the knowledge of genetic factors underlying the development of curd in the *Brassica* cauliflower was limited, primarily due to the perceived complexity of the curd trait (Crisp, 1982). The identification of an *A. thaliana* double mutant resembling the cauliflower phenotype of *Brassica* (Bowman *et al.*, 1993), and subsequent cloning of a mutated copy of *BoCAL* (Kempin *et al.*, 1995) provided an entry point which could be exploited to dissect the curding trait.

8.1.1 The role of *BoCAL-a* in cauliflower curd development

Four distinct pieces of evidence strongly suggest that *BoCAL-a* is involved in the development of curd in the *Brassica* cauliflower. Firstly the *A. thaliana* orthologue *CAL* is involved in the curding *A. thaliana ap1-1/cal-1* mutant. Secondly, the *BobCAL-a* allele possesses a premature termination codon in exon 5. Similar mutations in the *A. thaliana AP1* gene show strong LOF effects (Kempin *et al.*, 1995). Thirdly, the *BobCAL-a* allele derived from a cauliflower has been shown to co-segregate completely with the curd phenotype, in a DH population. Fourthly, diagnosis of this locus throughout the *B. oleracea* gene-pool shows specific alleles are at high frequency in curding crop types, suggesting an extremely strong correlation between this locus and the curd trait.

These lines of evidence strongly suggest that *BoCAL-a* is one of the genes that contribute to the development of curd. An alternative hypothesis is that a gene distinct from, yet closely linked to *BoCAL-a* is the contributing locus, and thus *BoCAL-a* is not involved, but simply represents a marker of curd development. If this were the

case, the segregation and population evidence reported above would be the result of linkage drag upon *BoCAL-a,* the result of active selection for the nearby locus. Such a proposition appears plausible, yet the development of similar models independently for two related species, both *A. thaliana,* and *B. oleracea,* suggest that a copy of *BoCAL* is likely to be one of the contributing loci for the development of curd in the *Brassica* cauliflower.

8.1.2 The Role of *BoAP1* in cauliflower floral induction and curd development

To address the question of mode of action of *BoAP1* in curd development a theoretical expression model may be proposed. A similar model was proposed following analysis of increasing copy number of the *LEAFY* (*LFY*) gene in *A. thaliana* (Blazquez *et al.*, 1997), which demonstrated that increasing copy number under the control of the endogenous promoter reduced the time taken to reach the required LFY threshold to initiate floral induction (fig. 8.1).

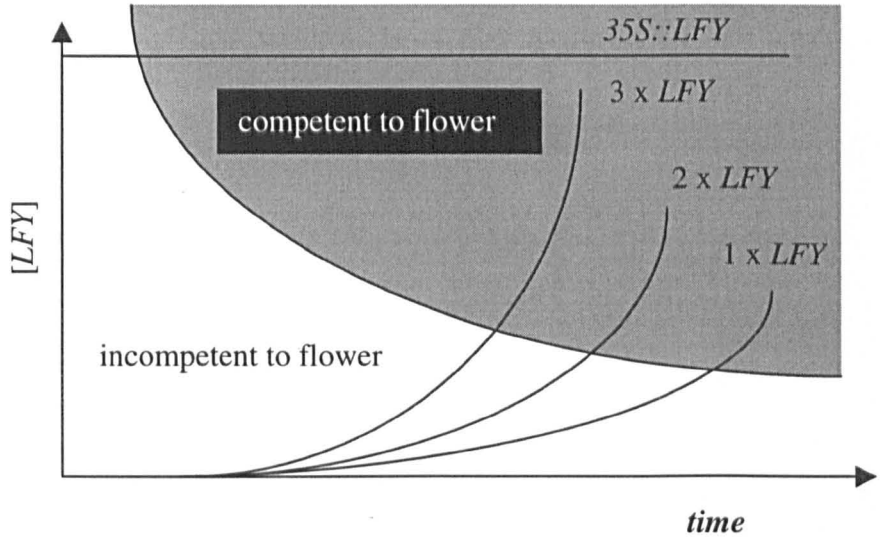


Figure 8.1: Experimental model showing effect of multiple copies of *LFY* on time to floral induction in *A. thaliana*. Reproduced from (Blazquez *et al.*, 1997). As copy number of *LFY* increases under the control of the endogenous promoter, the time to flowering is reduced. Constitutive expression results in extremely rapid flowering.

Application of this model to the naturally evolved situation regarding the replicated *BoAP1* loci in *Brassica oleracea* provides insights into the possible mechanism by which cauliflower curds develop (fig. 8.2).

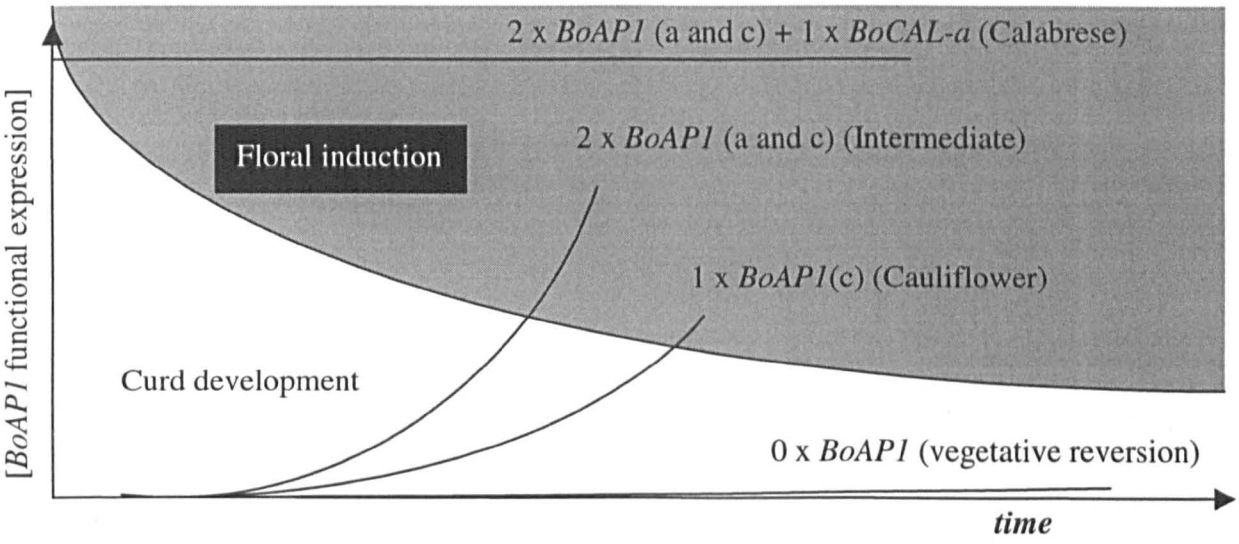


Figure 8.2: Theoretical model of role of *BoAP1* threshold requirement to induce floral induction in the *Brassica* cauliflower. At non-permissive temperatures *BoAP1* expression is undetectable (Anthony *et al.*, 1996) and vegetative reversion occurs on previously induced cauliflowers. Expression of *BoAP1-c* allows flowering to occur in cauliflowers, albeit at an extremely slow rate, thus resulting in curd development. A wild-type copy of both *BoAP1-a* and *BoAP1-c* (such as in half of the NxB population intermediate phenotypes) results in a more rapid induction of flowers, albeit following initial curd development. The introduction of a wild-type *BoCAL-a* (such as in the Calabrese broccoli parent) results in rapid floral induction with an extremely small curd stage (Kieffer *et al.*, 1998).

This model suggests that the role of *BoAP1*, as in *A. thaliana*, is to initiate floral induction. When moved to high temperatures, floral induction is halted in curding cauliflowers and *BoAP1* expression drops to undetectable levels (Anthony *et al.*, 1996). Thus with no expression of *BoAP1* (in a *BoCAL*-a mutant background), no flowering is initiated.

In cauliflowers grown at permissive temperatures, the threshold requirement for floral induction is reached, albeit slowly. This time delay is crucial, as it allows the undifferentiated proliferation of apical meristems, and *time* for the curd to develop.

Application of segregation evidence to the genetic model suggests that *BoAP1-c* is not directly responsible for the development of curd (chapter 4). This could suggest that a single copy of *BoAP1* (*BoAP1-c*) is functionally expressed in cauliflowers, perhaps accounting for later wild-type flower development.

NxB lines expressing an intermediate phenotype reach the threshold for floral induction more rapidly than cauliflowers.. For one half of the NxB intermediate lines, this could be due to the presence of functional *BoAP1-a* from the Calabrese broccoli parent. Thus with two functional loci (*BoAP1-a* and *BoAP1-c*) the threshold for floral induction is reach after a much reduced curding stage.

In plants expressing a wild-type *B. oleracea* inflorescence, functional expression of both of the *BoAP1* loci is in the genetic background of a wild-type *BoCAL-a* and the threshold for floral induction is reached following an almost undetectable curd stage (Kieffer *et al.*, 1998).

The recombination between *BoAP1-c* and the cauliflower curd phenotype (NxB 616b, chapter 4) suggests that this locus does not have such a significant effect on floral meristem identity as *BoAP1-a.*Thus although the presence of *BoAP1-c* expression may allow eventual floral induction, it is unable to compensate fully for any LOF of *BoAP1-a.* Unfortunately no reciprocal recombination events were observed in the population so we may only speculate as to the action of a Calabrese derived *BoAP1-a* in the background of a cauliflower derived *BoAP1-c.* If the genetic model is correct, such that *BoAP1-c* is not directly involved, this haplotype should preclude curd development due to the presence of a functional *BoAP1-a.*

8.1.3 The role of *BoAP1* function in flower development

Genetic data suggesting the direct involvement of *BoAP1-a,* but not *BoAP1-c,* in curd development may prove to be highly significant regarding the role of each locus. Further analysis of *BoAP1-c* may help to answer the question raised since the characterisation of the *ap1-1/cal-1 A. thaliana* mutant. If the models are the same in both *A. thaliana* and the *Brassica* cauliflower, why do cauliflowers always develop wild-type flowers?

Logical examination of the key roles of *BoCAL* and *BoAP1* function in floral induction (based on the *A. thaliana* model) suggests that *BoAP1-c* should be functional.

Assuming that *BoAP1-a* and *BoCAL-a* are proven to be involved in curd development, the arrest at a recognised ontological stage in cauliflowers (floral induction) suggests that these genes contain LOF mutations. At a gross level, LOF appears to be temporary though, as all plants eventually bolt and produce wild-type flowers. A complete LOF of all *BoAP1* loci suggests that cauliflowers, already LOF for *BoCAL-a,* would never overcome the arrest stage, in effect producing a lethal mutant phenotype as the plants would never set seed. This suggests that the persistence of cauliflower forms is as much dependent on the *lack* of complete LOF of *BoAP1* as it is on LOF of *BoCAL-a.*

A hypothesis could therefore suggest that *BoAP1-c* may act to rescue the cauliflower phenotype from sterility by allowing development to proceed to seed set, and therefore should be heavily selected in favour of functionality in the cauliflower gene-pool. This hypothesis could be tested utilising an SSR discovered in intron 1 of *BoAP1-c.* The oligonucleotide primers AP1CSSRF and AP1CSSRR (appendix 1) were designed from *BoAP1-c* genomic DNA (BAC BoB28L01), and amplify a fragment of 120-150bp. Preliminary analysis demonstrates that this SSR is polymorphic in the *B. oleracea* genepool. Characterisation of this locus through the diversity population in much the same way as for *BoCAL-a* (chapter 6), may help ascertain any selection of specific alleles of *BoAP1-c,* and the importance of its role, if any, in maintaining cauliflowers as a viable crop.

An alternative model could take into account a wider role for each expressed locus of *BoAP1* and *BoCAL.* Expansion of a threshold model for the development of curd, to include completely redundant loci would suggest that the cauliflower DJ7032 represents just one configuration of *BoAP1* allelic states that can promote curd development. In other cauliflowers, A reversal of functional status between BoAP1 loci could be envisaged whereby a mutation in *BoAP1-c* rather than *BoAP1-a* would promote

curding.. Thus a redundant system could be envisaged amongst the cauliflower genepool, whereby potentially lethal mutations (such as loss of floral meristem identity) could be buffered by replicated loci. Such buffering could allow partial complementation of the LOF of a specific copy, thus resulting in an overall reduction in levels of functional protein and a time lag before floral development proceeds. A similar hypothesis could be proposed for interactions between distinct *BoCAL* loci. Sequencing of distinct loci from the cauliflower accessions found to lack the *BobCAL-a* mutation may find either distinct mutations in *BoCAL-a,* or alternatively, a functional *BoCAL-a,* with LOF of other loci.

8.2 Objective 2: Utilise molecular genetic technology to address the conflicting debate regarding the classification of broccoli and cauliflower types within the *B. oleracea* genepool.

Analysis of the combined results of this research suggests that the gross inflorescence morphologies of the cauliflowers and broccoli can be characterised in terms of allelic status at a very few loci. This addresses quite critically the conventional view of a natural hierarchical relationship within the broccoli and cauliflower genepool (Gray, 1989; Synge, 1956). Cauliflowers and broccolis, appear to group together on overall genetic relatedness (Gray 1989, Massie 1998). This observation is supported by examination of the inflorescence morphologies of broccolis and cauliflowers. This suggests that the major distinguishing features can in theory be explained by interactions amongst three contributing factors, the allelic state of loci defined by *BoCAL-a* and *BoAP1-a,* and the relative contribution of Fused Inflorescence (fig. 8.3). Given changes at just two well defined genetic loci and in a third trait, there is little justification for the taxonomic status distinguishing cauliflowers (*B. oleracea* var. *botrytis*) and broccoli (var. *italica*).

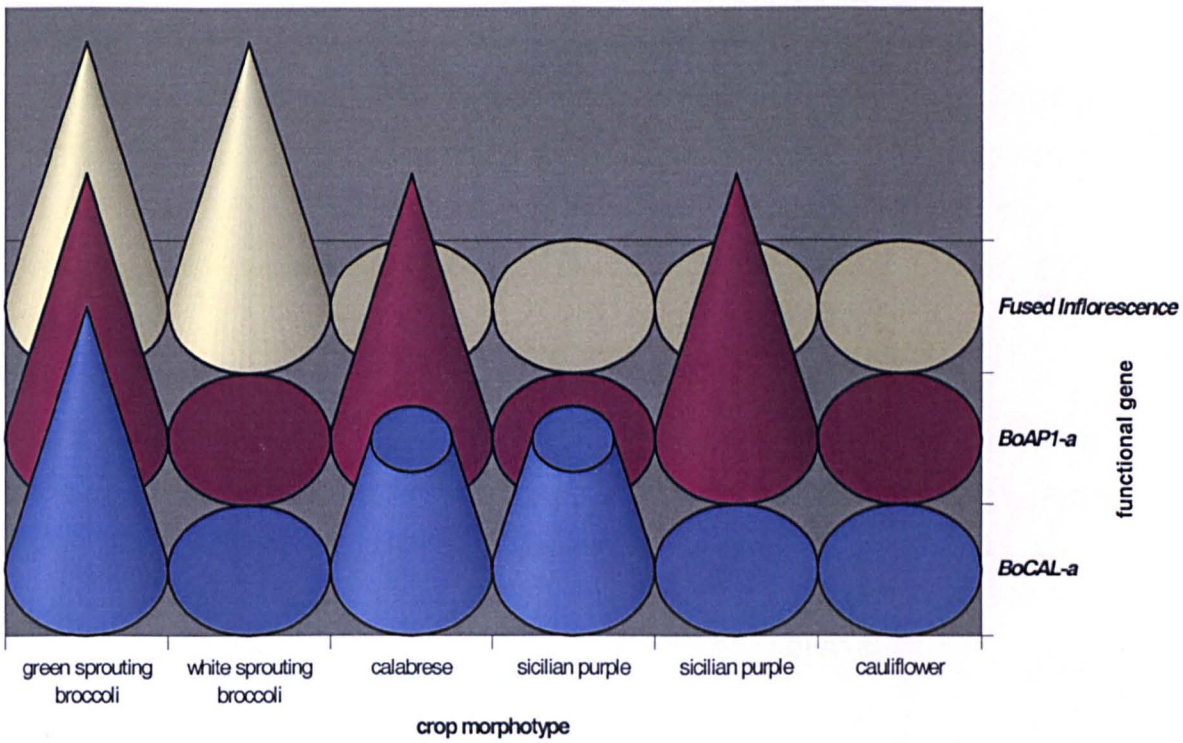


Figure 8.3: Theoretical model of how the functional status of *BoCAL-a* and *BoAP1-a,* allied to the status of inflorescence fusion may control the generic morphologies covering the entire range of crop morphotypes within broccolis and cauliflower. Cone denotes wild-type status, circle denotes affected status. N.B. half cone denotes evidence that suggests Calabrese may possess a weakened *BoCAL-a* allele (SSR156, chapter 6).

With hindsight, a system of hierarchical taxonomic classification applied to incompletely isolated, interfertile populations is bound to result in confusion. If there must exist an intra-specific delineation between morphotypes, evidence derived from this study suggests that at the very least, individuals currently covered by the varietal names of *botrytis* and *italica* should be combined within a single nomenclature class.

8.3 Objective 3: Compare the genomic distribution and the functional status of each copy of the meristem identity genes *BoCAL* and *BoAP1* in *Brassica oleracea*

The identification and characterisation of three distinct loci of *BoAP1* represents a significant advancement on previous knowledge, which up until very recently had assumed a single *BoAP1* locus. This result is made all the more significant given the novel discovery of two distinct expressed copies of *BoAP1* in the curd of cauliflower.

The ability to account for previously published cDNA clones and assign them to specific loci will be of practical use for further investigations into *Brassica* floral induction. The characterisation of these loci also provided a stringent test of the methods used during previous investigations, and strongly suggested that the published *BoAP1* cDNA sequence from cauliflower (Anthony *et al.*, 1995) is chimeric.

Two of the three *BoAP1* loci identified were mapped (*BoAP1-a* and *BoAP1-c*). The third locus, *BoAP1-b* could not be mapped, as no polymorphism was found in the parents of the mapping populations, despite the generation of over 2Kb of contiguous sequence. The reason for such a high level of sequence conservation at *BoAP1-b*, a locus which appears not to be expressed during floral induction in cauliflowers, and in which a premature termination codon is present in exon 4, remains to be found.

The failure to identify and characterise further copies of *BoCAL* via molecular genetic means was unfortunate, although it provided a good opportunity to examine the utility of cytogenetic analysis as an aid to physical mapping in *B. oleracea.* The three distinct loci of *BoCAL* suggested by FISH analysis are intriguing given the close physical proximity of two of the three hybridising regions. The two *AP1/CAL*-like loci identified by RFLP analysis are linked and in close proximity to *BoAP1-a* and *BoAP1-c*, although they are at least distinct from *BoAP1-c.* Further characterisation through *BoAP1* and *BoCAL* double-FISH probings would address whether these RFLP loci are likely to represent novel copies of *BoCAL*, if so, this would provide future researchers with an extremely interesting region of the genome to examine.

8.4 Objective 4: Utilise knowledge gained from molecular genetic analysis to examine the genetic relationship between broccoli and cauliflower and trace the domestication of the Brassica cauliflower curd trait.

The development of a possible domestication model based on a key putative contributing locus (*BoCAL-a*) represents an advancement over previous RFLP and RAPD based surveys. The ability to direct analysis specifically to follow a putative

contributing gene has provided supporting evidence regarding the co-segregation of specific alleles of this locus with the curding phenotype. In addition it has identified the Sicilian Purple type as a potential intermediate form during the domestication of the curd trait. Interestingly, although local varieties have been shown still to persist (Massie, 1998), this investigation suggests introgression of distinct alleles into morphologically distinct cauliflower populations has occurred.

To conclude, it would appear that cauliflower curd development could arise from a pivotal interaction between multiple copies of multiple related genes. In the population examined, two key loci (*BoCAL-a* and *BoAP1-a*) characterise curd development, although a third, *BoAP1-c,* may act to ensure wild-type flower development. The molecular genetic dissection of this delicately integrated system is hampered by the close relatedness of the loci involved and the large potential for epistasis between loci. Utilisation of synthetic oligonucleotide-cDNA microarray technology, which is able to distinguish between alleles of specific genes (Hacia *et al.,* 1996) could be utilised to distinguish between replicated loci. This approach may provide an opportunity to characterise the relative expression level of each locus, in comparisons of floral induction in cauliflowers and other *B. oleracea.* Experiments of this nature may also be applicable to examination of the development of the economically important trait of 'riceyness' in commercial cauliflowers, which may result from relative expression level differences of such genes.

8.5 Areas warranting further investigation

Several areas of research highlighted during this study require further investigation. The majority of experiments are concerned with testing and re-testing the genetic model of curding involving the *BoAP1-a* and *BoCAL-a* loci.

In light of comparative DNA sequence analysis which suggested that the published *BoAP1* cDNA sequence from cauliflower (Anthony *et al.,* 1995) is chimeric, completing the genomic sequencing of *BoAP1-a* from the cauliflower DJ7032 would provide a valuable reference. This would allow further characterisation of the possible role of *BoAP1-a* in cauliflower curd development.

Completion of DNA sequencing of each of the distinct loci of *BoAP1* is also required. The realisation of this task will provide the first opportunity to characterise in detail, interactions between specific triplicated loci in *B. oleracea.*

Analysis of the coding region of *BoAP1-a* has failed to definitively identify a mutational change that could be involved in the development of cauliflower curd. However, application of assays designed around each of the identified amino-acid substitutions to the diversity population may help narrow the possibilities. *BoAP1-a* also appears to be correctly spliced following transcription, suggesting that any change in function of *BoAP1-a* may lie in the promoter of this gene. A change in control of expression of *BoAP1-a* between cauliflower and other *B. oleracea* varietas would integrate well into the 'threshold' model of *BoAP1-a* involvement proposed by (Anthony *et al.*, 1996), and developed further above. This hypothesis would appear to be testable by comparing the relative expression level of this locus in the two intermediate classes of the NxB population (chapter 3). Such analysis could employ a bulk segregant method to filter out background effects derived from interactions with other segregating loci. The major stumbling block would be identification of comparable tissue given the compound structures of the intermediate inflorescences.

Further analysis of the involvement of *BoCAL* in cauliflower curd development is required. If the novel loci of *BoCAL* suggested by FISH analysis (chapter 5) can be identified and cloned, characterisation in a manner similar to that undertaken for *BoAP1* loci could provide significant information regarding the control of floral induction in a species possessing genomic replication.

Identification of several cauliflowers lacking the premature stop codon in exon 5 of *BoCAL-a* raises several important points. If *BoCAL-a* is directly involved in the development of curd, identification of contributing mutations from each of the three novel alleles linked to curding is required. Characterisation of such mutations may allow the development of further molecular markers for use by cauliflower breeders. The identification of one such allele *SSR153* in a particular pedigree of Australian cauliflowers, also raises the possibility of using such assays to trace the radiation of

germplasm from the centre of diversity to regions across the world. Alternatively, these accessions may represent key recombinants, thus providing evidence that would exclude *BoCAL-a* as being involved in the development of the curd phenotype of cauliflowers.

A more robust statistical analysis of the Fused Inflorescence trait is warranted. The preliminary evidence is encouraging, although further examination with replicated trials should be able to better quantify any environmental effects. Quantitative assessment could take into account time of onset, number of fused inflorescences per spike, and fusion below the cauline leaf, which were not addressed during this study.

8.5.1 Complementation of the cauliflower phenotype through transformation

Despite the large amount of circumstantial evidence supporting the genetic model, which can be accumulated through classical and molecular genetic means, the classical proof of the model would only be achieved through complementation of the observed phenotype. Transformation is the currently accepted mode of proof.

Transformation of *B. oleracea* is becoming more routine (Puddephat *et al.*, 1996), albeit with long generation times. However, with the development of rapid cycling lines the seed to seed time has been significantly reduced, down to 56 days (Senna, G King pers. comm.). Although it may be possible to transform *Brassica,* there are several potential pitfalls associated with utilising transformation as a mechanism for testing the genetic model proposed for curd development.

The evidence accumulating that the *BoCAL-a* allele from the cauliflower DJ7032 is a recessive LOF allele (presence of premature stop mutation, parental-Calabrese phenotype of NxB F1) suggests, at first glance, that complementation of *BoCAL-a* in cauliflowers with a wild-type cDNA is a viable option, and should result in development of 'intermediate' type curds in successful transformants. However, problems arise from the preliminary evidence which suggests that other copies of *BoCAL* are present in *B. oleracea.* Without also characterising the role of these additional loci in directing floral induction in cauliflower, it may prove difficult to defend the conjecture that the function complemented is solely that of *BoCAL-a.* A

similar situation can be envisaged regarding transformation of cauliflower with a functional *BoAP1* cDNA. Despite the characterisation of each of three loci, it would at present be impossible to say that functional complementation is limited to the action of one loci.

Perhaps conclusive evidence that would prove or disprove involvement of specific loci in the development of curd awaits the development of a system for conducting gene knockouts via targeted recombination. Although preliminarily tested in *A. thaliana* (Halfter *et al.*, 1992; Kempin *et al.*, 1997), no publications have yet been reported in the literature of application of this technology to *Brassica.* If such technology could be perfected then the LOF of each copy of *BoAP1* and *BoCAL* could be studied in isolation, and in combination via inter-breeding of mutants.

BIBLIOGRAPHY

Acton, T. B., Zhong, H. L., and Vershon, A. K. (1997). DNA-binding specificity of MCM1: Operator mutations that alter DNA- bending and transcriptional activities by a MADS box protein. *Molecular and Cellular Biology* **17,** 1881-1889.

Alvarez, J., Guli, C. L., Xiang-Hua, Y., and Smyth, D. R. (1992). *TERMINAL FLOWER*: a gene affecting inflorescence development in *Arabidopsis thaliana. The Plant journal,* **2,** 103-116.

Anthony, R. G. (1994). Isolation and characterisation of meristem identity genes expressed during cauliflower (*Brassica oleracea* var. *botrytis* L.) curd development. PhD thesis, School of Biological Sciences: University of Sussex), pp. 169.

Anthony, R. G., James, P. E., and Jordan, B. R. (1996). Cauliflower (*Brassica oleracea* var *botrytis* L) curd development: The expression of meristem identity genes. *Journal of Experimental Botany* **47,** 181-188.

Anthony, R. G., James, P. E., and Jordan, B. R. (1995). The cDNA sequence of a cauliflower *apetala-1/squamosa* homolog. *Plant Physiology* **108,** 441-442.

Anthony, R. G., James, P. E., and Jordan, B. R. (1993). Cloning and sequence analysis of a *flo/lfy* homologue isolated from cauliflower (*Brassica oleracea* L. var. *botrytis). Plant Molecular Biology* **22,** 1163-1166.

***Arabidopsis* Genome Initiative** (1999) http://www.stanford.edu

Armstrong, S. J., Fransz, P., Marshall, D. F., and Jones, G. H. (1998). Physical mapping of DNA repetitive sequences to mitotic and meiotic chromosomes of *Brassica oleracea* var. *alboglabra* by fluorescence *in situ* hybridization. *Heredity* **81,** 666-673.

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9,** 208-218.

Bailey, N. T. J. (1981). Statistical methods in biology, Second Edition. Hodder and Stoughton: London.

Bjorkman, T., and Pearson, K. J. (1998). High temperature arrest of inflorescence development in broccoli (*Brassica oleracea* var. *italica* L.). *Journal of Experimental Botany* **49,** 101-106.

Blazquez, M. A., Soowal, L. N., Lee, I., and Weigel, D. (1997). *LEAFY* expression and flower initiation in *Arabidopsis. Development* **124,** 3835-3844.

Bohuon, E. J. R., Keith, D. J., Parkin, I. A. P., Sharpe, A. G., and Lydiate, D. J. (1996). Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus. Theoretical and Applied Genetics* **92:,** 1-7.

Boukema, I. W., Van Hintum, T. J. L., and Astley, D. (1997). The creation and composition of the *Brassica oleracea* core collection. *IPGRI Plant Genetic Resources Newsletter* **111,** 29-32.

Bowman, J. L., J, A., Weigel, D., Meyerowitz, E. M., and Smyth, D. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119,** 721-743.

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis. Development,* **112,** 1-20.

Branca, F., and Iapichino, G. (1997). Some wild and cultivated Brassicaceae exploited in Sicily as vegetables. *Plant Genetic Resources Newsletter* **110,** 22-28.

Carr, S. M., and Irish, V. F. (1997). Floral homeotic gene expression defines developmental arrest stages in *Brassica oleracea* L. vars. *botrytis* and *italica. Planta* **201,** 179-188.

Cavell, A. C., Lydiate, D. J., Parkin, I. A. P., Dean, C., and Trick, M. (1998). Colinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* Genome. *Genome* **41,** 62-69.

Cho, S. C., Jang, S. H., Chae, S. J., Chung, K. M., Moon, Y. H., An, G. H., and Jang, S. K. (1999). Analysis of the C-terminal region of *Arabidopsis thaliana APETALA1* as a transcription activation domain. *Plant Molecular Biology* **40,** 419-429.

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* **353,** 31-37.

Conner, J. A., Conner, P., Nasrallah, M. E., and Nasrallah, J. B. (1998). Comparative mapping of the *Brassica S* locus region and its homeolog in *Arabidopsis:* Implications for the evolution of mating systems in the *Brassicaceae. The plant cell* **10,** 801-812.

Crisp, P. (1983). Pre-breeding of cruciferous vegetables. *Genetika* **15,** 337-353.

Crisp, P. (1982). The use of an evolutionary scheme for cauliflowers in the screening of genetic resources. *Euphytica* **31,** 725-734.

Crisp, P., and Gray, A. R. (1984). Breeding old and new forms of purple heading broccoli. *Cruciferae Newsletter* **9,** 17-18.

Crisp, P., Gray, A. R., and Jewell, P. A. (1975). Selection against the bracting defect of cauliflower. *Euphytica* **24,** 459-465.

Crisp, P., and Tapsell, C. R. (1993). Cauliflower, *Brassica oleracea* L. In Genetic improvement of vegetable crops, pp. 157-178. Edited by G. Kalloo and B. O. Bergh.:

Pergamon Press, Oxford.

De Nettancourt, D. (1977). Incompatibility in angiosperms, Volume 3. Springer-Verlag: Berlin.

Dias, J. S., Lima, M. B., Song, K. M., Monteiro, A. A., Williams, P. H., and Osborn, T. C. (1992). Molecular taxonomy of Portuguese tronchuda cabbage and kale landraces using nuclear RFLPs. *Euphytica* **58,** 221-229.

Dickson, M. H., and Lee, C. Y. (1980). Persistent white curd and other curd characters of cauliflower. *Journal of the American Society for Horticultural Science* **105,** 533.

Fujime, Y. (1983). Studies on thermal conditions of curd formation and development in cauliflower and broccoli, with special reference to abnormal curd development. *Memoirs of Faculty of Agriculture, Kagawa University* **40,** 117-123.

Fujime, Y., and Okuda, N. (1996). The physiology of flowering in brassicas, especially about cauliflower and broccoli. *Acta Horticulturae* **407,** 247-254.

Gallegly, M. E. (1956). Progress in breeding for resistance to clubroot of broccoli and cauliflower. *Phytopathology* **46,** 467.

Gates, R. R. (1953). Wild cabbages and the effects of cultivation. *Journal of Genetics* **51,** 365-372.

Goto, K., and Meyerowitz, E. M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes and Development* **8,** 1548-1560.

Grant, M. R., McDowell, J. M., Sharpe, A. G., De Torres Zabala, M., Lydiate, D. J., and Dangl, J. L. (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 15843-15848.

Gray, A. R. (1982). Taxonomy and evolution of broccoli *(Brassica oleracea* var. *italica). Economic Botany* **36,** 397-410.

Gray, A. R. (1989). Taxonomy and evolution of broccolis and cauliflower. *Baileya* **23,** 28-46.

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. A. and Collins, F. S. (1996) Detection of heterozygous mutations in *BRCA1* using high density oligonucleotide arrays and two colour fluorescence analysis. *Nature Genetics,* **14,** 441-447

Halfter, U., Morris, P.-C., and Willmitzer, L. (1992). Gene targeting in *Arabidopsis thaliana. Molecular and General Genetics* **231,** 186-193.

Hand, D. J., and Atherton, J. G. (1987). Curd initiation in the cauliflower. 1. Juvenility. *Journal of Experimental Botany* **38,** 2050-2058.

Honma, S., and Heecht, O. (1971). Green Ball-a new type of Cauliflower, *Michigan Agricultural Experimental Station research report.*

Hooper, W. D., and Ash, H. B. (1935). Translation of Cato, Marco Porcius (234-148 BC). In *'On Agriculture'*. Heinemann W.: London.

Hyams, E. (1971). Plants in the service of man. J M Dent and Sons: London.

Jack, T., Fox, G. L., and Meyerowitz, E. M. (1994). *Arabidopsis* homeotic gene *APETALA3* ectopic expression: Transcriptional and post-transcriptional regulation determine floral organ identity. *Cell* **76,** 703-716.

Jackson, S. A., Wang, M. L., Goodman, H. M., and Jiang, J. M. (1998) Application of Fibre-FISH in physical mapping of *Arabidopsis thaliana. Genome* **41,** 566-572

Jourdren, C., Barret, P., Brunel, D., Delourme, R., and Renard, M. (1996). Specific molecular marker of the genes controlling linolenic acid content in rapeseed. *Theoretical and Applied Genetics* **93,** 512-518.

Kaneko, T., Hirota, N., Yokoi, S., Kanatani, R., and Ito, K. (1999). Molecular marker for protein Z content in barley (*Hordeum vulgare* L.). *Breeding Science* **49,** 69-74.

Kempin, S. A., Liljegren, S. J., Block, L. M., Rounsley, S. D., and Yanofsky, M. F. (1997). Targeted disruption in *Arabidopsis. Nature* **389,** 802-803.

Kempin, S. A., Savidge, B., and Yanofsky, M. F. (1995). Molecular basis of the cauliflower phenotype in *Arabidopsis. Science* **267,** 522-525.

Kieffer, M., Fuller, M. P., and Jellings, A. J. (1998). Explaining curd and spear geometry in broccoli, cauliflower and 'romanesco': quantitative variation in activity of primary meristems. *Planta* **206,** 34-43.

Krizek, B. A., and Meyerowitz, E. M. (1996). Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ identity proteins. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 4063-4070.

Lagercrantz, U. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150,** 1217-1228.

Lagercrantz, U., and Lydiate, D. J. (1996). Comparative genome mapping in *Brassica. Genetics,* 1903-1910.

Lagercrantz, U., Putterill, J., Coupland, G., and Lydiate, D. (1996). Comparative mapping in *Arabidopsis* and *Brassica,* fine scale genome colinearity and congruence of genes controlling flowering time. *The Plant Journal* **9,** 13-20.

Lanner-Herrera, C., Gustafsson, M., Falt, A.-S., and Bryngelsson, T. (1996). Diversity in natural populations of wild *Brassica oleracea* as estimated by isozyme and RAPD analysis. *Genetic Resources and Crop Evolution* **43,** 13-23.

Levin, J. Z., and Meyerowitz, E. M. (1995). *UFO*: An *Arabidopsis* gene involved in both floral meristem and floral organ development. *The Plant Cell* **7,** 529-548.

Levy, Y. Y., and Dean, C. (1998). The transition to flowering. *The Plant Cell* **10,** 1973-1989.

Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S., and Yanofsky, M. F. (1999). Interactions among *APETALA1, LEAFY,* and *TERMINAL FLOWER1* specify meristem fate. *The Plant Cell* **11,** 1007-1018.

Lowman, A. C., and Purugganan, M. D. (1999). Duplication of the *Brassica oleracea APETALA1* floral homeotic gene and the evolution of the domesticated cauliflower. *Journal of Heredity* **In press.**

Lydiate, D., Sharpe, A., Lagercrantz, U., and Parkin, I. (1993). Mapping the *Brassica* genome. *Outlook on Agriculture* **22,** 85-92.

Ma, H. (1998). To be, or not to be, a flower - control of floral meristem identity. *Trends in Genetics* **14,** 26-32.

Ma, H., Yanofsky, M. F., and Meyerowitz, E. M. (1991). *AGL1-AGL6,* an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes and Development* **5,** 484-495.

Mandel, M. A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1. Nature* **360,** 273-277.

Mandel, M. A., and Yanofsky, M. F. (1995). A gene triggering flower formation in *Arabidopsis. Nature* **377,** 522-524.

Massie, I. H. (1998). Patterns of variation in the Italian landrace cauliflower and

broccoli. PhD thesis. University of London.: London, pp. 223.

Massie, I. H., Astley, D., and King, G. J. (1996). Patterns of genetic diversity and relationships between regional groups and populations of Italian landrace cauliflower and broccoli (*Brassica oleracea* L. var. *botrytis* L. and var. *italica* Plenck). *Acta Horticulturae* **407,** 45-53.

Mazur, B., Krebbers, E., and Tingey, S. (1999). Gene discovery and product development for grain quality traits. *Science* **285,** 372-375.

Meyerowitz, E. M. (1987). *Arabidopsis thaliana. Annual Review of Genetics* **21,** 93-111.

Mizukami, Y., and Ma, H. (1997). Determination of *Arabidopsis* floral meristem identity by *AGAMOUS. Plant Cell* **9,** 393-408.

Mozo, T., Fischer, S., MeierEwert, S., Lehrach, H., and Altmann, T. (1998). Use of the IGF BAC library for physical mapping of the *Arabidopsis thaliana genome. The Plant Journal.* **16,** 377-384.

Munster, T., Pahnke, J., DiRosa, A., Kim, J. T., Martin, W., Saedler, H., and Theissen, G. (1997). Floral homeotic genes were recruited from homologous MADS-box genes pre-existing in the common ancestor of ferns and seed plants. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 2415-2420.

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the *c-fos* serum response element. *Cell* **55,** 989-1003.

Pang, P. P., and Meyerowitz, E. M. (1987). *Arabidopsis thaliana*: A model system for plant molecular biology. *Bio/Technology* **5.**

Parcy, F., Nilsson, O., Busch, M. A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. *Nature* **395,** 561-566.

Pease, M. S. (1926). Genetic Studies in *Brassica oleracea. Journal of Genetics* **16,** 363-385.

Pellegrini, L., Tan, S., and Richmond, T. J. (1995). Structure of serum response factor bound to DNA. *Nature* **376,** 490-498.

Pidkowich, M. S., Klenz, J. E., and Haughn, G. W. (1999). The making of a flower: control of meristem identity in *Arabidopsis. Trends in Plant Science* **4,** 64-70.

Puddephat, I. J., Riggs, T. J., and Fenning, T. M. (1996). Transformation of *Brassica oleracea* L: A critical review. *Molecular Breeding* **2,** 185-210.

Quarrie, S. A., LazicJancic, V., Kovacevic, D., Steed, A., and Pekic, S. (1999).

Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany* **50,** 1299-1306.

Reichmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996). Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins *APETALA1, APETALA3, PISTILLATA,* and *AGAMOUS. Proceedings of the National Academy of the United States of America* **93,** 4793-4798.

Reichmann, J. L., and Meyerowitz, E. M. (1997). MADS domain proteins in plant development. *Biological Chemistry* **378,** 1079-1101.

Rong-Lin, W., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. *Nature* **398,** 236-239.

Sadik, S. (1962). Morphology of the curd of cauliflower. *American Journal of Botany* **49,** 290-297.

Salter, P. J. (1969). Studies of crop maturity in cauliflowers. I. Relationship between the times of curd initiation and curd maturity of plants within a cauliflower crop. *Journal of Horticultural Science* **44,** 129-140.

Salter, P. J., and James, J. M. (1974). Further studies on the effect of cold treatment of transplants on crop maturity characteristics of cauliflower. *Journal of Horticultural Science* **49,** 329-342.

Scheffler, J. A., Sharpe, A. G., Schmidt, H., Sperling, P., Parkin, I. A. P., Luhs, W., Lydiate, D. J., and Heinz, E. (1997). Desaturase multigene families of *Brassica napus* arose through gene duplication. *Theoretical and Applied Genetics* **94,** 583-591.

Scwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus.Science* **250,** 931-936

Sebastian, R. L., Howell, E. C., King, G. J., Marshall, D. F., and Kearsey, M. J. (1999). An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled haploid mapping populations. *Theoretical and applied genetics* **In press.**

Seyfarth, R., Feuillet, C., Schachermayr, G., Winzeler, M., and Keller, B. (1999). Development of a molecular marker for the adult plant leaf rust resistance gene *Lr35* in wheat. *Theoretical and Applied Genetics* **99,** 554-560.

Shannon, S., and Meeks-Wagner, D. R. (1993). Genetic interactions that regulate inflorescence development in *Arabidopsis. The Plant Cell* **5,** 639-655.

Shore, P., and Sharrocks, A. D. (1995). The MADS-Box family of transcription factors. *European Journal of Biochemistry* **229,** 1-13.

Slocum, M. K., Figdore, S. S., Kennard, W. C., Suzuki, J. Y., and Osborn, T. C. (1990). Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea. Theoretical and Applied Genetics* **80.**

Snogerup, S. (1980). The wild forms of the *Brassica oleracea* (2n = 18) and their possible relations to the cultivated ones. pp. 121-132, In *Brassica crops and wild allies.* Edited by S. Tsunoda, K. Hinata and C. Gomez-Campo. Tokyo: Japan Scientific Societies Press.

Song, K., Slocum, M. K., Osborn, T. C. (1995). Molecular marker analysis of genes-controlling morphological variation in *Brassica-rapa* (syn *campestris*). *Theoretical and Applied Genetics* **90,** 1-10.

Stam, P., and Van Ooijen, J. W. (1995). JoinMap (tm) version 2.0: Software for the calculation of genetic linkage maps. (Wageningen: CPRO-DLO).

Synge, P. M. (1956). Royal Horticultural Society dictionary of gardening. Edited by P. M. Synge. (Oxford: Clarendon Press).

Theissen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution* **43,** 484-516.

Thompson, K. F. (1976). Cabbages Kales etc. *Brassica oleracea* (Cruciferae). In *Evolution of crop plants,* pp. 49-52. Edited by N. W. Simmonds. (London: Longman group Ltd.).

Toxopeus, H. (1977). Outlines of breeding programs for clubroot resistance in *Brassica* crops. In *Woronin +100 conference,* pp. 124-136. Edited by, S. T. Buczacki and P. H. Williams. Madison, Wisconsin, USA,

Vriesenga, J. D., and Honma, S. (1971). Inheritance of seedling resistance to clubroot in *Brassica oleracea. Horticultural Science* **6,** 395.

Watts, L. E. (1966). Investigations on the inheritance and responses to selection of riceyness in early summer cauliflower. *Euphytica* **15,** 90-98.

Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., and Meyerowitz, E. M. (1992). *LEAFY* controls floral meristem identity in *Arabidopsis. Cell* **69,** 843-859.

Weigel, D., and Meyerowitz, E. M. (1993). Activation of floral homeotic genes in *Arabidopsis. Nature* **261,** 1723-1726.

Weigel, D., and Nilsson, O. (1995). A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377,** 495-500.

Weisaeth, G. (1974). Some problems and results in breeding for clubroot resistance

in cole crops. In *Eucarpia crucifer meeting*, SHRI, Dundee, UK, pp. 101-107.

West, A. G., and Sharrocks, A. D. (1999). MADS-box transcription factors adopt alternative mechanisms for bending DNA. *Journal of Molecular Biology* **286,** 1311-1323.

West, A. G., and Sharrocks, A. D. (1997). The role of DNA-bending in MADS-box transcription factor function. *Biochemical Society Transactions* **25,** S639.

West, A. G., Shore, P., and Sharrocks, A. D. (1997). DNA binding by MADS-box transcription factors: A molecular mechanism for differential DNA bending. *Molecular and Cellular Biology* **17,** 2876-2887.

Yeager, A. F. (1943). The characteristics of crosses between botanical varieties of cabbage, *Brassica oleracea. Proceedings of the American Society for Horticultural Science* **43,** 199-200.

Zeven, A. C. (1996). Sixteenth to eighteenth century depictions of cole crops, (*Brassica oleracea* L.), Turnips (*B. rapa* L. cultivar group vegetable turnip) and Radish (*Raphanus sativus* L.) from Flanders and the present day Netherlands. *Acta Horticulturae* **407,** 29-33.

APPENDIX 1: OLIGONUCLEOTIDE PRIMER SEQUENCES

Gene	Region	Orientation	Primer name	Primer sequence
BoAP1	exon 5	F	AP1B5F	GCA GAA CCA ACT TAT GTA
BoAP1	exon 8	R	AP1B8R	CCA AGG TTG CAG TTG TA
BoAP1	5'-exon 1	F	AP1BPF	GTC TGA GCT CTT CTT TAT CT
BoAP1	exon 2	F	AP1cDNAF	GAC AGC TTA TTG CAC CTG AG
BoAP1	exon 2	R	AP1cDNAR	CTC AGG TGC AAT AAG CTG TC
BoAP1	exon 1	R	AP1MAD1NTR	TTA GGG TAA ATC AGG TAG TC
BoAP1	exon 1	R	AP1MADSR	AGA ATG TCA CTT GTC TAT TG
BoAP1	exon 1	F	APIMADSF	GGT TCT CTT AGA GGA AAT AG
BoAP1	exon 7	R	GK96	TAA GGA TGC TGG ATT TGA TGC
BoAP1-a	exon 2	F	AP12F	CCA ATT TCT CTC CAT TAA C
BoAP1-a	exon 4	R	AP14R	GTG CCT AAA TTT CAA TAG AG
BoAP1-a	5'-exon 1	F	AP15PF	TTG GTT CAT ACC AAA GTC T
BoAP1-a	5'-exon 1	R	AP15PR	ATC CCT TTT AAA GGA ACT A
BoAP1-a	exon 5	R	AP15R	ACG ATT TTT AGC TCC CCT AT
BoAP1-a	intron 7	R	AP17R	AAG GAG ATG GCT GAT GAG AG
BoAP1-a	exon 8	R	AP18R	CAT ATG CAC ATG ATA TAG C
BoAP1-a	intron 1	F	AP1I1F	GAT AAG AAG CTA GCC CTA AT
BoAP1-a	intron 1	R	AP1I1R	AGG GCT AGC TTC TTA TCA TC
BoAP1-a	intron 6	F	AP1I6F	CAC TGT TCT CAC CAC ACT TC
BoAP1-a	intron 6	R	AP1I6R	GCT TGA AGT GTG GTG AGA AC
BoAP1-a	intron 5	F	API15F	ATA ACT GCG TGT AAG AAT CC
BoAP1-a	intron 2	R	API2R	GGA GAG AAA TTG GTT TAC AT
BoAP1-b	exon 4	F	AP1INDEL4F	TGA GCT AAA TAG GCT CTA AG
BoAP1-b	exon 4	R	AP1INDEL4R	GCT CCT TAG AGC CTA TTT AG
BoAP1-b	intron 6	F	INDEL6F	GTA CAA TAG TTT ATG AAA GGA
BoAP1-b	intron 6	R	INDEL6R	TCC TTT CAT AAA CTA TTG
BoAP1-c	intron 7	F	AP1667F	CGT ATG TTA ATG CGG TGG TC
BoAP1-c	intron 7	F	AP1667F	CGT ATG TTA ATG CGG TGG TC
BoAP1-c	intron 7	R	AP1667R	GAC CAC CGC ATT AAC ATA CG
BoAP1-c	intron 1	F	AP166i1F	ATC TTT GGT TCT CTA CTC
BoAP1-c	intron 1	R	AP166i1R	GAG TAG AGA ACC AAA GAT
BoAP1-c	intron 1	F	AP1CSSRF	CGA GCT TAT CAC GCT GTT GT
BoAP1-c	intron 1	R	AP1CSSRR	CCT CTC ATA GCG TTC CAG TA
BoCAL	exon 2	F	CALcDNAF	GTA CTA GAA CGC TAC GAG AG
BoCAL/BoAP1	exon 1	F	BOMB1F	GAA GAG GAT AGA RAA CAA GAT CA
BoCAL-a	exon 8	R	BOB2	GTC GAT ATA TGG CGA GTC C
BoCAL-a	intron 4	F	CAL4F	CAT ACA AAC GCA AAC ATC TC
BoCAL-a	5'-exon 1	F	CAL5PF	ATG AAC CTC GAG CTG CAC AC
BoCAL-a	exon 8	F	CAL8F	TTG CCG CAT GAA TGG ACT CG
BoCAL-a	intron 1	F	CALI1F	TGT GCC TTT GTT CGC CCT AA
BoCAL-a	intron 1	R	CALI1R	TTA GGG CGA ACA AAG GCA CA
BoCAL-a	intron 2	R	CALI2R	GAC AGA GTC TTG GAG ATC AT
BoCAL-a	intron 6	F	CALI6F	CAT GTA GGT GGG CTT TGT CC
BoCAL-a	intron 6	R	CALI6R	CGG ACA AAG CCC ACC TAC AT
BoCAL-a	intron 7	F	CALSSRF	GTT AAG TGT GGC GTT AG
BoCAL-a	intron 7	R	CALSSRR	CCT TGG TAC ATG CCA CTG AA
BoCAL-a	exon 7	F	LSCALF1	CCT TAC ATG GCA TCA TCT CC
BoCAL-a	exon 7	R	LSCALR1	GGA GAT GAT GCC ATG TAA GG

Appendix 2: DNA sequence alignments

Chinese White Kale A12DH *BoAP1-c* (exonic sequence and theoretical translation)

```
ATG GGG AGG GGT AGG GTT CAA CTG AAG AGG ATA GAG AAC AAG ATC AAT AGA CAA GTG ACA
 M   G   R   G   R   V   Q   L   K   R   I   E   N   K   I   N   R   Q   V   T

TTC TCC AAA AGA AGA GCT GGT CTT TTC AAG AAA GCT CAT GAG ATC TCT GTT CTC TGT GAT
 F   S   K   R   R   A   G   L   F   K   K   A   H   E   I   S   V   L   C   D

GCT GAA GTT GCC CTT GTT GTC TTC TCC CAT AAG GGG AAA CTC TTT GAG TAC TCC ACT GAT
 A   E   V   A   L   V   V   F   S   H   K   G   K   L   F   E   Y   S   T   D

TCT TGT ATG GAG AAG ATA CTG GAA CGC TAT GAG AGG TAC TCT TAC GCG GAG AGA CAG CTT
 S   C   M   E   K   I   L   E   R   Y   E   R   Y   S   Y   A   E   R   Q   L

ATT GCA CCT GAG TCC GAC GTC AAT ACG AAC TGG TCA ATG GAG TAC AAC AGG CTT AAG GCT
 I   A   P   E   S   D   V   N   T   N   W   S   M   E   Y   N   R   L   K   A

AAG ATT GAG CTT TTG GAG AGA AAC CAG AGG CAC TAT CTT GGG GAA GAC TTG CAA GCA ATG
 K   I   E   L   L   E   R   N   Q   R   H   Y   L   G   E   D   L   Q   A   M

AGC CCT AAA GAG CTC CAG AAT CTG GAA CAG CAG CTT GAC ACT GCC CTT AAG CAC ATC CGA
 S   P   K   E   L   Q   N   L   E   Q   Q   L   D   T   A   L   K   H   I   R

TCT AGA AAA AAC CAA CTT ATG TAC GAC TCC GTC AAT GAG CTC CAA AGA AAG GAG AAG GCG
 S   R   K   N   Q   L   M   Y   D   S   V   N   E   L   Q   R   K   E   K   A

ATA CAA GAA CAA AAC AGC ATG CTT TCT AAA CAG ATC AAG GAG AGG GAA AAG GTT CTT ATG
 I   Q   E   Q   N   S   M   L   S   K   Q   I   K   E   R   E   K   V   L   M

GCA CAA CAR GAG CAA TGG GAC CAG CAG AAC CAT GGC CAA AAT ATG CCT TCG CCG CCG CCC
 A   Q   Q   E   Q   W   D   Q   Q   N   H   G   Q   N   M   P   S   P   P   P

CCG CAR CAG CAT CAA ATC CAG CAT CCA TAC ATG CTC TYT CAT CAG CCA TCT CCT TTT CTC
 P   Q   Q   H   Q   I   Q   H   P   Y   M   L   S   H   Q   P   S   P   F   L

AAC ATG GGT GGC CTG TAT CAA GAA GAA GAT CCA ATG GCA ATG AGG AGG AAC GAC CTT GAT
 N   M   G   G   L   Y   Q   E   E   D   Q   M   A   M   R   R   N   D   L   D

TTA TCT CTT GAA CCC GTT TAC AAC TGC AAC CTT GGC TGC TCT GCG GCA TGA
 L   S   L   E   P   V   Y   N   C   N   L   G   C   S   A   A   *
```

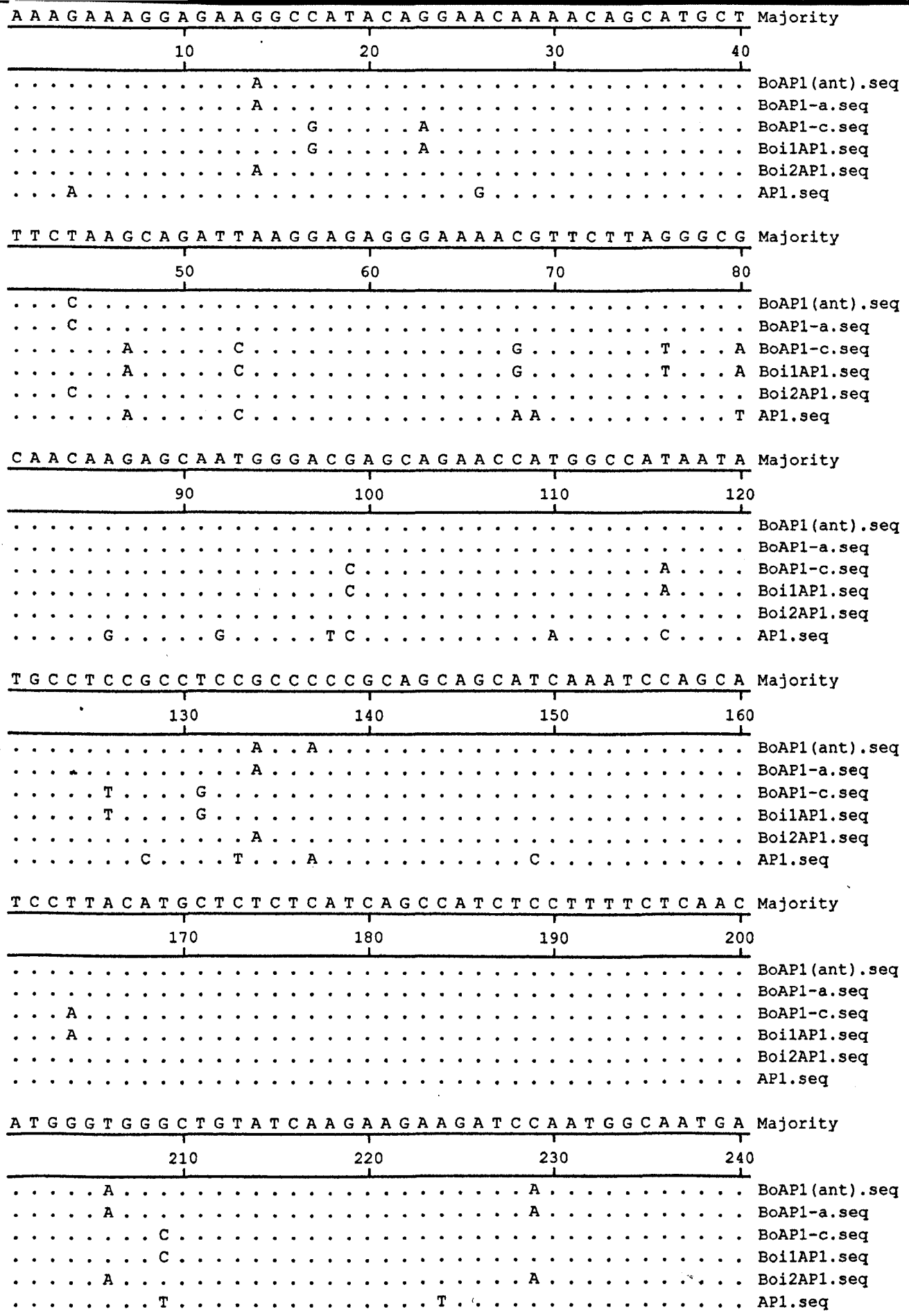
Chinese White Kale A12DH *BoAP1-a* (exonic sequence and theoretical translation)

```
ATG GGA AGG GGT AGG GTT CAG TTG AAG AGG ATA GAA AAC AAG ATC AAT AGA CAA GTG ACA
 M   G   R   G   R   V   Q   L   K   R   I   E   N   K   I   N   R   Q   V   T

TTC TCG AAA AGA AGA GCT GGT CTT ATG AAG AAA GCT CAT GAG ATC TCT GTT CTG TGT GAT
 F   S   K   R   R   A   G   L   M   K   K   A   H   E   I   S   V   L   C   D

GCT GAA GTT GCG CTT GTT GTC TTC TCC CAT AAG GGG AAA CTC TTT GAA TAC TCC ACT GAT
 A   E   V   A   L   V   V   F   S   H   K   G   K   L   F   E   Y   S   T   D

TCT TGT ATG GAG AAG ATA CTT GAA CGC TAT GAG AGA TAC TCT TAC GCC GAG AGA CAG CTT
 S   C   M   E   K   I   L   E   R   Y   E   R   Y   S   Y   A   E   R   Q   L

ATA GCA CCT GAG TCC GAC TCC AAT ACG AAC TGG TCG ATG GAG TAT AAT AGG CTT AAG GCT
 I   A   P   E   S   D   S   N   T   N   W   S   M   E   Y   N   R   L   K   A

AAG ATT GAG CTT TTG GAG AGA AAC CAG AGG CAC TAT CTT GGG GAA GAC TTG CAA GCA ATG
 K   I   E   L   L   E   R   N   Q   R   H   Y   L   G   E   D   L   Q   A   M

AGC CCT AAG GAA CTC CAG AAT CTA GAG CAA CAG CTT GAT ACT GCT CTT AAG CAC ATC CGC
 S   P   K   E   L   Q   N   L   E   Q   Q   L   D   T   A   L   K   H   I   R

TCT AGA AAA AAC CAA CTT ATG TAC GAC TCC ATC AAT GAG CTC CAA AGA AAG GAG AAA GCC
 S   R   K   N   Q   L   M   Y   D   S   I   N   E   L   Q   R   K   E   K   A

ATW CAG GAC MA? AAC AGC ATG CTT TCC AAG CAG ATT AAG GAG AGG GAA AAC GTT CTT AGG
 I   Q   D   ?   N   S   M   L   S   K   Q   I   K   E   R   E   N   V   L   R

GCG CAA CAA GAG CAA TGG GAC GAG CAG AAC CAT GGC CAT AAT ATG CCT CCG CCT TCA CCC
 A   Q   Q   E   Q   W   D   E   Q   N   H   G   H   N   M   P   P   P   S   P

CCG CAG CAG CAT CAA ATC CAG CAT CCT TAC ATG CTC TCT CAT CAG CCA TCT CCT TTT CTC
 P   Q   Q   H   Q   I   Q   H   P   Y   M   L   S   H   Q   P   S   P   F   L

AAC ATG GGG GGG CTG TAT CAA GAA GAA GAT CAA AWG GCA ATG AGG AGG AAC GAT CTC GAT
 N   M   G   G   L   Y   Q   E   E   D   Q   ?   A   M   R   R   N   D   L   D

CTG TCT CTT GAA CCC GTT TAC AAC TGC AAC CTT GGC TCG TTC GCC GCA TGA
 L   S   L   E   P   V   Y   N   C   N   L   G   S   F   A   A   *
```

Cauliflower DJ7032 *BoAP1-a* (partial exonic sequence and theoretical translation)

```
ATG GGA AGG GGT AGG GTT CAG NNG AAG ANG ATA GAA AAC AAG ATC AGT ANA CAA GTG ACA
 M   G   R   G   R   V   Q   ?   K   ?   I   E   N   K   I   S   ?   Q   V   T

TTC TCN AAA AGA ANA GCT GGT CTT ATG AAG AAN GCT CAT GAG ATC TCT GTT CTG TGT GAT
 F   ?   K   R   ?   A   G   L   M   K   ?   A   H   E   I   S   V   L   C   D

GCT GAA GTT GCG CTT GTN GTC TTC TCC CAT AAG GGG AAA CTC TNA GAA TAC TCC ACT GAT
 A   E   V   A   L   ?   V   F   S   H   K   G   K   L   ?   E   Y   S   T   D

TCT TGN NNN NNN NNN NNN NNN NAA CGC TAT GAG AGA TAC TCT TAC GCC GAG AGA CAG CTT
 S   ?   ?   ?   ?   ?   ?   ?   R   Y   E   R   Y   S   Y   A   E   R   Q   L

ATA GCA CCT GAG TCC GAC TCC AAT ACG AAC TGG TCG ATG GAG TAT AAT AGG CTT AAG GCT
 I   A   P   E   S   D   S   N   T   N   W   S   M   E   Y   N   R   L   K   A

AAG ATT GAG CTT TTG GAG AGA AAC CAG AGG CAC TAT CTT GGG GAA GAC TTG CAA GCA ATG
 K   I   E   L   L   E   R   N   Q   R   H   Y   L   G   E   D   L   Q   A   M

AGC CCT NAG GAA CTC CAG AAT CTA GAG CAA CAG CTT GAT ACT GCT CTT AAG CAC ATC CGC
 S   P   ?   E   L   Q   N   L   E   Q   Q   L   D   T   A   L   K   H   I   R

TCT AGA AAA AAC CAA CTT ATG TAC GAC TCC ATC AAT GAG CTC CAA AGA AAG GAG ANN NNN
 S   R   K   N   Q   L   M   Y   D   S   I   N   E   L   Q   R   K   E   ?   ?

NNN NNN NNN NNN NAC AGC ATG CTT TCC AAT CAG ATT AAG GAG AGG GAA AAC GTT CTT AGG
 ?   ?   ?   ?   ?   S   M   L   S   N   Q   I   K   E   R   E   N   V   L   R

GCG CAA CAA GAG CAA TGG GAC GAG CAG AAC CAT GGC CAT AAT ATG CCT CCG CCT CCA CCC
 A   Q   Q   E   Q   W   D   E   Q   N   H   G   H   N   M   P   P   P   P   P

CCG CAG CAG CAT CAN ATC CAG CAT CCT TAC ATG CTC TCT CAT CAG CCA TCT CCT TTT CTC
 P   Q   Q   H   ?   I   Q   H   P   Y   M   L   S   H   Q   P   S   P   F   L

AAC ATG GGA GGG CTG TAT CAA GAA GAA GAT CAA ATG GCA ATG AGG AGG AAC GAT CTC GAT
 N   M   G   G   L   Y   Q   E   E   D   Q   M   A   M   R   R   N   D   L   D

CTG TCT CTT GAA CCC GTT TAC AAC TGC AAC TTT GGG--- + 4 amino acids
 L   S   L   E   P   V   Y   N   C   N   F   G
```

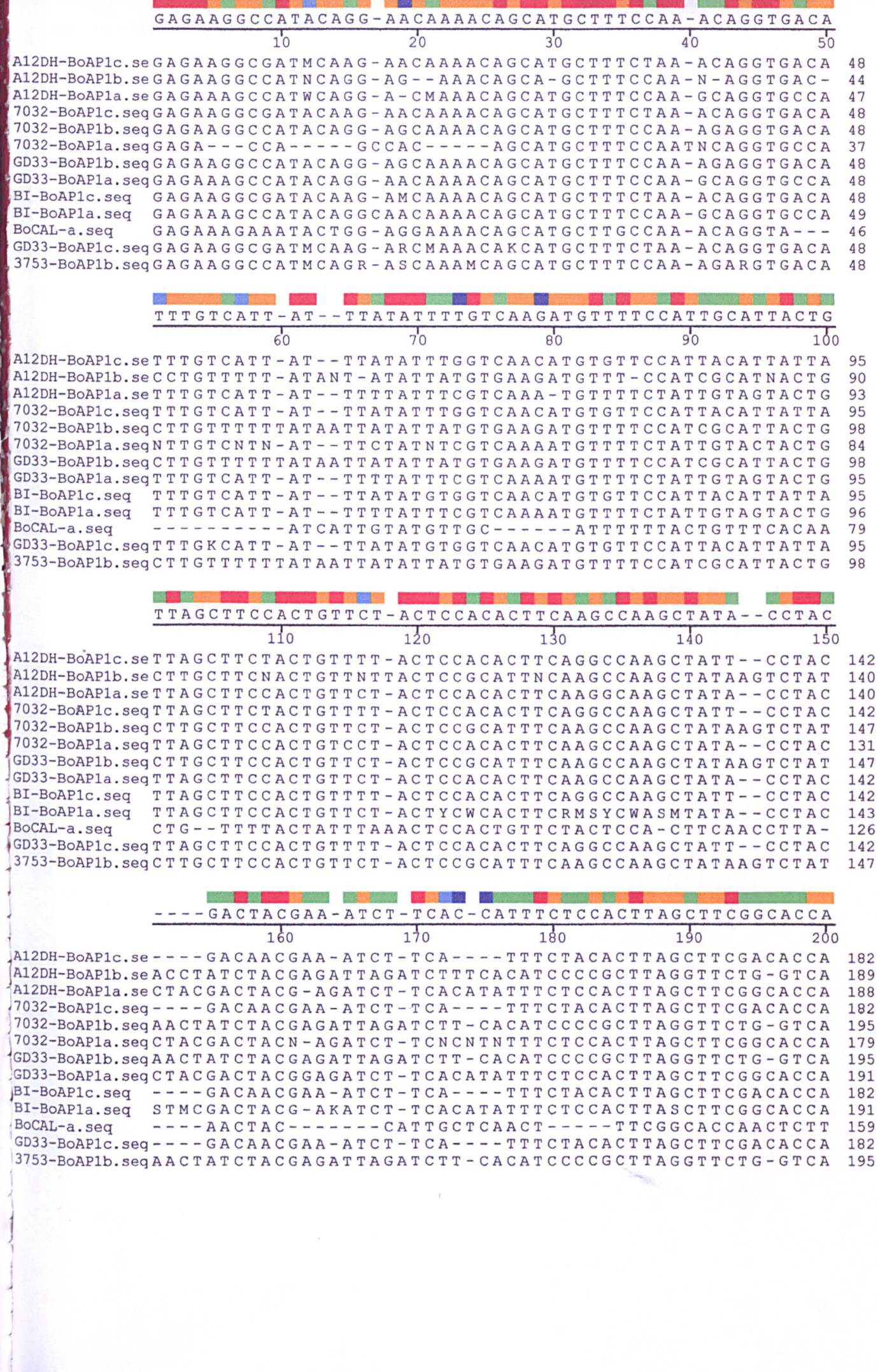
Following pages:- DNA sequence alignment from figure 4.6 and DNA sequence alignment from figure 4.7

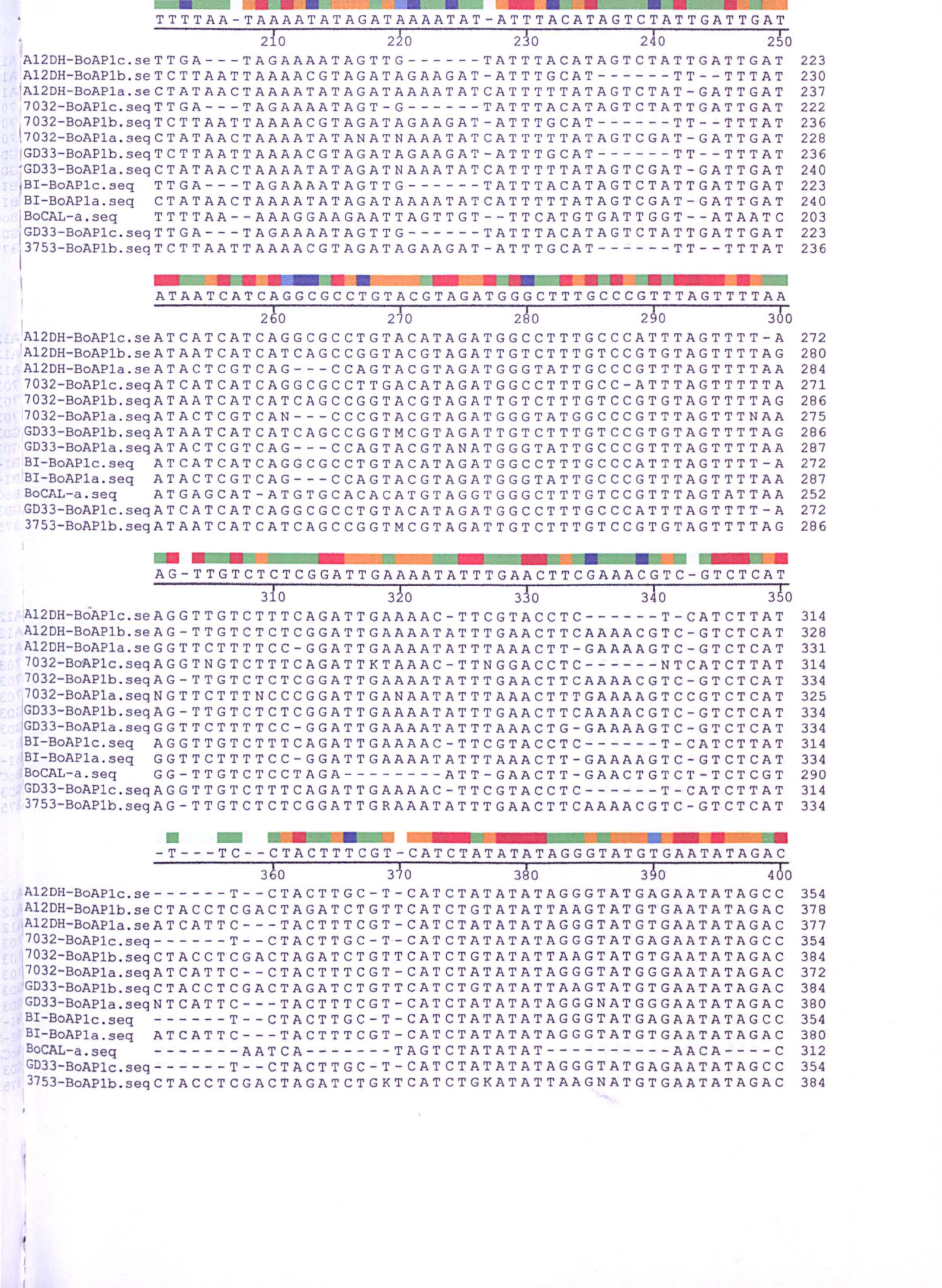

AAAGAAAGGAGAAGGCCATACAGGAACAAAACAGCATGCT Majority
10
20
30
40
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq
TTCTAAGCAGATTAAGGAGAGGGAAAACGTTCTTAGGGCG Majority
50
60
70
80
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq
CAACAAGAGCAATGGGACGAGCAGAACCATGGCCATAATA Majority
90
100
110
120
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq
TGCCTCCGCCTCCGCCCCCGCAGCAGCATCAAATCCAGCA Majority
130
140
150
160
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq
TCCTTACATGCTCTCTCATCAGCCATCTCCTTTTCTCAAC Majority
170
180
190
200
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq
ATGGGTGGGCTGTATCAAGAAGAAGATCCAATGGCAATGA Majority
210
220
230
240
BoAP1(ant).seq
BoAP1-a.seq
BoAP1-c.seq
Boi1AP1.seq
Boi2AP1.seq
AP1.seq

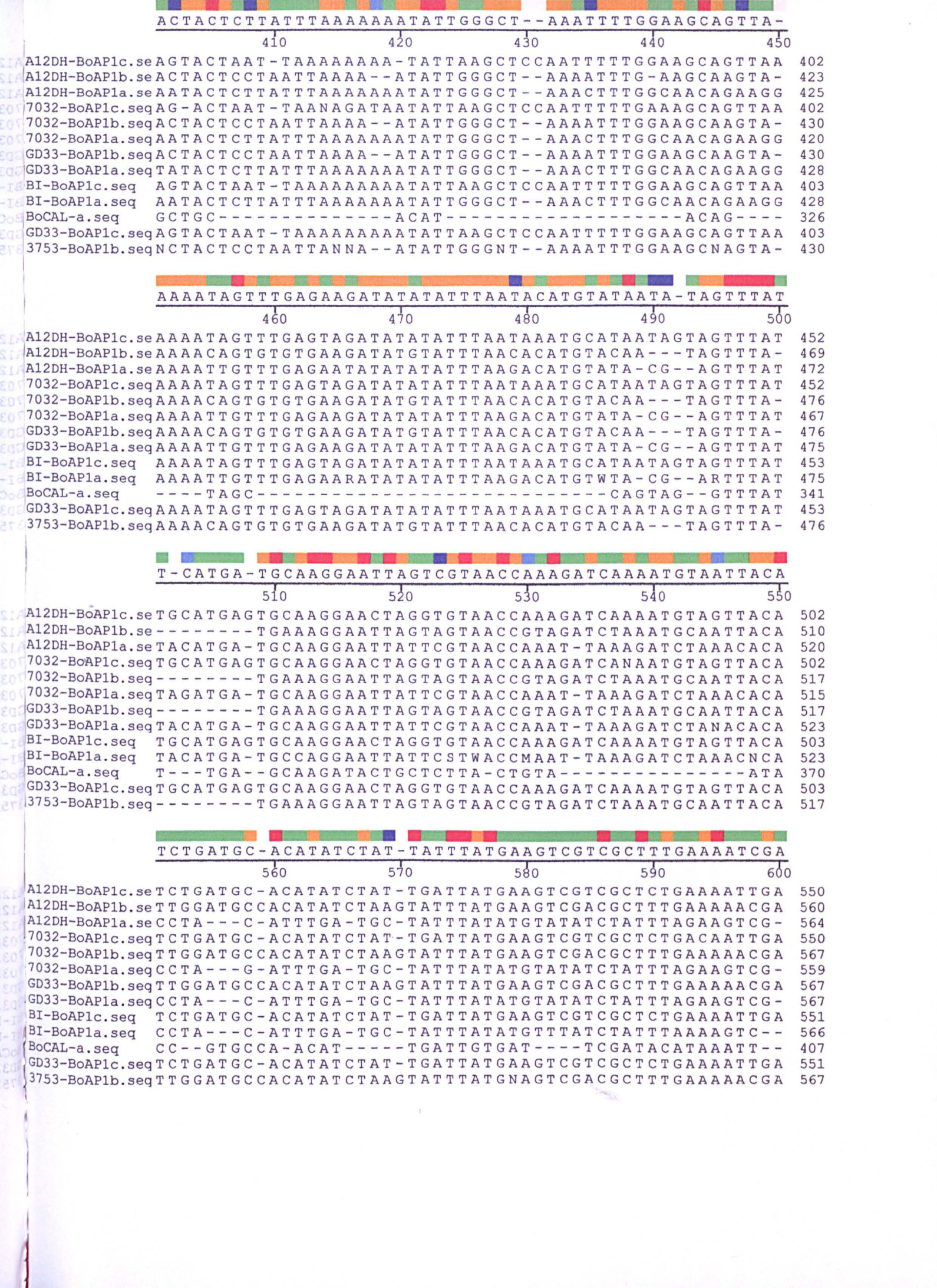
TEXT BOUND CLOSE TO THE SPINE IN THE ORIGINAL THESIS

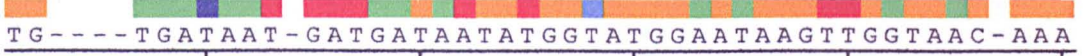
```
    G G A G G A A C G A T C T C G A T C T G T C T C T T G A A C C C G T T T A C A A  Majority
    |---------|---------|---------|---------|---------|---------|---------|---------|
              250       260       270       280
241 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  BoAP1(ant).seq
241 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  BoAP1-a.seq
241 . . . . . . . . . . C . . T . . . T . A . . . . . . . . . . . . . . . . . . .  BoAP1-c.seq
241 . . . . . . . . . . C . . T . . . T . A . . . . . . . . . . . . . . . . . . .  Boi1AP1.seq
241 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  Boi2AP1.seq
241 . . . - - - . T . . . . . . . . A . . . A . . . . . . . . . . . . . . . . . .  AP1.seq

    C T G C A A C C T T G G C X                                                    Majority
    |---------|---------|---|
              290
281 . . . . . . . . . . . . .                                                     BoAP1(ant).seq
281 . . . . . . . T . . . . G                                                     BoAP1-a.seq
281 . . . . . . . . . . . . G                                                     BoAP1-c.seq
281 . . . . . . . . . . . . .                                                     Boi1AP1.seq
281 . . . . . . . . . . . . . T                                                   Boi2AP1.seq
278 . . . . . . . . . . . . .                                                     AP1.seq
```

Decoration 'Decoration #1': Hide (as '.') residues that match the Consensus exactly.




TTTTAA-TAAAATATAGATAAAATAT-ATTTACATAGTCTATTGATTGAT
210 220 230 240 250
A12DH-BoAP1c.se TTGA---TAGAAAATAGTTG------TATTTACATAGTCTATTGATTGAT 223
A12DH-BoAP1b.se TCTTAATTAAAACGTAGATAGAAGAT-ATTTGCAT------TT--TTTAT 230
A12DH-BoAP1a.se CTATAACTAAAATATAGATAAAATATCATTTTTATAGTCTAT-GATTGAT 237
7032-BoAP1c.seq TTGA---TAGAAAATAGT-G------TATTTACATAGTCTATTGATTGAT 222
7032-BoAP1b.seq TCTTAATTAAAACGTAGATAGAAGAT-ATTTGCAT------TT--TTTAT 236
7032-BoAP1a.seq CTATAACTAAAATATANATNAAATATCATTTTTATAGTCGAT-GATTGAT 228
GD33-BoAP1b.seq TCTTAATTAAAACGTAGATAGAAGAT-ATTTGCAT------TT--TTTAT 236
GD33-BoAP1a.seq CTATAACTAAAATATAGATNAAATATCATTTTTATAGTCGAT-GATTGAT 240
BI-BoAP1c.seq TTGA---TAGAAAATAGTTG------TATTTACATAGTCTATTGATTGAT 223
BI-BoAP1a.seq CTATAACTAAAATATAGATAAAATATCATTTTTATAGTCGAT-GATTGAT 240
BoCAL-a.seq TTTTAA--AAAGGAAGAATTAGTTGT--TTCATGTGATTGGT--ATAATC 203
GD33-BoAP1c.seq TTGA---TAGAAAATAGTTG------TATTTACATAGTCTATTGATTGAT 223
3753-BoAP1b.seq TCTTAATTAAAACGTAGATAGAAGAT-ATTTGCAT------TT--TTTAT 236
ATAATCATCAGGCGCCTGTACGTAGATGGGCTTTGCCCGTTTAGTTTTAA
260 270 280 290 300
A12DH-BoAP1c.se ATCATCATCAGGCGCCTGTACATAGATGGCCTTTGCCCATTTAGTTTT-A 272
A12DH-BoAP1b.se ATAATCATCATCAGCCGGTACGTAGATTGTCTTTGTCCGTGTAGTTTTAG 280
A12DH-BoAP1a.se ATACTCGTCAG---CCAGTACGTAGATGGGTATTGCCCGTTTAGTTTTAA 284
7032-BoAP1c.seq ATCATCATCAGGCGCCTTGACATAGATGGCCTTTGCC-ATTTAGTTTTTA 271
7032-BoAP1b.seq ATAATCATCATCAGCCGGTACGTAGATTGTCTTTGTCCGTGTAGTTTTAG 286
7032-BoAP1a.seq ATACTCGTCAN---CCCGTACGTAGATGGGTATGGCCCGTTTAGTTTNAA 275
GD33-BoAP1b.seq ATAATCATCATCAGCCGGTMCGTAGATTGTCTTTGTCCGTGTAGTTTTAG 286
GD33-BoAP1a.seq ATACTCGTCAG---CCAGTACGTANATGGGTATTGCCCGTTTAGTTTTAA 287
BI-BoAP1c.seq ATCATCATCAGGCGCCTGTACATAGATGGCCTTTGCCCATTTAGTTTT-A 272
BI-BoAP1a.seq ATACTCGTCAG---CCAGTACGTAGATGGGTATTGCCCGTTTAGTTTTAA 287
BoCAL-a.seq ATGAGCAT-ATGTGCACACATGTAGGTGGGCTTTGTCCGTTTAGTATTAA 252
GD33-BoAP1c.seq ATCATCATCAGGCGCCTGTACATAGATGGCCTTTGCCCATTTAGTTTT-A 272
3753-BoAP1b.seq ATAATCATCATCAGCCGGTMCGTAGATTGTCTTTGTCCGTGTAGTTTTAG 286
AG-TTGTCTCTCGGATTGAAAATATTTGAACTTCGAAACGTC-GTCTCAT
310 320 330 340 350
A12DH-BoAP1c.se AGGTTGTCTTTCAGATTGAAAAC-TTCGTACCTC------T-CATCTTAT 314
A12DH-BoAP1b.se AG-TTGTCTCTCGGATTGAAAATATTTGAACTTCAAAACGTC-GTCTCAT 328
A12DH-BoAP1a.se GGTTCTTTTCC-GGATTGAAAATATTTAAACTT-GAAAAGTC-GTCTCAT 331
7032-BoAP1c.seq AGGTNGTCTTTCAGATTKTAAAC-TTNGGACCTC------NTCATCTTAT 314
7032-BoAP1b.seq AG-TTGTCTCTCGGATTGAAAATATTTGAACTTCAAAACGTC-GTCTCAT 334
7032-BoAP1a.seq NGTTCTTTNCCCGGATTGANAATATTTAAACTTTGAAAAGTCCGTCTCAT 325
GD33-BoAP1b.seq AG-TTGTCTCTCGGATTGAAAATATTTGAACTTCAAAACGTC-GTCTCAT 334
GD33-BoAP1a.seq GGTTCTTTTCC-GGATTGAAAATATTTAAACTG-GAAAAGTC-GTCTCAT 334
BI-BoAP1c.seq AGGTTGTCTTTCAGATTGAAAAC-TTCGTACCTC------T-CATCTTAT 314
BI-BoAP1a.seq GGTTCTTTTCC-GGATTGAAAATATTTAAACTT-GAAAAGTC-GTCTCAT 334
BoCAL-a.seq GG-TTGTCTCCTAGA--------ATT-GAACTT-GAACTGTCT-TCTCGT 290
GD33-BoAP1c.seq AGGTTGTCTTTCAGATTGAAAAC-TTCGTACCTC------T-CATCTTAT 314
3753-BoAP1b.seq AG-TTGTCTCTCGGATTGRAAATATTTGAACTTCAAAACGTC-GTCTCAT 334
-T---TC--CTACTTTCGT-CATCTATATATAGGGTATGTGAATATAGAC
360 370 380 390 400
A12DH-BoAP1c.se------T--CTACTTGC-T-CATCTATATATAGGGTATGAGAATATAGCC 354
A12DH-BoAP1b.se CTACCTCGACTAGATCTGTTCATCTGTATATTAAGTATGTGAATATAGAC 378
A12DH-BoAP1a.se ATCATTC---TACTTTCGT-CATCTATATATAGGGTATGTGAATATAGAC 377
7032-BoAP1c.seq------T--CTACTTGC-T-CATCTATATATAGGGTATGAGAATATAGCC 354
7032-BoAP1b.seq CTACCTCGACTAGATCTGTTCATCTGTATATTAAGTATGTGAATATAGAC 384
7032-BoAP1a.seq ATCATTC--CTACTTTCGT-CATCTATATATAGGGTATGGGAATATAGAC 372
GD33-BoAP1b.seq CTACCTCGACTAGATCTGTTCATCTGTATATTAAGTATGTGAATATAGAC 384
GD33-BoAP1a.seq NTCATTC---TACTTTCGT-CATCTATATATAGGGNATGGGAATATAGAC 380
BI-BoAP1c.seq------T--CTACTTGC-T-CATCTATATATAGGGTATGAGAATATAGCC 354
BI-BoAP1a.seq ATCATTC---TACTTTCGT-CATCTATATATAGGGTATGTGAATATAGAC 380
BoCAL-a.seq-------AATCA-------TAGTCTATATAT----------AACA----C 312
GD33-BoAP1c.seq------T--CTACTTGC-T-CATCTATATATAGGGTATGAGAATATAGCC 354
3753-BoAP1b.seq CTACCTCGACTAGATCTGKTCATCTGKATATTAAGNATGTGAATATAGAC 384


ACTACTCTTATTTAAAAAAATATTGGGCT--AAATTTTGGAAGCAGTTA-
410 420 430 440 450
A12DH-BoAP1c.se AGTACTAAT-TAAAAAAAA-TATTAAGCTCCAATTTTTGGAAGCAGTTAA 402
A12DH-BoAP1b.se ACTACTCCTAATTAAAA--ATATTGGGCT--AAAATTTG-AAGCAAGTA- 423
A12DH-BoAP1a.se AATACTCTTATTTAAAAAAATATTGGGCT--AAACTTTGGCAACAGAAGG 425
7032-BoAP1c.seq AG-ACTAAT-TAANAGATAATATTAAGCTCCAATTTTTGAAAGCAGTTAA 402
7032-BoAP1b.seq ACTACTCCTAATTAAAA--ATATTGGGCT--AAAATTTGGAAGCAAGTA- 430
7032-BoAP1a.seq AATACTCTTATTTAAAAAAATATTGGGCT--AAACTTTGGCAACAGAAGG 420
GD33-BoAP1b.seq ACTACTCCTAATTAAAA--ATATTGGGCT--AAAATTTGGAAGCAAGTA- 430
GD33-BoAP1a.seq TATACTCTTATTTAAAAAAATATTGGGCT--AAACTTTGGCAACAGAAGG 428
BI-BoAP1c.seq AGTACTAAT-TAAAAAAAAATATTAAGCTCCAATTTTTGGAAGCAGTTAA 403
BI-BoAP1a.seq AATACTCTTATTTAAAAAAATATTGGGCT--AAACTTTGGCAACAGAAGG 428
BoCAL-a.seq GCTGC----------ACAT----------ACAG---- 326
GD33-BoAP1c.seq AGTACTAAT-TAAAAAAAAATATTAAGCTCCAATTTTTGGAAGCAGTTAA 403
3753-BoAP1b.seq NCTACTCCTAATTANNA--ATATTGGGNT--AAAATTTGGAAGCNAGTA- 430
AAAATAGTTTGAGAAGATATATATTTAATACATGTATAATA-TAGTTTAT
460 470 480 490 500
A12DH-BoAP1c.se AAAATAGTTTGAGTAGATATATATTTAATAAATGCATAATAGTAGTTTAT 452
A12DH-BoAP1b.se AAAACAGTGTGTGAAGATATGTATTTAACACATGTACAA---TAGTTTA- 469
A12DH-BoAP1a.se AAAATTGTTTGAGAATATATATATTTAAGACATGTATA-CG--AGTTTAT 472
7032-BoAP1c.seq AAAATAGTTTGAGTAGATATATATTTAATAAATGCATAATAGTAGTTTAT 452
7032-BoAP1b.seq AAAACAGTGTGTGAAGATATGTATTTAACACATGTACAA---TAGTTTA- 476
7032-BoAP1a.seq AAAATTGTTTGAGAAGATATATATTTAAGACATGTATA-CG--AGTTTAT 467
GD33-BoAP1b.seq AAAACAGTGTGTGAAGATATGTATTTAACACATGTACAA---TAGTTTA- 476
GD33-BoAP1a.seq AAAATTGTTTGAGAAGATATATATTTAAGACATGTATA-CG--AGTTTAT 475
BI-BoAP1c.seq AAAATAGTTTGAGTAGATATATATTTAATAAATGCATAATAGTAGTTTAT 453
BI-BoAP1a.seq AAAATTGTTTGAGAARATATATATTTAAGACATGTWTA-CG--ARTTTAT 475
BoCAL-a.seq ----TAGC----------CAGTAG--GTTTAT 341
GD33-BoAP1c.seq AAAATAGTTTGAGTAGATATATATTTAATAAATGCATAATAGTAGTTTAT 453
3753-BoAP1b.seq AAAACAGTGTGTGAAGATATGTATTTAACACATGTACAA---TAGTTTA- 476
T-CATGA-TGCAAGGAATTAGTCGTAACCAAAGATCAAAATGTAATTACA
510 520 530 540 550
A12DH-BoAP1c.se TGCATGAGTGCAAGGAACTAGGTGTAACCAAAGATCAAAATGTAGTTACA 502
A12DH-BoAP1b.se --------TGAAAGGAATTAGTAGTAACCGTAGATCTAAATGCAATTACA 510
A12DH-BoAP1a.se TACATGA-TGCAAGGAATTATTCGTAACCAAAT-TAAAGATCTAAACACA 520
7032-BoAP1c.seq TGCATGAGTGCAAGGAACTAGGTGTAACCAAAGATCANAATGTAGTTACA 502
7032-BoAP1b.seq --------TGAAAGGAATTAGTAGTAACCGTAGATCTAAATGCAATTACA 517
7032-BoAP1a.seq TAGATGA-TGCAAGGAATTATTCGTAACCAAAT-TAAAGATCTAAACACA 515
GD33-BoAP1b.seq --------TGAAAGGAATTAGTAGTAACCGTAGATCTAAATGCAATTACA 517
GD33-BoAP1a.seq TACATGA-TGCAAGGAATTATTCGTAACCAAAT-TAAAGATCTANACACA 523
BI-BoAP1c.seq TGCATGAGTGCAAGGAACTAGGTGTAACCAAAGATCAAAATGTAGTTACA 503
BI-BoAP1a.seq TACATGA-TGCCAGGAATTATTCSTWACCMAAT-TAAAGATCTAAACNCA 523
BoCAL-a.seq T---TGA--GCAAGATACTGCTCTTA-CTGTA----------ATA 370
GD33-BoAP1c.seq TGCATGAGTGCAAGGAACTAGGTGTAACCAAAGATCAAAATGTAGTTACA 503
3753-BoAP1b.seq --------TGAAAGGAATTAGTAGTAACCGTAGATCTAAATGCAATTACA 517
TCTGATGC-ACATATCTAT-TATTTATGAAGTCGTCGCTTTGAAAATCGA
560 570 580 590 600
A12DH-BoAP1c.se TCTGATGC-ACATATCTAT-TGATTATGAAGTCGTCGCTCTGAAAATTGA 550
A12DH-BoAP1b.se TTGGATGCCACATATCTAAGTATTTATGAAGTCGACGCTTTGAAAAACGA 560
A12DH-BoAP1a.se CCTA---C-ATTTGA-TGC-TATTTATATGTATATCTATTTAGAAGTCG- 564
7032-BoAP1c.seq TCTGATGC-ACATATCTAT-TGATTATGAAGTCGTCGCTCTGACAATTGA 550
7032-BoAP1b.seq TTGGATGCCACATATCTAAGTATTTATGAAGTCGACGCTTTGAAAAACGA 567
7032-BoAP1a.seq CCTA---G-ATTTGA-TGC-TATTTATATGTATATCTATTTAGAAGTCG- 559
GD33-BoAP1b.seq TTGGATGCCACATATCTAAGTATTTATGAAGTCGACGCTTTGAAAAACGA 567
GD33-BoAP1a.seq CCTA---C-ATTTGA-TGC-TATTTATATGTATATCTATTTAGAAGTCG- 567
BI-BoAP1c.seq TCTGATGC-ACATATCTAT-TGATTATGAAGTCGTCGCTCTGAAAATTGA 551
BI-BoAP1a.seq CCTA---C-ATTTGA-TGC-TATTTATATGTTTATCTATTTAAAAGTC-- 566
BoCAL-a.seq CC--GTGCCA-ACAT-----TGATTGTGAT----TCGATACATAAATT-- 407
GD33-BoAP1c.seq TCTGATGC-ACATATCTAT-TGATTATGAAGTCGTCGCTCTGAAAATTGA 551
3753-BoAP1b.seq TTGGATGCCACATATCTAAGTATTTATGNAGTCGACGCTTTGAAAAACGA 567



```
                TG----TGATAAT-GATGATAATATGGTATGGAATAAGTTGGTAAC-AAA
                          610       620       630       640       650
A12DH-BoAP1c.seTGATGATGATAAT-GATG-TAATATGGCATGGAACAAGTTGGTAAC-AAA 597
A12DH-BoAP1b.seTG----------TAGATTATAATATGGTATGGAATAAGTTGATACCCNAA 600
A12DH-BoAP1a.seTGGCTTTGAAAATTGATGATGATATGGTATGGTATAAGTTGGTAMC-AAA 612
7032-BoAP1c.seqTGATGATGATAAT-GATG-TAATATGGCATGGAACAAGTTGGTANC-AAA 597
7032-BoAP1b.seqTG----------TAGATTATAATATGGTATGGAATAAGTTGATAAC-AAA 606
7032-BoAP1a.seqTGGCTTTGAAAATTGATGATGATATGGTATGGTATAAGTTGGTAAC-AAA 607
GD33-BoAP1b.seqTG----------TAGATTATAATATGGTATGGAATAAGTTGATAAC-AAA 606
GD33-BoAP1a.seqTGGCTTTGAAAATTGATGATGATATGGTATGGTATAAGTTGGTAAC-AAA 615
BI-BoAP1c.seq  TGATGATGATAAT-GATG-TAATATGGCATGGAACAAGTTGGTAAC-AAA 598
BI-BoAP1a.seq  CKGCTTTGAAAAT-GATGAAGATATNGTNWGGTATAAGTTKGTTAC-AA- 613
BoCAL-a.seq    ----------TAGTTGATCATAACGT------------TT---------A 425
GD33-BoAP1c.seqTGATGATGATAAT-GATG-TAATATGGCATGGAACAAGTTGGTAAC-AAA 598
3753-BoAP1b.seqTG----------TAGATTATNATATGGNATGGAATAAGTTGATAAC-AAA 606
```

```
                TTGGTGTGTGAAATTGAAACTTGTCAG
                          660       670
A12DH-BoAP1c.seTTGGTGTGTTAAACTGAGACTTGTCAG                        624
A12DH-BoAP1b.seTTGGTGTGTGAAATTGGAACTTGTCAG                        627
A12DH-BoAP1a.seCTGGTGTGTGAAATTGAAACTTGTCAG                        639
7032-BoAP1c.seqTTGGTGTGTTAAACTGAGACTTGTCAG                        624
7032-BoAP1b.seqTTGGTGTGTGAAATTGGAACTTGTCAG                        633
7032-BoAP1a.seqCTGGTGTGTGAAATTGAAACTTGTCAG                        634
GD33-BoAP1b.seqTTGGTGTGTGAAATTGGAACTTGTCAG                        633
GD33-BoAP1a.seqCTGGTGTGTGAAATTGAAACTTGTCAG                        642
BI-BoAP1c.seq  TTGGTGTGTTAAACTGAGACTTGTCAG                        625
BI-BoAP1a.seq  CTGGTGTGTGAAATTGAAACTTGTCCC                        639
BoCAL-a.seq    TCGGTATTTGAAATTGG-------TAG                        445
GD33-BoAP1c.seqTTGGTGTGTTAAACTGAGACTTGKCAG                        625
3753-BoAP1b.seqTTGGTGTGTGAAATTGGAACTTGTCAG                        633
```

Appendix 3: Reagent Suppliers

Amersham Life Sciences Ltd
Amersham Place
Little Chalfont
Buckinghamshire

Anachem Ltd
20 Charles Street
Luton
Bedfordshire

Biorad Laboratories
2000 Alfred Noble Drive
Hercules
California
USA

FMC Bioproducts
191 Thomaston Street
Rockland
Maine
USA

Fuji Photofilm co. Ltd
Tokyo
Japan

Genetic Research Instrumentation Ltd
Gene House
Dunmow Road
Felsted
Dunmow
Essex

Hoefer Scientific Instruments
654 Minnesota Street
San Fransisco
USA

Hybaid Ltd
111-113 Waldegrave Road
Teddington
Middlesex

Invitrogen BV
De Schelp 12 9351
NV Leek

The Netherlands

Life Technologies Ltd
3 Foutain Drive
Inchinnan Business Park
Paisley

New England Biolabs UK Ltd
Knowl Piece
Wilbury Way
Hitchin
Hertfordshire

Promega UK Ltd
Delta House
Chilworth Research Centre
Southampton

Qiagen Ltd
Boundary Court
Gatwick Road
Crawley
West Sussex

Roche Diagnostics Ltd
Bell Lane
Lewes
East Sussex

Sigma-Aldrich Company Ltd
Fancy Road
Poole
Dorset

UVP inc
5100 Walnut Grove Avenue
San Gabriel
California
USA