





Université de Lille

Ecole Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement

THÈSE

pour l'obtention du grade de

Docteur en Sciences de l'Université de Lille

Ingénierie des Fonctions Biologiques

Identification de nouveaux déterminants génétiques de la tolérance au gel chez *Pisum sativum*

par

Sana BEJI

Soutenue le 05 Mars 2019

Composition du jury :

Pr, Université de Picardie Jules Verne	Rapporteur
MdC, Université Paris Sorbonne	Rapporteur
Pr, AgroCampus Ouest	Examinateur
IR, INRA Toulouse	Examinateur
Pr, Université de Lille	Directeur de thèse
MdC, Université de Lille	Co-Directeur de thèse
CR, INRA Estrées-Mons	Invitée
	Pr, Université de Picardie Jules Verne MdC, Université Paris Sorbonne Pr, AgroCampus Ouest IR, INRA Toulouse Pr, Université de Lille MdC, Université de Lille CR, INRA Estrées-Mons







Charles Viollette Université de Lille

Ecole Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement

THÈSE

pour l'obtention du grade de

Docteur en Sciences de l'Université de Lille

Ingénierie des Fonctions Biologiques

Identification de nouveaux déterminants génétiques de la tolérance au gel chez *Pisum sativum*

par

Sana BEJI

Soutenue le 05 Mars 2019

Composition du jury :

Olivier Van Wuytswinkel	Pr, Université de Picardie Jules Verne	Rapporteur
Evelyne Téoulé	MdC, Université Paris Sorbonne	Rapporteur
Maria Manzanares-Dauleux	Pr, AgroCampus Ouest	Examinateur
Hélène Bergès	IR, INRA Toulouse	Examinateur
Jean-Louis Hilbert	Pr, Université de Lille	Directeur de thèse
Bruno Delbreil	MdC, Université de Lille	Co-Directeur de thèse
Isabelle Lejeune-Hénaut	CR, INRA Estrées-Mons	Invitée

Résumé

Chez le pois (*Pisum sativum* L.) la tolérance au gel est contrôlée par un nombre relativement restreint de loci quantitatifs (QTL, *Quantitative Trait loci*) mis en évidence par analyse de liaison dans des populations biparentales de cartographie. Le développement récent de SNP en grand nombre et d'outils de génotypage à haut débit offre la possibilité d'approfondir les études du déterminisme génétique responsables de la variation phénotypique du caractère d'intérêt. Au cours de cette thèse, nous avons étudié le déterminisme génétique de la tolérance au gel chez le pois par deux approches de génétique quantitative à savoir la cartographie de QTL par analyse de liaison, et la génétique d'association ou GWAS (*Genome Wide Association Study*).

Une re-détection de QTL de tolérance au gel a été réalisée sur un sous-ensemble de 76 lignées recombinantes (RILs, *Recombinant Inbred Lines*) issues de la population Champagne x Térèse (Pop2,164 RILs) avec 6486 marqueurs. L'analyse a permis de préciser l'intervalle de confiance des QTL détectés précédemment et d'identifier des marqueurs utilisables pour la cartographie fine. Une cartographie fine de l'un des principaux QTL préalablement identifié (WFD6.1), a été réalisée en utilisant un matériel quasi-isogénique (NILs, *Near Isogenic Lines*) à ce QTL, dont la création par rétrocroisements assistés par marqueurs avait été préalablement initiée. Les NILs étudiées ont montré un faible taux de recombinaison au niveau du locus ciblé, ce qui a réduit la résolution de la cartographie fine.

En parallèle, une analyse de génétique d'association, à partir d'une collection de 365 accessions de pois génotypés avec 11366 marqueurs SNP, a permis d'identifier 62 marqueurs en association significative avec la tolérance au gel. Ces résultats ont confirmé 3 QTL déjà détectés par analyse de liaison, dans de multiples environnements, sur les groupes de liaison (LG, Linkage Groups) III, V et VI. Ils ont également permis d'identifier un nouveau locus sur le LG II et deux loci sur les LGs I et VII, qui n'avaient auparavant été détectés que dans un seul environnement. De plus, nous avons identifié des haplotypes favorables à la tolérance au gel ainsi que les accessions représentatives de ces haplotypes. L'annotation des marqueurs candidats a permis d'identifier 50 gènes sous-jacents aux loci détectés par GWAS et ceux qui sont en déséquilibre de liaison fort (DL (r^2) > 0.8) avec les marqueurs significatifs en GWAS. Parmi ces gènes, un CBF (C-repeat Binding Factor), situé sur le groupe de liaison VI au niveau du QTL WFD6.1 a été identifié. En nous basant sur la synténie avec Medicago truncatula, nous avons émis l'hypothèse qu'il existerait un cluster de gènes CBF à cet endroit. Par conséquent, nous avons construit deux banques BAC (Bacterial Artificial Chromosome) à partir de l'ADN de Champagne et Térèse qui présentent des niveaux contrastés de tolérance au gel. Ces banques ont été criblées par des marqueurs dessinés sur les séquences des gènes CBF repérées dans le génome de référence Caméor. Les clones BAC porteurs des marqueurs cibles ont été ensuite séquencées par la technologie de séquençage PacBio®. La caractérisation des séquences obtenues fait partie des perspectives de ce travail.

Mots clés : *Pisum Sativum* L., tolérance au gel, cartographie de QTL, cartographie fine, RILs, NILs, GWAS, gènes candidats, banque BAC, gènes *CBF*.

Abstract

In pea (*Pisum sativum* L.), frost tolerance is controlled by a relatively small number of Quantitative Trait Loci (QTLs) identified by linkage analyses studies within biparental mapping populations. The recent discovery of high numbers of SNP and the development of high throughput genotyping tools offers the opportunity to further study the genetic determinism responsible for the phenotypic variation of the targeted trait. During this thesis, we investigated the genetic determinism of frost tolerance in pea by two quantitative genetic approaches namely linkage analysis QTL mapping and Genome Wide Association Study (GWAS).

A re-detection of frost tolerance QTLs was performed on a subset of 76 recombinant inbred lines extracted from the Champagne x Térèse population (Pop2, 164 RILs) and genotyped with 6486 markers. The analyses allowed to refine the confidence intervals of previously detected QTLs and to identify markers to be used in a fine mapping approach. Fine mapping of one of the formerly identified QTLs (WFD6.1) has been carried out using NILs (Near Isogenic Lines) for the targeted QTL, which creation by marker-assisted backcrossing has previously been initiated. Studied NILs showed a low rate of recombination at the targeted locus, which reduced the resolution of fine mapping.

Parallely, a genome wide association mapping, performed within a collection of 365 pea accessions genotyped with 11366 SNP markers, revealed 62 markers in significant association with frost tolerance. Results confirmed 3 QTLs already detected by linkage analyses studies on linkage groups (LGs) III, V and VI, in multiple environmental conditions. They also allowed to identify a new locus on LG II and two loci on LGs I and VII, which have formerly been detected in only one environment. In addition, GWAS allowed to identify favourable haplotypes for frost tolerance and representative accessions carrying these haplotypes. Annotation of candidate markers identified 50 genes underlying the GWAS-detected loci and markers in high linkage disequilibrium (LD (r^2) > 0.8) with associated markers. Among these genes, a particular interest for a CBF gene (C-repeat binding factor), located on the linking group VI at a position corresponding to the WFD6.1 QTL, was mentioned. On the basis of the syntenic relationship with Medicago truncatula, we hypothesized the presence of a cluster of CBF genes at this position. Therefore, we constructed two BAC (Bacterial Artificial Chromosome) libraries from the DNA of Champagne and Térèse, characterized by contrasted levels of frost tolerance. These libraries were screened by markers designed from the sequences of the CBF genes identified from the reference genome of Caméor. BAC clones carrying targeted markers were then sequenced by the PacBio® sequencing technology. Further characterization of the obtained sequences obtained is part of the perspectives of this work.

Keywords: *Pisum Sativum* L., frost tolerance, QTL mapping, fine mapping, RILs, NILs, GWAS, candidate genes, BAC library, *CBF* genes.

Remerciements

Ce travail de thèse n'aurait pas été possible sans le soutien de nombreuses personnes qui m'ont apporté leur aide directe ou indirecte durant ces années de thèse et je leur en suis redevable.

Je commencerai donc par remercier la Région Hauts-de-France, l'unité ICV-AFP pour avoir co-financé cette thèse, ainsi que le projet investissement d'avenir ANR PeaMUST pour avoir financé mes différents travaux de recherche et mes déplacements nationaux et internationaux.

Je remercie également les membres de mon jury de thèse de m'avoir fait l'honneur d'évaluer mon travail. Merci également à Marie-Laure Pilet, Nathalie Rivière, Grégoire Aubert, Gilles Boutet, Nadim Tayeh, Hélène Bergès et Fabrice Foucher, pour leur participation aux comités de suivi de thèse, leurs conseils avisés mais aussi pour avoir orienté ces travaux.

Je tiens à remercier mon directeur de thèse, Jean-Louis Hilbert, pour avoir généreusement accueilli mon projet de thèse. Mes remerciements vont également à Isabelle Lejeune-Hénaut et Bruno Delbreil, mes co-encadrants de thèse. Sans la confiance qu'ils m'ont témoignée au moment de l'inscription en thèse, je n'aurais peut-être pas sauté le pas. Ils m'ont aussi aidé à concevoir ce projet et à le mener à terme, en m'offrant les moyens nécessaires à son déroulement. Je les remercie enfin pour le temps précieux qu'ils ont pu consacrer au suivi de mon travail, et dans les corrections de ce manuscrit même pendant les vacances. Pour tout cela, je leur exprime aujourd'hui ma plus profonde reconnaissance. J'adresse aussi mes remerciements à Nasser Bahrman qui a contribué à enrichir et suivre mes travaux de thèse. Je le remercie aussi pour sa participation à l'écriture de l'article.

Je remercie vivement l'équipe AFP de l'INRA Estrées-Mons ; merci à Corinne, Odile et Emilie pour avoir contribué à la construction et le phénotypage du matériel quasi-isogénique utilisée dans mes travaux de thèse. Merci également à Martine Thomas pour tous les aspects administratifs.

Je voudrais aussi souligner, l'accueil chaleureux que j'ai reçu quand je suis arrivée au CNRGV, INRA de Toulouse, pour réaliser le criblage des banques BAC de pois. Je tiens à remercier Sonia Vautrin pour son implication dans le criblage des banques BAC et dans le séquençage des extrémités ainsi que les inserts entiers de clones BAC. Je remercie Hélène Bergès, directrice du CNRGV, pour tous ses conseils. Par ailleurs, je remercie Grégoire Aubert (Unité Agroécologie, INRA Dijon) qui a contribué à définir les séquences de marqueurs sur le génome de référence du pois pour le criblage des banque BAC.

Je remercie infiniment Raphaël, Wei, Julie, Laura, Auberi et Corinne pour leurs encouragements pour la fin de la thèse. Je leur exprime ma profonde reconnaissance et leur souhaite beaucoup de bien. Je tiens également à remercier de tout cœur Emilie qui a toujours été là lorsque le moral n'était pas au plus haut.

Mes remerciements vont enfin à ma famille, pour leur bienveillance, leur patience et leur confiance. Ce sont eux qui m'ont donné l'envie et l'énergie de faire cette course toute personnelle. Un énorme merci à mes parents qui se sont investis corps et âme pour me donner une éducation digne de ce nom et j'aimerai aujourd'hui les remercier pour tout ce qu'ils ont toujours fait pour moi. Merci à Hajer pour son soutien, pour avoir toujours cru en moi. Un merci spécial à Faiez qui m'a offert son soutien inconditionnel. Sans lui rien n'aurait été possible et ce document n'aurait pas vu le jour. Mohamed, quant à lui, a toujours su montrer qu'il était présent (des fois un peu trop) mais je pense que c'est le bon moment pour lui affirmer tout mon amour. Et bien sûr, je dédie cette thèse à mes deux adorables nièces, Yasmine et Rahma, mes plus grandes sources de bonheur, j'espère que la vie leur réserve le meilleur.

Table des matières

Contexte et objectifs de la thèse	
1. Contexte général	2
1.1. Le pois protéagineux, enjeux socio-économiques	2
1.2. Contexte scientifique de la thèse	
2. Objectifs et stratégie de la thèse	5
3. Structure du manuscrit	6
Chapitre I : Synthèse bibliographique	9
1. Comment identifier les déterminants génétiques impliqués dans la variation quantitatifs ?	n de caractères 10
1.1. Détection de QTL par analyse de liaison	
1.2. Génétique d'association	
1.3. Approche intégrative	
2. Outils génétiques et génomiques disponibles pour l'amélioration génétiqu	ue du pois 25
2.1. Ressources génétiques	
2.2. Ressources génomiques	
3. Les plantes face aux basses températures	
3.1. Les effets de basses températures sur la plante	
3.2. Stratégies de tolérance des plantes face aux basses températures	
3.3. Mécanismes d'acclimatation au froid	
3.4. Tolérance au froid du pois	
Chapitre II : Re-détection de QTL de tolérance au gel chez le pois et cartog	raphie fine 36
A. Re-detection of frost tolerance QTLs using a massive number	of markers 39
1. Material and methods	

1	.1. Plant material	
1	.2. Phenotyping	
1	.3. Phenotypic data Analysis	
1	.4. Genetic linkage map and Genotypic data	
1	.5. QTL analysis	
2.	Results	
2	2.1. Phenotyping	
2	2.2. QTL mapping for winter frost damages	
3.	Discussion	
4.	Conclusion	
B. A near	an attempt to refine the confidence interval of the frost tolerance QTL WF r-isogenic lines (NILs)	D6.1 using 50
1.	Material and methods	
1	.1. Development of NILs for WFD6.1 QTL	
1	.2. Verification of the genetic background of WFD6.1-NILs	
1	.3. Phenotyping of the NILs for winter frost damages	
2.	Results and discussion	
2	2.1. Genotype of the five near-isogenic mother plants of the BC5 generation	
2	2.2. Attempted fine-mapping of the frost tolerance WFD6.1 QTL	
3.	Conclusion	
Cha géné	pitre III : Dissection du déterminisme génétique de la tolérance au gel chez étique d'association	le pois par 64
1.	Abstract	
2.	Background	
3.	Materials and Methods	
3	3.1. Plant material	

	3.2	. Phenotyping	70
	3.3	. Phenotypic data analyses	71
	3.4	. Genotyping and quality control	71
	3.5	. Linkage disequilibrium estimation	72
	3.6	. Population structure and individual relatedness	72
	3.7	. Association mapping	.73
	3.8	. Local LD bloc estimation and favorable allele identification	.73
	3.9	. Identification of the frost tolerance candidate gene	.74
4	•	Results	.74
	4.1	. Statistical analyses of phenotypic data	.74
	4.2	. SNP genotyping of the association mapping collection	.75
	4.3	. LD analysis	.77
	4.4	. Population structure and kinship analyses	.77
	4.5	. Genome wide association mapping	.78
	4.6	. Favourable alleles exploration for frost tolerance in pea	. 84
	4.7	. Candidate gene analysis	. 84
5	•	Discussion	85
	5.1	. GWAS brings new insights into the determinism of frost tolerance in pea	. 85
	5.2. can	. GWAS detected frost tolerance-associated markers which are relevant putat didate genes	tive 89
	5.3	. GWAS provides new markers and new genitors to breed for frost tolerance in pea	.92
6	-	Conclusion	.93
7	•	Availability of supporting data	.94
8		Abbreviations	.94
9		Competing interests	.95

10. Authors' contributions	
11. Acknowledgements	
12. References	
Chapitre IV : Création de banques BAC pour la recherche de gènes jacents au QTL WFD6.1	candidats sous- 105
1. Introduction du chapitre	
2. Matériels et Méthodes	
2.1. Production du matériel végétal pour la construction des banques BAC	
2.2. Construction des banques BAC	
2.3. Estimation de la taille d'insert	
2.4. Criblage des banques BAC	
2.4.1. Méthodes d'identification de clones BAC positifs par hybridation	n de macroarrays
2.4.1.1. Production des membranes (macroarrays)	110
2.4.1.2. Préparation des sondes	111
2.4.1.3. Hybridation	
2.4.1.4. Analyse des résultats qualitatifs	113
2.4.1.5. Réarrangement des clones d'intérêts identifiés par criblage de validation par PCR	s macroarrays et
2.4.2. Méthodes de criblage de la banque BAC non réarrangée	114
2.5. Séquençage des BAC	116
2.5.1.Méthode de séquençage des	BAC-ends
2.5.2.Méthode de séquençage des clones BAC et traitement des	données brutes 116
2.6. Dessin d'amorces pour le contigage des BAC chevauchants	117
3. Résultats et discussions	

3.1. Caractérisation des deux banques BAC	117
3.2. Caractérisation des marqueurs utilisés pour le criblage des banques BAC	119
3.3. Criblage des banques BAC	122
3.4. Séquençage des clones identifiés par criblage des banques BAC	125
4. Conclusion	128
Conclusion générale	129
Références bibliographiques	133
Annexes	144
Annexe A : Additional files du chapitre III	145
Annexe B : Données supplémentaires du chapitre IV	188

Table des figures et des tableaux

Contexte et objectifs de la thèse

Figure I. 9 : Sché	ma présentant le processu	s de déshydratation	cellulaire causé par le gel	. Extrait
de Sharma et al.	(2005)			

Chapitre II : Re-détection de QTL de tolérance au gel chez le pois et cartographie fine

Figure II. 1: Strategy adopted for the fine mapping of the frost tolerance locus WFD6.1 38

 Table II. 1: List of QTLs identified for winter frost damage in *Pisum sativum* detected in 11

 experiment conditions

 46

 Table II. 2: Characteristics of markers used in the genotyping analysis of NILs in BC5

 generation

 58

 Table II. 3: Evaluation of WFD6.1 introgression and the return to Eden genetic background 59

Table II. 4: Phenotypic and genotypic evaluation of studied NILs under Theix/2014-2015 andOrsonville /2017-2018 experimental conditions

 Table II. 5: Analyses of variance (mean square) for winter frost damage scores of studied pea
 62

Chapitre III : Dissection du déterminisme génétique de la tolérance au gel chez le pois par génétique d'association

Figure III. 2: Comparative genetic map of genome wide association study (GWAS) loci identified in the present study and quantitative trait loci (QTL) previously detected for frost Table III. 2 : Bic-based comparison of the four models used to control the rate of false positive Table III. 3: Significant associations detected in the association mapping collection (363 Chapitre IV : Création de banques BAC pour la recherche de gènes candidats sousjacents au QTL WFD6.1 Figure IV. 2 : Production de membranes de nylon (macroarrays) de 22 x 22 cm avec un motif Figure IV. 3 : Stratégie de criblage des banque BAC en pool (extrait du site web du Centre National de Ressources Génomiques Végétales (CNRGV) (www.cnrgv.toulouse.inra.fr)...116 Figure IV. 4 : Résultats de migration sur gel d'agarose des produits de PCR des deux marqueurs CBF303_4831 et CBF303_4830 testés pour la température d'hybridation de 66°C......120 Figure IV. 5 : Résultats de migration sur gel d'agarose des produits de PCR des quatre marqueurs CBF426 6484, CBF303 4832, CBF211 3604 et CBF211 3609 testés pour la Figure IV. 6 : Résultat du criblage des banques BAC de Champagne et Térèse pour le marqueur Figure IV. 7 : Bilan des résultats du criblage des banques BAC (classiques et en pool) de Figure IV. 9 : Résultats des expériences de contiguage PCR pour vérifier le chevauchement des clones BAC entre les marqueurs CBF303 4832 et CBF303 4834 et entre les marqueurs CBF211 3604 et CBF211 3605 pour les deux génotypes......127 Tableau IV. 1 : Marqueurs utilisés pour le criblage des banques BAC de Champagne et Térèse Tableau IV. 2 : Série d'hybridations utilisées pour les banques BAC de Champagne et Térèse

Tableau IV. 3 : Caractéristiques des banques BAC réarrangées de Champagne et Térèse ... 118

Tableau IV. 4 : Caractéristiques des banques BAC en pool de Champagne et Térèse	119
Tableau IV. 5 : Conditions d'amplification par PCR des amorces utilisées	121
Tableau IV. 6 : Caractéristiques des données d'assemblage des séquences de la liste des c BAC identifiés et validé par PCR	lones

Liste des abréviations

ABA : Acide abscissique ABRE: Responsive Element Binding Factor ADN ou DNA Acide désoxyribonucléique AG : Acide Gibbérellique AP2/EREBP : APETALA2/Ethylene Responsive Element Binding Protein ARR: type-A Arabidopsis Response Regulator BAC: Bacterial Artificial Chromosome **BC:** Backcross **BLUPs: Best Linear Unbiased Predictions** BR : Brassinostéroïdes BZR1: BrassinaZole-Resistant 1 CBF: C-repeat Binding Factor **CIM:** Composite Intervel Mapping cM: centi Morgan COR: COld Regulated CRT/DRE : C-RepeaT/Dehydration-, high salt-, low temperature-Responsive Element Ct: Cycle Threshold CVg: Coefficient of Variation of the genetic variance DAPC : Discriminant Analysis of Principal Components DL ou LD : Déséquilibre de Liaison DREB: Dehydration Responsive Element Elf3: Early flowering 3 FD: Frost Damage Gpb : Giga base pairs GWAS: Genome Wide Association Study H² : heritabilité au sens large HD : Haploïde Double Hr : High response to photoperiod HSP (Heat Shock Protein) IC ou CI : Interval de Confiance INRA : Institut National de Recherche Agronomique

ICV-AFP : Institut Charles Viollette- Adaptation au Froid du Pois

JAZ : JAsmonate-Zim-domain K : matrice de Kinship LB : milieu de culture de Lysogeny Broth LD-LA: Linkage Disequilibrium-Linkage Analysis LEA: Late Embryogenesis Abundant LG: Linkage group LM: Linear model LMM: Linear Mixed Model LOD score Logarithm of Odds score LSMeans: Least square means MAGIC: Multi-parent Advanced Generation Inter-Cross NAM: Nested Association Mapping NGS: Next Generation Sequencing NIL: Near Isogenic Line Q : matrice de structure QTL : Quantitative Trait Locus ou Quantitative Trait Loci, suivant le contexte pb : paire de bases PCR: Polymerase Chain Reaction **RIL:** Recombinant Inbred Line **ROS: Reactive Oxygen Species** SAM ou MAS : Sélection Assistée par Marqueurs SE: Standard Error SNK: Student-Newman-Keuls SNP: Single Nucleotide Polymorphism SSR: Simple Sequence Repeat TILLING: Target Induced Local Lesions IN Genomes Tm: Melting temperature ou Température de fusion UMR : Unité Mixte de Recherché USC : Unité Sous Contrat Vg: variance génétique Vp: variance phénotypique WFD: Winter Frost Damage YAC: Yeast Artificial Chromosome

Contexte et objectifs de la thèse

1. Contexte général

1.1. Le pois protéagineux, enjeux socio-économiques

Le pois protéagineux, *Pisum sativum* L., constitue une source de protéines pour l'alimentation humaine et animale. En Europe, la culture du pois est un enjeu économique très important pour l'alimentation animale car elle pourrait limiter la dépendance européenne vis-à-vis du soja importé. Grâce à leur forte teneur en amidon (45 %) et en protéines (21 %), les pois secs sont des graines riches en énergie. Plus encore, leurs protéines sont particulièrement riches en lysine, un acide aminé essentiel dont les céréales sont peu pourvues. Ces deux éléments sont tout aussi essentiels à la croissance des animaux d'élevage (http://www.terresunivia.fr). De plus cette culture est importante d'un point de vue agronomique et environnemental. En effet, sa capacité à fixer l'azote atmosphérique génère un double intérêt, économique par une moindre utilisation d'engrais azotés de synthèse, et écologique par une limitation des fuites azotées et des gaz à effet de serre produits lors de la fabrication des engrais.

La France est le premier producteur du pois protéagineux en Europe avec 26 % de la production en 2017 (http://www.terresunivia.fr) (Figure 1). En France, le pois de printemps est le type cultural actuellement prédominant. Cependant, certains aspects négatifs sont liés au semis de printemps, notamment la durée relativement courte du cycle de développement qui limite la production de biomasse et donc le rendement, et l'occurrence en fin de cycle de stress hydriques et thermiques liés aux fortes températures qui sont à l'origine de fluctuations des rendements (Figure 2). Pour améliorer la productivité et surtout régulariser le rendement, la sélection s'est orientée vers le semis du pois protéagineux en automne pour allonger le cycle de développement tout en évitant les périodes de stress hydriques et thermiques en fin de cycle. De ce fait, le pois d'hiver peut être confronté à un facteur limitant spécifique de sa période de semis, à savoir les températures gélives en début de cycle et en particulier à partir de l'initiation florale, qui marque le début d'une période de plus grande sensibilité au gel. Le développement de variétés de pois d'hiver nécessite donc l'amélioration de la tolérance aux basses températures.



Figure 1 : Principaux pays producteurs de protéagineux et du soja dans l'union européenne en 2017 (en 1000 t) (http://www.terresunivia.fr).



Figure 2 : Évolution du rendement du pois protéagineux en France de 1961 à 2017 (FAOSTAT 2018).

1.2. Contexte scientifique de la thèse

Dans ce contexte, l'identification des déterminants génétiques responsables de la tolérance au gel chez le pois est une étape essentielle pour la mise en œuvre d'un programme de sélection assistée par marqueurs de lignées de pois améliorées pour ce caractère. Ce programme de recherche des déterminants génétiques est mis en œuvre par l'unité ICV-AFP (Institut Charles Viollette - Adaptation au Froid du Pois) et le programme d'innovation variétale est porté par

l'unité expérimentale GCIE Picardie (Grandes Cultures Innovation Environnement) sur le site INRA d'Estrées-Mons. Dans un premier temps, des détections de QTL (*Quantitative Trait Loci*) impliqués dans la tolérance au gel chez *Pisum sativum* ont été réalisées à partir d'expérimentations au champ (Lejeune-Hénaut et al. 2008; Klein et al. 2014) et en conditions contrôlées (Dumont et al. 2009) pour deux populations de lignées recombinantes issues respectivement d'un croisement entre Champagne et Térèse d'une part (Pop2, Lejeune-Hénaut et al. 2008; Dumont et al. 2009) et d'un croisement entre China et Caméor d'autre part (Pop9, Klein et al. 2014), en collaboration avec l'équipe INRA de l'unité Agroécologie de Dijon. Ces études ont abouti à l'identification de 5 principaux QTL dont 3 sont communs entre les trois séries d'expérimentations : WFD3.1, WFD5.1 et WFD6.1, localisés respectivement au niveau des groupes de liaison III, V et VI (Figure 3).



Figure 3 : Principaux QTL de tolérance au gel identifiés sur chez le pois par analyse de liaison. (D'après Lejeune-Hénaut et al. 2008; Dumont et al. 2009; Klein et al. 2014).

Les QTL ont été détectés à partir de deux populations de lignés recombinantes issues de croisement biparentaux entre une lignée tolérante et une lignée sensible au gel : Pop2 (Champagne x Térèse) et Pop9 (China x Caméor).

Dans un second temps, un programme de validation des trois QTL communs, mais aussi une cartographie fine de ces QTL, destinée à identifier les gènes candidats sous-jacents, ont été

engagés au sein de l'unité ICV-AFP. Ces actions s'appuient sur un matériel quasi-isogénique aux QTL, dont la création par rétrocroisements assistés par marqueurs a été initiée au sein du projet SAMPOIS (2007-2011) et poursuivie *via* le projet Investissement d'Avenir ANR-PeaMUST (2012-2020).

En parallèle, d'importantes ressources génétiques et génomiques ont été apportées à la communauté de recherche sur le pois par les unités Agroécologie (INRA, Dijon) et IGEPP (Institut de Génétique, Environnement et Protection des Plantes, INRA, Le Rheu). Ainsi, une carte génétique consensus de 15079 marqueurs incluant en particulier 12941 marqueurs SNP (*Single Nucleotide Polymorphism*) issus de la puce GenoPea Infinium® et 1339 SNP issus du projet PEAPOL (Duarte et al. 2014) a été récemment construite à partir d'environ 1440 génotypes de pois issus de 12 populations de lignées recombinantes (*Recombinant Inbred Lines* : RILs) (Tayeh et al. 2015a). En plus des ressources génomiques, l'unité Agroécologie a rassemblé une importante collection de ressources génétiques de pois comportant plus de 1000 accessions y compris des variétés, des mutants et des formes sauvages. L'ensemble de ces ressources rend possible l'identification et l'analyse de la variation nucléotidique à grande échelle chez le pois par l'approche de génétique d'association du génome entier (*Genome Wide Association Study* : GWAS).

2. Objectifs et stratégie de la thèse

Le travail de ma thèse s'inscrit dans le cadre général des recherches pour l'amélioration de la tolérance au gel chez le pois, qui sont menées au sein de l'unité ICV-AFP dans une équipe mixte INRA/Université de Lille. Le contexte général était d'exploiter les nouvelles ressources génétiques et génomiques développées chez *Pisum sativum* pour approfondir l'étude du déterminisme génétique de la tolérance au gel et identifier des stratégies pour accéder aux gènes candidats sous-jacents aux QTL impliqués dans le déterminisme du caractère.

Les objectifs scientifiques de ce travail ont été fixés comme suit :

- localiser plus précisément les QTL préalablement cartographiés, en considérant la disponibilité récente des marqueurs de type SNP en forte densité sur le génome du pois ;
- identifier des allèles et des génotypes favorables à la tolérance au gel par analyses de génétique d'association dans une collection de ressources génétiques de pois préalablement phénotypée et récemment génotypée à haute densité ;

 identifier par l'approche de génétique d'association des gènes candidats impliqués dans la tolérance au gel chez le pois.

Suite à ces premières analyses à l'échelle du génome, il a été décidé de se focaliser sur :

- la réduction de l'intervalle de confiance de l'un des principaux QTL de tolérance au gel localisé sur le groupe de liaison VI du pois (WFD6.1), en utilisant un matériel quasiisogénique au QTL d'intérêt en vue d'initier un clonage positionnel ;
- la construction des banques BAC (*Bacterial Artificial Chromosome*) de deux lignées de pois de tolérance contrastée au gel, afin d'établir une séquence physique de la zone réduite de WFD6.1 permettant d'identifier les gènes sous-jacents.

3. Structure du manuscrit

Après un premier chapitre dédié à une synthèse bibliographique, je présenterai la démarche de mon travail de thèse suivant une organisation en trois parties (Figure 4) : la re-détection de QTL de tolérance au gel chez le pois et la cartographie fine du QTL WFD6.1 (chapitre II), la dissection du déterminisme génétique de la tolérance au gel chez le pois par génétique d'association (chapitre III), et le développement de banques BAC pour la recherche de gènes candidats sous-jacents au QTL WFD6.1 (chapitre IV).

La synthèse bibliographique fournit les bases de génétique nécessaires à la compréhension des enjeux du travail qui a été réalisé pendant ma thèse. J'y présente les deux méthodes d'analyse utilisées pour étudier le déterminisme génétique d'un caractère quantitatif comme la tolérance au gel et visant à identifier des régions génomiques impliquées dans la variation phénotypique, à savoir la cartographie de QTL, d'une part, et l'analyse de génétique d'association, d'autre part. Les grands principes du clonage positionnel de QTL sont aussi évoqués. J'aborderai également les réponses mises en œuvre par les plantes pour répondre au stress froid.

Le deuxième chapitre est structuré en deux grandes parties : la première est consacrée à la redétection de QTL de tolérance au gel en utilisant un grand nombre de marqueurs pour tenter de réduire les intervalles de confiance des QTL précédemment identifiés, la seconde partie est consacrée à la validation de l'effet de l'un des principaux QTL, WFD6.1, localisé sur le groupe de liaison VI avec des lignées quasi-isogéniques (*Near Isogenic Lines* : NILs) et à la cartographie fine de ce QTL, cette dernière approche étant également destinée à la réduction de l'intervalle de confiance. Le troisième chapitre de ce mémoire est consacré à l'identification des loci responsables du contrôle génétique de tolérance au gel chez le pois, par une approche de génétique d'association. Cette dernière est basée sur la variabilité génétique pour la tolérance au gel observée dans une collection de ressources génétiques comportant 365 accessions de pois représentant des origines géographiques diverses. Dans ce chapitre il s'agit d'identifier des associations entre des variations de séquences nucléotidiques de gènes et l'intensité des dégâts de gel et de définir des haplotypes de marqueurs favorables à la tolérance ainsi que des géniteurs potentiels pour la sélection de la tolérance au gel chez le pois. L'originalité de ces travaux réside dans l'approche de génétique d'association utilisant du matériel génétique diversifié et des données de génotypage plus denses. Les résultats de ce chapitre sont rédigés sous forme d'un article qui va être soumis pour publication à *BMC Genomics*.

Dans le quatrième chapitre, je présente les premiers éléments d'une approche de construction de banques de BAC qui a été entreprise compte-tenu des résultats des deux premiers chapitres. Les objectifs, qui seront atteints à la suite de ma thèse, sont d'établir la carte physique de la zone réduite du QTL WFD6.1 pour un génotype de pois sensible au gel et un génotype tolérant, d'identifier les gènes sous-jacents et d'étudier leurs différences structurales.

Une conclusion générale et des perspectives de poursuite de mon travail sont proposées à la fin de ce manuscrit.



Figure 4 : Stratégie adoptée pour les travaux de thèse et correspondance avec les chapitres du manuscrit.

Chapitre I : Synthèse bibliographique

La première partie de cette étude bibliographique présentera les méthodes d'identification des déterminants génétiques impliqués dans la variation phénotypique d'un caractère quantitatif. La deuxième partie présentera l'espèce *Pisum sativum* (L.), matériel d'étude de la thèse, ainsi que les différentes ressources biologiques et génomiques disponibles pour l'amélioration génétique du pois. La dernière partie se penchera sur les effets du stress froid sur les plantes et sur les réponses physiologiques et moléculaires mises en œuvre par ces dernières.

1. Comment identifier les déterminants génétiques impliqués dans la variation de caractères quantitatifs ?

L'amélioration génétique des plantes a pour objectif d'associer dans un même génotype le maximum d'allèles favorables. Cet objectif peut être particulièrement difficile à atteindre pour les caractères complexes, polygéniques, dits caractères quantitatifs. Le développement des marqueurs génétiques, avec d'abord les isozymes puis les marqueurs moléculaires, marque une étape importante de l'évolution des méthodes d'amélioration des plantes. Les marqueurs moléculaires en grand nombre permettent d'identifier les déterminants génétiques des caractères quantitatifs et peuvent être utilisés pour sélectionner rapidement les génotypes qui possèdent les caractéristiques recherchées.

Les principes de détection de régions génomiques liées à la variation phénotypique des caractères quantitatifs (QTL, *Quantitative Trait Loci*), sont connus depuis plusieurs années mais font intervenir des calculs très lourds. Ce sont les progrès de l'informatique avec la naissance des microprocesseurs en 1971 qui ont permis le développement des biostatistiques. Cette amélioration du temps de calcul est contemporaine de l'émergence de la biologie moléculaire moderne et en combinant ces deux avancées les scientifiques ont pu développer des cartes génétiques haute densité. Un enjeu majeur a alors été de cartographier des QTL sur ces cartes génétiques puis d'identifier les gènes, responsables de ces variations.

Dans cette partie bibliographique nous présenterons, les principales méthodes d'analyse utilisées pour étudier le déterminisme génétique d'un caractère quantitatif. Selon le matériel caractérisé, deux approches principales permettent la recherche des associations entre la variation phénotypique d'un caractère d'intérêt et une région génomique : en premier lieu la cartographie par analyse de liaison génétique dans des descendances de croisements contrôlés, en second lieu la cartographie par analyse du déséquilibre de liaison au sein de collections de ressources génétiques, dite aussi cartographie par analyse d'association.

1.1. Détection de QTL par analyse de liaison

La cartographie a pour objectif de localiser un locus (ou un ensemble de loci) dont la variation allélique est associée à la variation d'un caractère quantitatif. Comme les allèles des QTL ne sont pas identifiables (leur position est à priori inconnue), l'information disponible doit venir de marqueurs moléculaires proches (dont les localisations sont connues grâce à la carte génétique). On conclut à la présence d'un QTL quand sont observées des différences phénotypiques significatives entre individus qui ont hérité différents allèles de marqueurs de leurs parents.

La première étape utilisée afin d'identifier avec précision une association entre les marqueurs génétiques et le phénotype d'intérêt est la réalisation d'une carte génétique (Figure I.2). Pour cela, une population de cartographie doit être construite à partir d'une descendance issue d'un croisement entre deux parents caractérisés par une diversité phénotypique pour le caractère d'intérêt. La cartographie de QTL est donc réalisée avec des populations en ségrégation. Différents types de population de cartographie de QTL peuvent être utilisés (Figure I.1) comme des populations F2, des populations d'haploïdes doublés (HD), de lignées recombinantes (Recombinant Inbred Lines : RILs), ou de rétrocroisement avec un parent récurrent (backcross : BC). Cependant, il est conseillé d'utiliser les populations de RILs ou de lignées HD pour les analyses génétiques car elles sont plus stables que les populations F2 et sont immortelles (Ferreira et al. 2006). Une population de RILs est un ensemble de lignées homozygotes dont chacune contient une combinaison chromosomique unique issue de la recombinaison des gamètes parentaux. Ce type de population est produit par autofécondations successives d'un hybride F1. Les lignées HD sont produites par autofécondation d'une plante régénérée à partir de cellules haploïdes reproductrices, souvent des grains de pollen d'un individu F1, ayant subi un doublement chromosomique (revue par Collard et al. 2005; Li et al. 2005; Huang et al. 2006).

Les individus de la population étudiée doivent être ensuite génotypés à l'aide d'un grand nombre de marqueurs moléculaires. Il est important que les deux parents de la population de cartographie soient polymorphes non seulement au niveau phénotypique mais aussi au niveau moléculaire afin de pouvoir suivre la ségrégation des marqueurs moléculaires. La ségrégation des marqueurs par recombinaison, durant la méiose subie par les individus de la population de cartographie, est détectée et une carte génétique peut alors être construite. Le nombre des évènements de recombinaison entre deux marqueurs est transformé en distance génétique additive (Collard et al. 2005).



Figure I. 1 : Principaux types de populations biparentales pour l'analyse de liaison. Extrait de Collard et al. (2005).



Détection de QTL

Une carte génétique est construite avec tous les marqueurs. Différentes méthodes statistiques permettent ensuite de déterminer la localisation de chaque QTL.

Cartographie fine

Les régions contenant des QTL sont densifiées avec de nouveaux marqueurs moléculaires. De nouveaux individus sont phénotypés et génotypés afin d'augmenter la probabilité de trouver des évènements de recombinaison entre les marqueurs et le QTL.

Cartographie physique

- La séquence du génome du pois n'est pas complètement disponible
- → Construction d'une banque d'ADN génomique (BAC)
- Criblage de la banque avec les marqueurs les plus proche encadrant le QTL
- Séquençage et assemblage des clones d'ADN
- Construction d'une carte physique locale du locus d'intérêt
- Caractérisation des gènes présents dans ce locus

Clonage

Il y a clonage du gène lorsque celui qui est responsable du phénotype est identifié.

Figure I. 2 : Différentes étapes dans la détermination des bases moléculaires d'un caractère quantitatif (la tolérance au gel) par cartographie génétique via la recherche d'association entre marqueurs et caractère.

La population de cartographie doit également être phénotypée pour le caractère d'intérêt si possible dans des environnements multiples afin de prendre en compte la variation environnementale (Figure I.2).

La cartographie de QTL par analyse de liaison génétique s'effectue par la détection d'association entre le phénotype et le génotype du marqueur. Chaque marqueur permet séparer



une population de RILs en deux groupes caractérisés par leur génotype au locus marqueur. Si le locus marqueur et le QTL sont très proches, la probabilité de leur recombinaison est très faible et ils sont transmis ensemble dans la descendance lors de la méiose. Le principe de la détection de QTL est d'observer dans la descendance RIL, s'il existe ou non une différence entre les moyennes phénotypiques des deux groupes caractérisés par un allèle différent au locus marqueur. Si cette différence existe, on considère que le QTL et le locus marqueur sont liés, ce qui permet de situer le QTL sur la carte génétique.

Différentes méthodes statistiques peuvent être employées pour la détection de QTL. La méthode la plus simple consiste en une analyse marqueur par marqueur (*Single marker*). On dispose, pour u ensemble de marqueur des informations de génotypage de chaque individu. Pour chaque marqueur, les individus sont classés en fonction de leur génotype. Une analyse de variance à un facteur (le marqueur testé) est réalisée en comparant les moyennes phénotypiques des deux classes génotypiques pour le caractère mesuré. Une différence significative entre les moyennes phénotypiques des classes indique que le marqueur utilisé est lié à un QTL contrôlant le trait. Le coefficient de détermination (R²) associé au marqueur évalue la part de la variation phénotypique expliquée par le QTL pour le caractère (Collard et al. 2005).

Une seconde méthode de détection de QTL est la cartographie d'intervalle simple (*Simple Interval Mapping* : SIM). Il s'agit de comparer les classes de la population correspondant aux allèles de deux marqueurs adjacents, ce qui permet de tenir compte du taux de recombinaison entre ces deux marqueurs. Cette méthode compense les données manquantes entre deux marqueurs en rajoutant une probabilité d'avoir un génotype en chaque position de la carte génétique, par exemple tous les 1 cM. Cette approche est basée sur la méthode du maximum de vraisemblance pour estimer la position et l'effet du QTL. A chaque position sur la carte, le rapport entre la vraisemblance de l'hypothèse « présence d'un QTL » et l'hypothèse nulle (absence d'un QTL) est calculé. Ce rapport, ou plus généralement sa fonction logarithmique, le LOD score (*Logarithm of odds*), est alors représenté graphiquement en fonction du paramètre de position (Figure I.3) (Charmet 2011). La position la plus probable du QTL est celle qui maximise la vraisemblance du modèle avec présence d'un QTL, ce qui correspond à une valeur maximale du LOD score (Charmet 2011). La valeur seuil, à partir de laquelle un QTL est considéré comme significatif, peut être déterminée en réalisant un certain nombre de permutations pour chaque population et chaque trait (Szalma et al. 2007). L'intervalle de

confiance du QTL est déterminé lorsque la valeur du LOD est abaissée de 1 par rapport à la position correspondant au LOD maximum (LOD drop off) (Charmet 2011).



Figure I. 3 : Représentation graphique de la statistique LOD score le long d'un chromosome par cartographie d'intervalle avec le logiciel MultiQTL. Extrait de Charmet (2011).

Cependant, la cartographie d'intervalle simple suppose qu'il existe au plus un QTL dans l'intervalle entre deux marqueurs successifs, et même pour un seul chromosome. En effet, le locus possédant la plus grande significativité est choisi comme QTL. Des QTL à effets forts peuvent alors masquer la présence de QTL à effets plus faibles. En prenant comme cofacteurs dans un modèle multi-marqueurs des QTL déjà détectés on peut révéler des QTL à effet plus faibles. Ce modèle multi marqueurs peut être une cartographie par intervalle composite (*Composite Interval Mapping* : CIM) ou une analyse marqueur par marqueur en complétant le modèle d'analyse de variance avec les cofacteurs déjà détectés. Ces méthodes permettent à la fois un gain de puissance permettant de la détection de QTL à effets plus faibles et une meilleure précision en termes d'intervalles de confiance plus étroits (Figure I.4). (Charmet 2011).



Figure I. 4 : Comparaison des méthodes de cartographie d'intervalle simple (SIM) et composite (CIM): tracé du rapport de vraisemblance (LR) en fonction de la position le long du chromosome. Extrait de Charmet (2011).

Notons que l'approche de cartographie de QTL ne permet pas d'identifier directement le gène responsable de la variation phénotypique, mais des segments chromosomiques compris entre deux marqueurs et qui peuvent contenir de nombreux gènes candidats. La résolution de cette approche, en termes de taille du fragment chromosomique identifié contenant le QTL, dépend du nombre de marqueurs disponibles mais surtout du nombre de recombinaisons. Celui-ci sera fonction du nombre d'individus produits dans la descendance.

Afin d'identifier le(s) gène(s) sous-jacent(s) à un QTL donné, la région génétique correspondante doit être affinée jusqu'à pouvoir isoler un gène ou une région génomique de petite taille contenant le polymorphisme responsable. Pour atteindre cet objectif, il est nécessaire d'augmenter le nombre de recombinaisons dans la région en produisant un plus grand nombre d'individus recombinants. Dans le cas où le génome de l'espèce travaillée n'est pas encore séquencé, il est nécessaire de passer ensuite par une étape de construction d'une carte physique locale comprenant une série de fragments d'ADN clonés (principalement *Bacterial Artificial Chromosome* : BAC , ou *Yeast Artificial Chromosome* : YAC) (Eggen 2000). Les marqueurs appartenant à la région affinée contenant le QTL peuvent être utilisés doivent contenir des inserts chevauchants formant une séquence complète (contig) de la région d'intérêt. Ensuite, un inventaire de tous les gènes présents dans ce contig est réalisé afin d'identifier le(s) gène(s) d'intérêt impliqué dans la variation phénotypique du caractère étudié (Eggen, 2000).

Malgré le fait que la cartographie de QTL continue d'être une stratégie de choix pour identifier les régions du génome responsables des variations quantitatives chez les plantes, elle présente certains inconvénients. Tout d'abord ces études sont réalisées sur des populations biparentales, où les deux parents sont, la plupart du temps, homozygotes. Cela implique qu'on teste seulement l'effet de deux allèles à chaque locus. De plus, au cours de leur construction, les populations de cartographie subissent souvent une seule méiose efficace générant les évènements de recombinaison, les autres générations servant juste à fixer ces évènements. Une façon de surmonter ce problème est de réaliser des croisement multiparentaux (Blanc et al. 2006). Il s'agit de produire une population de RIL à partir de croisements entre plusieurs parents, ces lignées étant désignées sous le nom *Multiparent Advanced Generations Inter-Cross lines* ou lignées MAGIC (Kover et al. 2009). Les lignées MAGIC présentent l'avantage de capturer une grande part de la diversité génétique et phénotypique de l'espèce, du fait du

nombre plus élevé de lignées parentales. Le nombre plus élevé de croisements et de générations dans le cas de ces lignés permet aussi d'augmenter le nombre de recombinaisons dans la descendance, améliorant ainsi la résolution de la cartographie (Kover et al. 2009).

1.2. Génétique d'association

La cartographie par déséquilibre de liaison (DL), ou génétique d'association, utilise comme échantillon une collection d'individus non apparentés, exploitant ainsi un nombre d'événements de recombinaison plus grand que ce qui peut être obtenu avec n'importe laquelle des populations de cartographie de liaison précédemment citées (Figure I.5). Ces recombinaisons sont le résultat d'une accumulation au cours de l'histoire évolutive des accessions composant les collections utilisées. Détectées grâce à de nombreux marqueurs maintenant disponibles pour la plupart des espèces, elles conduisent à une résolution supérieure dans la localisation des QTL. La génétique d'association peut également être mise en oeuvre avec un pas de temps plus court que la cartographie de liaison, notamment grâce au temps épargné pour la construction des populations de cartographie et de cartographie fine (Figure I.5 A). De plus, le nombre d'allèles disponible au sein des populations utilisées en génétique d'association est également plus important que les deux allèles généralement représentés au sein des populations biparentales utilisées pour la cartographie par analyse de liaison (revue par (Zhu et al. 2008) (Figure I.5). Ces avantages ont fortement motivé, depuis quelques dizaines années, le développement de cette méthodologie pour l'étude du déterminisme génétique des caractères quantitatifs chez les plantes.

Les individus de la collection d'étude d'association sont génotypés pour différents marqueurs puis des corrélations sont recherchées entre le phénotype et les allèles aux marqueurs. Un autre avantage potentiel des études de génétique d'association est basé sur l'hypothèse qu'un des marqueurs génotypés correspond au polymorphisme causal, ce qui n'est cependant vérifié que si le nombre de marqueurs disponibles permet un génotypage exhaustif.

La cartographie par génétique d'association s'appuie sur une propriété génétique des populations naturelles : le déséquilibre de liaison (DL). Le DL représente l'association non aléatoire entre les allèles à différents loci, dans une population donnée. Il correspond à un écart par rapport aux fréquences attendues pour deux loci indépendants au sein d'une population respectant l'équilibre de Hardy-Weinberg (population panmictique, c'est-à-dire dans laquelle les croisements entre individus sont aléatoires, sans phénomènes de mutation, migration ou
sélection). Dans une telle population, le DL observé n'est dû qu'à la distance génétique entre les deux locus. Différentes statistiques permettent d'estimer le DL entre deux loci ; r^2 et D' sont les plus couramment utilisées. On considère deux loci avec les allèles A et a pour l'un et B et b pour l'autre. Soit F_A, F_B, F_a et F_b les fréquences alléliques et F_{AB}, F_{ab}, F_{Ab} et F_{aB} les fréquences génotypiques. D représente l'écart entre les fréquences génotypiques observées et attendues. Il est calculé de la façon suivante : $D = D_{AB} = (F_{AB} - F_AF_B)$. Le coefficient statistique $r^2 = (D_{AB})^2$ / (F_A F_B F_a F_b) (Hill and Robertson 1968) et le coefficient statistique $D' = |D_{AB}| / \min (F_A F_B,$ $F_a F_b)$ si $D_{AB} < 0$ ou $D' = |D_{AB}| / \min (F_A F_b, F_a F_B)$ si $D_{AB} > 0$ (Hedrick 1987) reflètent différents aspects du DL (Figure I.6). L'estimateur statistique r^2 résume à la fois l'histoire des mutations et des recombinaisons alors que D' mesure uniquement l'histoire des recombinaisons. Cependant, D' est fortement affecté par la taille de l'échantillon car il présente un biais lorsqu'il est utilisé pour comparer des loci ayant de faibles fréquences alléliques.



Figure I. 5 : Comparaison schématique de différentes stratégies de cartographie. Extrait de Yu and Buckler (2006) et Zhu et al. (2008).

(A) Comparaison au niveau de la résolution, de durée de la recherche et du nombre d'allèles évalué. BC : Backcross. (B) Comparaison entre la population utilisée en cartographie par analyse de liaison (à gauche) et la collection utilisée en cartographie par analyse de déséquilibre de liaison (à droite). (D'après Zhu et al. (2008)). Les deux approches de cartographie par analyse de liaison et par étude d'association reposent toutes deux sur la co-transmission du polymorphisme fonctionnel. La différence est que, pour l'analyse de liaison il n'y a que quelques possibilités de recombinaison au sein de la descendance (a, population issue d'un croisement parental), ce qui se traduit par une résolution de cartographie relativement faible, alors que pour la cartographie d'association (b, collection représentée par ses haplotypes), la recombinaison historique et la diversité génétique naturelle ont été exploitées pour la cartographie à haute résolution.



Figure I. 6 : Comparaison de différents scénarios expliquant le déséquilibre de liaison (DL) entre deux loci polymorphes liés, affectés par l'histoire évolutive et par des évènements de mutation et de recombinaison, avec leur impact sur les statistiques r² et de D'. Extrait de Flint-Garcia et al. (2003).

(A) Les deux loci présentent une histoire mutationnelle similaire sans recombinaison et r^2 et D' ont leur valeur maximale (= 1). (B) Le DL est dû à deux évènements de mutation successifs ayant eu lieu sur deux loci différentes sans recombinaison entre les loci, r^2 et D' ont des valeurs différentes. (C) la recombinaison entre les loci, ayant subi chacun un évènement de mutation, tend à annuler le DL, r^2 et D' ont leur valeur minimale (= 0).

On peut imaginer que pour des populations non panmictiques, le DL avec un locus d'intérêt sera d'autant plus fort qu'on sera physiquement proche de celui-ci. La cartographie par DL identifie donc des marqueurs moléculaires qui sont liés au polymorphisme causal même si celui-ci n'est pas génotypé. Ainsi, l'étendue du DL peut affecter la résolution de la cartographie par génétique d'association et jouer également sur la densité de marqueurs nécessaire pour détecter des associations génétiquement liées au gène causal. Si on estime que, pour une population donnée, le déséquilibre de liaison s'étend sur une longue distance (plusieurs centimorgans), alors des marqueurs proches séparés de quelques centaines de paires de bases apporteront une information redondante, et il sera inutile de génotyper la population pour tous ces marqueurs. Par conséquent, on pourra identifier la plupart des QTL d'association en génotypant la population d'intérêt avec un nombre réduit de marqueurs répartis sur tout le génome. Cette approche est nommée analyse du génome entier (Genome wide analysis study : GWAS) (Figure I.7). Par contre, si le DL (r^2) s'étend sur une courte distance, la résolution de la cartographie sera importante, mais un grand nombre de marqueurs moléculaires sera requis pour une analyse globale du génome. Il sera néanmoins possible de s'intéresser à des régions particulières qui seront typées avec une densité plus grande de marqueurs selon une approche de type gène candidat (Candidate Gene Analysis : CGA) (Figure I.7).



Figure I. 7 : Relation entre l'étendue du déséquilibre de liaison (DL) et la résolution des études d'association. Extrait de Rafalski (2002).

(a) : le DL diminue lentement avec l'éloignement du gène responsable du phénotype (ovale rouge). Dans ce cas, une faible densité de marqueurs (représentés par des barres verticales rouges) est suffisante pour identifier les marqueurs associés (jaune flèches). (b) : le DL diminue très rapidement autour du gène responsable, et une densité de marqueurs beaucoup plus grande est nécessaire pour identifier un marqueur associé (flèche jaune).

L'étendue du DL peut être affectée par de nombreux facteurs tels que la fréquence allélique, la mutation, le taux de recombinaison ainsi que la structure et la taille de la population (Rafalski

and Morgante 2004). En effet, le DL présent dans une population naturelle va être créé par la mutation qui va induire l'apparition de nouveaux allèles. Le DL pourra ensuite augmenter avec des facteurs influençant les fréquences alléliques des nouveaux allèles tels que la migration et la sélection. Le seul facteur capable de diminuer le DL entre deux locus est la recombinaison. La sélection et la diminution du nombre de recombinaisons (liaison physique) vont augmenter le DL local alors que la dérive génétique, due à la faible taille des populations, et le mélange de populations génétiquement différenciées vont augmenter le DL au niveau du génome entier (DL global) (Rafalski and Morgante 2004). En effet, la présence de sous-populations ou d'apparentements entre les individus va augmenter le DL à longue distance, éventuellement entre marqueurs portés par des chromosomes différents et va, par conséquent, augmenter le nombre de fausses associations identifiées entre le génotype et le caractère d'intérêt. Il est donc important de prendre en compte la structure génétique de l'échantillon étudié lorsqu'on recherche des associations avec un caractère, et surtout, lorsque ce caractère influence la structure. Cette structuration, lorsqu'elle est ignorée dans les modèles statistiques d'analyses de génétique d'association, produit l'effet décrit plus haut : la distribution des allèles différenciant les populations va être virtuellement corrélée avec le phénotype, pour tous les traits ayant des moyennes différentes selon la population. Il devient donc difficile de faire la distinction entre les vrais allèles responsables d'un phénotype et les allèles associés au phénotype par le seul effet du biais statistique.

Ainsi, la première étude utilisant la génétique d'association chez les végétaux en prenant en compte dans les analyses d'association une matrice de structure de la population (Q), a été réalisée par Thornsberry et collaborateurs (2001). Cette étude a permis de mettre en évidence une association significative entre le gène Dwarf8 et la précocité de floraison chez le maïs. La structure de la population (matrice Q) peut être estimée par le nombre de sous-populations et la probabilité d'appartenance des individus à une sous-population, estimés avec la méthode bayésienne mise en œuvre avec le logiciel STRUCTURE (Pritchard et al. 2000), ou par les coordonnées des individus issues d'une analyse discriminante en composante principale (*Discriminant Analysis of Principal Components* : DAPC), qui est une méthode multivariée et mise en œuvre avec le package R Adegenet (Jombart et al. 2010). L'utilisation des matrices de composantes principales ou de composantes issues d'une analyse DAPC, ont également été proposées comme alternative aux matrices de populations Q (Price et al. 2006; Zhu and Yu 2009; Bauchet et al. 2017).

Toutefois, ces deux méthodes ne sont pas efficaces pour capturer la structure des populations complexes. Il est donc courant d'associer à la matrice Q, une matrice d'apparentement entre les individus, ou matrice de kinship (k). Yu et collaborateurs (2005) ont développé ce modèle pour diminuer le taux de faux positifs dans les études d'association chez le maïs en utilisant les données d'apparentement entre individus. L'analyse de l'association entre marqueur et phénotype est alors réalisée en utilisant un modèle linéaire mixte (MLM) :

 $Y = X\alpha + Q\beta + Ku + \varepsilon$, où Y est le phénotype, X les allèles au marqueur, α l'effet fixe du locus, Q la matrice de structure, β l'effet fixe de la population, K la matrice d'apparentement, u l'effet aléatoire de l'apparentement et ε la résiduelle.

La matrice K définit le degré de covariance génétique entre paires d'individus. Cependant, en prenant en compte ces coefficients comme cofacteurs dans les modèles d'analyses association, des marqueurs en association significative avec le phénotype d'intérêt pourraient être ne pas être détectés. En effet, dans le modèle statistique d'analyse d'association, les marqueurs interviennent 2 fois : un effet fixe pour le locus testé et un effet aléatoire avec la matrice de kinship qui est déjà est estimée avec tous les marqueurs testés en génétique d'association. Si on teste un marqueur en déséquilibre de liaison avec beaucoup d'autres marqueurs, et que tous ces marqueurs sont utilisés pour estimer la kinship, alors on risque de perdre de la puissance statistique car l'effet du marqueur testé est « gommé » par la matrice de kinship. Pour surmonter ce risque et augmenter la puissance de détection, Rincent et collaborateurs (2014) ont proposé une nouvelle méthode d'estimation de la matrice d'apparentement permettant de gagner en puissance de détection particulièrement dans les régions à fort déséquilibre de liaison. Il s'agit de générer une matrice d'apparentement entre individus par chromosome en prenant en compte tous les marqueurs sauf ceux du chromosome testé. De ce fait, un certain nombre de matrices d'apparentement (Kchr) correspondant au nombre des chromosomes du génome de l'espèce étudiée, sont utilisées pour l'analyse d'association par chromosome.

1.3. Approche intégrative

Récemment, une approche intégrative de cartographie, nommées LD-LA (*Linkage Disequilibrium-Linkage Analysis*), s'est développée à partir de populations multi-parentales. Cette approche combine l'utilisation de marquage dense au niveau des parents, analogue à ce qui est fait pour la génétique d'association, avec la cartographie classique de QTL (par analyse de liaison) intrapopulation en ségrégation. Cette méthode présente l'intérêt de tirer parti des avantages des approches de liaison et d'association en incluant notamment la connaissance de

la structuration des populations, la prise en compte d'évènements de recombinaison anciens et récents et la détection des allèles rares (Pascual et al. 2016). La population NAM (*Nested Association Mapping*) développée chez le maïs (Yu et al. 2008) permet de combiner ces différents avantages. Ce dispositif est composé de population de lignées recombinantes obtenues en croisant la lignée de référence B73 utilisée comme pivot avec un ensemble de 25 parents (Figure I.8). L'ensemble des lignées parentales a été reséquencé pour générer une forte densité de marqueurs (1,6 millions). Les descendances recombinantes ont été ensuite génotypées avec un marquage moins dense (Tian et al. 2011). Ce panel à grand effectif (5000 lignées) a été évalué par analyse de génétique d'association afin d'identifier les déterminants génétiques de la date de floraison et d'autres caractères agronomiques, tels que la résistance aux maladies ou l'architecture des feuilles (Buckler et al. 2009; Kump et al. 2011; Tian et al. 2011).





 \times : Croisement avec B73 (Lignée de référence); SSD (*Single seed Descent*) : descente à une seule graine ; \otimes : Autofécondation pendant 5 générations

En conclusion, nous disposons de plusieurs techniques de génétique quantitative ayant chacune leurs avantages et inconvénients. Par exemple, les populations de cartographie permettent l'identification de régions chromosomique assez larges avec assez peu de marqueurs tandis que les approches de génétique d'association permettent une identification plus précise de régions chromosomiques à l'échelle du gène, mais sont contraintes par les effets de structure ou d'hétérogénéité génétique et allélique. Le choix de l'utilisation d'une technique doit se faire en fonction des objectifs et des questions, mais aussi de l'information (phénotypique, génomique) déjà disponible.

2. Outils génétiques et génomiques disponibles pour l'amélioration génétique du pois

2.1. Ressources génétiques

Le pois est une espèce appartenant à la famille des Légumineuses ou Fabacées et au genre *Pisum*. La famille des Fabacées regroupe diverses espèces de plantes herbacées annuelles, réparties en cinq genres : *Lathyrus* L. (gesse/pois doux, environ 160 espèces), *Lens* (lentilles, 4 espèces), *Pisum* L. (pois, 3 espèces), et *Vicia* L. (vesces, environ 140 espèces) et le genre monotypique Vavilovia (Smýkal et al. 2011). La classification du genre *Pisum*, regroupe trois espèces : *Pisum sativum*, *Pisum fulvum* et *Pisum abyssinicum* (Cieslarová et al. 2011).

Des collections de ressources génétiques, conservant entre 200 et 9000 accessions de pois sauvages et cultivés, sont réparties dans les banques de gènes d'une vingtaine de pays dans le monde (Warkentin et al. 2015). Des collections de référence de pois ont été constituées à partir des données passeports de ces ressources et de quelques caractères d'intérêt (origine géographique, morphologie, architecture de la plante, qualité de la graine, tolérance aux stress biotiques et abiotiques...). C'est le cas de la collection de référence de 372 accessions constituée par le Centre des Ressources Biologique (CRB) de l'INRA de Dijon (Burstin et al. 2015), utilisée dans le cadre de ma thèse.

Profitant de la diversité présente dans les collections de pois disponibles, de nombreuses populations de cartographie ont été construites au cours des dernières décennies et des efforts ont été déployés pour identifier les loci impliqués dans le contrôle de caractères d'intérêt et les introgresser dans du matériel en sélection. Ces populations de lignées recombinantes ont été développées pour étudier des caractères d'intérêt comme la résistance aux maladies ou la tolérance au gel. (Lejeune-Hénaut et al. 2008; Hamon et al. 2013; Klein et al. 2014). Des lignées quasi-isogéniques comportant des QTL de tolérance aux stress biotiques ou abiotiques ont été construites dans différents fonds génétiques (Hascoët et al. 2014; Lavaud et al. 2015). De nombreuses données génotypiques (marqueurs SSRs et SNP) et phénotypiques sont désormais disponibles sur la collection de référence, les populations de cartographie et les lignées quasi-isogéniques. Par ailleurs, trois populations de mutants de Tilling (*Targeting*)

induced local lesions in genomes) ont également été développées à l'INRA par les unités Agroécologie, IJPB (Institut Jean-Pierre Bourgin) et IPS2 (Institut des Sciences des Plantes Paris Saclay) : environ 5000 familles sont issues du génotype Térèse, 4800 familles sont issues de Caméor (Dalmais et al. 2008) et 3700 familles sont issues de 336/11. Ces ressources de lignées mutantes sont très importantes pour entreprendre des approches de validation fonctionnelle car le pois est extrêmement difficile à transformer génétiquement (Švábová and Griga 2008; Warkentin et al. 2015).

2.2. Ressources génomiques

Le pois est une espèce diploïde (2n = 14) dont le génome présente une taille d'environ 4,45 Gpb. Cette taille importante et un contenu riche en éléments répétés (50 à 60 %) ont sans aucun doute contribué à retarder la réalisation de la carte physique du génome de cette espèce (Tayeh et al. 2015b).

Récemment, plusieurs programmes nationaux et internationaux ont développé diverses ressources génomiques de grand intérêt en tirant parti des technologies de séquençage de nouvelle génération (Next Generation Sequencing : NGS). De nombreux marqueurs SNP issus du séquençage du transcriptome ou du génome entier ont ainsi été développés (Duarte et al. 2014; Leonforte et al. 2014; Sindhu et al. 2014; Tayeh et al. 2015a; Boutet et al. 2016). Ces ressources ont permis la densification de cartes génétiques du pois et le génotypage de collections de ressources génétiques. Actuellement, il existe chez le pois au minimum 52 cartes génétiques construites grâce au génotypage de populations de lignées biparentales F2 ou RIL et comprenant jusqu'à 8503 marqueurs (Tayeh et al. 2015a). A partir de ces cartes génétiques, des cartes consensus ont été construites afin d'offrir une cartographie plus dense en marqueurs avec une meilleure couverture du génome. Parmi ces cartes, citons plus particulièrement les deux cartes consensus utilisées dans le cadre de ma thèse : d'une part, la carte de Duarte et collaborateurs (2014) qui combine les données moléculaires de 4 populations de cartographie et comporte 1340 marqueurs SNP développés à partir du séquençage d'ADNc de huit lignées et d'autre part, la carte de Tayeh et collaborateurs (2015a). Cette dernière reprend une partie des données de la précédente en y ajoutant 12802 SNP générés à partir de séquences d'ADN génomique de 16 génotypes, alignées sur des contigs d'ADNc du génotype de référence Caméor représentant l'atlas du génome du pois (Alves-Carvalho et al. 2015). Elle combine les données de génotypage de 12 populations de RIL pour un total de 15079 marqueurs. La puce GenoPea Infinium® BeadChip de 13204 SNP qui a permis d'établir la carte consensus de Tayeh et collaborateurs (2015a) a également été utilisée pour génotyper la collection de référence de 372 accessions qui sert de base à l'étude de génétique d'association de ma thèse.

Pour compléter ces ressources génomiques, deux banques BAC étaient jusqu'à présent disponibles pour *Pisum sativum*. Ces deux banques correspondent à l'accession PI 269818 (Coyne et al. 2007) et au cultivar Caméor (http://cnrgv.toulouse.inra.fr/fr).

3. Les plantes face aux basses températures

L'étude de certaines plantes ayant la capacité de développer la tolérance aux différents stress biotique et abiotique tels que les températures extrêmes, la sécheresse et le taux élevé de salinité va aider à améliorer et à stabiliser la productivité agricole lors d'exposition à ces stress. Un des grands stress environnementaux est le froid qui peut réduire la croissance des plantes voir entrainer leur mort et donc impacter négativement la productivité agricole (Xiong et al. 2002). De ce fait, la compréhension des mécanismes de tolérance à ce stress constitue un enjeu économique majeur.

Toutes les plantes ne présentent pas la même sensibilité à l'intensité du froid. Les plantes d'origine tropicale ne sont pas capables de s'adapter au froid positif entre 0 et 15 °C (*chilling*) ni négatif (*freezing*), alors que les plantes issues de milieu tempéré peuvent être tolérantes au froid négatif si elle sont au préalable soumises à de basses températures non gélives pendant une période qualifiée d'acclimatation au froid (Sharma et al. 2005).

3.1. Les effets de basses températures sur la plante

En réponse au stress froid, le ralentissement de la croissance observé permet à la plante de diminuer les effets du stress, par exemple par l'utilisation des réserves glucidiques pour la protection cellulaire. En outre, le froid donne lieu à un stress secondaire dû à l'accumulation de dérivés réactifs de l'oxygène (*Reactive Oxygen Species* : ROS) : le stress oxydatif (Rihan et al. 2017) (Figure I. 10). En effet, au cours d'un stress froid, l'énergie lumineuse absorbée par la plante excède la capacité de cette dernière à utiliser cette énergie au travers des réactions photosynthétiques ou à la dissiper sous forme radiative par la fluorescence ou la chaleur. Ceci se traduit par un déséquilibre entre la quantité de lumière reçue et l'énergie consommée par la plante (photoinhibition) impliquant la formation des ROS qui regroupent des radicaux libres d'oxygène et des non-radicaux qui sont des agents oxydants ou facilement convertis en radicaux

 $(O2^{-}, OH^{-} \text{ et } H_2O_2 \dots)$ (Halliwell 2006). Toutes ces perturbations physiologiques peuvent aboutir à l'inhibition de la croissance et la mort de la plante (Yadav 2010).

Quand la température diminue en dessous de 0°C, la glace se forme dans l'espace intercellulaire du tissu de la plante car le point de congélation du liquide intercellulaire est plus haut que celui du liquide intracellulaire. En conséquence, l'eau se déplace de l'intérieur de la cellule vers les espaces intercellulaires conduisant à la déshydratation cellulaire (Figure I.9) (Thomashow 1998; Sharma et al. 2005).



Figure I. 9 : Schéma présentant le processus de déshydratation cellulaire causé par le gel. Extrait de Sharma et al. (2005).

La différence de potentiel de soluté conduit à la formation de glace dans les espaces intercellulaires. Cela provoque une chute du potentiel hydrique dans l'espace extracellulaire, ce qui finit par induire un mouvement de l'eau vers l'extérieur de la cellule, entraînant une déshydratation de la cellule.

Le processus, si la température décroit, se poursuit par la formation de la glace à l'intérieur de la cellule ce qui est létal, car le grossissement des cristaux de glace intracellulaires perturbe l'intégrité du protoplasme vivant jusqu'à perforer la membrane plasmique. La diminution rapide de température peut causer la création de cristaux de glace intracellulaire (Guy 2003). Pour tolérer cela les plantes ont donc développé différentes stratégies.

3.2. Stratégies de tolérance des plantes face aux basses températures

Le niveau de tolérance au froid dépend essentiellement de deux stratégies : l'échappement et/ou la tolérance aux températures froides non gélives via le phénomène d'acclimatation. La première stratégie consiste à décaler les stades phénologiques les plus sensibles aux basses

températures, notamment l'initiation florale et la floraison, et/ou à modifier sa morphologie afin que les organes les plus sensibles ne soient pas soumis au stress froid. L'échappement permet donc à la plante d'éviter les contraintes du stress froid par une bonne adaptation de son cycle de développement à la période et durée de survenue des basses températures.

3.3. Mécanismes d'acclimatation au froid

L'acclimatation au froid survient lors de la croissance aux basses températures non gélives et permet d'induire une tolérance au gel et ainsi de minimiser les dommages causés par des températures moins favorables (Sharma et al. 2005). La plante met en place des mécanismes de tolérance assurant la protection de la cellule. En effet, le rétablissement de l'équilibre osmotique, la protection des membranes et la détoxification du milieu sont souvent les objectifs des stratégies de tolérance. Un schéma de la réponse des plantes face au stress froid est présenté figure I.10.

Les cellules des plantes perçoivent le stress froid à travers des récepteurs membranaires qui initient un signal, ensuite transmis par l'intermédiaire de seconds messagers comme le calcium et les dérivés réactifs de l'oxygène (figure I.10). L'augmentation du calcium dans le cytosol dû au stress froid est analysée par des senseurs de calcium qui provoquent une cascade de phosphorylations donnant lieu à l'activation des gènes impliqués dans la réponse au stress. Ces gènes permettent la mise en place de mécanismes pour la survie de la plante tels que le rétablissement de l'équilibre osmotique, la détoxification par élimination des dérivés réactifs de l'oxygène et la protection contre la dénaturation des membranes et des protéines. De plus, l'accumulation de saccharose et d'autres molécules de sucres simples qui surviennent pendant l'acclimatation au froid pourraient participer à la stabilisation de la membrane, car ces molécules peuvent protéger les membranes in vitro contre le dommage induit par le froid (Rihan et al. 2017; Kazemi-Shahandashti and Maali-Amiri 2018).

Plus de 170 gènes codant des facteurs de transcription sont impliqués dans l'acclimatation au froid chez *Arabidopsis thaliana* (Thomashow 2010). Ils appartiennent pour la majorité aux familles de facteurs de transcription suivantes : *bHLH, bZIP, HSF, AP2/ERF, NAC, WRKY, DOF, MADS, MYB* et *TCP2* (Chawade et al. 2007). Les modalités d'activation ou de répression de la plupart de ces facteurs de transcription en réponse au signal froid restent peu comprises sauf pour la famille des *AP2/ERF* qui est la famille la plus étudiée avec la classe des facteurs de transcription *DREB* (*Dehydration Responsive Element*) et la sous-classe *CBF/DREB1*. Les

produits de l'expression des gènes *CBF/DREB1* (*DRE-Binding factor 1*) se fixent au niveau des éléments *cis* semblables à *CRT* (*C-RepeaT*) et *DRE* (*Dehydration-, high salt-, low temperature-Responsive Element*), ayant RCCGAC comme séquence centrale (Liu et al. 1998).

Chez *Arabidopsis thaliana*, 3 *CBF* sont impliqués dans la tolérance au froid : *CBF1*, *CBF2 et CBF3*, également appelés respectivement *DREB1B*, *DREB1C* et *DREB1A* qui furent les premiers gènes *CBF/DREB1* à être clonés (Stockinger et al. 1997; Gilmour et al. 1998; Liu et al. 1998). Des gènes *CBF/DREB1* ont été identifiés chez de nombreuses autres espèces herbacées ou ligneuses, capables ou non de s'acclimater au froid. Ces gènes s'expriment très précocement lors de la phase d'acclimatation et ils activent l'expression de gènes *COR* (*COld Responsive*) (Jaglo-Ottosen et al. 1998; Shinwari et al. 1998) possèdent les motifs de régulation CRT/DRE (Stockinger et al. 1997). Actuellement près de 4000 gènes COR sont identifiés.

Le nombre de gènes *CBF/DREB1* varie d'une espèce à l'autre mais aussi à l'échelle intraspécifique. Le génome d'*Arabidopsis thaliana* comprend 6 gènes *CBF/DREB1* (Sakuma et al. 2002) comme celui de *Populus trichocarpa* (Benedict et al. 2006). Il y en a 17 pour *Eucalyptus grandis* (Cao et al. 2015) et plus chez *E. gunnii* (Nguyen et al. 2017), alors que les génomes des céréales tempérées présentent 25 gènes répartis en 10 groupes (revue par (Tondelli et al. 2011).

Des clusters de gènes *CBF/DREB1* répétés en tandem ont pu être mis en évidence chez plusieurs espèces et ont souvent été localisés au niveau des QTL d'acclimatation ou de tolérance au froid. Chez *Arabidopsis thaliana*, le gène *CBF2* a été proposé comme candidat majeur pour FTQ4, un QTL de tolérance au gel après acclimatation au froid (Alonso-Blanco et al. 2005). Chez *Triticum monococcum*, au moins 11 gènes *CBF/DREB1* ont été identifiés au niveau du *QTL* de tolérance au gel Fr-Am2 (Miller et al. 2006). Chez *Hordeum vulgare*, un cluster comportant 11 gènes *CBF/DREB1* colocalise avec le QTL de tolérance au gel Fr-H2 (Skinner et al. 2006).Chez *Medicago truncatula*, un cluster de 12 gènes *CBF* colocalise avec un QTL de tolérance au gel (Tayeh et al. 2013b).



Figure I. 10 : Schéma général de la réponse des plantes au stress froid. Adapté de Wang et al. (2003) et Rihan et al. (2017).

Les différents facteurs de transcriptions impliqués dans la réponse au froid entrainent de nombreuses modifications de l'expression du génome impliquées dans le rétablissement de l'équilibre osmotique, la protection des macromolécules et la détoxification (figure I.10). Cela conduit à :

- l'accumulation de protéines synthétisées comme des protéines hydrophiles telles les déhydrines (protéines appartenant au groupe II de la famille *LEA (Late Embryogenesis Abundant)* qui jouent un rôle essentiel dans la protection des protéines contre la dénaturation et dans la stabilisation des membranes face aux pertes excessives d'eau (Kosová et al. 2007), des protéines de choc thermique HSP (*Heat Shock Protein*) qui interviennent dans la protection des membranes cellulaires, le repliement des protéines dénaturées et empêchent l'agrégation des protéines (Janská et al. 2010) et des protéines antigel capables d'interagir avec les cristaux de glace et de ralentir leur croissance et leur recristallisation limitant ainsi les risques de dégâts physiques liés au gel (Griffith et al. 2005) ;
- l'accumulation de solutés (sucres, acides aminés...) qui représente un mécanisme commun de réponse à plusieurs stress abiotiques, y compris le froid (revue par Bhandari and Nayyar 2014) ; l'accumulation de sucres au niveau du cytoplasme contribue en effet à réduire les différences de potentiel hydrique entre le milieu intracellulaire et le milieu extracellulaire où les cristaux de glace se forment, stabiliser les structures membranaires, stabiliser les protéines au cours du gel et neutraliser les espèces réactives de l'oxygène (revue par Bhandari and Nayyar 2014).
- des modifications membranaires avec l'augmentation de l'insaturation des lipides membranaires, le raccourcissement des chaînes carbonées, et/ou la modification du rapport lipides/protéines (revue par Barrero-Gil and Salinas 2017).
- De nombreuses réponses au froid sont également liées aux régulateurs de croissance qui ont une action sur l'acclimatation au froid, soit via leurs actions sur le régulon CBF (revue par Eremina et al. 2016a; Barrero-Gil and Salinas 2017) soit de manière indépendante comme la réponse au froid, via la voie dépendante de l'acide abscissique (ABA). Dans ce cas, le froid induit une accumulation d'ABA qui conduit à l'expression des ABF (*ABA responsive element Binding Factor*) qui induisent eux-mêmes des gènes effecteurs porteurs d'une séquence *cis-acting*: ABRE (*ABA Responsive Element*) (Rihan et al. 2017; Kazemi-Shahandashti and Maali-Amiri 2018)

L'acide gibbérellique (AG) est impliqué dans la réponse au stress régulée par la voie des CBF. Les composants clés de l'AG sont les protéines nucléaires localisées (DELLA) qui inhibent la croissance des plantes. Cependant, des progrès récents ont indiqué que les protéines DELLA jouent un rôle dans de nombreux aspects de la croissance des plantes, en particulier ceux qui sont affectés par les stress environnementaux (Achard et al. 2008). Ces auteurs ont étudié la relation entre les voies CBF1 et l'AG et ont montré que l'expression constitutive de *CBF1* désactive le rôle de l'AG, ce qui entraîne un nanisme chez les plantes et un retard de floraison.

Le methyljasmonate et l'acide jasmonique peuvent également réguler positivement les voies de signalisation via leurs actions sur la voie des *CBF* et leurs effets inhibiteurs sur l'expression des gènes *JAZ (JAsmonate-Zim-domain)* (revue par Shi et al. 2015; Sharma and Laxmi 2016; Hu et al. 2017).

L'éthylène est un régulateur négatif de la signalisation par le froid induite par les *CBF*. Il favorise la stabilité du facteur de transcription EIN3 qui se fixe au niveau des régions promotrices des gènes de régulation de la réponse au froid telles que les gènes *CBF*. Il intervient également via les gènes *ARR5*, *ARR7* et *ARR15* (*type-A Arabidopsis Response Regulator*) (Shi et al. 2012). Plus récemment, le gène *ERF105*, un facteur de transcription de réponse à l'éthylène, a été identifié comme impliqué en liaison avec les *CBF* dans la réponse au froid (Bolt et al. 2017).

Récemment, l'implication des brassinostéroïdes dans l'acclimatation au froid a été mise en évidence (Eremina et al. 2016b). Ils contrôlent le facteur de transcription CESTA impliqué dans l'expression constitutive des *CBF* et le facteur de transcription *BZR1* (BrassinaZole-Resistant 1) qui régule positivement l'expression de *CBF1*. *BZR1* est également impliqué dans la régulation de gènes de réponse au froid, indépendamment du régulon CBF (Li et al. 2017).

L'ABA, l'acide gibbérellique, les cytokinines et le jasmonate interviennent également sur les voies indépendantes des *CBF*, en particulier par leurs actions inhibitrices sur la photosynthèse pour l'acide jasmonique, et la division et l'expansion cellulaire pour l'acide gibbérellique (revue par Wingler 2015).

3.4. Tolérance au froid du pois

Le pois a développé plusieurs stratégies de tolérance au froid (Figure I.11). La stratégie d'évitement est caractéristique des lignées fourragères qui sont capables d'échapper aux principales périodes de gel de l'hiver tout en allongeant leur développement végétatif, et en retardant en particulier leur initiation florale qui marque le début d'une période de plus grande sensibilité au gel. Dans ce sens, un intérêt particulier est réservé au gène *Hr* (*High response to photoperiod*) contrôlant l'intensité de la réponse à la photopériode. En effet, l'allèle dominant *Hr* retarde le passage à l'initiation florale sous un régime de jours courts et permet le maintien

de la plante à l'état végétatif depuis l'automne jusqu'au printemps suivant (Lejeune-Hénaut et al. 1999). Des QTL impliqués dans la tolérance au gel hivernal (WFD : *Winter Frost Damage*) ont été détectés au sein d'une première population de lignées recombinantes (Champagne x Térèse, Pop2, Lejeune-Hénaut et al. 2008). Trois de ces QTL ont été identifiés dans toutes les conditions expérimentales : WFD3.1, WFD5.1, WFD6.1, respectivement sur les groupes de liaison III, V et VI et le locus *Hr* a été identifié comme le marqueur le plus associé avec le QTL WFD3.1. Par la suite, la colocalisation entre WFD3.1 et *Hr* a été confirmée par l'étude de la population China x Caméor (Pop9, Klein et al. 2014).

Les lignées qui portent l'allèle *Hr* sont tolérantes au gel et utilisées comme géniteurs dans le programme de sélection de pois d'hiver coordonné par l'unité expérimentale INRA GCIE (Grandes Cultures Innovation Environnement). Weller et collaborateurs (2012) ont pu mettre en évidence que le gène *Hr* est un orthologue du gène *Elf3 (Early flowering 3)* d'*Arabidopsis* connu par son implication dans la fonction de l'horloge circadienne. Le gène de nanisme *Le,* colocalisé au QTL *WFD*3.2 et qui conduit à un port en rosette pendant l'hiver serait également à relier à une réponse d'évitement.



Figure I. 11 : Modèle proposé des composantes de la tolérance au froid chez le pois. En noir : QTL identifiés dans Pop2 et Pop9, en rouge QTL identifié dans Pop9 uniquement. D'après Delbreil (2018).

Une tolérance au gel à l'état végétatif qui est en relation étroite avec la capacité d'acclimatation au froid a également été mise en évidence par Bourion et collaborateurs (2003), ce qui a permis de développer l'étude de ce caractère agronomique complexe au sein de notre équipe. Nous faisons l'hypothèse que les autres QTL détectés dans Pop2 et Pop9 (Lejeune-Hénaut et al. 2008; Dumont et al. 2009; Klein et al. 2014) participent au déterminisme génétique de la tolérance à l'état végétatif.

Les deux génotypes Champagne et Térèse utilisés dans la construction de la population de lignées recombinantes Pop2 permettent de dissocier les variations spécifiques de l'acclimatation au froid des variations liées à la croissance à basses températures (*chilling*) et à la croissance de chaque génotype. Champagne est une lignée fourragère dérivée d'une population française locale. Elle présente le phénotype de floraison [Hr], c'est-à-dire une initiation florale retardée sous jours courts et considérée comme une lignée tolérante au gel portant les allèles favorables aux principaux QTL de tolérance. Par contre, Térèse n'est que légèrement réactif à la durée du jour, ce qui révèle son phénotype [hr]. C'est une variété française de pois de printemps, sensible au gel et donc généralement incapable de survivre en hiver. Ces deux génotypes ont servi de support aux études de protéomique et transcriptomique réalisées dans notre équipe, qui ont permis de mettre en évidence les principaux mécanismes impliqués dans l'acclimatation au froid chez le pois (Dumont et al. 2011; Lucau-Danila et al. 2012; Grimaud et al. 2013; Legrand et al. 2013).

Chapitre II : Re-détection de QTL de tolérance au gel chez le pois et cartographie fine

Introduction

This chapter takes place in a program of elucidation of the genetic determinism of frost tolerance in pea using quantitative genetics approaches as QTL detection and fine mapping. The aims of the experiments presented here were to validate the effect of the WFD6.1 QTL in frost tolerance by phenotyping quasi-isogenic lines as well as to reduce the size of its confidence interval by a fine mapping procedure.

The strategy adopted in this chapter is summarized in Figure II.1. After the identification of QTLs for frost tolerance in the Pop2 population (Champagne x Térèse), a marker-assisted backcross program was carried out to transfer favorable alleles at one of the main frost tolerance QTL, namely WFD6.1, in a frost sensitive genetic background, namely Eden. Near isogenic Lines (NILs) were there by produced, which had become homozygous in the targeted region with either the allele of the tolerant donor line, or the allele of the sensitive recipient line. After seed increase, the phenotyping of NILs in multiple field experiments offer the possibility to validate the effect of the WFD6.1 QTL in a sensitive genetic background.

Beyond QTL validation, NILs will help to reduce the size of the frost tolerance locus. Thus, when a line was identified as carrying the alleles of interest in its heterozygous region, descendants are then generated and genotyped to search for crossing-overs in this locus. These recombinant plants are then self-pollinated to be fixed if necessary, then phenotyped to more precisely localize the region implicated in frost tolerance.

In parallel, a re-detection of frost tolerance QTLs using a massive number of markers was performed in order to precise the confidence interval of previous detected QTLs, and to identify suitable candidate markers which will be added to minimize the size of the locus by using NILs.

The results presented in this first chapter of the thesis concerned:

- Re-detection of frost tolerance QTLs using a massive number of markers
- Construction of NILs at the frost tolerance WFD6.1 QTL
- Genomic and phenotypic evaluation of NILs
- An attempted fine-mapping of the frost tolerance WFD6.1 QTL



Figure II. 1: Strategy adopted for the fine mapping of the frost tolerance locus WFD6.1.

A. Re-detection of frost tolerance QTLs using a massive number of markers

1. Material and methods

1.1. Plant material

The plant material used in the present study is represented by a subset (n= 76) of the (164) recombinant imbred lines (RILs) of *Pisum sativum* which were identified as Pop2 and have been used in previous QTL (Lejeune-Hénaut et al. 2008; Dumont et al. 2009) and mapping studies (Aubert et al. 2006; Duarte et al. 2014; Tayeh et al. 2015a). The mapping population Pop2 was obtained by single seed descent from a large outcrossed F2 population. This population was originally designed to maximize the segregation of frost tolerance differences in pea. It was constructed from crossing two unrelated lines, Champagne and Térèse, characterized by a contrasted behavior against frost. The parental line Champagne is a frost tolerant and winter hardy pea variety. It is known as a forage line exhibiting the flowering gene *Hr* responsible for delaying floral initiation under short days (Lejeune-Hénaut et al. 1999). In contrast, the parental line Térèse is a spring dry frost sensitive variety, with no significant response to photoperiod revealing its *hr* phenotype (Lejeune-Hénaut et al. 1999).

1.2. Phenotyping

The QTL analysis was conducted using phenotypic data which has already been used in Lejeune-Hénaut et al. (2008). Lejeune-Hénaut et al. (2008) evaluated frost tolerance of the Pop2 under field experiment in 11 location x year conditions. The experiments were carried out at the INRA experimental stations of Mons, Clermont Ferrand-Theix and Dijon, during the growing season of 2000/2001, 2001/2002 and 2002/2003. In addition, experiments were also conducted at the INRA experimental stations of Colmar and Lusignan in 2001/2002. Plots were sown in a randomized complete block design with two replicates in 2000/2001 and three replicates in the other experimental conditions. The collection was assessed for frost tolerance by visual estimation of winter frost damages after each winter frost period. A score was assigned to a plot as a whole, based on the extent of necrotic areas of the aerial parts of the plants according to a scale ranging from 0 to 5 where 0 represented no damage and 5 a dead plant (Lejeune-Hénaut et al. 2008).

Overall, 11 traits constituted the phenotyping data for the QTL analysis. Traits were named with the location of experimental station and then the year of growing season.

1.3. Phenotypic data Analysis

Phenotypic datasets for frost damages, obtained from the 76 Pop2 lines were analyzed using the SAS package ((SAS 1999) cited in Lejeune-Hénaut et al. (2008)). For each trait, a linear model (LM) was used: $Y_{ij} = \mu + \text{geno}_i + \text{rep}_j + e_{ij}$, where Y_{ij} is the value of frost damages recorded for the genotype *i* at the replicate *j*. μ is the mean, geno_i is the fixed genetic effect of the genotype *i*, rep_j is the fixed replicate effect of the replicate *j* and e_{ij} is the residual effect. Least Square Means (LSMeans) for each genotype-trait combination were calculated from each LM analysis using SAS program and were used for further QTL analysis. Histograms of LSMeans frequency distributions were drawn using the R 3.5.0 software (Hervé 2014; R Core Team 2014).

1.4. Genetic linkage map and Genotypic data

The Pop2 genotyping data has already been available for the QTL mapping, and the genetic data are detailed in Tayeh et al. (2015a) and Duarte et al. (2014). The total length of Pop2 map, described in Tayeh et al. (2015a), is 945 cM distributed on seven linkage groups (LG). Genetic distances (in cM) between markers were estimated according to Haldane mapping function. The number of markers used for QTL detection is 6486 markers.

1.5. QTL analysis

QTL analysis was performed with Windows QTL Cartographer (WinQTLCart) version 2.5 (Wang et al. 2012) using a composite interval mapping (CIM) with a forward regression model (model 6). Forward and backward regression method with a probability in and probability out set each one to a value of 0.05, were selected for the QTL analysis. The windows size was set up to a value of 10 cM on either side of the markers flanking the test site. The genome scan interval, characterized by the Walk speed value, was set to a value of 1 cM. LOD significance threshold was determined by the software by computing 500 permutations tests ($\alpha = 0.05$) for each of the studied traits. The higher resulting threshold, LOD = 4.7, was chosen for QTL analysis. For each QTL, confidence intervals (CIs) were calculated by WinQTLCart as the interval corresponds to a 2-LOD drop-down. In addition, the percentage of phenotypic variation (\mathbb{R}^2) explained by each individual QTL and the corresponding additive effect were also

estimated using the software. Overall, for each linkage group, consensual positions of a global QTL are revealed by overlapping all individual CIs corresponding to QTL associated with a studied trait. Thus, the consensus CIs of global QTLs and their corresponding allelic effect were estimated for each trait. LOD peaks at QTL intervals were considered as putative QTL locations on the linkage map. The QTL map was drawn using Mapchart version 2.32 software (Voorrips 2002) was used to draw global QTLs on the genetic linkage map.

2. Results

2.1. Phenotyping

The subpopulation Pop2 composed by 76 RILs exhibited quantitative variation for winter frost damages (Figure II.2). The distribution of most of WFD scores for the 11 experimental condition was skewed towards sensitive phenotype values (Figure II.2 C, E, F, H and I). Based on the scale 0-5, the adjusted means of WFD scores (LsMeans), corresponding for the three growing seasons studied, from an average of 0.89 to 4.5 for Mons, from 0.44 to 4.66 for Clermont-Ferrand and from 0.78 to 4.66 for Dijon. The WFD values for Colmar and Lusignan experiment ranged from 1.67 to 5 and 1.33 to 5 respectively.





The remaining histograms A to I indicate the experiments carried out at the 3 experimental stations of Mons, Clermont-Ferrand and Dijon, during the growing season of 2000/2001, 2001/2002 and 2002/2003 respectively. J and K indicate the two experiments conducted in 2001/2002 at the experimental stations of Colmar and Lusignan respectively.

2.2. QTL mapping for winter frost damages

QTL analyses using the subpopulation Pop2 (76 RILs) were performed on LSMeans values for the 11 traits corresponding to WFD scores in different experimental conditions as explained before. QTL mapping resulted in the detection of 20 QTLs located on LGIII, LGV and LGVI as shown in Figure II.3 and table II.1. All QTLs on each linkage group had an overlapping confidence interval (CI) allowing to define a global QTL corresponding to the same genomic region with a consensual CI for each LGs (Figure II.4). Thus, three genomic region related to frost tolerance located on LGIII, LGV and LGVI were named respectively WFD.III WFD.V and WFD.VI, with the corresponding CI (2-LOD drop-down) of 15.2, 8.7 and 12.6 cM respectively (Table II.1). Each consensual QTL was defined with at least three experimental conditions. Overall, LOD values ranged from 4.9 (mon0203 QTL on LGV) to 13.6 (dij0102 QTL on LGIII). The two consensual QTLs (WFD.V and WFD.VI) located on LGV and LGVI explained relatively the same phenotypic variance of QTL peak (R^2) with an average of R^2 equal to 27.4% and 29% respectively (Table II.1). WFD.III QTL was the most important one which was detected among 9 experiment conditions with a high R^2 ranged from 29.1 to 55.4%. All QTLs had a positive value of the additive effect for Térèse allele, indicating that the Térèse allele contribute to increase frost damages (Table II.1). Thus, favourable alleles for tolerance to winter frost damage underlying all detected QTLs were those given by the Champagne parent.



Figure II. 3 : QTL likelihood curve of LOD score for winter frost damage traits identified on *Pisum sativum* genome using cartographer version 2.5.

The y-axis represents likelihood-ratio test statistic (LOD score). The x-axis represents the locations of molecular markers (in cM) used in genotyping. Each linkage group is separating by a vertical line. The corresponding horizontal line (black) shows the significance threshold for the 11 studied traits. Each peak of the curve exceeding the significance threshold represents a QTL associating with the target trait.

GLIII



Figure II. 4: *Pisum sativum* genetic map and QTLs for winter frost damage (WFD) trait detected with a subset of 76 RILs of the Pop2 population in 11 experiment conditions.

Only linkage groups containing QTLs are showed. Markers names are indicated on the right of each linkage group (LG). For each LG, only some markers closest to QTLs intervals are shown on the genetic map. Marker positions are indicated in cM Haldane on the left on each LG. Individual QTLs (red hatched box) are identified by the corresponding experimental location abbreviation following the consecutive year of growing season. Each individual QTL is defined by two confidence intervals corresponding to 1-LOD (box) and 2-LOD (whisker) drop-down. Markers mapped in red corresponding to a 2-LOD drop-down and mentioned by the symbols WFD.III, WFD.V and WFD.VI. The blue boxes present the QTLs detected by Lejeune-Hénaut et al. (2008) with the whole population Pop2 (164 RILs).

Linkage group	QTL identification	LExperimentMarkers closest to the 2-LOD interval, marker at the QTL peak (in bold)		QTL peak position and 2- LOD drop-down interval (cM)	LOD ^a	R ² (%) ^b	Additive effect ^c	
3	WFD.III	mon0102	PsCam056487_37300_1026,	38.6 (35,0-44,0)	6.6	32.5	+0.52	
			PsCam057866_38405_252,					
			PsCam054784_36091_1160					
		mon0203	PsCam056676_37447_180, AA175-282,	34.6 (33,0-35,0)	5.8	29.1	+0.60	
			PsCam056487_37300_1026					
		mon0203	PsCam056487_37300_1026,	38.6 (35,0-44,0)	6.6	32.5	+0.52	
			PsCam057866_38405_252,					
			PsCam036264_21412_1640					
		cle0102	PsCam056487_37300_1026,	38.6 (36,1-39,5)	12.7	53.1	+1.02	
			PsCam057866_38405_252,					
			PsCam057845_38387_1581					
		cle0203	PsCam056487_37300_1026,	38.6 (36,1-39,9)	12.8	53.4	+1.08	
			PsCam057866_38405_252,					
			PsCam057845 38387 1581					
		dij0001	PsCam056487 37300 1026,	38.6 (35,5-40,0)	8.5	39.6	+0.75	
		-	PsCam057866_38405_252,					
			PsCam057845 38387 1581					
		dij0102	PsCam056487 37300 1026,	38.6 (36,8-39,9)	13.6	55.4	+0.87	
		-	PsCam057866_38405_252,					
			PsCam057845 38387 1581					
		dij0203	PsCam056487 37300 1026,	38.6 (37,6-41,7)	9.4	43.0	+0.90	
		-	PsCam057866 38405 252,					
			PsCam050518_33096_205					
		col0102	PsCam056487 37300 1026,	38.6 (36,4-39,8)	10.4	46.1	+0.77	
			PsCam057866_38405_252,					
			PsCam057845_38387_1581					

Table II. 1: List of QTLs identified for winter frost damage in *Pisum sativum* detected in 11 experiment conditions

-		col0102	PsCam057845_38387_1581,	42.9 (42,0-47,3)	7.2	34.8	+0.66
			PsCam050518_33096_205,				
			PsCam042960_26997_2211				
		lus0102	PsCam056487_37300_1026,	37.9 (36,0-41,3)	10.1	45.2	+0.52
			PsCam026085_15023_289,				
			PsCam057845_38387_1581				
		lus0102	PsCam036264_21412_1640,	46.4 (45,0-48,2)	6.1	30.6	+0.42
			PsCam052477_34775_186,				
			PsCam056640_37419_697				
5	WFD.V	mon0001	Ps001692, DHPS1 , PsCam000454_396_539	88.7 (84,2-90,9)	5.5	28.1	+0.48
		mon0203	PsCam036576_21710_2893, DHPS1,	88.7 (87,0-91,4)	4.9	25.3	+0.57
			PsCam000454_396_539				
		dij0203	Ps001915, PsCam044478_28344_1306 ,	85.9 (82,7-87,0)	5.7	29.0	+0.75
			PsCam036576_21710_2893				
6	WFD.VI	mon0001	PsCam055783_36734_228,	37.4 (29,8-39,0)	5.5	27.9	+0.46
			PsCam048554_31268_288,				
			PsCam003061_2398_595				
		mon0001	PsCam053897_35641_548,	49.9 (44,0-56,0)	5.6	28.3	+0.46
			PsCam051457_33949_1016,				
			PsCam056887_37599_419				
		mon0203	PsCam038121_23170_70, AD141,	46.8 (43,4-53,1)	5.9	29.8	+0.60
			M16_1300				
		dij0001	PsCam058552_38960_139, Ps001775 ,	50.8 (48,0-53,8)	6.0	30.3	+0.65
			PsCam049563_32190_2592				
		dij0203	PsCam038121_23170_70, AD141,	46.8 (43,4-54,3)	5.7	28.7	+0.72
			PsCam049563_32190_2592				

Table II.1: List of QTLs identified for winter frost damage in Pisum sativum (L.) detected in 11 experiment conditions (Continued)

^a LOD is Empirical LOD threshold value at the 0.05 alpha level by 500 permutation test; LOD is the log-likelihood at the QTL peak position. The LOD threshold, based on 500 permutations and a type of error of 5% were 4.7; ^b R^2 is the percentage of phenotypic variation explained by the QTL; ^c is the allelic value of Térèse; the additive effect indicates the effect of Térèse allele in comparison to Champagne allele. Thus, when this value is positive, the Térèse allele increase trait value.

3. Discussion

Several QTLs described by Lejeune-Hénaut et al. (2008) with 164 Pop2 lines were used to validate previous QTLs and to have more precise information of QTL positions. In the present study, phenotypic and genotypic data set collected from the subset of 76 Pop2 lines allowed to confirm three of the four consistent WFD QTL detected with the entire population by (Lejeune-Hénaut et al. 2008), namely WFD.III, V and VI. Common markers between the two QTL analyses were located within their peak QTLs thus that confirm collinearity of these three consistent QTLs: AA175-282 for WFD.III, DHPS1 for WFD.V and AD141 for WFD.VI. Other novel SNP markers closest to each QTL peak were identified allowing to precise QTL position. Therefore, the use of accurate marker density for this study showed differences in confidence intervals (CIs) of common QTLs comparing to the previous reported WFD QTLs especially the QTL located on LGV and LGVI which were refined respectively from 16.1 cM (WFD5.1) to 8.7 cM (WFD.V) and from 19.2 cM (WFD.III) was detected with higher size (15.2 cM) than that previously identified in Lejeune-Hénaut et al. (2008) (WFD3.1, 8 cM) (Figure II.4).

For a given QTL, many genes could be found within the confidence interval focusing on markers with a known function localized on the genetic map (Tayeh et al. 2015a). For the three detected QTLs WFD.III, V and VI, 732 markers were annotated with a known function according to Tayeh et al. (2015a). Among these annotated markers, it is important to focus on 6 markers which encoded to the *trypsin inhibitor activity* gene whose 1 marker was located on LGIII and LGVI, and 4 markers located on LGVI. In fact, it will be difficult to consider that these markers will be good candidate markers for breeding program for improvement of frost tolerance in pea since many studies have shown that trypsin inhibitor activity decreases the nutritional value of pea seeds (Page et al. 2002). Thus, caution should be exercised in the use of these QTLs in a markers-assisted breeding program.

The present study confirmed that WFD.III QTL colocalized with the flowering locus *Hr* (*High response to photoperiod*) as showed previously by (Lejeune-Hénaut et al. 2008). Besides, Weller et al. (2012) demonstrated that *Hr* locus is an orthologue of the Arabidopsis *Early flowering 3*, *Elf3*, thus it would be a relevant candidate for WFD.III QTL as it allows plants to be maintained in a vegetative state under short days and thus to escape the main winter freezing periods.

In addition, four transcription factors called C-repeat binding factor (CBF), have been annotated in the confidence interval of the QTL for WFD.VI. These CBF genes will be good candidate genes for frost tolerance as they have a major role in freezing tolerance. In fact, it has been demonstrated that the CBF genes regulate cold acclimation response and which in turn promotes tolerance to freezing. Many studies have reported significant role of CBF/DREB1 genes in freezing tolerance in different plant species. Tayeh et al. (2013b) demonstrated that a major freezing tolerance QTL region on Medicago truncatula chromosome 6 was governed by a cluster of 12 tandemly duplicated CBF/DREB1 genes. In addition to that, 9 markers underlying CIs of detected QTLs have been annotated as a MYB transcription factor proteins which known as a vital family of transcription factors playing critical roles in plant development and stress response (Ambawat et al. 2013; Wang et al. 2017). In fact, Shinwari et al. (1998) showed that the promoters of CBFs contain MYB recognition sequences, suggesting that a MYB-related transcription factor may also be involved in the cold induction of CBFs. In that sense, Chen et al. (2013) suggested that MYB14 participates in freezing tolerance in Arabidopsis by affecting expression of *CBF* genes. Indeed, the authors demonstrated that the knock-down of MYB14 by artificial microRNA increased the tolerance to freezing stress (Chen et al. 2013).

Moreover, 11 markers located in the confidence interval of the three QTLs detected in the present study were located within a functional gene encoding a *brassinosteroid receptor* which known to have a role in increased plant tolerance against freezing stress (Eremina et al. 2016b).

4. Conclusion

In summary, our results confirm previous study of QTL mapping for frost tolerance in field conditions from scoring winter frost damages trait, especially three main QTLs located on LGIII, V and VI, namely WFD.III, WFD.V and WFD.VI respectively. This study provided important data for positional cloning of genetic determinants of frost tolerance in pea underlying these three main QTLs. Indeed, by using a massive number of SNP markers, suitable candidate markers have been identified and can be used in future studies in marker assisted selection (MAS) and near-isogenic lines (NILs) creation. These candidate markers will help especially in the reduction of CIs of genomic regions associated with trait of interest.

B. An attempt to refine the confidence interval of the frost tolerance QTL WFD6.1 using near-isogenic lines (NILs)

1. Material and methods

1.1. Development of NILs for WFD6.1 QTL

In order to develop NILs containing the frost tolerance WFD6.1 located on LGVI, a cross was made between a freezing sensitive genotype named Eden and a Recombinant Inbred Line (RIL) selected from the Pop2 population named RIL2-38. This RIL was chosen because it contained the favorable allelic version of the targeted region, i.e. the allelic version from the tolerant line Champagne and additionally because its genetic background was composed by the sensitive line alleles (Térèse) which could limit the presence of favorable alleles at other QTL regions than the targeted QTL. The procedure for introgression of the WFD6.1 QTL is shown in figure II.5. The donor line RIL2-38 was crossed with the recipient line Eden. Plants from this cross (F1) were then backcrossed five times (BC5) with the recipient parent. At each backcross generation, plants were genotyped using microsatellite and RAPD markers for the first two backcrosses (AA200, AD159, AD60, AD59, I01-600: Lejeune-Hénaut et al. 2008) and two SSCP markers for backcrosses 3 to 5 (MTIC153F2 and NT6011: Tayeh et al. 2013a) in order to identify plants with heterozygous genotypes at the targeted QTL. Then, BC5 progenies were self-pollinated for two generations to select BC5S1 or BC5S2 plants carrying homozygous alleles at markers linked to WFD6.1 and potential recombinants within the targeted region. During the selfing generations, genotyping was carried out with markers described in Duarte et al. (2014) and in Tayeh et al. (2015a). From the BC5S1 or BC5S2 progenies, selected homozygous plants were self-pollinated every year to increase the number of seeds necessary for phenotyping tests.

Lines were compared between each other and with the parental plants (control lines) to identify homozygous haplotypes presenting a recombination event at the targeted region. Then, NILs were classified into groups based on the genotypic data for tested markers at the QTL. NILs of the same group had the same genotype for all tested markers.

Chapitre II : Re-detection de QTL de tolérance au gel chez le pois et cartographie fine



Figure II. 5: Marker-assisted backcrossed scheme used to develop NILs at the frost tolerance QTL WFD6.1.

1.2. Verification of the genetic background of WFD6.1-NILs

A high density genotypic analysis was performed to compare the genome of the five different near-isogenic mother plants of the BC5 generation. This comparison allowed to evaluate the size of the introgression at the targeted position on LGVI, and to evaluate the return to the Eden

Chapitre II : Re-detection de QTL de tolérance au gel chez le pois et cartographie fine

genetic background for the rest of the genome. Along with the analysis of the BC5 mother plants, RIL2-38 was also compared to its two parental lines namely Champagne and Térèse.

Genotyping was performed by molecular characterization using the GenoPea Infinium® BeadChip which comprises 13204 SNPs distributed over all the pea genome. The consensus map (Tayeh et al. 2015a) was used to order the markers along the seven pea linkage groups and to establish the graphical genotypes of the BC5 mother plants and controls.

SNPs were analyzed individually for each of the genotypes using the Illumina GenomeStudio® software taking into account the polymorphism observed between the parental lines Eden and RIL2-38. The parental line RIL2-38, as it was issued from the cross between Champagne (Ch) and Térèse (Te) as described previously, should have the same genotypes as observed with Champagne or Térèse. After the analysis of genotyping data (Figure II.6), the results were further exploited by constructing a graph to summarize the genotype of the studied pea lines (Figure II.7). Marker positions on the consensus map as well as the genotyping data of the NILs, were used to realize the graphic genotypes of studied lines. The percentage of the donor genome at the QTL WFD6.1 region and the percentage of the return to the recipient genetic background outside the targeted QTL region were measured for each near-isogenic line.



Genotyping table extracted from GenomeStudio® software								
Marker name	Interpretation of tested marker	Eden (Ed)	RIL2-38	F1_Ed x RIL2-38	Champagne (Ch)	Térèse (Te)	F1_Ch x Te	
PsCam057977_38501_1113	Polymorphic	AA	BB	AB	AA	BB	AB	

Genotyping table extracted from GenomeStudio® software								
Marker name	Interpretation of	Eden	RIL2-38	F1_Ed x	Champagne	Térèse	F1_Ch	
	tested marker	(Ed)		RIL2-38	(Ch)	(Te)	x Te	
PsCam000024_19_1205	Monomorphic	BB	BB	BB	BB	BB	BB	

Figure II. 6: Check of the genotyping with GenomeStudio® software. Four examples of analysis performed after checking the position of the control lines (Champagne and Térèse), and the parental lines of studied NILs (RIL2-38 and Eden) in the graph.

(a): Polymorphic; The marker is polymorphic between the two parental lines of studied NILs; in this example RIL2-38: BB, Eden: AA and the F1: AB. The RIL2-38 line has the allele of Térèse. (b): Monomorphic; The marker is monomorphic between the two parental lines of studied NILs; in this example RIL2-38: BB, Eden: BB and the F1: BB. The marker is also monomorphic for the control lines Champagne and Térèse.

Genotype AA; Genotype BB and Genotype AB


Figure II.6: Check of the genotyping with GenomeStudio® software. Four examples of analysis performed after checking the position of the control lines (Champagne and Térèse), and the parental lines of studied NILs (RIL2-38 and Eden) in the graph (*Continued*).

(c): The marker is polymorphic between the two parental lines of studied NILs; in this example RIL2-38: BB, Eden: AA and the F1: AB. But, this marker is monomorphic for the control lines Champagne (Ch) and Térèse (Te): in fact, RIL2-38 line has a different allele from his parental lines Ch and Te. (d): Not operable; error of genotyping.

Genotype AA; Genotype BB and Genotype AB



Figure II. 7 : Graphic genotypes of the five set of NILs at WFD6.1 QTL

Genotypes of parent lines of studied NILs:

RIL2-38 (alleles from Te: or Ch:; exogenous allele: other than those from Ch and Te); Eden (Ed:; homozygous AA or BB);

F1: Ed x RIL2-38 (: heterozygous AB); Monomorphic marker:

Genotypes of NILs:

E Polymorphic marker showing the return to the Eden genetic background; Polymorphic marker showing an introgression of the donor RIL2-38 allele (heterozygous form);

E: Polymorphic marker showing a different allele from Te and Ch presented by RIL2-38; 🗌 : Unknown genotype

Confidence interval of winter frost damage (WFD) QTLs detected by (Lejeune-Hénaut et al. 2008) in centiMorgan (cM)

Confidence interval of WFD QTLs re-detected in this study in centiMorgan (cM)

^a Genetic position of the last polymorphic marker used in the analysis (Pop2 genetic map, Tayeh et al. 2015a); ^b Total length of the linkage group (Pop2 genetic map, Tayeh et al. 2015a)



Figure II.7: Graphic genotypes of the five set of NILs at WFD6.1 QTL (Continued).

1.3. Phenotyping of the NILs for winter frost damages

To evaluate the NILs for their performance under winter frost events, NILs with homozygous alleles of RIL2-38 and/or Eden for the target interval (BC5S3-6) were sown in the fields in seven experimental conditions. The experiments were carried out at the INRA experimental stations of Clermont-Ferrand (Theix), Mons and Dijon, as well as at the Agri-Obtentions and Rouergue Auvergne Gévaudan Tarnais (RAGT) experimental stations of Orsonville and Rodez respectively. Experiments were conducted in the four consecutive growing seasons 2014/2015, 2015/2016, 2016/2017 and 2017/2018. In this study, we present only experiments that presented significant differences of freezing effects on plants, especially the two experimental conditions of Theix in 2014/2015 and Orsonville in 2017/2018. The two parents, RIL2-38 and Eden, were included in all blocks as controls. Plants were grown in plots following a randomized complete block design with two replicates for all experimental conditions. The effect of winter frost was carried out by visual estimation of winter frost damages after each winter frost period with scores varying from 0 to 5 as described in Devaux and Lejeune-Hénaut (2010). Statistical analyses of phenotypic data were performed using R software. For each experimental condition; phenotypic data were analyzed using the linear model 'aov' function in the 'lme4' package, considering the WFD scoring data as explanatory variables and the genotype and replicate as fixed factors. Then, phenotypic means from recombinant, non-recombinant and control lines were compared between each other using the Student-Newman-Keuls (SNK) comparison test $(\alpha = 5\%)$ of the 'SNK' function available in the 'agricolae' package (de Mendiburu 2017).

For fine mapping, recombinant lines at the QTL interval were compared with their phenotyping data to identify informative recombination breakpoints.

2. Results and discussion

2.1. Genotype of the five near-isogenic mother plants of the BC5 generation

In order to analyze genotyping data results of the five near-isogenic mother plants of the BC5, a representation of molecular markers distributed on the seven linkage groups was performed (Figure II.7). The genotype of these plants were compared with their parents to verify whether the QTL of frost tolerance WFD6.1 is well introgressed in the recipient background. As a first step, the results of each linkage groups containing the three main frost tolerance QTLs WFD3.1, WFD5.1 and WFD6.1 will be detailed, then those not containing QTLs will be also exploited.

The evaluation of the number of polymorphic, monomorphic and non-operable SNP markers against the total number of studied SNPs was performed for each of the seven pea linkage groups (Table II.2). The figure II.7 presents the result of genotyping analysis for near-isogenic mother plants and control lines. Only polymorphic markers between Eden and RIL2-38 were showed in order to simplify the presentation on the graph. These markers are coded by colors representing allelic variability. Thus, for each near-isogenic mother plants, it is possible to identify regions of the genome that have returned to the Eden background (light blue color corresponding to the two Eden alleles) or that have introgressed an allele of the parent RIL2-38 (light green color corresponding to an Eden allele and an RIL2-38 allele showing heterozygosis, since the genotyping was done at BC5 generation).

	Number of total markers used for the genotyping analysis	Number of polymorphic markers between Eden and RIL2-38	Number of monomorphic markers between Eden and RIL2-38	Number of not operable markers
LGI	1460	307	1100	53
LGII	1877	445	1412	20
LGIII	2295	771	1458	66
LGIV	1948	404	1519	25
LGV	1653	388	1217	48
LGVI	1665	612	1018	35
LGVII	2086	467	1571	48

Table II. 2: Characteristics of markers used in the genotyping analysis of NILs in BC5 generation

The Table II.3 presents the percentages of the targeted QTL introgression in the heterozygosis form and the return to the recipient genetic background for all studied near-isogenic mother plants.

The LGVI containing the targeted QTL showed that near-isogenic mother plant (NIMP) 1, 2 and 3 introgressed more than the size of WFD6.1 confidence interval, while NIMP 4 and NIMP 5 introgressed a part of it (21%) (Figure II.7 and Table II.3). Concerning the other frost tolerance QTLs located on LGIII and LGV, the graph indicates that there is a total return to the Eden genetic background within and outside the two QTL WFD3.1 and WFD5.1 confidence interval for each of studied plants.

Chapitre II : Re-detection de QTL de tolérance au gel chez le pois et cartographie fine

Isogenic- mother plant in BC5 generation	Rate of the return to Eden genetic background at the WFD3.1 CI*	Rate of the return to Eden genetic background at the WFD5.1 CI*	Rate of the introgression of RIL2-38 allele at the WFD6.1 CI*	Rate of the return to Eden genetic background for all the genome
1	100%	100%	116%	100%
2	100%	100%	175%	100%
3	100%	100%	119%	74%
4	100%	100%	21%	95%
5	100%	100%	21%	100%

Table II. 3: Evaluation of WFD6.1 introgression and the return to Eden genetic background

*confidence interval

The second linkage group (LGII) presents a total return of the genetic background Eden. Overall, LGIV showed a relative return for Eden genome except one marker for all tested plants which was heterozygous. Regarding LGI, the graph indicates that all genotypes of near-isogenic mother plant returned to Eden genetic background except the mother plant 3 which showed an introgression of RIL2-38 alleles with a percentage of 26% (Table II.3). Linkage group VII showed a total return to the Eden genetic background for the studied plants 1, 2, 3 and 5. For this linkage group, the near-isogenic mother plant 4 has two regions of heterozygosis with 5% of introgression of RIL2-38 alleles (Table II.3).

Furthermore, the analysis of the genotyping results showed the presence of novel alleles conferred by the parental line RIL2-38. Indeed, the RIL2-38 presented other alleles than those of its parental lines Térèse and Champagne (showed by dark orange in Figure II.7). These exogenous alleles could be explained by a contamination by exogenous pollen during the production of Pop2 population. Regarding the five near-isogenic mother plant, we noted that the plant 2 and 4 inherited some of these exogenous alleles which were located in the LGVI and LGVII respectively.

From theses results of backcrossing series, we can conclude that the isogenic-mother plant 3 was the most suitable plant for further studies for further analysis of fine mapping, as it presented a full return for the genetic background with an introgression of the tolerance alleles at the targeted region without any exogenous alleles.

2.2. Attempted fine-mapping of the frost tolerance WFD6.1 QTL

Following the genotyping and the availability of seeds, four groups of near isogenic plants were selected for fine mapping. Each group were presented by NILs which have the same genotypes at the homozygous form for tested markers at the targeted region. The table II.4 showed the genotyping results of selected sets of NILs genotyped by markers belonging the confidence interval of targeted QTL. Based on markers used in this study, screened groups of NILs correspond to 3 different haplotypes which one carried a recombination event within the confidence interval of targeted QTL. This recombinant haplotype was represented by two groups of NILs (E601 and E6C) exhibiting the same recombination between the two markers MTIC153F2 and Ps001357. Thus, only one recombinant haplotype will contribute to refine the confidence interval of WFD6.1 QTL. For the two other groups of NILs, the first one (E6A) presented non-recombinant haplotype with all Eden alleles in the targeted region contrary to the last set of NILs which presented a marker haplotype with all RIL2-38 alleles.

The four groups of NILs with the parental lines Eden and RIL2-38 were phenotyped for winter frost damages under two experimental conditions. In the experimental condition of Theix/2014-2015, not all groups of NILs were tested for frost tolerance because of a limited number of seeds (Table II.4). The analyses of variance (Table II.5) showed significant variation between tested plants for winter frost damage. The block effect is not significant (P > 0.05) for the three experimental conditions which highlighted the homogeneity of the experimental design. The error, presenting the variability due to all other uncontrolled effects, had low values which indicated that the model was accurate.

			Ger	notype at	WFD6.	1 QTL			-	
		Marker name	PsCam028287	Ps001552	PsCam037030	MTIC153F2	Ps001357	Ps001393		
		Genetic position*	46.9	47.4	49.1	50.8	50.8	54		
		Genetic position**	45.7	47.1	47.8	47.1	47.1	51.9	Mean (± SD) of W of SN	VFD trait and group IK test
		Mother plant***							Theix/2014-2015	AO/2017-2018
	Eden		AA	GG	GG	TT	CC	GG	$2.50\pm0.70~b$	4.75 ± 0.35 b
	RIL2-38		GG	AA	AA	CC	AA	AA		3.12 ± 0.53 a
	E601	5	GG	AA	AA	CC	CC	GG	$2.35\pm0.70\ b$	3.50 ± 0.00 a
Group of NILs	E6A	1,3 and 5		GG			CC	GG	$3.50 \pm 0.70 \text{ b}$	4.75 ± 0.35 b
	E6B	1 and 2		AA			AA	AA	1.00 ± 0.00 a	3.50 ± 0.00 a
	E6C	4	GG		AA	CC		GG	1.00 ± 0.00 a	

Table II. 4: Phenotypic and genotypic evaluation of studied NILs under Theix/2014-2015 and Orsonville /2017-2018 experimental conditions

*Genetic position in the consensus map (in cM); **Genetic position in the Pop2 map (in cM); ***Isogenic-mother plant in BC5 generation

	Theix	x/2014-15	Orso	Orsonville/2017-18				
	Df	Mean square	Pr	Df	Mean square	Pr		
Genotype	4	2.29	1.08E-02*	4	1.18	6.01E-03*		
Replicate	1	0.9	7.05E-02	1	0.30	8.00E-02		
error	4	0.15		4	0.05			

 Table II. 5: Analyses of variance (mean square) for winter frost damage scores of studied pea lines under different conditions

* Significant at the threshold of P < 0.05

Near-isogenic groups with recombinant and non-recombinant haplotypes were compared between each other and with the two parental lines to distinguish significantly the behavior of NILs under frost conditions. Results of the statistical mean comparison (SNK test) of WFD scores for the different haplotypes of NILs and parental lines were presented in the table II.4.

Comparisons of the phenotypes and marker genotypes of tested plants revealed that:

- The recipient line Eden had the highest values of frost damage scores with an average of 3.62 for the two studied conditions. For the donor parent RIL2-38, tested only at the Orsonville/2017-2018 condition a significant difference of phenotyping scores compared to Eden has been observed with a frost damage score of 3.12. Thus, RIL2-38 alleles, inherited from Champagne, could decreased frost damage scores compared to Eden which was consistent with the allele effect estimated for this QTL (Table II.4.).

- Near-isogenic lines of the group E6A, which were fixed with Eden alleles in the confidence interval of WFD6.1, are ranking in the same phenotypic group as the sensitive line Eden with an average of WFD of 4.33.

- In the experiment condition Orsonville /2017-2018 where the parental line RIL2-38 was tested as control line with NILs, we noted that the groups of NILs E601 and E6B were ranking in the same phenotypic group as the donor line RIL2-38. Furthermore, this result enabled to note that the recombinant haplotype presented by the near-isogenic group E601 was included in the same phenotype group as the non-recombinant haplotype E6B which are fixed with RIL2-38 alleles for all tested markers. Consequently, by eliminating the markers Ps001357 and Ps001393, WFD6.1 interval could be delineated to a little smaller region between the two markers PsCam028287 (45.7 cM) and MTIC153F2 (50.8 cM). In this reduced region the phenotype was in coherence with genotypic information for all tested lines.

3. Conclusion

From these results, we can conclude that targeted region coincided with one recombination event in the fine-mapping population that affected the reduction of the confidence interval of WFD6.1. The low recombination within this targeted region limited the resolution we achieved and discouraged further efforts for screening of additional recombinants in this genetic background. In fact, this absence or very limited occurrence of recombination events could be due to the proximity of the centromere. In that sense, Limborg et al. (2016) and (Tayeh et al. 2015a) declared that the rate of recombination and distribution of markers over linkage groups usually indicates the position of centromeres: low recombination frequencies and so large marker clusters are often observed in centromeric regions. In case of our study, (Tayeh et al. 2015a) showed that there was a large cluster of 291 markers in the middle of pea LGVI suggesting an absence of recombination and explained by a high number of co-localizing markers. In that sense, a maximum of 58 and 47 markers co-localizing at the same positions of 47.1 cM and 47.8 cM respectively, was observed on the individual genetic map Pop2 of LG IV.

Article associé

Genome wide association mapping highlights new loci determining frost tolerance in pea and expands the range of sources of tolerance

Sana Beji¹, Véronique Fontaine², Rosemonde Devaux³, Martine Thomas³, Sandra Negro^{5,6}, Nasser Bahrman², Mathieu Siol⁴, Grégoire Aubert⁴, Jean-Louis Hilbert¹, Bruno Delbreil¹, Isabelle Lejeune-Hénaut^{2*}

¹ USC ICV-AFP, Université Lille, 59655, Villeneuve d'Ascq, France

² USC ICV-AFP, INRA, 80203, Estrées-Mons, France

³ UE GCIE, INRA, 80203, Estrées-Mons, France

⁴ UMR Agroécologie, INRA, 21065, Dijon, France

⁵ UMR GQE, INRA, 91190, Gif sur Yvette, France

⁶ UMR MIA-Paris, AgroParisTech, 75005, Paris, France

SB: sana.beji@univ-lille.fr

VF: veronique.fontaine@inra.fr

RD: rosemonde.devaux@inra.fr

MT: martine.thomas@inra.fr

SN: sandra.negro@inra.fr

NB: nasser.bahrman@inra.fr

MT: matthieu.siol@inra.fr

GA: gregoire.aubert@inra.fr

JLH: jean-louis.hilbert@univ-lille.fr

BD: bruno.delbreil@univ-lille.fr

ILH: isabelle.lejeune-henaut@inra.fr

* Corresponding author:

Isabelle Lejeune-Hénaut USC ICV-AFP, INRA 2, Chaussée Brunehaut, Estrées-Mons BP 50136, 80203 Péronne, France E-mail: isabelle.lejeune-henaut@inra.fr Tel: +333 22 85 75 07

1. Abstract

Background: Frost is a limiting abiotic stress for the winter pea crop (*Pisum sativum* L.) and identifying the genetic determinants of frost tolerance is a major issue to breed varieties for cold northern areas. Quantitative trait loci (QTLs) have previously been detected from bi-parental mapping populations, giving an overview of the genome regions governing this trait. The recent development of high-throughput genotyping tools for pea brings the opportunity to undertake association genetics studies in order to capture a higher allelic diversity within large collections of genetic resources as well as to identify more precise confidence intervals for QTLs thanks to numerous markers.

Methods: In this study, a genome wide association study (GWAS) was performed using a set of 365 pea accessions. Phenotyping was carried out by scoring frost damages in field and controlled conditions. Besides, the association mapping collection was also genotyped using an Illumina Infinium® BeadChip, which allowed to collect data for 11366 single nucleotide polymorphism (SNP) markers.

Results: GWA analyses identified 62 SNPs significantly associated with frost tolerance and distributed over six of the seven pea linkage groups (LGs). These results confirmed 3 QTLs that were already mapped in multiple environments on LG III, V and VI with bi-parental populations. They also allowed to identify one locus, on LG II, which has not been detected yet and two loci, on LGs I and VII which have formerly been detected in only one environment. Fifty candidate genes corresponding to annotated significant SNPS, or SNPS in strong linkage disequilibrium with the formers, were found to underly the frost damage (FD)-related loci detected by GWAS. Additionally, the analyses allowed to define favourable haplotypes of markers for the FD-related loci and their corresponding accessions within the association mapping collection.

Conclusion: This study led to identify FD-related loci as well as corresponding favourable haplotypes of markers and representative pea accessions that might to be used in winter pea breeding programs. Among the candidate genes highlighted at the identified FD-related loci, the results also encourage further attention to the presence of *CBF* transcriptions factors as potential genetic determinants of the frost tolerance locus on LGVI.

Keywords: Frost damages, frost tolerance, Genome Wide Association Study (GWAS), Pea (*Pisum sativum* L.), Quantitative trait loci (QTL), haplotypes of markers, candidate genes.

2. Background

The dry pea (*Pisum sativum* L.) is the second grain legume after soybean, with a world production of about 14.3 millions tons grown on 7.6 millions hectares in 2016 [1]. Average seed yield worldwide was about 1883 kg/ha in 2016, with the highest yields achieved in the Northern European countries [1]. Dry peas are an important nutritional source which provides high quality protein for humans and for animal feeding [2]. In addition to the economic importance of pea seeds, pea crops have beneficial environmental impacts, mainly due to their ability to fix atmospheric nitrogen. They do not need nitrogen fertilizers and therefore help reducing N2O emissions [3]. For the past decades, spring sowing has been the most common method of cultivation for dry pea. However, the relatively short duration of the development cycle and various stresses, particularly water stress and high temperatures at the end of the development cycle, are at the origin of grain yield losses and variation [4]. Nowadays, winter peas are being developed in order to obtain higher and more stable yields. They are however limited by freezing temperatures during the winter time and the development of winter pea genotypes able to overcome freezing periods is thus desirable.

Mechanisms of tolerance to freezing temperatures have already been reviewed in many plant species. Plants can tolerate freezing temperatures using non-exclusive strategies: freezing escape and cold acclimation. Indeed, plants can escape freezing stress by delaying sensitive phenological stages, particularly floral initiation and flowering, given that frost sensitivity increases after floral initiation [5, 6]. Cold acclimation is the process by which certain plants increase their frost tolerance in response to low non freezing temperatures [reviews : 7, 8, 9]. The *CBF* (C-repeat /dehydration responsive element Binding Factor) genes have an important role in plant cold acclimation. These genes have been isolated first from Arabidopsis thaliana and belong to the AP2/EREBP (APETALA2/Ethylene-Responsive Element Binding Protein) family of transcription factors [10, 11]. In Arabidopsis thaliana, the CBF pathway is characterized by rapid cold induction of *CBF* genes which alter the expression of CBF-targeted genes, known as the *CBF* regulon, which in turn contribute to an increase in freezing tolerance [12]. Many studies have reported the significant role of *CBF* genes in freezing tolerance of

herbaceous and woody plant species. Among these studies, the biological role of the *CBF* pathway for freezing tolerance has also been underlined by the colocalization of *CBF* genes with freezing tolerance quantitative trait loci (QTL) ([Arabidopsis: 13, temperate cereals : 14, 15-17, forage grasses : 18, legumes : 19]. Moreover, within the temperate cereals, *CBF* genes underlying FR2, a major homeologous frost tolerance QTL in barley, diploid and hexaploid wheat, are known to be structured in a cluster of tandemly duplicated genes [20]. In legumes a similar feature was found in *Medicago truncatula*, where a cluster of 12 tandemly duplicated *CBF* genes was shown to match with a major freezing tolerance QTL on chromosome 6 [19].

The identification of genomic regions controlling frost tolerance has initially been completed for cultivated species through the assessment of mapping populations and QTL mapping. In pea, QTL mapping studies for frost damages have been conducted in multiple field environments as well as in controlled conditions. QTLs were detected using two populations of recombinant inbred lines (RILs), namely RIL2 and RIL9, respectively derived from crosses between Champagne (frost tolerant) and Térèse (frost sensitive) on the one hand [21, 22] and China (frost tolerant) and Caméor (frost sensitive) on the other hand [23]. Four common QTLs were detected within both populations. The corresponding regions were located on linkage groups (LG) III (two independant positions), V and VI. They explained altogether a major part of the phenotypic variation and were repeatable across environmental conditions. Three other loci were specific to one or other of the two populations and detected in fewer environments (RIL2 : one locus on LG I, RIL9 : one locus on LGVII, detected in one environment ; RIL9 : one locus on LGV, detected in two environments).

The resolution and accuracy with which QTL mapping can identify causal genetic determinism of the considered traits is limited by the high confidence intervals and the relatively low total number of recombination events presented in bi-parental mapping populations. In addition to QTL mapping, association studies have emerged as a complementary approach to dissect quantitative traits by exploiting natural genetic diversity and ancestral recombination events present in germplasm collections. Used for long in human genetic research, they have been adapted for genetic dissection in plants, taking advantage of the development of highthroughput genotyping resources in number of species [24]. Genome wide associations studies (GWAS) are based on the detection of linkage disequilibrium (LD) between genetic markers and genes controlling the phenotype of interest. They rely on high-density genetic maps allowing an increased resolution of detection generating more precise QTL positions than bi-

parental QTL mapping and give access to multiple allelic variation through the exploration of natural genetic diversity. In addition, GWAS can offer a powerful genomic tool for breeding plants by the identification of associated markers tightly linked to targeted genomic regions which can be used for marker-assisted selection. In the recent years, GWAS has been conducted in many plant species to dissect complex quantitative traits among which winter survival and frost tolerance [25-33]. High throughput genotyping resources now available in pea [34] have also allowed to carry out GWAS in order to dissect the genetic determinism of resistance to Aphanomyces euteiches, plant architecture and frost tolerance. Desgroux et al. [35] studied associations for resistance to A. euteiches in 175 pea lines using a high-density SNP genotyping consisting of 13.2 SNPs from the developed GenoPea Infinium ® BeadChip [36]. The authors identified several markers significantly associated with A. euteiches harboring relevant putative candidate genes and they were able to refine the confidence interval of QTLs previously detected in bi-parental populations. Using the same SNP resource and a 266 pea collection including the 175 former lines, they also identified genomic intervals significantly associated with plant architecture and resistance to A. euteiches, of which 8 were overlapping for both traits [37]. In a different genetic background composed by a set of 672 pea accessions genotyped with 267 SSR markers, Liu et al. [38] detected 7 SSRs significantly associated with frost tolerance of which one was located on LG VI and was shown to co-localize with a gene involved in the metabolism of glycoproteins in response to chilling stress.

The present study aimed at reconsidering the regions of the genome that control frost tolerance in pea using a GWAS approach. A diversified collection consisting of 365 pea accessions was evaluated for frost damages in the field and in controlled conditions. GWAS was performed using genome wide distributed SNPs generated from the 13.2K GenoPea Infinium® BeadChip [36]. Then, haplotype analysis was conducted and the phenotype differences among all haplotypes were tested by ANOVA. By this way, favourable haplotypes for frost tolerance in pea, as well as accessions carrying these haplotypes were identified, aiming to be used for future breeding programs to improve frost tolerance in pea. Finally, numbers of candidate genes underlying defined haplotypes related to the frost tolerance were determined.

3. Materials and Methods

3.1. Plant material

The association mapping collection, here after named the collection, consisted of 365 accessions (Additional file 1) from the pea reference collection, described in Burstin et al. [39]. *Pisum* accessions from the collection represents a large genetic diversity ranging from wild peas (*Pisum fulvum, P. humile, P. elatius, P. speciosum, P. transcaucasicum* and *P. abyssinicum*) and local populations to breeding lines and cultivars. This collection also represents a variability of genotypes based on the type of sowing (winter *vs* spring peas) and the type of end-use (fodder, field, mangetout, preserve and garden peas). Reference accessions being the parents of bi-parental populations formerly used in QTL studies for frost tolerance [21-23], i.e. Champagne, China, Térèse and Caméor, are included in the association mapping collection.

3.2. Phenotyping

Frost tolerance of the collection was evaluated under field and controlled conditions. The field experiment was carried out at the INRA experimental station of Clermont Ferrand-Theix, France, during the growing season of 2007-2008. Plots were sown in a randomized complete block design with three replicates. The collection was assessed for frost tolerance by visual estimation of winter frost damages after each winter frost period. A score was assigned to a plot as a whole, based on the extent of necrotic areas of the aerial parts of the plants according to a scale ranging from 0 to 5 where 0 represented no damage and 5 a dead plant [21]. Frost damages observations were realized at four dates in 2008: January, 4th and 15th, March, 28th and April, 10th.

A frost experiment was also conducted in a controlled environment chamber using the standardized test described previously by Dumont et al. [22]. Pea accessions were arranged in a randomized complete block design with three replicates. Briefly, the plants at the stage of 2^{nd} - 3^{rd} leaf were first treated with a regime of 11 days of cold acclimation at 10°C/2°C (day/night) with a 10 h photoperiod. The frost treatment was then carried out at 6°C/-8°C with 8 h of daylight during four days. After frost, a recovery period was applied with a temperature regime of 16°C/5°C and, 10 h of daylight during 8 days. Frost tolerance was evaluated by scoring frost damages at the end of the recovery period with the same scoring scale as the one used to

evaluate the frost damages in the field experiment, except that scores was attributed to single plants instead of plots.

Overall, 5 traits constituted the phenotyping data for the GWAS. Traits were abbreviated as follows: FD_CC: frost damages in the controlled conditions experiment, FD_Field_date1, FD_Field_date2, FD_Field_date3 and FD_Field_date4: winter frost damages evaluated in the field experiment at the first, second, third and fourth date respectively.

3.3. Phenotypic data analyses

The phenotypic data were analyzed with the R 3.5.0 software [40, 41] using a linear mixed model to obtain estimates of variance components, heritability (H²), as well as best linear unbiased predictions (BLUPs) of adjusted means. The following Linear Mixed Model (LMM) was used: $Y_{ij} = \mu + \text{geno}_i + \text{rep}_j + e_{ij}$, where Y_{ij} is the value of frost damages recorded for the genotype *i* at the replicate *j*. μ is the mean, geno_i is the random genetic effect of the genotype *i*, rep_j is the fixed replicate effect of the replicate *j* and e_{ij} is the residual effect. The model was carried out using the R function 'lmer' of package lme4 [42, 43]. Heritability (H²) was estimated using the following formula: $H^2 = V_g / [V_g + (V_{res} / n_{rep})]$, where V_g is the genotypic variance component, V_{res} is the residual variance component and n_{rep} is the number of replicates taking into account the missing values. BLUPs for each genotype-trait combination were calculated from each LMM analysis using the function "ranef" implemented in lme4 package of R software [43] and were used for GWA analysis.

3.4. Genotyping and quality control

The collection was genotyped using 11366 SNPs which were included in the Illumina Infinium[®] BeadChip 13.2K SNPs as described in [36]. These SNPs were all located in gene-context sequences and derived from separated transcripts [44]. The consensus genetic map from Tayeh et al. [36] was used as the genetic framework for the association analyses.

The genotyping matrix, which was composed of a set of 11366 SNPs and 365 pea accessions, was filtered using Plink v.1.9 software [45, 46]. Accessions and SNP markers with call rate below 0.90 as well as SNP markers with minor allele frequency (MAF) below 0.05 were excluded from the GWA analysis. After quality control checking, a data set of 10739 SNPs and 363 accessions was kept for further analyses. The resulting data set was further imputed using

Beagle v.3.3.2 software [47]. Beagle applies a Markov model to the hidden states (the haplotype phase and the true genotype) along the chromosome using an EM (Expectation-Maximization) algorithm that iteratively updates model parameters to maximize the model likelihood up to the moment where convergence is achieved. Finally, a genotyping matrix consisting of 10739 SNPs and 363 accessions with no missing data was used for GWAS.

3.5. Linkage disequilibrium estimation

The estimates of linkage disequilibrium (LD) within the association mapping collection were determined by the squared allele-frequency correlations (r^2) for pairs of loci as described in Weir [48]. LD analysis between pairs of SNP markers was calculated using a set of 10739 SNPs without any missing data in a sliding window of 900 markers using Plink v.1.9 software [45, 46]. Then, intrachromosomal LD quantification and graphical representation of LD decay were accomplished using R 3.5.0 software [40, 41]. The LD decay of the collection was measured as the genetic distance (cM) where the average r^2 decreased to half its maximum value.

3.6. Population structure and individual relatedness

To control false positive associations, population structure and individual relatedness (Kinship) among accessions of the collection were taken into account by fitting marker based structure and kinship matrices in the association models [49]. Kinship and population structure were estimated using a matrix data composed of 363 accessions and a set of 2962 markers without any missing data and corresponding to non-redundant genetic positions randomly selected on the consensus map. The coefficients of kinship between pairs of accessions were calculated by FaST-LMM software [50]. Two alternative approaches were considered to estimate the kinship matrix as described by Rincent et al. [51]. In the first one, the kinship was estimated with all the markers that are not located on the same linkage group (LG) than the tested SNP. Thus, seven kinship matrices were estimated, each being specific to a linkage group; these matrices were noted K_{LGx} with x corresponding to the number of linkage group tested. Such an approach aims at increasing power of detection of significant markers in GWAS particularly in regions of high LD. In the second approach, correlation between markers took into account all the 2962 markers and the kinship matrix was noted K.

The discriminant analysis of principal components (DAPC) method developed by Jombard et al. [52] and implemented into the Adegenet R package [53-55] was used to cluster accessions

on the basis of their genotype. This method aims at identifying and describing clusters of genetically related individuals without prior knowledges of groups. First, the optimal number of genetic clusters (k) was determined through the 'K-means' method. The number of clusters was allowed to vary from one to 20 during the determination of the optimal value of k, based on the Bayesian Information Criterion (BIC). The most likely number of clusters was chosen on the basis of the lowest associated BIC. Then, the principal component analysis (PCA) step of DAPC was performed through maximization of the 'a-score' criterion and the optimal number of principal components (PCs) was obtained after 100 iterations. Finally, DAPC was performed considering the most likely number of clusters (k) and the optimal number of PCs identified. The two first coordinates of DAPC results were used as covariates (Q matrix) in the GWAS to correct the association tests for false positives.

3.7. Association mapping

BLUPs corresponding to the phenotypic data collected for each accession were used to identify marker-trait association using Linear Mixed Model (LMM) analyses accounting for kinship matrix (K or K_{LG}) with or without population structure matrix (Q) as random effect(s). Four models were therefore compared for their capacity to fit the data: (1) a LMM using the kinship matrix K, (2) a LMM corrected for kinship using the K_{LG} matrices, (3) a LMM including the K and Q matrices and (4) a LMM using both K_{LG} and Q matrices. For each frost damages trait, the best model was chosen, according to the method used in Courtois et al. [56], by comparing the likelihoods of each model using the Bayesian Information Criterion (BIC). The model with the smallest BIC was selected. All analyses were performed using LMM provided by the FaST-LMM version algorithm [50]. The threshold to declare an association significant was set at a probability level of the p-value inferior to 4.65E-06, i.e. -log10 (p) > 5.33, which corresponded to the Bonferroni threshold (0.05/ number of tested SNPs). To represent the association results, Manhattan plots were drawn using the package QQman implemented in R software [57].

3.8. Local LD bloc estimation and favorable allele identification

Local LD blocks were defined around all the significant trait-associated markers which included markers in strong linkage disequilibrium (LD; $r^2 > 0.8$) with the targeted associated marker(s). One to eight LD blocks on the one hand and a unique associated marker, on the other hand, composed the significant loci identified. At each significant locus, haplotypes were identified, among the accessions of the collection, according to the non-imputed genotyping data

corresponding to the list of markers significantly detected by GWAS and linked markers. Haplotypes showing missing data loci as well as SNP with heterozygous genotypic data were excluded from further analysis. Besides, haplotypes represented by less than 5 % of the total number of accessions were also removed from the analysis. Based on the results of association mapping, the allelic effect corresponding to the minor allele (a_{eff}) of markers significantly associated with the frost damage traits were analyzed: if a_{eff} had a negative value, the minor allele of the associated marker was considered to decrease frost damage (favourable allele for frost tolerance); if a_{eff} had a positive value, the minor allele of the associated marker was considered to increase frost damage (unfavourable allele for frost tolerance). For each significant locus and each corresponding trait, the values of frost damages of the different haplotypes were compared using an analysis of variance with a nested design for 'haplotype/genotype', followed by a Student-Newman-Keuls (SNK) comparison test using the package 'agricolae' implemented in R [58]. Then, favourable and unfavourable haplotypes at each significant locus were defined as follows: the favourable haplotypes should show a significant lower frost damage mean score while unfavourable haplotypes should show a significant higher frost damage mean score. Finally, we listed representative accessions for each favourable and unfavourable haplotype based on the following conditions: each accession should (i) show a mean score of frost damages inferior to 1 for favourable haplotypes and superior to 4 for unfavourable ones, and (ii) be representative according to condition i) for at least two traits, in the case where the haplotype was associated with two or more traits.

3.9. Identification of the frost tolerance candidate gene

Predicted genes at significant GWAS markers and markers in LD ($r^2 > 0.8$) with significant associated markers, were identified based on the annotation data developed by Tayeh et al. [36], using BLASTx searches against known and predicted proteins of *Pisum sativum, Medicago truncatula, Glycin max and Arabidopsis thaliana*. Besides, it has been verified that the annotation data realized by Tayeh et al. has not changed since 2015.

4. Results

4.1. Statistical analyses of phenotypic data

Statistical analyses of frost damages scores showed highly significant genetic variation for all studied traits and the coefficient of genetic variation ranged between 37.6 for FD_Field_date4

and 74.4 for FD_Field_date2 (Table III.1). The estimates of broad sense heritabilities (H²) were high for all traits, varying from 0.84 to 0.89. In addition, frost damages observed in controlled conditions and in the field for the date 3 and 4 showed the highest mean scores of 2.8, 2.4 and 2.4 respectively (Table III.1). Frequency distributions of BLUPs values for each trait tended to fit normal curves within the CRB population (Additional file 2).

4.2. SNP genotyping of the association mapping collection

After quality control, the genotyping data comprised a total of 10739 polymorphic SNPs with imputed missing data and a MAF greater than 5%. Each linkage group contained 1533 SNPs on average. The distribution of SNPs varied within and between LGs (Additional files 3 and 4). Linkage group III showed the highest density of SNPs (1887 SNPs), while LG I was the least dense (1220 SNPs). Mean MAF in the association mapping collection varied from 0.29 on LG I to 0.31 on LG V and LG VI. Only 7% of SNPs had a MAF less than or equal to 0.1 (Additional file 5). The distribution of SNP markers across the different LGs was dense and no gap between adjacent SNPs exceeded 1.7 cM, except on LGI and LGV which presented respectively gaps of 2.3 cM (for the interval position 48.4-50.7 cM) and 3.7 cM (for the interval position 0.1-3.7 cM) (Additional file 3). In addition, the map of the 10739 SNPs used for GWAS showed an average number of 28 SNPs mapped at the same position.

Trait	Number of accessions	Number of observations	Min	Max	Mean	SE	Vg	CVg	H²
FD_CC	363	1087	0.00	5.00	2.80	0.03	1.25	39.93	0.89
FD_Field_date1	363	1089	0.00	5.00	1.43	0.03	0.87	65.23	0.89
FD_Field_date2	363	1048	0.00	5.00	1.39	0.03	1.07	74.42	0.89
FD_Field_date3	363	1083	0.00	5.00	2.40	0.04	1.65	53.52	0.87
FD_Field_date4	363	849	0.00	5.00	2.41	0.03	0.82	37.57	0.84

Table III. 1 : Statistical parameters of the pea collection for the five observed traits

Traits are abbreviated as follows: FD_CC: Frost damages in the controlled conditions experiment, FD_Field_date1, FD_Field_date2, FD_Field_date3 and FD_Field_date4: Frost damages in the field experiment at the first, second, third and fourth date of observation, respectively. Minimum (Min), maximum (Max) and mean values, standard error (SE), genetic variance (Vg), coefficient of variation of the genetic variance (CVg) expressed as $\sqrt{(Vg)/mean}$ and broad-sense heritability (H²) are shown for each trait.

4.3. LD analysis

The distribution of the estimate of LD (r^2) according to the genetic distance for each linkage group as well as for the whole genome is presented in Additional file 6. The r^2 value rapidly decreased as the genetic distance increased. The LD decay, estimated as the distance for which r^2 decreases to half of its maximum level (0.22), was equal to 0.9 cM for the whole genome. Considering the LGs individually, the LD decay ranged from 0.3 cM for LG IV to 1.4 cM for LG V.

4.4. Population structure and kinship analyses

To avoid false positive results in association analysis, structure and kinship of the association mapping collection were analyzed using 2967 non-redundant positions markers. Following the analysis of the BIC profile and using the 'a-score' criterion, the optimal number of clusters was fixed to 7 (Additional file 7, A) and the optimal number of principal components (PCs) was set to 6 (Additional file 7, B). Therefore, these 6 PCs and 7 clusters were used for discriminant analysis of principal components. The distribution of individuals into the 7 clusters is represented along the two first axes of DAPC on Figure S5-C (Additional file 7). The main passport data (Additional file 1) seemed to be related to the discrimination of the clusters. The cluster 2, comprising mainly wild peas and local populations from Africa and Middle or Far-East, was totally separated from the other clusters. The majority of accessions from clusters 1, 3, 4 and 6 were registered as spring sowing types while the winter sowing types were essentially gathered in clusters 5 and 7. These last two clusters differed for their end-use, the cluster 5 being mainly composed by field peas (72 %) and the cluster 7 by fodder peas (82 %).

Within the kinship matrix (K), estimated for the whole genome, more than half of the kinship coefficient values were close to zero and around 86 % of them were less than 0.1. Less than 2 % of these values were larger than 0.5. For the seven kinship matrices specific to each linkage group (K_{LG}), the kinship coefficient values were ranging similarly than for the K matrix (Additional file 8). These results indicated a weak relatedness between accessions and suggested that the majority of the accessions are genetically diverse, which was beneficial for subsequent GWAS mapping. From these results, the two first coordinates of DAPC results (Q matrix) and relatedness matrices (K matrix and K_{LG} matrix) were used as covariates for subsequent association analyses.

4.5. Genome wide association mapping

The comparison of BIC values of the four GWA-models tested, showed that the linear mixed model which included both Q and K_{LG} matrices as covariates was the optimal model for the following traits: FD_CC, FD_field_date1, FD_field_date2 and FD_field_date3. Whereas the best fitting model for FD_field_date4 was the linear mixed model which included only K_{LG} matrices (Table III.2). The Manhattan plots of the association mapping results, run with the best model for each trait, are presented in Figure III.1. A total of 62 markers were significantly associated with any of the studied traits at the Bonferroni threshold -log10 (p) > 5.33. Frost Damage (FD)-associated markers were distributed on all linkage groups except LG IV. These SNPs exhibited a minimum allele frequency (MAF) ranging from 0.13 to 0.5. The number of markers associated with FD_CC, FD_field_date1, FD_field_date2, FD_field_date3 and FD_field_date4 were 40, 4, 6, 3 and 17 respectively. The highest p value (9.67E-11) was showed by the significant locus on LG V (Table III.3).

Trait	LMM1]	LMM2	Ι	LMM3	LMM4		
	-2ln(L)	BIC	-2ln(L)	BIC	-2ln(L)	BIC	-2ln(L)	BIC	
FD_CC	704.71	4073.94	701.16	4070.4	695.27	4064.51	691.68	4060.92	
FD_Field_date1	658.87	4028.11	658.4	4027.63	675.69	4026.92	656.14	4025.38	
FD_Field_date2	735.42	4076.81	732.06	4073.45	731.26	4072.65	728.85	4070.24	
FD_Field_date3	837.83	4207.06	833.87	4203.1	836.8	4206.04	831.51	4200.74	
FD_Field_date4	482.59	3489.84	481.45	3488.7	483.23	3490.48	482.53	3489.78	

Table III. 2 : Bic-based comparison of the four models used to control the rate of false positive associations.

LMM1: linear mixed model including the K kinship matrix; LMM2: linear mixed model including the K_{LG} kinship matrix; LMM3: linear mixed model including the K kinship matrix and the population structure matrix (Q); LMM4: linear mixed model including the K_{LG} kinship matrix and the population structure matrix (Q). The model with the lowest BIC value (bold) is considered to be the best choice for the target trait.



Figure III. 1 : Manhattan plots of markers associated with the five frost damage traits.

The plots show the p-values (p) for association between a phenotypical trait and each tested marker (expressed as the negative decimal logarithm of p, y-axis) plotted against the linkage group position of the marker (x-axis). Plots above the red horizontal line indicate the genome wide significance with the Bonferroni threshold ($-\log 10$ (p) > 5.33). A, B, C and D are the plots for the four evaluations of FD in the field experiment, corresponding to the 4 dates of damages observation. E is the plot for the evaluation of FD in the controlled conditions experiment.

Table III. 3 : Significant associations detected in the association mapping collection (363 accessions) for the five observed traits

Traits are coded as described in Table III.1. Significance threshold of the marker-trait association is $-\log_{10}(p) > 5.33$ as described in the Material and Methods section. For each associated SNP, the allelic effect of the minor allele as well as the favourable allele are listed. MAF*: Minor allele frequency; Allelic effect**: Allele effect of the minor allele. SE***: standard error of the allelic effect. Alleles of four cultivars which are the progenitors of bi-parental mapping populations for frost tolerance are listed to illustrate contrasted genotypes for this trait: Champagne and China (frost tolerant), Térèse and Caméor (frost sensitive).

LG	Associated marker	Genetic	Associated	log ₁₀	Minor	Allelic	SE	Favourable	Allele of referen	Allele of reference accessions		
		position (cM)	trait	(<i>p</i>)	allele (MAF)*	effect **	***	allele	Champagne	Térèse	China	Caméor
Ι	PsCam020765_11567_1428	47.5	FD_Field_date1	5.57	B (0.13)	0.18	0.04	А	AA	BB	AA	AA
II	PsCam043698_27686_2011	20.5	FD_CC	5.45	B (0.39)	0.18	0.04	А	AA	BB	AA	AA
III	PsCam049061_31705_537	131.1	FD_CC	5.55	B (0.50)	-0.29	0.06	В	AA	BB	AA	BB
III	PsCam035617_20792_637	132	FD_CC	6.63	A (0.50)	-0.37	0.07	А	BB	AA	BB	AA
			FD_Field_date1	5.47	A (0.50)	-0.31	0.07	А				
III	PsCam037922_22979_691	132.1	FD_CC	6.63	A (0.50)	-0.37	0.07	А	BB	AA	BB	AA
			FD_Field_date1	5.47	A (0.50)	-0.31	0.07	А				
V	PsCam051352_33858_733	53.6	FD_Field_date4	7.29	A (0.25)	-0.27	0.05	А	AA	BB	BB	AA
V	PsCam048068_30823_2326	56.9	FD_Field_date4	6.21	B (0.33)	-0.26	0.05	В	BB	AA	BB	AA
V	PsCam002670_2123_708	58.6	FD_Field_date4	6.99	A (0.19)	-0.23	0.04	А	AA	BB	BB	BB
V	PsCam051635_34107_368	58.6	FD_Field_date4	9.21	A (0.27)	-0.33	0.05	А	AA	BB	BB	BB
V	PsCam011361_7735_182	60.2	FD_Field_date4	5.88	A (0.26)	-0.25	0.05	А	AA	BB	BB	BB
V	PsCam012913_8715_718	60.2	FD_Field_date4	5.43	A (0.32)	-0.22	0.05	А	AA	BB	AA	BB
V	PsCam049838_32453_1430	60.2	FD_Field_date4	7.17	B (0.28)	-0.26	0.05	В	BB	AA	BB	AA
V	PsCam042222_26315_1205	61	FD_Field_date4	10.0	A (0.24)	-0.38	0.06	А	AA	BB	BB	BB
VI	PsCam021402_12023_752	41.4	FD_Field_date4	5.84	A (0.21)	-0.23	0.05	А	AA	BB	BB	BB
VI	PsCam043553_27552_265	43	FD_Field_date4	5.95	A (0.22)	-0.25	0.05	А	AA	BB	BB	BB
VI	PsCam022275_12527_210	47.4	FD_CC	6.21	B (0.23)	-0.33	0.06	В	BB	AA	AA	AA
		47.4	FD_Field_date1	5.64	B (0.23)	-0.29	0.06	В				
		47.4	FD_Field_date4	6.86	B (0.23)	-0.27	0.05	В				
VI	PsCam003541_2731_574	48	FD_CC	6.51	A (0.44)	0.21	0.04	В	BB	BB	AA	BB
VI	PsCam050834_33394_133	48.5	FD_CC	5.35	A (0.33)	-0.25	0.05	Α	AA	BB	AA	BB
VI	PsCam001662_1375_798	49.1	FD_CC	7.38	A (0.31)	-0.35	0.06	Α	AA	BB	AA	BB
VI	PsCam043847_27818_181	49.1	FD_CC	5.90	A (0.27)	-0.29	0.06	Α	AA	BB	AA	BB
VI	PsCam045419_29091_815	49.1	FD CC	7.57	B (0.37)	-0.24	0.04	В	BB	AA	BB	AA

VI	PsCam048404_31138_2325	49.1	FD_CC	6.02	B (0.31)	-0.30	0.06	В	BB	AA	BB	AA
VI	PsCam001662_1375_798	49.1	FD_Field_date4	6.08	A (0.31)	-0.24	0.05	А	AA	BB	AA	BB
VI	PsCam004890_3696_1580	49.1	FD_Field_date4	5.82	A (0.22)	-0.24	0.05	А	AA	BB	BB	BB
VI	PsCam023246_13111_1125	49.1	FD_Field_date4	5.40	B (0.21)	-0.25	0.05	В	BB	AA	AA	AA
VI	PsCam037082_22189_1302	49.1	FD_Field_date4	5.40	B (0.22)	-0.23	0.05	В	BB	AA	AA	AA
VI	PsCam053556_35439_228	49.1	FD_Field_date4	5.39	B (0.31)	-0.21	0.04	В	BB	AA	BB	AA
VI	PsCam007060_5248_2156	50.1	FD_CC	7.66	B (0.40)	0.22	0.04	А	AA	AA	AA	BB
VI	PsCam026381_15234_450	50.1	FD_CC	6.33	B (0.47)	0.21	0.04	А	AA	AA	BB	BB
VI	PsCam037879_22937_948	50.1	FD_CC	9.90	A (0.43)	0.26	0.04	В	BB	BB	BB	AA
VI	PsCam001302_1089_1249	50.2	FD_CC	6.50	B (0.46)	-0.22	0.04	В	BB	BB	AA	AA
VI	PsCam003337_2589_220	50.2	FD_CC	7.34	A (0.39)	0.22	0.04	В	BB	BB	BB	AA
VI	PsCam037603_22679_2024	50.2	FD_CC	7.59	A (0.38)	0.22	0.04	В	BB	BB	BB	AA
VI	PsCam040539_25267_719	50.2	FD_CC	7.59	A (0.38)	0.22	0.04	В	BB	BB	BB	AA
VI	PsCam035398_20586_1094	50.6	FD_CC	7.63	B (0.50)	-0.25	0.04	В	BB	AA	BB	AA
VI	PsCam039054_24037_586	50.6	FD_CC	6.30	B (0.41)	0.21	0.04	А	AA	AA	AA	BB
VI	PsCam049925_32537_2703	50.7	FD_CC	6.57	B (0.41)	0.22	0.04	А	AA	AA	AA	BB
VI	PsCam038477_23510_469	50.8	FD_CC	7.67	B (0.39)	-0.27	0.05	В	BB	AA	BB	AA
VI	PsCam001148_968_692	51.1	FD_CC	6.06	B (0.50)	0.22	0.04	А	AA	AA	BB	AA
VI	PsCam006586_4888_245	51.1	FD_CC	6.00	A (0.41)	0.20	0.04	В	BB	BB	AA	BB
VI	PsCam017327_10745_93	51.1	FD_CC	8.40	B (0.48)	-0.27	0.04	В	BB	BB	BB	AA
VI	PsCam020872_11644_1756	51.1	FD_CC	7.31	B (0.45)	0.23	0.04	А	AA	AA	BB	AA
VI	PsCam025448_14564_2314	51.1	FD_CC	6.08	A (0.29)	-0.22	0.04	А	AA	BB	BB	AA
VI	PsCam033562_19160_307	51.1	FD_CC	6.04	A (0.48)	0.21	0.04	В	BB	BB	AA	BB
VI	PsCam036704_21832_970	51.1	FD_CC	7.12	B (0.48)	0.22	0.04	А	AA	AA	BB	AA
VI	PsCam038677_23691_325	51.1	FD_CC	6.59	B (0.43)	0.20	0.04	А	AA	AA	BB	AA
VI	PsCam044418_28295_579	51.1	FD_CC	6.61	A (0.41)	0.21	0.04	В	BB	BB	AA	BB
VI	PsCam052151_34533_410	51.1	FD_CC	6.16	A (0.47)	0.21	0.04	В	BB	BB	AA	BB
VI	PsCam053189_35238_569	51.1	FD_CC	7.30	B (0.49)	0.23	0.04	А	AA	AA	BB	AA
VI	PsCam042568_26621_42	51.1	FD_Field_date4	5.43	A (0.21)	-0.25	0.05	А	AA	BB	BB	BB
VI	PsCam035356_20546_778	52.5	FD_CC	5.99	B (0.43)	0.20	0.04	А	AA	AA	BB	AA
VI	PsCam040326_25099_1923	52.5	FD_CC	5.91	A (0.38)	0.20	0.04	В	BB	BB	AA	BB
VI	PsCam011774_8038_200	52.8	FD_CC	6.76	A (0.49)	-0.23	0.04	А	AA	BB	BB	AA
VI	PsCam023852_13541_264	52.8	FD_CC	6.71	B (0.46)	-0.24	0.04	В	BB	AA	AA	BB
VI	PsCam026055_15000_192	52.8	FD_CC	5.70	B (0.44)	-0.22	0.04	В	BB	AA	AA	BB
VII	PsCam001108_940_48	72.9	FD_Field_date2	5.74	B (0.48)	0.20	0.04	A	BB	AA	AA	AA
		72.9	FD_Field_date3	5.34	B (0.48)	0.21	0.05	A				

Chapitre III : Dissection du déterminisme génétique de la tolérance au gel chez le pois par génétique d'association

VII	PsCam050827_33387_242	82.4	FD_Field_date2	6.27	B (0.21)	0.21	0.04	А	AA	AA	AA	AA
VII	PsCam037927_22984_97	82.5	FD_Field_date2	5.40	B (0.41)	0.19	0.04	Α	BB	AA	AA	BB
VII	PsCam038378_23415_721	86.4	FD_CC	6.18	A (0.46)	0.22	0.04	В	BB	BB	BB	AA
VII	PsCam042427_26489_524	87.1	FD_CC	5.67	B (0.41)	0.20	0.04	Α	AA	AA	AA	BB
VII	PsCam000301_265_1519	87.9	FD_Field_date2	7.41	B (0.19)	0.23	0.04	Α	AA	AA	AA	BB
		87.9	FD_Field_date3	5.72	B (0.19)	0.22	0.05	Α				
VII	PsCam045010_28743_3365	88	FD_Field_date2	5.73	B (0.21)	0.20	0.04	Α	AA	AA	AA	BB
VII	PsCam004928_3732_3087	89.3	FD_Field_date2	7.46	B (0.28)	0.25	0.04	Α	AA	AA	AA	AA
		89.3	FD_Field_date3	5.40	B (0.28)	0.24	0.05	Α				

4.6. Favourable alleles exploration for frost tolerance in pea

Twelve LD blocks were defined around the 62 significant FD-associated markers which included all markers in LD ($r^2 > 0.8$) with the FD-associated markers (Additional file 9). FDassociated markers which were not in significant LD with any other SNP, and thus did not constitute a LD block, were also kept for further analysis. Finally, 73 SNPs, covering six socalled FD-related loci distributed on LG I, II, III, V, VI and VII, were kept to identify favourable and unfavourable haplotypes for frost tolerance. Then, for each of the six FD-related loci, markers haplotypes (two to nine) and corresponding representative accessions were identified (Additional file 10 and 11). For each locus, the effect of the different allele combinations was tested thanks to a variance analysis and a multiple comparison test of phenotypic mean effects (Additional file 10). These analyses allowed to identify 8 favourable haplotypes over the 6 FDrelated regions carrying favourable alleles at each FD-associated marker except the haplotypes VI.5 and VII.4 which contained each 1 unfavourable allele, among 31 and 3 FD-associated markers respectively (Additional file 10 and 11). Accessions carrying favourable haplotypes presented lower values of frost damages ranging between 0.00 ± 0.00 (haplotype VII.4 for the trait FD Field date2) and 2.65 ± 0.05 (haplotype III.1 for the trait FD CC). Six groups of accessions carrying unfavourable haplotypes were also identified for which 100 % of unfavourable alleles were observed over the significant FD-associated markers detected by GWAS. Typical accessions carrying favourable haplotypes and showing a mean score of frost damage ≤ 1 were mainly winter fodder peas (e.g. Black seeded, Champagne, Melrose, Blixt104) while whose carrying unfavourable haplotypes with a mean score of frost damage ≥ 4 were mainly spring garden peas (e.g. Automobile, Caroubel, Cennia, Ersling) (Additional file 11).

4.7. Candidate gene analysis

For candidate gene analysis, we considered SNPs significantly associated with the five traits that were detected by GWAS and markers in high LD ($r^2 > 0.8$) with associated marker(s), which represented a total of 73 SNPs. Fifty of these SNPs corresponded to genes with putative protein functions and one encoded an expressed protein with unknown function. The annotation process gave no result for 22 SNPs (Additional file 12).

5. Discussion

5.1. GWAS brings new insights into the determinism of frost tolerance in pea

In the present study, 73 markers associated with frost tolerance, i.e. 62 markers significantly detected by GWAS and 11 markers in high LD ($r^2>0.8$) with one or the other of the 62 markers, were located among all the linkage groups of pea, except LG IV (Table III.3).

Comparing the map positions with those of frost tolerance QTLs previously described, 3 regions corresponding to 63 among the 73 markers, were found to co-localize with 3 main QTLs previously detected by linkage mapping in two bi-parental populations for the same trait, namely WFD 3.2, WFD 5.1 and WFD 6.1 [21-23] (Figure III.2). These 3 QTLs were repeatedly detected in 10 to 11 field conditions. Moreover, the position corresponding to WFD 6.1 seems to also match with an EST marker recently found to be associated with frost tolerance in pea [38] : indeed, Liu et al. identified 7 marker-trait associations within a collection of 672 accessions, among which the marker EST1109 was located on LG VI within a functional gene that has a high homology with a gene encoding an alpha-mannosidase in *M. truncatula*; we checked the 1646 transcript-derived SNP markers mapped on LG VI by Tayeh et al. [36] and found that a marker corresponding to an alpha-mannosidase-like protein was located at 43.2 cM on the consensus map; consequently, as the FD-associated markers detected on LG VI in the present study are comprised between 41.4 and 52.8 cM, it is likely that the LG VI positions of both association studies coincide. Similarly, this study validated the QTL previously detected on LG III (WFD 3.2 in RIL2 [21], III.2 in RIL9 [23]) with a higher resolution than previous linkage mapping studies by the identifications of three main significant FD-associated markers, and whose positive alleles decreased the frost damage of accessions by an average of 0.33 (Table III.3). All favourable alleles were carried by the sensitive genotypes Térèse and Caméor unlike the tolerant genotypes Champagne and China which had all undesirable alleles underlying this locus (Table III.3). Altogether, the consistency of the 3 positions detected in both bi-parental mapping and association genetics reinforces their interest for breeding. The results presented here constitute an additional step towards the identification of underlying genes potentially involved in the control of frost tolerance, thanks to refined intervals provided by GWAS. We will discuss the candidate genes suggested at the LG VI position in the next section.

Chapitre III : Dissection du déterminisme génétique de la tolérance au gel chez le pois par génétique d'association





Only linkage groups (LGs) with significant frost damages (FD)-associated markers detected by GWAS are presented. On each LG, SNP markers are shown on the right and genetic distances between markers are indicated in cM on the left. Frost tolerance loci detected in the present study are shown in red: FD-associated markers are in a red and underlined font; markers in linkage disequilibrium (LD; $r^{2}>0.8$) with associated marker(s) are in a red and non-underlined font; the LD blocks identified by GWAS are drawn as red bars on the right of each LG. QTLs represented by blue and green bars were detected in the Champagne x Térèse [21, 22] and China x Caméor [23] populations respectively. For presentation purposes, only markers at the vicinity of significant loci and a few markers distributed along LGs are shown.



LG VII



Figure III. 2: Comparative genetic map of genome wide association study (GWAS) loci identified in the present study and quantitative trait loci (QTL) previously detected for frost tolerance in pea (*Continued*).

Unlike the correspondences with bi-parental mapping positions presented above, the present GWA study did not highlight any co-localization with a major QTL, namely WFD3.1, which was however found to be responsible for up to 52 and 19 % of the winter frost damage variation within the RILs populations derived from Champagne x Térèse (RIL2) and China x Caméor (RIL9), respectively [21, 23]. The flowering gene *Hr* (*High response to photoperiod*), an orthologue of the Arabidopsis *Early flowering 3*, *Elf3* [59], was shown to be a relevant

candidate for this QTL as it allows plants to be maintained in a vegetative state under short days and thus to escape the main winter freezing periods. It is likely that, in the case of the WFD3.1 position corresponding to the Hr gene, a strong correlation may have emerged between the population structure, probably biased by the allelic variation at the Hr locus, on the one hand, and the frost damage trait, on the other hand. This hypothesis relies on the observation that Hr may have been the target of natural selection for frost tolerance. Weller et al. [59] speculated that the hr mutation may have arisen within an ancestral pea lineage originating from the Near East domestication center and bearing itself the Hr allele. The hr mutation probably permitted summer cropping in areas characterized by colder winters and is therefore highly represented in many domesticated lines of *Pisum sativum* at the origin of the current spring peas. To explore the hypothesis of undetection of a true association due to confounding with the population structure, we have verified the distribution of the Hr alleles, represented by their Elf3 genotype, within the DAPC clusters of the association mapping population (Additional file 13). The Hr accessions, homozygous for the dominant allele Hr, are the main components of the clusters 2 (84%) and 7 (96%) and they represent 58% of the cluster 1. They are thus over represented in the three clusters gathering most of the winter sowing-type accessions, which may have contributed to a correlation between the frost tolerance trait and the population structure. At the opposite, the hr accessions, homozygous for the recessive allele hr, are the main components of the clusters 3 (99 %), 4 (100 %), 5 (100 %) and 6 (96 %), which are mostly spring sowingtype accessions. Consequently, we can suggest that the correction for the population structure (Q matrix) might have result in a structuring marker that probably cannot be detected by further association analysis using this sample of accessions. This kind of result has already been reported by Visioni et al. [27] who found that 2 of the3 most significant SNPs of their study, tightly linked to major known genetic determinants of cold tolerance in barley, were undetected by GWAS if a correction by the structure was used. It was particularly the case for a SNP linked to Vrn-H1, a developmental locus governing barley vernalization requirement, which is for long a candidate for the frost tolerance locus Fr-H1 but whose effect was suspected to be confounded with the population structure. To overcome this point, NAM-like linkage populations with biparental crosses in a reference design could be an interesting plant material for association mapping, in order to minimize population structure which may be necessary for dissecting the most structured traits.

Comparatively with consistent QTLs previously detected in bi-parental populations, the present GWA study also pinpointed three loci which have either not been detected yet (one region on

LG II) or formerly detected in only one environment (two regions located on LG I and LG VII, respectively). The significant marker on LG II is supported only by one experiment in controlled conditions in our study, but allowed to identify two clear favourable and unfavourable haplotypes (Additional file 10). The two markers significantly associated with frost tolerance on LG I are located at 47.5 cM on the consensus map, which lies in the projected confidence interval of a previous QTL, namely WFDcle.a, identified in one field condition with the Hr subpopulation extracted from RIL2 [22]. This colocalization slightly reinforces the consistency of this LG I position. In the same way, the 8 significant markers identified on LG VII with this study were located between 72.9 and 89.3 cM on the consensus map, which overlap with one former QTL, namely FD.c, which was detected once in a controlled chamber experiment [22], with the same Hr subpopulation. Thus, the colocalizing region on LG VII relies now on three independent experiments. With their 672-accessions collection, Liu et al. [38] identified one and two markers significantly associated with frost tolerance on LG I and LG VII, respectively. It would be interesting to check the localization of these markers on the consensus map used here, as their position on the genetic map used by the authors [60] is not reported. Additionally, to these coincident positions for the frost tolerance trait on LGI and LG VII within different experimental conditions, the LG I and the LG VII regions detected in this study seems to also overlap with already detected QTLs for resistance to Aphanomyces euteiches. Indeed, both markers of the FD-related LD block I.1 lie in the confidence interval of the QTL Ae-Ps1.1 identified by Hamon et al. [61], when the latter is projected on the consensus map used in the present work. Besides, the FD-related locus on LG VII includes the LD block VII.16 associated to the resistance to Aphanomyces euteiches [35], with which it shares the FDassociated marker PsCam038378 23415 721. The FD-related locus on LG VII deserves a particular attention for a further use in breeding for frost tolerance because accessions carrying the favourable haplotype underlying this locus (haplotype VII.4) shows very low scores of frost damages, ranging from 0 to 0.96, both in the field and controlled conditions experiments (Additional file 10). The relationships between haplotypes at this locus and the values for frost tolerance and Aphanomyces euteiches resistance will however have to be more precisely explored, to check if both traits can be bred favourably at this locus.

5.2. GWAS detected frost tolerance-associated markers which are relevant putative candidate genes
Fifty of the 73 FD-related SNPs detected in this study corresponded to genes with putative protein functions (Additional file 12). Among the diverse protein functions predicted, some are already related to the acquisition of frost tolerance in the literature.

Comparison of map positions has shown that the FD-related locus detected on LG VI in this study colocalizes with the previous QTL WFD 6.1, which is itself orthologous to a major QTL for frost tolerance in *M. truncatula* (Mt-FTQTL6) [62]. Tayeh et al moreover showed that Mt-FTQTL6 covers a region containing a cluster of twelve *CBF* genes tandemly duplicated [19]. In the present study, the marker PsCam037030 22140 221 on LG VI (49.1 cM), which is in high LD with all the significant FD-associated markers belonging to the LD block VI.2, corresponds to the CBF14 gene in M. truncatula (Additional file 12). Besides that, three other close (PsCam050192 32788 145, **SNPs** PsCam039597 24507 724 and PsCam025621 14701 176), located in the interval situated between 50.2 and 50.4 cM on the consensus genetic map, have also been annotated by Tayeh et al. as CBF8, CBF14 and CBF7, following a blast search against *M. truncatula* protein sequences [36]. The first two of these markers were excluded from the GWA analysis for the following reasons: (i) PsCam050192 32788 145 showed a minor allele frequency lower than 0.05; (ii) PsCam039597 24507 724 had non-exploitable results of genotyping within the collection marker, (Grégoire Aubert, personal communication). The third namely PsCam025621 14701 176 was not found to be significantly associated with the frost damage traits. Given these results and the previous findings of Tayeh et al. concerning Mt-FTQTL6 [19], CBF genes located in the LD-block VI.2, or at its vicinity, are also relevant candidates determining frost tolerance at this locus in pea. The potential role of *CBF* genes, and particularly *CBF14*, has already been highlighted in cereals. In wheat, Zhu et al. [63] showed that the natural variation for frost tolerance is mainly associated with a frost resistance 2 (FR2) locus including tandemly replicated CBF genes that regulates the expression of cold-regulated genes. Additionally, these authors proved that an increased copy number of CBF14 was frequently associated with the tolerant haplotype of the locus FR-A2 and with higher CBF14 transcript levels in response to cold. Novák et al. [64] showed that CBF14 genes contribute to enhance frost tolerance during cold acclimation in cereals.

Three FD-associated markers detected in the present study at the LG III, LGV and LGVI positions also appear in the literature to be implied in the crosstalk between plant hormone signaling in the cold stress response and the *CBF* regulon. These three markers, namely

PsCam035617_20792_637, PsCam048068_30823_2326 and PsCam011774_8038_200 correspond, in *Pisum sativum*, to a functional gene encoding a brassinosteroid receptor (Additional file 12). In Arabidopsis, Eremina et al. [65] provided evidence that brassinosteroids contribute to the control freezing tolerance. Indeed, these authors showed that brassinosteroid-deficient mutants of Arabidopsis were hypersensitive to freezing stress, whereas an activation of the brassinosteroid signaling pathway increased freezing tolerance both before and after cold acclimation. Furthermore, two brassinosteroids-responsive transcription factors have also been characterized as direct regulators of *CBF* expression through their binding to the promoters of these genes [65, 66]. In cultivated plants, the role of brassinosteroids is so far documented for the response to the chilling stress, as reviewed by Anwar et al. [67].

The SNP PsCam037922 22979 691, which 0.1 is only cM from apart PsCam035617 20792 637 just mentioned above, also deserves attention. This marker is one of the three associated markers underlying LD block III.1, which co-localizes with WFD 3.2. It has been shown to correspond to Le in Pisum sativum, a gene encoding for the gibberellin 3beta-hydroxylase enzyme which produces bioactive gibberellin (Additional file 12). Recessive le mutants at this locus are impaired in the production of gibberellin and produce a dwarf phenotype [68, 69]. In Arabidopsis, Achard et al. [70] undertook a molecular and genetic approach to evaluate the interaction between the *CBF1*-dependent cold acclimation pathway and the gibberellin pathway. They proposed a model in which the induction of CBF1 expression by low temperature affects the gibberellin metabolism via upregulation of gibberellin-2oxydase gene transcripts. The following reduction in bioactive gibberellin causes a higher accumulation of DELLAs, a family of nuclear growth-repressing proteins which in turns restrains plant growth. The Le gene has already been proposed as a candidate for the WFD3.2 and III.3 QTLs identified in the RIL2 and RIL9 populations, respectively. We considered the haplotypes of the parents of both populations at the corresponding FD-associated locus in this study and found that the favourable haplotype (III.1) was beared by Térèse and Caméor while the unfavourable haplotype (III.2) was beared by Champagne and China. This observation is consistent with the favourable and unfavourable alleles determined by QTL mapping in biparental populations. As the three genes constituting the FD-associated locus lie within a 1 cM interval, neither a synergistic effect nor linkage can be excluded.

Finally, three markers underlying the FD-related locus on LGVII were located within genes related to the synthesis of soluble sugars (PsCam001108_940_48 : endo-beta-1,3-glucanase,

PsCam037927_22984_97 : endo-1,4-beta-glucanase and PsCam004928_3732_3087 : betaglucosidase G2) (Additional file 12). As accumulation of soluble sugars during cold acclimation is well documented in many plants, we can suggest that the LG VII-locus may have a role in frost tolerance by accumulating sugars in plant tissues during cold acclimation. Several roles for sugars in protecting cells from freezing injury have been proposed, including functioning as cryoprotectants for specific enzymes, as molecules promoting membrane stability and as osmolytes to prevent excessive dehydration during freezing, as reviewed by Xin and Browse [71].

5.3. GWAS provides new markers and new genitors to breed for frost tolerance in pea

In the present study, 6 loci related to frost tolerance in pea were identified. At these 6 FD-related loci, 8 favourable haplotypes, carrying the highest number of favourable alleles at the FDassociated markers detected by GWAS, were significantly associated with the lowest scores of frost damages, i.e. the highest levels of frost tolerance. We identified 13 accessions cumulating 3 (Champagne), 4 (Winterberger, Glacier, Côte d'Or and Picar) or 5 (Booster*aw7, Melrose, Blixt 7, Holly 9, Black seeded, Holly 17, Blixt 109 and Cheyenne) favourable haplotypes at the 6 FD-related loci, as shown in Additional file 14. Twelve among these 13 accessions belong to the cluster 7 which was shown to be totally isolated from the other clusters by the DAPC analysis. They are mainly fodder peas which can be considered as relatively close to wild peas also present in the cluster 7 and among which frost tolerant accessions have already been identified for breeding. The same 12 accessions also carry the Hr allele which was formerly shown to be favourable to frost tolerance [21]. Cheyenne, a French variety released in 1998 is the only accession carrying the hr allele and its mean frost damage record is slightly superior to the record of the other accessions in the present study. Excepted Cheyenne, the 12 accessions carrying 3 to 5 five favourable haplotypes showed a mean frost damage score ranging from 0.2 to 0.53 and no information allows to decide that a given locus plays a minor part than another. One common point of 11 of the 12 accessions is that they carry the unfavourable haplotype at the FD-related locus on LGIII comprising the Le gene. But rather than indicating a minor effect of the favourable haplotype at this locus, genetically linked to the dwarf *le* allele, this feature is to relate to the observation that field-autumn-sown Hr lines remain dwarf until a longer spring daylength has also triggered off the switch from the prostrate to the erected growth habit, which suggests an epistatic effect of Hr upon the expression of the dwarfism [21]. In winter field breeding material (like Booster*aw7) or cultivars (like Cheyenne), the le allele which has been

bred to reduce lodging also brought a favourable allele for frost tolerance. Comparatively to the above-described material, we also identified 4 accessions (Ersling, Automobile, Cennia and Caroubel) carrying all the unfavourable alleles at the 6 FD-related loci of the present study, additionally to the unfavourable allele *hr*. These 4 spring cultivars presented a mean frost damage score ranging from 3.14 and 4.63. Our results can help choosing tolerant progenitors and following favourable haplotypes through marker-assisted breeding.

Furthermore, the FD-associated locus on LGV was found to overlap with the confidence interval of the frost tolerance QTL WFD 5.1 earlier detected within the RIL2 population [21] (Figure III.2). Comparatively to the linkage mapping method, with which a confidence interval of 16.6 cM was obtained, the GWA study enabled to refine the confidence interval to 7.4 cM (53.6 to 61 cM on the consensus map). This refined LG V locus presents a particular interest in breeding as it may provide markers to break the genetic relationship between the frost tolerance position and the neighbouring locus governing the seed trypsin activity. Trypsin inhibitors are known to be unfavorable for animal feed because they decrease the digestibility of protein [72]. The locus responsible for the seed trypsin activity (Tri) has been mapped on LG V [73] at the vicinity of WFD 5.1 [21]. The favourable alleles at both loci are generally in repulsion. On the consensus map, this locus is represented by three markers, annotated as *trypsin inhibitor* genes [36] and located between 67.0 and 67.3 cM, 6 cM apart from the frost tolerance locus detected by GWAS. Thus, it seems possible to select favourable alleles for frost tolerance corresponding to the FD-associated markers detected by GWAS on LG V together with recessive alleles of the markers encoding for the trypsin inhibitor genes.

6. Conclusion

In the present study, GWAS enabled to confirm QTLs significantly associated with frost tolerance such as WFD 3.2, WFD 5.1 and WFD 6.1. It also allowed to identify one region on LG II, which has not been detected yet and provided significant associations for two regions on LGI and LG VII that were formerly detected in only one environment. The results showed that GWAS is an effective strategy to identify markers precisely defining frost tolerance loci, which can be useful to breed for antagonistic traits as it is the case for the frost tolerance and Tri loci on LG V which are in linkage disequilibrium and in a repulsion phase. Our results also highlight that GWAS enables to find new sources of frost tolerance within collections of pea genetic resources. Finally, the present GWA study also brought to light the presence of *CBF*

transcriptions factors as potential genetic determinants of the frost tolerance locus on LG VI, with one *CBF*-annotated marker being in high LD with significant FD-associated markers of the locus and three additional *CBF*-annotated markers mapped at the vicinity. As 12 tandemly duplicated *CBF* genes were already found to be relevant candidates underlying the orthologous frost tolerance QTL on *Medicago truncatula* chromosome 6, the hypothesis of an equivalent genomic organization in pea deserves to be tested thanks particularly to the upcoming knowledge of the physical map for this species.

7. Availability of supporting data

The additional files supporting the results of this article are presented in Annexe A at the end of the manuscript.

8. Abbreviations

AP2/EREBP: APETALA2/Ethylene Responsive Element Binding Protein

BLUPs: Best Linear Unbiased Predictions

CBF: C-repeat Binding Factor

cM: centi Morgan

COR: COld Regulated

CVg: Coefficient of Variation of the genetic variance

DAPC: Discriminant Analysis of Principal Components

Elf3: Early flowering 3

FD: Frost Damage

FD_CC: Frost damages in the controlled conditions experiment

FD_Field_date1: Frost damages in the field experiment at the first date of scoring

FD_Field_date2: Frost damages in the field experiment at the second date of scoring

FD_Field_date3: Frost damages in the field experiment at the third date of scoring

FD_Field_date4: Frost damages in the field experiment at the fourth date of scoring

GWAS: Genome Wide Association Study

H²: broad sense heritability

Hr: High response to photoperiod

INRA : Institut National de la Recherche Agronomique

K: Kinship matrix

KLG: Kinship matrix corresponding to a given Linkage Group

LD: Linkage Disequilibrium LG: Linkage Group LMM: Linear Mixed Model MAF: Minor Allele Frequency NAM: Nested Association Mapping NIL: Near Isogenic Line PCs: Principal Components PCA : Principal Components Analysis Q: Structure matrix QTL : Quantitative Trait Locus **RIL** : Recombinant Inbred Line SE: Standard Error SNK: Student-Newman-Keuls SNP: Single Nucleotide Polymorphism SSR: Simple Sequence Repeat Vg: Genetic variance WFD: Winter Frost Damage

9. Competing interests

The authors declare that they have no competing interests.

10. Authors' contributions

VF, RD and MT collected phenotypic data in the field and controlled conditions experiments. GA coordinated the genotyping of the association mapping collection and provided genotyping data. SB performed the phenotypic and genetic analyses, as well as association mapping and haplotypes analysis and wrote the manuscript. MS and SN assisted in analysing genetic data by testing GWAS models and providing R scripts, respectively. ILH and BD conceived and coordinated the study. ILH managed the technical and financial supplies of the study within the PeaMUST project. JLH, BD and ILH coordinated the overall progress and financial support of the study. ILH, BD and NB reviewed and contributed to draft the manuscript. All authors read and approved the final manuscript.

11. Acknowledgements

This work was supported by the French National Research Agency (ANR) through the project Investissement d'Avenir PeaMUST and by the region Nord-Pas-de-Calais through the scholarship of Sana Beji. The authors would like to thank the INRA experimental units of Estrées-Mons, Dijon and Theix, France, for their contribution to the field experiments. They are also grateful to Catherine Desmetz (UMR Agroécologie, INRA) for additional genotyping of the reference collection and Renaud Rincent (UMR GDEC, INRA) for his advices in genome wide association analyses.

12. References

1. FAOSTAT Food and Agriculture Organization of the United Nations - Statistic Division. 2018.

2. Graham, P. and Vance, C. Legumes: importance and constraints to greater use. Plant Physiol 2003, 131:872 - 877.

3. Jeuffroy, M.H., Baranger, E., Carrouee, B., de Chezelles, E., Gosme, M., Henault, C., Schneider, A., and Cellier, P. Nitrous oxide emissions from crop rotations including wheat, oilseed rape and dry peas. Biogeosciences 2013, 10(3):1787-1797.

4. Benezit, M., Biarnes, V., and Jeuffroy, M.H. Impact of climate and diseases on pea yields: what perspectives with climate change? OCL Oilseed.Fats Crops Lipids 2017, 24(1).

5. Lejeune-Hénaut, I., Bourion, V., Eteve, G., Cunot, E., Delhaye, K., and Desmyter, C. Floral initiation in field-grown forage peas is delayed to a greater extent by short photoperiods, than in other types of European varieties. Euphytica 1999, 109(3):201-211.

6. Fowler, D.B., Breton, G., Limin, A.E., Mahfoozi, S., and Sarhan, F. Photoperiod and Temperature Interactions Regulate Low-Temperature-Induced Gene Expression in Barley. Plant Physiol. 2001, 127(4):1676-1681.

7. Thomashow, M. Plant cold acclimation, freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 1999, 50:571 - 599.

8. Ruelland, E., Vaultier, M., Zachowski, A., and Hurry, V. Cold signalling and cold acclimation in plants. Adv Bot Res 2009, 49:35 - 146.

9. Rihan, H.Z., Al-Issawi, M., and Fuller, M.P. Advances in physiological and molecular aspects of plant cold tolerance. J. Plant Interact. 2017, 12(1):143-157.

10. Stockinger, E., Gilmour, S., and Thomashow, M. Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci U S A 1997, 94:1035 - 1040.

11. Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J. 1998, 16(4):433-442.

12. Thomashow, M. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiol 2010, 154:571-577.

13. Alonso-Blanco, C., Gomez-Mena, C., Llorente, F., Koornneef, M., Salinas, J., and Martinez-Zapater, J. Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. Plant Physiol 2005, 139:1304 - 1312.

14. Toth, B., Galiba, G., Feher, E., Sutka, J., and Snape, J.W. Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. Theor. Appl. Genet. 2003, 107(3):509-514.

15. Vagujfalvi, A., Galiba, G., Cattivelli, L., and Dubcovsky, J. The cold regulated transcriptional activator Cbf3 is linked to the frost tolerance locus Fr-A2 on wheat chromosome 5A. Mol Gen Genome 2003, 269:60-67.

16. Knox, A.K., Li, C.X., Vagujfalvi, A., Galilba, G., Stockinger, E.J., and Dubcovsky, J. Identification of candidate CBF genes for the frost tolerance locus Fr-A(m)2 in Triticum monococcum. Plant Molecular Biology 2008, 67(3):257-270.

Pasquariello, M., Barabaschi, D., Himmelbach, A., Steuernagel, B., Ariyadasa, R.,
 Stein, N., Gandolfi, F., Tenedini, E., Bernardis, I., Tagliafico, E., Pecchioni, N., and Francia,
 E. The barley Frost resistance-H2 locus. Funct. Integr. Genomics 2014, 14(1):85-100.

18. Sandve, S.R., Kosmala, A., Rudi, H., Fjellheim, S., Rapacz, M., Yamada, T., and Rognli, O.A. Molecular mechanisms underlying frost tolerance in perennial grasses adapted to cold climates. Plant Science 2011, 180(1):69-77.

19. Tayeh, N., Bahrman, N., Sellier, H., Bluteau, A., Blassiau, C., Fourment, J., Bellec, A., Debelle, F., Lejeune-Henaut, I., and Delbreil, B. A tandem array of CBF/DREB1 genes is located in a major freezing tolerance QTL region on Medicago truncatula chromosome 6. BMC Genomics 2013, 14(1):814.

20. Tondelli, A., Francia, E., Barabaschi, D., Pasquariello, M., and Pecchioni, N. Inside the CBF locus in Poaceae. Plant Science 2011, 180(1):39-45.

21. Lejeune-Hénaut, I., Hanocq, E., Béthencourt, L., Fontaine, V., Delbreil, B., Morin, J., Petit, A., Devaux, R., Boilleau, M., Stempniak, J.J., Thomas, M., Lainé, A.L., Foucher, F., Baranger, A., Burstin, J., Rameau, C., and Giauffret, C. The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. Theor. Appl. Genet. 2008, 116(8):1105-1116.

Dumont, E., Fontaine, V., Vuylsteker, C., Sellier, H., Bodele, S., Voedts, N., Devaux, R., Frise, M., Avia, K., Hilbert, J.L., Bahrman, N., Hanocq, E., Lejeune-Henaut, I., and Delbreil, B. Association of sugar content QTL and PQL with physiological traits relevant to frost damage resistance in pea under field and controlled conditions. Theor. Appl. Genet. 2009, 118(8):1561-1571.

23. Klein, A., Houtin, H., Rond, C., Marget, P., Jacquin, F., Boucherot, K., Huart, M., Riviere, N., Boutet, G., Lejeune-Henaut, I., and Burstin, J. QTL analysis of frost damage in pea suggests different mechanisms involved in frost tolerance. Theor. Appl. Genet. 2014, 127(6):1319-1330.

24. Gupta, P.K., Kulwal, P.L., and Jaiswal, V., Association Mapping in Crop Plants: Opportunities and Challenges, in Advances in Genetics, Vol 85, T. Friedmann, J.C. Dunlap, and S.F. Goodwin, Editors. 2014, Elsevier Academic Press Inc: San Diego. p. 109-147.

25. Li, Y.L., Bock, A., Haseneyer, G., Korzun, V., Wilde, P., Schon, C.C., Ankerst, D.P., and Bauer, E. Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. BMC Plant Biology 2011, 11.

26. Zhao, Y.S., Gowda, M., Wurschum, T., Longin, C.F.H., Korzun, V., Kollers, S., Schachschneider, R., Zeng, J., Fernando, R., Dubcovsky, J., and Reif, J.C. Dissecting the genetic architecture of frost tolerance in Central European winter wheat. J. Exp. Bot. 2013, 64(14):4453-4460.

27. Visioni, A., Tondelli, A., Francia, E., Pswarayi, A., Malosetti, M., Russell, J., Thomas, W., Waugh, R., Pecchioni, N., Romagosa, I., and Comadran, J. Genome-wide association mapping of frost tolerance in barley (Hordeum vulgare L.). BMC Genomics 2013, 14.

28. Tondelli, A., Pagani, D., Ghafoori, I.N., Rahimi, M., Ataei, R., Rizza, F., Flavell, A.J., and Cattivelli, L. Allelic variation at Fr-H1/Vrn-H1 and Fr-H2 loci is the main determinant of frost tolerance in spring barley. Environmental and Experimental Botany 2014, 106:148-155.

29. Yu, X.Q., Pijut, P.M., Byrne, S., Asp, T., Bai, G.H., and Jiang, Y.W. Candidate gene association mapping for winter survival and spring regrowth in perennial ryegrass. Plant Science 2015, 235:37-45.

30. Ali, M.B.M., Welna, G.C., Sallam, A., Martsch, R., Balko, C., Gebser, B., Sass, O., and Link, W. Association Analyses to Genetically Improve Drought and Freezing Tolerance of Faba Bean (Vicia faba L.). Crop Science 2016, 56(3):1036-1048.

31. Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N., and Martsch, R. Identification and Verification of QTL Associated with Frost Tolerance Using Linkage Mapping and GWAS in Winter Faba Bean. Frontiers in Plant Science 2016, 7.

32. Tumino, G., Voorrips, R.E., Rizza, F., Badeck, F.W., Morcia, C., Ghizzoni, R., Germeier, C.U., Paulo, M.J., Terzi, V., and Smulders, M.J.M. Population structure and genome-wide association analysis for frost tolerance in oat using continuous SNP array signal intensity ratios. Theor. Appl. Genet. 2016, 129(9):1711-1724.

33. Wurschum, T., Longin, C.F.H., Hahn, V., Tucker, M.R., and Leiser, W.L. Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. Plant J. 2017, 89(4):764-773.

34. Tayeh, N., Aubert, G., Pilet-Nayel, M.L., Lejeune-Henaut, I., Warkentin, T.D., and Burstin, J. Genomic Tools in Pea Breeding Programs: Status and Perspectives. Frontiers in Plant Science 2015, 6.

35. Desgroux, A., L'Anthoene, V., Roux-Duparque, M., Riviere, J.P., Aubert, G., Tayeh, N., Moussart, A., Mangin, P., Vetel, P., Piriou, C., McGee, R.J., Coyne, C.J., Burstin, J., Baranger, A., Manzanares-Dauleux, M., Bourion, V., and Pilet-Nayel, M.L. Genome-wide association mapping of partial resistance to Aphanomyces euteiches in pea. BMC Genomics 2016, 17.

36. Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., Bérard, A., Houtin, H., Rond, C., Kreplak, J., Boucherot, K., Martin, C., Baranger, A., Pilet-Nayel, M.-L., Warkentin, T.D., Brunel, D., Marget, P., Le Paslier, M.-C., Aubert, G., and Burstin, J. Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high density, high resolution consensus genetic map. The Plant Journal 2015:n/a-n/a.

37. Desgroux, A., Baudais, V.N., Aubert, V., Le Roy, G., de Larambergue, H., Miteul, H., Aubert, G., Boutet, G., Duc, G., Baranger, A., Burstin, J., Manzanares-Dauleux, M., Pilet-Nayel, M.L., and Bourion, V. Comparative Genome-Wide-Association Mapping Identifies Common Loci Controlling Root System Architecture and Resistance to Aphanomyces euteiches in Pea. Frontiers in Plant Science 2018, 8.

38. Liu, R., Fang, L., Yang, T., Zhang, X.Y., Hu, J.G., Zhang, H.Y., Han, W.L., Hua, Z.K., Hao, J.J., and Zong, X.X. Marker-trait association analysis of frost tolerance of 672 worldwide pea (*Pisum sativum* L.) collections. Sci Rep 2017, 7.

39. Burstin, J., Salloignon, P., Chabert-Martinello, M., Magnin-Robert, J.B., Siol, M., Jacquin, F., Chauveau, A., Pont, C., Aubert, G., Delaitre, C., Truntzer, C., and Duc, G. Genetic diversity and trait genomic prediction in a pea diversity panel. BMC Genomics 2015, 16.

40. R Core Team R : A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2014. Accessed 2016 7 Jun; http://www.R-project.org].

41. Hervé, M. RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-36. 2014. Accessed 2016 7 Jun; http://CRAN.R-project.org/package=RVAideMemoire].

42. Bates, D., Machler, M., Bolker, B.M., and Walker, S.C. Fitting Linear Mixed-Effects Models Using Ime4. J. Stat. Softw. 2015, 67(1):1-48.

43. Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., and Fox, J. lme4: Linear Mixed-Effects Models using 'Eigen' and S4. R package version 1.1-18-1. 2018. Accessed 2018 17 Aug; https://CRAN.R-project.org/package=lme4].

44. Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A.-L., Jacquin, F., Klein, A., Martin, C., Boucherot, K., Kreplak, J., da Silva, C., Moreau, S., Gamas, P., Wincker, P., Gouzy, J., and Burstin, J. Full-length de novo assembly of RNA-seq data in pea (*Pisum*

sativum L.) provides a gene expression atlas and gives insights into root nodulation in this species. The Plant Journal 2015, 84(1):1-19.

45. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007, 81(3):559-575.

46. Purcell, S.M. PLINK 1.9. 2017. Accessed 2017 13 Feb; https://www.cog-genomics.org/plink2].

47. Browning, B.L. and Browning, S.R. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am. J. Hum. Genet. 2009, 84(2):210-223.

48. Weir, B., . Genetic Data Analysis II: Methods for Discrete Population Genetic Data.1996, Sunderland, Massachussets: Sinauer Associates, Inc. 437.

49. Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 2006, 38(2):203-208.

50. Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. FaST linear mixed models for genome-wide association studies. Nat. Methods 2011, 8(10):833-U94.

51. Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A.E., Malvar, R.A., Moreno-Gonzalez, J., Nicolas, S., Madur, D., Combes, V., Dumas, F., Altmann, T., Brunel, D., Ouzunova, M., Flament, P., Dubreuil, P., Charcosset, A., and Mary-Huard, T. Recovering Power in Association Mapping Panels with Variable Levels of Linkage Disequilibrium. Genetics 2014.

52. Jombart, T., Devillard, S., and Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. Bmc Genetics 2010, 11.

53. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008, 24(11):1403-1405.

54. Jombart, T. and Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 2011, 27(21):3070-3071.

55. Jombart, T., Kamvar, Z., Collins, C., Lustrik, R., Beugin, M., Knaus, B., Solymos, P., Mikryukov, V., Schliep, K., Maié, T., Morkovsky, L., Ahmed, I., Cori, A., Calboli, F., and Ewing, R. adegenet: Exploratory Analysis of Genetic and Genomic Data version 2.1.1. 2018. Accessed 2018 08 Jun; https://cran.r-project.org/web/packages/adegenet/].

56. Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., Frouin, J., Rouan, L., Goze, E., Kilian, A., Ahmadi, N., and Dingkuhn, M. Genome-Wide Association Mapping of Root Traits in a Japonica Rice Panel. PLoS ONE 2013, 8(11).

57. Turner, S. qqman: Q-Q and Manhattan Plots for GWAS Data. R package version 0.1.4.
2017. Accessed 2018 20 Aug; https://CRAN.R-project.org/package=qqman].

58. de Mendiburu, F. agricolae: Statistical Procedures for Agricultural Research. R package version 1.2.8. 2017. Accessed 2018 07 Jun; https://cran.r-project.org/web/packages/agricolae/index.html].

Weller, J.L., Liew, L.C., Hecht, V.F.G., Rajandran, V., Laurie, R.E., Ridge, S., Wenden,
 B., Vander Schoor, J.K., Jaminon, O., Blassiau, C., Dalmais, M., Rameau, C., Bendahmane,
 A., Macknight, R.C., and Lejeune-Hénaut, I. A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proceedings of the National Academy of Sciences 2012, 109(51):21158-21163.

Sun, X., Yang, T., Hao, J., Zhang, X., Ford, R., Jiang, J., Wang, F., Guan, J., and Zong,
X. SSR genetic linkage map construction of pea (*Pisum sativum* L.) based on Chinese native varieties. The Crop Journal 2014, 2(2):170-174.

61. Hamon, C., Baranger, A., Coyne, C.J., McGee, R.J., Le Goff, I., L'Anthoëne, V., Esnault, R., Rivière, J.-P., Klein, A., Mangin, P., McPhee, K.E., Roux-Duparque, M., Porter, L., Miteul, H., Lesné, A., Morin, G., Onfroy, C., Moussart, A., Tivoli, B., Delourme, R., and Pilet-Nayel, M.-L. New consistent QTL in pea associated with partial resistance to Aphanomyces euteiches in multiple French and American environments. Theor. Appl. Genet. 2011, 123(2):261-281.

62. Tayeh, N., Bahrman, N., Devaux, R., Bluteau, A., Prosperi, J., Delbreil, B., and Lejeune-Henaut, I. A high-density genetic map of the Medicago truncatula major freezing tolerance QTL on chromosome 6 reveals colinearity with a QTL related to freezing damage on *Pisum sativum* linkage group VI. Mol Breeding 2013, 32:279 - 289.

63. Zhu, J., Pearce, S., Burke, A., See, D.R., Skinner, D.Z., Dubcovsky, J., and Garland-Campbell, K. Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. Theor. Appl. Genet. 2014, 127(5):1183-1197.

64. Novak, A., Boldizsar, A., Gierczik, K., Vagujfalvi, A., Adam, E., Kozma-Bognar, L., and Galiba, G. Light and Temperature Signalling at the Level of CBF14 Gene Expression in Wheat and Barley. Plant Mol. Biol. Rep. 2017, 35(4):399-408.

65. Eremina, M., Unterholzner, S.J., Rathnayake, A.I., Castellanos, M., Khan, M., Kugler, K.G., May, S.T., Mayer, K.F.X., Rozhon, W., and Poppenberger, B. Brassinosteroids participate in the control of basal and acquired freezing tolerance of plants. Proceedings of the National Academy of Sciences of the United States of America 2016, 113(40):E5982-E5991.

66. Li, H., Ye, K., Shi, Y., Cheng, J., Zhang, X., and Yang, S. BZR1 Positively Regulates Freezing Tolerance via CBF-Dependent and CBF-Independent Pathways in Arabidopsis. Molecular Plant 2017, 10(4):545-559.

67. Anwar, A., Liu, Y.M., Dong, R.R., Bai, L.Q., Yu, X.C., and Li, Y.S. The physiological and molecular mechanism of brassinosteroid in response to stress: a review. Biol. Res. 2018, 51.

68. Lester, D.R., Ross, J.J., Davies, P.J., and Reid, J.B. Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell 1997, 9(8):1435-1443.

69. Martin, D.N., Proebsting, W.M., and Hedden, P. Mendel's dwarfing gene: cDNAs from the Le alleles and function of the expressed proteins. Proceedings of the National Academy of Sciences of the United States of America 1997, 94(16):8907-8911.

70. Achard, P., Gong, F., Cheminant, S., Alioua, M., Hedden, P., and Genschik, P. The Cold-Inducible CBF1 Factor–Dependent Signaling Pathway Modulates the Accumulation of the Growth-Repressing DELLA Proteins via Its Effect on Gibberellin Metabolism. The Plant Cell Online 2008, 20(8):2117-2129.

71. Xin, Z. and Browse, J. Cold comfort farm: the acclimation of plants to freezing temperatures. Plant Cell and Environment 2000, 23(9):893-902.

72. Wiseman, J., Al-Mazooqi, W., Welham, T., and Domoney, C. The apparent ileal digestibility, determined with young broilers, of amino acids in near-isogenic lines of peas (*Pisum sativum* L.) differing in trypsin inhibitor activity. Journal of the Science of Food and Agriculture 2003, 83(7):644-651.

73. Page, D., Duc, G., Lejeune-Henaut, I., and Domoney, C. Marker-assisted selection of genetic variants for seed trypsin inhibitor contents in peas. Genetics 2003, 35:19-21.

1. Introduction du chapitre

La cartographie fine du QTL de tolérance au gel WFD6.1, comme décrit au cours du premier chapitre du manuscrit, ne nous a permis de réduire l'intervalle de confiance du QTL qu'à une taille de 3,9 cM. Cela est probablement dû au fait que sa position est proche d'une région centromérique et que le taux de recombinaison à cette zone est par conséquent très réduit, ce qui ne permet pas d'obtenir des lignées recombinantes indispensables pour l'étape de la cartographie fine. Nous avons donc entrepris une approche de clonage en utilisant les informations liées à la synténie avec *Medicago truncatula* obtenues lors de travaux précédents (Tayeh et al. 2013a) et celles obtenues par les analyses de génétique d'association (cf. Chapitre III).

Tayeh et al. (2013b) ont réalisé une cartographie fine d'un QTL de tolérance au gel chez *Medicago truncatula*, le QTL Mt-FTQTL6 qui est synténique du QTL WFD6.1 de *Pisum sativum*. Ces auteurs ont identifié 20 gènes candidats sous-jacents à l'intervalle de confiance réduit de Mt-FTQTL6, dont un cluster de 12 gènes de facteurs de transcription C-repeat binding factors (CBF) organisés en tandem (Tayeh et al. 2013b). Cette famille de facteurs de transcriptions est connue pour son rôle dans la régulation de la majorité des gènes impliqués dans la tolérance au froid chez les plantes (cf. Chapitre I).

Dans le deuxième chapitre de ce manuscrit portant sur l'analyse par génétique d'association d'une collection de ressources génétiques de pois pour la tolérance au gel, nous avons identifié, au niveau du groupe de liaison VI du pois, un marqueur correspondant à un gène *CBF*, qui est en déséquilibre de liaison avec 5 marqueurs significativement associés à la tolérance au gel.

Suite à ces résultats, nous avons émis l'hypothèse qu'il existerait un cluster de gènes *CBF* sousjacents à WFD6.1 impliqué dans la tolérance au gel chez le pois. Par conséquent, nous avons cherché à identifier la séquence de cette région afin d'évaluer le nombre et le type de *CBF* présents au locus WFD6.1 et d'étudier les éventuelles différences entre un génotype sensible et un génotype tolérant au gel. Ceci peut s'avérer utile pour l'analyse du rôle de ces gènes mais également en tant que marqueurs dans la sélection assistée par marqueurs pour augmenter la tolérance au gel.

Pour faire de la cartographie physique locale au locus d'intérêt, la méthode que nous avons employée est le développement de deux banques BAC (Bacterial Artificial Chromosome) à

partir de l'ADN des deux génotypes de *Pisum sativum* présentant des phénotypes contrastés pour la tolérance au gel. Puis, ces deux banques ont été criblées par des marqueurs *CBF*. Les clones identifiés ont été séquencés afin d'obtenir une carte physique de la région d'intérêt pour chacun des génotypes étudiés.

Dans ce chapitre de thèse, nous présentons la construction des banques BAC, leur caractérisation, les résultats du criblage de ces banques ainsi que ceux du séquençage et de l'assemblage des clones positifs porteurs des gènes candidats.

Ce chapitre rapporte les premières étapes de la construction de la carte physique du locus d'intérêt. Le développement d'un contig consensus pour chacun des deux génotypes étudiés, par le contiguage des séquences assemblées des clones BAC positifs préalablement identifiés, ainsi que leur caractérisation restent à réaliser.

2. Matériels et Méthodes

2.1. Production du matériel végétal pour la construction des banques BAC

Deux banques BAC ont été construites à partir de deux génotypes de *Pisum sativum* (Champagne et Térèse) différents pour leur niveau de tolérance au gel et pour leur génotype au locus WFD6.1.

Les deux génotypes ont été mis en culture dans une chambre climatique à 20 ° C, avec un cycle lumière/obscurité de 16/8 h. Après 12 jours (Figure IV.1), 20 g de feuilles de jeunes plantes ont été récoltées pour chaque génotype et congelées dans l'azote liquide suivant le protocole décrit dans l'annexe 1. A la fin du prélèvement, tous les échantillons ont été conservés à -80 °C jusqu'au moment de l'extraction de l'ADN de haut poids moléculaire.



Figure IV. 1 : Photo du matériel végétal le jour du prélèvement.

2.2. Construction des banques BAC

Après l'extraction de l'ADN de haut poids moléculaire à partir des noyaux des jeunes feuilles des deux génotypes, ce dernier est préparé pour être inséré dans un vecteur bactérien. L'ADN est digéré partiellement par l'enzyme de restriction HindIII, et les fragments résultants sont séparés sur gel d'agarose, puis ceux dont la taille est comprise entre 100 et 250 kb sont sélectionnés. Ces fragments sont purifiés et utilisés pour une ligation dans un vecteur de clonage pAGIBAC HindIII afin de transformer des cellules d'Escherichia coli électro-compétentes. Le vecteur utilisé est caractérisé d'une part par un gène de résistance à un antibiotique (chloramphénicol) qui permettra à la bactérie porteuse de ce plasmide de ne pas être sensible à l'effet de cet antibiotique, et d'autre part par le site d'insertion de l'insert situé sur le gène lacZ qui code la □-galactosidase de l'opéron lactose. L'inactivation de ce gène par l'insertion du fragment d'ADN du pois est révélée par l'absence de l'activité enzymatique
-galactosidase. Ceci est mis en évidence par le substrat galactoside (X-Gal) dont la couleur passe de l'incolore au bleu quand il est clivé par la 🗆 -galactosidase. Pour pouvoir métaboliser le X-gal, la cellule doit être exposée à un inducteur, l'IPTG (isopropylthio-□-D-galactoside). Ensuite, les cellules transformées sont mises en culture sur un milieu solide sélectif en présence de chloramphénicol, IPTG et X-Gal. Les colonies blanches sont sélectionnées et repiquées à l'aide d'une station robotique à haut débit QPix2 XT (Genetix), et réarrangées en microplaques de 384 puits

contenant un milieu de croissance LB (milieu de *Lysogeny Broth*). Après incubation pendant 12h à 37°C, les plaques sont conservées à -80°C pour les analyses ultérieures.

Une stratégie complémentaire de construction des banques BAC a également été utilisée pour les deux génotypes de pois. Elle utilise un protocole sans réarrangement : on parle aussi de banque BAC en pool ou de banques BAC "à façon". Les bactéries transformées avec un plasmide recombinant sont maintenues en mélanges de clones (pool). Cette approche permet l'isolement de clones d'intérêt, tout en diminuant le nombre de réactions nécessaires au criblage d'une banque par PCR. Les protocoles suivis pour la construction des banques BAC avec et sans réarrangement sont présentés en annexe 2.

2.3. Estimation de la taille d'insert

L'ADN de quelques colonies sélectionnées au hasard au sein des deux banques est isolé à l'aide du kit de purification rapide Nucleobond Xtra Midi plus (Macherey Nagel) afin d'estimer la taille de leurs inserts (Annexe 3). L'ADN de chaque BAC est digéré avec *Not*I pour exciser les inserts (Annexe 4). L'ADN digéré est ensuite séparé dans un gel d'agarose par électrophorèse en champ pulsé, la taille des inserts estimée et le pourcentage de clones vides calculé à l'aide du logiciel Gene Tools. Enfin, la couverture génomique de la banque (en kb) est déduite par la formule suivante :

 $Couvertureg\acute{e}nomique(kb) = \frac{Taillemoyennedesinserts(kb) \times nombredeclones}{Tailledug\acute{e}nome(kb)}$

2.4. Criblage des banques BAC

2.4.1. Méthodes d'identification de clones BAC positifs par hybridation de macroarrays

Cette méthode repose sur la complémentarité des bases constitutives des acides nucléiques et sur la réversibilité du processus de séparation (dénaturation) et de réassociation (renaturation) des deux brins d'ADN complémentaires. Lorsqu'il y a renaturation (reformation des liaisons hydrogènes entre les deux brins) on parle aussi d'hybridation. Les macroarrays portent l'ADN de bactéries transformées avec lequel des sondes marquées peuvent s'hybrider lorsque les conditions sont optimales. En cas d'homologie forte, l'hybride reste stable même en condition

de contraintes importantes (lavages). Le signal radioactif (capté par l'écran du phosphoimager) permet d'identifié le clone hybridé.

2.4.1.1. Production des membranes (macroarrays)

Des membranes de colonies hautes densité (macroarrays) sont préparées à partir des banques de BAC de Champagne et Térèse en utilisant la station de travail robotisée Q-Pix (Genetix). Le robot piqueur utilise une tête comportant 384 aiguilles qui prélève des aliquots de bactéries dans une plaque de culture et les dépose de façon ordonnée sur des membranes de nylon de 22x22 cm ayant un motif de 7x7 (Figure IV.2). Sur chaque membrane ou filtre, des clones uniques sont déposés en deux exemplaires selon un motif particulier permettant d'identifier les coordonnées des clones positifs lors du criblage. Ces clones sont cultivés à 37 ° C pendant 17 h. Les membranes sont ensuite traitées comme suit : (1) humidification par capillarité avec un mesh (drain) puis la membrane est placée dans une solution d'hybridation (SSC 6x + SDS 1,5% + Denhardt 5x), (2) l'ensemble membrane-mesh est roulé dans un tube à hybridation contenant 50 ml de la solution d'hybridation de façon à ce que la membrane se déroule au maximum, en se plaquant contre la paroi du tube, et (3) incubation au four pendant 15 min à 50° C.

La qualité des membranes est vérifiée après production par pré-hybridation à 68°C pendant 1h30 à 6 heures, à l'aide d'une sonde témoin correspondant à l'ADN de sperme de saumon préalablement dénaturée à 95°C pendant 10 min.



Figure IV. 2 : Production de membranes de nylon (macroarrays) de 22 x 22 cm avec un motif de 7 x 7 (www.cnrgv.toulouse.inra.fr).

2.4.1.2. Préparation des sondes

Les sondes utilisées pour cribler les banques BAC dans le cadre du clonage positionnel du locus WFD6.1 ont été définies à partir des marqueurs moléculaires dessinés sur les séquences de gènes *CBF* repérées dans le génome de référence de la lignée de pois Caméor grâce aux premières annotations de la carte physique en préparation au sein de l'unité Agroécologie de Dijon (données non publiées). Ces gènes ont été identifiés par homologie avec les gènes candidats de *Medicago truncatula* sous-jacents au QTL Mt- FTQTL6, synténique du QTL WFD6.1 de *Pisum sativum*. Les principales caractéristiques de ces marqueurs sont présentées dans le tableau IV.1. Au total, nous disposons de 13 marqueurs appartenant à 3 scaffolds de la carte physique de Caméor (Tableau IV.1 et Annexe 7). Parmi ces sondes, notons la présence de 2 marqueurs appartenant au bloc de déséquilibre de liaison VI.4 associé à la tolérance au gel par analyse d'association (PsCam037752 et PsCam050192, Tableau IV-1 et Additional file 9).

Avant de procéder à l'étape de criblage des banques BAC, les sondes ont été testées par PCR sous 2 conditions, 62°C et 66°C, avec de l'ADN génomique de Térèse et Champagne. Les produits PCR obtenus ont ensuite été purifiés et migrés sur gel d'agarose pour être dosés.

Pour le marquage des sondes nous procédons de la manière suivante : après amplification par PCR, les sondes sont marquées par incorporation d'alpha-P³³-dCTP selon la méthode de random priming. Chaque sonde, marquée est ensuite purifiée sur colonne Sephadex (Annexe 5).

Scaffold/Contig de la	м	Taille de	C I
carte physique de Caméor	Marqueur	(pb)	Sonde
Scaffold426	CBF426_6484	467	CBF426_6484_F
Contig31084	(= PsCam037752; 50,2 cM)	40/	CBF426_6484_R
Scaffold303	CDE202 4920	405	CBF303_4830_F
Contig23826	CDI 505_4050	ч <i>у</i> 5	CBF303_4830_R
Scaffold303	CBF303 /831	465	CBF303_4831_F
Contig23826	001000_1001	105	CBF303_4831_R
Scaffold303	CBF303 4832	435	CBF303_4832_F
Contig23845	001303_1032	155	CBF303_4832_R
Scaffold303	CBF303_4834	474	CBF303_4834_F
Contig23859	(= PsCam050192 ; 50,2 cM)	1,1	CBF303_4834_R
Scaffold211 Contig17830	CBF211_3603	439	CBF211_3603_F
		107	CBF211_3603_R
Scaffold211	CBF211_3604	484	CBF211_3604_F
Contig17830			CBF211_3604_R
Scaffold211 Contig17831	CBF211_3605	531	CBF211_3605_F
			CBF211_3605_R
Scaffold211 Contig17831	CBF211_3606	419	CBF211_3606_F
			CBF211_3606-R
Scaffold211	CBF211_3607	480	CBF211_3607_F
Contig17834			CBF211_3607_R
Scaffold211 Contig17834	CBF211_3608	419	CBF211_3608_F
			CBF211_3608_R
Scaffold211	CBF211 3609	444	CBF211_3609_F
Contig17837	_		CBF211_3609_R
Scaffold211	CBF211_3610	402	CBF211_3610_F
Contig17837			CBF211_3610_R

Tableau IV. 1 : Marqueurs utilisés pour le criblage des banques BAC de Champagne et Térèse

2.4.1.3. Hybridation

La sonde radiomarquée et purifiée est ajoutée dans le tube à hybridation puis incubée pendant une nuit à 68°C. Ensuite, nous procédons au lavage des membranes en deux étapes successives :

(1) un premier lavage pendant 15 min à 50°C à l'aide de 50 ml d'une solution de lavage (SSC 2x + SDS 0,1%), et (2) un deuxième lavage pendant 30 min à 50°C à l'aide de 50 ml d'une solution de lavage moins concentrée (SSC 0,5x + SDS 0,1%). Après hybridation, les macroarrays sont exposées pendant 2 jours et révélées à l'aide d'un PhosphoImager. Le protocole d'hybridation des macroarrays est détaillé dans l'annexe 6.

Pour limiter le travail laborieux du protocole du criblage par hybridation, nous avons procédé à des séries d'hybridations sur la base de la stratégie suivante. En parallèle de l'hybridation de certaines membranes avec un seul marqueur pour chacun des deux génotypes étudiés, nous avons parfois utilisé simultanément deux marqueurs pour cribler une même membrane par génotype. Le tableau IV.2 résume les séries de criblage par hybridation réalisées pour les deux génotypes étudiés.

	Macroarray / Marqueur(s) utilisés
	Macroarray1 / CBF426_6484 + CBF211_3604
Série d'hybridation1	Macroarray2 / CBF303_4832
	Macroarray3 / CBF211_3609
Série d'hybridation 2	Macroarray4 / CBF211_3607
	Macroarray5 / CBF211_3603 + CBF211_3610
Série d'hybridation 3	Macroarray6 / CBF303_4831
	Macroarray7 / CBF211_3610

Tableau IV. 2 : Série d'hybridations utilisées pour les banques BAC de Champagne et Térèse

2.4.1.4. Analyse des résultats qualitatifs

Nous identifions les clones positifs grâce au logiciel High Denstity Filter Reader version 3. Ce logiciel appose une grille sur les images obtenues, détecte la quantité de pixels par unité de surface et enfin détermine les coordonnées des clones contenant la séquence d'intérêt. Un clone positif correspond à deux signaux radioactifs.

2.4.1.5. Réarrangement des clones d'intérêts identifiés par criblage des macroarrays et validation par PCR

Les clones positifs sont identifiés dans les microplaques de la banque BAC du pois stockées à -80°C, pour être par la suite transférés et réorganisés dans une microplaque « fille » contenant un milieu de culture approprié. Ce réarrangement est assuré par la station robotique QPix2 XT qui garantit la traçabilité et la filiation des échantillons réarrangés. La microplaque « fille »

contenant les clones d'intérêts est ensuite scellée puis incubée à 37°C pendant 20 heures sous agitation (125 rpm). Enfin, les clones positifs sont validés par PCR avec les amorces utilisées lors de l'hybridation des microarrays. Un clone positif est validé si : (i) Le Ct (*Cycle Threshold*) correspond à celui du témoin (ADNg de Térèse ou Champagne), (i) la température de fusion (Tm, *Melting Temperature*) correspond au Tm attendu, et (iii) la taille du fragment amplifié correspond à celui de l'ADNg du témoin.

Pour surmonter les étapes laborieuses du protocole de criblage des banques par hybridation, nous avons utilisé la stratégie de cross validation. Les clones positifs issus des premières séries d'hybridation sont validés par PCR non seulement par les marqueurs utilisés comme sonde pour le criblage par hybridation, mais aussi par d'autres marqueurs dans l'objectif de détecter des clones positifs à plusieurs marqueurs.

2.4.2. Méthodes de criblage de la banque BAC non réarrangée

Le schéma de la figure IV.3 résume les différentes étapes de la stratégie de criblage de banque non réarrangée. En premier lieu, l'ADN des clones présents dans les pools de clones est amplifié par une ADN polymérase particulière : Whole Genome polymérase Phi29. L'enzyme Phi29 est une ADN polymérase issue d'un bactériophage qui amplifie de façon exponentielle différents types d'échantillons ADN simple ou double brin(s). L'amplification est réalisée à partir d'amorces aléatoires de 6 pb qui s'hybrident sur l'ADN dénaturé simultanément sur plusieurs sites. Les fragments obtenus ont une taille comprise entre 1 et 70 kb.

Les amplifiâts sont ensuite dilués au 1/200e pour être utilisés comme matrice pour le criblage par PCR en temps réel afin d'identifier le(s) pool(s) contenant le clone porteur de la séquence du marqueur cible.

Un pool positif est repéré si : le Ct est correct par rapport à l'ADNg du témoin, son Tm correspond au Tm attendu et la taille du fragment amplifié est correcte par rapport au témoin d'ADNg. Dans ce cas, le pool doit être systématiquement revalidé individuellement pour limiter le risque de faux positifs. Si le pool est validé, on peut alors passer à l'étape suivante de validation sur pool bruts bactériens.

Le clone d'intérêt est isolé de l'aliquot du pool identifié et validé individuellement suivant les étapes suivantes :

- Deux dilutions de l'aliquot positif (1000^e et 10000^e) sont étalées sur une grande boîte de Petri appelé Q-Tray (le format est adapté au robot Q-Pix qui intervient dans les étapes suivantes) et sont mises en culture 20 heures à 37°C. Puis en fonction du résultat obtenu, si besoin, re-étaler le nombre requis de Qtrays pour pouvoir piquer sur plaques 384, deux fois le nombre de plaques théoriques pour retrouver le clone d'intérêt en adaptant si besoin la dilution. Le nombre de plaques théoriques est le nombre de clones uniques estimés dans le pool divisé par 384.
- Le repiquage est ensuite réalisé grâce au robot Q-Pix en sélectionnant des clones bactériens isolés selon leurs caractéristiques morphologiques parmi un ensemble de clones poussés sur la boite de Petri. Les plaques ensemencées sont ensuite scellées puis incubées à 37 °C pendant 20 heures sous agitation.
- Le criblage se fait en réalisant la stratégie de pooling plaque pour chaque plaque ensemencée par le QPix2 XT, dans une plaque réservoir contenant 25ml de milieu LB + chloramphénicol + glycérol. Cette stratégie consiste à mélanger tous les puits d'une plaque dans une plaque réservoir qui sera par la suite notre matrice de criblage. Cette méthode permet d'éviter un grand nombre de criblage PCR. Après croissance à 37°C pendant 18h sous agitation (45 rpm), les pools obtenus sont testés par PCR. Si une plaque positive est repérée, des pools ligne-colonne secondaires sont produits pour identifier la coordonnée du clone recherché. C'est la même démarche que pour le pooling plaque, sauf que cette fois-ci ce sont les puits d'une même ligne qui sont mélangés dans une plaque réservoir « ligne » et les puits d'une même colonne qui sont mélangés dans une plaque réservoir « colonne ». Après croissance à 37°C pendant 18h, l'ensemble des pools dilués au demi est criblé par PCR.
- Les clones repérés sont systématiquement validés individuellement sur colonies isolées et remis en culture stock (tube ou plaques) dans l'optique de leur caractérisation (séquençage des extrémités (BES : *BAC End Sequences*), séquençage pleine longueur, estimation de la taille).



Figure IV. 3 : Stratégie de criblage des banque BAC en pool (extrait du site web du Centre National de Ressources Génomiques Végétales (CNRGV) (www.cnrgv.toulouse.inra.fr).

2.5. Séquençage des BAC

2.5.1. Méthode de séquençage des BAC-ends

Le séquençage des BES permet de positionner les clones BAC les uns par rapports aux autres. Il consiste à séquencer les extrémités de l'insert cloné dans le vecteur pAGIBAC. L'utilisation de ce vecteur a permis d'amorcer le séquençage des BES avec les séquences d'amorces universelles M13rev et T7 déjà présentes sur le vecteur de part et d'autres du site d'insertion. L'extraction de l'ADN des clones a été réalisée en plaques de 96 puits, avec le kit d'extraction Nucleobond Xtra Midi plus (Macherey Nagel ref 740410.100). Les échantillons sont au préalable amplifiés par PCR avant d'être séquencés par le CNRGV en collaboration avec la plateforme Génomique de l'INRA de Toulouse (Get-PlaGe-GENOTOUL via la méthode Sanger (avec l'automate ABI3730-Applied Biosystems).

2.5.2. Méthode de séquençage des clones BAC et traitement des données brutes

Le séquençage complet des clones BAC, issue de la banque réarrangée et non réarrangée, a été réalisé via la technologie de séquençage PacBio® RSII (kit de chimie P6-C4). L'ADN des

clones BAC est isolé et préparé en appliquant la méthode Macherey-Nagel permettant d'obtenir de l'ADN de très bonne qualité (annexe 3).

Après une étape de démultiplexage, l'assemblage des séquences (reads) générées par le séquenceur PacBio® est réalisé selon le standard PacBio® workflow d'assemblage génomique hiérarchique (HGAP : Hierarchical Genome Assembly Process) (revue par Chin et al. (2013)) suivant les principales étapes suivantes :

- préassemblage et correction des reads bruts,
- détection et élimination des séquences du génome de la bactérie hôte E.coli,
- assemblage des reads,
- nettoyage des reads en supprimant les séquences trop petites et de faible qualité.

Enfin, les contigs obtenus sont validés grâce à l'alignement des BES sur les extrémités des séquences assemblées.

2.6. Dessin d'amorces pour le contigage des BAC chevauchants

Des couples d'amorces sont définis sur la séquence des BES, afin de pouvoir cribler les clones BAC et de mettre en évidence le chevauchement ou non des clones. Pour avoir des amorces les plus spécifiques possible, les séquences répétées ont été masquées à l'aide d'une application Web «repeatmasker.org (http://www.repeatmasker.org/). Les bases nucléiques définies comme de l'ADN répété sont détectées et remplacées par des « N ». La séquence résultante est par la suite récupérée en format FASTA. Ensuite, le design des amorces est effectué à l'aide de l'application web Primer3 (http://primer3.ut.ee/) avec les paramètres suivants : une longueur de la paire d'amorces de 20 pb, une température de fusion (Tm) de 60°C et une longueur d'amplicon de 400 pb avec un %GC aux alentours de 50%.

3. Résultats et discussions

3.1. Caractérisation des deux banques BAC

Les tableaux IV.3 et IV.4 résument les principales caractéristiques des banques BAC classiques et en pool correspondant aux deux génotypes étudiés. Les banques BAC classiques de Champagne et Térèse (tableau IV.3) sont composées chacune de 55 296 clones répartis en 288 microplaques à 384 puits. Une sélection aléatoire de clones a été analysée pour les deux banques

Champagne et Térèse. Elle a déterminé que le pourcentage de clones sans insert était respectivement de 0% et 0,1 % pour les banques de Champagne et Térèse. Les tailles moyennes d'inserts des BAC de Champagne et Térèse sont de 127 kb et 139 kb, respectivement. Ces deux banques correspondent à une couverture du génome de 1,63 et de 1,34 fois respectivement pour Champagne et Térèse. Pour la banque de Térèse, nous notons une couverture du génome (1.34 x) plus faible que celle escomptée (1,60 x). Ceci est dû à un problème de repiquage des clones par la station robotique QPix2 XT pour certaines microplaques de la banque de Térèse. En ce qui concerne les deux banques en pool, grâce à la stratégie de pooling nous notons un nombre plus élevé de clones obtenus par rapport à la banque BAC classique. En effet un total de 144 576 clones, contenant des inserts de 103 kb en moyenne a été obtenu pour la banque en pool de Champagne. Pour la banque de Térèse, 248 408 clones contenant des inserts de 112 kb en moyenne ont été obtenus. Ces deux banques correspondent à une couverture expérimentale finale de 3,18 x et 5,89 x respectivement pour le génome de Champagne et de Térèse (Tableau IV.4).

	Banque BAC de	Banque BAC de
	Champagne	Térèse
Nombre de clones	55 296	55 296
Nombre de plaques à 384 puits	144	144
Taille moyenne des inserts (kb)	127	139
Clones vides (%)	0	10
Couverture du génome (théorique) 1,63	1,60
Couverture du génome (finale)	1.63	1.34

Tableau IV. 3 : Caractéristiques des banques BAC réarrangées de Champagne et Térèse

	Banque BAC de Champagne	Banque BAC de Térèse
Nombre de clones	144 576	248 408
Nombre de pools	94	96
Nombre de clones par pool	1 538	2 587
Taille moyenne des inserts (kb)	103	112
Clones vides (%)	8	9
Couverture du génome (théorique)	3,46 x	6,47 x
Couverture du génome (finale)	3,18 x	5,89 x

Tableau IV. 4 : Caractéristiques des banques BAC en pool de Champagne et Térèse

3.2. Caractérisation des marqueurs utilisés pour le criblage des banques BAC

Avant de commencer tout criblage, les conditions PCR optimales ont été établies pour chaque marqueur en faisant une PCR en gradient de température à 62 °C et à 66 °C correspondant à la température d'hybridation. Les Tm et Ct obtenus en PCR pour les 13 marqueurs étudiés sont présentés dans le tableau IV.5. Les résultats montrent que seules les réactions PCR des deux marqueurs CBF303_4830 et CBF303_4831 présentent des problèmes (Figure IV.4). Pour le marqueur CBF303_4831, nous avons été confrontés à des difficultés à amplifier la séquence de l'ADN génomique témoin de Champagne et Térèse pour les deux conditions de températures. Par contre pour le marqueur CBF303_4830, nous notons une amplification en condition de température d'hybridation de 66°C pour Champagne et une très légère amplification pour Térèse. Cependant, la taille du fragment amplifié est égale à 850 pb ce qui ne correspond pas à la taille attendue (495 pb).

Au final, la température de 66°C est choisie pour les expériences ultérieures de PCR étant donné que cette température assure une bonne spécificité et une amplification suffisante. La figure IV.5 montre un exemple des résultats de migration sur gel d'agarose des produits PCR de 4 marqueurs testés pour la température de 66°C. Les tailles de tous les fragments ont été en correspondance avec la taille de l'amplicon attendu.

Chapitre IV. Création de banques BAC pour la recherche de gènes candidats sous-jacents au QTL WFD6.1



Figure IV. 4 : Résultats de migration sur gel d'agarose des produits de PCR des deux marqueurs CBF303 4831 et CBF303 4830 testés pour la température d'hybridation de 66°C



Figure IV. 5 : Résultats de migration sur gel d'agarose des produits de PCR des quatre marqueurs CBF426_6484, CBF303_4832, CBF211_3604 et CBF211_3609 testés pour la température d'hybridation de 66°C

Dans un deuxième temps, plusieurs expériences de PCR ont été réalisées pour remédier au problème d'amplification des deux marqueurs CBF303_4830 et CBF303_4831 en utilisant deux autres couples d'amorces et en testant 16 températures d'hybridation allant de 55°C à 65°C. Dans cette expérience, l'ADN génomique de Caméor a été ajouté comme témoin en plus de celui de Champagne et Térèse. Les résultats de cette expérience révèlent qu'il n'a pas été possible d'obtenir une amplification de la séquence de CBF303_4831 sur l'ADN de Champagne et Térèse, ce qui pourrait être expliqué par un mésappariement entre les amorces et l'ADN de Champagne et Térèse et indiquer la spécificité de Caméor pour ce marqueur. Concernant le marqueur CBF303_4830, les résultats de PCR avec la nouvelle paire d'amorce

confirment le résultat trouvé précédemment, ce qui pourrait s'expliquer par le fait que les deux amorces dessinées ne sont pas spécifiques au marqueur.

Marqueur	Température d'hybridation testée en PCR	Tm*	Ct**
	62°C	82,5	-
CDF420_0404	66°C	82,5	25,5
CBF303_4830	62°C	73 - 80,5	-
	66°C	73	36
CDE202 4921	62°C	-	-
CBF303_4831	66°C	-	-
CDE202 4922	62°C	80,5	22,76
CDF303_4032	66°C	73	24,3
CDE202 4924	62°C	82,5 - 87,5	22,63
CDF303_4034	66°C	80,5	25,4
CBF211_3603	62°C	82,5	75
	66°C	82,5 - 87,5	19
CDE211 2604	62°C	80	21
CDF211_3004	66°C	82,5	19,3
CBF211_3605	62°C	82,5	23
	66°C	80,5	26,5
CDE211 2(0(62°C	82	21,8
CBF211_5000	66°C	82,5	20,7
CDE211 2607	62°C	81,5 - 82	20,56
CBF211_360/	66°C	82	21,22 - 21,44
CBF211_3608	62°C	81,5 - 82	21,8
	66°C	82	20,7
CBF211_3609	62°C	85,5	22
	66°C	85,5	23,6
CRE211 2610	62°C	84,5	21,6
CBF211_3010	66°C	84	22,25

Tableau IV. 5 : Conditions d'amplification par PCR des amorces utilisées

*Tm (Melting temperature) : température de fusion ; ** Ct : Cycle Threshold

3.3. Criblage des banques BAC

Onze marqueurs définis sur la région d'intérêt et sur les séquences de CBF repérées dans le génome de référence *Ps* Caméor, ont été validés par amplification PCR sur l'ADN génomique de Champagne et Térèse. Des clones contenant ces marqueurs ont été sélectionnés dans les deux banques BAC de Champagne et Térèse par la technique d'hybridation sur macroarray suivie par la validation par amplification PCR de la séquence d'intérêt.

Dans la partie suivante, je présenterai les résultats du criblage des banques BAC de Champagne et Térèse pour le marqueur CBF303_4832 à titre d'exemple (Figure IV.6), puis je synthétiserai les résultats de criblage des deux banques (Figure IV.7).

La sonde correspondant au marqueur CBF303_4832 a permis d'identifier 3 clones de Térèse (Térèse-16K02, Térèse-16O06 et Térèse-99I01) et un seul clone de Champagne (Champagne-45L05) sur lesquels le marqueur CBF303_4832 a été amplifié (Figure IV.6.a). Ces quatre clones ont été ensuite identifiés dans les deux banques BAC stockées à -80°C, pour être validés par amplification PCR en utilisant les amorces du marqueur CBF303_4832 et en ajoutant comme témoin l'ADN génomique de Champagne et Térèse. Les résultats de PCR ont permis de valider les trois clones Térèse-16K02, Térèse-99I01 et Champagne-45L05, étant donné que le Ct (*Cycle Threshold*) et la température de fusion (Tm) correspondaient bien à ceux de l'ADN témoin. Après amplification, les produits de PCR ont migré sur gel d'agarose pour s'assurer que la taille des amplifiâts des clones validés correspondait à celle de l'ADN des témoins (Figure IV.6.b). Au final pour le marqueur CBF303_4832, l'hybridation de la sonde sur le clone Térèse-16O06 s'est révélée aspécifique et donc inexploitable.

La Figure IV.7 montre un bilan des résultats du criblage des deux banques BAC (classique et en pool) pour les deux génotypes étudiés. Au moins un clone BAC a été criblé positivement pour chaque marqueur sur les banques BAC, révélant ainsi la présence de la séquence ciblée, sauf pour deux marqueurs : CBF211_3609 et CBF211_3610 sur le génotype Térèse. Pour remédier à ces clones manquants, la première idée a été d'augmenter la couverture génomique sur ce locus en exploitant la banque BAC en pool, afin d'augmenter les chances d'identifier des clones positifs. Malgré cela, les séries successives de criblages n'ont pas permis d'identifier de clones positifs. Ce problème pourrait être lié à un mésappariement entre les amorces des deux marqueurs et l'ADN cible qui pourrait causer des soucis de répétabilité de la PCR. En effet, les amorces ont été dessinées sur un autre génotype que Térèse et la présence de mutations

ponctuelles de type SNP pourrait causer un mésappariement de l'amorce. Pour remédier à cela, nous avons redéfini de nouvelles amorces en faisant varier les conditions, mais à nouveau nous n'avons pas réussi à identifier de clones positifs au sein de la banque Térèse pour ces deux marqueurs.



Figure IV. 6 : Résultat du criblage des banques BAC de Champagne et Térèse pour le marqueur CBF303 4832.

(a) Hybridation des macroarrays de Champagne et Térèse par la sonde de CBF303_4832 radiomarquée. Les clones positifs sont entourés en rouge. Chaque rond rouge entoure les deux copies d'un même clone (positif) sur lesquels s'est hybridée la sonde de CBF303_4832. (b) Résultat de validation des clones positifs identifiés en hybridation par PCR. Les amplifiâts des clones positifs ont migré sur gel d'agarose. (1) : clones validés, (2) : clone non validé, (3) : ADN génomique des témoins Champagne et Térèse, (4) : marqueur de taille.

Chapitre IV. Création de banques BAC pour la recherche de gènes candidats sous-jacents au QTL WFD6.1



Figure IV. 7 : Bilan des résultats du criblage des banques BAC (classiques et en pool) de Champagne et Térèse.

Les amorces des sondes utilisées pour le criblage (flèches oranges) ont été dessinées sur les séquences de gènes CBF repérées dans le génome de référence de la lignée Caméor (rectangles gris) grâce aux premières annotations de la carte physique de ce génotype (UMR Agroécologie, données non publiées). Ces gènes ont été identifiés par homologie avec les gènes candidats de Medicago truncatula sous-jacents au QTL de tolérance au gel Mt- FTQTL6, synténique du QTL WFD6.1 de *Pisum sativum*. Les clones Champagne-ng-16A14 et Champagne-ng-7O13 proviennent du criblage de la banque en pool de Champagne. Tous les clones BAC présents dans cette figure ont été validés par PCR.

3.4. Séquençage des clones identifiés par criblage des banques BAC

Les résultats d'assemblage et les critères de qualité des séquences PacBio® de 13 clones BAC sélectionnés sont présentés dans le tableau IV.6. Les résultats sont validés pour 12 des 13 clones assemblés présentant chacun une valeur de qualité (VQ) satisfaisante, une taille de contig cohérente avec la taille obtenue en estimation sur gel. De plus, les séquences assemblées ont été validées par alignement des séquences BES en extrémités des 12 contigs. Cependant, pour le clone BAC Térèse-117I04, nous avons obtenu deux contigs au lieu d'un contig unique. En alignant les séquences BES correspondant à ce clone, on note que la séquence BES-M13 s'aligne bien sur la fin du contig Térèse-117I04-Utg1. La séquence de l'amorce CBF426_6484_F du marqueur cible pour ce clone (CBF426_6484) s'aligne également sur le contig Térèse-117I04-Utg1 en position 82 kb. Par contre, la séquence BES-T7 n'a pas pu être alignée. Ceci pourrait être dû à un assemblage incomplet de quelques centaines de bases en fin d'assemblage. Cependant cette hypothèse est peu plausible puisque la somme des tailles des 2 contigs semble cohérente avec la taille estimée sur gel.

Tableau IV. 6 : Caractéristiques des données d'assemblage des séquences de la liste des clones BAC identifie	és
et validé par PCR	

Clone BAC	Couverture moyenne (X)	VQ*	Taille totale du contig (kb)	Estimation de la taille de l'insert après migration sur gel (kb)
Champagne-31I03	93	48,47	126	131
Champagne-45L05	127	48,69	109	108
Champagne-67L13	186	48,68	93	90
Champagne-74D03	101	48,66	138	137
Champagne-ng-16A14	224	48,72	97	96
Champagne-ng-7013	317	48,71	117	118
Champagne-91B08	73	48,74	130	132
Térèse-15I13	113	48,72	143	142
Térèse-16K02	200	48,68	147	144
Térèse-23A03	78	48,61	141	143
Térèse-80C08	158	48,71	153	149
Térèse-99I01	152	48,73	148	147
Térèse-117I04-Utg0	139	48,6	78	173
Térèse-117I04-Utg1	160	48,62	101	173

*Valeur de la qualité (Probabilité que le nucléotide soit correct : VQ40 = 99,99%, VQ50 = 99,99%) ; seuil CNRGV 48,0
Chapitre IV. Création de banques BAC pour la recherche de gènes candidats sous-jacents au QTL WFD6.1

Dans un second temps, pour comprendre la difficulté à obtenir un contig unique pour le clone BAC Térèse-117I04, l'alignement des deux contigs Térèse-117I04-Utg0 et Térèse-117I04-Utg1 a été analysé. Le résultat est que la séquence de Térèse-117I04-Utg1 s'aligne entre 1 à 7 kb sur la séquence de Térèse-117I04-Utg0 entre 71 à 78kb (99,6% d'identité). Donc, nous pouvons en déduire que le chevauchement n'a pas été validé en cours de l'assemblage, probablement à cause d'une zone répétée introduisant des erreurs sur cette région. En effet, l'alignement du contig Térèse-117I04-Utg1 sur lui-même (Figure IV.8) montre la présence d'une région répétée en début de contig susceptible d'introduire des erreurs dans l'assemblage.



Figure IV. 8 : Alignement du contig Térèse-117I04-Utg1 sur lui-même.

La construction du contig consensus, a été initiée par des expériences de contiguage PCR pour vérifier le chevauchement des clones BAC, notamment entre les marqueurs CBF303_4832 et CBF303_4834 et entre les marqueurs CBF211_3604 et CBF211_3605 pour les deux génotypes étudiés. Pour ce faire, des amorces ont été définies sur les extrémités des séquences BES des clones concernés. Après validation des PCR sur les ADN génomiques témoins de Champagne et Térèse, chaque couple d'amorces a été testé sur les clones d'intérêt par amplification PCR. Les résultats de ce test ont permis de mettre en évidence qu'il n'y a pas de chevauchement entre les clones porteurs de ces 2 gènes pour les deux génotypes étudiés. D'autre part, la continuité de la couverture du scaffold 211 pour le génotype Champagne, a été confirmée par ce test. En effet, selon les résultats de PCR le clone Champagne-74D03 chevauche bien les clones Champagne-139A24 et Champagne-ng-7O13.



Figure IV. 9 : Résultats des expériences de contiguage PCR pour vérifier le chevauchement des clones BAC entre les marqueurs CBF303_4832 et CBF303_4834 et entre les marqueurs CBF211_3604 et CBF211_3605 pour les deux génotypes.

Le chevauchement des clones BAC entre les marqueurs étudiés est présenté par un rond rouge.

4. Conclusion

Ce chapitre de thèse porte sur la construction de deux banques BAC dans l'objectif d'obtenir les séquences nucléotidiques du locus de tolérance au gel du groupe de liaison VI du pois (WFD6.1) en faisant l'hypothèse qu'il comportait des gènes *CBF* pour le génotype tolérant Champagne et le génotype sensible Térèse.

Nous avons criblé les deux banques BAC grâce à des marqueurs spécifiques du locus d'intérêt pour identifier des clones BAC portant les fragments de la région recherchée. Cependant, une partie de ce locus s'est avérée plus complexe à cribler en particulier pour le génotype Térèse pour deux marqueurs. Pour remédier à ce problème, nous avons essayé de construire une banque BAC en pool caractérisée par une plus grande couverture du génome. Or, même avec cette méthode complémentaire, n'avons pas réussi à identifier de clones correspondant à ces deux marqueurs pour la lignée Térèse. Par contre, nous avons réussi à vérifier le chevauchement de deux clones BAC qui n'avaient pas de marqueurs en commun. La perspective envisageable pour remédier à ce problème de criblage serait d'essayer de combler le gap avec de nouvelles séries de criblage de clones en suivant la stratégie de la marche chromosomique, c'est à dire à partir d'amorces dessinées sur les clones BAC bordant le gap.

Par ailleurs, les clones positifs sélectionnés ont été séquencés par la méthode de PacBio®, et les contigs obtenus seront assemblés pour obtenir une séquence consensus par génotype qui sera caractérisée dans des études ultérieures.

Conclusion générale

L'objectif général de ce travail était d'exploiter les nouvelles ressources génétiques et génomiques développées chez *Pisum sativum* pour approfondir l'étude du déterminisme génétique de la tolérance au gel. Pour ce faire, nous avons étudié la réponse au froid en nous appuyant sur deux méthodes d'analyses à savoir la recherche de QTL par analyse de liaison au sein d'une population biparentale de cartographie et la génétique d'association au sein d'une collection de diversité génétique, afin d'obtenir une vision à la fois complète et précise du déterminisme génétique de la réponse au froid. En complément, nous avons mené une étude de cartographie fine dans le but de réduire l'intervalle de confiance de l'un des principaux QTL de tolérance au gel localisé sur le groupe de liaison VI du pois (WFD6.1) puis de réaliser son clonage positionnel. Même si cette démarche n'a pas donné les résultats escomptés, nous avons pu initier la construction d'une carte physique locale du QTL WFD6.1 en nous basant sur les résultats de génétique d'association et sur la synténie avec *Medicago truncatula*. Les résultats de cette thèse ouvrent des perspectives de recherche pour continuer à préciser les déterminants génétiques de la tolérance au gel chez le pois mais donnent aussi de nouveaux outils pour la sélection.

Ce travail a tout d'abord permis une avancée significative dans l'identification des déterminants génétiques sous-jacents au QTL WFD6.1. Après une re-détection avec un plus grand nombre de marqueurs (chapitre II.A), nous avons entrepris une démarche de cartographie fine qui s'est traduite par une réduction de l'intervalle de confiance de 19,2 à 3,9 cM (chapitre II.B). Nous avons observé un faible taux de recombinaison au niveau de la région ciblée, vraisemblablement en lien avec la position du QTL à proximité du centromère, ce qui ne nous a pas permis de mener à bien la cartographie fine. L'approche de génétique d'association (chapitre III) nous a cependant permis d'identifier, parmi les marqueurs associés à ce locus, un marqueur correspondant à un gène CBF, représentant d'une famille de facteurs de transcription connus pour être impliqués dans la régulation de la réponse au froid. Sur la base de ce résultat et en utilisant les informations liées à la synténie avec Medicago truncatula (Mt) obtenues lors de travaux précédents (Tayeh et al. 2013a), nous avons entamé le clonage positionnel du QTL WFD6.1 par une approche de construction et de criblage de banques BAC pour un génotype tolérant au gel (Champagne) et un génotype sensible (Térèse) (chapitre IV). Les éléments obtenus dans le cadre de ma thèse vont permettre d'établir la séquence consensus de Champagne et Térèse pour ce locus sous réserve de réaliser un criblage complémentaire des banques BAC de Térèse pour obtenir la séquence complète de la zone étudiée. Les séquences des deux génotypes permettront d'identifier tous les gènes sous-jacents et plus particulièrement de comparer leurs contenus en gènes *CBF*, leurs nombres respectifs de copies et leur niveau d'homologie gène à gène entre les deux lignées.

La réduction des intervalles de confiance par re-détection et/ou génétique d'association et la production de lignées quasi-isogéniques aux QTL (QTL-NILs), réalisées dans le cadre de ma thèse, constituent également des étapes importantes pour le clonage positionnel d'autres zones du génome impliquées dans la tolérance au gel. C'est le cas pour les QTL WFD3.1 et WFD5.1, qui présentent en outre l'intérêt d'intervenir également dans le déterminisme des résistances à l'ascochytose et/ou à la pourriture racinaire liée à *Aphanomyces euteiches* (Prioul et al. 2004; Hamon et al. 2011; Desgroux et al. 2016). Pour le locus significatif détecté par génétique d'association sur le groupe de liaison VII, qui colocalise avec un QTL de résistance à *A. euteiches*, c'est réciproquement le clonage positionnel entrepris par l'unité IGEPP dans le cadre du projet PeaMUST qui apportera des éléments sur le contenu en gènes. Les résultats prometteurs obtenus récemment par l'approche de génétique d'association (Desgroux et al 2016, le présent travail) seront complétés prochainement par l'étude d'une population constituée dans le cadre de PeaMUST, génotypée pour environ 1 million de SNP et phénotypée vis à vis de différents stress biotiques et abiotiques.

Ce travail de thèse contribue enfin à compléter la panoplie des outils utilisables en sélection. Tout d'abord, l'effet des QTL WFD3.1, WFD5.1 et WFD 6.1, précédemment détectés dans de multiples environnements, a été validé grâce à l'évaluation au champ des lignées quasiisogéniques aux QTL (QTL-NILs) créées dans les fonds génétique Eden et Isard (chapitre II.B). Cette étape de validation confirme l'intérêt pour les sélectionneurs de pois d'hiver de mettre en œuvre une sélection assistée par marqueurs pour chacune de ces trois zones. Le pyramidage en cours des trois QTL dans le cultivar Eden devrait en outre permettre de tester la plus-value apportée par le cumul des trois zones, au moins dans un fond génétique "printemps". En second lieu, notre étude fournit un nombre important de marqueurs SNP localisés dans les intervalles de confiance des QTL re-détectés par analyse de liaison ou détectés par GWAS. Cette information est d'ores et déjà utilisée dans le projet PEAMAS pour caractériser des panels de diversité et du matériel en sélection chez les sélectionneurs français de pois d'hiver. Cette analyse donnera un aperçu des allèles favorables déjà présents dans le matériel en sélection, permettra de tester le polymorphisme SNP utilisable pour les programmes de sélection assistée par marqueurs et identifiera des géniteurs potentiels dans les collections. Du fait de la réduction des intervalles de confiance et de la caractérisation d'une importante collection ressources génétiques, une attention particulière pourra être portée aux haplotypes de marqueurs favorables définis par notre approche de génétique d'association et aux accessions représentatives de ces haplotypes au sein de la collection utilisée.

Références bibliographiques

Achard P, Gong F, Cheminant S, Alioua M, Hedden P, Genschik P (2008). The Cold-Inducible CBF1 Factor–Dependent Signaling Pathway Modulates the Accumulation of the Growth-Repressing DELLA Proteins via Its Effect on Gibberellin Metabolism. The Plant Cell. 20:2117-2129

Alonso-Blanco C, Gomez-Mena C, Llorente F, Koornneef M, Salinas J, Martínez-Zapater JM (2005). Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. Plant Physiology. 139:1304-1312

Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot A-L, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J, da Silva C, Moreau S, Gamas P, Wincker P, Gouzy J, Burstin J (2015). Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. The Plant Journal. 84:1-19

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013). MYB transcription factor genes as regulators for plant responses: an overview. Physiology and molecular biology of plants : an international journal of functional plant biology. 19:307-321

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J (2006). Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume Medicago truncatula. Theor Appl Genet. 112:1024-1041

Barrero-Gil J, Salinas J (2017). CBFs at the Crossroads of Plant Hormone Signaling in Cold Stress Response. Molecular Plant. 10:542-544

Bauchet G, Grenier S, Samson N, Bonnet J, Grivet L, Causse M (2017). Use of modern tomato breeding germplasm for deciphering the genetic control of agronomical traits by Genome Wide Association study. Theoretical Applied Genetics. 130:875-889

Benedict C, Skinner JS, Meng R, Chang Y, Bhalerao R, Huner NPA, Finn CE, Chen THH, Hurry V (2006). The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in Populus spp. J Plant Cell Environ. 29

Bhandari K, Nayyar H (2014) Low Temperature Stress in Plants: An Overview of Roles of Cryoprotectants in Defense. Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment, pp 193-265

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theoretical Applied Genetics. 113:206-224

Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T (2017). ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. 40:108-120

Bourion V, Lejeune-Hénaut I, Munier-Jolain N, Salon C (2003). Cold acclimation of winter and spring peas: carbon partitioning as affected by light intensity. European Journal of Agronomy. 19:535-548

Boutet G, Alves Carvalho S, Falque M, Peterlongo P, Lhuillier E, Bouchez O, Lavaud C, Pilet-Nayel M-L, Rivière N, Baranger A (2016). SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. BMC Genomics. 17:121

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009). The Genetic Architecture of Maize Flowering Time. Science. 325:714-718

Burstin J, Salloignon P, Chabert-Martinello M, Magnin-Robert J-B, Siol M, Jacquin F, Chauveau A, Pont C, Aubert G, Delaitre C, Truntzer C, Duc G (2015). Genetic diversity and trait genomic prediction in a pea diversity panel. BMC Genomics. 16:105

Cao PB, Azar S, SanClemente H, Mounet F, Dunand C, Marque G, Marque C, Teulières C (2015). Genome-wide analysis of the AP2/ERF family in Eucalyptus grandis: an intriguing over-representation of stress-responsive DREB1/CBF genes. PloS one. 10:e0121041-e0121041

Charmet G (2011) De la détection de QTL à la génétique d'association. Le Sélectionneur Français, p 29

Chawade A, Bräutigam M, Lindlöf A, Olsson O, Olsson B (2007). Putative cold acclimation pathways in Arabidopsis thaliana identified by a combined analysis of mRNA coexpression patterns, promoter motifs and transcription factors. BMC genomics. 8:304-304

Chen Y, Chen Z, Kang J, Kang D, Gu H, Qin G (2013). AtMYB14 Regulates Cold Tolerance in Arabidopsis. Plant molecular biology reporter. 31:87-97

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods. 10:563

Cieslarová J, Smýkal P, Dočkalová Z, Hanáček P, Procházka S, Hýbl M, Griga M (2011). Molecular evidence of genetic diversity changes in pea (*Pisum sativum* L.) germplasm after long-term maintenance. Genetic Resources Crop Evolution. 58:439-451

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica. 142:169-196

Coyne CJ, McClendon MT, Walling JG, Timmerman-Vaughan GM, Murray S, Meksem K, Lightfoot DA, Shultz JL, Keller KE, Martin RR, Inglis DA, Rajesh PN, McPhee KE, Weeden NF, Grusak MA, Li CM, Storlie EW (2007). Construction and characterization of

two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. Genome. 50:871-875

Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savois V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, Thompson R, Bendahmane A (2008). UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. Genome biology. 9:R43-R43

de Mendiburu F (2017) agricolae: Statistical Procedures for Agricultural Research. R package version 1.2.8. pp https://cran.r-project.org/web/packages/agricolae/index.html

Delbreil B (2018) Tolérance au froid du pois. Mémoire de HDR Ecole doctorale Sciences de la Matière, du Rayonnement et de l'Environnement Université de Lille, p 126

Desgroux A, L'Anthoëne V, Roux-Duparque M, Rivière JP, Aubert G, Tayeh N, Moussart A, Mangin P, Vetel P, Piriou C, McGee RJ, Coyne CJ, Burstin J, Baranger A, Manzanares-Dauleux M, Bourion V, Pilet-Nayel ML (2016). Genome-wide association mapping of partial resistance to Aphanomyces euteiches in pea. BMC Genomics. 17:124

Devaux R, Lejeune-Hénaut I (2010) Score card for frost Damage. Physiology oif the pea crop. Sciences Publishers, Enfield, USA, p 270:180

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant J-P, Pichon J-P, Pilet-Nayel M-L, Boutet G (2014). Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. BMC Genomics. 15:126

Dumont E, Bahrman N, Goulas E, Valot B, Sellier H, Hilbert JL, Vuylsteker C, Lejeune-Hénaut I, Delbreil B (2011). A proteomic approach to decipher chilling response from cold acclimation in pea (*Pisum sativum* L.). Plant Sci. 180:86-98

Dumont E, Fontaine V, Vuylsteker C, Sellier H, Bodèle S, Voedts N, Devaux R, Frise M, Avia K, Hilbert JL, Bahrman N, Hanocq E, Lejeune-Hénaut I, Delbreil B (2009). Association of sugar content QTL and PQL with physiological traits relevant to frost damage resistance in pea under field and controlled conditions. Theor Appl Genet. 118:1561-1571

Eggen A (2000). Cartographie fine d'un gène et clonage positionnel. Productions Animales, HS 2000:136

Eremina M, Rozhon W, Poppenberger B (2016a). Hormonal control of cold stress responses in plants. Cellular Molecular Life Sciences. 73:797-810

Eremina M, Unterholzner SJ, Rathnayake AI, Castellanos M, Khan M, Kugler KG, May ST, Mayer KFX, Rozhon W, Poppenberger B (2016b). Brassinosteroids participate in the control of basal and acquired freezing tolerance of plants. Proceedings of the National Academy of Sciences of the United States of America. 113:E5982-E5991

FAOSTAT (2018). FAOSTAT data base. Food and agriculture organisation of the United Nations - statistic

Ferreira A, Silva MFd, Silva LdCe, Cruz CD (2006). Estimating the effects of population size and type on the accuracy of genetic maps. Genetics and Molecular Biology. 29:187-192

Flint-Garcia SA, Thornsberry JM, IV ESB (2003). Structure of Linkage Disequilibrium in Plants. Annu Rev Plant Biol. 54:357-374

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998). Low temperature regulation of Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J. 16

Griffith M, Lumb C, Wiseman SB, Wisniewski M, Johnson RW, Marangoni AG (2005). Antifreeze proteins modify the freezing process in planta. Plant physiology. 138:330-340

Grimaud F, Renaut J, Dumont E, Sergeant K, Lucau-Danila A, Blervacq A-S, Sellier H, Bahrman N, Lejeune-Hénaut I, Delbreil B, Goulas E (2013). Exploring chloroplastic changes related to chilling and freezing tolerance during cold acclimation of pea (*Pisum sativum* L.). Journal of Proteomics. 80:145-159

Guy CL (2003). Freezing tolerance of plants: current understanding and selected emerging concepts. Canadian Journal of Botany. 81:1216-1223

Halliwell B (2006). Reactive Species and Antioxidants. Redox Biology Is a Fundamental Theme of Aerobic Life. Plant Physiology. 141:312-322

Hamon C, Baranger A, Coyne CJ, McGee RJ, Le Goff I, L'Anthoëne V, Esnault R, Rivière J-P, Klein A, Mangin P, McPhee KE, Roux-Duparque M, Porter L, Miteul H, Lesné A, Morin G, Onfroy C, Moussart A, Tivoli B, Delourme R, Pilet-Nayel M-L (2011). New consistent QTL in pea associated with partial resistance to Aphanomyces euteiches in multiple French and American environments. Theoretical Applied Genetics. 123:261-281

Hamon C, Coyne CJ, McGee RJ, Lesné A, Esnault R, Mangin P, Hervé M, Le Goff I, Deniot G, Roux-Duparque M, Morin G, McPhee KE, Delourme R, Baranger A, Pilet-Nayel M-L (2013). QTL meta-analysis provides a comprehensive view of loci controlling partial resistance to Aphanomyces euteichesin four sources of resistance in pea. J BMC Plant Biology. 13:45

Hascoët E, Jaminon O, Devaux C, Blassiau C, Bahrman N, Bochard A (2014) Toward fine mapping of frost tolerance QTL in pea. 2nd PeaMUST Annual Meeting, Dijon, France

Hedrick PW (1987). Gametic Disequilibrium Measures: Proceed With Caution. Genetics. 117:331

Hervé M (2014) RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-36. pp http://CRAN.R-project.org/package=RVAideMemoire

Hill WG, Robertson A (1968). Linkage disequilibrium in finite populations. Theoretical Applied Genetics. 38:226-231

Hu Y, Jiang Y, Han X, Wang H, Pan J, Yu D (2017). Jasmonate regulates leaf senescence and tolerance to cold stress: crosstalk with other phytohormones. Journal of Experimental Botany. 68:1361-1369

Huang XQ, Cloutier S, Lycar L, Radovanovic N, Humphreys DG, Noll JS, Somers DJ, Brown PDJT (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (Triticum aestivum L.). Theoretical Applied Genetics. 113:753-766

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. . 280:104-106

Janská A, Maršík P, Zelenková S, Ovesná J (2010). Cold stress and acclimation – what is important for metabolic adjustment? 12:395-405

Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics. 11:94

Kazemi-Shahandashti S-S, Maali-Amiri R (2018). Global insights of protein responses to cold stress in plants: Signaling, defence, and degradation. Journal of Plant Physiology. 226:123-135

Klein A, Houtin H, Rond C, Marget P, Jacquin F, Boucherot K, Huart M, Rivière N, Boutet G, Lejeune-Hénaut I, Burstin J (2014). QTL analysis of frost damage in pea suggests different mechanisms involved in frost tolerance. Theoretical Applied Genetics. 127:1319-1330

Kosová K, Vítámvás P, Prášil IT (2007). The role of dehydrins in plant response to cold. Biologia Plantarum. 51:601-617

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009). A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. PLOS Genetics. 5:e1000551

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics. 43:163

Lavaud C, Lesné A, Piriou C, Le Roy G, Boutet G, Moussart A, Poncet C, Delourme R, Baranger A, Pilet-Nayel M-L (2015). Validation of QTL for resistance to Aphanomyces euteiches in different pea genetic backgrounds using near-isogenic lines. Theoretical Applied Genetics. 128:2273-2288

Legrand S, Marque G, Blassiau C, Bluteau A, Canoy AS, Fontaine V, Jaminon O, Bahrman N, Mautord J, Morin J, Petit A, Baranger A, Rivière N, Wilmer J, Delbreil B, Lejeune-Hénaut I (2013). Combining gene expression and genetic analyses to identify candidate genes involved in cold responses in pea. J Plant Physiol. 170:1148-1157 Lejeune-Hénaut I, Bourion V, Etévé G, Cunot E, Delhaye K, Desmyter C (1999). Floral initiation in field-grown forage peas is delayed to a greater extent by short photoperiods, than in other types of European varieties. Euphytica. 109:201-211

Lejeune-Hénaut I, Hanocq E, Béthencourt L, Fontaine V, Delbreil B, Morin J, Petit A, Devaux R, Boilleau M, Stempniak J-J, Thomas M, Lainé A-L, Foucher F, Baranger A, Burstin J, Rameau C, Giauffret C (2008). The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. Theoretical Applied Genetics. 116:1105-1116

Leonforte A, Sudheesh S, Cogan N, Salisbury P, E Nicolas M, Forster J, Kaur S (2014) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.)

Li H, Ye K, Shi Y, Cheng J, Zhang X, Yang S (2017). BZR1 Positively Regulates Freezing Tolerance via CBF-Dependent and CBF-Independent Pathways in Arabidopsis. Molecular Plant. 10:545-559

Li Z, Mu P, Li C, Zhang H, Li Z, Gao Y, Wang X (2005). QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. Theoretical Applied Genetics. 110:1244-1252

Limborg MT, McKinney GJ, Seeb LW, Seeb JE (2016). Recombination patterns reveal information about centromere location on linkage maps. Mol Ecol Resour. 16:655-661

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. J Plant Cell. 10

Lucau-Danila A, Toitot C, Goulas E, Blervacq AS, Hot D, Bahrman N, Sellier H, Lejeune-Hénaut I, Delbreil B (2012). Transcriptome analysis in pea allows to distinguish chilling and acclimation mechanisms. Plant Physiology and Biochemistry. 58:236-244

Miller AK, Galiba G, Dubcovsky J (2006). A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-Am2 in Triticum monococcum. Mol Genet Genomics. 275:193-203

Nguyen HC, Cao PB, San Clemente H, Ployet R, Mounet F, Ladouce N, Harvengt L, Marque C, Teulieres C (2017). Special trends in CBF and DREB2 groups in Eucalyptus gunnii vs Eucalyptus grandis suggest that CBF are master players in the trade-off between growth and stress resistance. 159:445-467

Page D, Aubert G, Duc G, Welham T, Domoney C (2002). Combinatorial variation in coding and promoter sequences of genes at the Tri locus in *Pisum sativum* accounts for variation in trypsin inhibitor activity in seeds. Molecular Genetics Genomics. 267:359-369

Pascual L, Albert E, Sauvage C, Duangjit J, Bouchet J-P, Bitton F, Desplat N, Brunel D, Le Paslier M-C, Ranc N, Bruguier L, Chauchard B, Verschave P, Causse M (2016).

Dissecting quantitative trait variation in the resequencing era: complementarity of bi-parental, multi-parental and association panels. Plant Science. 242:120-130

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 38:904

Prioul S, Frankewitz A, Deniot G, Morin G, Baranger A (2004). Mapping of quantitative trait loci for partial resistance to Mycosphaerella pinodes in pea (*Pisum sativum* L.), at the seedling and adult plant stages. Theoretical Applied Genetics. 108:1322-1334

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics. 155:945-959

R Core Team (2014) R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, pp http://www.R-project.org

Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology. 5:94-100

Rafalski A, Morgante M (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends in Genetics. 20:103-111

Rihan HZ, Al-Issawi M, Fuller MP (2017). Advances in physiological and molecular aspects of plant cold tolerance. Journal of Plant Interactions. 12:143-157

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, Dumas F, Altmann T, Brunel D, Ouzunova M, Flament P, Dubreuil P, Charcosset A, Mary-Huard T (2014). Recovering Power in Association Mapping Panels with Variable Levels of Linkage Disequilibrium. Genetics. 197:375-387

Sakuma Y, Qiang L, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun. 290

SAS (1999). SAS/STAT User's Guide, version 8. SAS Institute Inc., Cary, NC, 3884 p

Sharma M, Laxmi A (2016). Jasmonates: Emerging Players in Controlling Temperature Stress Tolerance. Frontiers in plant science. 6:1129-1129

Sharma P, Sharma N, Deswal R (2005). The molecular biology of the low-temperature response in plants. Bioessays. 27:1048-1059

Shi Y, Ding Y, Yang S (2015). Cold Signal Transduction and its Interplay with Phytohormones During Cold Acclimation. Plant and Cell Physiology. 56:7-15

Shi Y, Tian S, Hou L, Huang X, Zhang X, Guo H, Yang S (2012). Ethylene signaling negatively regulates freezing tolerance by repressing expression of CBF and type-A ARR genes in Arabidopsis. The Plant Cell. 24:2578-2595

Shinwari ZK, Nakashima K, Miura S, Kasuga M, Seki M, Yamaguchi-Shinozaki K, Shinozaki K (1998). AnArabidopsisGene Family Encoding DRE/CRT Binding Proteins Involved in Low-Temperature-Responsive Gene Expression. Biochemical and Biophysical Research Communications. 250:161-170

Sindhu A, Ramsay L, Sanderson L-A, Stonehouse R, Li R, Condie J, Shunmugam ASK, Liu Y, Jha AB, Diapari M, Burstin J, Aubert G, Tar'an B, Bett KE, Warkentin TD, Sharpe AG (2014). Gene-based SNP discovery and genetic mapping in pea. Theoretical Applied Genetics. 127:2225-2241

Skinner JS, Szűcs P, von Zitzewitz J, Marquez-Cedillo L, Filichkin T, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2006). Mapping of barley homologs to genes that regulate low temperature tolerance in Arabidopsis. Theoretical and Applied Genetics. 112:832-842

Smýkal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, Ellis NTH (2011). Phylogeny, phylogeography and genetic diversity of the Pisum genus. Plant Genetic Resources. 9:4-18

Stockinger EJ, Gilmour SJ, Thomashow MF (1997). Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C repeat/DRE, a cisacting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci U S A. 94

Švábová L, Griga M (2008). The effect of cocultivation treatments on transformation efficiency in pea (*Pisum sativum* L.). Plant Cell, Tissue Organ Culture. 95:293-304

Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW, Holland JB (2007). QTL mapping with near-isogenic lines in maize. Theoretical Applied Genetics. 114:1211-1228

Tayeh N, Aluome C, Falque M, Jacquin F, Klein A, Chauveau A, Bérard A, Houtin H, Rond C, Kreplak J, Boucherot K, Martin C, Baranger A, Pilet-Nayel M-L, Warkentin TD, Brunel D, Marget P, Le Paslier M-C, Aubert G, Burstin J (2015a). Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, highresolution consensus genetic map. Plant J. 84:1257-1273

Tayeh N, Aubert G, Pilet-Nayel M-L, Lejeune-Hénaut I, Warkentin TD, Burstin J (2015b). Genomic Tools in Pea Breeding Programs: Status and Perspectives. 6

Tayeh N, Bahrman N, Devaux R, Bluteau A, Prosperi JM, Delbreil B, Lejeune-Hénaut I (2013a). A high-density genetic map of the Medicago truncatula major freezing tolerance QTL on chromosome 6 reveals colinearity with a QTL related to freezing damage on *Pisum sativum l*inkage group VI. Mol Breeding. 32

Tayeh N, Bahrman N, Sellier H, Bluteau A, Blassiau C, Fourment J, Bellec A, Debellé F, Lejeune-Hénaut I, Delbreil B (2013b). A tandem array of CBF/DREB1 genes is located in a major freezing tolerance QTL region on Medicago truncatula chromosome 6. BMC Genomics. 14:814

Thomashow MF (1998). Role of Cold-Responsive Genes in Plant Freezing Tolerance. Plant Physiology. 118:1-8

Thomashow MF (2010). Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiol. 154

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler Iv ES (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics. 28:286

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics. 43:159

Tondelli A, Francia E, Barabaschi D, Pasquariello M, Pecchioni N (2011). Inside the CBF locus in Poaceae. Plant Science. 180:39-45

Voorrips RE (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. Journal of Heredity. 93:77-78

Wang S, Basten CJ, Zeng ZB (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh

Wang W, Vinocur B, Altman A (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta. 218:1-14

Wang Y, Wang Q, Liu M, Bo C, Wang X, Ma Q, Cheng B, Cai R (2017). Overexpression of a maize MYB48 gene confers drought tolerance in transgenic arabidopsis plants. Journal of Plant Biology. 60:612-621

Warkentin TD, Smýkal P, Coyne CJ, Weeden N, Domoney C, Bing D-J, Leonforte A, Xuxiao Z, Dixit GP, Boros L, McPhee KE, McGee RJ, Burstin J, Ellis THN (2015) Pea. In: AM DR (ed) Grain Legumes. Springer-Verlag New York

Weller JL, Liew LC, Hecht VFG, Rajandran V, Laurie RE, Ridge S, Wenden B, Vander Schoor JK, Jaminon O, Blassiau C, Dalmais M, Rameau C, Bendahmane A, Macknight RC, Lejeune-Hénaut I (2012). A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proceedings of the National Academy of Sciences. 109:21158-21163

Wingler A (2015). Comparison of signaling interactions determining annual and perennial plant growth in response to low temperature. 5

Xiong L, Lee H, Ishitani M, Tanaka Y, Stevenson B, Koiwa H, Bressan RA, Hasegawa PM, Zhu J-K (2002). Repression of stress-responsive genes by FIERY2, a novel transcriptional regulator in Arabidopsis. Proceedings of the National Academy of Sciences. 99:10899-10904

Yadav S (2010). Cold stress tolerance mechanisms in plants. A review. Agronomy for Sustainable Development. 30:515-527

Yu J, Buckler ES (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology. 17:155-160

Yu J, Holland JB, McMullen MD, Buckler ES (2008). Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics. 178:539-551

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics. 38:203

Zhu C, Gore M, Buckler E, Yu J (2008) Status and Prospects of Association Mapping in Plants

Zhu C, Yu J (2009). Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics. 182:875-888

Annexes

Annexe A : Additional files du chapitre III

Additional file 1 : Table S1. Description of the association mapping collection
Additional file 2: Figure S1. Distribution of Best Linear Unbiased Prediction (BLUP) values for the five traits observed within the pea collection
Additional file 3: Figure S2. Distribution of 10739 SNPs along the <i>Pisum sativum</i> linkage groups
Additional file 4: Table S2. Description of the <i>Pisum sativum</i> linkage groups (LGs) used in the present study
Additional file 5: Figure S3. Distribution of minor allele frequencies (MAF) for 10739 SNP markers within the 363 pea accessions
Additional file 6: Figure S4. Scatterplot showing the linkage disequilibrium (LD) decay estimated in the association mapping collection
Additional file 7: Figure S5. Population structure of the pea association mapping collection based on Discriminant Analysis of Principal Components (DAPC) analysis
Additional file 8: Figure S6. Distribution of the kinship coefficients between accessions of the association mapping collection. 162
Additional file 9: Table S3. Description of linkage disequilibrium (LD) blocks per linkage group in the association mapping collection. 163
Additional file 10: Table S4. Marker haplotype analysis of the association mapping collection
Additional file 11: Table S5. Typical accesssions carrying favourable and unfavourable haplotypes
Additional file 12: Table S6. Candidate genes annotated for markers significantly detected by GWAS and markers in linkage disequilibrium (LD; $r^2 > 0.8$) with the former ones
Additional file 13: Table S7. The correspondence of the genotyping results of accessions at the Hr locus (Elf3 SNP marker)
Additional file 14: Table S8. Description of accessions cumulating 3 to 5 favourable haplotypes vs accessions cumulating 6 unfavourable haplotypes at the frost damage (FD)-associated loci of the GWA study

Additional file 1 : Table S1. Description of the association mapping collection.

This table presents the list of the pea accessions composing the association mapping collection with their end-use, cultivation status, geographical origin and sowing type. The 'cluster' column shows at which cluster each of the 363 pea accessions has been attributed on the basis of the discriminant analysis of principal components (DAPC). The description of the pea accessions is extracted from Burstin et al. [39].

CRB CodeAccession Name / Donor CodeEnd		End-use	Cultivation status	Geographical origin	Sowing type	Cluster
CRB001	ABAQUE	field	cultivar	North-Europe	spring	4
CRB002	BIRTE	field	cultivar	North-Europe	spring	4
CRB003	CHINE-D367	Unknown	local population	Far-East	Unknown	6
CRB004	CHINE-D368	Unknown	local population	Far-East	Unknown	6
CRB005	CHINE-D371	Unknown	local population	Far-East	Unknown	1
CRB006	DOVE	field	cultivar	North-Europe	winter	5
CRB007	FELTHAM FIRST	garden	cultivar	North-Europe	spring	3
CRB009	KOROZA	field	cultivar	North-Europe	spring	1
CRB010	LUMINA	field	cultivar	North-Europe	spring	4
CRB011	KATHMANDOU DISTRICT 1	Unknown	local population	Far-East	Unknown	6
CRB012	KATHMANDOU DISTRICT 2	Unknown	local population	Far-East	Unknown	4
CRB014	WBH 1458	Unknown	cultivar	North-Europe	Unknown	1
CRB015	Pisum sativum-AFGHANISTAN JI86	Unknown	wild	Middle-East	Unknown	2
CRB016	Pisum sativum-BOLIVIA JI228	Unknown	wild	South- America	Unknown	6
CRB017	Pisum sativum-GREECE JI267	Unknown	local population	South-Europe	winter	7
CRB018	ALASKA	garden	cultivar	North- America	spring	6
CRB019	WISCONSIN PERFECTION	Unknown	cultivar	North- America	Unknown	3
CRB020	Pisum sativum PLP 70 JI1267	Unknown	local population	Far-East	Unknown	2
CRB023	PESOL JI1831	Unknown	local population	South-Europe	Unknown	3
CRB024	Pisum sativum-MEXICO JI1844	Unknown	local population	South- America	winter	7
CRB025	Pisum sativum-USSR JI2200	Unknown	local population	North-Europe	Unknown	1
CRB026	Pisum sativum-ZAIRE JI2376	Unknown	wild	Africa	Unknown	6
CRB027	Pisum sativum-ZAMBIA JI2383	Unknown	local population	Africa	Unknown	6
CRB029	USATYJ JI113	Unknown	germplasm	North-Europe	Unknown	3
CRB031	Pisum elatius JI1075	Unknown	wild	Middle-East	winter	2
CRB034	Pisum elatius JI1703	Unknown	wild	Unknown	Unknown	1
CRB035	Pisum humile JI1794	Unknown	wild	Middle-East	Unknown	2
CRB039	ATAR JI2202	Unknown	local population	Middle-East	Unknown	2
CRB040	WILD TUNESIAN JI2263	Unknown	local population	Africa	Unknown	2
CRB041	WT304 JI2473	Unknown	wild	Middle-East	Unknown	2
CRB044	Pisum fulvum JI2523	Unknown	wild	Middle-East	Unknown	2
CRB047	Pisum transcaucasicum JI2546	Unknown	wild	South-Europe	Unknown	2

CRB048	Pisum speciosum-LIBYA JI2605	Unknown	wild	Africa	Unknown	2
CRB053	86-22	field	breeding	North-Europe	winter	5
CRB054	87-317	field	breeding	North-Europe	winter	5
CRB055	87-500	field	breeding	North-Europe	winter	5
CRB056	ASSAS	fodder	cultivar	North-Europe	winter	7
CRB057	AW2	fodder	local population	North-Europe	winter	7
CRB058	BLIXT 7	fodder	breeding	North-Europe	winter	7
CRB059	BLIXT 104	fodder	breeding	North-Europe	winter	7
CRB060	BLIXT 109	fodder	breeding	North-Europe	winter	7
CRB061	BLIXT157	fodder	Unknown	Unknown	winter	7
CRB062	BLIXT 484	fodder	breeding	North-Europe	winter	7
CRB063	BOOSTER	field	cultivar	North-Europe	winter	1
CRB064	BOOSTER*AW7	field	breeding	North-Europe	winter	7
CRB065	BRIDGE	field	cultivar	North-Europe	spring	4
CRB066	BUL179	garden	breeding	North-Europe	spring	3
CRB067	CE*CF	fodder	breeding	North-Europe	winter	7
CRB068	COTE D'OR	fodder	local population	North-Europe	winter	7
CRB069	DOLMEN	field	cultivar	North-Europe	winter	4
CRB070	GA91H2	field	breeding	Unknown	winter	4
CRB071	CHEYENNE	field	breeding line	Unknown	winter	4
CRB072	MUENCHENER BANATER WINTERERBSE	fodder	Unknown	North-Europe	winter	7
CRB073	X78122-sbm-SCOUT	preserve	germplasm	North- America	spring	3
CRB074	ORANGE POD-orp	fodder	germplasm	Unknown	Unknown	4
CRB075	ARTHRITIC-art JI2122	Unknown	germplasm	Unknown	Unknown	4
CRB076	SENTINEL	fodder	cultivar	North-Europe	spring	3
CRB077	MAGNUS	fodder	cultivar	North-Europe	spring	6
CRB078	CELESTE	field	cultivar	North-Europe	spring	4
CRB081	GSP981	Unknown	local population	Unknown	Unknown	-
CRB082	GSP1013	preserve	breeding	North- America	Unknown	3
CRB084	GSP1016	preserve	breeding	North- America	Unknown	3
CRB085	GSP1017	preserve	breeding	America	Unknown	3
CRB086	GSP1023	preserve	breeding	America	Unknown	3
CRB087	GSP1078	preserve	breeding	North- America	Unknown	3
CRB088	MN314	preserve	breeding	North- America	Unknown	3
CRB089	GSP1091	preserve	breeding	North- America	Unknown	3
CRB090	GSP1097	preserve	breeding	America	Unknown	3
CRB091	HOLLY 2	Unknown	Unknown	Unknown	Unknown	7
CRB092	HOLLY 3	fodder	breeding	North-Europe	winter	7
CRB093	HOLLY 4	fodder	breeding	North-Europe	winter	7
CRB094	HOLLY 5	fodder	breeding	North-Europe	winter	7
CRB095	HOLLY 6	fodder	breeding	North-Europe	winter	7
CRB096	HOLLY 7	Unknown	Unknown	Unknown	winter	7
CRB097	HOLLY 8	fodder	breeding	North-Europe	winter	7
CRB098	HOLLY 9	fodder	breeding	North-Europe	winter	7

CRB099	HOLLY 13	fodder	breeding	North-Europe	winter	1
CRB100	HOLLY 17	fodder	breeding	North-Europe	winter	7
CRB101	HOLLY 18	fodder	breeding	North-Europe	winter	7
CRB102	ICEBERG	field	cultivar	North-Europe	winter	5
CRB103	MISTRAL	field	cultivar	North-Europe	winter	5
CRB104	MONITOR	field	cultivar	North-Europe	winter	5
CRB105	NEVE	field	cultivar	North-Europe	winter	7
CRB106	NEVE*b195-5	field	breeding	North-Europe	winter	7
CRB107	PICAR	fodder	local population	North-Europe	winter	7
CRB108	PURSAN	field	cultivar	North-Europe	spring	4
CRB109	RITMO	garden	cultivar	North-Europe	spring	4
CRB110	SYDNEY	field	cultivar	North-Europe	spring	4
CRB111	TOPOLIANSKI	fodder	local population	Unknown	winter	7
CRB112	VERTOP	field	cultivar	North-Europe	Unknown	4
CRB113	WNC 23Z	fodder	local population	Unknown	winter	1
CRB114	WNC Z61Z	fodder	local population	Unknown	winter	7
CRB115	XSARA	field	cultivar	North-Europe	spring	4
CRB116	US IDAHO W6 8955	Unknown	cultivar	North-	Unknown	4
CRB117	RAN SEKEREC	Unknown	cultivar	South-Europe	Unknown	6
CRB118	MN108	Unknown	cultivar	North- America	Unknown	3
CRB119	90-2079	garden	breeding	North- America	spring	3
CRB120	K4797	Unknown	Unknown	North-Europe	Unknown	6
CRB121	79-2022	garden	cultivar	North- America	spring	1
CRB122	Aa86	Unknown	breeding	North-Europe	Unknown	1
CRB123	PI271936	Unknown	Unknown	North-Europe	Unknown	1
CRB124	PI273279	Unknown	Unknown	South- America	Unknown	1
CRB125	FENN	fodder	breeding	North- America	Unknown	7
CRB126	PI512066	Unknown	breeding	North- America	Unknown	7
CRB127	CAPELLA	Unknown	cultivar	North-Europe	Unknown	1
CRB128	HOHENHEIMER PINK- FLOWERED PI180693	fodder	Unknown	North-Europe	Unknown	1
CRB129	MN313	Unknown	breeding	North- America	Unknown	3
CRB130	552	garden	breeding	North- America	spring	3
CRB131	PH14-119	preserve	cultivar	North- America	Unknown	3
CRB132	74SN4	preserve	breeding	North- America	Unknown	3
CRB133	74SN5	preserve	breeding	North- America	Unknown	1
CRB134	DARK SKIN PERFECTION	garden	cultivar	North-Europe	spring	3
CRB135	90-2131	garden	breeding	North- America	spring	3
CRB136	SHRAT	Unknown	cultivar	Far-East	Unknown	2
CRB137	BEZELYA	Unknown	cultivar	Middle-East	Unknown	2
CRB138	PI196023	Unknown	cultivar	Africa	Unknown	6
CRB139	PI196033	Unknown	cultivar	Africa	Unknown	6
CRB140	PI358630	Unknown	wild	Africa	Unknown	6

CRB141	PI471171	Unknown	Unknown	Far-East	Unknown	3
CRB142	P58	Unknown	Unknown	Unknown	Unknown	5
CRB143	0-2016	garden	breeding	North- America	spring	3
CRB144	0-2042	garden	breeding	North- America	spring	3
CRB145	0-2072	garden	breeding	North- America	spring	1
CRB146	0-2073	garden	breeding	North- America	spring	3
CRB147	0-2093	garden	breeding	North- America	spring	1
CRB148	PI179461	Unknown	cultivar	Middle-East	Unknown	1
CRB149	PI193841	Unknown	cultivar	Africa	Unknown	6
CRB150	PI193846	Unknown	cultivar	Africa	Unknown	6
CRB151	PI212112	Unknown	wild	Middle-East	Unknown	2
CRB152	MOSHONG 3	Unknown	cultivar	Middle-East	Unknown	2
CRB153	GRANT	garden	cultivar	North-Europe	spring	3
CRB154	SUNDANCE	garden	cultivar	North-Europe	spring	3
CRB155	MINI 93	Unknown	cultivar	North- America	Unknown	3
CRB157	WSU 23	Unknown	cultivar	North- America	Unknown	3
CRB158	WSU 28	Unknown	cultivar	North- America	Unknown	3
CRB159	VANTAGE	Unknown	cultivar	Unknown	Unknown	3
CRB160	21	garden	breeding	North-Europe	winter	5
CRB161	34R	garden	breeding	North-Europe	winter	5
CRB162	100	garden	breeding	North-Europe	winter	5
CRB163	194	field	breeding	North-Europe	winter	5
CRB165	418	field	breeding	North-Europe	winter	5
CRB166	5031	field	breeding	North-Europe	winter	5
CRB168	5085	field	breeding	North-Europe	winter	5
CRB169	5099	field	breeding	North-Europe	winter	5
CRB170	AA1	Unknown	breeding	North-Europe	spring	6
CRB171	eM	garden	breeding	North-Europe	spring	6
CRB172	aC	Unknown	breeding	North-Europe	spring	6
CRB174	533	garden	breeding	North-Europe	spring	6
CRB175	585	Unknown	breeding	North-Europe	spring	4
CRB179	AUSTRALIA	Unknown	cultivar	Unknown	Unknown	1
CRB180	ENGLISH	fodder	Unknown	North-Europe	Unknown	1
CRB181	IREGI SARGA	fodder	cultivar	North-Europe	winter	6
CRB182	MOROZOUSTOICHIVYL	Unknown	Unknown	North-Europe	winter	7
CRB184	OSZI TKB	fodder	cultivar	North-Europe	winter	1
CRB185	PRECOCE DE NAPLES	garden	cultivar	South-Europe	spring	6
CRB186	ROSAKRONE	fodder	cultivar	North-Europe	spring	1
CRB188	TASMANIE	Unknown	Unknown	Unknown	Unknown	1
CRB189	MAZARE	Unknown	Unknown	South-Europe	Unknown	6
CRB190	KIRIN 40	mangetout	cultivar	Far-East	spring	3
CRB191	POL 6806	Unknown	Unknown	Unknown	Unknown	1
CRB192	BLACK SEEDED	fodder	Unknown	Unknown	winter	7
CRB193	CUZCO 1	garden	local	South- America	Unknown	6
CRB194	BEAN PEA	Unknown	cultivar	Unknown	Unknown	1
CRB195	CCO 2	Unknown	Unknown	South-Europe	Unknown	3
CRB196	POIS AMERICAIN COUSIN	Unknown	breeding	North- America	Unknown	3

I	CRB198	AGAT	Unknown	cultivar	North-Europe	Unknown	4
	CRB200	HAITI COLORE	fodder	local population	South- America	Unknown	1
	CRB201	AFGHANISTAN ASIATICUM	fodder	wild	Middle-East	Unknown	2
	CRB202	NEPAL A	fodder	wild	Far-East	Unknown	2
	CRB203	BRAUN YEMEN	fodder	Unknown	Middle-East	Unknown	2
	CRB204	ABYSSINICUM VAVILOVANIUM	fodder	wild	Unknown	Unknown	2
	CRB205	CAPSICUM	fodder	Unknown	Middle-East	Unknown	1
	CRB206	MEXIC 1	Unknown	Unknown	South- America	Unknown	6
	CRB207	POIS IRANIEN 57 B	Unknown	Unknown	Middle-East	Unknown	1
	CRB209	CEYLAN	Unknown	Unknown	Unknown	Unknown	3
	CRB210	BUTARO 1A	Unknown	Unknown	Unknown	Unknown	6
	CRB212	CLAVIER 2	garden	breeding	Unknown	spring	1
	CRB213	KATMANDOU	Unknown	local	Far-Fast	Unknown	6
	CDD214		c inkilo wii	population	Tul Eust		6
	CRB214	CAROUBEL	mangetou	cultivar	Unknown	spring	6
	CRB216	NORMAND	mangetout	cultivar	North-Europe	spring	6
	CRB218	SUGAR SNAP	mangetout	cultivar	North- America	spring	3
	CRB219	ALGERA	fodder	cultivar	Africa	Unknown	1
	CRB220	BLAUWSCHOKKER	fodder	cultivar	North-Europe	spring	6
	CRB222	JAYGEE	fodder	cultivar	North-Europe	spring	4
	CRB223	TRIFFID	fodder	cultivar	Unknown	spring	3
	CRB228	R 4-53 DJNOVIC	Unknown	breeding	South-Europe	Unknown	3
	CRB229	TRAPPER	garden	cultivar	North- America	spring	3
	CRB231	ERSLING	garden	cultivar	North-Europe	spring	6
	CRB232	FINLANDE	garden	Unknown	Unknown	spring	6
	CRB233	AUTOMOBILE	garden	cultivar	North-Europe	spring	3
	CRB234	TEMPO	garden	cultivar	North-Europe	spring	3
	CRB236	KOMPACTA	garden	cultivar	South-Europe	spring	3
	CRB239	PYGMA	garden	cultivar	North-Europe	spring	6
	CRB240	CLAUSE H4	fodder	Unknown	Unknown	spring	7
	CRB241	SOUCHET	garden	cultivar	North-Europe	spring	6
	CRB244	SINIKKA	garden	cultivar	North-Europe	spring	1
	CRB245	HATIVER	garden	cultivar	North-Europe	winter	5
	CRB246	WICO	garden	cultivar	Unknown	spring	6
	CRB247	LAFONTAINE	garden	cultivar	North-Europe	spring	3
	CRB248	LE DELICIEUX	garden	cultivar	North-Europe	spring	3
	CRB249	TELEPHONE A RAMES	garden	cultivar	North-Europe	spring	3
	CRB250	FIN DE LA BIEVRE	garden	cultivar	North-Europe	spring	6
	CRB252	MERVEILLE D'ETAMPES (syn=SERPETTE GUILLOTEAUX)	garden	cultivar	North-Europe	spring	6
ļ	CRB253	TIMO	fodder	cultivar	North-Europe	spring	1
	CRB254	TRIANON	garden	cultivar	North-Europe	spring	3
ļ	CRB256	ROSAMU	garden	cultivar	North-Europe	spring	6
ļ	CRB258	RODOGUNE	garden	cultivar	North-Europe	spring	6
	CRB259	BURPEANA	garden	cultivar	Unknown	spring	3
	CRB262	NFG KRUPP PELUSCHKE	fodder	cultivar	North-Europe	spring	1
ļ	CRB263	URSS 2759	fodder	Unknown	North-Europe	Unknown	2
ļ	CRB265	PIVER	fodder	cultivar	Unknown	winter	7
	CRB267	SOLIDO	fodder	cultivar	North-Europe	Unknown	4
ļ	CRB268	GLOIRE DE CORREZE	garden	cultivar	North-Europe	spring	6
٠							•

CRB270	PETIT PROVENCAL	garden	cultivar	North-Europe	spring	3
CRB271	NORLI	mangetout	cultivar	North-Europe	spring	6
CRB272	DESIREE	fodder	cultivar	North-Europe	spring	6
CRB273	LIVIOLETTA	fodder	cultivar	Unknown	spring	1
CRB274	SETCHEY	fodder	cultivar	Unknown	Unknown	1
CRB276	CHIPEAU	garden	cultivar	North-Europe	spring	3
CRB277	REMINI	garden	cultivar	North-	spring	3
CRB278	DOORDRAGER	garden	cultivar	North-Europe	spring	3
CRB279	ETA	garden	cultivar	North-Europe	spring	3
CRB283	BINGEFORS	fodder	cultivar	North-Europe	Unknown	6
CRB285	MIRAVIL	garden	cultivar	North-Europe	spring	3
CRB287	MINGOMARK	garden	cultivar	North-Europe	spring	3
CRB289	MERVEILLE DE WEISSENFELS	garden	cultivar	North-Europe	snring	3
CRB290	MIDFREEZER	garden	cultivar	North-	spring	3
		C		America North-	1 0	
CRB291	MULTIRESISTANT	garden	breeding	America	Unknown	3
CRB293	st100	Unknown	breeding	Unknown	Unknown	3
CRB294	st+af+tl	Unknown	breeding	Unknown	Unknown	3
CRB295	tl100	Unknown	breeding	Unknown	Unknown	6
CRB296	667	garden	breeding	North-Europe	spring	6
CRB297	776	field	breeding	North-Europe	spring	4
CRB299	AMAC	field	cultivar	North-Europe	winter	7
CRB300	AMINO	field	cultivar	North-Europe	spring	6
CRB301	FIGARO	field	breeding	North-Europe	spring	4
CRB303	CADOR	garden	cultivar	North-Europe	spring	6
CRB304	VAVILOV-D265	fodder	local population	North-Europe	spring	1
CRB305	PRECLAMEX	garden	cultivar	North-Europe	spring	6
CRB306	CAMÉOR	garden	cultivar	North-Europe	spring	3
CRB307	BALLET	field	cultivar	North-Europe	spring	4
CRB308	SKOROSPELYI GIBRID N92 K4819	field	local population	North-Europe	spring	1
CRB309	CERISE-ce,CR	fodder	germplasm	North-Europe	spring	1
CRB310	CHINA	field	local population	Far-East	winter	2
CRB311	SOMMETTE	garden	cultivar	North-Europe	spring	6
CRB312	FINETTE	garden	cultivar	North-Europe	spring	4
CRB313	ULADOVSKIJ 9	Unknown	Unknown	Unknown	Unknown	-
CRB315	EIFFEL	field	cultivar	North-Europe	spring	4
CRB316	FRISSON	field	cultivar	North-Europe	winter	5
CRB317	MESSIRE	field	cultivar	North-Europe	spring	4
CRB321	PILET	garden	cultivar	North-Europe	spring	6
CRB323	VERCAS	garden	cultivar	North-Europe	spring	3
CRB324	GUISANTE2	garden	local population	South-Europe	spring	1
CRB325	GUISANTE1	garden	local population	South-Europe	spring	1
CRB326	ARVELLO1	garden	local population	South-Europe	spring	1
CRB329	CLAUSELAN	garden	cultivar	North-Europe	spring	6
CRB330	Pisum sativum-ETHIOPIA JI281	Unknown	local population	Africa	Unknown	2
CRB331	GUISANTE	garden	local	South-Europe	spring	1
CRB333	FROGEL	garden	cultivar	North-Europe	winter	5

CRB334	GRADUS	garden	cultivar	North-Europe	spring	3
CRB335	JUWEL	garden	cultivar	North-Europe	spring	3
CRB336	MINI	garden	cultivar	North- America	spring	3
CRB337	LINCOLN	garden	cultivar	North-Europe	spring	3
CRB340	WISCONSIN MERIT	garden	cultivar	North- America	spring	3
CRB341	MERVEILLE DE KELVEDON	garden	cultivar	North-Europe	spring	3
CRB342	STARCOVERT	garden	cultivar	North-Europe	spring	3
CRB344	WELLENSIEK'S WHITE INDENT JI805	Unknown	germplasm	North-Europe	Unknown	6
CRB345	Pisum sativum-ETHIOPIA JI1594	Unknown	wild	Africa	Unknown	6
CRB346	MONGOLIA	Unknown	local population	Far-East	Unknown	1
CRB347	YELLOW POLLEN	Unknown	germplasm	Unknown	Unknown	1
CRB348	PINKISH WHITE	Unknown	germplasm	North-Europe	Unknown	1
CRB349	PROGRETA	field	cultivar	North-Europe	spring	4
CRB350	HAMMARLUNDS K LINE JI146	Unknown	germplasm	South-Europe	Unknown	2
CRB351	Pisum transcaucasicum JI45	Unknown	wild	South-Europe	Unknown	2
CRB352	Pisum sativum-ETHIOPIA JI1431	Unknown	local population	Africa	Unknown	1
CRB353	COSTA RICA	garden	germplasm	South- America	spring	6
CRB354	SPRUT K8349	field	cultivar	North-Europe	spring	1
CRB355	GREITUKAI K4834	Unknown	cultivar	North-Europe	spring	1
CRB356	HOHENHEIMER GRUNE VIKTORIA K4252	field	local population	North-Europe	spring	6
CRB357	ZHALYUKAI K4269	field	local population	North-Europe	spring	1
CRB358	ROSTOVSKII VYSOKII K4356	field	local population	North-Europe	spring	1
CRB359	KHIRPOSINSKII K5127	Unknown	local population	North-Europe	spring	1
CRB360	K30	Unknown	local population	North-Europe	spring	6
CRB361	K1666	Unknown	local	North-Europe	spring	1
CRB362	NORD K8290	field	cultivar	North-Europe	spring	4
CRB363	K4088	Unknown	local population	North-Europe	spring	1
CRB364	K4170	fodder	local	North-Europe	spring	1
CRB365	E11	Unknown	cultivar	Unknown	Unknown	1
CRB366	BENDI	garden	local population	Far-East	spring	6
CRB367	ER BLANCH	garden	local population	Far-East	spring	6
CRB368	HECAI	garden	local population	Far-East	spring	6
CRB369	YANGWAN	garden	local population	Far-East	spring	6
CRB370	WANDOU 2	garden	local population	Far-East	spring	6
CRB371	WANDOU 1	garden	local population	Far-East	spring	6
CRB372	ULADOVSKIJ 10	garden	breeding	North-Europe	spring	6
CRB374	Pisum humile JI241	Unknown	wild	Middle-East	Unknown	2
CRB375	CENNIA	field	cultivar	North-Europe	spring	6
CRB376	RONDO	garden	cultivar	North-Europe	spring	4

CRB377	661	field	breeding	North-Europe	spring	4
CRB378	320*247=BP	field	breeding	North-Europe	winter	4
CRB379	C413=AP	field	breeding	North-Europe	winter	5
CRB380	C437	field	breeding	North-Europe	winter	5
CRB381	Pisum elatius JI1089	Unknown	wild	Middle-East	Unknown	2
CRB382	CD5MM	field	breeding	North-Europe	spring	4
CRB383	CE101=FP	field	breeding line	Unknown	winter	5
CRB384	CF100=GP	field	breeding	North-Europe	Unknown	7
CRB385	CHAMPAGNE	fodder	cultivar	North-Europe	winter	7
CRB386	Pisum sativum-ETHIOPIA JI174	Unknown	local population	Africa	Unknown	6
CRB388	M163*247=CP	field	breeding line	Unknown	winter	5
CRB389	DANTO	field	cultivar	europe	spring	4
CRB390	DP	fodder	germplasm	North-Europe	winter	7
CRB394	JI103	Unknown	landrace	Unknown	Unknown	2
CRB395	PRIEKULSKIJ-341-def JI1184	Unknown	cultivar	North-Europe	Unknown	6
CRB397	KEERAUPEA	fodder	landrace	Unknown	Unknown	1
CRB398	WIRAIG JI190	Unknown	local population	Africa	Unknown	2
CRB399	Aa82 JI252	Unknown	wild	Africa	Unknown	2
CRB400	JI261	Unknown	wild	Unknown	Unknown	2
CRB401	MOSHONG 1	Unknown	local population	Middle-East	Unknown	2
CRB402	KATRIN	field	cultivar	North-Europe	spring	4
CRB403	L808	fodder	wild	Africa	Unknown	2
CRB405	MELROSE	fodder	cultivar	North- America	winter	7
CRB406	MERCURE	field	breeding line	Unknown	winter	5
CRB407	MULTIFREEZER	garden	cultivar	North- America	spring	3
CRB408	OPTIMA	field	cultivar	North-Europe	spring	4
CRB409	P1259	Unknown	germplasm	Unknown	winter	7
CRB410	PF31905	field	breeding	europe	spring	4
CRB411	PF35323	field	breeding	europe	spring	4
CRB412	PUGET	garden	cultivar	Unknown	spring	3
CRB413	TERESE	field	cultivar	europe	spring	4
CRB414	VIRGO	field	cultivar	North-Europe	spring	1
CRB415	Wav F502	garden	breeding	North-Europe	spring	3
CRB416	Wav F750	garden	breeding	North-Europe	spring	3
CRB418	ALASKOR	garden	cultivar	North-Europe	spring	3
CRB419	ALDOT	garden	cultivar	North- America	spring	6
CRB420	ARAVIS	field	cultivar	North-Europe	winter	4
CRB421	AUSTRIAN WINTER	fodder	cultivar	Unknown	winter	7
CRB423	BACCARA	field	cultivar	North-Europe	spring	4
CRB424	BLIXT 460	fodder	breeding	North-Europe	winter	7
CRB425	BLIXT 461	fodder	breeding	North-Europe	winter	7
CRB426	BLIXT 468	fodder	breeding	North-Europe	winter	7
CRB427	BLIXT 195	fodder	breeding	North-Europe	winter	7
CRB429	CCA	field	cultivar	South-Europe	winter	7
CRB430	CHEYENNE	field	cultivar	North-Europe	winter	4
CRB431	FENN	fodder	cultivar	North- America	winter	7
CRB432	FINALE	field	cultivar	North-Europe	spring	4
CRB433	FRILENE	field	cultivar	North-Europe	winter	5

CRB435	GLACIER	garden	cultivar	North- America	spring	7
CRB436	HOLLY 11	fodder	local population	North-Europe	winter	7
CRB437	KARNOBAT	field	cultivar	South-Europe	winter	7
CRB438	KAZAR	field	cultivar	North-Europe	winter	5
CRB439	MIR 12	Unknown	cultivar	Unknown	spring	4
CRB440	RAFALE	field	cultivar	North-Europe	winter	5
CRB442	TORSDAG	field	cultivar	North-Europe	spring	1
CRB443	VENDEVIL	field	cultivar	North-Europe	winter	7
CRB444	VICTOR	field	cultivar	North- America	winter	5
CRB445	VITALIS	garden	cultivar	North-Europe	spring	3
CRB446	WINTERBERGER	Unknown	local population	North-Europe	winter	7
CRB448	CHEMIN LONG JI296	garden	cultivar	North-Europe	spring	6
CRB449	CLAMART HATIF	garden	cultivar	North-Europe	spring	6
CRB450	MICHAUX DE PARIS	garden	cultivar	North-Europe	spring	6
CRB451	CHAMPION D'ANGLETERRE	garden	cultivar	Unknown	spring	1
CRB452	PLEIN LE PANIER	garden	cultivar	North-Europe	spring	3
CRB453	SERPETTE D'AUVERGNE	garden	cultivar	North-Europe	spring	6
CRB454	CAROUBY DE MAUSSANE	mangetout	cultivar	North-Europe	spring	6
CRB455	CORNE DE BELIER	mangetout	cultivar	North-Europe	spring	6



Additional file 2: Figure S1. Distribution of Best Linear Unbiased Prediction (BLUP) values for the five traits observed within the pea collection.

A: frost damages in the controlled conditions experiment. B, C, D and E: frost damages in the field experiment at the date 1, 2, 3 and 4 respectively.

Annexes



Additional file 3 : Figure S2. Distribution of 10739 SNPs along the Pisum sativum linkage groups. Number of SNPs per position are indicated as grey horizontal bars. Genetic distance in cM is shown on the y-axis and number of SNPs per position is shown on the x-axis.



Additional file 3: Figure S2. Distribution of 10739 SNPs along the *Pisum sativum* linkage groups (*Continued*).

Additional file 4: Table S2. Description of the *Pisum sativum* linkage groups (LGs) used in the present study. For each LG, are shown the number of SNP markers, the genetic length (in cM, from Tayeh et al. [36]) and the average minor allele frequency (MAF).

LG	Number of SNP	Genetic lenght	Average MAF
	markers	(cM)	
Ι	1220	93.2	0.29
II	1584	114.1	0.3
III	1888	135.3	0.3
IV	1641	115.4	0.3
V	1332	113.3	0.31
VI	1398	101	0.31
VII	1676	122.4	0.3



Additional file 5: Figure S3. Distribution of minor allele frequencies (MAF) for 10739 SNP markers within the 363 pea accessions.



Additional file 6: Figure S4. Scatterplot showing the linkage disequilibrium (LD) decay estimated in the association mapping collection.

The LD decay across each linkage group (LG) and the overall LD decay across the genome (All LG) are shown. The r2 values of LD between pairs of markers considered are plotted as a function of the genetic position in cM. Red curves represent the estimated LD decay. Blue dashed horizontal lines represent half of the maximum LD value. Blue dashed vertical lines represent the estimated genetic distance (cM) at which the LD decay dropped to half of its maximum. LD decay rate is represented as the point of intersection between the two dashed lines.



Additional file 7: Figure S5. Population structure of the pea association mapping collection based on Discriminant Analysis of Principal Components (DAPC) analysis.

(A)Number of clusters vs BIC values. The x-axis shows the potential numbers of clusters that could structure the population. On the y-axis is represented the BIC value associated with each number of clusters. (B) Number of principal components (PCs) vs a-score criterion. The x-axis shows the potential numbers of PCs that could be used in the principal component analysis (PCA) step of DAPC. On the y-axis is given the a-score criterion associated with each number of PCs. The optimal number of PCs, represented by a red bar, was obtained after 100 permutations. (C) Scatterplot showing the distribution of the association mapping collection along the first two principal components of the DAPC. Accessions are represented by dots and genetic clusters as inertia ellipses coded from 1 to 7. The bottom right inset shows eigenvalues of the six principal components in relative magnitude, ordered from 1 to 6 from the left to the right.


Additional file 8: Figure S6. Distribution of the kinship coefficients between accessions of the association mapping collection.

The first histogram (A) describes the distribution of the kinship coefficients within the K matrix, calculated with all markers of the genome. The remaining histograms (B, C, D, E, F, G and H) describe the kinship coefficients within each of the seven KLG matrices calculated as explained in the material and methods section of the manuscript, for example the kinship matrix KLG1 was estimated with all the markers except those that are located on the first linkage group.

Additional file 9: Table S3. Description of linkage disequilibrium (LD) blocks per linkage group in the association mapping collection.

A LD block consists in a series of at least 2 markers which are in significant LD (r2 > 0.8) with at least one trait-associated marker (underlined marker). LD blocks are named in consecutive numerical order following their linkage group (LG) name. cM*: genetic position, in centiMorgan of each marker along the genetic map of the corresponding linkage group; LD (r^2) **: r^2 value of each marker with the other markers of the same LD block.

L G	LD block	Associated marker	Associated trait			N	larker i	n LD (r ² >0.8)	with t	rait-ass	ociateo	l mark	er		
Ι	I.1				PsCam020765_11567_1428	PsCam025082_14309_170										
				cM*	47.5	47.5										
		PsCam020765_11567_14 28	FD_Field_date1	LD (r ²)**	1.00	0.92										
III	III.1			aM*	PsCam049061_31705_537	PsCam035617 20792 637	2 PsCam037922 22979 691									
				cM*	131.1	132	132.1									
		PsCam049061_31705_53 7	FD_CC	LD (r ²)**	1.00											

		PsCam035617_20792_63 7	FD_CC. FD_Field_date1		0.91	1.00							
		PsCam037922_22979_69 1	FD_CC. FD_Field_date1		0.91	1.00	1.00						
V	V.1				PsCam051352_33858_733	PsCam051635_34107_368	PsCam042222 26315 1205			-			
				cM*	53.6	58.6	61						
		PsCam051352_33858_73 3	WFD_Field_date4	LD (r ²)**	1.00								
		PsCam051635_34107_36 8	WFD_Field_date4		0.70	1.00							
		PsCam042222_26315_12 05	WFD_Field_date4		0.82	0.81	1.00						
VI	VI.1				PsCam021402 12023 752	PsCam043553 27552 265							
				cM*	41.4	43							
		PsCam021402_12023_75 2	WFD_Field_date4	LD (r ²)**	1.00								

	PsCam043553_27552_26		1	0.01	1.00											
	5	WFD_Field_date4		0.81	1.00											
VI.2				⁹ sCam022275_12527_210	² sCam006662_4945_2210	2sCam004890 3696 1580	9sCam023246 13111 1125	² sCam037082 22189 1302	PsCam057485_38078_2761	9sCam037030_22140_221	?sCam009869_6623_546	sCam037565_22643_550	sCam034572_19953_2034	9sCam036921_22033_689	9sCam042568 26621 42	PsCam009542_6362_1386
			cM*	47.4	47.4	49.1	49.1	49.1	49.1	49.1	49.1	49.1	50.6	50.9	51.1	55.9
	PsCam022275_12527_21 0	FD_CC. FD_Field_date1. FD_Field_date4		1.00	0.81				0.74	0.82	0.79	0.63	0.82	0.79		0.73
	PsCam004890_3696_158 0	FD_Field_date4		0.79	0.83	1.00			0.76	0.84	0.77	0.80	0.85	0.79		0.76
	PsCam023246_13111_11 25	FD_Field_date4	LD (r ²)**	0.87	0.88	0.87	1.00		0.82	0.93	0.83	0.70	0.94	0.85		0.84
	PsCam037082_22189_13 02	FD_Field_date4		0.80	0.87	0.83	0.89	1.00	0.78	0.86	0.79	0.71	0.87	0.81		0.80
	PsCam042568_26621_42	FD_Field_date4		0.82	0.86	0.85	0.94	0.87	0.80	0.90	0.81	0.68	0.95	0.87	1.00	0.87
V1.3				PsCam043847_27818_181	PsCam048404 31138 2325	PsCam001662 1375 798	PsCam050022_32630_1411									
			cM*	49.1	49.1	49.1	49.1									

	PsCam043847_27818_18 1	FD_CC		1			0.70						
	PsCam048404_31138_23 25	FD_CC	LD (r ²)**	0.84	1		0.83						
	PsCam001662_1375_798	FD_Field_date4	_	0.72	0.88	1	0.75						
VI.4				PsCam007060_5248_2156	PsCam026381 15234 450	PsCam037879_22937_948	PsCam003337_2589_220	PsCam037603_22679_2024	PsCam040539 25267 719				
			cM*	50.1	50.1	50.1	50.2	50.2	50.2				
	PsCam007060_5248_215 6	FD_CC		1.00									
	PsCam026381_15234_45 0	FD_CC		0.70	1.00								
	PsCam037879_22937_94 8	FD_CC	LD (r ²)**	0.83	0.84	1.00							
	PsCam003337_2589_220	FD_CC		0.99	0.69	0.82	1.00						
	PsCam037603_22679_20 24	FD_CC		0.95	0.66	0.79	0.96	1.00					
	PsCam040539_25267_71 9	FD_CC		0.95	0.66	0.79	0.96	1.00	1.00				

VI.5				PsCam039054_24037_586	PsCam049925_32537_2703	PsCam043101_27130_869	PsCam036192_21340_510					
			cM*	50.6	50.7	50.9	50.9					
	PsCam039054_24037_58 6	FD_CC	LD (r ²)**	1.00		0.93	0.90					
	PsCam049925_32537_27 03	FD_CC		0.96	1.00	0.92	0.89					
VI.6				PsCam006586_4888_245	PsCam020872 11644 1756	PsCam033562 19160 307	PsCam044418_28295_579					
			cM*	51.1	51.1	51.1	51.1					
	PsCam006586_4888_245	FD_CC		1.00								
	PsCam020872_11644_17 56	FD_CC]	0.76	1.00							
	PsCam033562_19160_30 7	FD_CC	LD (r ²)**	0.75	0.80	1.00						
	PsCam044418_28295_57 9	FD_CC		0.87	0.85	0.77	1.00					

VI	7				PsCam036704_21832_970	PsCam053189_35238_569	PsCam035454_20637_2339					
				cM*	51.1	51.1	51.1					
		PsCam036704_21832_97 0	FD_CC	LD (r ²)**	1.00		0.89					
		PsCam053189_35238_56 9	FD_CC		0.86	1.00	0.84					
VI	8				PsCam011774_8038_200	PsCam023852 13541 264						
				cM*	52.8	52.8						
		PsCam011774_8038_200	FD_CC		1.00							
		PsCam023852_13541_26 4	FD_CC	LD (r ²)**	0.81	1.00						

VI I	VII.1				PsCam000301_265_1519	PsCam045010_28743_3365						
				cM*	87.9	88						
		PsCam000301_265_1519	FD_Field_date2. FD_Field_date3	I.D. (2) ##	1.00							
		PsCam045010_28743_33 65	FD_Field_date2	LD (r ²)**	0.83	1.00						

Additional file 10: Table S4. Marker haplotype analysis of the association mapping collection.

For each linkage group (LG), the list of markers significantly detected by GWAS and markers in linkage disequilibrium (LD; $r^2 > 0.8$) with the former ones is shown. The third line shows genetic positions from the consensus map of Tayeh et al. [36]. The fourth line indicates the LD blocks composed and named as mentioned in the legend of Additional file 9. The following lines show the allelic composition of haplotypes defined by LD blocks and individual associated markers at each of the 6 frost tolerance loci on linkage groups (LGs) I, II, III, V, VI and VII. For each frost damage (FD)-associated marker, the favourable allele is in red font and the unfavourable allele in blue font. Haplotypes are named in consecutive numerical order following their linkage group name; only haplotypes without missing values or heterozygous markers and carried by more than 3 % of the lines from the associated with marker(s) in the linkage group are shown. Significant differences between haplotypes were assayed by a SNK means comparison test; Favourable haplotypes are shown by a red background and unfavourable haplotypes by a blue background, regarding the SNK test. Haplotypes with a white background are classified in intermediate groups.

L	G																						VI																																			
м	arker	PsCam037030 22140 221		PSCam05/082_22189_1502	PsCam037565_22643_550	PsCam057485_38078_2761	PsCam034572_19953_2034	PsCam036921 22033 689	PsCam042568 26621 42		PsCam009542_6362_1386	PsCam003541 2731 574	PsCam050834 33394 133	PsCam001662 1375 798	Perfam043847 27818 181	3000 00110 F0F0F0	LaCallio4040404 21120 2220	Pscam050022_32630_1411	PsCam045419 29091 815	PsCam053556 35439 228	PsCam007060 5248 2156	PsCam026381 15234 450	PsCam037879 22937 948	PsCam003337 2589 220	PsCam037603 22679 2024	PsCam040539 25267 719	PsCam001302 1089 1249	PsCam035398 20586 1094	Defferen 030054 24037 586	P. C. 040005 2002 200	rscamu49925 525512105	PsCam036192_21340_510	PsCam043101_27130_869	PsCam038477 23510 469	PsCam001148 968 692	PsCam006586 4888 245	PsCam020872 11644 1756	PsCam033562 19160 307	PsCam044418 28295 579	PsCam025448 14564 2314	PsCam035454_20637_2339	PsCam036704 21832 970	PsCam053189 35238 569	PsCan038677 23691 325	PsCam035356 20546 778	PsCam040326 25099 1923	PsCam011774 8038 200	Pscam025522 15241 204	261 00001 00001 0	Number of accessions arrying the haplotype	Associa Mean (ted (± SE)	traits with sig) of associated haplotypes a	nificant 1 1 traits fo nd grouj	narke or sig o of Si	er detected nificantly r NK test	l by GV eprese	WAS ented
Ge	enetic	17	1.5	3	7	7	9,0	6,0		: :	S.	~	5,5	-				-	-	Ξ,	0,1	1	1,	.2	.2	.2	0	.6	Y Y	2	2	S,	<u>s</u>	8	,1	-,	,1	,1	-, ,	-		Γ,	-	١,	5	S, I	<u>ه</u> ه	0 0	0				FD Field	FD F	hla	FD Field	FD	Field
ро	sition (cM)	4		4 •	4	4	Š	Š	ŝ	i	8	48	48	4	4		ř -	4	4	4	50	Š	Š	Ň	Š	Š	N N	Š	V	i i	<u> </u>	ň	ы	Ň	S.	5	5	5	6	5	5	5	5	S I	5	5	ŝ	6 6	5		FD_C	С	date1	date	2	date3		ate4
LI	D block		V.	1.2												VI.3	<u>,</u>						V	1.4							VI.	5					VI.	6			V	1.7					VI.8	3					unter	unt	-			
z Ha	aplotype VI.1	BB	3 A	AB	BB	BB	BB	BB	BE	B	Β.	AA	BB	BB	BI	3 A	A A	A.	AA	AA	BB	BB	AA	AA	. AA	. AA	A.A.	A A	A BI	BB	BE	B	AA .	AA 1	BB .	AA I	BB .	AA A	A E	BB E	BBE	BB E	BB E	BB I	BB A	AA E	BB A	A A	A 9	1	3.50 ± 0.02	5 d	1.91 ± 0.06 c				2.94 ±	: 0.05 b
Ha Ha	aplotype VI.2	AA	A B	B A	A	AA	AA	AA	A.	A A	LA 1	BB	AA	AA	. A.	ΑB	BE	B	BB	BB	AA	AA	BB	BB	BB	BB	BE	BE	8 A.	A A	A A	A I	BB	BB /	AA I	BB .	AA I	BB E	B A	AA A	A A	AA A	AA A	AA A	AA B	BB A	AA B	B Bl	B 3	3	0.99 ± 0.00	5 a	0.18 ± 0.04 a				1.13 ±	0.08 a
e Ha	aplotype VI.3	BB	3 A	AB	B	BB	BB	BB	BE	3 B	B	BB	BB	BB	BI	3 A	A A	A.	AA	AA	AA	AA	BB	BB	BB	BB	BB	A	A A	A A	A A	A I	BB	AA .	AA I	BB .	AA I	BBB	BB	BB A	A A	AA A	AA A	AA A	AA B	BB	BB A	A A	A 2	3	2.78 ± 0.11	l c	1.39 ± 0.10 b				2.92 ±	0.10b
d Ha	aplotype VI.4	BB	3 A	AB	B	BB	BB	BB	BE	3 B	B	BB	BB	BB	BI	3 A	A A	A I	BB	AA	AA	AA	BB	BB	BB	BB	BE	BE	A	A A	A A	AI	BB	BB /	AA I	BB .	AA I	BB B	B A	AA A	A A	AA A	AA A	AA A	AA B	BB A	AA B	B BI	B 1	8	2.67 ± 0.12	2 c	1.88 ± 0.11 c				2.97 ±	0.10b
TH Ha	aplotype VI.5	AA	ΑB	B A	A	AA	AA	AA	AA	A A	LA 1	BB	AA	AA	A	ΑB	BE	B	BB	BB	AA	AA	BB	BB	BB	BB	BE	BE	3 A	A A	A A	A I	BB	BB /	AA :	BB .	AA I	BB B	BB	BB A	A A	AA A	AA B	BB /	AA B	BB A	AA B	B BI	B 1	2	1.26 ± 0.12	2 b	0.30 ± 0.09 a				1.04 ±	0.16 a

	LG		I	П		Ш						V							1	/11									
	Marker	PsCam020765_11567_1428	PsCam025082_14309_170	PsCam043698 27686 2011	PsCam049061 31705 537	PsCam035617_20792_637	PsCam037922_22979_691	PsCam051352 33858 733	PsCam051635_34107_368	PsCam042222 26315 1205	PsCam048068 30823 2326	PsCam002670_2123_708	PsCam011361 7735 182	PsCam012913 8715 718	PsCam049838 32453 1430	PsCam001108 940 48	PsCam050827 33387 242	PsCam037927_22984_97	PsCam038378 23415 721	PsCam042427 26489 524	PsCam000301_265_1519	PsCam045010 28743 3365	PsCam004928_3732_3087	Number of accessions carrying the haplotype	Associ Mean (± SE	ated traits with) of associated tr and	significant mar raits for signific d group of SNK	ker detected by antly represent test	GWAS ed haplotypes
	Genetic position (cM) LD block	47,5	41,5	20,5	131,1	132	132,1	<mark>53,</mark> 6	9; 85 V.1	61	56,9	58,6	60,2	60,2	60,2	72,9	82,4	82,5	86,4	87,1	A 87,9	8 1.1	89,3		FD_CC	FD_Field_dat el	FD_Field_dat e2	FD_Field_dat e3	FD_Field_dat e4
	Haplotype I.1	AA	AA										I						-					314		1.29 ± 0.03 a			
	Haplotype I.2	вв	BB																					45		2.31 ± 0.07 b			
	Haplotype II.1			AA																				218	2.59 ± 0.05 a				
	Haplotype II.2			BB																				140	3.13 ± 0.05 b				
	Haplotype III.1				BB	AA	AA																	177	2.65 ± 0.05 a	1.31 ± 0.04 a			
	Haplotype III.2				AA	BB	BB																	178	2.90 ± 0.05 b	1.51 ± 0.05 b			
	Haplotype V.1							BB	BB	BB	AA	BB	BB	BB	AA									217					2.96 ± 0.03 c
	Haplotype V.2							AA	AA	AA	BB	AA	AA	AA	BB									60					1.36 ± 0.06 a
bes	Haplotype V.3							AA	AA	AA	BB	BB	AA	AA	BB									13					1.23 ± 0.13 a
t a	Haplotype V.4							BB	BB	BB	BB	BB	BB	AA	BB									10					2.47±0.19 b
H	Haplotype VII.1															AA	AA	AA	BB	AA	AA	AA	AA	93	2.41 ± 0.06 b		1.11±0.05 b	2.09 ± 0.07 b	
	Haplotype VII.2															BB	AA	AA	BB	AA	AA	AA	AA	28	2.89 ± 0.12 c		1.66 ± 0.12 c	2.88 ± 0.13 d	
	Haplotype VII.3															AA	AA	AA	AA	BB	AA	AA	AA	27	$3.11\pm0.12~\text{cd}$		1.17 ± 0.08 b	2.37 ± 0.11 bc	
	Haplotype VII.4															BB	AA	BB	BB	AA	AA	AA	BB	26	0.96 ± 0.05 a		0.00 ± 0.00 a	0.32 ± 0.06 a	
	Haplotype VII.5															BB	BB	BB	AA	BB	BB	BB	BB	25	4.03 ± 0.08 f		3.17 ± 0.13 d	4.29 ± 0.12 e	
	Haplotype VII.6															AA	BB	BB	AA	BB	AA	AA	AA	10	3.38 ± 0.15 de		1.53 ± 0.15 c	2.60 ± 0.20 cd	
	Haplotype VII.7															BB	AA	AA	AA	AA	AA	AA	AA	11	3.12 ± 0.17 cd		1.69 ± 0.17 c	3.03 ± 0.27 d	
	Haplotype VII.8															BB	BB	BB	AA	BB	AA	AA	AA	11	3.47 ± 0.21 e		1.75 ± 0.16 c	2.66 ± 0.19 cd	
	Haplotype VII.9															BB	AA	BB	BB	AA	AA	AA	AA	10	3.27 ± 0.27 de		1.81 ± 0.28 c	2.76 ± 0.33 cd	

Additional file 10: Table S4. Marker haplotype analysis of the association mapping collection (Continued).

Additional file 11: Table S5. Typical accesssions carrying favourable and unfavourable haplotypes.

Favourable haplotypes are shown by a yellow background. Unfavourable haplotypes are shown by a blue background.

Haplotype	Typical accesssions carrying favourable and unfavourable haplotypes
Haplotype I.1	JI267, JI2473, JI2523, 86-22, 87-500, BLIXT7, BLIXT104, BLIXT157, BLIXT484, BOOSTER*AW7, CE*CF, HOLLY 2, HOLLY 3, HOLLY 4, HOLLY 5, HOLLY 17, HOLLY 18, MISTRAL, PICAR, WNC 23Z SPP ARVENSE 1809, WNC Z61Z SPP ARVENSE Z009, PI512066, 21, 34R, 5031, 5085, K5284, BLACK SEEDED, CLAUSE H4, st+af+tl, 320*247=BP, C413=AP, CF100=GP, CHAMPAGNE, JI261, MELROSE, P1259, BLIXT460, BLIXT195, CCA, CHEYENNE, FENN, GLACIER, HOLLY11, VENDEVIL, WINTERBERGER, DOVE, JI1075, 87-317, ASSAS, AW2, BLIXT109, COTE D'OR, GA91H2, MUENCHENER BANATER WINTEREBSE, HOLLY 6, HOLLY 9, HOLLY 13, MONITOR, NEVE*b195-5, TOPOLIANSKI, XSARA, BEZELYA, P58, 194, 5099, HATIVER, FIGARO, CAMÉOR, FRISSON, JI1431, CE101=FP, DP, ALASKOR, ARAVIS, AUSTRIAN WINTER, BLIXT461, KAZAR, MIR 12, CHINE-D367, WISCONSIN PERFECTION, JI1844, BOOSTER, BUL179, DOLMEN, SENTINEL, GSP1016, HOLLY 7, HOLLY 8, ICEBERG, NEVE, US IDAHO W6 8955, PI269763, 418, ENGLISH, TASMANIE, POL 6806, AFGHANISTAN ASIATICUM, NEPAL A, PIVER, EIFFEL, WISCONSIN MERIT, JI45, RONDO, 661, CD 5 MM, MERCURE, BLIXT468, FINALE, FRILENE, KARNOBAT, TORSDAG, VICTOR, ABAQUE, BIRTE, CHINE-D371, KATHMANDOU DISTRICT 2, USATYJ, JI1794, BRIDGE, ORANGE POD-orp, GSP1017, GSP1078, GSP1091, SYDNEY, VERTOP, MN 108, PI271936, FENN, 0-2042, PI193841, MAZARE, AGAT, CAPSICUM, JAYGEE, LAFONTAINE, DESIREE, SETCHEY, CHIPEAU, 776, BALLET, CHINA, MESSIRE, E175, C437, JI1089, DANTO, JI1184, PF35323, BACCARA
Haplotype I.2	ERSLING
Haplotype II.1	BOOSTER*AW7, 5031, FENN, DOVE, CHINA, CHEYENNE, HATIVER, WNC Z61Z SPP ARVENSE Z009, FENN, BLIXT157, 87- 317, J11075, HOLLY 7, HOLLY 17, BLIXT104, 5099, BLIXT461, HOLLY 6, HOLLY 9, MONITOR, 86-22, CHEYENNE, HOLLY 5, HOLLY 18, HOLLY11, ASSAS, BLIXT7, BLIXT484, HOLLY 8, BLACK SEEDED, CLAUSE H4, MELROSE, P1259, AUSTRIAN WINTER, BLIXT468, BLIXT195, VENDEVIL, VICTOR
Haplotype II.2	K8349, KIRIN 40, MERVEILLE DE KELVEDON, AUTOMOBILE, FELTHAM FIRST, MERVEILLE DE WEISSENFELS, JI1267, MINGOMARK, MIRAVIL, BUTARO 1A, WavF750, JI1831, E386=JI805, BRIDGE, K4356, CAPELLA, PI180693, SHRAT, KEERAUPEA, PI179461, CENNIA, PH14-119, MERVEILLE D'ETAMPES, TORSDAG, LUMINA, WavF502, JI2376, 90-2079, CAROUBEL, JI281

Haplotype III.1	BOOSTER*AW7, 5031, DOVE, CHEYENNE, HATIVER, WNC Z61Z SPP ARVENSE Z009, 87-317, 5099, MONITOR, 86-22, ALASKOR, ICEBERG, VENDEVIL, VICTOR
Haplotype III.2	PI180693, AUTOMOBILE, ERSLING
Haplotype V.1	533, CAROUBEL, CENNIA, CORNE DE BELIER, K4356, K4626, K8349, MICHAUX DE PARIS, RODOGUNE, SOMMETTE, VIRGO
Haplotype V.2	AW2, BLIXT104, BLIXT109, BLIXT157, BLIXT7, CF100=GP, CHAMPAGNE, CHEYENNE, COTE D'OR, DP, HOLLY 17, HOLLY 2, HOLLY 3, HOLLY 6, HOLLY 9, KAZAR, MELROSE, NEVE*b195-5, PICAR, TASMANIE, TOPOLIANSKI, VICTOR, WINTERBERGER
Haplotype V.3	BLACK SEEDED, BLIXT461, BLIXT484, GLACIER, HOLLY 4, HOLLY11, KARNOBAT
Haplotype VI.1	AUTOMOBILE, CAROUBEL, CENNIA, ERSLING
Haplotype VI.2	87-317, AW2, BLACK SEEDED, BLIXT109, BLIXT7, BOOSTER*AW7, CHAMPAGNE, CHEYENNE, COTE D'OR, DOVE, FENN, GLACIER, HOLLY 17, HOLLY 18, HOLLY 3, HOLLY 5, HOLLY 8, HOLLY 9, JI1844, K5284, MELROSE, MONITOR, P1259, PICAR, TOPOLIANSKI, VENDEVIL, WINTERBERGER
Haplotype VI.5	BLIXT104, BLIXT157, BLIXT460, CLAUSE H4, DP, HOLLY 6, NEVE*b195-5
Haplotype VII.4	86-22, BLACK SEEDED, BLIXT104, BLIXT109, BLIXT157, BLIXT484, BLIXT7, BOOSTER*AW7, CF100=GP, CLAUSE H4, COTE D'OR, DP, FENN, GLACIER, HOLLY 17, HOLLY 4, HOLLY 6, HOLLY 9, HOLLY11, MELROSE, NEVE, NEVE*b195-5, P1259, PI512066, PICAR, WINTERBERGER
Haplotype VII.5	667, AUTOMOBILE, CADOR, CENNIA, EM, ERSLING, FELTHAM FIRST, FINLANDE, MERVEILLE DE KELVEDON, MERVEILLE DE WEISSENFELS, MINGOMARK, MIRAVIL, POIS AMERICAIN COUSIN, SOMMETTE, TRAPPER, WavF502

Additional file 12: Table S6. Candidate genes annotated for markers significantly detected by GWAS and markers in linkage disequilibrium (LD; $r^2 > 0.8$) with the former ones.

SNP markers are shown according to their order on the consensus genetic map. Associated SNP markers detected by GWAS are underlined. LD blocks are named as described in Additional file 9. The annotation of markers was extracted from Tayeh et al. [36] which was itself obtained following Blastx search against P. sativum, M. truncatula, G. max, and A. thaliana protein sequences.

									Anno	tation following Blastx (e	e-value cu	tt-off of 1e-10) against protein sequence	es from:	
LG	LD block	Marker	Genetic position	Associated trait	Pea gene	Reference		P. sativum		M. truncatula		G. max		A. thaliana
			(CM)				e- value	predicted protein function	e- value	predicted protein function	e- value	predicted protein function	e- value	predicted protein function
Ι	I.1	PsCam020765_11567_1428	47,5	FD_Field_date1			8,99E- 64	copper amine oxidase	0	Copper amine oxidase	0	copper amine oxidase 1-like	0	Copper amine oxidase family protein
		PsCam025082_14309_170	47,5						1,06E- 73	60S ribosomal protein L27a-3			5,04E- 65	Ribosomal protein L18e/L15 superfamily protein
II		PsCam043698_27686_2011	20,5	FD_CC					5,01E- 16	Early nodulin 93 protein			8,65E- 19	early nodulin-related
III	III.1	PsCam049061_31705_537	131,1	FD_CC			2,74E- 80	heat shock protein hsp70	0	Heat shock 70 kDa protein 4L	0	heat shock 70 kDa protein 14-like	0	Heat shock protein 70 (Hsp 70) family protein
		PsCam035617_20792_637	132	FD_CC, FD_Field_date1			3,42E- 118	brassinosteroid receptor	0	Receptor-like protein kinase	0	receptor-like protein kinase-like	0	PEP1 receptor 1
		PsCam037922_22979_691	132,1	FD_CC, FD_Field_date1	PsGA3ox1 (=LE)	Martin et <i>al.</i> , 1997; Lester et al., 1997			2,12E- 48	Protein SRG1	0	gibberellin 3-beta-dioxygenase 1-like	1,62E- 134	gibberellin 3-oxidase 1
V	V.1	PsCam051352_33858_733	53,6	FD_Field_date4					1,51E- 135	PsbP domain- containing protein	9,36E- 125	psbP domain-containing protein 6, chloroplastic-like	1,25E- 110	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein
		PsCam051635_34107_368	58,6	FD_Field_date4										
		PsCam042222_26315_1205	61	FD_Field_date4					0	Eukaryotic translation initiation factor	0	eukaryotic translation initiation factor 5-like isoform X1	7,78E- 145	Translation initiation factor IF2/IF5
		PsCam048068_30823_2326	56,9	FD_Field_date4			8,28E- 48 brassinosteroid receptor				0	probable inactive leucine-rich repeat receptor-like protein kinase At3g03770-like isoform X1	0	Leucine-rich repeat protein kinase family protein
		PsCam002670_2123_708	58,6	FD_Field_date4										

		PsCam011361_7735_182	60,2	FD_Field_date4								
		PsCam012913_8715_718	60,2	FD_Field_date4			0	DNA cross-link repair protein SNM1	0	DNA cross-link repair protein SNM1-like isoform X1	3,24E- 173	DNA repair metallo-beta- lactamase family protein
		PsCam049838_32453_1430	60,2	FD_Field_date4			0	Nodulation receptor kinase	0	probable receptor-like protein kinase At1g30570-like isoform X1	3,89E- 148	Protein kinase superfamily protein
VI	VI.1	PsCam021402_12023_752	41,4	FD_Field_date4			0	Polyadenylation and cleavage factor-like protein			0	PCF11P-similar protein 4
		PsCam043553_27552_265	43	FD_Field_date4			9,79E- 160	Thaumatin-like protein	2,32E- 124	thaumatin-like protein-like	1,21E- 110	Pathogenesis-related thaumatin superfamily protein
	VI.2	PsCam006662_4945_2210	47,4				0	YTH domain family protein			3,68E- 93	evolutionarily conserved C- terminal region 4
		PsCam022275_12527_210	47,4	FD_CC, FD_Field_date1, FD_Field_date4								
		PsCam004890_3696_1580	49,1	FD_Field_date4					1,02E- 93	titin-like isoform X2	1,33E- 32	HSP20-like chaperones superfamily protein
		PsCam009869_6623_546	49,1									
		PsCam023246_13111_1125	49,1	FD_Field_date4			1,39E- 66	BAHD acyltransferase DCR	0	BAHD acyltransferase DCR-like	0	HXXXD-type acyl-transferase family protein
		PsCam037030_22140_221	49,1				1,17E- 70	C-repeat binding factor 14			1,13E- 39	dehydration response element B1A
		PsCam037082_22189_1302	49,1	FD_Field_date4			4,07E- 35	Protein SET DOMAIN GROUP	0	protein SET DOMAIN GROUP 41- like isoform X1	8,05E- 92	SET domain-containing protein
		PsCam037565_22643_550	49,1									
		PsCam057485_38078_2761	49,1				0	Pyruvate kinase	0	plastidial pyruvate kinase 2-like	0	plastidic pyruvate kinase beta subunit 1
		PsCam034572_19953_2034	50,6									
		PsCam036921_22033_689	50,9				0	Enolase	0	enolase-like	0	Enolase
		PsCam042568_26621_42	51,1	FD_Field_date4			0	ATP-dependent Clp protease ATP-binding subunit clpX	0	ATP-dependent Clp protease ATP- binding subunit clpX-like, mitochondrial-like	0	ATP-dependent Clp protease
		PsCam009542_6362_1386	55,9				0	Cyclic nucleotide gated channel	0	probable cyclic nucleotide-gated ion channel 20, chloroplastic-like isoform X1	0	cyclic nucleotide-binding transporter 1
		PsCam003541_2731_574	48	FD_CC								

	PsCam050834_33394_133	48,5	FD_CC	1,03E- 50	APL, partial	3,97E- 52	Two-component response regulator ARR1			8,04E- 98	Homeodomain-like superfamily protein
VI.3	PsCam001662_1375_798	49,1	FD_CC			1,74E- 65	Nodulin-related integral membrane protein DUF125	1,11E- 88	vacuolar iron transporter homolog 2- like	1,05E- 74	Vacuolar iron transporter (VIT) family protein
	PsCam043847_27818_181	49,1	FD_CC								
	PsCam048404_31138_2325	49,1	FD_CC	0	glutathione reductase	2,23E- 36	Dihydrolipoyl dehydrogenase	0	glutathione reductase, cytosolic-like isoform X2	0	glutathione-disulfide reductase
	PsCam050022_32630_1411	49,1		1,71E- 172	sucrose transport protein SUT1	2,37E- 173	SUT1-1	0	sucrose transport protein SUC1-like	7,78E- 165	sucrose-proton symporter 2
	PsCam045419_29091_815	49,1	FD_CC			0	Heat shock protein	0	heat shock cognate 70 kDa protein- like isoform X2	0	heat shock cognate protein 70- 1
	PsCam053556_35439_228	49,1	FD_Field_date4								
VI.4	PsCam007060_5248_2156	50,1	FD_CC			4,17E- 19	Protein SYM1	5,49E- 87	protein SYM1-like	3,28E- 84	Peroxisomal membrane 22 kDa (Mpv17/PMP22) family protein
	PsCam026381_15234_450	50,1	FD_CC								
	PsCam037879_22937_948	50,1	FD_CC			4,36E- 112	Palmitoyltransferase ZDHHC9	0	protein S-acyltransferase 8-like isoformX1	1,22E- 163	DHHC-type zinc finger family protein
	PsCam003337_2589_220	50,2	FD_CC								
	PsCam037603_22679_2024	50,2	FD_CC			2,59E- 98	Ripening regulated protein DDTFR18	0	protein TRANSPARENT TESTA 12-like isoform X1	2,08E- 145	MATE efflux family protein
	PsCam040539_25267_719	50,2	FD_CC								
	PsCam001302_1089_1249	50,2	FD_CC			2,08E- 102	F-box family protein	1,29E- 159	F-box protein PP2-A13-like	2,03E- 112	phloem protein 2-A13
	PsCam035398_20586_1094	50,6	FD_CC			0	ABC transporter B family member	0	ABC transporter B family member 28-like	0	non-intrinsic ABC protein 8
VI.5	PsCam039054_24037_586	50,6	FD_CC	4,20E- 24	hsr203J homolog	9,09E- 90	Gibberellin receptor GID1	4,47E- 133	carboxylesterase 1-like isoform 1	1,96E- 99	carboxyesterase 20
	PsCam049925_32537_2703	50,7	FD_CC			3,36E- 96	Mitochondrial metalloendopeptidase OMA1			7,15E- 174	Peptidase family M48 family protein
	PsCam036192_21340_510	50,9				0	Pentatricopeptide repeat-containing protein	9,84E- 160	putative pentatricopeptide repeat- containing protein At5g40405-like isoform X1	9,69E- 165	Tetratricopeptide repeat (TPR)-like superfamily protein

		PsCam043101_27130_869	50,9									
		PsCam038477_23510_469	50,8	FD_CC	7,21E- 18	DNA-binding PD1-like protein					1,07E- 29	AT-hook motif nuclear- localized protein 1
		PsCam001148_968_692	51,1	FD_CC			1,66E- 144	Pentatricopeptide repeat-containing protein	1,11E- 116	pentatricopeptide repeat-containing protein At2g40240, mitochondrial- like	4,80E- 78	Tetratricopeptide repeat (TPR)-like superfamily protein
F	VI.6	PsCam006586_4888_245	51,1	FD_CC								
		PsCam020872_11644_1756	51,1	FD_CC								
		PsCam033562_19160_307	51,1	FD_CC			0	PHD finger protein	0	PHD finger protein At1g33420-like isoform X1	0	RING/FYVE/PHD zinc finger superfamily protein
		PsCam044418_28295_579	51,1	FD_CC			2,59E- 75	Peptide chain release factor			2,00E- 73	Class I peptide chain release factor
ŀ		PsCam025448_14564_2314	51,1	FD_CC								
	VI.7	PsCam035454_20637_2339	51,1				0	Chaperone protein dnaJ			4,68E- 145	DNAJ heat shock N-terminal domain-containing protein
		PsCam036704_21832_970	51,1	FD_CC					0	pentatricopeptide repeat-containing protein At1g33350-like	0	Pentatricopeptide repeat (PPR) superfamily protein
		PsCam053189_35238_569	51,1	FD_CC								
		PsCam038677_23691_325	51,1	FD_CC	1,09E- 46	hsr203J homolog	6,56E- 178	Neutral cholesterol ester hydrolase	1,07E- 142	probable carboxylesterase 18-like	4,82E- 116	carboxyesterase 18
		PsCam035356_20546_778	52,5	FD_CC	0	cytosine-5 DNA methyltransferase	0	DNA (cytosine-5)- methyltransferase	2,96E- 59	DNA (cytosine-5)-methyltransferase CMT2-like isoform X3	0	methyltransferase 1
ŀ		PsCam040326_25099_1923	52,5	FD_CC								
	VI.8	PsCam011774_8038_200	52,8	FD_CC	8,04E- 25	brassinosteroid receptor	2,74E- 126	ССР	1,72E- 108	probable LRR receptor-like serine/threonine-protein kinase At3g47570-like isoform X2	1,90E- 68	Leucine-rich repeat protein kinase family protein
		PsCam023852_13541_264	52,8	FD_CC								
		PsCam026055_15000_192	52,8	FD_CC								
II		PsCam001108_940_48	72,9	FD_Field_date2, FD_Field_date3	3,82E- 96	beta-1,3- glucanase	6,85E- 169	Endo-beta-1 3- glucanase	2,40E- 156	glucan endo-1,3-beta-glucosidase- like	4,20E- 97	beta-1,3-glucanase 2
F		PsCam050827_33387_242	82,4	FD_Field_date2								
		PsCam037927_22984_97	82,5	FD_Field_date2	3,02E- 159	endo-1,4-beta- glucanase	3,07E- 164	Endo-beta-1 4- glucanase	0	endoglucanase 8-like	4,84E- 173	glycosyl hydrolase 9B1

	PsCam038378_23415_721	86,4	FD_CC								
	PsCam042427_26489_524	87,1	FD_CC			3,94E- 148	Serine/arginine repetitive matrix protein	2,00E- 84	serine/arginine repetitive matrix protein 1-like isoform X1	9,89E- 67	splicing factor PWI domain- containing protein
VII.	PsCam000301_265_1519	87,9	FD_Field_date2, FD_Field_date3					0	mitochondrial tRNA-specific 2- thiouridylase 1-like isoform X1	0	transferases;tRNA (5- methylaminomethyl-2- thiouridylate)- methyltransferases
	PsCam045010_28743_3365	88	FD_Field_date2							0	Putative glycosyl hydrolase of unknown function (DUF1680)
	PsCam004928_3732_3087	89,3	FD_Field_date2, FD_Field_date3			1,82E- 110	Beta-glucosidase G2	0	probable serine/threonine-protein kinase GCN2-like	0	protein kinase family protein

Additional file 13: Table S7. The correspondence of the genotyping results of accessions at the Hr locus (Elf3 SNP marker).

CRB Code	Accession Name / Donor Code	End-use	Cultivation status	Geographical origin	Sowing type	Cluster	ELF3- Marker
CRB099	HOLLY 13	fodder	breeding	North-Europe	winter	1	Hr
CRB212	CLAVIER 2	garden	breeding	Unknown	spring	1	Hr
CRB122	Aa86	Unknown	breeding	North-Europe	Unknown	1	Hr
CRB145	0-2072	garden	breeding	North- America	spring	1	hr
CRB147	0-2093	garden	breeding	North- America	spring	1	hr
CRB133	74SN5	preserve	breeding	North- America	Unknown	1	hr
CRB186	ROSAKRONE	fodder	cultivar	North-Europe	spring	1	Hr
CRB262	NFG KRUPP PELUSCHKE	fodder	cultivar	North-Europe	spring	1	Hr
CRB244	SINIKKA	garden	cultivar	North-Europe	spring	1	Hr
CRB355	GREITUKAI K4834	Unknown	cultivar	North-Europe	spring	1	Hr
CRB148	PI179461	Unknown	cultivar	Middle-East	Unknown	1	Hr
CRB194	BEAN PEA	Unknown	cultivar	Unknown	Unknown	1	Hr
CRB365	E11	Unknown	cultivar	Unknown	Unknown	1	Hr
CRB009	KOROZA	field	cultivar	North-Europe	spring	1	hr
CRB354	SPRUT K8349	field	cultivar	North-Europe	spring	1	hr
CRB414	VIRGO	field	cultivar	North-Europe	spring	1	hr
CRB442	TORSDAG	field	cultivar	North-Europe	spring	1	hr
CRB063	BOOSTER	field	cultivar	North-Europe	winter	1	hr
CRB253	TIMO	fodder	cultivar	North-Europe	spring	1	hr
CRB273	LIVIOLETTA	fodder	cultivar	Unknown	spring	1	hr
CRB219	ALGERA	fodder	cultivar	Africa	Unknown	1	hr
CRB274	SETCHEY	fodder	cultivar	Unknown	Unknown	1	hr
CRB184	OSZI TKB	fodder	cultivar	North-Europe	winter	1	hr
CRB451	CHAMPION D'ANGLETERRE	garden	cultivar	Unknown	spring	1	hr
CRB014	WBH 1458	Unknown	cultivar	North-Europe	Unknown	1	hr
CRB127	CAPELLA	Unknown	cultivar	North-Europe	Unknown	1	hr
CRB179	AUSTRALIA	Unknown	cultivar	Unknown	Unknown	1	hr
CRB347	YELLOW POLLEN	Unknown	germplasm	Unknown	Unknown	1	Hr
CRB348	PINKISH WHITE	Unknown	germplasm	North-Europe	Unknown	1	Hr
CRB309	CERISE-ce,CR	fodder	germplasm	North-Europe	spring	1	hr
CRB308	SKOROSPELYI GIBRID N92 K4819	field	local population	North-Europe	spring	1	Hr
CRB357	ZHALYUKAI K4269	field	local population	North-Europe	spring	1	Hr
CRB358	ROSTOVSKII VYSOKII K4356	field	local population	North-Europe	spring	1	Hr
CRB364	K4170	fodder	population	North-Europe	spring	1	Hr
CRB113	WNC 23Z	fodder	local population	Unknown	winter	1	Hr
CRB324	GUISANTE2	garden	population	South-Europe	spring	1	Hr
CRB325	GUISANTE1	garden	local population	South-Europe	spring	1	Hr
CRB326	ARVELLO1	garden	population	South-Europe	spring	1	Hr
CRB331	GUISANTE	garden	local population	South-Europe	spring	1	Hr

CRB359	KHIRPOSINSKII K5127	Unknown	local population	North-Europe	spring	1	Hr
CRB361	K1666	Unknown	local population	North-Europe	spring	1	Hr
CRB363	K4088	Unknown	local population	North-Europe	spring	1	Hr
CRB025	Pisum sativum-USSR JI2200	Unknown	local population	North-Europe	Unknown	1	Hr
CRB304	VAVILOV-D265	fodder	local population	North-Europe	spring	1	hr
CRB200	HAITI COLORE	fodder	local population	South- America	Unknown	1	hr
CRB005	CHINE-D371	Unknown	local population	Far-East	Unknown	1	hr
CRB346	MONGOLIA	Unknown	local population	Far-East	Unknown	1	hr
CRB352	Pisum sativum-ETHIOPIA JI1431	Unknown	local population	Africa	Unknown	1	hr
CRB128	HOHENHEIMER PINK- FLOWERED PI180693	fodder	Unknown	North-Europe	Unknown	1	Hr
CRB180	ENGLISH	fodder	Unknown	North-Europe	Unknown	1	Hr
CRB205	CAPSICUM	fodder	Unknown	Middle-East	Unknown	1	Hr
CRB124	PI273279	Unknown	Unknown	South- America	Unknown	1	Hr
CRB188	TASMANIE	Unknown	Unknown	Unknown	Unknown	1	Hr
CRB191	POL 6806	Unknown	Unknown	Unknown	Unknown	1	Hr
CRB207	POIS IRANIEN 57 B	Unknown	Unknown	Middle-East	Unknown	1	Hr
CRB123	PI271936	Unknown	Unknown	North-Furone	Unknown	1	hr
CRB034	Pisum Flatius II1703	Unknown	wild	Unknown	Unknown	1	Hr
CRB137	BEZEL VA	Unknown	cultivar	Middle-East	Unknown	2	Hr
CRB157	MOSHONG 3	Unknown	cultivar	Middle-East	Unknown	2	hr
CRB350	HAMMARI UNDS K LINE II1/6	Unknown	germplasm	South-Europe	Unknown	2	Hr
CRB304	IIAMMAREONDS & EINE JI140	Unknown	landrace	Unknown	Unknown	2	111 hr
CRD574	51105	Ulikilowii	local		UIKIIOWII	2	111
CRB310	CHINA	field	population local	Far-East	winter	2	Hr
CRB020	Pisum sativum PLP 70 J11267	Unknown	population	Far-East	Unknown	2	Hr
CRB039	ATAR JI2202	Unknown	population	Middle-East	Unknown	2	Hr
CRB040	WILD TUNESIAN JI2263	Unknown	population	Africa	Unknown	2	Hr
CRB330	Pisum sativum-ETHIOPIA JI281	Unknown	population	Africa	Unknown	2	Hr
CRB398	WIRAIG JI190	Unknown	population	Africa	Unknown	2	Hr
CRB203	BRAUN YEMEN	fodder	Unknown	Middle-East	Unknown	2	Hr
CRB263	URSS 2759	fodder	Unknown	North-Europe	Unknown	2	Hr
CRB201	AFGHANISTAN ASIATICUM	fodder	wild	Middle-East	Unknown	2	Hr
CRB202	NEPAL A	fodder	wild	Far-East	Unknown	2	Hr
CRB204	ABYSSINICUM VAVILOVANIUM	fodder	wild	Unknown	Unknown	2	Hr
CRB403	L808	fodder	wild	Africa	Unknown	2	Hr
CRB015	Pisum sativum-AFGHANISTAN JI86	Unknown	wild	Middle-East	Unknown	2	Hr
CRB047	Pisum Transcaucasicum JI2546	Unknown	wild	South-Europe	Unknown	2	Hr
CRB351	Pisum Transcaucasicum JI45	Unknown	wild	South-Europe	Unknown	2	Hr
CRB381	Pisum Elatius JI1089	Unknown	wild	Middle-East	Unknown	2	Hr
CRB399	Aa82 JI252	Unknown	wild	Africa	Unknown	2	Hr
CRB400	JI261	Unknown	wild	Unknown	Unknown	2	Hr
CRB031	Pisum Elatius JI1075	Unknown	wild	Middle-East	winter	2	Hr

	nknown	2	hr
CRB374 Pisum Humile JI241 Unknown wild Middle-East Ur	nknown	2	hr
CRB066 BUL179 garden breeding North-Europe spr	oring	3	hr
CRB11990-2079gardenbreedingNorth- Americaspin	oring	3	hr
CRB130552gardenbreedingNorth- Americaspin	oring	3	hr
CRB13590-2131gardenbreedingNorth- Americaspin	oring	3	hr
CRB1430-2016gardenbreedingNorth- Americaspin	oring	3	hr
CRB144 0-2042 garden breeding North- America spr	oring	3	hr
CRB146 0-2073 garden breeding North- America spr	oring	3	hr
CRB415 Wav F502 garden breeding North-Europe spi	oring	3	hr
CRB416 Wav F750 garden breeding North-Europe spi	oring	3	hr
CRB291 MULTIRESISTANT garden breeding North- America	nknown	3	hr
CRB084 GSP1016 preserve breeding North- America Ur	nknown	3	hr
CRB085 GSP1017 preserve breeding North- America Ur	nknown	3	hr
CRB086 GSP1023 preserve breeding North- America Ur	nknown	3	hr
CRB087 GSP1078 preserve breeding North- America Ur	nknown	3	hr
CRB088 MN314 preserve breeding North- America Ur	nknown	3	hr
CRB089 GSP1091 preserve breeding North- America Ur	nknown	3	hr
CRB090 GSP1097 preserve breeding North- America Ur	nknown	3	hr
CRB132 74SN4 preserve breeding North- America Ur	nknown	3	hr
CRB129 MN313 Unknown breeding North- America Ur	nknown	3	hr
CRB196 POIS AMERICAIN COUSIN Unknown breeding America Ur America	nknown	3	hr
CRB228 R 4-53 DJNOVIC Unknown breeding South-Europe Ur	nknown	3	hr
CRB293 st100 Unknown breeding Unknown Ur	nknown	3	hr
CRB294 st+at+tl Unknown breeding Unknown Ur	nknown	3	hr
CRB076 SENTINEL fodder cultivar North-Europe spi	oring	3	hr
CRB223 IRIFFID fodder cultivar Unknown spi	oring	3	hr
CRB007 FELTHAM FIRST garden cultivar North-Europe spi	oring	3	hr
CRB134 DARK SKIN PERFECTION garden cultivar North-Europe spi	oring	3	hr
CRB134SUNDANCEgardencultivarNorth-Europesp.CRB229TRAPPERgardencultivarNorth- Americasp.	oring	3	nr hr
CRB233 AUTOMOBILE garden cultivar North-Europe sp	ring	3	hr
CRB234 TEMPO garden cultivar North-Europe sp	ring	3	hr
CRB236 KOMPACTA garden cultivar South-Europe sp	ring	3	hr
CRB247 LAFONTAINE garden cultivar North-Europe sp	ring	3	hr
CRB247 LA ONTAINE garden cultivar North-Europe sp.	ring	3	hr
CRB249 TELEPHONE A RAMES garden cultivar North-Europe sp	ring	3	hr
CRB254 TRIANON garden cultivar North-Europe sp	ring	3	hr
CRB259 BURPEANA garden cultivar Unknown sn	ring	3	hr
CRB270 PETIT PROVENCAL garden cultivar North-Furone sn	ring	3	hr
CRB276 CHIPEAU garden cultivar North-Europe sp	ring	3	hr
CRB277REMINIgardencultivarNorth- Americaspi	oring	3	hr

CRB278	DOORDRAGER	garden	cultivar	North-Europe	spring	3	hr
CRB279	ETA	garden	cultivar	North-Europe	spring	3	hr
CRB285	MIRAVIL	garden	cultivar	North-Europe	spring	3	hr
CRB287	MINGOMARK	garden	cultivar	North-Europe	spring	3	hr
CRB289	MERVEILLE DE WEISSENFELS	garden	cultivar	North-Europe	spring	3	hr
CRB290	MIDFREEZER	garden	cultivar	North- America	spring	3	hr
CRB306	CAMÉOR	garden	cultivar	North-Europe	spring	3	hr
CRB323	VERCAS	garden	cultivar	North-Europe	spring	3	hr
CRB334	GRADUS	garden	cultivar	North-Europe	spring	3	hr
CRB335	JUWEL	garden	cultivar	North-Europe	spring	3	hr
CRB336	MINI	garden	cultivar	North- America	spring	3	hr
CRB337	LINCOLN	garden	cultivar	North-Europe	spring	3	hr
CRB340	WISCONSIN MERIT	garden	cultivar	North- America	spring	3	hr
CRB341	MERVEILLE DE KELVEDON	garden	cultivar	North-Europe	spring	3	hr
CRB342	STARCOVERT	garden	cultivar	North-Europe	spring	3	hr
CRB407	MULTIFREEZER	garden	cultivar	North- America	spring	3	hr
CRB412	PUGET	garden	cultivar	Unknown	spring	3	hr
CRB418	ALASKOR	garden	cultivar	North-Europe	spring	3	hr
CRB445	VITALIS	garden	cultivar	North-Europe	spring	3	hr
CRB452	PLEIN LE PANIER	garden	cultivar	North-Europe	spring	3	hr
CRB190	KIRIN 40	mangetout	cultivar	Far-East	spring	3	hr
CRB218	SUGAR SNAP	mangetout	cultivar	North- America	spring	3	hr
CRB131	PH14-119	preserve	cultivar	North- America	Unknown	3	hr
CRB019	WISCONSIN PERFECTION	Unknown	cultivar	North- America	Unknown	3	hr
CRB118	MN108	Unknown	cultivar	North- America	Unknown	3	hr
CRB155	MINI 93	Unknown	cultivar	North- America	Unknown	3	hr
CRB157	WSU 23	Unknown	cultivar	North- America	Unknown	3	hr
CRB158	WSU 28	Unknown	cultivar	North- America	Unknown	3	hr
CRB159	VANTAGE	Unknown	cultivar	Unknown	Unknown	3	hr
CRB029	USATYJ JI113	Unknown	germplasm	North-Europe	Unknown	3	Hr
CRB073	X78122-sbm-SCOUT	preserve	germplasm	North- America	spring	3	hr
CRB023	PESOL JI1831	Unknown	local population	South-Europe	Unknown	3	hr
CRB141	PI471171	Unknown	Unknown	Far-East	Unknown	3	hr
CRB195	CCO 2	Unknown	Unknown	South-Europe	Unknown	3	hr
CRB209	CEYLAN	Unknown	Unknown	Unknown	Unknown	3	hr
CRB297	776	field	breeding	North-Europe	spring	4	hr
CRB301	FIGARO	field	breeding	North-Europe	spring	4	hr
CRB377	661	field	breeding	North-Europe	spring	4	hr
CRB382	CD5MM	field	breeding	North-Europe	spring	4	hr
CRB410	PF31905	field	breeding	europe	spring	4	hr
CRB411	PF35323	field	breeding	europe	spring	4	hr
CRB070	GA91H2	field	breeding	Unknown	winter	4	hr
CRB378	320*247=BP	field	breeding	North-Europe	winter	4	hr
CRB001	ABAQUE	field	cultivar	North-Europe	spring	4	hr
CRB002	BIRTE	field	cultivar	North-Europe	spring	4	hr

CRR010	LUMINA	field	cultivar	North-Furone	spring	4	hr
CRB065	BRIDGE	field	cultivar	North-Europe	spring	4	lli hr
CRB078	CELESTE	field	cultivar	North-Europe	spring	4	hr
CRB108	PURSAN	field	cultivar	North-Europe	spring	4	hr
CRB110	SVDNEV	field	cultivar	North-Europe	spring	4	hr
CRB115	YSARA	field	cultivar	North-Europe	spring	4	hr
CRB307	BALLET	field	cultivar	North-Europe	spring	4	hr
CRB315	FIFFI	field	cultivar	North-Europe	spring	4	hr
CRB317	MESSIDE	field	cultivar	North Europe	spring	4	lli hr
CRB3/0	PROGRETA	field	cultivar	North-Europe	spring	4	hr
CRB362	NOPD K 8200	field	cultivar	North Europe	spring	4	lli hr
CRB380	DANTO	field	cultivar	europe	spring	4	hr
CRB402	KATRIN	field	cultivar	North-Europe	spring	4	hr
CRB402	OPTIMA	field	cultivar	North Europe	spring	4	lli hr
CRD400	TÉDÈSE	field	cultivar	North-Europe	spring	4	lli hr
CRD413		field	cultivar	North Europa	spring	4	lli hr
CRD423		field	cultivar	North Europe	spring	4	lli hr
CRD452	VERTOR	field	cultivar	North Europe	Unknown	4	lli hr
CRD112		field		North Europe	Ulikilowii	4	111 1
CRD420		field	cultivar	North Europe	winter	4	lll hr
CRD420	AKAVIS CHEVENNE	field	cultivar	North Europe	winter	4	lll hr
CRD430		foddor	cultivar	North Europe	witter	4	lll hr
CRD222		fodder	cultival	North Europe	spring	4	111 1
CRB207	BITMO	rodder	cultivar	North Europe	Unknown	4	nr hr
CRB109		garden	cultivar	North Europe	spring	4	nr ha
CRB312	FINETIE	garden	cultivar	North-Europe	spring	4	nr 1
CKB5/6	KONDO	garden	cultivar	North-Europe	spring	4	nr
CDD 420	MID 10	TT.1	. 1/:	TT.1		4	1
CRB439	MIR 12	Unknown	cultivar	Unknown North	spring	4	hr
CRB439 CRB116	MIR 12 US IDAHO W6 8955	Unknown Unknown	cultivar cultivar	Unknown North- America	spring Unknown	4	hr hr
CRB439 CRB116 CRB198	MIR 12 US IDAHO W6 8955 Agat	Unknown Unknown Unknown	cultivar cultivar cultivar	Unknown North- America North-Europe	spring Unknown Unknown	4 4 4	hr hr hr
CRB439 CRB116 CRB198 CRB074	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp	Unknown Unknown Unknown fodder	cultivar cultivar cultivar germplasm	Unknown North- America North-Europe Unknown	spring Unknown Unknown Unknown	4 4 4 4	hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122	Unknown Unknown Iodder Unknown	cultivar cultivar cultivar germplasm germplasm	Unknown North- America North-Europe Unknown Unknown	spring Unknown Unknown Unknown Unknown	4 4 4 4 4	hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2	Unknown Unknown Iodder Unknown Unknown	cultivar cultivar cultivar germplasm germplasm local population	Unknown North- America North-Europe Unknown Unknown Far-East	spring Unknown Unknown Unknown Unknown	4 4 4 4 4 4	hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22	Unknown Unknown fodder Unknown Unknown field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe	spring Unknown Unknown Unknown Unknown winter	4 4 4 4 4 4 5	hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317	Unknown Unknown fodder Unknown Unknown field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe	spring Unknown Unknown Unknown Unknown winter winter	4 4 4 4 4 4 5 5	hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500	Unknown Unknown fodder Unknown Unknown field field field	cultivar cultivar germplasm germplasm local population breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe	spring Unknown Unknown Unknown Unknown winter winter winter winter	4 4 4 4 4 5 5 5 5	hr hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194	Unknown Unknown fodder Unknown Unknown field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding	Unknown America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031	Unknown Unknown fodder Unknown Unknown field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085	Unknown Unknown fodder Unknown Unknown field field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099	Unknown Unknown fodder Unknown Unknown field field field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Winter winter winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP	Unknown Unknown fodder Unknown Unknown field field field field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown winter winter winter winter winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437	Unknown Unknown fodder Unknown Unknown field field field field field field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Winter winter winter winter winter winter winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21	Unknown Unknown fodder Unknown Unknown field field field field field field field field field garden	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Unknown winter winter winter winter winter winter winter winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R	Unknown Unknown fodder Unknown Unknown field field field field field field field field garden garden	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Unknown winter winter winter winter winter winter winter winter winter winter winter winter winter winter winter winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB161	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100	Unknown Unknown fodder Unknown Unknown field field field field field field field field garden garden garden	cultivar cultivar cultivar germplasm germplasm local population breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Unknown winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB162 CRB183	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP	Unknown Unknown fodder Unknown Unknown field field field field field field field field garden garden garden field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Unknown Winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB161 CRB162 CRB383 CRB388	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP M163*247=CP	Unknown Unknown fodder Unknown Unknown field field field field field field field garden garden garden field field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe Unknown	spring Unknown Unknown Unknown Unknown Unknown winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB161 CRB162 CRB383 CRB388 CRB406	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP M163*247=CP MERCURE	Unknown Unknown fodder Unknown Unknown field field field field field field field garden garden garden field field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe Unknown	spring Unknown Unknown Unknown Unknown Unknown winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB162 CRB162 CRB383 CRB388 CRB406 CRB006	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP M163*247=CP MERCURE DOVE	Unknown Unknown fodder Unknown Unknown field field field field field field field garden garden garden field field field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe Unknown Unknown	spring Unknown Unknown Unknown Unknown Unknown Winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB162 CRB383 CRB388 CRB406 CRB006 CRB102	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP M163*247=CP MERCURE DOVE ICEBERG	Unknown Unknown fodder Unknown Unknown field field field field field field field field garden garden garden field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe Unknown Unknown Unknown	spring Unknown Unknown Unknown Unknown Unknown Winter winter	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h
CRB439 CRB116 CRB198 CRB074 CRB075 CRB012 CRB053 CRB054 CRB055 CRB163 CRB166 CRB168 CRB169 CRB379 CRB380 CRB160 CRB161 CRB162 CRB383 CRB388 CRB406 CRB102 CRB102 CRB103	MIR 12 US IDAHO W6 8955 AGAT ORANGE POD-orp ARTHRITIC-art JI2122 KATHMANDOU DISTRICT 2 86-22 87-317 87-500 194 5031 5085 5099 C413=AP C437 21 34R 100 CE101=FP M163*247=CP MERCURE DOVE ICEBERG MISTRAL	Unknown Unknown fodder Unknown Unknown field field field field field field field garden garden garden field field field field field field field field field field field	cultivar cultivar cultivar germplasm germplasm local population breeding	Unknown North- America North-Europe Unknown Far-East North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe North-Europe Unknown Unknown Unknown North-Europe North-Europe	spring Unknown Unknown Unknown Unknown Unknown Winter wint	4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	hr hr hr hr hr hr hr hr hr hr hr hr hr h

hr

hr

hr

hr

hr

hr

hr

hr

hr

hr hr

hr

hr

hr

hr

Hr

Hr

hr

hr

hr

hr

hr

hr hr

hr

hr

hr

hr

hr

hr

hr

hr hr

hr

hr hr

hr

hr hr

hr

hr

hr

hr hr

hr

hr

hr

hr

hr

hr

hr

CRB316	FRISSON
CRB433	FRILENE
CRB438	KAZAR
CRB440	RAFALE
CRB444	VICTOR
CRB245	HATIVER
CRB333	FROGEL
CRB142	P58
CRB171	eM
CRB174	533
CRB296	667
CRB372	ULADOVSKIJ 10
CRB170	AA1
CRB172	aC
CRB295	tl100
CRB077	MAGNUS
CRB239	PYGMA
CRB300	AMINO
CRB375	CENNIA
CRB220	BLAUWSCHOVVED
CRD220	DESIDEE
CRD2/2	DESIREE
CRD203	
CKD101	IKEOI SAKOA
CRB018	ALASKA
CRB185	PRECOCE DE NAPLES
CRB231	ERSLING
CRB241	SOUCHET
CRB246	WICO
CRB250	FIN DE LA BIEVRE
CRB256	ROSAMU
CRB258	RODOGUNE
CRB268	GLOIRE DE CORREZE
CRB303	CADOR
CRB305	PRECLAMEX
CRB311	SOMMETTE
CRB321	PILET
CRB329	CLAUSELAN
CRB419	ALDOT
CRB448	CHEMIN LONG JI296
CRB449	CLAMART HATIF
CRB450	MICHAUX DE PARIS
CRB214	CAROUBEL
CRB216	NORMAND
CRB271	NORLI
CRB454	CAROUBY DE MAUSSANE
CRB117	RAN SEKEREC
CRB138	PI196023
CRB139	PI196033
CRB149	PI193841
CRB150	PI193846
CRB395	PRIEKULSKIJ-341-def JI1184

field	cultivar
field	cultivar
garden	cultivar
garden	cultivar
Unknown	Unknown
garden	breeding
Unknown	breeding
Unknown	breeding
Unknown	breeding
fodder	cultivar
garden	cultivar
field	cultivar
field	cultivar
fodder	cultivar
garden	cultivar
mangetou	cultivar
mangetout	cultivar
mangetout	cultivar
mangetout	cultivar
Unknown	cultivar

North-Europe	winter	5
North-Europe	winter	5
North-Europe	winter	5
North-Europe	winter	5
North-	winter	5
America	. ,	-
North-Europe	winter	5
North-Europe	winter	5
Unknown	Unknown	5
North-Europe	spring	6
Unknown	Unknown	6
North-Europe	spring	6
North-Europe	Unknown	6
North-Europe	winter	6
North-	spring	6
America South Europe	enring	6
North Europe	spring	6
North Europe	spring	0
North-Europe	spring	0
Unknown	spring	6
North-Europe	spring	6
North- America	spring	6
North-Europe	spring	6
North-Europe	spring	6
North-Europe	spring	6
Unknown	spring	6
North-Europe	spring	6
North Europe	spring	6
North Europe	spring	6
South Europe	Unknowe	6
A fries	Unknown	6
Africa	Unknown	0
Africa	Unknown	0
Africa	Unknown	6
Africa	Unknown	6
North-Europe	Unknown	6

CRB344	WELLENSIEK'S WHITE INDENT JI805	Unknown	germplasm	North-Europe	Unknown	6	Hr
CRB353	COSTA RICA	garden	germplasm	South- America	spring	6	hr
CRB356	HOHENHEIMER GRUNE VIKTORIA K4252	field	local population	North-Europe	spring	6	hr
CRB366	BENDI	garden	local population	Far-East	spring	6	hr
CRB367	ER BLANCH	garden	local population	Far-East	spring	6	hr
CRB368	HECAI	garden	local population	Far-East	spring	6	hr
CRB369	YANGWAN	garden	local population	Far-East	spring	6	hr
CRB370	WANDOU 2	garden	local population	Far-East	spring	6	hr
CRB371	WANDOU 1	garden	local	Far-East	spring	6	hr
CRB193	CUZCO 1	garden	local population	South- America	Unknown	6	hr
CRB360	K30	Unknown	local population	North-Europe	spring	6	hr
CRB003	CHINE-D367	Unknown	local population	Far-East	Unknown	6	hr
CRB004	CHINE-D368	Unknown	local population	Far-East	Unknown	6	hr
CRB011	KATHMANDOU DISTRICT 1	Unknown	local population	Far-East	Unknown	6	hr
CRB027	Pisum sativum-ZAMBIA JI2383	Unknown	local population	Africa	Unknown	6	hr
CRB213	KATMANDOU	Unknown	local population	Far-East	Unknown	6	hr
CRB386	Pisum sativum-ETHIOPIA JI174	Unknown	local population	Africa	Unknown	6	hr
CRB232	FINLANDE	garden	Unknown	Unknown	spring	6	hr
CRB120	K4797	Unknown	Unknown	North-Europe	Unknown	6	hr
CRB189	MAZARE	Unknown	Unknown	South-Europe	Unknown	6	hr
CRB206	MEXIC 1	Unknown	Unknown	South- America	Unknown	6	hr
CRB210	BUTARO 1A	Unknown	Unknown	Unknown	Unknown	6	hr
CRB016	Pisum sativum-BOLIVIA JI228	Unknown	wild	South- America	Unknown	6	hr
CRB026	Pisum sativum-ZAIRE JI2376	Unknown	wild	Africa	Unknown	6	hr
CRB140	PI358630	Unknown	wild	Africa	Unknown	6	hr
CRB345	Pisum sativum-ETHIOPIA JI1594	Unknown	wild	Africa	Unknown	6	hr
CRB384	CF100=GP	field	breeding	North-Europe	Unknown	7	Hr
CRB064	BOOSTER*AW7	field	breeding	North-Europe	winter	7	Hr
CRB106	NEVE*b195-5	field	breeding	North-Europe	winter	7	Hr
CRB125	FENN	fodder	breeding	North- America	Unknown	7	Hr
CRB058	BLIXT 7	fodder	breeding	North-Europe	winter	7	Hr
CRB059	BLIXT 104	fodder	breeding	North-Europe	winter	7	Hr
CRB060	BLIXT 109	fodder	breeding	North-Europe	winter	7	Hr
CRB062	BLIXT 484	fodder	breeding	North-Europe	winter	7	Hr
CRB067	CE*CF	fodder	breeding	North-Europe	winter	7	Hr
CRB092	HOLLY 3	fodder	breeding	North-Europe	winter	7	Hr
CRB093	HOLLY 4	fodder	breeding	North-Europe	winter	7	Hr
CRB094	HOLLY 5	fodder	breeding	North-Europe	winter	7	Hr
CRB095	HOLLY 6	fodder	breeding	North-Europe	winter	7	Hr
CRB097	HOLLY 8	fodder	breeding	North-Europe	winter	7	Hr
CRB098	HOLLY 9	fodder	breeding	North-Europe	winter	7	Hr

CRB100 HOLLY 17 CRB101 HOLLY 18 CRB424 BLIXT 460 CRB425 BLIXT 461 CRB426 BLIXT 468 CRB427 BLIXT 195 PI512066 CRB126 CRB429 CCA CRB437 KARNOBAT CRB056 ASSAS CRB385 CHAMPAGNE CRB405 MELROSE **CRB421** AUSTRIAN WINTER CRB431 FENN CRB435 GLACIER CRB105 NEVE CRB443 VENDEVIL CRB390 DP CRB409 P1259 CRB057 AW2 CRB068 COTE D'OR CRB107 PICAR CRB111 TOPOLIANSKI WNC Z61Z CRB114 CRB436 HOLLY 11 CRB017 Pisum sativum-GREECE JI267 CRB024 Pisum sativum-MEXICO JI1844 CRB446 WINTERBERGER CRB240 CLAUSE H4 **CRB061** BLIXT157 MUENCHENER BANATER **CRB072** WINTERERBSE CRB192 BLACK SEEDED **CRB091** HOLLY 2 CRB096 HOLLY 7 CRB182 MOROZOUSTOICHIVYL CRB397 KEERAUPEA CRB121 79-2022 CRB136 SHRAT **CRB401** MOSHONG 1 CRB035 Pisum Humile JI1794 CRB041 WT304 JI2473 CRB048 Pisum Speciosum-LIBYA JI2605 CRB151 PI212112

fodder breeding fodder breeding fodder breeding fodder breeding fodder breeding fodder breeding Unknown breeding field cultivar field cultivar fodder cultivar fodder cultivar fodder cultivar fodder cultivar fodder cultivar garden cultivar field cultivar field cultivar fodder germplasm Unknown germplasm local fodder population local Unknown population local Unknown population local Unknown population fodder Unknown fodder Unknown fodder Unknown fodder Unknown Unknown Unknown Unknown Unknown Unknown Unknown fodder landrace garden cultivar Unknown cultivar local Unknown population Unknown wild Unknown wild wild Unknown Unknown wild

North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North- America	Unknown	7	Hr
South-Europe	winter	7	Hr
South-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North- America	winter	7	Hr
Unknown	winter	7	Hr
North- America	winter	7	Hr
North- America	spring	7	Hr
North-Europe	winter	7	hr
North-Europe	winter	7	hr
North-Europe	winter	7	Hr
Unknown	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
North-Europe	winter	7	Hr
Unknown	winter	7	Hr
Unknown	winter	7	Hr
North-Europe	winter	7	Hr
South-Europe South-	winter	7	Hr
America	winter	7	Hr
North-Europe	winter	7	Hr
Unknown	spring	7	Hr
Unknown	winter	7	Hr
North-Europe	winter	7	Hr
Unknown	winter	7	Hr
Unknown	Unknown	7	Hr
Unknown	winter	7	Hr
North-Europe	winter	7	Hr
Unknown	Unknown	1	Hr/hr
North- America	spring	1	Unknown
Far-East	Unknown	2	Unknown
Middle-East	Unknown	2	Unknown
Middle-East	Unknown	2	Unknown
Middle-East	Unknown	2	Unknown
Africa	Unknown	2	Unknown
Middle-East	Unknown	2	Unknown

CRB082	GSP1013	preserve	breeding	North- America	Unknown	3	Unknown
CRB153	GRANT	garden	cultivar	North-Europe	spring	3	Unknown
CRB175	585	Unknown	breeding	North-Europe	spring	4	Unknown
CRB071	CHEYENNE	field	breeding line	Unknown	winter	4	Unknown
CRB165	418	field	breeding	North-Europe	winter	5	Unknown
CRB252	MERVEILLE D'ETAMPES (syn=SERPETTE GUILLOTEAUX)	garden	cultivar	North-Europe	spring	6	Unknown
CRB453	SERPETTE D'AUVERGNE	garden	cultivar	North-Europe	spring	6	Unknown
CRB455	CORNE DE BELIER	mangetout	cultivar	North-Europe	spring	6	Unknown
CRB265	PIVER	fodder	cultivar	Unknown	winter	7	Unknown
CRB299	AMAC	field	cultivar	North-Europe	winter	7	Unknown

Additional file 14: Table S8. Description of accessions cumulating 3 to 5 favourable haplotypes vs accessions cumulating 6 unfavourable haplotypes at the frost damage (FD)-associated loci of the GWA study.

Accessions are identified by their CRB code and accession name. At the linkage groups (LGs) I, II, III, V, VI and VII, the favourable haplotypes are shown by a yellow background and unfavourable haplotypes by a blue background, as described in the legend of Additional file 10. The same colour code has been used to describe the favourable (yellow background: Hr) and unfavourable (blue background: hr) allele for the Hr gene, as determined in [21]. The mean frost damage score observed in the field experiment as well as its standard error (SE) are given. The end-use, cultivation status, geographical origin, sowing type and cluster number are extracted from Additional file 1.

CRB Code	Accession Name	LGI	LGII	LGIII	LGV	LGVI	LGVII	Hr	Mean Frost Damage_Field	SE	End-use	Cultivation status	Geographical origin	Sowing type	Cluster (k)
CRB385	Champagne			•					0.53	0.16	fodder	cultivar	North-Europe	winter	7
CRB446	Winterberger								0.38	0.15	Unknown	local population	North-Europe	winter	7
CRB435	Glacier								0.42	0.14	garden	cultivar	North-America	spring	7
CRB068	Côte d'Or								0.43	0.12	fodder	local population	North-Europe	winter	7
CRB107	Picar								0.50	0.15	fodder	local population	North-Europe	winter	7
CRB064	Booster*aw7								0.20	0.10	field	breeding	North-Europe	winter	7
CRB405	Melrose								0.30	0.12	fodder	cultivar	North-America	winter	7
CRB058	Blixt 7								0.33	0.12	fodder	breeding	North-Europe	winter	7
CRB098	Holly 9								0.33	0.11	fodder	breeding	North-Europe	winter	7
CRB192	Black seeded								0.40	0.13	fodder	unknown	unknown	winter	7
CRB100	Holly 17								0.42	0.13	fodder	breeding	North-Europe	winter	7
CRB060	Blixt 109								0.52	0.16	fodder	breeding	North-Europe	winter	7
CRB071	Cheyenne								0.82	0.24	field	cultivar	North-Europe	winter	4
CRB231	Ersling								4.63	0.12	garden	cultivar	North-Europe	spring	6
CRB233	Automobile								4.43	0.16	garden	cultivar	North-Europe	spring	3
CRB375	Cennia								3.63	0.29	field	cultivar	North-Europe	spring	6
CRB214	Caroubel								3.14	0.27	mangetout	cultivar	unknown	spring	6

Annexe B : Données supplémentaires du chapitre IV

Annexe 1 : Protocole de récolte du matériel végétal pour la construction de banque BAC	189
Annexe 2 : Construction de banques BAC	190
Annexe 3 : Mise en culture Midi-prép BAC - méthode Macherey-Nagel	214
Annexe 4 : Caractérisation de clones BAC : digestion avec l'enzyme de restriction Not nigration en électrophorèse en champ pulsé	I et 218
Annexe 5 : Marquage d'une sonde par Random Priming	220
Annexe 6 : Hybridation de macroarrays	222
Annexe 7 : Marqueurs utilisés pour le criblage des deux banques BAC de pois (Champagn Férèse)	e et 227



Centre INRA de Toulouse - CNRGV

Annexes

Annexe 1 : Protocole de récolte du matériel végétal pour la construction de banque BAC

Preparation and harvest of plant material for BAC library construction

DNA is isolated from young seedlings or young leaves.

During the last 4 days before harvest plants must be grown in obscurity to avoid risk of chloroplastic contamination and starch accumulation.

40g of fresh tissue are required to produce the BAC library.

Please send us a picture of plant material at harvest time.

Harvesting protocol

- Dispose an aluminium sheet on ice.
- Cut young seedlings (0.5 cm up to the soil level) or leaves with scissors previously sterilized by 70° ethanol.
- As soon as seedlings/leaves are cut put them on the cooled aluminium sheet.

- When the harvested seedlings/leaves quantity reaches 10g shut the aluminium sheet and <u>write quickly on it the</u> <u>amount harvested</u>, the day of the harvest and the name of the cultivar. Then put this packet in liquid nitrogen and keep it in until the end of harvesting.

- Proceed the same way until harvesting of all seedlings/leaves.
- Consolidate all aluminium packets in a single plastic bag.
- Print and fill the traceability sheet (page 2) and staple it to the plastic bag.
- Store the aluminium packets at -80°C.



Annexe 2 : Construction de banques BAC

NB : Ce MO est rédigé en Anglais car il s'adresse à la fois au personnel du laboratoire mais également aux collaborateurs étrangers qui viennent se former à cette technique au CNRGV

1. Aim and scope

This document describes step by step the protocol used to construct plant BAC libraries at CNRGV.

Depending on the type of plant, nature and quality of tissues, needs in mean inserts size, various adjustments could have to be done.

2. References

Construction of Plant Bacterial Artificial Chromosome (BAC) Libraries: An Illustrated Guide. 2nd Edition (2002)

Daniel G. Peterson, Jeffrey P. Tomkins, David A. Frisch, Rod A. Wing and Andrew H. Paterson. Journal of agricultural genomics 5, 2000

An improved method for plant BAC library construction.

Luo M, Wing RA. Methods Mol Biol. 2003;236:3-20.

Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size.

Chalhoub B, Belcram H, Caboche M. Plant Biotechnol J. 2004 May;2(3):181-8.

3. Health and safety

At any step : wear gloves and lab coat

When using dangerous chemicals (notified in the following): work under Sorbonne hood

When using liquid nitrogen: wear special gloves and protection glasses

4. Constraints

The whole protocol last about 3 weeks from DNA extraction to BAC clones characterization. The schedule have to be set up with the technical responsible of CNRGV.

5. Registration and traceability

All steps must be registered in GenoLIMS.

Every BAC library production must be registered in eZpublish.

6. Protocol

See next page

EXTRACTION: Isolation of high molecular weight nuclear DNA

J1. ISOLATION OF NUCLEI

<u>1</u>.Prepare Solutions for 20g of vegetal material:

Keep these solutions at 4°C before and during the extraction

For 20 g of material, 400 ml of SEB buffer is enough.



Centre INRA de Toulouse - CNRGV

Annexes

Note : For species with a lot of sugar and polyphenol, it is possible to increase the PVP and Betamercaptoéthanol concentration until 2% of each and it is necessary to prepare 600ml.

Tampon SEB	400 ml	600ml
TKE (0.1M Tris Base, 1M KCl,0.1M EDTA, pH 9.4-9.5) autoclavé	40 ml	60 ml
Sucrose	67.8 g	102.72 g
Spermidine	0.400 mL	0.600 mL
Spermine tetrahydrochloride stocké à 4°C	0.14 g	0.21 g
Ascorbic acid	0.4 g	0.6 g
PVP40 (0,25%)	1 g	1.5 g
Sodium diethyldithiocarbamate trihydrate stocké à -20°C	0.51 g	0.78 g
H20 stérile – qsp <i>final volume</i>	Qsp 400 mL	Qsp 600 mL
Tampon SEB + Bétamercaptoéthanol à 0.2% v/v		
Tampon SEB	299.9 ml	499.9 ml
Betamercaptoéthanol	0.66 mL	1 mL
Final volume	Qsp 300 mL	Qsp 500 mL
Tampon SEB+BME/ 10%v/v Triton in SEB+BME		
Tampon SEB+BME	13.5 mL	
Triton 100X	1.5 mL	-
Final volume	15 ml	-
Tampon de lyse		
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé	100 ml	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine	100 ml	-
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml	100 ml 1 g 0.01 g	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml Ndiéthyldithiocarbamate 0.13% w/v	100 ml 1 g 0.01 g 0.13 g	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml Ndiéthyldithiocarbamate 0.13% w/v ascorbic acid 0,1 % w/v	100 ml 1 g 0.01 g 0.13 g 0,10g	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml Ndiéthyldithiocarbamate 0.13% w/v ascorbic acid 0,1 % w/v EGTA 6mM	100 ml 1 g 0.01 g 0.13 g 0,10g 0.228g	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml Ndiéthyldithiocarbamate 0.13% w/v ascorbic acid 0,1 % w/v EGTA 6mM L-Lysine 200 mM	100 ml 1 g 0.01 g 0.13 g 0,10g 0.228g 2.924g	
Tampon de lyse EDTA 0.5M pH 9.1 autoclavé Nlausyl sarcosine Protéinase K 0.1mg/ml Ndiéthyldithiocarbamate 0.13% w/v ascorbic acid 0,1 % w/v EGTA 6mM L-Lysine 200 mM Final volume	100 ml 1 g 0.01 g 0.13 g 0,10g 0.228g 2.924g 100 ml	

<u>1 H before to begin the extraction</u>

Get liquid nitrogen with cryoprotective gloves – glasses - at LIPM or LGC or UPS-CNRS – IFR40 Note in the table (by following the link below) the volume taken with the date:



Centre INRA de Toulouse - CNRGV

Annexes

Y:\DIFFUSER OUTILS\DOC ASSOCIES\Docs construction banques\Suivi N2 liquide_JF.xls

Precool a centrifuge with « godet » for falcon 50 ml at 4°C.

Prepare the material necessary for the extraction

1 large mortar with a large pestle; 4 falcons 50 ml; 8 squares of cheesecloth (one filter of cheesecloth = 2 squares of cheesecloth superimposed folded funnel - shaped held in place by a agraffe); 8 squares of miracloth (one filter of miracloth = 2 squares of miracloth superimposed folded funnel - shaped held in place by an agraffe); 4 little funnels; 1 beaker of 2L; 1 spatula ; a large sheet of absorbent paper = protected bench ;1 tray filled with ice

1. Take 20 g of frozen tissue out of the freezer.

2. Place the frozen tissue in a mortar containing liquid nitrogen. Grind the tissue to a powder with the mortar and pestle.

3. Transfer the powder into ice-cold **SEB+BME** in a beaker of 2L. For every gram of tissue, use approximately 10 ml of SEB+BME (*e.g.*, place 20 g of frozen tissue in 200 ml of ice-cold SEB+BME).

4. Place the mixture on ice for 12 minutes. During this time, swirl very gently the contents of the beaker at least every two minutes (each swirl time = 20 sec).

 \Box **Note**: Nuclei in sucrose-based buffers must be handled with extreme care. The absence of divalent cations coupled with the extreme osmotic conditions in the SEB+BME make the nuclei extremely fragile.

If roughly agitated, the nuclei will break.

5. Filter the homogenate through 2 layers of Cheesecloth and 2 layers of Miracloth on a funnel into 50 ml tubes maintained on ice.

6. Add 1/20 volume of **SEB+BME/Triton** to the flask(s), and leave it on ice for **10 min**. During this time, swirl the contents of the flask every two minutes.

7. Spin the tubes in a centrifuge at specific speed (according to the genome size) for 20 min (4°C).

Genome size (Mb)	100-250	250-1000	1000-5000	6000-10000	11000-15000	>15000
Speed (g)	3000	2400	1500	1200	900	650

8. Very gently decant the supernatants. Add 10 ml of SEB+BME to each pellet, and gently resuspend nuclei. Consolidate the nuclear suspensions into as few 50 ml tubes as possible. To each tube add SEB+BME to 50 ml, and centrifuge tubes at the same speed defined previously **for 15 min** (4°C).

 \Box **Note :** If the starting material looks full of sugar or secondary metebolites (like polyphenol), wash the nuclei at least 2 times and prepare 600 ml of SEB buffer.

9. Repeat step 8 until all of the nuclear suspensions has been consolidated into a single tube.

10. Decant the supernatant, and resuspend the nuclei in 20 ml of SEB (without β -mercaptoethanol).

Centrifuge as described above, and very gently remove all but 1-2 ml of the supernatant with a large pipettor.

□ □ **Note:** according to the nuclei concentration expected and the pellet size, you have to adjust the volume.

11. Gently resuspend the pellet in the residual SEB.

12. Proceed to II.

II. PREPARATION OF PLUGS



Centre INRA de Toulouse – CNRGV

SUPPLIES, EQUIPMENT, AND REAGENTS : SEB; Lysis Buffer; InCert agarose; plug molds

1. In a 50 ml tube, mix 0.15 g of **InCert** agarose with 10 ml of SEB. Transfer the flask into a microwave and heat the agarose/buffer mixture until all of the agarose has gone into solution (this requires various quick boilings, with interrupts). Place the flask in a 45°C water bath.

2. Place agarose plug molds on ice. If plug molds are being re-used, make sure that they are clean

(thoroughly washed and rinsed with ethanol or washed and UV-sterilized) and that fresh tape has been placed on the bottom of the wells.

3. Place the tube containing nuclei in the 45°C water bath for 10 minutes.

4. Mix an equal volume of the agarose solution with the pre-warmed nuclei. Using a 1ml pipet

with a relatively large-bore tip, place the nuclei/agarose mixture into the wells of the pre-chilled plug

molds (1 well volume = 80μ l). Place the ice bucket containing the plug molds in a refrigerator.

Allow plugs to solidify for 30 minutes.

5. Push plugs out of plug molds into 50 ml of Lysis Buffer. BioRad plug molds come with a small

plastic tab designed for pushing plugs out of the molds. If the nuclear isolation was performed as described above, the plugs should be white to light yellow in color.

6. Incubate the plugs in Lysis Buffer at 50°C for 24 hours.

J2. Washing of plugs

Drain the Lysis Buffer from the plugs, and

add 50 ml of fresh Lysis Buffer. Incubate the plugs at 50°C for an additional 24 hours.

J3. Washing, preservation of plugs and DNA quality control

I. WASHING STEPS

Washing buffer 1	
EDTA à 0.5 M - Ph 9.1	50 ml
Washing buffer 2	
EDTA à 0.05 M - Ph8	100 ml

Remove lysis buffer

Add 50ml of washing buffer 1 – Keep at 50°C during 1H

Empty the buffer 1, add **50 ml of Washing buffer 2**

Keep 1h at 4°C. Let the falcon lying in ice on the table with gentle agitation.

After, change a last time with fresh Washing buffer 2 and store at 4°C.

At this step DNA plugs can be stored at 4°C for weeks.

II. DNA QUALITY CONTROL

PURPOSE



Centre INRA de Toulouse - CNRGV

Annexes

In several cases it is recommended to test the DNA quality using a short electrophoresis program to determine if a purification can be useful to eliminate small or degraded DNA fragments.

The DNA must be tested if:

A high quantity of DNA is expected (for example : sunflower or wheat)

Every time the DNA of a new species is extracted

This step is **not recommended** when a low quantity of DNA is expected (for example: maize or *Arabidopsis*).

EXPERIMENTAL PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS: agarose plugs containing genomic DNA; 0.25X TBE; agarose; PFGE Lambda Ladder

1. Make a 1.0% agarose D3 (Euromedex) in 0.25X TBE using the large BioRad CHEF gel casting stand and the 15-tooth gel comb.

2. Cut one plug in 2 equal pieces and save one ½ plug in a 1.5ml Eppendorf tube with EDTA 0.05M pH8.

3. In the agarose gel, enlarge several wells to get 3 places following the figure below.

Place the other ¹/₂ plug using the pointed end of a spatula in the enlarge well n°1.

Don't seal the wells and don't add the ladder at this step.



4. Run the gel using DNA purification parameters: volts/cm = 5.0, included angle = 120°, initial switch time = 0.5 sec, final switch time = 1.5 sec, ramping = linear, running time = 1 hour.

5. After the migration, take off the gel and recover the purified $\frac{1}{2}$ plug with the pointed end of a spatula. Put it in the **enlarged well n°2**. Add the remaining $\frac{1}{2}$ plug (without purification) in the **enlarge well n°3** and the PFGE Lambda Ladder in the **two well from either side**.

6. Run the gel using the following parameters: volts/cm = 6.0, included angle = 120°, initial switch time = 1.0 sec, final switch time = 40.0 sec, ramping = linear, running time = 18 hours

7. Stain, destain, and examine the gel to determine the quantity and quality of the extracted DNA.

Workcases to determine if a purification is needed before the BAC library construction:



Centre INRA de Toulouse - CNRGV

Annexes



1: purified 1/2 plug - degradated/small DNA

- 2: purified ¹/₂ plug remaining DNA
- 3 : ¹/₂ plug no purification

<u>Case 1</u>: High quantity of DNA.and degradation. In the well $n^{\circ}1$ we can see the eliminated DNA and the remaining DNA in the well $n^{\circ}2$. **Purification recommended.**

Case 2: Low quantity of DNA. Purification not recommended even if we can see degradation on the gel.

8. If the purification is not recommended: skip to next step J4

If the purification is recommended: Proceed to J4 washes and purify 15 plugs following J3 II steps n° 1 to 4.

Use 3 plugs for the test restriction digest and the remaining 12 for 2 BAC libraries constructions (6 + 6). Make sure DNA plugs used for test restriction digest and BAC library construction were purified IN THE SAME EXPERIMENT.

J4. Preparation of plugs for restriction digest+ Digestion test

TE 10/10		
Reacts	Solution fille	TE 10/10
Tris-HCl Ph8 1M	10 mM	10 ml
EDTA Ph 8.0 0.5M	10 mM	20 ml
H20 stérile	-	1000 ml

PMSF solution		
Reacts	PMSF 0.1M	!! Note the date of manufacturing!!
PMSF	0.17g	Preservation max 15 days



Centre INRA de Toulouse – CNRGV

Annexes

Ethanol pur	10 ml	at -20°C

Washing buffer 3		
Stock solution	Volume	
PMSF 0.1M	100 µL	
TE 10/10	100 mL	

<u>. Preparation of plugs for Digestion test + restriction digest(Sizing I) – to do in the morning as soon as possible</u>

Refer to the previous tables for the preparation of buffers.

Select the number of desired plugs: 4 (for the digestion test) + x (for the sizing 1).

 \square **Note:** if you produced plugs that were concentrated (20 plugs from 20 g of plant material), you can use only 3 plugs for this test because the degraded DNA will be visible with a ¹/₄ plug

Transfer these plugs in a falcon having **50 ml of Washing buffer 3** and wash 1H on ice with gentle agitation.

Repeat this operation with other 50mL of Washing buffer 3.

Remove the Washing buffer 3, add 50 ml of TE 10/10 and wash 30 min on ice with gentle agitation.

Repeat this operation three times with other 50mL of TE10/10.

Finally fill the falcon with 50mL of TE10/10 and stock at 4°C.

Remove each time the solutions in the specific liquid PMSF waste container.

J4. Test restriction digest

PURPOSE

Construction of a BAC library requires generation of relatively high molecular weight restriction fragments. Such restriction fragments will serve as inserts in BAC construction. In general, fragments between 125 kb and 200 kb are desirable. To obtain fragments in this size range, the high molecular weight DNA in the agarose plugs must be partially digested with a restriction enzyme. To determine the conditions that yield a maximum percentage of fragments between 125 and 200 kb, a series of partial restriction digests is performed.

Be sure to proceed to the DNA quality test before the test restriction digest and to DNA purification if needed.

EXPERIMENTAL PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS : agarose plugs containing genomic DNA ; 0.25X TBE ; agarose ; PFGE Lambda Ladder ; H3M; 10X SB buffer (*HindIII, BamHI*) or 10X NEB4 buffer (*EcoRI*) ; enzyme *Hind*III, *BamH*I or *EcoRI* ; 70% ethanol ; microscope slides and coverglasses (slides and coverglasses should be sprayed with 70% ethanol and dried with a Kimwipe prior to use); CHEF gel apparatus ; regular CHEF gel casting stand ; 15-tooth gel comb

To prepare a new spermidine stock solution: see annex a


Annexes

	Buffer H3M or B1M	Buffer E1M
	For <i>HindIII</i> or <i>BamH1</i> restriction digest test	For <i>EcoRI</i> restriction digest test
Buffer SB 10x	1.3 ml	-
Buffer NEB4 10x	-	1.3 ml
Spermidine 40mM	1.3 ml	1.3 ml
BSA 20mg/ml	65 μL	65 μL
DTT 1M	13 μL	13 µL
Ultra-pure sterilized water	10.205 ml	10.205 ml
Final volume	13 ml	13 ml
	9.75 ml H3M for incubation + 13 x 250 µl for digestion	9.75 ml H3M for incubation + 13 x 250 µl for digestion

□ □ **Note :** Use the same tubes of enzyme and same lot number of BSA for the digestion test and for the sizing

1. Make a 1.0% agarose D3 (Euromedex) in 0.25X TBE using the **large BioRad CHEF gel casting stand and the 15-tooth gel comb.** Place the PFGE Lambda Ladder in lanes 1 and 15, respectively.

2. Transfer 4 of the plugs ("test plugs") into 9.75 ml of buffer. Incubate on ice for 1 hour under gentle shaking.

3. Place in 13 x 2 ml microfuge tubes in a tube rack, and label the tubes 1 to 13.

Add 250 µl of H3M to each tube.

4. Place tubes 1 to 13 on ice.

5. The next step (cutting of plugs) must be performed on ice. Using a spatula, transfer the four test plugs onto a microscope slide. Using a clean coverglass, cut each plug laterally and transversely to produce four pieces of roughly equal size. There should be now sixteen ¹/₄ plug pieces.

6. Still on ice, macerate (thoroughly chop) each 1/4 plug piece with a coverglass. (8 pieces / 1/4 plug).







7. Place 4 of the macerated $\frac{1}{4}$ plugs pieces into tube 1. Place each of the remaining macerated $\frac{1}{4}$ plug pieces in one of the remaining ten tubes (2-12).

8. In a 1.5 ml microfuge tube prepare the enzyme dilution:



Annexes

In the case of an HindIII or BamHI digestion

Dilution 1 -1U/µL	Volume (µL)
<i>HindIII</i> or <i>BamHI</i> 10U/µL	5
SB buffer 10x	5
Sterilized H2O	40
Final volume	50

Dilution 2- 0.1 U/µL	Volume (µL)
Dilution 1 -1U/µL	20
SB buffer 10x	20
Sterilized H2O	160
Final volume	200

In the case of an EcoRI digestion

Dilution 1 -1U/µL	Volume (µL)
EcoRI I 20U/µL	3
NEB4 buffer 10x	6
Sterilized H2O	51
Final volume	60

Dilution 2- 0.1 U/µL	Volume (µL)
Dilution 1 -1U/µL	20
NEB4 buffer 10x	20
Sterilized H2O	160
Final volume	200

9. Using **Dilution 2 of** *HindIII*, *BamHI* or *EcoRI*, add a specific quantity of *Hind*III, *BamHI* or *EcoRI* to each of the 13 tubes as directed in TABLE 1. Le point 0.6 est important !!!



Annexes

Tubes N°	Final concentration in the tube (U/mL)	Final quantity in the tube (U for 250µL)	Volume dilution 2 used (0.1 U/µl)	plug
1	0	0	0	1 plug
2	0	0	0	¹ / ₄ plug
3	0.2	0.05	0.5	
4	0.4	0.1	1	-
5	0.6	0.15	1.5	-
6	0.8	0.2	2	-
7	1	0.25	2.5	¹ / ₄ nlug
8	1.5	0.375	3.75	,4 prug
9	2	0.5	5	
10	3	0.75	7.5	
11	4	1	10	
12	5	1.25	12.5	
13	10	2.5	25	

10. Gently mix the contents of each tube and incubate tubes on ice for **1 hour** under gentle shaking. This allows the enzyme to infiltrate the agarose cubes.

11. Gently agitate each tube and place the tubes in a **37°C** water bath. Remove the tubes from the water bath after **EXACTLY 10 min**, and immediately place the tubes on ice.

12. Add 30 µl of 0.5 M EDTA (pH 8.0) to each tube (this inhibits further enzyme activity) and gently agitate tubes to promote contact between the agarose and the EDTA. Keep tubes on ice.

13. Using the pointed end of a spatula, transfer the undigested macerated 1 plug in tube 1 into well 2 of the gel. Likewise, transfer the macerated $\frac{1}{4}$ plug in tube 2 to well 3, the $\frac{1}{4}$ plug in tube 3 to well 4,

etc. Seal wells with leftover melted 1% TBE 0.25X agarose (50°C).

Run the gel using the following parameters: volts/cm = 6.0, included angle = 120°, initial switch time = 1.0 sec, final switch time = 40.0 sec, ramping = linear, running time = 18 hours.

14. Stain, destain, and examine the gel to determine the optimal enzyme concentration to produce fragments between 100 and 250 kb.

The 1-2 lanes should not show degraded DNA.

Help for choose the optimal concentration of enzyme



Annexes





Extraction of 20g 25/09/2012

36 plugs producted

J5. Restriction digest for SizingI

PURPOSE

Once the optimal conditions for producing fragments between 100 and 250 kb are determined, a mass digestion using several plugs is performed. The partially digested DNA from these plugs then can be used as insert DNA in construction of a BAC library.

 \square **Note :** avoid to load too much DNA because small DNA used to be stuck with the large one. A typical insert DNA quantity recovered is between 100 and 300 ng.

EXPERIMENTAL PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS : agarose plugs containing genomic DNA; H3M; 10X SB buffer (*HindIII, BamHI*) or 10X NEB4 buffer (*EcoRI*) ; *HindIII, BamHI* or *EcoRI*; microscope slides and coverglasses (slides and coverglasses should be sprayed with 70% ethanol and dried with a Kimwipe prior to use)

1. Transfer the **PMSF-treated plugs** (5-10 plugs) into a 50 ml tube containing **20 ml of H3M, B1M or E1M.** Incubate on ice for **1 hour** under gentle shaking.

2. On ice, macerate each plug using clean coverglasses (~ 32-40 pieces per plug). Place all macerated plugs into 2 ml tubes containing 1 ml of H3M, B1M or E1M (1 plug per tube).

	H3M or B1M			E1M
	For <i>HindIII</i>	or <i>BamHI</i> Si	zing 1	For <i>EcoRI</i> Sizing 1
Plug nb	<7	8-12	13-20	
Buffer SB 10x	2 ml	3 ml	4 ml	-



Annexes

Buffer NEB2.1 10x			-	4 ml
Spermidine	2 ml	3 ml	4 ml	4 ml
BSA	100 µl	150 µl	200 µl	200 µl
DTT	20 µl	30 µl	40 µl	40 µl
Ultra-pure sterilized water	15.88 ml	23.82 ml	30.76 ml	30.36 ml
Final volume	20 ml	30 ml	40 ml	40 ml

3. Add HindIII, BamHI or EcoRI so that the final enzyme concentration is the same as the optimal enzyme

concentration determined. USE THE SAME TUBE OF ENZYME, BSA AND SAME TYPE OF BUFFER

USED IN THE TEST DIGEST!

NB: Several different concentrations around the optimal concentration are usually chosen (e.g., 1.5, 2 and 2.5

U/ml if the optimal concentration is 2U/ml)

Example:

Optimal final Concentration of enzyme U/mL per tube	Volume dilution 2 of enzyme at 0.1 U/mL add at each tube containing 1mL of H3M, B1M or E1M	Number of plugs choosen
1	10 μL	5
1.5	15 μL	4

4. Gently mix, and incubate the tube on ice for 1 hour under gentle shaking.

5. Place in a 37°C water bath. Remove the tube from the water bath after EXACTLY 10 min and place on ice.

6. Immediately add 150µl 0.5 M EDTA (pH 8.0) per tube(this inhibits further enzyme activity)

and gently agitate to promote contact between the agarose and the EDTA. Keep on ice

\Box \Box Proceed with the first size selection

J6. First size selection

PURPOSE

Now that the DNA in the plugs has been digested, it is important to check and make sure that the

DNA is of an appropriate length for BAC library construction. DNA longer than 210 kb and much of the DNA < 125 kb is removed during this first size selection. Conversely, most of the DNA between 125 and 210 kb is sequestered and taken through a second size selection.

 \square \square **Note :** be careful to the quantity and migration. Shift the window if too much material

EXPERIMENTAL PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS) : agarose plugs containing partially-digested genomic DNA; 0.25X TBE; agarose; PFGE Lambda Ladder; scalpel ; CHEF gel apparatus; regular CHEF gel casting stand; 15-tooth gel comb

1. Make a 1.0% Seakem Gold agarose gel in 0.25X TBE using the large BioRad CHEF gel casting stand and

the 15- tooth gel comb.

2. Using a clean scalpel, cut and remove the agarose between 4-5 of the center wells of the gel to produce **one large "slot well"**. Make sure that the agarose lining the bottom of the wells is removed and that the leading edge



of the slot is parallel to the leading edge of the wells (FIGURE 1a). The plug pieces will be loaded into the slot well.

3. Add a small amount of melted agarose to the slot well so that the bottom of the well is completely sealed. Allow the agarose to solidify.

4. Collect the plug pieces from the digestion tubes. Using the pointed end of a spatula, transfer the macerated, digested plug pieces into the slot well. Press the pieces up against the leading edge of the slot well (FIGURE 1b).

5. Place the PFGE Lambda Ladder in lanes flanking the slot well (FIGURE 1b).

6. Seal the slot well and the wells containing ladder DNA with melted agarose. Allow the agarose to solidify.

7. Remove the gel from the casting stand, and wipe any agarose off the bottom of the base plate.

8. Place the base plate and overlying gel in the BioRad CHEF electrophoresis chamber. The unit **should be cleaned**

(3 baths of 15min: the first with 0.01M NaOH, and the last two with distilled water) and finally should contain 2.5 L of fresh 0.25X TBE cooled to $12 \square C$.

9. Run the gel using the following parameters: <u>Bulk 1</u> : volts/cm = 6.0, included angle = 120°, initial switch time = 1.0 sec, final switch time = 40.0 sec, run time = 13 hours, ramping = linear

<u>Bulk 2</u> : volts/cm = 5.0, included angle = 120° , initial switch time = 4.0 sec, final switch time = 5.0 sec, run time = 8 hours, ramping = linear

10. Using a ruler as a "straight-edge" cut the gel with as shown in FIGURE 1c.

11. Stain and destain the two peripheral (flanking) pieces of the gel with SyberSafe. On the Safe Imager box, align the flanking gel pieces.

12. Use tips to mark the flanking gel pieces at 125 kb and 210 kb as shown in FIGURE 1d.

13. On a clean Qtray cover, reconstruct the gel by placing the unstained "center piece" between the two flanking gel pieces (FIGURE 1e). Using a ruler as a guide, cut the center gel piece as shown in FIGURE 1f so that you have three gel blocks : a center gel block containing DNA between 125 and 210 kb and two "end pieces".

□ □ Note : Never expose the center block to UV light as this will break the size-selected DNA!

14. Cut the center gel block transversely at two places to yield three blocks (referred to from bottom to

top as "1", "2", and "3") of approximately equal size (FIGURE 1g). Stain the end pieces with BET, partially reconstruct the gel using the various stained pieces, and photograph. The unstained pieces should not be included in the reconstruction.

Place blocks 1,2,3 back in the BioRad CHEF electrophoresis chamber.

□ □ **Note :** Continue with the second size selection before pausing.

The Fraction 1 correspond to the 125-150 Kb

The Fraction 2 correspond to the 150-180 Kb

The Fraction 3 correspond to the 180-210 Kb

FIGURE 1 - Performing the first size selection



Annexes



(a) Prepare a "slot well" by joining 4-6 wells together. If necessary, the slot-well can be extended anteriorly to provide more space for the plug pieces. Seal the bottom of the slotwell with a thin layer of melted agarose.



(b) Place plug pieces into the slot-well. Load the PFGE lambda ladder in wells flanking the slot well. Seal all wells with melted agarose.



(c) Divide the gel into three pieces as shown. The two flanking pieces should each contain a small part of the slot well as well as a DNA ladder. Stain the flanking pieces with ethidium bromide. Do not stain the center piece (the gel piece containing most of the genomic DNA).



(d) Align the stained flanking gel pieces and examine on a UV light box. If the digestion worked as planned, most of the genomic DNA should lie between 125 and 250 kb. Using a scalpel, make incisions in the flanking pieces marking the 125, 150, 200 and 250 kb borders



(e) Reconstruct the gel by placing the unstained center pieces between the two flanking gel pieces.



French Plant Genomic Resource Center

Centre INRA de Toulouse – CNRGV

Annexes

J7. Second size selection

PURPOSE

Though the first size selection gets rid of most of the DNA < 100 kb, some small DNA molecules do get trapped by the longer DNA molecules (this is especially true when the DNA concentration in the plugs is relatively high). The second size selection increases the chance that most of these smaller DNA molecules are eliminated prior to electroelution and ligation.

EXPERIMENTAL PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS : agarose cubes *1*, *2*, and *3* containing size-selected DNA; 0.25X TBE; agarose; SeakemGold agarose, LMP agarose (Seaplaque® Agarose); PFGE Lambda Ladder; scalpel; microscope slides; coverglasses; CHEF gel apparatus; large CHEF gel casting stand; 30-tooth gel comb

1. Make a 1.0% Seakem Gold agarose gel in 0.25X TBE using the **large BioRad CHEF gel casting stand and a 30-tooth gel comb**. Fill the stand until it is near overflowing. Allow the gel to thoroughly solidify.

2. Using a clean scalpel, construct three "slot wells. Extend each slot well anteriorly to allow

enough room for one of the end products of the first size selection (*i.e.*, blocks *1*, *2*, and *3*). Make sure that the agarose lining the bottom of the wells is removed and that the leading edge of each slot well is parallel to the leading edge of the comb-made wells (FIGURE 2a).

3. Using a scalpel and a ruler as a "straight-edge", cut a large block of agarose out of the upper-central

region of the gel. The cuts should be made exactly as shown in FIGURE 2a. Remove all agarose from this region. Completely fill the resulting cavity with 1.0% low melting point (LMP) Seaplaque agarose in 0.25X TBE. Allow the LMP agarose to solidify.

4. Place the PFGE Lambda Ladder in lanes flanking the slot wells.

5. Place blocks *1*, *2*, and *3* in their original top/bottom orientation (see FIGURE 1g) in the three slot wells as shown in FIGURE 2a. Add LMP agarose to fill any remaining space in each slot well.

Gently move each agarose block back and forth to dislodge any bubbles, and position each block so that it touches the leading edge of its respective slot well.

6. Seal the wells containing ladders with melted LMP agarose. Allow the LMP agarose to solidify.

7. Remove the gel from the casting stand, and wipe any agarose off the bottom of the base plate.

8. Place the base plate and overlying gel in the BioRad CHEF Mapper apparatus. The unit should

contain 2.5 L of fresh 0.25X TBE cooled to $12 \Box C$.

9. Run the gel using the following parameters: 5.0 v/cm, included angle = 120° , initial switch time = 3.0 sec, final switch time = 4.5 sec, run time = 13 hours, ramping = linear.

10. Using a ruler as a "straight-edge", cut the gel with a scalpel as shown in FIGURE 2b. There will be seven gel pieces (referred to as 1-7 from left to right). Using a scalpel, place identification

notches at the bottom of some of the gel pieces as shown. These notches allow easy reassembly of the gel later in the protocol.

11. Stain and destain the odd-numbered pieces of the gel (i.e., the gel pieces containing the PFGE

Lambda Ladder). On the Safe Imager box, realign the stained pieces as shown in FIGURE 2c. While observing the illuminated gel, use tips to mark DNA (mark DNA bigger than 100 kb) in each of the stained pieces (FIGURE 2c).



 \Box **Note :** Never expose the even-numbered gel pieces to UV light as this will break the size-selected DNA and make it unclonable!

12. On a clean Qtray cover on a workbench, reconstruct the gel by placing the unstained evennumbered pieces between their flanking odd-numbered stained pieces (FIGURE 2d). Extend the marks on each stained gel piece into adjacent unstained (even-numbered) gel piece. Place the stained gel pieces aside.

13. Using a scalpel, coverglass, or razor blade, connect the incisions on each even-numbered gel piece as shown in FIGURE 2e. You should now have three unstained gel pieces containing DNA. The DNA from these pieces (referred to from left to right as α , β and δ) will be isolated. Place the gel pieces in 50 ml polypropylene centrifuge tubes containing 1X TAE. **Incubate at least 2 hours in 1X TAE at 4°C before electroelution**. Change the 1X TAE buffer at least 2 times.

14. Stain and destain the remaining gel pieces. Based on the notches made earlier, reconstruct the gel on a UV light box. If desired, photograph the reconstructed gel for documentation purposes.

(a) Using a scalpel, join adjacent comb-made wells to form three slot-wells. Place a thin layer

of LMP agarose in the bottom of each slot- well

and allo 0.5 cm arose to solidify. Insert blocks 1 ,2 and 0.5 cm arose slot-wells. Place the PFGE lambda ladder into wells flanking the slot-wells as shown. Be sure to push x, y, and z to the leading edge of their respective slot-wells. Seal all wells

with LMP agarose. Using a ruler as a guide, use a scalpel to cut out a large block of agarose from the center of the gel (see diagram). Fill the resulting

FIGURE 2 - The second size selection.



(b) Cutting the gel After electrophoresis, divide the gel into seven pieces as shown above. The pieces are referred to as 1-7 from left to right. Cut notches into the bottom of pieces 3-7 as Notches facilitate rapid shown. reconstruction of the gel. Stain the odd-numbered pieces with ethidium bromide. Do not stain the even-numbered pieces as contain genomic DNA these which may be used in subsequent ligation reactions





Annexes



(c) Align the stained (odd-numbered) gel pieces and examine on a UV light box as shown. Using a scalpel, make incisions to mark the 125 kb (pink lines) and 250 kb (green lines) borders.

(d) Reconstruct the gel by placing the unstained even-numbered pieces between the stained odd-numbered pieces. Use the notches at the bottom of pieces 3-7 to facilitate proper assembly of the gel. Extend the incisions at 125 kb and 250 kb on each odd-numbered gel piece into adjacent even-numbered gel pieces (green and pink lines).
* Note: The center gel piece should never be exposed to UV light from the light box as this can break the DNA!





Annexes



(e) Using a scalpel or razor blade, connect the appropriate incisions on each even-numbered gel piece as shown. The resulting center blocks containing DNA between 125 and 250 kb are referred to as α , β , and δ , respectively. Subsequently, cut each of the three center blocks longitudinally to produce three blocks of roughly equal size (dotted lines). Place the three α blocks in one centrifuge tube, the three β blocks in a second tube, and the three δ blocks in a third tube. The DNA in these blocks may be suitable for ligation.

J8. Isolation of size-selected DNA from agarose

PURPOSE

High molecular weight restriction fragments must be removed from surrounding agarose before they can be used in ligation reactions.

Using the Bio-Rad Model 422 Electroelution System

SUPPLIES, EQUIPMENT, AND REAGENTS: agarose cubes $\Box \alpha$, β and δ)containing size-selected DNA; 1X TAE; Miracloth \Box squares; BioRad Model 422 Electro-Eluter electroelution system; BioRad PowerPac \Box 1000 power supply

1. Incubate collection cups and miracloth pieces in 1X TAE at 65° during 1h.

2. Pick out three assembly joints and place them on a Q-tray cover so that the large orifice of each is pointing up. Place a collection cup under each assembly joint and gently fit the collection cup into the small (dorsal) orifice of the assembly joint (FIGURE B.2).

3. Place assembly joints, glass tubes, collection cups, and the cuvette rack on the Q-tray cover. Add 1X TAE to each of the three assembly joint/collection cup combinations until a reverse meniscus is visible at the top of each assembly joint. Gently place a circular piece of sterile Miracloth onto the reverse meniscus of each assembly joint (FIGURE B.2). Insert a glass tube (frosted side down) into each of the three assembly joints. The three resulting structures are referred to as electroelution cuvettes (see FIGURE B.2).

4. Fill each cuvette with 1X TAE. Place the three agarose blocks in the cuvettes (agarose blocks must have been previously soaked in 1X TAE during (at least) 2 hours). Let agarose blocks drift through the buffer until they come to lie on the Miracloth stretched across the bottom of the glass tubes. If a block does not naturally drift to the Miracloth boundary, gently use a glass rod to push it to this interface. If the cube is too wide to fit in a cuvette, do not force it through the cuvette. Rather, trim the cube until it will fit in the cuvette. Remove air bubbles inside the cuvettes.

5. Insert each cuvette into a gasket of the cuvette holder as shown in FIGURE B.3. Place gasket plugs in gaskets that are not being used.



6. Place the cuvette holder into the main buffer tank (FIGURE B.4). Add 1X TAE to the main buffer tank until each assembly joint on the electroelution cuvettes is completely submerged.

7. Fill the upper buffer chamber with 1X TAE. Each cuvette should be full of buffer (FIGURE B.4).

8. Place a stir bar in the bottom of the main buffer tank, and set the apparatus on a stir plate. Activate the stir bar so that the buffer in the main tank is **gently** agitated.

9. Place the lid on the main buffer tank. Make sure that the red (positive) electrode is next to the red dot on the side of the main buffer tank and that the black electrode is next to the black dot on the main buffer tank (FIGURE B.4).

10. Insert the electrode plug at the end of the electrode wires into an appropriate power supply

11. Each cuvette requires 10 mA of power. Thus if three cuvettes are used, the power should be set at 30 mA. If the cuvettes are properly loaded, the voltage should start off at about 80-100 v. Higher voltages indicate possible problems with the cuvettes (*e.g.*, trapped air bubbles). Periodically check the voltage. If voltages > 110 are seen upon starting the apparatus, turn off the power supply, remove the main buffer tank lid, and check to see that the agarose cubes are resting on the Miracloth stretched across the bottom of the glass tubes. Dislodge any bubbles trapped below the agarose cubes or otherwise present in the cuvettes.

If the voltage drifts to > 130 v over the course of an hour, decrease the amperage until the voltage is between 80-100 v.

12. Run the electroeluter for 2 hours.

13. Interchange electrodes and let the DNA run in opposite direction for 90s.

14. Recover electroeluted DNA with large-orifice pipet tips for α,β and δ blocks. TAKE GREAT CARE IN **PERFORMING THIS STEP**. The volume recovered should be approximately 70-80 µl.

Place F1, F2 and F3 DNA samples directly in the well which will be used for the ligation reaction in the thermocycler. After quantification the volume in excess will be removed and placed in a proper tube for each fraction to be conserved at -20°C.



Annexes





FIGURE B.2 - Assembly and loading of an electroelution cuvette. (a) Insert a collection cup into the small aperture of an assembly joint. (b) Add 1X TAE to the cup/joint unit until a reverse meniscus is visible at the top of the joint. (c) Place a Miracloth square onto the reverse meniscus. (d) Insert the frosted end of a glass tube into the large aperture of the assembly joint. Push the tube in as far as it will go. The Miracloth should be

stretched across the bottom of the glass tube. (e) The cuvette is now completely assembled and ready for loading. (f) Fill the cuvette with sterile 1X TAE buffer. Drop an agarose cube containing size-selected DNA into the cuvette. The cube should drift through the buffer until it rests on the Miracloth stretched across the bottom of the glass tube. Use a glass rod to gently push the cube to the bottom of the tube if it does not drift down by itself.



FIGURE IS'S Recovery of DNA. Once electroniation is complete, in the control in the part in the part in the upper buffer tank using a sterile Pasteur pipet. (a) Using the pipet, gently remove all buffer above the surface of the Miracloth at the bottom of each glass tube. PERFORM THIS STEP SLOWLY AND CAREFULLY! IF THE BUFFER IS ASPIRATED TOO RAPIDLY OR IF SUCTION IS APPLIED TO THE MIRACLOTH ITSELF, DNA WILL BE PULLED OUT OF THE COLLECTION CUP! (b) Remove a cuvette from the cuvette holder. Gently remove the glass tube, Miracloth, and agarose cube from the assembly joint. A droplet of liquid containing the DNA should be visible at the bottom of the collection cup. (c) Using a wide-orifice plastic pipet ip, slowly remove the DNA solution and place it in a sterile 1.5 ml microfuge tube. Approximately 0.4 ml of DNA solution should be recovered from each cuvette.



J8.POST-ISOLATION PROCEDURES

SUPPLIES, EQUIPMENT, AND REAGENTS (see CHAPTER 2 for details): DNA isolated from α,β and $\delta \square$ agarose cubes (see above); 1X TAE; agarose; 70% ethanol in a spray bottle, 1X uncut lambda DNA; submarine \square mini-gel apparatus; UV light box equipped with camera or image-capture system; blue juice; ethidium bromide

1. Prepare a 1% agarose submarine mini-gel in 1X TAE. In preparing the gel, use a comb with at least 7 wide teeth.

2. Place 0.5, 1.0, 2.0, 3.0, 4.0 μ l and 6.0 μ l of 5ng/ μ l uncut lambda DNA (*i.e.*, 2.5 ng, 5 ng, 10 ng, 15 ng, 20 ng and 30 ng) in separate 0.65 ml microfuge tubes.



Annexes

Each person must prepare and maintain its own solution of lambda DNA at -20°C.

Think to adapt the range of lambda DNA on the basis of visible gel.

To prepare a new uncut lambda DNA stock solution: see annex b

3. Take a 5 μ l aliquot from F1 insert DNA and place the solution in a 0.65 ml microfuge tube. Add 1 μ l of blue juice. Do the same for the F2 and F3 samples.

4. Submerge the mini-gel in 1X TAE buffer in an appropriate mini-gel apparatus. Load the gel as shown in FIGURE 3. Run the gel at 100 v for 15-20 min.

30 minutes in BET – 30 minutes in water

Stain and photograph the gel. Based on comparison of the relative fluorescence in the sample and standard lanes, an estimate of the concentration of each sample can be made (FIGURE 3).

Multiplication of a sample's volume by its concentration gives an estimate of the total amount of insert DNA in that sample.

5. Proceed to high sense qubit quantification to correlate the DNA concentration to the gel estimation. Gently homogenise the DNA with a wide bore tip and make two measures per fraction: top and bottom and calculate the average concentration.

See notice : <u>..\..\DOCUMENTS QUALITE CNRGV\Activites ACT\Notices\NT 1258 ACT 068-VA Dosage de</u> <u>l'ADN double brin sur le fluoromètre Qubit 3.0.doc</u>

□ □Note : Be careful to correlate the Qubit quantification to the gel profile before proceed to ligations.

5. Proceed to ligations.

 \Box **Note :** High MW (insert) DNA samples are quite unstable. Though they can usually be left at 4°C overnight, it is best to perform ligation immediately after checking the DNA concentration.

FIGURE 3



DG Peterson 9/99

J8.Ligation

SUPPLIES, EQUIPMENT, AND REAGENTS:

Dephosphorylated *HindIII* pIndigoBAC5, *BamHI* pIndigoBAC5 or *EcoRI* Copy control pCC1BAC vector stock; insert DNA; T4 ligase; 5X T4 ligase buffer



1. Place insert DNA, the dephosphorylated vector stock solution, and the T4 ligase 10X buffer on ice. Allow the buffer and the vector stock to thaw **on ice**. Keep T4 ligase in the -20° C freezer-box.

2. Set up ligation reactions as described below. The number of reactions that can be prepared depends upon the nanograms of insert DNA available. Make reactions in tubes that will fit in a thermocycler.

Handle insert DNA and ligations with large-orifice pipet tips.

Quantity of DNA insert	100 ng
Ligase 5U/µL	2 μL
Ligase buffer 10X	adapt to final volume
Vector 25ng/µL	adapt to insert DNA quantity
H20	Complete at the final volume
Final volume	100-150 μL

 \square **Note :** The DNA concentration is important and it is important to maintain it around 1-1.5 ng/µl

3. Gently swirl each reaction tube to mix the tube's contents. DO NOT VORTEX OR AGITATE

VIOLENTLY AS THIS MAY SHEAR THE INSERT DNA.

- 4. Incubate the ligation reactions at 16°C for 10h (in a themocycler).
- 5. "heat kill" the enzyme : 65°C for 20 min.

J9- Mass transformation

SUPPLIES, EQUIPMENT, AND REAGENTS:

ELECTROMAX® DH10B[™] competent cells ; ligated DNA ; desalting tubes (100mM Glu, 1% agarose) SOC ; X/I/C petri dishes ; glass plating rod ; electroporator (Gibco BRL Cell Porator) with cuvettes, culture tubes

Handle ligations with large-orifice pipet tips

1. Desalt the ligations during **1h15** on ice.

Preparation of desalting tubes	100mM Glu, 1% agarose	
Glucose solution 1 M	5 ml	
Standart agarose	0.5g	
H20	45 mL	
Final volume	50 mL	

To avoid DNA manipulation before transformation, synchronize the desalting step and the transformation. Just 10

min before the end of the desalting step, think to thaw competent cells on ice. After that you will be able to sample the volume you need from each desalting tube for the transformation. Finally remove the remaining desalted ligation products and place them in proper microfuge tubes for conservation.

Store the ligation at 4°C for 2 weeks. And then store it at -20°C.

In the case of a first transformation (efficiency test)

2. Prepare 5 cuvettes, 3 culture tubes. Put SOC at 37°C with the different volume in 2 ml tubes cf table. Thaw competent cells on ice for **10 minutes.**

Volume bactérie en µL	Volume ligation dessalée en µL	SOC Volume
-----------------------	--------------------------------	------------

French Plant Genomic Resource Center

Centre INRA de Toulouse – CNRGV

Annexes

F1	40	6	800 µL
F2	40	6	800 μL
F3	20	3	400 μL

3. Dispatch in 3 tubes respectively 40,40 and 20 µl of competent cells

Add x μ l of desalted ligation to x μ l competent cells cf table. Tap the tubes gently to mix contents and keep on ice for **10 minutes** before proceeding with electroporation.

4. Tap the tubes gently and Divide each tube into 2 or 1 cuvettes (2 x 23 μ l or 1 x 23 μ L) and electropore :

Biorad micropulser : Take care to put the EC3 program

Gibco Cell Porator : Electropore at 325-330V

5. Immediately recover electroporated cells (with large-orifice pipet tips) in 2 ml pre-heated SOC.

6. Immediately place culture tubes in a 37°C incubator shaker set at 250 rpm for exactly one hour.

7. For each culture tube: spread 50 μ l on a petri dish. Add 10% of glycerol to the volume of remaining culture and stock at -80°C.

For each culture tube, think to spread 25 μ l for F1 and 50 μ l for F2 and F3 on a petri dish if you have recovered little material after elution.

In the case of a transformation for real construction of a BAC library

2. Prepare 15 cuvettes, 3 culture tubes with 2 Ml of SOC medium. Put the 3 tubes at 37°C. Thaw competent cells on ice for **10 minutes**.

3. Add 15 μ l of desalted ligation to 100 μ l competent cells. Tap the tubes gently to mix contents and keep on ice for **10 minutes** before proceeding with electroporation.

4. Tap the tubes gently and divide each tube into 5 cuvettes (5x 23 μ l) and electropore :

Biorad micropulser : Take care to put the EC3 program

Gibco Cell Porator : Electropore at 325-330V

5. Immediately recover electroporated cells (with large-orifice pipet tips) in 2 ml pre-heated SOC.

6. Immediately place culture tubes in a 37°C incubator shaker set at 250 rpm for exactly one hour.

7. For each culture tube: spread 50 μ l on a petri dish. Add 200 μ l of glycerol to the 1.9ml remaining culture and stock at -80°C.



Annexes

Annex a: Spermidine stock solution

To prepare 30 tubes at 40mM concentration (30 x 1.5ml = 45 ml stock solution)

Cm = 0.925 g / ml (25°C) MM = 145.25 g / mol 1 mL = 0.925g = 6.37.10⁻³ mol 1 ml pure solution: $0.925g = 6.37.10^{-3}$ mol 1000 ml pure solution \Leftrightarrow 40 mMol 45 ml pure solution \Leftrightarrow 1.8 mMol \Leftrightarrow <u>0.283mL spermidine QSP 45mL</u> Prépare 1.5ml aliquots <u>Annex b:</u> Uncut lambda DNA stock solution 6µL lambda DNA solution (500ng/µl) + 100µL Loading buffer + 494 µl H2O = 600µl (5ng/µl) Prepare 100µL aliquots



Mélanger en inversant

Annexe 3 : Mise en culture Midi-prép BAC - méthode Macherey-Nagel

Matériel utilisé : Centrifugeuse Hearaeus Multifuge X3R 230V, ThermoFisher

Kit utilisé : Nucleobond Xtra Midi plus, Macherey Nagel ref 740410.100 (100 réactions)

NB : Faire par série de 6 (centrif de 6 pots), la lyse de la 2^{em} série peut démarrer pendant la centrif de la neutralisation de la 1^{er} série.

1- Préculture overnight (12h) à 37°C agitation 250 rpm

Ensemencer 1µL du clone à extraire dans 4mL de milieu LB Glycerol + Cm $(12.5\mu g/\mu L)$

Conserver la préculture à -80°C jusqu'à validation du résultat final (séquençage 454)

2- Culture overnight (16 h) à 37 °C agitation 250 rpm

100 ml LB + Cm 12.5 μ g/ μ L inoculé avec une préculture de 100 μ L

- 3- Mettre le buffer ELU et la LYSE dans une étuve à 50 °C.
- 4- Transférer 50mL de culture dans des tubes falcons de 50mL
- Centrifuger à 4 °C **15 min à 6000 g**

Jeter le surnageant dans un flacon bouché (à autoclaver)

Transférer les 50mL de culture restant sur le culot de bactéries

Centrifuger à 4 °C **15 min à 6000 g**

Jeter le surnageant dans un flacon bouché (à autoclaver)

Enlever le reste de surnageant

Stocker les culots à - 20 °C, ou poursuivre la manip en mettant les culots sur la glace

Attention ne pas congeler les culots à -20°C pour faire du séquençage car trop de contamination

d'Escherichia coli continuer la manip.

Autoclaver les erlens ayant contenus des bactéries

5- Resuspension du culot bactérien dans buffer **RES** (RES + RNaseA 60mg/L, stocké à 4 °C)

Ajouter 8 ml de buffer RES Agitation à la main (culot bien homogéneisé) puis stocker les tubes dans la glace.

6- Lyse

Ajouter 8 ml de buffer LYS (préalablement chauffé à 50 °C) le tube 7-8 fois.

NE PAS VORTEXER.

Incuber à température ambiante 4 min strictes

7- Neutralisation

Ajouter 8 ml de buffer **NEU**. Mélanger en inversant le tube 7-8 fois jusqu'à l'obtention d'un mélange blanc (plus aucunes traces de bleu de Lyse)

Centrifuger à 4 °C. **15 min à 5000 g**

Après centrifugation, garder les tubes droits pour éviter le mélange lysat-précipité

8- Filtration



Annexes

- (préparation des colonnes pendant l'étape de la centrifugation) \sim 45 min
- + 12 ml EQU Buffer : lavage des colonnes à déposer sur la collerette du filtre
- + Lysat : chargement environ 25 ml, avec pipette 25ml + pipetboy
- (Si quelques débris, veiller à les déposer sur la collerette du filtre)
- + 5 ml EQU buffer : lavage de la colonne sur la collerette
- Puis retirer le filtre de la colonne <u>NE PAS OUBLIER (poubelle biologique)</u>
- + 8 ml **WASH** buffer : lavage de la colonne avec 8 ml de WASH.
- Essuyer l'extrémité de la colonne avec un kimwipe afin d'éliminer un maximum de WASH buffer
- + 5 ml ELU buffer : élution (préalablement chauffé à 50 °C), en ayant mis sous la colonne
- un Falcon 50 mL NE PAS OUBLIER.
- 9- Précipitation ADN
- + 3.5 ml isopropanol (0.7 volumes) sous sorbonne. Mélanger doucement en inversant le tube 4-5 fois.
- Centrifuger à 4 °C (on peut laisser plus longtemps) 45 min à 12 000 g
- 10- Lavage EtOH 70%

Jeter le surnageant dans flacon Déchets Isopropanol.

Culot + 3 ml EtOH froid (stocké à 4 °C).

Centrifuger à 4 °C (on peut laisser plus longtemps) 45 min à 12 000 g

Jeter le surnageant (déchets non spécifiques) et localiser le culot d'ADN sur le tube avec un marqueur. Bien

sécher le culot d'ADN avec kimwipe, ou 10 min ou + sous ventilation hotte minimum 1 heure. PAS UNE

SEULE GOUTTE D'EtOH.

- 11- Culot ADN + 150 μl TE, reprendre avec pipette et pointe plusieurs fois pour bien homogénéiser. Laisser reprendre overnight à 4 °C sur portoir penché vers la localisation du marqueur.
- 12- Le lendemain, récupérer les 150μL de chaque échantillon dans des tubes à vis qui serviront au dosage Xpose le jour même et à la digestion des BACs ou sinon les congeler à -20°C.
- 13-Digestion des BACs pour validation des tailles, on peut faire la digestion NotI 1 heure après avoir mis les culots + TE à 4°C

Caractérisation de l'ADN extrait

<u>Plusieurs étapes :</u>

- Test contamination ARNs

Déposer 2 μ L d'ADN extrait + 1 μ L de tampon de charge 6X sur un gel d'agarose 2% (TAE 1X) pour vérifier des traces éventuelles d'ARNs :

- Pas de trace d'ARNs \rightarrow OK
- Trace d'ARNs → Ajouter 1 µL de RNase à 10mg/ml (stockée à 4°C frigo salle 454 solution mère de RNase à 100mg/ml QIAGEN) dans chaque tube contenant des ARNs mélanger doucement et incuber à 37°C 1 heure puis refaire un gel pour confirmation.

- Estimation de la concentration en ADN : dosage au Xpose

Déposer 2 μ L d'ADN extrait sur une slide 40 et ne pas oublier 2 μ L de TE qui serviront de blanc. **Référence notice**



Annexes

- Estimation de la taille des clones BACs

1- Digestion enzymatique avec avec l'enzyme NotI

A partir de la concentration d'ADN obtenue au Xpose, calculer la quantité à pipeter pour obtenir une concentration

à 100ng/µL. Les digestions se font dans des microplaques 96 puits Greiner à fond conique (plaques de dilutions). Préparation du MIX :

	n=1
Tampon 10X	2
FastNotI	0,1
H2O	17,4

10X FastDigest Green Buffer THERMO SCIENTIFIC FERMENTAS réf :11836863

FastDigest NotI THERMO SCIENTIFIC FERMENTAS réf: 15300221

Ajouter 19.5 μ L de mix/puits + Vol ADN à 100ng/ μ L

Sceller la plaque avec un film aluminium, mélanger en tapotant et centrifuger puis incuber 40 minutes à 37°C sur

plaque agitante (vitesse 40 rpm).

 $\label{eq:preparer} Préparer une autre plaque avec 100 ng de BACs non digérés (15 \mu L H2O + 3 \mu L tampon de charge 6X + Vol ADN) et laisser à 4°C jusqu'au dépôt sur gel PFGE$

Préparation du mix :

	1X	
Tampon 10X	2µl	NB : Tampon 10X Biolabs NEB4
BSA 100X	0,2µl	
NotI (NEB)	0,11µl	
H ₂ O	16,69µl	
Volume Total	19µl	

Ajouter 1µl et 0.5µl de chaque midi-prep dans la microplaque puis mettre 19µl de mix dans chaque puit.

- 2- Electrophorèse en champs pulsé (PFGE)
 - Préparer un grand gel 0,8% Pulsed Field Certified Agarose (Biorad) (2g/250ml TBE 0.25x peigne de 30 ou 45 dents).
 - Ajouter 2µl de tampon de charge « maison », ou 5µL de tampon de charge 6X Fermentas, par puits.
 - Mettre le ladder : 1mm de MidRange PFG Marker N3551S Biolabs (ou 1mm de Lambda ladder PFG Marker N03405 Biolabs) (enlever les bulles, ou sceller avec de l'agarose se conserve à 20°C et se présente sous forme de seringue).
 - Immerger le gel dans la cuve (TBE 0.25X <u>cuve branchée à 12 °C au préalable</u>).
 - Transférer **la totalité** de chaque digestion dans les puits du gel Alterner pour chaque échantillon BACs digérés et non digérés
 - Faire migrer O/N en appliquant le programme suivant (prog. Miniprep) :

6 v/cm, included angle = 120°, initial switch time = 5 sec, final switch time = 15 sec, run time = 16 hours, ramping = linear



Annexes

- Colorer le gel 1h dans BET puis le rincer 1h dans H20.
- Estimer la taille des inserts avec le logiciel Genetools. Référence notice



Annexe 4 : Caractérisation de clones BAC : digestion avec l'enzyme de restriction NotI et migration en électrophorèse en champ pulsé

Les digestions enzymatiques de clones BACs avec l'enzyme de restriction NotI permettent d'exciser le fragment d'ADN du vecteur afin de pouvoir en estimer la taille en électrophorèse en champ pulsé.

Remarque : La FastNot1 permet de faire une digestion en 40 min

Digestion enzymatique :

Les digestions se font dans un volume de **20µl** dans des microplaques 96 puits à fond conique (type PCR classique) à partir d'ADN de clones BACs extrait en mini-prep (NT 1258 ACT 023 -VF Mini-prép BAC en plaques- methode Macherey-Nagel).

Préparation du mix pour n réactions :

	1 réaction
10X Fast Digest Bufffer (Green)*	2µL
Enz. NotI/FastNot1	0,1µL
H2O	10,9µL
Volume Total	13µL

* Le tampon 10X contient déjà le tampon charge. Utiliser préférentiellement le tampon 10X « green » qui contient déjà le tampon de charge. Si ce n'est pas possible, utiliser le tampon classique et ajouter du tampon de charge avant dépôt sur gel.

Mettre 13µl de mix dans chaque puits.

Ajouter 7µl d'ADN

Sceller la plaque avec un film aluminium ou plastique, centrifuger et mettre à **37**°C sous agitation 40 rpm pendant **40 minutes**.

Migration :

La migration se fait en électrophorèse en champs pulsé (PFGE) dans du **TBE 0.25X** dans un grand gel **0,8%** de **Pulsed Field Certified Agarose** (Biorad, volume à préparer 250ml).

Centrifuger la plaque quelques secondes.

Ajouter 4µL de tampon de charge 6X par puits si nécessaire.

Immerger le gel dans la cuve et ajouter 10µL de ladder maison aux extrémités.

Déposer chaque digestion dans les puits du gel.

Faire migrer le gel en appliquant le programme suivant :

6 v/cm, included angle = 120°, initial switch time = 5 sec, final switch time = 15 sec, run time = 16 hours, ramping = linear



Annexes

Coloration et lecture du gel :

Colorer le gel pendant 1h dans un bain de BET. Rincer le gel pendant 1h dans un bain d'eau. Prendre une photo sous UV et procéder à l'estimation de la taille des inserts : NT 1258 ACT 037 Utilisation de GeneTools



Annexes

Annexe 5 : Marquage d'une sonde par Random Priming

I - Marquage d'une sonde : produit PCR unique ou mélange

Utilisation du kit "Ready-To-Go DNA Labelling Beads (-dCTP)" GE-Healthcare

<u>Principe</u> : Sondes marquées radioactivement d'environ au-moins 250 nucléotides de long en incorporant des $[\alpha^{-33}P]dCTP$ (dans certains cas où les sondes seront très spécifiques, le marquage pourra se faire sur des fragments de plus petite taille, 100 nt minimum). Les sondes sont marquées séparément puis poolées si nécessaire.

Préchauffer un bain marie à sec à 100°C et un autre à 37°C

Prép. de la sonde dans un tube à vis :

- 100-150 ng ADN

- H₂O stérile qsp 45.5 µl

Dénaturer 10 min à 100 °C

Refroidir immédiatement dans la glace 1 min puis centrifuger brièvement

Ajouter dans le tube contenant l'ADN (tubes dans la glace)

- bille contenant l'enzyme klenow et les hexanucléotides (dATP, dTTP, dGTP)

- $[\alpha$ -³³P]dCTP 50µCi 4.5 µl (pointes longues 10µl pour éviter contaminations)

Vortexer doucement

1h30 à 5 h à 37°C

II - Purification d'une sonde marquée

Utilisation du kit "Illustra Tm ProbeQuantTM G-50 Micro Columns "GE-Healthcare.

<u>Principe</u> : La colonne Sephadex G50 permet de retenir tous les fragments d'ADN dont la taille est inférieure à 50 nucléotides et les dNTPs non incorporés.

Resuspendre la résine en tapotant du doigt la colonne G50

Casser la fermeture inférieure de la colonne.

Placer la colonne G50 dans un tube 2 ml sans bouchon (tubes usagés venant du kit).

Centrifuger 2 min à 0.7 rcf (795 g, 3000 rpm)

Jeter le tube et introduire la colonne G50 dans un tube 2 ml sans bouchon (tube neuf)

Déposer au centre de la colonne les 50 µl de la sonde marquée sans toucher à la résine

Centrifuger 2 min à 0.7 rcf.

Jeter la colonne dans la poubelle de déchets radioactifs solides

Transférer les sondes marquées dans des tubes 1.5ml à vis.



Annexes

III - Comptage sonde

 $1~\mu l$ de sonde purifiée + 4 ml de liquide à scintillation dans les tubes à scintillation

Bien homogénéiser avant comptage au compteur à scintillation Triathler



Annexe 6 : Hybridation de macroarrays

1. Objet et domaine d'application

Ce mode opératoire concerne les hybridations de sondes radioactives marquées avec le radioélément P33 de macroarrays de type bactérien et rappelle les précautions à adopter lors de cette manip.

2. Documents de références et documents associés

EN 1258 ACT 004	Tri des déchets et traitement verrerie
EN 1258 ACT 007	Fiche d'enlèvement des déchets
EN 1258 ACT 006	Utilisation P33.
NT 1258 ACT 005	Marquage terminal d'une sonde oligonucleotidique
NT 1258 ACT 006	Marquage d'une sonde par Random Priming
NT 1258 ACT 033	Utilisation du storm

3. Hygiène et sécurité

Pour l'hybridation/lavage :

Port de protections individuelles adaptées à la manipulation des radioéléments : gants, chaussures fermées.

Poste de travail adéquat : écran de protection adapté aux radioéléments utilisés, container déchets radioactifs (solides et liquides) accessibles et sécurisés, plan de travail facilement décontaminable et protégé (Benchguard).

Contrôle à priori, à posteriori et en cours de manip par sonde Geiger de l'opérateur (notamment gants), des pipettes, du plan de travail et de son environnement immédiat.

Utilisation des cônes à filtres, pipettes et matériels de paillasses étiquetées (sigle « danger radioactif).

A la fin de chaque manip, vider les poubelles de paillasse contenant les radioéléments dans la grande poubelle prévue pour les déchets radioactifs et changer la feuille de Benchguard nous servant de plan de travail pour la manipulation des radioéléments, si nécessaire

L'ensemble des tubes, bechers, mesh contaminés sont à rassembler dans la bassine prévue à cet effet et laisser tremper au-moins 1 à 2 jours dans un bain de « décon » à 5 %, puis rincer à l'eau du robinet.

4. Principe de la méthode

Cette méthode repose sur la complémentarité des bases constitutives des acides nucléiques (ADN ou ARN) et sur la réversibilité des processus de séparation (dénaturation) et de réassociation (renaturation) de 2 brins d'ADN complémentaires. Lorsqu'il y a renaturation (reformation des liaisons hydrogènes entre les deux brins) on parle aussi d'hybridation.Les macroarrays portent l'ADN de bactéries transformées ; dans des conditions optimales des sondes (petits fragments d'ADN marqués) peuvent s'y hybrider. En cas d'homologie forte, l'hybride reste stable même en condition de contraintes importantes (lavages). Le signal radioactif (capté par l'écran du phosphoimager) permet d'identifier le clone hybridé.

Matériels et produits nécessaires

Macroarrays Intercalaire de type « mesh » ou drain. Sonde radiomarquée Papier benchguard Grand bac en plastique



Bain sec à 98 °C. Four à hybridation Tubes à hybridation Centrifugeuse de paillasse Vortex Bain marie à agitation Julabo Container pour déchets liquides P33. Container pour déchets solides P33.

<u>SSC6X</u>

	Concentration finale	Concentration solution mère	Volume de solution mère à prélever pour 2L
SSC	6 X	20X	600 mL
H20	-	-	1400 mL

Solution de pré hybridation et d'hybridation

Respecter l'ordre des solutions à ajouter (cf. tableau). Attendre la dissolution complète du SDS (flacon 2L couché dans bain marie à 50 °C) avant de rajouter le denhardt.

	Concentration finale	Concentration initiale	Volume de solution initiale à prélever pour 2L
H20 stérile	-	-	1250 mL
SSC	6 X	20X	600 mL
SDS	0,5% (w/v)	20%	50 mL
Denhardt	5 X	100X	100 mL

Dénaturer 10 min à 95 °C l'ADN de sperme de saumon, puis refroidir dans la glace et ajouter extemporanément à la solution lors de l'étape de pré hybridation.

Annexes



Solution de Lavage 1

	Concentration finale	Concentration solution	Volume de solution
		mère	mère à prélever pour 2L
SSC	2 X	20X	200 mL
SDS	0,1% (w/v)	20 %	10 mL
H20	-	-	1790 mL

Solution de Lavage 2

	Concentration finale	Concentration solution mère	Volume de solution mère à prélever pour 2L
SSC	0.5 X	20X	50 mL
SDS	0,1% (w/v)	20 %	10 mL
H20	-	-	1940 mL

6. Contraintes de la méthode

- La durée minimum de cette manipulation est de 2 jours.

Préparation des membranes avant hybridation : 3 h

Hybridation des membranes : 6 h à 1 nuit.

Exposition des membranes en cassette : 4h30 à 1 semaine

- La manipulation des radioéléments doit se faire de manière rigoureuse selon les règles d'hygiène et de sécurité, en se souciant constamment des contaminations possibles de l'ensemble des matériaux utilisés : vérification au compteur geiger des gants, des pipettes utilisées, et autres matériels.

7. Contenu du mode opératoire.

• <u>Préparation des membranes</u> (après le marquage des sondes cf NT1258ACT006 Marquage d'une sonde par random priming):

Préchauffer le four à hybridation à 50 °C.

- Humidifier par capillarité un mesh (drain) puis la membrane, dans un bain de SSC6X (bac).
- Rouler de haut en bas l'ensemble membrane-mesh, nom mb en bas et les placer dans le tube à hybridation contenant **50 mL de SSC6X** préchauffé à 50 °C, bouchon à gauche dans les 2 fours (sens de rotation différent selon les fours)
- Veiller à ce que les membranes se déroulent au maximum, en se plaquant contre la paroi du tube à hybridation
- Incuber à 50 °C, 15'.
- Jeter le SSC6X dans l'évier.

• Préhybridation :

Augmenter la température du four à hybridation jusqu'à 68 °C.

- Dénaturer 10' à 95 °C l'ADN de sperme de saumon (prévoir 50µl de solution mère à 100 mg/mL par tube d'hybridation)
- Refroidir dans la glace. Centrifuger brièvement



Annexes

- Mettre 50 mL de tampon d'hybridation par tube d'hybridation.
- Ajouter 50µL d'ADN de sperme de saumon [100mg/ml] dénaturé dans le tube d'hybridation.
- Préhybrider à 68 °C 1 h 30 à 6 heures (en général de 10h à 15h).

• Hybridation :

NE PAS OUBLIER DE PURIFIER LA SONDE MARQUEE PAR COLONNE G-50 (cf NT1258ACT006 Marquage d'une sonde par random priming). Prévoir au minimum 10 Millions de cpm.

- Dénaturer 10' à 95° C la sonde marquée, refroidir 1 min dans la glace. Centrifuger brièvement.
- Ajouter la sonde dans le tube à hybridation. Si le volume de la sonde est inférieur à 30µl, la mettre directement dans le tampon en penchant le tube précautionneusement.
- Hybrider 1 nuit à 68 °C.

• Lavages :

- Préchauffer les solutions de lavages à 50 °C minimum 30 min avant dans le bain-marie.
- Baisser la température du four à 50 °C pour moins se brûler les doigts avec les tubes.

Solution de Lavage 1

- Vider la solution d'hybridation dans les déchets radioactifs liquides.
- Mettre 50 ml de solution lavage 1 par tube : 15' à 50 °C
- Vider la solution 1 dans le container « déchets radioactifs ».

Solution de Lavage 2

- Mettre 50 ml de solution lavage 2 par tube : 30' à 50 °C
- Vider la solution 2 dans le container « déchets radioactifs ».

• Exposition et révélation des membranes :

- Sortir la membrane avec son meshe dans un bac vide avec des pinces puis déposer la membrane sur une feuille de Benchguard. Les meshes seront laissés dans le bac avec du Decon pour décontamination.
- Laisser 3-5 minutes minimum le Benchguard absorber l'excès de liquide, tout en gardant humides les membranes. Ne pas les faire sécher complètement car il est difficile de les déshybrider parfaitement par la suite.
- Ensacher chaque membrane dans un film plastique type « saran » de manière hermétique (préalablement étendu sur la paillasse en faisant le moins possible de plis et de bulles d'air). Le plus pratique est de ne pas replier le saran sur les côtés à condition qu'il n'y ait pas de liquide restant sur les membranes.
- Vérifier l'hybridation avec le compteur.
- Jeter les feuilles de Benchguard dans la poubelle si elles sont radioactives, déchets solides radioactifs. Les feuilles sont cependant utilisées plusieurs fois avant d'être jetées.

Les membranes sont ensachées dans un film plastique afin de protéger les écrans des cassettes.

- Exposer 4h30 minimum les membranes dans une cassette (1 jour en moyenne selon le marquage, 1 semaine maximum), utiliser un écran érasé.
- Scanner les écrans avec un phosphoimager (résolution conseillée 200 μm). Compter environ 6 minutes de scan.

Transformer les photos. gel en .tif avec ImageQuant pour analyser les filtres avec HDFR3.



Annexes

<u>Nettoyage meshes et tubes</u>

Vérification au compteur de la non contamination du matériel. Si nécessaire, essuyer avec du decon jusqu'à décontamination ; s'il y a toujours contamination, laisser encore 1 semaine avec du decon.

Nettoyage de tout le matériel à grande eau, minimum 1 jour après utilisation. Chaque mesh est rincée 3-4 fois par va-et-vient sous le jet du robinet de chaque côté. On remplit d'eau chaque tube jusqu'à ce qu'il n'y ait plus de mousse de decon, puis rinçage encore 1-2 fois à l'intérieur et bien frotter l'extérieur pour enlever tout dépôt de tampon.

• Déshybridation des membranes.

Pensez à vider toutes les solutions de déshybridation-lavages dans la poubelle déchets liquides radioactifs.

- Préchauffer le four à hybridation avec table agitante à 60 °C
- Retirer le saran recouvrant les membranes et le jeter dans la poubelle déchets radioactifs solides
- Introduire avec des pinces les membranes séparées d'un intercalaire de type mesh dans un bain de NaOH à 0.5 mol/L.
- Déshybrider **30' à 60 °C**
- Vider la solution de NaOH dans le bidon déchets non spé (ce bidon est étiquetté avec la bande radioactive ; une fois plein, le bidon est transporté dans la soute radioactive pour décroissance avant de le donner au prestataire qui ramasse les bidons de déchets non spé).
- Rincer 2 fois, 5' avec SSC 6X sur le plateau agitant à température ambiante, vider la solution dans l'évier entre chaque rinçage.
- Déposer les membranes sur des feuilles de papier Benchguard
- Laisser sécher
- Conserver chaque membrane entre deux feuillets bleus dans des boîtes de type Q-Tray.

8. Enregistrement et traçabilité

Noter à chaque utilisation le volume consommé en radioéléments. Chaque lot de radioélément doit posséder sa propre fiche d'enregistrement, chacune présente dans le classeur fiche de vie de salle radioactivité. Tout nouveau lot doit être également enregistré dans le tableau excel du Centre INRA sur le serveur « Samba », répertoire radioprot / suivi stock.xls.

Pour chaque hybridation, noter l'ensemble des informations dans le fichier suivi hybridations macroarrays qui se trouve dans le dossier : Y:\BIO/Hybridation Macroarrays/SUIVI HYBRIDATIONS/

Enregistrer dans un fichier propre à la collection étudiée, l'ensemble des photos .gel et .tif obtenues pour chaque hybridation, suivant le cheminement suivant :

Y:\DiffuserOutilsGenomiques\Espèce \Prestation\ CriblFILTR_200911_xx_BléCS_nom

Enregistrer les résultats de validation des clones par PCR dans un dossier résultats



Annexes

Annexe 7 : Marqueurs utilisés pour le criblage des deux banques BAC de pois (Champagne et Térèse)

Les amorces des sondes utilisées pour le criblage ont été dessinées sur des séquences de gènes CBF repérées dans le génome de référence de la lignée Caméor (UMR Agroécologie, données non publiées).

• Séquence du Marqueur 1 : CBF426_6484

Pisum_scaffold00426_01

Pisum_contig031084

PsCam037752

EVM_all_unmasked.6484

>Pisum_contig031084 Pisum_contig031084:67820..69694 (- strand) class=gene length=1875

AAAAAACAACCAAACCAAACACCATGTGTGAAACAAATCCAAATCCTAATCCTAATGAAACCAACAACATGATCCTCACCAACCCTTT ${\sf CAATGATCCTAACATGTCTTCCCTATATTGTAGATCCATGATTTCAATGCTTTTCCTTGGTCACCTCAATGACCTTGCTCTTGCCGGTG$ GATCCCTTGCTGTTGGCAAACATCACTGGCTACTCCATTCTCTCTGGTCTTGCTGTTGGCATGGAACCAATTTGTGGCCAAGCTT TTTACATGAAAAAATCTACTTTTGTTTTTCAACCAAGATGAAGAAAATTGCAACAATGGCTCAAACTTATATTCTCTACTCAATCCCTGAT CTTGTTGCTCAATCTTTCATACACCCTTTGAGAATTTACCTTAGGACTCAATCCATAACTCTTCCTCTCACACTCTTTGTGCAACTTTTTCAA TCATTCTCCACATACCCATCAATTATTTTCTAGTCTTTTACCTCAATTTGGGTATTAAAGGTGTTGCCTTAAGTGGGGTGTGGACAAATT AGTGGTTACCACTTCTAAACCTCGCCGTTCCAAGCTGCATTTCTGTATGTTTGGAATGGTGGTGGTATGAATTTATGACTATCTTTTGTG GCTACCTAATCAACCCAAGATCAAGTGTTGCATCAATGGGTGTGTTGATTCAAATCACTTCATTGATGTACATTTTACCTTACTCACTA AGCAATAGTGTCTCAACAAGAGTTGGTAACATGATAGGTGCACAAAAGCCATCAAAAGCTAAAACTTTCCGCAATTTCCGGCTTATCAT ATTACCTTAACATCAATGGTTTTACCAATTATAGGTTTGTGTGAACTAGGTAACTGTCCTCAAACAACAGGCTGTGGTGTTCTTAGAGG AACAGCAAGACCTAAAGTTGGTGCTAACATAAACCTAGGATGTTTTTATCTTGTCGGAATGCCGGTTTCTGTTTGTCTAGCCTTCTATGATTGGAAATTTGAAGCTTCAAGAGCAAGGAAACTAGTGAATGGATCAGAAAGTGGTGGAGAATTGATCAAGACAAGTTAATTGGTT AAATATATGAAATTAATATTAGTATATATTTGTTCAAAAAATTTACCTAGTCAAAAATATATACGATTTTAAACAACATTTGAATG

CBF426_6484_FGTTTCCTCACTCTGTTTCACCTC

CBF426_6484_RGAATGGTTCTTTGAAGACAAAGACC

467bp

Tm 60°C

• Séquence du Marqueur 2 : CBF303_4830

Pisum_scaffold00303_3

Pisum contig023826

EVM_all_unmasked.4830

>Pisum_contig023826 Pisum_contig023826:808..1976 (- strand) class=gene length=1169



Annexes

CBF303_4830_F GGTATAAGAGATGGCAAAGTCCG CBF303_4830_RTCTCTACTCCAAAGTGTGGTACG 495bp Tm60°C

CBF303_4830_F2 CBF303_4830_R2 526bp 60°C

GTGGTTGTTATCTTCAGTGTGAGG CATGAGCGTAGAATCGGTAGCTG

• Séquence du Marqueur 3 : CBF303_4831

EVM_all_unmasked.4831

>Pisum contig023826 Pisum contig023826:3197..7801 (+ strand) class=gene length=4605 GCCGCCGCAGCCGGCCGGCCGGCCGGGCCGGGCCTCTGTTTTGGAAGGAGGAGACTCGGCTGCAGACTAGTTGTCCTTGCGCCC CTGCGCGCTCGAGGTTATTCCGGTAATGAACGAAGGCTTCGGATCGCGCGTATTCCTTGCGCAGTAAGCTACATAGTGGGAAGAAACTAAAGGTTTTAAATGTGAGCCTCGATGGATCGCGAAGGATTGCGCCTCTTATCTACTGCAATGAGAGACTAGCAAGGATTTCATGCTAG CACTTGCTAGAGACTGGTAAGCCACTGGTGCACACAGTACCCTGTCCTTAACAAAGGGGGTGCGATATATGCCATATATAGTAGATGT AGTCGGCCAAGGCTCCCAATGGGCAAGATGTCTGACGTGCACTTGATCGCTTTTCGTAGCCGCATCTGGTCTTTTTCGGGTACTAATCT ${\tt CCTGAAGAACCAGATTGAGCCTGCCTCTAAGCCGTCGTCTGCGTACCCGGACGTTTAATAGGCCCTCATGCACCAGACCACTTCCCAG}$ ${\tt CGCACACGGTCATTACCATTTTCCTCGCATTATTTCATGTTCAATCTGGCGTGCTATGGTTTTGTTTTCGACCTGGCTCTGTATGTTTGCG}$ ATCGCATAGTTGTGCAGTTCACTGCATGCAGTTGTCGTCGCCGAGCTGAAGTTCTTATCATCGCTCACAAAGACTGCAGCGAGTCGGCC GGGCAACACGGAACAGATGCACAAATCCACGTCTGTTCACCTATGAGCAAATCTTCTATTCATCCCAAATCGAGAGCCACAACGAGAC GGGTAATCTCAAAAATTATGAATACAATTTCACAATTCACATCCTA CCTTCCTTTCCCTATTGCATTTCGGCAACTCGCAAGATTGTGAAATAAGGCTCCAGCTTTCAAATTTTAGCACACACTCAACCGCAAGA $\mathsf{CAAACGGGCTTCAGTATCAATTTTGGGACAAACAGTATAGCAGGGGTGCAGACAACTGAATGATGGTTAACACCAAATGCAAAGGTT$ ATATCCGTCGCTGTAGGCTGGTCCAGCATAAAGCTGCCACCCAGAGATCTAAGCATGGAAGAAAATTCACATCAACTTGCATGGAACA CGAGTGGGTTGGCGAGCGCCAGCGTAAGGACATCGTCGAACACCAGAGAATATCACAAGAGACGACGACGACAAGTCCGATCTCTAGG TTGGTTCGACACCAAATTGTCCACACCCTAAGTTCCAAAAAATCATATCTAGCAACACAATGAGGTAATTGCACAAATCTTTATACCA



Annexes

AAATGGAGAGATCGAATGAAAGGAGTGCCTCCATTACCCTCGGACATATAAACATTATCTTCGAAGGCTCATATTTGGAACTAGTTCC GTTCCACGTGTTACGAATTCCTTCTATCCATCAAGGAATGGCAACCACGGATGTTTACCTCCGGCGATTATTCCGCAGTCACTTAGTCC GTCACCAACGAACAATCATTTGTATTCTATCAATGTGTGGGAATCATAGAGTGCCAGCAGCCTAATTGCGGTTAATATTCTACACTTGGG GTTATGCACTATCGCACAAATTGATGAAACATGAAGCACATAGCACAGTGCAAGAGGTGATTACTAATCATACTCCTTACCCCAGCGA TCCTCTTCTCTGGCTCCTCCAAGTCAATTCGGATGAGGATGTTTTCTCGTCGACCCATTCGTAGGTCACCACCCATGGGCAGGACATTC AACCTCTATTTTTTCCCTTCCGAGTTTCAAAACCTTTTTCTATGTTCCGCGTTAATCATTGCACTAGATCTCACATTTGGTTTTCAACATT TTTTTTCTTTTATTGAGCAAGTAGGCTTCTCGACCAGCTCGCTGTCTGGACCCAGTTCCACCAACAGCGGGTATGTACATTAATCGCGA TGAGGAGCTTCAAACATAAACCGTATCCATTTCGGTGCTTTCGCCCCCGCGCATTCCGTCAGTCCACAAATCGAATGGCGAGCACCG GTAGACATGGATTCAGATCATCGTGACCGTTACTTGTAGTAGGGAAAAGTGTTTCTTGGATTCGGCCATGCAAGCCCTACCGAATTTTAACAGGTAGACGCGTTATGCCATAACAACAACACAAGACACTGAACAACGTTTTGCTGGATTCGCCCTGTCCTGAGGTTCTTTTATTAGTC GCCATTGCGTTTCCGAAAGGAGCCAGAACTGTCTAAACTTACACACCCAGTCGGA<u>GCAAAGTAGCATAGCTGTGCATGG</u>GGCAAGCATG GTGGAAGCTAACAATCCAGCATCCATACTTTCCCTTTCCTCTCCACAATACAATCTATCAACGTTTCAACGTTATTTTGCTATACTTTCG GAACAATAAGTGAGGAAAGAATCGCCAGACCACTTGCCCGTATCGTACCACTAGACATTCACCCTCCACATGGAGTAATCCTCCTAGT GCTCTTACATCCATCTCCTCAGCTTATGATCGTCAACGTTGTATTGGACCAAGCAGTTGACAGGGAGTTTACAAGACACATATCATCTT ATTTAGAAAGAAATAAGGGATTAGTTCGAGCATGTTTACCACACTTGGATTTGACCCAGAAAATCAGTGTGGGAATCAAGTGGATTTG TTTTGGCCGTGTGATGTGTGCTGAACCAGATGGCCTTATGCTTGCCAAATGTGGATTATGCTTCGTGTCACGTTAG

CBF303_4831_FGATGTCTCTCACTGGTTTCACAG

CBF303_4831_RATGCACAGCTATGCCATGTTGC

485bp

Tm 60°C

CBF303_4831_F2	CTATATATCCAGCGTGCACCAAGG
CBF303_4831_R2	CACGATTGGATAGTGATGTTGTAAGC
Tm 61°C 465 hn	

• Séquence du Marqueur 4 : CBF303_4832

EVM_all_unmasked.4832

>Pisum_contig023845 Pisum_contig023845:21910..23948 (- strand) class=gene length=2039



Annexes

CBF303_4832_FGGACTTGTCGGAATCAAGGATC

CBF303_4832_RTGCAACTTCTGTACTTTCTCTAACC

435bp

Tm60°C

• Séquence du Marqueur 5 : CBF303_4834

Pisum_contig023859

PsCam050192

EVM_all_unmasked.4834

CBF303_4834_FGATAAATGGGTATGCGAAATGAGAG

CBF303_4834_RGTTCCATAGTGATACCTCGTCATC

474bp

Tm 60°C

• Séquence du Marqueur 6 : CBF211_3603

 $Pisum_scaffold00211$

Pisum_contig017821 Pisum_contig017847

EVM_all_unmasked.3603

>Pisum_contig017830 Pisum_contig017830:5021..6362 (+ strand) class=gene length=1342 ATGGTCGGTGCTGGATTTGCTTTTGCAGGATTAATGGTATACATTTGGTTCAATTTTGGCCTGTGCATGATGGCATGGTAGGGAGAATG TTTGATGTTACTAACTTGTGAACATGTTCAATTAAGTCTCAATGTTGCTGAATTTCCAGGATTTATAGCAGTGTTTGGAGTATGGACCTG



Annexes

CBF211_3603_FTAGCCAATCCTGTAGGTCAGTG

CBF211_3603_RACGGTTGATGACAGATTGATTAGC

439bp

Tm 60°C

• Séquence du Marqueur 7 : CBF211_3604

EVM_all_unmasked.3604

>Pisum_contig017830 Pisum_contig017830:6577..7778 (+ strand) class=gene length=1202

CBF211_3604_FCTTTTGCTGATCAGAGAGCCTAC

CBF211_3604_RTGAGGTTCCTTCTGTGAGTAGAC

484bp

Tm 60°C

• Séquence du Marqueur 8 : CBF211_3605

EVM_all_unmasked.3605



Annexes

AACAATTCCGCACGCATCACACTGCTATGCAGTACCTCTATAATTCATATCAGCAATAGAAAAGAATCAGATAAAGGTGGAAGTAGAA GAAGAGCAAGAGCAATACTCATTCCCGAAAAACCCCCGCCACCGATTTTCAAATTTGCCTTGAATAGCCAGACGTCTATTTACTTCACA AGATGGATTTTCCTCAGATTTGCGCTATTTCAGTGTCGTTGCTTGGTGGGAAAGGTGAATGCCATAGCTTATGCTTCTGAAGCTTGC ATCAGTGATCCAAGCTCAATTCTAAGTCCATCATTTTGGATTCATGTGCATATTGTGATCCCAGAGAGGCCAACTGAATGTGCAGTATT ${\sf CAATGTCATTATCATTTGCATCAGTGATCCAAGCTCAATTCTAAGTCCATCATTTTGGATTCATGTGCATATTGTGATCCCAGAGAGGCC}$ TGGCAGTGTGAGCATGAGTGGCGGCAGAATATTTCAGTTGCTACAGAGTGGCTATCAGCGGTGTCTCCGTATCGGTGGATTTCATCCA AATCAGGTGTTCTTGCCAATTCAAATTCAGTATTTGAGAAAAACTTTACTTCACCAACAATTGACACTTCAGCTCCAACAACAGTCTCA AGCTCATGTAACACAATGCATGGTACAGTAGAAAAACCCAGGAGTGATGGTGTTCGCAATTAAGGTTCTCTAAAGTTATGATGGAAACC ATGTAGTTCATCTGTCCTATCTCGAGGTTTATAACGATACTGTAAGAGATTTGATTTCACCTGGAAGGCCTCTTGTTTTGAGAGCAGCA GGCCTTACGCAATACAAAGTTTACTCCGCCGATAAAGTGATGGCTAACGAAGTCGTTCCTTTAGCATTGTCTTGATCGGAACCAAGCTC AAGGATCGTCTTCTAGGAATTCATCTGAACACAATGCTCGTGGAGAAAAGCAGATGTTGCTTGAAGTTGTTGATGCTGCTGACGGGAC GGCAAGCGCTTCATGCGTGAAGGAACATGTTGTGAAAAACCACACTGGATGCATCTCAATCTAGCCAAGTTGCGACTCCTAGAGATATT GAGGATTTTGGCAGGTCTTTAAGACCAAACACTTTCTTGCATTATAATTTCTCTACACTAAATCAATTCCAGTCTAAGAAAAATATGGA TATCAATCCTATTGATCAAGATGTCAAACAATTCAAAGTATCAGATGATGTTGGGGGATAGACAGTTTGATTCTAATCATGGACAACGGTC ATACGGATACACCTACACAGTTGAAGATGTTTTGGTTCTGTTTTTAAGGACCATGGTGGTAGTGACATTGACACTAGTAGCGGCATTCC AATCACAGTTCATGAGGCAATAG

CBF211_3605_F GGTTCTCTAAAGTTATGATGGAAACC CBF211_3605_RGTCTATCCCCAACATCATCTGATAC 531bp Tm 60°C

Zone répétée

• Séquence du Marqueur 9 : CBF211_3606

EVM_all_unmasked.3606

>Pisum_contig017831 Pisum_contig017831:49758..50559 (+ strand) class=gene length=802

CBF211_3606_FATTACGAGTTCCAGAGCGACAAC CBF211_3606_RCATCCCAAGATTTGCAGAATGGTC

419bp

Tm 60°C

• Séquence du Marqueur 10 : CBF211_3607

EVM_all_unmasked.3607

>Pisum_contig017834 Pisum_contig017834:2079..3343 (- strand) class=gene length=1265


Centre INRA de Toulouse – CNRGV

Annexes

$CBF211_3608_FTCAATGTCATATCAGATTCTCCTCG$

CBF211_3608_RCGTTATAAACCTCGAGATAGGACAG

480bp

Tm 60°C

• Séquence du Marqueur 11: CBF211_3608

EVM_all_unmasked.3608

>Pisum_contig017834 Pisum_contig017834:4778..5627 (+ strand) class=gene length=850

Remarque: Ce marqueur eut être criblé par les mêmes amorces de celles du marqueur CBF211_3606 (EVM all unmasked.3606)

CBF211 3608 FATTACGAGTTCCAGAGCGACAAC

CBF211_3608_RCATCCCAAGATTTGCAGAATGGTC

419bp

Tm 60°C

• Séquence du Marqueur 12: CBF211_3609

EVM_all_unmasked.3609

>Pisum_contig017837 Pisum_contig017837:15002..15694 (- strand) class=gene length=693



Centre INRA de Toulouse – CNRGV

Annexes

 $CBF211_3609_FTGGGTGTGTGAAATGAGAGAGC$

CBF211_3609_RACATCTCAACATCTTCAAGGTCTTC

444bp

Tm 60°C

• Séquence du Marqueur 13: CBF211_3610

EVM_all_unmasked.3610

>Pisum_contig017837 Pisum_contig017837:32116..32871 (+ strand) class=gene length=756

ATGTTTCCAACTAACAGCTCTTCCACTTCACAACCCAGTTCCTCAAAACCTTCCACTTCCTCCAACAACAATGCCAAAATTTAGAGGT CTCTCATGACATGTCAGTTTTCAACGAGGAACTGCGGGTTAGC<u>TGCTAGTACACCGAAGAAACGTG</u>CAGGGAGGAAGAAGATCAAGGA GACTCGCCACCCTGTGTATAGGGGTGTAAGGAGGAGGAACTTAGATAAATGGGTCTGTGAAATGAGGGAGCCAAACAAGAAGAACTAA AATTTGGCTAGGAACTTTTCCGACAGCTGAGATGGCAGCCGAGCTCATGATGTTGCAGCATTGGCATTAAGAGGTCGTTACGCTTGTC TCAACTTTAGAGACTCCGCATGGCGACTCCCTGTTCCAGCCACTACCGAGACAAAAGATATACAAAAGGCAGCTGCACTGGCCGCCGA AGCTTTCAGACCAGCCAACAAGACTTTAGAACCAGATGAGGCTTTTAGAACCAGAACAAAAGATATACAAAAGGCAGCTGCACTGGCCGCCGA *ACAA*GACTTTAATATCTAATGACATTGACACGGCAGTAGCTGCAACAACAGAAGAGAACATTTAGTTTCGTATGGAAGAGAAGAAGA AGAGAAAGAGTTGAGCATTCCAGAAATGTTAAAGAATATGGCACTAATGTCCCCAACACATAGCTTGGGACATGAGGATGAACATAT TGATGCAGACTTTGAAGATGTTGAGGTGTCACTATGGAAGTTTTCAATTTGA

CBF211_3610_FTGCTAGTACACCGAAGAAACGTG CBF211_3610_RTGTCTGGTTCGAAAGCATCATCTG 402bp Tm 60°C