

MEBAREK LAMARA

**ANALYSE COMPARATIVE DE LA DIVERSITÉ
GÉNÉTIQUE ET DE LA STRUCTURE DES
POPULATIONS CHEZ L'ORGE (*HORDEUM VULGARE*
L.) À L'AIDE DE MARQUEURS SSR, DART ET DU
PEDIGREE**

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en biologie végétale
pour l'obtention du grade de Maître ès Sciences (M.Sc.)

DÉPARTEMENT DE PHYTOLOGIE
FACULTÉ DES SCIENCES DE L'AGRICULTURE ET DE L'ALIMENTATION
UNIVERSITÉ LAVAL
QUÉBEC
2010

Résumé

La diversité génétique de 92 cultivars d'orges canadiennes a été mesurée en utilisant deux types des marqueurs moléculaires (SSR et DArT) et le pedigree. Au total, 368 allèles ont été identifiés aux 50 locus SSR et utilisés pour créer une matrice de distance génétique. Le nombre d'allèles par locus variait de 2 à 13 ($\bar{x} = 7,36$) et l'indice PIC variait entre 0,34 et 0,86 ($\bar{x} = 0,69$). Pour les marqueurs bialléliques DArT, la matrice de distance génétique s'est basée sur 971 marqueurs dont les indices PIC variaient entre 0,06 et 0,50 ($\bar{x} = 0,39$). Une troisième matrice de distance a été calculée basée sur le coefficient de parenté. Un groupement des génotypes a été effectué basé sur les trois matrices de distances génétiques et les dendrogrammes obtenus ont montré des relations génétiques parmi les cultivars d'orge. Une comparaison de la similarité topologique des trois dendrogrammes réalisée en utilisant l'indice de congruence a montré une très bonne concordance entre les trois dendrogrammes. Des analyses statistiques ont par ailleurs montré une corrélation fortement significative entre les matrices SSR et DArT ($r = 0,80$, $p < 0,002$) mais une plus faible corrélation positive du pedigree avec les deux types de marqueur ($r = 0,46$, $p < 0,002$; $r = 0,52$, $p < 0,002$). Ces informations comparant les résultats de différentes méthodes d'estimation de la diversité génétique seront utiles pour l'amélioration et la conservation des ressources génétiques d'orge.

Abstract

Genetic diversity of 92 Canadian barley cultivars was measured using two types of molecular markers (SSRs and DArTs) and pedigree data. A total of 368 alleles were identified at 50 SSR loci and used to create a genetic distance matrix. The number of alleles per locus ranged between 2 and 13 ($\bar{x} = 7.36$) and PIC values ranged from 0.34 to 0.86 ($\bar{x} = 0.69$). For the biallelic DArT markers, the genetic distance matrix was based on 971 markers whose PIC values ranged between 0.06 and 0.50 ($\bar{x} = 0.39$). A third distance matrix was computed based on the kinship coefficient, itself based on pedigree information. Clustering of genotypes was performed based on each of three genetic distance matrices and the dendrograms obtained showed the genetic relationships among barley cultivars. The topological similarity of the three dendrograms was estimated using a congruence index and showed that the three dendrograms are in very good agreement. Statistical analyses also showed a highly significant correlation between the SSR and DArT matrices ($r = 0.80$, $p < 0.002$) and lower yet significant correlations of the pedigree data with both marker types ($r = 0.46$, $p < 0.002$; $r = 0.52$, $p < 0.002$). Information obtained by comparing results of different genetic diversity estimation methods will be particularly important for the improvement and conservation of barley genetic resources.

Avant-Propos

Je remercie mon directeur de recherche, Monsieur François Belzile, qui m'a donné l'opportunité de travailler dans son laboratoire, et m'a confié un projet de recherche très stimulant. Je le remercie pour la qualité de son encadrement et sa grande disponibilité tout au long du projet. Daignez agréer, Monsieur Belzile, mes respectueuses salutations.

Je remercie également tous les membres de l'équipe qui ont contribué de près ou de loin à la réalisation de ce travail. Il s'agit de:

Monsieur Tung Huynh Thanh, pour son aide et sa patience dans le traitement des données. Les échanges de point de vue au plan scientifique et son amitié m'ont été très bénéfiques.

Madame Suzanne Marchand, pour sa collaboration et ses conseils ont été d'un grand apport pour la réussite de ce travail.

Madame Martine Jean, pour sa disponibilité et ses précieux conseils au moment de la présentation du projet de recherche et du séminaire de fin d'études.

Je remercie la COOP Fédérée pour le soutien matériel du projet.

Mes remerciements s'adressent aussi à mes camarades et collègues, Patricio Esteves, Mahamadou Sangaré, Elmer Iquira, Gabriel Richard, Li Yi Zhang, Marie-Pier Dubé, Mélanie Fortier, Mélanie Gauthier et Maxime Bastien, qui ont contribué à maintenir la joie et le dynamisme dans le laboratoire.

Je tiens à remercier toute ma famille en Algérie et mes amis, Farid Mezmiz et Youcef Boucherim, qui m'ont toujours soutenu.

Table des matières

Résumé	i
Abstract.....	ii
Avant-Propos.....	iii
Table des matières	iv
Liste des figures.....	vii
1. CHAPITRE I : INTRODUCTION GÉNÉRALE.....	1
1.1 Introduction générale	2
2. CHAPITRE II : REVUE DE LA LITTÉRATURE.....	4
2.1 Revue de la littérature	5
2.1.1 Les marqueurs génétiques utilisés dans l'estimation de la diversité	5
2.1.1.1 Les marqueurs morphologiques	5
2.1.1.2 Les marqueurs biochimiques.....	5
2.1.1.3 Les marqueurs de l'ADN	6
2.1.2 Caractéristiques des principaux types de marqueurs moléculaires employés chez l'orge	7
2.1.2.1 Les marqueurs RFLP.....	7
2.1.2.2 Les marqueurs microsatellites (SSR).....	9
2.1.2.3 Les marqueurs RAPD.....	11
2.1.2.4 Les marqueurs AFLP	13
2.1.2.5 Les marqueurs SNP	15
2.1.2.6 Les marqueurs DArT.....	18
2.1.3 Comparaison des marqueurs moléculaires employés chez l'orge	19
2.1.4 Mesures de la diversité	20
2.1.4.1 Diversité intraspécifique	20
2.1.4.2 Diversité interspécifique	21
2.1.4.3 Distance génétique	22
2.1.5 Représentation graphique	22
2.1.6 Le pedigree	23
2.1.7 Hypothèses de recherches.....	24
2.1.8 Objectifs de recherche	24
3. CHAPITRE III: MANUSCRIT	26
3.1 Résumé.....	28
3.2 Abstract	29
3.3 Introduction.....	30
3.4 Materials and methods	31
3.4.1 Plant materials	31

3.4.2	DNA extraction.....	32
3.4.3	SSR and DArT analyses	32
3.4.4	Pedigree analysis	33
3.4.5	Comparisons of tree topology and matrices	33
3.5	Results.....	34
3.5.1	Genetic diversity assessed with SSR and DArT markers.....	34
3.5.2	Genetic distances	35
3.5.3	Genetic relationships between cultivars	35
3.6	Discussion.....	36
	References.....	40
4.	CHAPITRE IV : CONCLUSION GÉNÉRALE	51
4.1	Conclusion générale.....	52
	Références	54
	LISTE DES ANNEXES	59

Liste des tableaux

Tableau 2-1 - Comparaison des six marqueurs moléculaires les plus utilisés chez l'orge.....	20
Table 3-1 - List of Canadian barley cultivars used in this work. Row number (R), geographic region of origin: (W for Western Canada, E for Eastern Canada) are indicated.....	43
Table 3-2 - Summary data on the polymorphism and efficiency of SSR and DArT markers in a collection of 92 Canadian barley lines.....	34
Table 1S - Linkage group, position (cM), number of alleles detected, PIC, and repeat motif of the 50 SSRs used in this study.....	47

Liste des figures

Figure 2-1 - Le polymorphisme détecté par les marqueurs RFLP..	9
Figure 2-2 - Le polymorphisme révélé par les marqueurs SSR.	11
Figure 2-3 - Le polymorphisme révélé par les marqueurs RAPD.	13
Figure 2-4 - Le polymorphisme révélé par les marqueurs AFLP.	15
Figure 2-5 - Polymorphisme révélé par les marqueurs SNP (A) Hybridation d'oligonucléotides spécifiques d'allèles, (B) Réactions d'extension d'amorce, (C) Ligature d'oligonucléotides, (D) Clivage invasif.....	17
Figure 2-6 - Le polymorphisme révélé par les marqueurs DArT.	19
Figure 3-1- Correlation between pairwise genetic distance values computed using SSR, DArT and pedigree data	44
Figure 3-2 - Relationships between barley lines visualized by Principal-coordinate analysis (PCoA) using SSR data (A) or DArT data (B).....	45
Figure 3-3 - Dendrograms of 92 Canadian barley varieties obtained using SSR (A), DArT (B) or pedigree (C) data. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).....	46
Figure 1S - Dendrogram of 92 Canadian barley varieties obtained using SSR markers. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).....	48
Figure 2S - Dendrogram of 92 Canadian barley varieties obtained using DArT markers. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).....	49
Figure 3S - Dendrogram of 92 Canadian barley varieties obtained using pedigree data. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).....	50

Liste des annexes

LISTE DES ANNEXES	59
ANNEXE I - Plant DNA Extraction Protocol for DArT	60

CHAPITRE I : INTRODUCTION GÉNÉRALE

1.1 Introduction générale

L'orge (*Hordeum vulgare* L.) est l'une des plus anciennes plantes cultivées par les humains. Elle résulterait de deux événements de domestication distincts qui se seraient produits dans le « croissant fertile » au Moyen-Orient et, plus à l'Est, en Asie centrale (Morrell et Clegg 2007). Aujourd'hui, elle constitue l'une des céréales les plus cultivées au monde, puisqu'elle occupe le quatrième rang en matière de production céréalière (après le maïs, le riz et le blé) avec une production de 137 millions de tonnes (Agriculture et Agroalimentaire Canada 2007).

Cette graminée possède des caractéristiques qui font d'elle une culture très appréciée pour une variété d'usages. On évalue qu'environ 70 % de l'orge produite sert à l'alimentation des animaux, 20 % au maltage et 5 % à l'alimentation humaine (Agriculture et Agroalimentaire Canada 2005). L'orge n'est pas uniquement une céréale d'importance économique mais aussi un modèle pour des études génétiques et physiologiques (Koornneef et coll. 1997). C'est une espèce autogame, diploïde ($2n = 14$) avec une taille de génome de 5,5 pg d'ADN par noyau haploïde, qui correspond à $5,3 \times 10^9$ pb (Bennett et Smith 1976).

L'intensification et l'industrialisation tant de la production que de l'amélioration variétale ont engendré une réduction importante de la diversité génétique chez l'orge et plusieurs autres plantes cultivées au cours des dernières décennies (Russell et coll. 2004). Ces nouvelles lignées génétiquement homogènes sont souvent plus sensibles aux attaques pathogènes et moins résistantes aux stress liés aux conditions environnementales (Asins et Carbonell 1989). Il importe donc de mesurer, de maximiser et d'exploiter au mieux la diversité génétique dans le cadre d'un programme d'amélioration génétique. Pour un sélectionneur, il serait très utile de connaître le degré de parenté génétique entre deux lignées qu'il envisage de croiser. Ainsi, il pourrait choisir des parents offrant les caractères agronomiques recherchés mais tout en ayant le contraste génotypique le plus grand de façon à maximiser la diversité génétique au sein de la progéniture en sélection.

Cette diversité peut être estimée sur la base de plusieurs critères dont le phénotype, le pedigree et à l'aide de marqueurs moléculaires (Matus et Hayes 2002). Depuis une

vingtaine d'années, les marqueurs moléculaires sont devenus les outils de choix dans la mesure de la variabilité génétique en raison de leurs nombreuses qualités. D'abord, ils révèlent directement les modifications du patrimoine génétique, c'est-à-dire de l'ADN, qu'elles se traduisent ou non par une modification phénotypique. Ensuite, contrairement aux marqueurs traditionnels (morphologiques et biochimiques), les marqueurs moléculaires ne sont pas influencés par les fluctuations de l'environnement, ils sont indépendants de l'organe analysé et du stade de développement de la plante et ils ne sont pas épistatiques. Finalement, il est devenu possible de procéder au génotypage des individus à l'aide de centaines, voire de milliers, de marqueurs, ce qui confère à ces analyses une précision inégalée.

Nous postulons, dans le cadre de cette étude, qu'une comparaison de la diversité génétique obtenue à l'aide des marqueurs moléculaires SSR et DArT avec celle du pedigree peut permettre de documenter avec précision les liens qui existent entre les divers groupes de lignées d'orges cultivées au Canada. Cette information peut s'avérer très utile pour la gestion des ressources génétiques et leur utilisation raisonnée dans les programmes de sélection.

CHAPITRE II : REVUE DE LA LITTÉRATURE

2.1 Revue de la littérature

2.1.1 Les marqueurs génétiques utilisés dans l'estimation de la diversité

La diversité génétique est le résultat de plusieurs phénomènes dont la sélection, les mutations, la migration et la dérive génétique. Ces forces évolutives ont toutes des effets sur l'importance et l'organisation de la variabilité génétique. Les effets de ces forces portent sur l'ampleur et l'organisation du polymorphisme : la mutation et la migration augmentent la diversité alors que la sélection et la dérive la diminuent. L'utilisation de la diversité génétique dans un programme de sélection passe inévitablement par son estimation et par le choix du type de marqueur susceptible de la traduire le plus fidèlement possible. Les marqueurs génétiques sont des caractéristiques héréditaires qui renseignent sur le génotype de l'individu qui les porte (de Vienne 1998). Trois types sont largement utilisés pour l'évaluation de la variabilité génétique, à savoir les marqueurs morphologiques, les marqueurs biochimiques et les marqueurs de l'ADN.

2.1.1.1 Les marqueurs morphologiques

En sélection classique, les sélectionneurs recherchent des marqueurs simples du génome des plantes pour suivre les variations de différents caractères d'intérêt agronomique. Chaque lignée est examinée à travers son aspect phénotypique, c'est-à-dire la résultante de l'expression de ses gènes dans un milieu donné. Les marqueurs morphologiques ont été retenus dans un premier temps parce qu'ils présentaient l'avantage d'être immédiatement disponibles et nécessitent seulement un équipement simple. Ils représentent des variations de type qualitatif (couleur..), morphologiques (forme), des résistances à des maladies ou des ravageurs. Ces critères sont utilisés pour décrire et identifier les lignées et les variétés chez les végétaux. Cependant, ces marqueurs sont peu polymorphes et en général dominants, ils ont fréquemment une base génétique complexe, sont limités en nombre et leur expression est souvent fortement influencée par l'environnement.

2.1.1.2 Les marqueurs biochimiques

Il est important pour le sélectionneur de connaître l'aspect purement génétique du caractère

étudié, indépendamment de son expression. A cet effet, de nombreux marqueurs biochimiques ou moléculaires ont été développés pour l'orge depuis une vingtaine d'années. Les marqueurs biochimiques, généralement des isozymes ou des protéines de réserve, ont un pouvoir de discrimination plus élevé que les marqueurs morphologiques, traduisent directement l'activité des gènes, sont neutres et peuvent être révélés dans plusieurs organes à différents stades de développement (Buth et Murphy 1999; Bank et coll. 2001). Ils sont mis en évidence par électrophorèse, sur la base des propriétés de migration des protéines, et révélés par la coloration histochimique spécifique des enzymes qui sont analysées. Les protéines ont l'avantage d'être plus nombreuses que les marqueurs morphologiques, mais dépendent également de l'organe considéré (de Vienne 1998).

Les limitations majeures à l'emploi des isozymes sont le faible nombre de locus susceptibles d'être révélés (rarement plus de quelques dizaines) et la spécificité de l'organe analysé. De plus, les marqueurs biochimiques sont soumis à l'influence de l'environnement. En conséquence, la capacité d'utiliser un marqueur pourra varier selon l'environnement dans lequel la plante aura été cultivée, une situation qui est évidemment indésirable.

2.1.1.3 Les marqueurs moléculaires

Un bon marqueur doit être à hérédité simple, polymorphe, co-dominant, non épistatique, neutre, et insensible au milieu (Bretting et Widrechner 1995; de Vienne 1998). Les marqueurs moléculaires répondent bien à ces critères et, contrairement aux marqueurs associés à des caractéristiques morphologiques ou biochimiques, révèlent directement les modifications du patrimoine génétique, c'est-à-dire de l'ADN, qu'elles se traduisent ou non par une modification phénotypique. Ce polymorphisme de l'ADN résulte de l'accumulation graduelle de mutations, produisant soit des substitutions nucléotidiques, soit des insertions ou délétions d'une ou plusieurs bases. L'analyse par ces marqueurs peut être réalisée tout au long du développement de la plante. Ces marqueurs ont le potentiel d'exister en nombre illimité, couvrant le génome entier, et ils ne sont pas influencés par l'environnement.

2.1.2 Caractéristiques des principaux types de marqueurs moléculaires employés chez l'orge

De nombreux marqueurs moléculaires ont été mis à la disposition des chercheurs, et de nouveaux sont régulièrement décrits. Vu l'impossibilité de dresser une liste exhaustive de la totalité de ces marqueurs, nous décrirons, par ordre chronologique de leur mise au point, un certain nombre de marqueurs qui sont ou ont déjà été les plus utilisés chez l'orge : RFLP («Restriction Fragment Length Polymorphism»), SSR («Simple Sequence Repeats»), RAPD («Random Amplified Polymorphic DNA») et AFLP («Amplified Fragment Length Polymorphism»). Récemment, deux nouvelles technologies de marqueurs moléculaires à haut débit ont été développées chez l'orge: les marqueurs SNP («Single Nucleotide Polymorphism») et DArT (« Diversity Arrays Technology »).

2.1.2.1 Les marqueurs RFLP

La technique RFLP développée par Botstein et coll. (1980) a été la première technique utilisée pour l'analyse de la diversité génétique chez l'orge (Saghai-Maroo et coll. 1984). Elle repose sur la mise en évidence de la variabilité de la séquence nucléotidique de l'ADN génomique après digestion par des enzymes de restriction (Figure 2-1). Les fragments générés par digestion enzymatique de l'ADN de deux individus (identifiés A et B dans la figure) sont séparés selon leur taille sur gel d'agarose, dénaturés et transférés par capillarité sur une membrane de nylon. La position relative des fragments d'ADN est conservée durant le transfert. La dernière étape consiste en l'hybridation moléculaire avec une sonde d'ADN marquée préalablement.

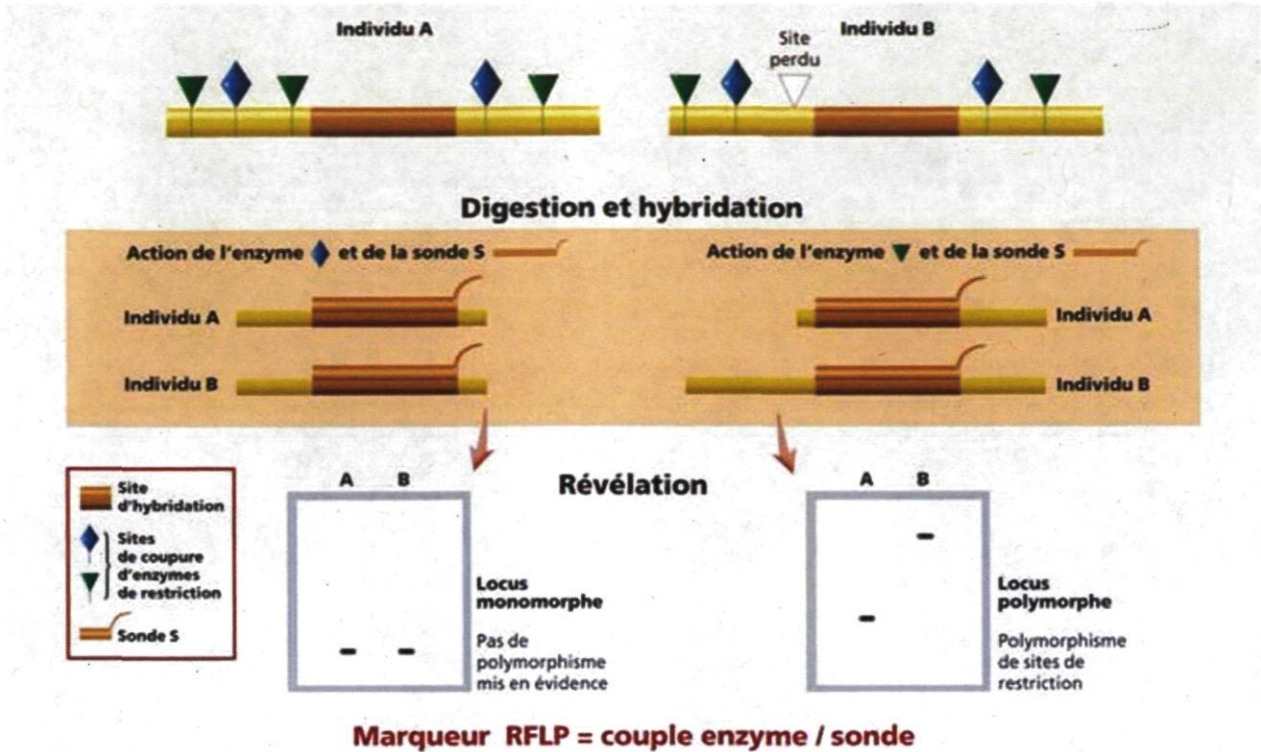
Dans un premier cas de figure illustré, la sonde employée permet de détecter une région du génome où la digestion (avec l'enzyme dont les sites de restriction sont indiqués à l'aide du symbole ♦) donne des fragments de restriction identiques pour les deux individus. Les deux profils révélés après électrophorèse ne permettent donc pas de les distinguer. Avec ce couple enzyme ♦/sonde S, aucun polymorphisme n'est mis en évidence entre ces deux individus. En revanche, pour l'enzyme ▼, l'individu B présente une mutation au niveau d'un site de restriction, entraînant la perte de ce site. Ainsi, dans un second cas de figure, la digestion avec cette enzyme et l'emploi de la même sonde met en évidence un fragment plus petit chez l'individu A que chez

l'individu B. Dans ce cas, le couple enzyme ∇ /sonde S permet ainsi de mettre à jour le polymorphisme existant au niveau des sites de restriction dans cette région du génome.

Les marqueurs RFLP présentent plusieurs des qualités espérées d'un bon marqueur moléculaire. Ils peuvent être mono-locus ou multi-locus, codominants, bi- ou multialléliques. Ils permettent une analyse génétique complète, présentent un très bon potentiel de transfert entre espèces, même assez éloignées, grâce à certaines sondes de séquence très conservée (Helentjaris et coll. 1985) et sont généralement très reproductibles.

Les RFLP ne sont cependant pas des marqueurs parfaits pour un travail à grande échelle. En effet, cette technique prend du temps (plusieurs jours de travail en laboratoire), nécessite beaucoup d'ADN pour chaque analyse (quelques microgrammes) et présente un taux de polymorphisme plutôt faible chez les céréales, particulièrement chez le blé (Powell et coll. 1996). Finalement, les étapes de transfert et d'hybridation empêchent une automatisation du travail.

Néanmoins, en raison des qualités décrites plus haut (et en l'absence d'alternatives plus attrayantes), de nombreux travaux en amélioration des plantes ou en analyse de la diversité génétique ont fait appel à ces marqueurs. Chez l'orge, cette technique a été largement utilisée pour les analyses de la diversité (Melchinger et coll. 1994; Casas et coll. 1998), et plusieurs cartes génétiques ont été réalisées (de la Pena et coll. 1999; Graner et coll. 1991; Zhu et coll. 1999). Ces marqueurs ont été utilisés par certains laboratoires pour l'identification et l'isolement de gènes contrôlant des caractères d'intérêt agronomique en exploitant la synténie entre le blé et le riz (Kurata et coll. 1994).



(GNIS, 1999)

Figure 2-1 - Le polymorphisme détecté par les marqueurs RFLP.

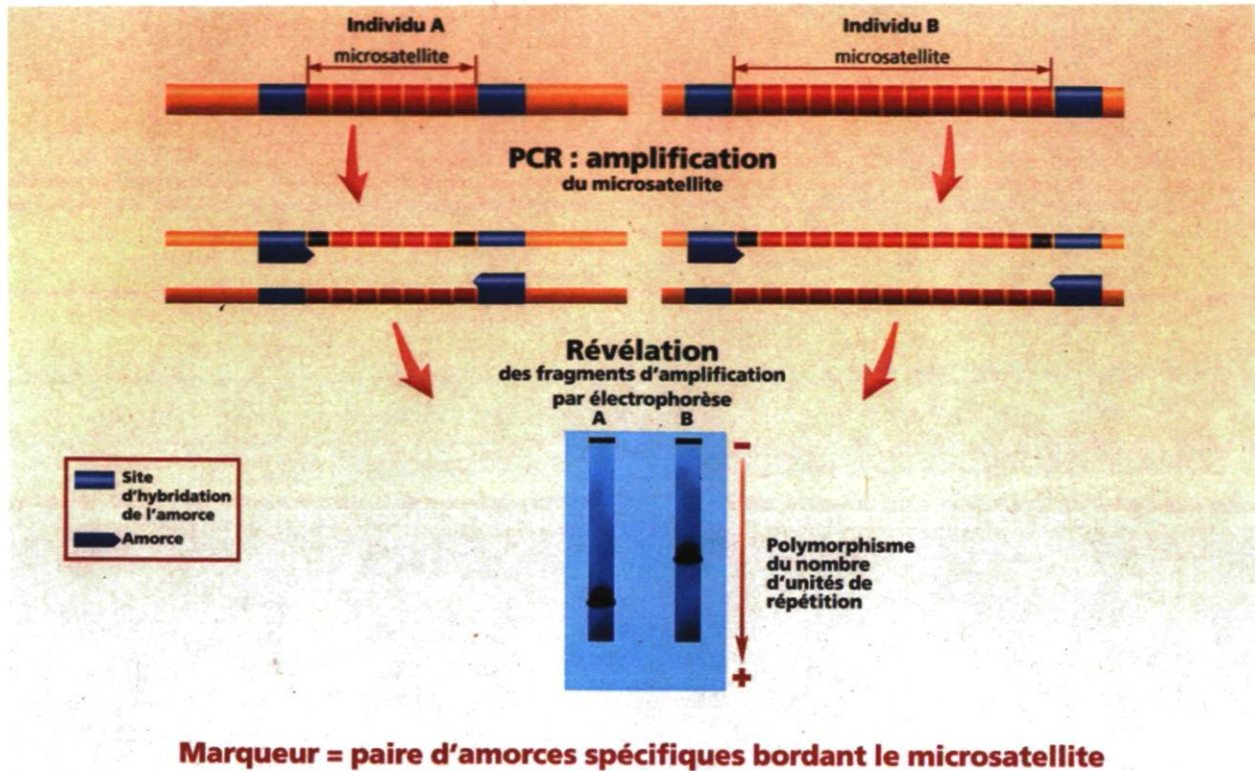
2.1.2.2 Les marqueurs microsatellites (SSR)

La découverte dans le génome nucléaire de nombreuses zones hypervariables a ouvert la voie au développement de nouveaux marqueurs pour la cartographie du génome humain. Ces zones sont appelées microsatellites (Litt et Litty 1989; Tautz 1989). Les microsatellites sont constitués de répétitions en tandem de motifs mono-, di-, tri-, ou tétra-nucléotidiques sur une longueur inférieure à 100 paires de bases (Tautz 1989). De tels motifs sont très nombreux dans le génome des organismes eucaryotes. Les SSR présentent un taux de polymorphisme extrêmement élevé qui repose sur la variation du nombre d'unités de répétition (n) constituant le microsatellite. Comme l'illustre la figure 2-2, chaque locus microsatellite est encadré par des séquences qui lui sont spécifiques (en bleu) et qui permettent son amplification par PCR («Polymerase Chain Reaction»). L'analyse des produits amplifiés s'effectue sur gel de polyacrylamide et le polymorphisme entre

les individus se traduit par une différence de taille des allèles. Un individu B, possédant plus de copies du motif répété (rectangles rouges) que l'individu A, produira un amplicon de taille supérieure qui migrera plus lentement que l'amplicon obtenu pour l'individu A.

Les microsatellites sont des marqueurs codominants, extrêmement polymorphes, faciles à manipuler et permettant le multiplexage. Ils requièrent très peu d'ADN (quelques nanogrammes) et sont bien répartis dans le génome. Ces marqueurs sont très reproductibles et moins lourds d'utilisation que les RFLP ou AFLP (décrits plus bas) avec une bonne résolution analytique. Un inconvénient reste cependant que leur développement nécessite un travail considérable dont les principales étapes sont : 1) le criblage d'une banque d'ADN génomique avec une sonde correspondant au motif répété recherché ; 2) le séquençage des clones positifs pour caractériser la séquence des régions bordant chaque microsatellite ; et 3) le développement d'amorces spécifiques pour chaque microsatellite et la vérification que ce microsatellite est polymorphe chez un groupe d'individus de l'espèce en question.

Ces marqueurs ont été très utiles pour des études de diversité génétique, phylogénétiques, ainsi que pour l'évaluation des ressources génétiques et la construction de cartes génétiques. Chez l'orge, leur grand nombre, leur polymorphisme et la facilité du génotypage font des SSR d'excellents marqueurs pour l'analyse de la diversité génétique (Saghai-Marouf et coll. 1994; Struss et Plieske 1998; Matus et Hayes 2002), le génotypage (Russell et coll. 1997; Pillen et coll. 2000) et la cartographie génétique (Liu et coll. 1996; Ramsay et coll. 2000).



(GNIS, 1999)

Figure 2-2 - Le polymorphisme révélé par les marqueurs SSR.

2.1.2.3 Les marqueurs RAPD

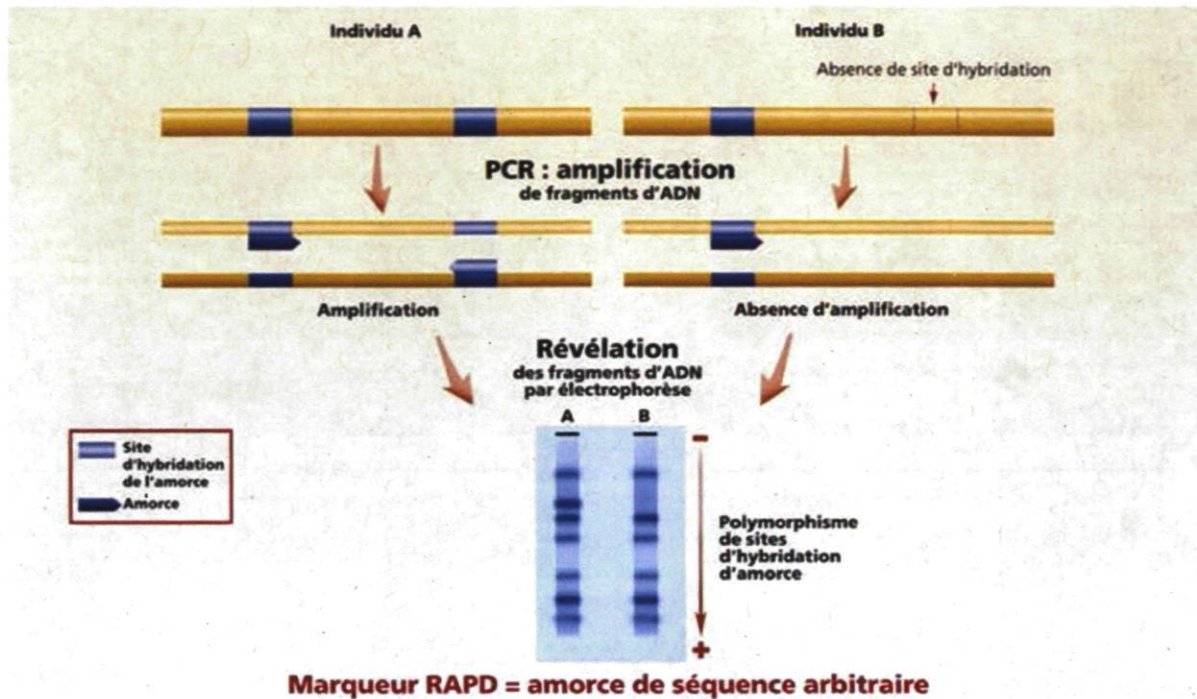
La technique RAPD développée par Williams et coll. (1990) consiste en l'amplification par PCR de fragments de l'ADN génomique en utilisant des amorces arbitraires de taille courte (10 pb). Une amorce RAPD permet généralement l'amplification d'une dizaine de fragments. Les fragments produits par cette amplification sont séparés par électrophorèse sur un gel d'agarose. Pour deux individus A et B (Figure 2-3), si deux sites d'hybridation sont proches et sur deux brins complémentaires d'ADN, il y aura amplification, comme c'est le cas de l'individu A. En revanche, si ces deux sites sont trop éloignés, il ne peut y avoir amplification, comme dans le cas de l'individu B. Ainsi, on observe la présence d'une bande sur le gel pour A et l'absence pour B de cette même bande. Le polymorphisme décelé est dû à des mutations soit dans les régions

amplifiées soit au niveau des sites de fixation des amorces. Le polymorphisme révélé est un polymorphisme de sites d'hybridation d'amorce traduit par la présence ou l'absence de la bande.

Cette technique est rapide et d'une faible technicité, elle génère un grand nombre de fragments à partir d'une faible quantité d'ADN et ne nécessite pas des informations génétiques ou génomiques préalables. Les marqueurs RAPD sont des marqueurs multilocus, bialléliques, dominants.

Bien que cette technique soit simple, le désavantage est que les marqueurs sont dominants et ne sont pas reproductibles (Williams et coll. 1990; Welsh et McClelland 1990). Les problèmes de reproductibilité des profils entre laboratoires (utilisation de conditions PCR non stringentes) ont défavorisé leur utilisation au profit des AFLP, plus récents.

Les marqueurs RAPD ont été utilisés chez plusieurs espèces pour l'analyse de la diversité génétique, la caractérisation de ressources génétiques, l'identification de cultivars et la cartographie des génomes. Chez l'orge, ces marqueurs ont été utilisés pour déterminer la diversité génétique (Baum et coll. 1997; Matus et Hayes 2002), en cartographie génétique (Dávila et coll. 1999) et pour l'identification de cultivars (Tinker et coll. 1993). Dans ce dernier exemple, les auteurs ont montré qu'avec neuf marqueurs RAPD polymorphes et un total de sept amorces arbitraires on pouvait distinguer 27 lignées d'orge cultivées.



(GNIS, 1999)

Figure 2-3 - Le polymorphisme révélé par les marqueurs RAPD.

2.1.2.4 Les marqueurs AFLP

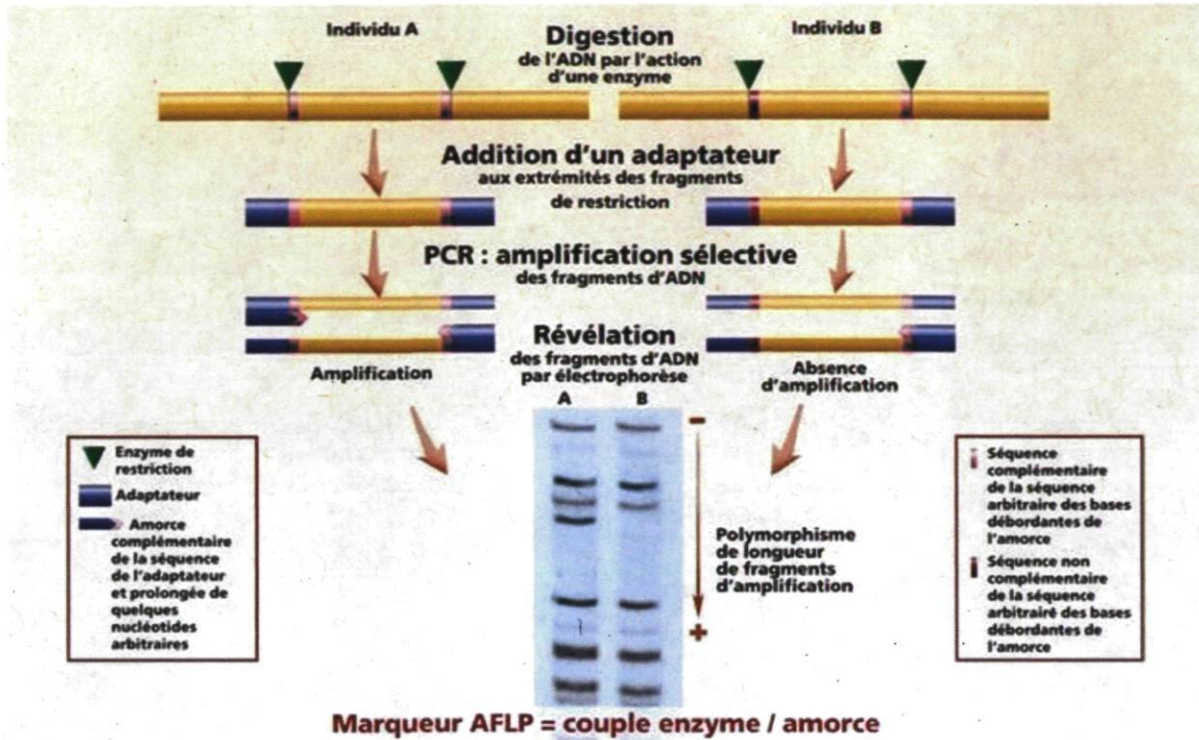
Développée par Vos et coll. (1995), cette technique utilise à la fois les enzymes de restriction et l'amplification PCR. Après la digestion de l'ADN génomique par deux enzymes de restriction, des adaptateurs de séquence connue et spécifiques des deux sites de restriction sont ajoutés aux extrémités des fragments de restriction (Figure 2-4). Une fois que les fragments de restriction sont ainsi bordés de séquences connues, il devient possible de les amplifier par PCR. Grâce à l'emploi d'amorces sélectives qui ne peuvent amplifier qu'une petite portion des milliers de fragments de restriction obtenus, seule une centaine de fragments sont amplifiés. Ces fragments sont séparés par électrophorèse sur gel de polyacrylamide dénaturant. Le polymorphisme révélé (de type présence ou absence de bandes) est provoqué par des changements dans les sites de restriction ou les régions bordantes. Plusieurs techniques de génotypage dérivent de la technique AFLP. Deux

d'entre elles, l'ADNc-AFLP et la SSAP («Sequence-Specific Amplification Polymorphism») détectent le polymorphisme présent au niveau des gènes.

Les marqueurs AFLP sont des marqueurs bialléliques dominants (Vos et coll. 1995) et génèrent un grand nombre de bandes polymorphes. Ils sont aussi extrêmement reproductibles et aucune information préalable de séquence ou génération de sonde ne sont nécessaires ce qui les rend très utiles pour créer rapidement des cartes génétiques. Toutefois, la dominance et les difficultés techniques liées au marquage par AFLP limitent son utilisation à grande échelle pour des applications comme la sélection assistée par marqueurs.

La technique AFLP est largement utilisée dans le domaine végétal en tant que technique d'empreintes génétiques hautement informatives et permet de comparer des génotypes génétiquement très proches. Elle connaît aussi une large application dans l'identification des cultivars et la détermination de leurs relations phylogénétiques, la cartographie des génomes et la saturation d'une région du génome au voisinage d'un gène en vue de son clonage.

Les AFLP ont été utilisés notamment dans des études de diversité chez de nombreuses espèces cultivées comme le maïs (Smith et coll. 1993) et le blé (Soleimani et coll. 2002). Chez l'orge, ces marqueurs ont été utilisés pour la cartographie génétique (Becker et coll. 1995; Hayes et coll. 1997; Powell et coll. 1997) ainsi que pour la saturation d'une région du génome proche d'un gène d'intérêt comme le gène *vrs1* (He et coll. 2004) ou le gène *Mlo* (Simons et coll. 1997).



(GNIS, 1999)

Figure 2-4 - Le polymorphisme révélé par les marqueurs AFLP.

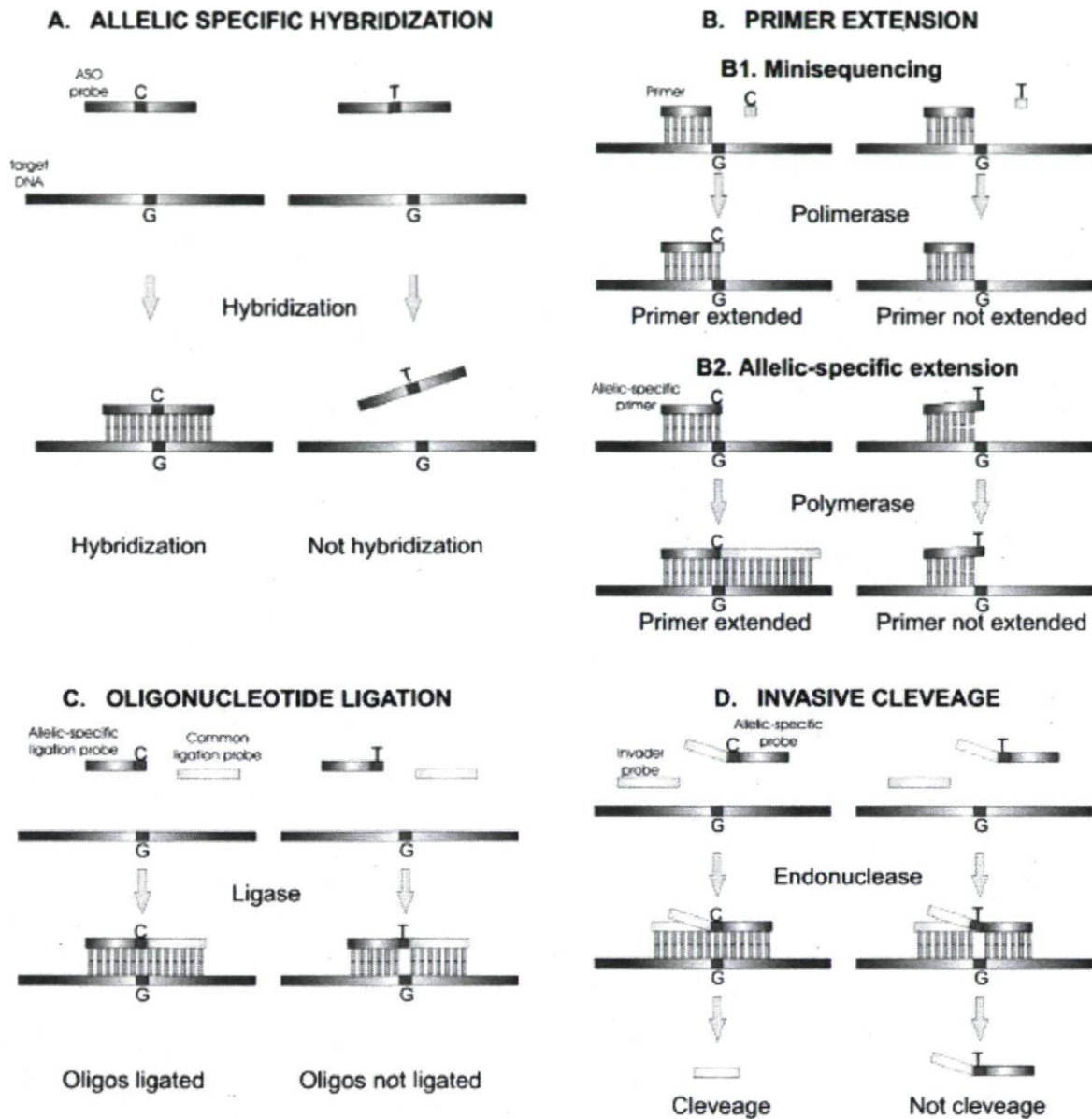
2.1.4.3 Les marqueurs SNP

Les polymorphismes mononucléotidiques, mieux connus sous leur acronyme anglais SNP (pour « Single Nucleotide Polymorphism » et prononcé « snip ») constituent la forme de variation la plus simple et la plus fréquente au sein des génomes. Ils ont été principalement développés et utilisés pour cartographier des maladies complexes chez les humains (Collins et coll. 1998). Les SNP sont une forme de polymorphisme engendré par une mutation ponctuelle qui crée différents allèles à un même locus. Ces mutations ponctuelles entraînent une différence dans une séquence d'ADN. Les SNP présentent de 2 à 4 allèles (pour un site donné), mais en général seulement deux bases alternatives sont observées pour un même site. Ce sont donc généralement des marqueurs bialléliques et leur fréquence peut être très élevée chez les plantes avec un SNP à tous les 100 ou même 50 bases chez le maïs (Rafalski 2002). Ce polymorphisme constitue une source de variabilité importante à exploiter.

Il existe actuellement de nombreuses techniques visant à mettre en évidence la présence de SNP dans l'ADN. Il n'existe pas de méthode parfaite de détection et de mise en évidence de SNP. Parmi la multitude de techniques disponibles, les techniques à haut débit dont les microarrays ou puces à ADN sont de plus en plus utilisés. Ceux-ci permettent un criblage d'un grand nombre de SNP en parallèle. Néanmoins cette technique est encore peu abordable. Selon Sobrino et coll. (2005), la majorité des analyses de génotypage des SNP peuvent être assignées à l'un de quatre groupes basés sur des méthodes moléculaires: l'hybridation allèle-spécifique, l'extension d'amorce, la ligation d'oligonucléotides et le clivage invasif (Figure 2-5). Il y a plusieurs méthodes de détection pour l'analyse des produits de chaque type de réaction (fluorescence, luminescence, mesure de la masse, etc...).

Cette technologie présente un potentiel d'automatisation très supérieur aux technologies précédentes (RFLP, RAPD, AFLP, SSR...). Ainsi, il est possible de génotyper des milliers jusqu'à plus d'un million de SNP par échantillon d'ADN dans le cas de l'humain (Jeffrey 2008). Par contre, le développement de telles collections de marqueurs nécessite un travail important de développement de marqueurs via la caractérisation de la séquence du génome dans plusieurs régions, voire la totalité, du génome.

Comparés aux autres marqueurs moléculaires, les SNP sont aujourd'hui des marqueurs largement utilisés dans des programmes de sélection de par leur abondance et leur répartition homogène au sein des génomes. Des marqueurs SNP ont été développés chez de nombreuses espèces d'intérêt agronomique. Leur utilisation relativement routinière, et la diversité des techniques de génotypage constituent les avantages principaux de ce type de marqueurs (Gupta et coll. 2001). Cette abondance permet la création de cartes génétiques de haute densité comprenant typiquement plus de mille marqueurs. À cause de leur grand nombre et de leur faible taux de mutations comparés aux microsatellites, les SNP peuvent servir de marqueurs pour des études d'association génétique, de déséquilibre de liaison, de sélection assistée par marqueurs, d'analyse de la diversité génétique, d'analyses phylogénétiques, ou pour l'identification des cultivars et la caractérisation des ressources génétiques (Rafalski 2002). Chez l'orge, ces marqueurs ont été utilisés pour étudier la diversité génétique (Bundock et Henry 2004; Chiapparino et coll. 2004; Varshney et coll. 2008). La cartographie par les SNP basée sur les gènes exprimés en réponse aux stress abiotiques a aussi été rapportée (Rostoks et coll. 2005).



(Sobrinho et coll. 2005)

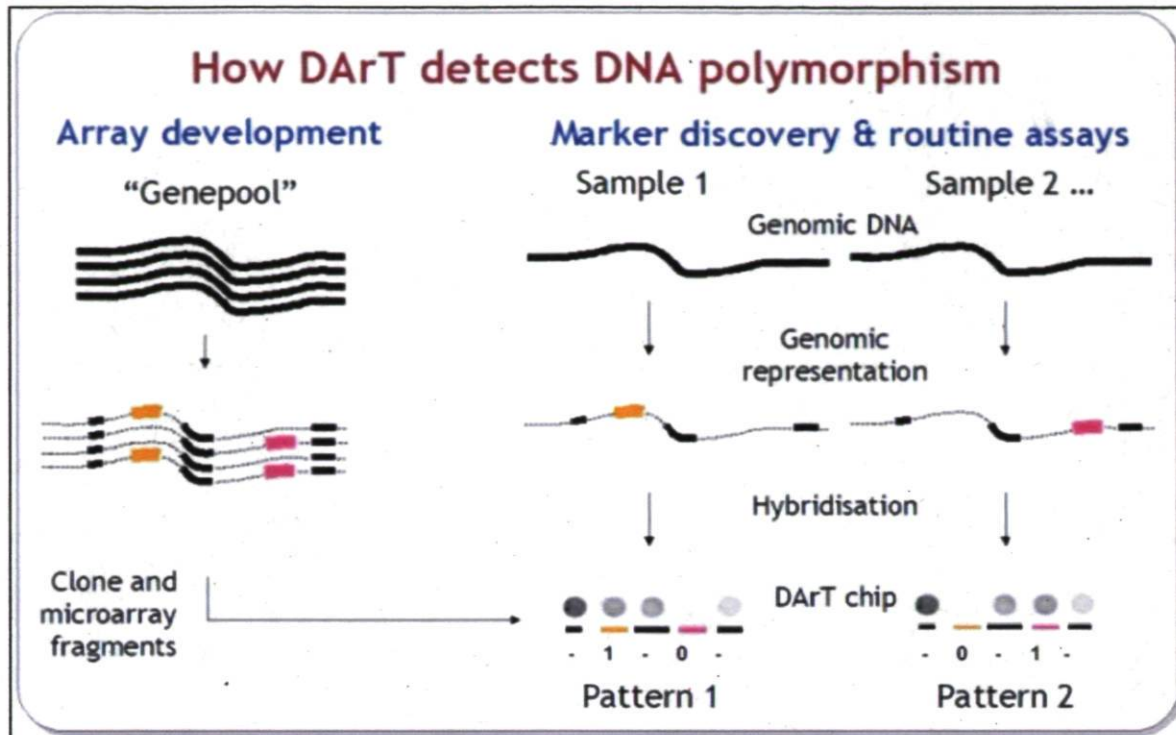
Figure 2-5 - Polymorphisme révélé par les marqueurs SNP (A) Hybridation d'oligonucléotides spécifiques d'allèles, (B) Réactions d'extension d'amorce, (C) Ligature d'oligonucléotides, (D) Clivage invasif.

2.1.2.6 Les marqueurs DArT

La méthode DArT, pour «Diversity Arrays Technology», a été décrite initialement par Jaccoud et coll. (2001). Tout comme pour les SNP, elle utilise des «puces à ADN» pour révéler le polymorphisme. Ces marqueurs sont, à plusieurs égards, analogues à des marqueurs AFLP puisqu'ils reposent également sur une amplification sélective d'un sous-ensemble de tous les fragments de restriction obtenus suite à la digestion de l'ADN génomique avec une paire d'enzymes de restriction. Ils diffèrent des AFLP en matière de méthode de révélation des polymorphismes. Dans le cas des AFLP, les amplicons polymorphes sont séparés sur gel de polyacrylamide et on peut y visualiser la présence ou l'absence d'un amplicon donné. Dans le cas des marqueurs DArT, une collection d'amplicons polymorphes (préalablement identifiés chez une collection de génotypes) ont été fixés sur une puce à ADN. En analyse DArT (Figure 2-6), suite à la digestion enzymatique et à l'amplification sélective pratiquées sur chaque échantillon d'ADN à l'étude, les amplicons sont marqués et hybridés à la puce. Si un individu A produit un amplicon particulier (couleur jaune), celui-ci reconnaîtra son complément sur la puce et il en résultera un signal positif (symbolisé par le «1»). Si l'amplicon est absent, aucun signal d'hybridation ne sera obtenu à cette position sur la puce (d'où un «0» dans ce cas).

Les marqueurs DArT sont des marqueurs à haut débit et la technique ne requiert pas d'information de séquence, mais les clones générés sont prêts à séquencer. Cette technique requiert une faible quantité d'ADN et elle est caractérisée par une excellente reproductibilité. Les seuls inconvénients liés à l'utilisation de ces marqueurs sont la dominance et la complexité technique.

Les marqueurs DArT sont présentement applicables sur plusieurs espèces pour la caractérisation rapide des ressources génétiques, la cartographie génétique, la sélection assistée par marqueurs et la diversité génétique. Chez l'orge, les marqueurs DArT ont été utilisés pour la réalisation de cartes génétiques de haute densité (Wenzl et coll. 2004; Wenzl et coll. 2006; Hearnden et coll. 2007) et la cartographie des QTL pour la fusariose de l'épi (Zhang et coll. En rédaction). Ces marqueurs permettent d'examiner la diversité génétique chez l'orge à une échelle et une vitesse inégalées.



(www.triticarte.com.au)

Figure 2-6 - Le polymorphisme révélé par les marqueurs DArT.

2.1.3 Comparaison des marqueurs moléculaires employés chez l'orge

Il est possible de comparer le polymorphisme obtenu avec les différentes techniques en utilisant des paramètres usuels de génétique des populations tels que le pourcentage de locus polymorphes, le nombre moyen d'allèles par locus et l'indice de diversité de Nei.

Une étude comparative entre quatre types de marqueurs a montré que les marqueurs SSR montrent le polymorphisme le plus élevé (100%), suivi des RFLP (83,2%), des RAPD (63,3%) et des AFLP (46,6%) si l'indice de bandes polymorphes est pris comme mesure (Russell et coll. 1997). Nous comparons, dans le Tableau 2-1, les différentes caractéristiques des six principales techniques de marquage moléculaire qui ont été utilisées pour identifier la diversité génétique chez l'orge. Les critères utilisés sont basés sur l'expérience et les résultats décrits dans la littérature.

Tableau 2-1 - Comparaison des six marqueurs moléculaires les plus utilisés chez l'orge

	RFLP	SSR	RAPD	AFLP	SNP	DArT
-Quantité d'ADN (µg)	10	0,05	0,02	0,5-1,0	0,05	0,05
-Qualité d'ADN	Haute	Moyenne	Haute	Moyenne	Haute	Moyenne
-PCR	Non	Oui	Oui	Oui	Oui	Oui
-Nombre de locus polymorphes analysés	1-3	1-3	1,5-50	20-200	1	>100
-Codominance	Oui	Oui	Non	Non	Oui	Non
-Facilité d'usage	Difficile	Facile	Facile	Facile	Facile	Facile
-Automatisation	Faible	Grande	Moyenne	Moyenne	Grande	Grande
-Reproductibilité	Grande	Élevée	Faible	Élevée	Élevée	Élevée
-Coût de développement	Faible	Élevé	Faible	Moyen	Élevé	Élevé
-Coût par analyse	Élevé	Faible	Faible	Moyen	Faible	Faible

2.1.4 Mesures de la diversité

Comme nous l'avons vu tout au long de la description des diverses techniques de marquage moléculaire, l'étude de la diversité génétique est une des applications les plus répandues pour les marqueurs moléculaires. Le choix du marqueur approprié dépendra en partie du matériel au sein duquel on cherche à mesurer la diversité génétique. Selon le but de l'étude, on peut être intéressé à mesurer la diversité génétique entre des espèces (niveau interspécifique) ou encore entre divers individus ou populations au sein d'une même espèce (niveau intraspécifique).

2.1.4.1 Diversité intraspécifique

La diversité intraspécifique est la diversité à l'intérieur d'une même espèce. Cette diversité est à l'origine de la création de nouvelles variétés. Plusieurs marqueurs moléculaires ont été utilisés pour l'évaluation de cette diversité au sein d'un matériel génétique et lorsqu'on opte pour un type

de marqueur moléculaire, il faudra s'assurer que celui-ci soit utile en raison du nombre de locus polymorphes disponibles ainsi que la diversité allélique présente à ces locus. L'indice de diversité génétique de Nei sert à décrire combien un marqueur moléculaire est informatif au sein d'un matériel génétique.

2.1.4.2 Diversité interspécifique

Pour décrire la diversité génétique des populations, Wright (1951) a développé l'indice F_{ST} qui est la corrélation entre deux gamètes tirés au hasard et provenant de deux sous-populations différentes. Cet indice s'applique aux locus bialléliques et en l'absence de sélection. Lorsque les locus sont multialléliques, on utilise une généralisation de cet indice, dénommé G_{ST} de Nei (1973).

L'analyse de la structuration hiérarchique des populations peut être abordée à l'aide de différents paramètres. Le principe général repose sur la décomposition de la variabilité observée entre les différents niveaux de regroupement des individus. L'analyse de variance sur les données moléculaires, AMOVA (Excoffier et coll. 1992), relève du même principe de décomposition de la diversité présente à chaque niveau hiérarchique ; elle facilite aussi l'identification des niveaux qui procurent le plus de variabilité. L'AMOVA présente l'intérêt aussi d'être basée sur une mesure de diversité intégrant des mesures de distance entre entités et de proposer des procédures d'inférence. Les tests de permutation de données permettent de s'affranchir de la normalité des distributions. Les analyses multivariées conduites sur les données alléliques ou génotypiques sont de bons outils pour représenter la variabilité observée. Elles permettent d'exploiter les génotypes individuels et donc d'évaluer la diversité intrapopulation et la différenciation interpopulation par des analyses hiérarchiques menés sur chaque axe.

Les analyses multivariées sont aussi de bons outils pour révéler la différenciation génétique entre sous-populations pour des caractères quantitatifs en tenant en compte des corrélations entre caractères.

2.1.4.3 Distance génétique

Le choix d'une mesure appropriée pour évaluer la diversité génétique et déterminer les relations entre les individus dans une population ainsi que l'utilisation d'une méthode statistique appropriée sont aussi importants que le choix du marqueur utilisé. Les mesures de similarité (S) et les distances génétiques (D) sont les plus généralement utilisées. Ces mesures servent à mesurer la ressemblance ou la dissemblance entre deux individus et on peut passer de l'une à l'autre par une simple transformation linéaire où $D = 1 - S$. De très nombreux indices de distance ont été identifiés sur les données binaires et la plupart varient de 0 à 1. Parmi les plus utilisés, nous mentionnerons le coefficient de Jaccard (Jaccard 1908), l'indice de Sokal et Michener (1958), la distance corrigée de Rogers (Wright 1978) et la distance de Nei et Li (1979). Ces indices sont choisis en fonction des caractéristiques des marqueurs utilisés. Les distances sont d'autant plus précises que le nombre de locus qui sert à les établir est élevé.

2.1.5 Représentation graphique

Pour une représentation graphique de la structure de la diversité génétique entre les lignées ou individus, la matrice de distances obtenue peut être représentée sous forme de dendrogramme. Les lignées sont agrégées de telle sorte que les plus proches génétiquement soient aussi les plus proches sur le dendrogramme. Diverses méthodes d'analyse de la structure sont disponibles mais la plus simple est celle qui utilise l'algorithme UPGMA («Unweighted Pair Group Method using Arithmetic Averages»), qui génère un seul arbre à partir d'une matrice des distances génétiques. Pour réduire l'information d'un ensemble de données multidimensionnelles à seulement deux ou trois axes, les techniques d'Analyse en Composantes Principales (PCA) et d'Analyse en Coordonnées Principales (PCoA) sont utilisées pour visualiser les relations entre les lignées. De nombreux programmes d'analyse phylogénétique peuvent être utilisés (PAUP, PHYLIP, Arlequin,...).

2.1.6 Le pedigree

Le pedigree constitue en quelque sorte l'arbre généalogique d'une variété ou d'un individu. La connaissance du pedigree des variétés fournit plusieurs types d'informations. Il peut être utilisé pour prévoir des potentiels de performance ou d'adaptation, pour calculer un coefficient de parenté entre deux lignées ou individus, pour guider le choix des parents en vue de croisements, ou encore pour suivre la transmission des variants d'un gène spécifique (Tinker et Deyl 2005). Beaucoup d'études de diversité génétique se sont appuyées sur l'étude du pedigree, la distance entre deux lignées étant calculée au moyen d'un coefficient de parenté. Ce coefficient de parenté a été défini par Cox et coll. (1985) comme la probabilité que deux allèles à un locus soient identiques par descendance. Cette approche exige de tracer l'histoire évolutive de chaque lignée, ce qui nécessite une grande quantité d'information, laquelle n'est pas toujours disponible.

Comment se comparent les mesures de similarité (ou « dissimilarité »/distance) obtenues avec les marqueurs moléculaires et l'étude du pedigree? Malheureusement, il n'y a pas de réponse simple ou unique à cette question, plusieurs situations ayant été rencontrées ce qui rend difficile la détermination de ce qui représente le mieux la « réalité ». La similarité, évaluée d'une part avec les marqueurs moléculaires et d'autre part par examen du pedigree, a été étudiée chez plusieurs céréales à l'aide de différents marqueurs. Dans l'ensemble, les corrélations entre les deux types d'informations sont faibles mais significatives. Chez l'orge, une corrélation de $r = 0,51$ a été obtenue entre les marqueurs RAPD et le pedigree (Tinker et coll. 1993). Pour les marqueurs RFLP, elle a été estimée à $r = 0,21$ pour l'orge d'hiver et à $r = 0,42$ pour l'orge de printemps (Graner et coll. 1994). Récemment, une faible corrélation $r = 0,24$ a été rapportée entre la distance génétique mesurée avec des SNP et la distance mesurée sur la base du pedigree (Soleimani et coll. 2007).

La faiblesse des corrélations observées pourrait s'expliquer par le fait que plusieurs facteurs liés au développement du cultivar tel que la sélection, la dérive génétique, ne sont pas pris en compte. Il est possible que cela découle aussi du faible nombre de marqueurs employés par ces auteurs et du nombre de générations examinées lors de l'étude du pedigree, ainsi que du calcul des coefficients d'apparentement qui repose sur des hypothèses irréalistes. Dans tous les cas, plus la concordance

sera élevée entre les diverses descriptions des relations qui existent entre un groupe de lignées, plus on aura confiance qu'il s'agit là d'une description juste de ces relations.

2.1.7 Hypothèses de recherches

En raison de leur développement récent, peu de travaux ont fait appel aux marqueurs DArT pour étudier la diversité génétique chez l'orge. Par contre, en raison de leurs avantages, il y a fort à parier qu'ils seront de plus en plus employés à de telles fins. Il nous a donc semblé important de déterminer dans quelle mesure ces nouveaux marqueurs fournissent une description de la distance génétique entre lignées d'orge qui est conforme avec celle obtenue soit à l'aide des marqueurs SSR ou de l'analyse du pedigree, compte tenu qu'il s'agit là des deux approches les plus fréquemment employées chez l'orge. C'est donc dans ce contexte que les deux hypothèses suivantes ont été formulées :

Hypothèse 1:

Les marqueurs SSR, DArT et le pedigree permettent tous les trois de mesurer la diversité génétique et de déterminer la structure de la population au sein d'une collection représentative des orges cultivées au Canada.

Hypothèse 2:

Les marqueurs DArT constituent une approche plus rapide et économique pour mesurer la diversité génétique que les SSR ou le pedigree.

3.4.8 Objectifs de recherche

Volet 1:

Mesurer la diversité génétique et déterminer la structure de la population au sein d'une collection de 92 lignées canadiennes d'orge à l'aide de :

- marqueurs SSR

- marqueurs DArT
- l'analyse du pedigree

Volet 2:

Comparer les trois approches en matière de :

- Concordance
- Efficacité (temps requis, coût)

Le but est d'analyser l'intérêt de ces types de marqueurs (SSR et DArT) et du pedigree pour des études de diversité en comparant la façon dont ils permettent individuellement de structurer la diversité parmi une collection d'orges canadiennes.

CHAPITRE III: MANUSCRIT

Comparative analysis of genetic diversity in barley assessed by SSR, DArT and pedigree data

MANUSCRIT

Ce manuscrit a été rédigé dans le cadre d'un projet de maîtrise en biologie végétale. Il sera soumis sous peu à une revue avec comité de lecture.

Les auteurs :

- Mebarek Lamara a été un artisan central et pris part à toutes les étapes de cette recherche (choix des marqueurs, réalisation des expériences, analyse des résultats et rédaction du manuscrit).
- Li Yi Zhang a contribué à l'analyse des résultats.
- Suzanne Marchand a contribué à l'analyse des résultats.
- Nicholas Tinker a contribué à l'analyse des résultats.
- François Belzile a contribué au choix des marqueurs, à l'analyse des résultats et à la rédaction du manuscrit.

3.1 Résumé

La domestication et, plus récemment, la sélection moderne des plantes ont contribué à une réduction importante de la diversité génétique chez plusieurs plantes cultivées, une homogénéité accrue qui peut être nuisible à la réponse des plantes aux défis posés par les pathogènes et le changement rapide du climat. Il est ainsi important de mesurer, maximiser et exploiter autant que possible la diversité génétique dans des programmes de sélection des plantes. L'objectif de cette étude était de mesurer la diversité génétique et la structure des populations parmi 92 cultivars canadiens d'orge en utilisant deux types de marqueurs moléculaires (SSR et DArT) et le pedigree. Un total de 368 allèles ont été identifiés à 50 locus SSR. Le nombre d'allèles par locus variait de 2 à 13 ($\bar{x} = 7,36$) et l'indice PIC variait entre 0,34 et 0,86 ($\bar{x} = 0,69$). Un groupement des génotypes a été effectué basé sur une matrice de distance génétique et le dendrogramme obtenu a montré les relations génétiques entre les cultivars d'orge. Pour les marqueurs bialléliques DArT, la matrice de distance génétique s'est basée sur 971 marqueurs dont les indices PIC variaient entre 0,06 et 0,50 ($\bar{x} = 0,39$). De même, une matrice de distance génétique a été employée pour produire un dendrogramme DArT. Une troisième matrice de distance a été calculée basée sur le coefficient de parenté et employée pour produire un troisième dendrogramme. La similarité topologique des trois dendrogrammes a été estimée en utilisant un indice de congruence et a montré que les trois dendrogrammes présentent une très bonne concordance. Les analyses statistiques ont également montré une corrélation très significative entre les matrices SSR et DArT ($r = 0,80$, $p < 0,002$) mais une faible corrélation positive entre la matrice basée sur le pedigree et les matrices basées sur les deux types de marqueurs ($r = 0,46$, $p < 0,002$; $r = 0,52$, $p < 0,002$). L'information obtenue en comparant les résultats de différentes méthodes d'estimation de la diversité génétique sera utile pour l'amélioration et la conservation des ressources génétiques chez l'orge.

3.2 Abstract

Domestication and, more recently, modern plant breeding have contributed to important reductions in genetic diversity in many crops and this increased homogeneity may be detrimental in responding to challenges posed by both pathogens and rapidly changing climate. It is thus important to measure, maximize and exploit as well as possible the genetic diversity in plant breeding programs. The aim of this study was to measure genetic diversity and population structure among 92 Canadian barley cultivars using two types of molecular markers (SSRs and DArTs) and pedigree data. A total of 368 alleles were identified at 50 SSR loci. The number of alleles per locus ranged between 2 and 13, ($\bar{x} = 7.36$) and PIC values ranged from 0.34 to 0.86 ($\bar{x} = 0.69$). Clustering of genotypes was performed based on the genetic distance matrix and the dendrogram obtained showed the genetic relationships among barley cultivars. For the biallelic DArT markers, the genetic distance matrix was based on 971 markers whose PIC values ranged between 0.06 and 0.50 ($\bar{x} = 0.39$). Similarly, a genetic distance matrix was used to produce a DArT dendrogram. A third distance matrix was computed based on the kinship coefficient and was used to produce a third dendrogram. The topological similarity of the three dendrograms was estimated using a congruence index and showed the three dendrograms to be in very good agreement. Statistical analysis also showed a highly significant correlation between the SSR and DArT matrices ($r = 0.80$, $p < 0.002$) compared to lower yet significant correlations of the pedigree data with both marker types ($r = 0.46$, $p < 0.002$; $r = 0.52$, $p < 0.002$). Information obtained from comparing results of different genetic diversity estimation methods should be for the improvement and conservation of barley genetic resources.

Key words: Barley, Genetic distance, Kinship coefficient, Congruence index, SSR, DArT

3.3 Introduction

Barley (*Hordeum vulgare* L.) is one of the oldest cultivated crops in the world (Morrell and Clegg 2007). Today, it constitutes the fourth most cultivated cereal in the world after maize, rice and wheat; (FAO, <http://www.fao.org/>, 2008). In this crop as in others, genetic diversity has declined following domestication and intense selection in modern plant breeding programs: it is highest in wild barley, *H. vulgare* subsp. *spontaneum*, intermediate in landraces or parents of mapping populations, and lowest in elite cultivars (Manninen and Nissila 1997, Matus and Hayes 2002). Recently, it was estimated that nucleotide diversity in elite European germplasm was less than half that found in wild barley lines (Russell et al. 2004). As a consequence, when these genetically homogeneous or closely-related lines are grown over vast areas, this increases the risk of large-scale epidemics or crop failure in adverse environmental conditions (Asins and Carbonell 1989; Martin et al. 1991). It is therefore important to accurately measure, maximize and exploit the genetic diversity both in the conservation of plant genetic resources and in breeding work.

Genetic diversity can be estimated in various ways including the examination of pedigrees and the use of molecular markers (Soleimani et al. 2007; Mikel and Kolb, 2008). Pedigree analysis has been used to assess the genetic diversity in North American barley cultivars (Martin et al. 1991; Mikel and Kolb 2008). Unfortunately, pedigree information can sometimes be erroneous or incomplete and it does not necessarily reflect the underlying genetics. Two lines selected from the same pair of parents could conceivably be quite different or quite similar depending on the alleles that were inherited and selected for.

In contrast, molecular markers allow one to assess relationships directly at the DNA level. The extent of their informativeness may depend on the nature of the marker, their number, genome coverage, and the population under investigation (Soleimani et al. 2002, 2007). Several different molecular markers can be used in genetic analysis. Their great number, polymorphism and ease of genotyping make microsatellites (or SSR for “simple sequence repeats”) excellent markers for the analysis of genetic diversity in barley (Matus and Hayes 2002). Recently, however, high throughput genotyping platforms, such as single nucleotide polymorphisms (SNPs; Rostoks et al.

2005) and diversity array technology (DArT; Wenzl et al. 2004) markers have been developed in barley.

In addition to providing information on the amount of genetic diversity in various gene pools, pedigree and marker information are crucial to properly describing population structure within sets of germplasm. The accurate description of population structure has become extremely important with the advent of association genetics (Nordborg and Weigel 2008). Indeed, when insufficiently accounted for, population structure can lead to the spurious detection of associations between markers and phenotypes of interest (Thornsberry et al. 2001).

In this context, we felt it would be of interest to measure genetic diversity in a collection of Canadian barley lines using two types of molecular markers (SSR and DArT), and to compare these results to those obtained using pedigree data. The information obtained on this collection of lines, chosen to be representative of the germplasm in use in barley breeding programs, provide a snapshot of the extant genetic diversity and show that all three types of data used to assess genetic diversity and population structure provide a highly consistent portrait of population structure within this germplasm.

3.4 Materials and methods

3.4.1 Plant materials

Ninety-two (92) Canadian barley varieties were used in this study and are listed in Table 3-1. These were selected to be representative of the barley germplasm developed in Canadian breeding programs in the last 20 years, both in eastern (37 varieties) and western (55 varieties) Canada. A first subset included sixty-four (64) lines that were extensively characterized in the early 1990s (Kong et al. 1994; Kong et al. 1995; Narasimhalu et al. 1998). A second subset comprised 28 lines that were developed more recently. In total, the collection included 41 2-row and 51 6-row varieties.

3.4.2 DNA extraction

Four seeds of each line were germinated on moist filter paper in a Petri dish and, after four days, young leaves (5 cm) of each barley genotype were collected and placed in a single well of a 96-well extraction plate containing two metallic beads. Leaf tissue was frozen by setting the plate in 1 cm of liquid nitrogen during 15 min. Leaves were ground to a fine powder using a mixer mill (Tissue Lyser; Qiagen, Hilden, Germany) set to 27 oscillations/sec (4 to 5 periods of 30 seconds each). The plate was centrifuged at 4,000 rpm (2,600 xg) during 15 sec. DNA was then extracted essentially according to the protocol suggested by Triticarte (http://triticarte.com.au/pdf/DArT_DNA_isolation.pdf) (Annexe I). Concentration of the DNA was estimated and adjusted to 50 ng/ μ l.

3.4.3 SSR and DArT analyses

Analysis with the SSR markers was carried out on all 92 lines using a set of 50 microsatellites providing a good coverage of the barley genome (Table 1S). PCR amplifications were carried out in a final volume of 10 μ l containing: 1 μ l of DNA, 1.5 μ l of 10x PCR buffer, 1 μ l of dNTPs (2 mM), 0.5 μ l (2.5 μ M) of the forward (M13-tailed) and reverse primers, 0.2 μ l (2 pM) of universal M13 primer labeled with IRD700 and 0.15 μ l of Taq DNA polymerase. DNA amplification was performed in a thermocycler (T Gradient, Biometra, Göttingen, Germany) programmed as follows: 1 cycle of 94°C for 3 min, 58°C for 1 min, and 72 °C for 1 min; 30 cycles of 94°C for 45 sec, 55°C for 45 sec, and 72°C for 45 sec.; 1 cycle of 72°C for 5 min. After PCR amplification, 2 μ l of formamide electrophoresis loading buffer were added to each well. Then, 0.8 μ l of PCR product was denatured and separated by electrophoresis in a 6.5% polyacrylamide gel. The bands were visualized on an automatic DNA analyzer (LI-COR Global Edition IR² DNA Analyzer) set to 1,500 V (temperature = 50°C) during 1:45 to 2:30 (according to the size of PCR products). As for the DArT markers, DNA was sent to Triticarte Ltd (Yarralumia, Australia) for genotyping, using Barley PstI(BstNI) version 2.0.

Matrices of binary data obtained for SSR and DArT markers were used to generate Nei and Li (1979) genetic distance matrices. Clustering of genotypes was performed using the unweighted

pair group method with arithmetic average (UPGMA) (Sneath and Sokal 1973) based on a genetic distance matrix with PAUP (version 4.0 b10). Principal Coordinate Analysis (PCoA) was performed with the PAST software (version 1.30) based on the Dice similarity index (Dice 1945). The polymorphism information content (PIC) indice was calculated for each marker using a 0/1 data matrix, using the following formula: $PIC = 1 - \sum P_i^2$, where P_i is the frequency of the i th SSR allele (Smith et al. 2000). The average number of alleles per locus and the effective number of alleles per locus were calculated as reported by Morgante et al. (1994). The total number of effective alleles and the efficiency index were calculated according to Pejic et al. (1998).

3.4.4 Pedigree analysis

The pedigree of the lines was obtained from several sources in the literature, a barley pedigree database (Baum et al. 1985), descriptions of the cultivars and personal communications with breeders. The pedigree of all cultivars was obtained and the ancestry of each cultivar was traced back as far as possible. The coefficient of relationship (“Kinship coefficient”) (r) between the lines was calculated using Winkin2[©] according to Tinker and Mather (1993). A kinship coefficient matrix was obtained and transformed into a distance matrix (using $d = 1 - r$), which was in turn used to perform a cluster analysis as described above.

3.4.5 Comparisons of tree topology and matrices

The topological similarity of the three dendrograms was estimated using a congruence index (as per de Vienne et al. 2007) and a p -value was thus computed. Genetic distance matrices of SSR, DArT and pedigree were compared by the MAXCOMP routine of NTSYS-pc (Rohlf 2000). The normalized Mantel Z statistic (Mantel 1967) was used to determine the level of significance of the observed correlation between the matrices.

3.5 Results

3.5.1 Genetic diversity assessed with SSR and DArT markers

To assess the genetic diversity in Canadian barley, a set of 92 cultivars (see Table 3-1 for a list and description) was genotyped using SSR and DArT markers. In total, 50 SSR loci broadly distributed over the seven linkage groups were used. The chromosomal position, number of alleles detected and the repeat motif for each SSR are detailed in Table 1S. A total of 368 alleles were detected and the number of alleles per locus varied from 2 (HVM67, WMC1E8) to 13 (Bmag0013, EBmac0775), for an average of 7.4 per locus and an average number of effective alleles of 3.77 per locus (Table 3-2). Polymorphism information content (PIC) values were typically quite high and ranged between 0.34 (Ebmac273b) and 0.86 (EBmac0775), with a mean of 0.69. Of 2,304 probes on the barley DArT chip, a total of 971 proved polymorphic and yielded a total of 1,942 alleles (2 alleles per locus), with an average number of effective alleles of 1.68 per locus. The PIC values ranged from 0.06 to 0.50 (theoretical maximum for a biallelic marker), for an average of 0.39. The number of informative DArT markers per linkage group varied between 52 (4H) and 160 (2H) and thus provided extensive coverage of the barley genome. The levels of polymorphism detected with each marker system are summarized and compared in Table 3-2.

Table 3-2 - Summary data on the polymorphism and efficiency of SSR and DArT markers in a collection of 92 Canadian barley lines.

Parameters	Marker system	
	SSR	DArT
Number of assay units	50	1
Number of polymorphic bands	368	971
Number of loci	50	2,304
Average number of alleles per locus	7.4	2
PIC average	0.6	0.3
Effective number of alleles per locus	3.7	1.7
Essay efficiency index	3.7	1,634.5

3.5.2 Genetic distances

Pair-wise genetic distances were computed using the molecular markers (SSR/DArT) or pedigree data (see Materials and Methods). In the latter case, a matrix of genetic distances was derived from kinship coefficients established on the basis of the pedigree for each line. Genetic distance values ranged from 0.03-0.36 for SSR, 0.002-0.09 for DArT and 0.29-1 for pedigree, with an average of 0.15, 0.04 and 0.88 for SSR, DArT and pedigree, respectively. The lowest genetic distance (0.002) was observed between 'Galt' and 'Melvin' revealed by DArT markers. The highest genetic distances (1) were observed in the pedigree matrix between different pair of lines with no shared ancestors, and they skewed the distribution towards highest values.

To explore the similarity of the relationships revealed using the three types of data, the similarity between distance matrices based on SSR, DArT and the kinship coefficient was measured by pair-wise comparison among the three genetic distance matrices using a normalized Mantel statistic. The highest correlation ($r = 0.80$, $p < 0.002$) was observed between the SSR and DArT matrices, indicating that both sets of markers yielded highly similar estimates of genetic distances between lines. Lower but significant positive correlations were obtained between the DArT and pedigree matrices ($r = 0.52$, $p < 0.002$) and between the SSR and pedigree matrices ($r = 0.46$, $p < 0.002$) (Figure 3-1). Of the three methods employed, analysis of the pedigree is that which offered the least resolution.

3.5.3 Genetic relationships between cultivars

The relationships between lines were first established by performing a principal coordinate analysis (PCoA) using both the DArT and SSR data (Figure 3-2). Three main conclusions can be drawn from this figure. Firstly, the collection of barley lines fall into two main groups clearly reflecting row-type (separated along the x-axis). Only two genotypes, 'Joly' and 'Niska', both 6-row types, landed in between these two clusters. Secondly, the 2-row group is much more diverse (broadly spread along the y-axis) than the 6-row type, with the genotypes from Eastern Canada located above the axis and those from Western Canada below the axis. Finally, the results obtained with both marker technologies were in excellent agreement.

To further refine the relationships between the lines, we conducted an UPGMA cluster analysis based on the genetic distance matrices. A dendrogram was obtained for each type of data and the general topology of these trees is compared in Figure 3-3. As can be seen in all three trees, the lines were clearly clustered into two large groups based on the type of spike: 2-row (blue) or 6-row (red). Geography constitutes an additional factor governing the grouping of lines in these trees. The lines from western Canada (dark blue or red) mostly clustered together as did lines from eastern Canada (light blue or red). A detailed view of each individual tree can be seen in the supplemental figures.

A comparison of the topological similarity of the three trees using the index of congruence as per de Vienne et al. (2007) showed that the greatest congruence was observed between the SSR and DArT trees ($I_{\text{cong}} = 2.62$ and $p = 2.60 \times 10^{-16}$). Congruence was lower, but still highly significant between the DArT- and pedigree-based trees ($I_{\text{cong}} = 1.98$ and $p = 9.27 \times 10^{-10}$) and between the SSR- and pedigree-based trees ($I_{\text{cong}} = 1.84$ and $p = 1.19 \times 10^{-8}$). In all three comparisons of topology, low p values indicate that the trees are significantly more congruent than expected by chance alone.

3.6 Discussion

Knowledge of existing genetic diversity allows for more efficient and effective use of resources in plant breeding programs. Given the proliferation of molecular markers used for assessing the genetic diversity in barley and their advantages over traditional morphological characters, comparisons between techniques can be helpful in comparing methods for assessing genetic diversity. Given the relative novelty of DArT markers, we were interested in comparing them with SSRs, the most frequently used type of molecular marker in barley, as well as with pedigree data.

We first compared how informative the two molecular marker technologies had proven. On an individual marker basis, SSRs proved to be much more informative than DArT markers as the average number of alleles per SSR locus was 7.4 compared to 2 for each DArT marker. Accordingly, average PIC values were almost twice as high for SSRs relative to DArTs (0.69 and 0.39, respectively). These results are in very good agreement with previously reported data for SSR and DArT markers in populations of comparable size and composition. For example, Matus

and Hayes (2002) reported 5.8 alleles/SSR locus and a mean PIC value of 0.45 in a set of 96 elite breeding lines of barley from a single barley breeding program. More recently, Varshney et al. (2008) reported a mean PIC value of 0.58 using a set of 28 SSRs on barley lines from six different countries. As for DArT markers, to the best of our knowledge, the only published data are those of Wenzl et al. (2004) who reported a mean PIC value of 0.38, a result almost identical to the one reported here.

Despite the fact that more information is provided by a single SSR marker than a single DArT marker, a more adequate appreciation of the efficiency of the two marker systems must, however, take into account the number of markers contributing information in a given analysis. The highest assay efficiency index value was observed with DArT markers (1,634.5) compared to the SSR markers (3.7). In a similar analysis carried out in maize inbred lines using RFLP, RAPD, SSR and AFLP, Pejic et al. (1998) found that the efficiency index calculated for AFLPs (61.9) was more than ten-fold higher than for SSRs (4.4) and the other methods. This high assay efficiency index results from the ability of the DArT platform to simultaneously screen over 2,000 potentially polymorphic loci in barley in a single assay.

The average genetic distance between pairs of barley cultivars was found to differ markedly. With DArT markers, a very low average genetic distance (0.04; range: 0.002-0.09) was obtained, whereas pedigree information yielded an average distance of 0.88 (range: 0.29-1). SSRs produced a range of pairwise distances between 0.03 and 0.36 for a mean of 0.15. The very low distances seen between lines using DArT markers likely reflects the limited number of alleles (2) that can be observed for a given marker. Thus, for any two lines, a large proportion of markers will happen to carry the same allele. In contrast, using SSRs, the greater number of alleles will increase the probability that two lines differ for any given marker. At the other extreme, pedigree information leads to a highly biased distribution of genetic distances (illustrated in Figure 3-1), with a large number of pairs showing a distance of 1 indicating that no common ancestors were found in the pedigree.

This bias in the distribution of distances based on pedigree information also impacts heavily the correlation between distances. As was noted in Figure 3-1, pedigree vs molecular marker, correlations were the weakest (0.46 and 0.52), whereas correlations between distances based on molecular data were much higher (0.8). These correlations between distance matrices are very

similar to the correlations reported in spring wheat by Almanza-Pinzón et al. (2003) between a similarity matrix derived from pedigree information and matrices derived from either SSR (0.44) or AFLP markers (0.54). In many previous studies, even weaker correlations (0.1-0.25) have been reported between measures of genetic distance or similarity based on pedigrees and molecular markers such as SNPs (Soleimani et al. 2007), RAPDs (Manninen and Nissila 1997), RFLPs (Graner et al. 1994; Dávila et al. 1998) or AFLPs (Ellis et al. 1997; Schut 1997). Moderate correlations, similar to the ones we found, were observed by Tinker et al. (1993) using RAPDs ($r = 0.51$) and Casas et al. (1998) using RFLP markers ($r = 0.54-0.67$). Various assumptions that are made in the calculation of genetic relatedness based on the pedigree may introduce inaccuracies (Souza and Sorrells, 1989; Cox and Murphy, 1990). Among these, we can mention 1) equal parental contribution to progeny, 2) the absence of selection pressure or genetic drift and 3) the absence of relatedness of parents with unknown pedigree. Moreover, pedigree information itself is prone to errors as cultivars may be known by more than one name (synonyms) or one name may have been used for more than one cultivar or genotype (homonyms, e.g., landraces). In addition, pedigrees may be wrongly recorded or uninformative, especially if encoded lines are included (Ellis et al. 1997).

Our results also demonstrate that dendrograms obtained with the two types of marker data are highly congruent as measured using the congruency index developed by de Vienne et al. (2007). The topology of the pedigree-based tree was also found to exhibit a highly significant, albeit lower, level of congruence to both trees based on molecular data.

The dendrograms reveal that this collection of Canadian barley cultivars and breeding lines is clearly divided into two major groups reflecting the architecture of the spike. With very few exceptions, 2- and 6-row cultivars fell neatly into two different branches of the trees. Such a clear division has also been documented in a collection of American barley cultivars as reported by Hayes and Szücs (2006) using SNP markers and Zhang et al. (2009) using DArT markers. This is, however, quite different from the situation reported by Rostoks et al. (2005) among European barley cultivars, where growth habit (winter vs spring) was found to be the primary determinant of population structure. This suggests that breeders tend to make crosses predominantly within one or the other pool and this marked population structure will need to be taken into account carefully

when planning association analyses in such germplasm as previously discussed by Zhang et al. (2009).

In conclusion, this study demonstrates the usefulness of DArT markers for the study of genetic diversity in a collection of barley cultivars. In addition to rapidly providing information on a large number of loci, and thus conferring excellent genome coverage, DArT markers have been reported to be 10-fold less expensive per data point than SSRs (Xia et al. 2005). Our results also suggest that DArT markers provide an accurate picture of population structure within barley germplasm collections, information that is of critical importance in devising well-designed association analyses.

Acknowledgements

The financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and of La Coop Fédérée are acknowledged. The authors wish to thank N. Tinker (AAFC-Ottawa) for providing the Winkin2[®] software. We would like also to thank Dr. Bill Legge (AAFC-Brandon) and Dr. Brian Rosnagel (Crop Development Centre, University of Saskatchewan) for providing pedigree data.

References

- Almanza-Pinzón M. I., Khairallah M., Fox P. N. and Warburton M. L. 2003. Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. *Euphytica* 130: 77–86.
- Asins M.J. and Carbonell E.A. 1989. Distribution of genetic variability in a durum wheat world collection. *Theor. Appl. Genet.* 77: 287–294.
- Baum B.R., Bailey L.G. and Thompson B.K. 1985. Barley register. Agriculture Canada Publ. 1783B, Ottawa, Canada.
- Casas A.M., Igartua E. and Molina-Cano J.L. 1998. Genetic diversity of barley cultivars grown in Spain, estimated by RFLP, similarity and coancestry coefficients. *Plant Breed.* 117: 429–435.
- Cox T.S., and Murphy J.P. 1990. The effect of parental divergence on F2 heterosis in winter wheat crosses. *Theor. Appl. Genet.* 79: 241–250.
- Dávila J.A., Sanchez de la Hoz M.P., Loarce Y. and Ferrer E. 1998. The use of random amplified microsatellite polymorphic DNA and coefficients of parentage to determine genetic relationships in barley. *Genome* 41: 477–486.
- de Vienne D.M., Giraud T. and Martin O.C. 2007. A congruence index for testing topological similarity between trees. *Bioinformatics* 23: 3119–3124.
- Dice L.R. 1945. Measures of the amount of ecological association between species. *Ecology*.26: 297–302.
- Ellis R.P., McNicol J.W., Baird E., Booth A., Lawrence P., Thomas B. and Powell W. 1997. The use of AFLPs to examine genetic relatedness in barley. *Mol. Breed.*3: 359–369.
- FAOSTATdata. 2008. <http://faostat.fao.org/site/291/default.aspx>.
- Graner A., Ludwig W.F. and Melchinger A.E. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. *Crop Sci.* 34: 1199–1205.
- Hayes P. and Szücs P. 2006. Disequilibrium and association in barley: Thinking outside the glass. *Proc. Natl. Acad. Sci. USA.* 103: 18385–18386.
- Kong D., Choo T.M., Jui P., Ferguson T., Therrien M.C., Ho K. M., May K.W. and Narasimhalu P. 1994. Genetic variation and adaptation of 76 Canadian barley cultivars. *Can. J. Plant Sci.* 74: 737–744.
- Kong D., Choo T.M., Jui P., Ferguson T., Therrien M.C., Ho K. M., May K.W. and Narasimhalu P. 1995. Variation in starch, protein, and fibre of Canadian barley cultivars. *Can. J. Plant Sci.* 75: 865–870.

- Manninen O. and Nissila E. 1997. Genetic diversity among Finnish six-rowed barley cultivars based on pedigree information and DNA markers. *Hereditas* 126: 87–93.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209–220.
- Martin J.M., Blake T.K. and Hockett E.A. 1991. Diversity among North American spring barley cultivars based on coefficient of parentage. *Crop Sci.* 31: 1131–1137.
- Matus I.A. and Hayes P.M. 2002. Genetic diversity in three groups of barley germplasm assessed by simple sequence repeat. *Genome* 45: 1095–1106.
- Mikel M.A. and Kolb F. L. 2008. Genetic diversity of Contemporary North American Barley. *Crop Sci.* 48: 1399–1407.
- Morgante M., Rafalsky J.A., Biddle P., Tingey S. and Olivieri A. M. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome* 37: 763–769.
- Morrell P. and Clegg M.T. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. USA.* 104: 3289–3294.
- Narasimhalu P., Kong D. and Choo T.M. 1998. Straw yields and nutrients of seventy-five Canadian barley cultivars. *Can. J. Animal Sci.* 78: 127–134.
- Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* 76: 5269–5273.
- Nordborg M. and Weigel D. 2008. Next-generation genetics in plants. *Nature* 456: 720–723.
- Pejic I., Ajmone-Marsan P., Morgante M., Kozumplik M., Castiglioni P., Taramino G. and Motto M. 1998. Comparative analysis of genetic similarity among maize inbredlines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theor. Appl. Genet.* 97: 1248–1255.
- Rohlf F.J. 2000. NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System, version 2.1. Exeter Software Publisher, Setauket, NY.
- Rostoks N., Mudie S., Cardle L., Russell J., Ramsay L., Booth A., Svensson J.T., Wanamaker S.I., Walia H., Rodriguez E.M., Hedley P.E., Liu H., Morris J., Close T.J., Marshall D.F. and Waugh R. 2005. Genome-wide SNPs discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genomics* 274: 515–527.
- Russell J., Booth A., Fuller J., Harrower B., Hedley P., Machray G. and Powell W. 2004. A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome* 47: 389–398.
- Schut J.W. 1997. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. *Theor. Appl. Genet.* 95: 1161–1168.

- Smith J.S.C., Kresovich S., Hopkins M.S., Mitchell S.E., Dean R.E., Woodman W.L., Lee M. and Porter K. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Sci.* 40: 226–232.
- Sneath P.H.A. and Sokal R.R. 1973. Numerical taxonomy: the principles and practice of numerical classification. (W. H. Freeman: San Francisco).
- Soleimani V.D., Baum B.R. and Johnson D.A. 2007. Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data. *Genet. Res. Crop Evol.* 54: 83–97.
- Soleimani V.D., Baum B.R. and Johnson D.A. 2005. Genetic diversity among barley cultivars assessed by sequence-specific amplification polymorphism. *Theor. Appl. Genet.* 110: 1290–1300.
- Soleimani V.D., Baum B.R. and Johnson D.A. 2002. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.]. *Theor. Appl. Genet.* 104:350–357.
- Souza E. and Sorrells M.E. 1989. Pedigree analysis of North American oat cultivars released from 1951 to 1985. *Crop Sci.* 29:595–601.
- Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D. and Buckler E.S. 4th. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics* 28: 286–289.
- Tinker N.A. and Mather D.E. 1993. KIN: Software for computing kinship coefficients. *J. Hered.* 84: 238.
- Tinker N.A., Fortin M.G. and Mather D.E. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theor. Appl. Genet.* 85: 976–984.
- Varshney R.K., Thiel T., Sretenovic-Rajcic T., Baum M., Valkoun J., Guo P., Grando S., Ceccarelli S. and Graner A. 2008. Identification and validation of a score set of informative genic SSR and SNP markers for assaying functional diversity in barley. *Mol. Breeding* 22: 1–13.
- Wenzl P., Carling J., Kudrna D., Jaccoud D., Huttner E., Kleinhofs A. and Kilian A. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. USA.* 101: 9915–9920.
- Xia L., Peng K., Yang S., Wenzl P., De Vicente M.C., Fregene M. and Killian A. 2005. DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theor. Appl. Genet.* 110: 1092–1098.
- Zhang L. Y., Marchand S., Tinker N. A. and Belzile F. 2009. Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theor. Appl. Genet.* 119: 43–52.

Table 3-1 – List of Canadian barley cultivars used in this work. Row number (R), geographic region of origin: (W for Western Canada, E for Eastern Canada) are indicated.

Cultivar	R	Origin	Cultivar	R	Origin	Cultivar	R	Origin
AB241-1	2	W	CFO284AA6	2	E	Léger	6	E
Abee	2	W	Chambly	6	E	Loyola	6	E
AC Kings	2	E	Chapais	6	E	Manley	2	W
AC Klinck	6	E	Conquest	6	W	Maskot	6	E
AC Metcalfe	2	W	Craig	2	E	Massey	6	E
AC Ranger	6	W	Creme	6	W	Melvin	6	W
AC Rosser	6	W	Cyane	6	E	Micmac	2	E
AC Westech	6	E	Deuce	2	W	Mingo	6	E
Albany	2	E	Diamond	6	W	Morrison	2	E
Altona	6	E	Duel	6	W	Newdale	2	W
Argyle	6	W	Duke	6	W	Niska	6	W
B1215	2	W	Ellice	2	W	Noble	6	W
B1602	6	W	Elrose	2	W	Norbert	2	W
Bedford	6	W	Empress	6	W	OAC Kawartha	6	E
Betzes	2	W	Étienne	6	E	OAC Kippen	6	E
Birka	2	E	Fairfield	2	W	Perth	6	E
Bonanza	6	W	Galt	6	W	Ponoka	2	W
Bridge	2	W	Gateway 63	6	W	Rodeo	2	E
Brier	6	W	Harrington	2	W	Sabina	6	E
Bruce	6	E	Heartland	6	W	Samson	6	W
Brucefield	6	E	Hector	2	W	Stein	2	W
Cadette	6	E	Helena	2	E	Symko	2	E
CDC Battleford	6	W	Herta	2	W	Synasolis	6	E
CDC Bold	2	W	Iona	2	E	Tankard	6	W
CDC Clyde	6	W	Island	2	E	TR251	2	W
CDC Coalition	2	W	Johnston	6	W	TR306	2	W
CDC Copeland	2	W	Joly	6	E	Virden	6	W
CDC Cowboy	2	W	Klages	2	W	Vivar	6	W
CDC Dolly	2	W	Klondike	6	W	Volla	2	E
CDC Earl	6	W	Laurier	6	E	Yielder	6	E
Centennial	2	W	Leduc	6	W			

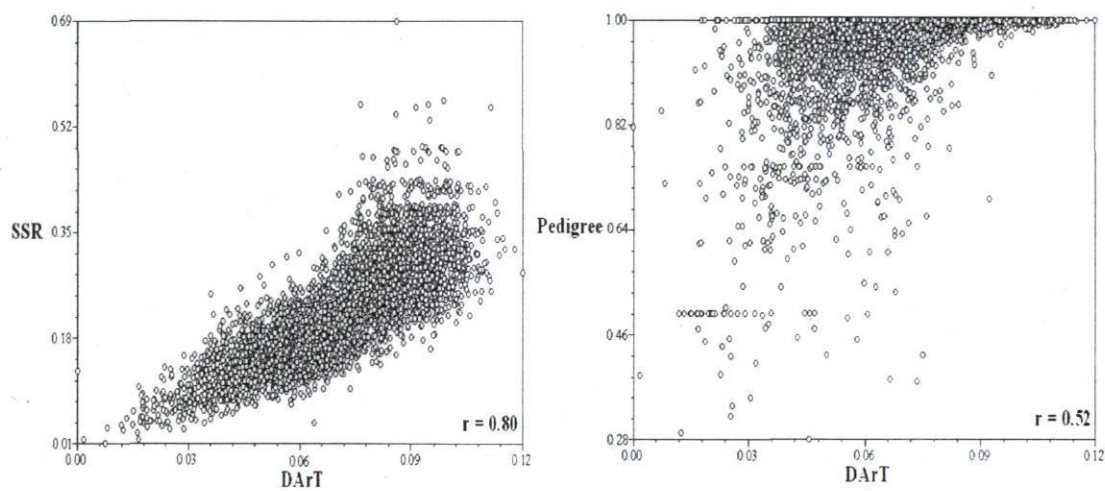


Figure 3-1 - Correlation between pairwise genetic distance values computed using SSR, DArT and pedigree data.

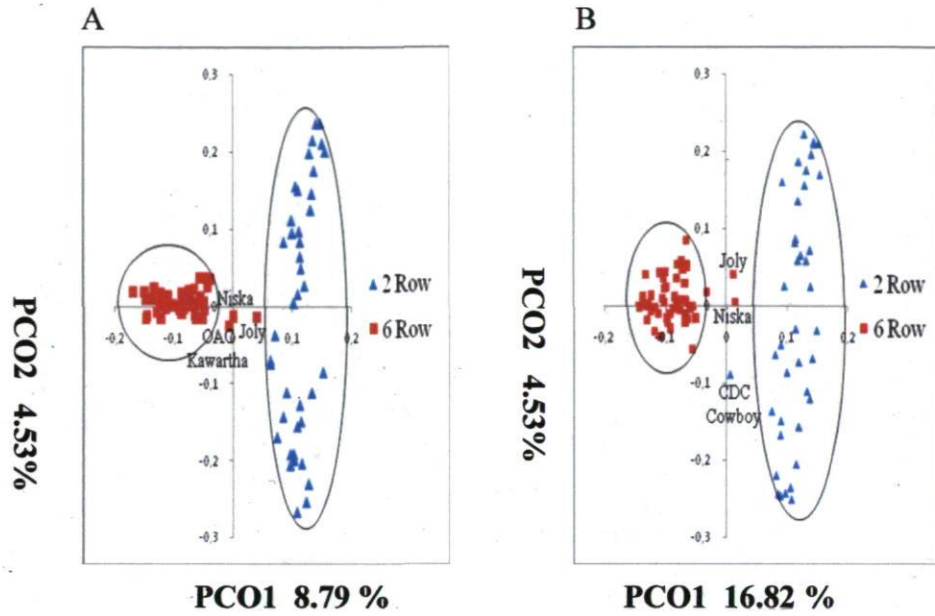


Figure 3-2 - Relationships between barley lines visualized by Principal coordinate analysis (PCoA) using SSR data (A) or DArT data (B).

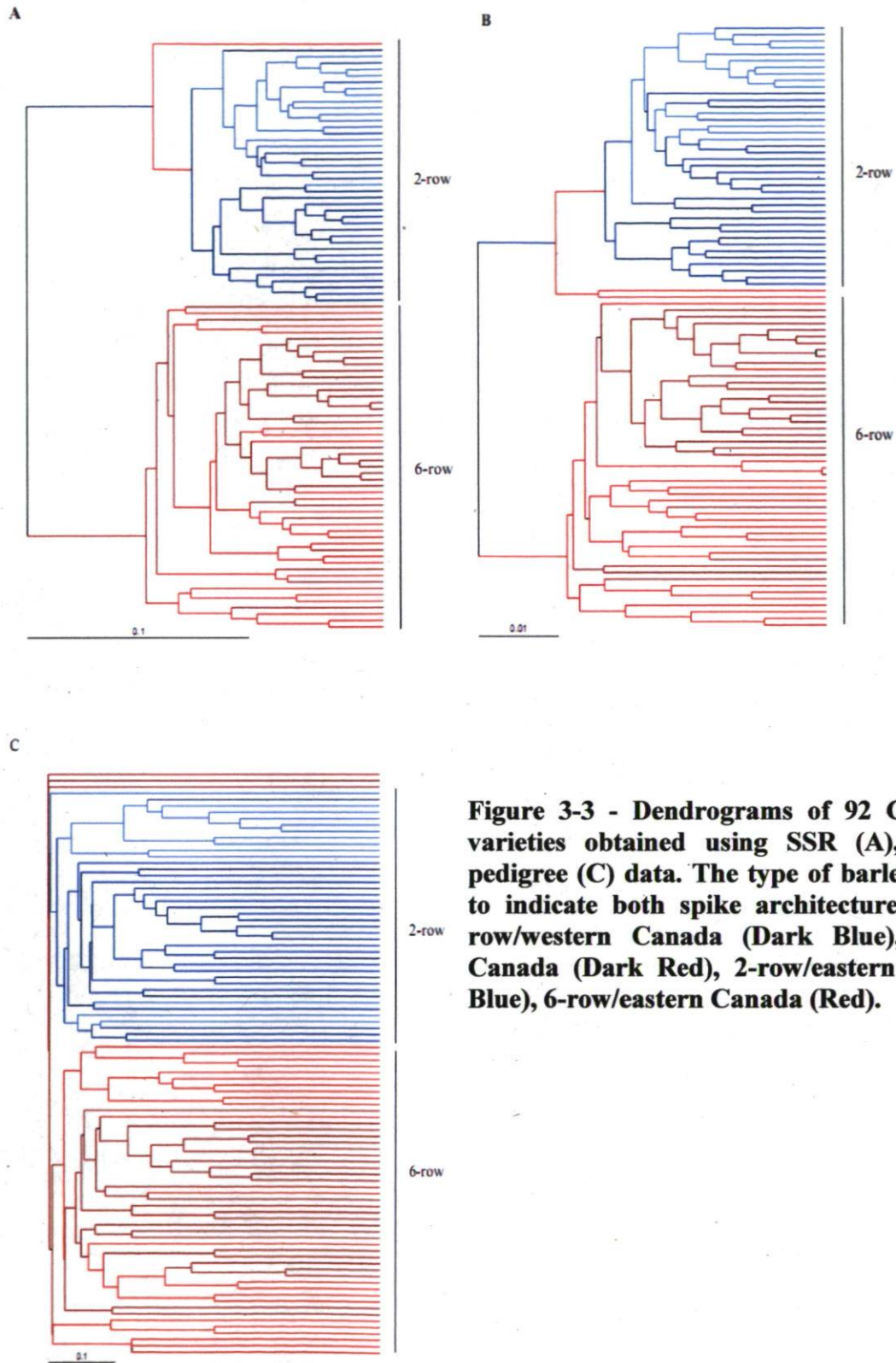


Table 1S - Linkage group, position (cM), number of alleles detected, PIC, and repeat motif of the 50 SSRs used in this study.

SSR	Linkage group	Position (cM)	No. of alleles	PIC	Repeat motif
Bmag0504	1H	49.18	4	0.54	(AG)6,(GAGG)4,(GA)7
Bmag0718	1H	50.9	5	0.64	(GA)18,(AG)6
Bmag0103c	1H	60.42	4	0.42	(AG)22
Bmac0090	1H	63.66	7	0.73	(AC)20
EBmac0501	1H	64.84	7	0.75	(AC)13
Bmac0032	1H	73.74	11	0.82	(AC)7T(CA)15(AT)9
Bmac0154	1H	87.83	8	0.73	(AT)19/(AC)6
WMC1E8	1H	131.93	2	0.49	(AC)24
Bmac0134	2H	10.87	10	0.83	(AC)28
Bmac0093	2H	67.01	8	0.76	(AC)24
Bmag0140	2H	76.26	4	0.69	(AG)15
EBmac0415	2H	117.86	6	0.59	(AC)17
HVM54	2H	122.41	6	0.73	(GA)14
Bmag0749	2H	147.93	4	0.65	(AG)11
Bmag0138a	3H	52.99	6	0.74	(AG)6(AG)15
Bmag0225	3H	75.54	10	0.79	(AG)26
Bmag0841	3H	91.21	7	0.70	(TC)8(CCTC)4,(CT)11
Bmag0606	3H	101.77	9	0.79	(CT)22
Bmag0013	3H	113.70	13	0.82	(CT)21
Bmag0877	3H	124.98	7	0.60	(GA)15
EBmac0541	3H	137.16	5	0.58	(AC)9
HVM62	3H	139.43	6	0.60	(GA)11
EBmac0708	3H	142.33	11	0.78	(AT)5(GT)16CT(GT)
Bmag0853	3H	144.10	6	0.55	(GA)15
Bmac0273d	4H	22.40	4	0.39	(AC)20(AG)20
HVM40	4H	22.40	6	0.40	(GA)7
EBmac0906	4H	54.98	10	0.55	(GC)5GGG(GT)16
EBmac0775	4H	64.37	13	0.86	(TG)4TT(TG)17(AG)10
EBmac0635	4H	93.06	7	0.76	(AC)23
EBmac0701	4H	96.17	9	0.79	(AC)23
EBmac0788	4H	97.67	10	0.78	(TG)23
HVM67	4H	120.50	2	0.47	(GA)11
Bmac0273b	5H	42.83	3	0.34	(AC)20(AG)20
Bmac0303b	5H	53.10	10	0.80	(AG)13(AC)21
Bmag0812	5H	90.32	7	0.80	(CT)26
GMS061	5H	122.09	5	0.67	(GA)16
Bmag0222	5H	144.14	8	0.76	(AC)9(AG)17
GMS001	5H	183.25	5	0.76	(CT)7TTT(CT)2
Bmag0870	6H	50.5	6	0.74	(TC)33
Bmag0807	6H	56.11	6	0.68	(TC)18
Bmac0040b	6H	61.07	11	0.85	(ATC)6N3(ATC)3
Bmag0613	6H	69.82	9	0.80	(GA)17
Bmag0571a	6H	105.45	10	0.83	(GA)25
Bmag0007	7H	20.62	11	0.76	(AG)16(AC)16
EBmac0603	7H	35.39	9	0.71	(CA)10
Bmag0516	7H	79.31	10	0.81	(TC)8(TATC)7(TC)19
Bmac0273a	7H	80.78	4	0.70	(AC)20(AG)20
Bmag0120	7H	97.00	11	0.82	(AG)15
EBmac0755	7H	125.21	11	0.74	(AC)16
Bmag0135	7H	147.51	5	0.70	(AG)10GG(AG)12

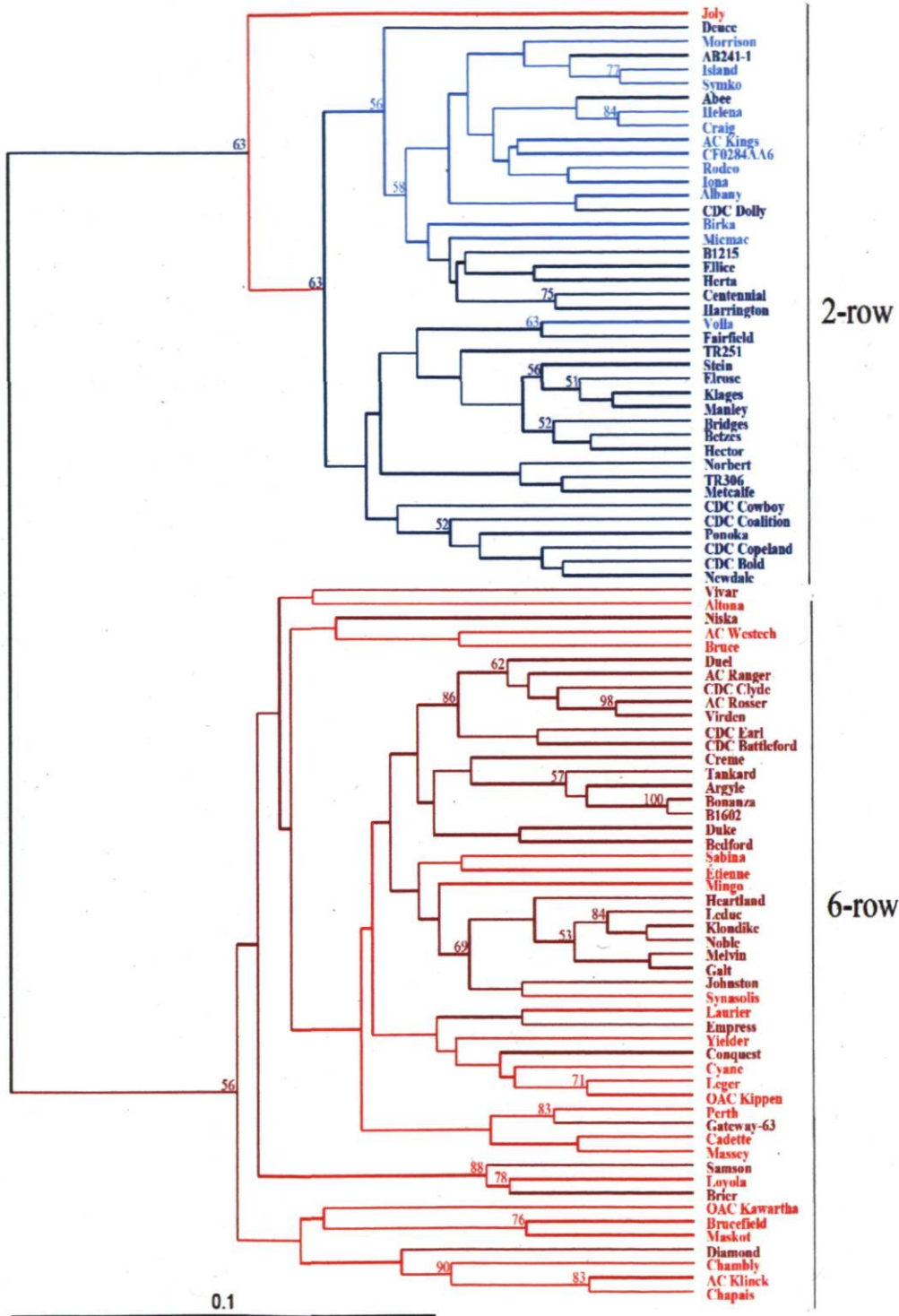


Figure 1S - Dendrogram of 92 Canadian barley varieties obtained using SSR markers. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).

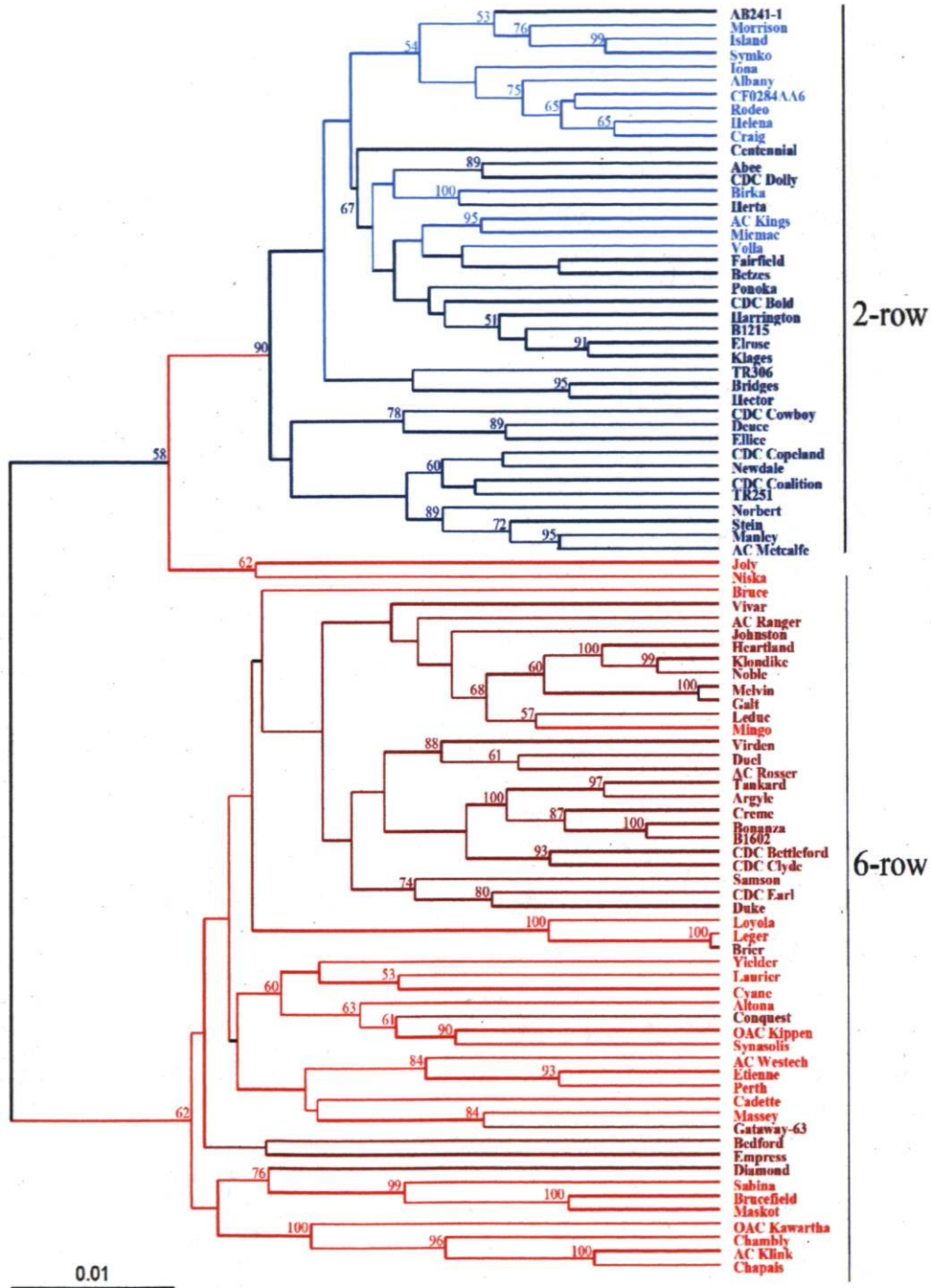


Figure 2S - Dendrogram of 92 Canadian barley varieties obtained using DArT markers. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).

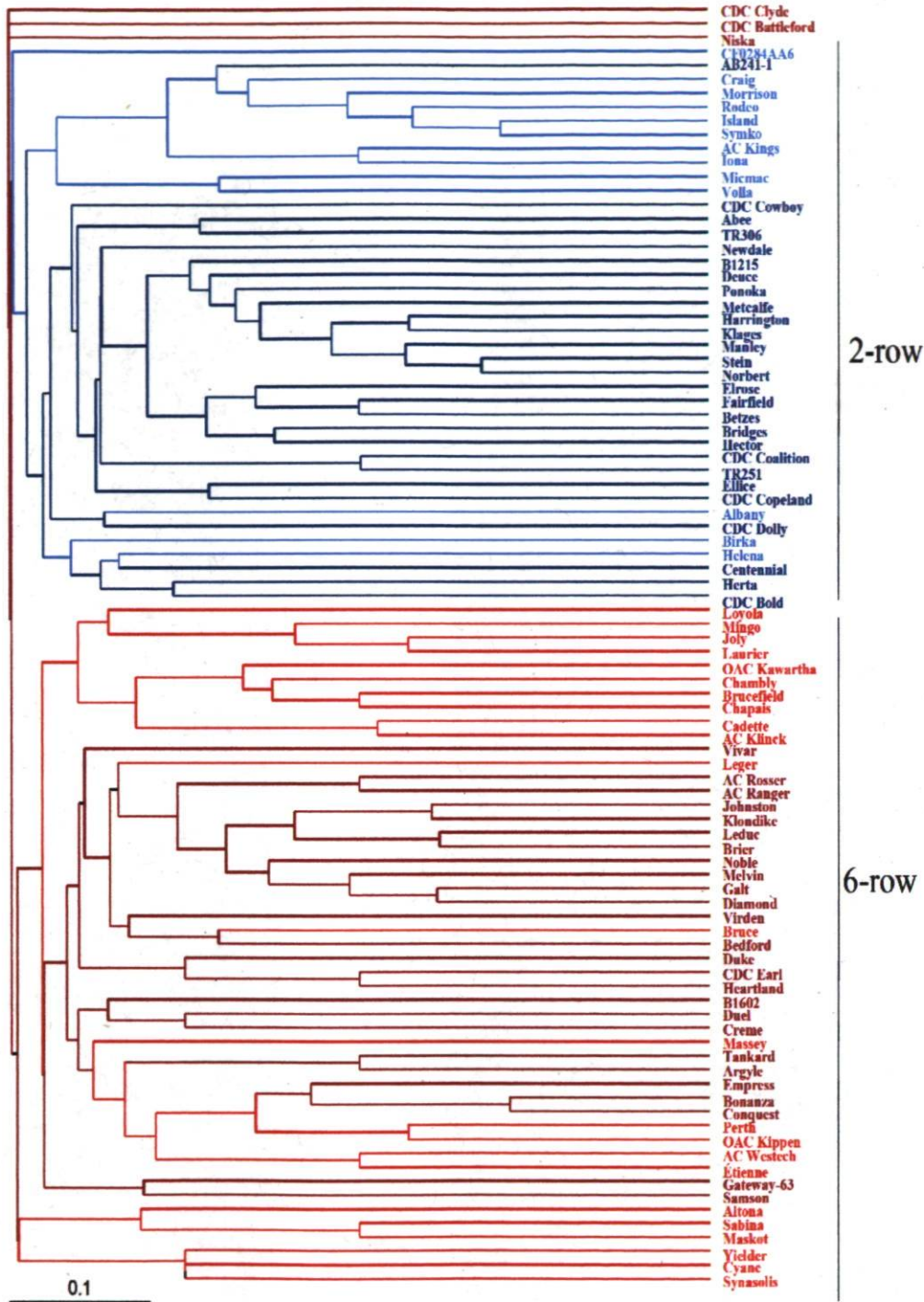


Figure 3S - Dendrogram of 92 Canadian barley varieties obtained using pedigree data. The type of barley is color-coded to indicate both spike architecture and origin: 2-row/western Canada (Dark Blue), 6-row/western Canada (Dark Red), 2-row/eastern Canada (Light Blue), 6-row/eastern Canada (Red).

CHAPITRE IV : CONCLUSION GÉNÉRALE

4.1 Conclusion générale

L'objectif principal de ce travail était de comparer deux types de marqueurs (SSR et DArT) et le pedigree pour leurs capacités à décrire les relations entre les lignées d'orge canadiennes. La connaissance des relations qui existent entre les génotypes permet, pour le sélectionneur, de faire un meilleur choix des parents à l'origine du croisement de départ pour mieux exploiter la diversité génétique dans un programme de sélection. Au meilleur de notre connaissance, notre travail est la première étude comparative effectuée entre la nouvelle génération des marqueurs moléculaires à haut débit (DArT), les marqueurs conventionnels (SSR) et le pedigree pour l'étude de la diversité génétique chez l'orge. Cette étude a permis la validation des marqueurs DArT et montré leur utilité dans des analyses visant à déterminer les relations génétiques qui existent entre les lignées.

Dans le premier volet de cette étude, 50 marqueurs SSR ont permis d'identifier un total de 368 allèles. Le nombre d'allèles par locus variait de 2 à 13 ($\bar{x} = 7,36$) et les valeurs de PIC variaient entre 0,34 à 0,86 ($\bar{x} = 0,69$). L'analyse de groupement à partir d'une matrice de distance génétique a permis de déterminer le regroupement des génotypes sous la forme d'un arbre phylogénétique. Pour les marqueurs DArT, une matrice de distance génétique a été produite à partir de 971 marqueurs polymorphes dont l'indice PIC variait de 0,06 à 0,50 ($\bar{x} = 0,39$). Les distances génétiques ont été utilisées ensuite pour produire un second arbre. Une troisième matrice de distance, obtenue à partir d'un coefficient de parenté (« Kinship coefficient ») calculé sur la base du pedigree des lignées, a permis de produire un troisième arbre. Ces trois méthodes ont permis de déterminer les liens entre ces lignées. L'analyse de groupement a permis de séparer la population en deux grand groupes en fonction du type d'orge (2 rangs vs 6 rangs). L'origine géographique (est, ouest) était aussi un facteur important dans le regroupement.

Dans un deuxième volet de l'étude, une comparaison de la similarité topologique des trois arbres a été réalisée en utilisant l'indice de congruence de de Vienne et coll. (2007). Cette comparaison a montré que les trois arbres étaient significativement plus semblables qu'attendu par le simple jeu du hasard. L'analyse statistique a montré également une importante corrélation entre les matrices de distance des marqueurs SSR et DArT ($r = 0,80$; $p < 0,002$), mais une plus faible corrélation entre les matrices DArT et pedigree ($r = 0,52$; $p < 0,002$), ou entre les matrices SSR et pedigree

($r = 0,46$; $p < 0,002$). Des trois méthodes employées, l'analyse du pedigree est celle qui offrait la moins grande résolution. La très bonne concordance obtenue entre les arbres produits à l'aide de marqueurs (SSR, DArT) et du pedigree des lignées contraste avec la piètre congruence rapportée récemment par Soleimani et coll. (2007). En effet, ces auteurs ont rapporté chez l'orge une faible concordance entre les résultats d'analyses génétiques effectuées à l'aide de marqueurs S-SAP et SNP ainsi qu'avec le pedigree. Il est possible que cela découle du faible nombre de marqueurs employés par ces auteurs et du faible nombre de générations examinées lors de l'étude du pedigree.

En termes d'efficacité, les marqueurs DArT sont plus efficaces que les marqueurs SSR pour l'étude de la diversité génétique. Les marqueurs DArT permettent d'obtenir des résultats en 15 jours tandis qu'il en faut au moins cinq mois pour obtenir des résultats similaires avec les SSR. De plus, l'utilisation des SSR dans le cadre de l'étude de la diversité génétique nécessite un travail d'optimisation et de génotypage assez laborieux. Une évaluation sommaire de la situation a permis de mettre en évidence que l'utilisation des marqueurs DArT est beaucoup plus économique que celle des marqueurs SSR. Il faut environ entre 10.000 et 30.000 \$ pour réaliser une telle analyse avec les SSR, contre seulement 6000 \$ pour les marqueurs DArT. Bien que l'utilisation du pedigree ne nécessite pas de travail au laboratoire, la recherche d'informations reste laborieuse et problématique. De surcroît, le pedigree génère des résultats parfois incomplets ou imprécis, ce qui le rend très peu efficace pour l'étude de la diversité génétique. Partant de ces constats, les marqueurs DArT représentent un outil de choix pour l'analyse de la diversité génétique et l'identification des cultivars chez l'orge. Ils présentent aussi l'avantage d'être cartographiés et donc leur utilisation dans des études d'association à des caractères d'intérêt agronomiques chez l'orge permettront d'identifier avec plus de précision les régions chromosomiques responsables.

Nos résultats montrent qu'avec un nombre suffisant de marqueurs, des portraits concordants sont obtenus et permettent de documenter avec précision les liens qui existent entre les divers groupes de lignées d'orges cultivées au Canada. Cette information s'avère particulièrement importante en vue de la gestion des ressources génétiques et de leur utilisation raisonnée dans les programmes de sélection.

Références

- Agriculture et Agroalimentaire Canada. 2007. Orge fourragère: situation et perspectives. Le Bulletin bimensuel Volume 20 Numéro 15. www.agr.gc.ca/mad-dam/ accessible le 08 novembre 2007.
- Agriculture et Agroalimentaire Canada. 2005. Orge brassicole : situation et perspective. Le bulletin bimensuel Volume 18 Numéro 4. <http://www.agr.gc.ca/mad-dam/>.
- Asins M.J. and Carbonell E.A. 1989. Distribution of genetic variability in a durum wheat world collection. *Theor. Appl. Genet.* 77: 287–294.
- Bank H.V.D., Bank M.V.D and Wyk B.E.V. 2001. A review of the use of allozyme electrophoresis in plant systematics. *Biochem. Syst. Ecol.* 29: 469–483.
- Bennett M.D. and Smith J.B. 1976. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London, Series B- Biological Sciences* 274: 227–274.
- Baum B.R., Nevo E., Johnson D.A. and Beiles A. 1997. Genetic diversity in wild barley (*Hordeum spontaneum* C. Koch) in the Near East: a molecular analysis using Random Amplified Polymorphic DNA (RAPD) markers. *Genet. Res. Crop Evol.* 44: 147–157.
- Becker J., Vos P., Kuiper M., Salamini F. and Heun M. 1995. Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.* 249: 65–73.
- Botstein D., White R. L., Skolnick M. and Davis R. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314–331.
- Bretting P.K. and Widrechner M.P. 1995. Genetic markers and plant genetic resource management. In : Janick J. ed. *Plant Breeding Reviews*. New York: John Wiley & Sons Inc. 13: 11–86.
- Bundock P.C. and Henry R.J. 2004. Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. *Theor. Appl. Genet.* 109: 543–551.
- Buth D.G. and Murphy R.W. 1999. The use of isozyme characters in systematic studies. *Biochem. Syst. Ecol.* 27: 117–129.
- Casas M.A., Igartua E., Valles M.P. and Molina-Cano J.L. 1998. Genetic diversity of barley cultivars grown in Spain by RFLP, similarity and co-ancestry coefficients. *Plant Breed.* 117: 429–435.
- Chiapparino E., Lee D. and Donini P. 2004. Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. *Genome* 47: 414–420.
- Collins F. S., Brooks L. D. and Chakravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 12: 1229–1231.

- Cox T.S., Hkiang Y.T., Gorman M.B. and Rogers D.M. 1985. Relationships between coefficient of parentage and genetic indices in the soybean. *Crop Sci.* 25: 529–532.
- Dávila J.A., Loarce Y. and Ferrer E. 1999. Molecular characterization and genetic mapping of random amplified microsatellite polymorphism in barley. *Theor. Appl. Genet.* 98: 265–273.
- de la Pena R.C., Smith K.P., Capettini F., Muehlbauer G.J., Gallo-Meagher M., Dill-Macky R., Somers D.A. and Rasmusson D.C. 1999. Quantitative trait loci associated with resistance to *Fusarium* head blight and kernel discoloration in barley. *Theor. Appl. Genet.* 99: 561–569.
- de Vienne D. 1998. *Biotechnologies végétales. Les marqueurs moléculaires et leurs applications.* CNED. Paris. 118pp.
- Excoffier L., Smouse P. and Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- Graner A., Ludwig W.F. and Melchinger A.E. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. *Crop Sci.* 34: 1199–1205.
- Graner A., Jahoor A., Schondelmaier J., Siedler H., Pillen K., Fischbeck G., Wenzel G. and Herrmann R.G. 1991. Construction of an RFLP map of barley. *Theor. Appl. Genet.* 83: 250–256.
- Gupta P.K., Rustgi S. and Mir R.R. 2008. Array-based high-throughput DNA markers for crop improvement. *Heredity* 101: 5–18.
- Hayes P.M.J., Ceroni H., Witsenhoer M., Kuiper M., Zabeau K., Sato A., Kleinhofs D., Kudrna A., Kilian M., Saghai-Marooof D. Hoffman and the North American Barley Genome Mapping Project. 1997. Characterizing and exploiting genetic diversity and quantitative traits in barley (*Hordeum vulgare*) using AFLP markers. *J. Agric. Genomics*, Volume 3, Article 2. ([//research/jag/papers97/paper297/indexp297.html](http://research/jag/papers97/paper297/indexp297.html)).
- He C., Sayed-Tabatabaei B.E. and Komatsuda T. 2004. AFLP targeting of the 1-cM region conferring the *vrs1* gene for six-rowed spike in barley, *Hordeum vulgare* L. *Genome* 47: 1122–1129.
- Hearnden P. R., Eckermann P. J., McMichael G. L., Hayden M. J., Eglinton J. K. and Chalmers K. J. 2007. A genetic map of 1,000 SSR and DArT markers in a wide barley cross. *Theor. Appl. Genet.* 115: 383–391.
- Helentjaris T., King G., Slocum M., Sidenstrang C. and Wegman S. 1985. Restriction fragment length polymorphism as probes for plant diversity and their development as tools for applied plant breeding. *Plant Mol. Biol.* 5: 109–118.
- Jaccard P. 1908. Nouvelles recherches sur la distribution florale. *Bull. Doc. Vaudoise Sci. Nat.* 44: 223–270.
- Jaccoud D., Peng K., Feinstein D. and Killian A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucl. Acids Res.* Vol. 29, No 4 e25.

- Jeffrey P. 2008. SNP genotyping: six technologies that keyed a revolution. *Nature* 5: 447–453.
- Koornneef M., Alonso-Blanco C. and Peeters A.J.M. 1997. Genetic approach in plant physiology. *New Phytol.* 137: 1–8.
- Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y. and Gale M. 1994. Conservation of genome structure between rice and wheat. *Nat. Bio/technology* 12: 276–278.
- Litt M. and Luty J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397–401.
- Liu Z.W., Biyashev R.M. and Saghai-Marooif M.S. 1996. Development of simple sequence repeat DNA markers and their integration into a barley linkage map. *Theor. Appl. Genet.* 93: 869–876.
- Matus I.A. and Hayes P.M. 2002. Genetic diversity in three groups of barley germplasm assessed by simple sequence repeat. *Genome* 45: 1095–1106.
- Melchinger A.E., Graner A., Singh M. and Messmer M.M. 1994. Relationships among European Barley Germplasm: I. Genetic Diversity among Winter and Spring Cultivars Revealed by RFLPs. *Crop Sci.* 34: 1191–1199.
- Morrell P. and Clegg M.T. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. USA.* 104: 3289–3294.
- Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* 10: 5269–5273.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA.* 70: 3321–3323.
- Pillen K., Binder A., Kreuzkam B., Ramsay L., Waugh R., Forster J. and Leon J. 2000. Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. *Theor. Appl. Genet.* 101: 652–660.
- Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S. and Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225–238.
- Powell W., Thomas W.T.B., Baird E., Lawrence P., Booth A., Harrower B., McNicol J.W. and Waugh R. 1997. Analysis of quantitative traits in barley by use of Amplified Fragment Length Polymorphisms. *Heredity* 79: 48–59.
- Rafalski A. 2002. Application of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5: 94–100.
- Ramsay L., Macaulay M., Ivanissevich S., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvevsson S., Morgante M., Massari A., Maestri E., Marmiroli N., Sjakste T., Ganai M., Powell W. and Waugh R. 2000. A simple sequence repeat-based linkage map of barley. *Genetics* 156: 1997–2005.

- Russell J., Booth A., Fuller J., Harrower B., Hedley P., Machray G. and Powell W. 2004. A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome* 47: 389–398.
- Russell J.R., Fuller J.D., Macaulay M., Hatz B.G., Jahoor A., Powell W. and Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor. Appl. Genet.* 95: 714–722.
- Rostoks N., Mudie S., Cardle L., Russell J., Ramsay L., Booth A., Svensson J.T., Wanamaker S.I., Walia H., Rodriguez E.M., Hedley P.E., Liu H., Morris J., Close T.J., Marshall D.F. and Waugh R. 2005. Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genomics.* 274: 515–527.
- Saghai-Marooif M.A., Biyashev R.M., Yang G.P., Zhang Q. and Allard R.W. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. *Proc. Natl. Acad. Sci. USA.* 91: 5466–5470.
- Saghai-Marooif M.A., Soliman K.M., Jørgensen R.A. and Allard R.W. 1984. Ribosomal DNA spacer-length in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA.* 81: 8014–8018.
- Simons G., Van der Lee T., Diergaarde P., Van Daelen R., Groenendijk J., Frijters A., Büschges R., Hollricher K., Töpsch S., Schulze-Lefert P., Salamini F., Zabeau M. and Vos P. 1997. AFLP-based fine mapping of the *Mlo* gene to a 30-kb DNA segment of the barley genome. *Genomics* 44: 61–70.
- Smith J.S.C., Zabeau M. and Wright S. 1993. Association among inbred lines as revealed by RFLPs and by a thermocycling methodology, Amplified Fragment Length Polymorphisms (AFLPs). *Maize Genet. Newsl.* 68: 62–64.
- Sobrino B., Brion M. and Carracedo A. 2005. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. Int.* 154: 181–194.
- Sokal R.R. and Michener C.D. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38: 1409–1438.
- Soleimani V.D., Baum B.R. and Johnson D.A. 2007. Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data. *Genet. Res. Crop Evol.* 54: 83–97.
- Soleimani V.D., Baum B.R. and Johnson D. A. 2005. Genetic diversity among barley assessed by sequence-specific amplification polymorphism. *Theo. Appl. Genet.* 110: 1290–1300.
- Soleimani V.D., Baum B.R., Jonson D.A. 2002. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp *durum* (Desf.) Husn.]. *Theor. Appl. Genet.* 104: 350–357.
- Struss D. and Plieske J. 1998. The use of microsatellite markers for detection of genetic diversity in barley populations. *Theor. Appl. Genet.* 97: 308–315.

- Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl. Ac. Res.* 16: 6463–6471.
- Tinker N.A. and Deyl J.K. 2005. A Curated Internet Database of Oat Pedigree. *Crop Sci.* 45: 2269–2272.
- Tinker N.A., Fortin M.G. and Mather D.E. 1993. Random amplified polymorphic DNA and pedigree relationship in spring barley. *Theor. Appl. Genet.* 85: 976–984.
- Varshney R.K., Thiel T., Sretenovic-Rajcic T., Baum M., Valkoun J., Guo P., Grando S., Ceccarelli S. and Graner A. 2008. Identification and validation of a score set of informative genic SSR and SNP markers for assaying functional diversity in barley. *Mol. Breeding* 22: 1–13.
- Vos P., Hogers R., Bleeker M., Reijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kupier M. and Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* 23: 4407–4414.
- Welsh J. and McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18: 7213–7218.
- Wenzl P., Li H., Carling J., Zhou M., Raman H., Paul E. et al. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* 7: 206.
- Wenzl P., Carling J., Kudrna D., Jaccoud D., Huttner E., Kleinhofs A. and Kilian A. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. USA.* 101: 9915–9920.
- Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A. and Tingey S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 22: 6531–6535.
- Wright S. 1978. *Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations.* University of Chicago Press. Chicago.
- Wright S. 1951. The genetical structure of populations. *Ann. Eugenics* 15: 323–354.
- Zhu H., Gilchrist L., Hayes P., Kleinhofs A., Kudrna D., Liu Z., Prom L., Steffenson B., Toojinda T. and Vivar H. 1999. Does function follow form? Principal QTLs for Fusarium head blight (FHB) resistance are coincident with QTLs for inflorescence traits and plant height in a doubled-haploid population of barley. *Theor. Appl. Genet.* 99: 1221–1232.

LISTE DES ANNEXES

ANNEXE I – Plant DNA Extraction Protocol for DArT

BUFFER STOCK SOLUTIONS

EXTRACTION BUFFER STOCK

	<u>To make 500 ml:</u>
0.35 M sorbitol	31.9 g sorbitol
0.1 M TrisHCl pH 8.0	50 ml 1M TrisHCl pH 8.0
5 mM EDTA pH 8.0	5 ml 0.5 M EDTA pH 8.0
	fill up to 500 ml MiliQ H ₂ O

LYSIS BUFFER STOCK

	<u>To make 500 ml:</u>
0.2 M Tris HCl pH 8.0	100 ml 1M Tri HCl pH 8.0
0.05 M EDTA pH 8.0	50 ml 0.5 M EDTA pH 8.0
2M NaCl	200 ml 5 M NaCl
2% CTAB	10 g CTAB
	fill up to 500 ml with MilliQ H ₂ O

SARCOSYL STOCK 5% (w/v)

FRESH BUFFER WORKING SOLUTION*:

0.5 % (w/v) sodiumdisulfite (= sodium metabisulfite)

2 % (w/v) PVP-40 (K29-32) Sigma

dissolve in required volume of extraction buffer stock; add same volume of lysis buffer stock and 0.4 volume of extraction (=lysis) buffer stock of sarcosyl stock.

For example to make 120 ml:

Add 0.6 g sodiumdisulfite (= sodium metabisulfite) and 2.4 g PVP-40 (K29-32) to 50 ml extraction buffer stock and dissolve; add 50 ml lysis buffer stock and 20 ml sarcosyl stock.

For example to make 30 ml:

Add 0.15 g sodiumdisulfite (= sodium metabisulfite) and 0.6 g PVP-40 (K29-32) to 12.5 ml extraction buffer stock and dissolve; add 12.5 ml lysis buffer stock and 5 ml sarcosyl stock

*This buffer may settle into two layers on standing. Heat to 65°C and shake immediately before adding to

PROTOCOL

For 15 ml Sarsted tubes:

- Aliquot 6 ml of freshly prepared preheated to 65°C well mixed “fresh buffer solution” and place tubes to the 65°C incubator or water bath, (3, 4 days old “fresh buffer solution” works fine),
- Grind required amount (same across all samples) of plant material in mortar and pestle under liquid nitrogen to fine powder,
- Suspend powder in 6 ml “fresh buffer solution” kept at 65°C (make sure there are no clumps, vortex if necessary),
- Incubate at 65°C for 1 h (can extend for another 30 min), invert tubes in every 20 minutes or incubi with gentle shaking,
- Cool down for 5 min and add 6 ml of chloroform: isoamyl alcohol (24 : 1) mixture,
- Mix well for 30 min,
- Spin 20 min, 3000 x g, RT,
- Transfer water phase to fresh tube, add same volume of ice cold isopropanol and invert tube ~ 10 times, nucleic acids should become visible,
- Spin 30 min, 3000 x g, RT,
- Discard supernatant, wash pellet with 2 ml 70 % EtOH,
- Discard EtOH, dry pellet and dissolve in 200 µl – 500 µl 1 x TE (10 mM TrisHCl pH 8.0, 1 mM EDTA pH 8.0),
- Check DNA quality and quantity on 0.8 % agarose gel. (If RNA quantity is several fold less than

DNA, RNase treatment is not necessary for DArT applications).

For 2 ml Eppendorf tubes:

- Aliquot 1 ml of freshly prepared preheated to 65°C, well mixed “fresh buffer solution” and place tubes to the 65°C incubator or water bath, (3, 4 days old “fresh buffer solution” works fine),
- Grind required amount (same across all samples) of plant material in mortar and pestle under liquid nitrogen to fine powder,
- Suspend powder in 1 ml “fresh buffer solution” kept at 65°C (make sure there are no clumps, vortex if necessary),
- Incubate at 65°C for 1 h (can extend for another 30 min), invert tubes in every 20 minutes or incubate with gentle shaking,
- Cool down for 5 min and add 1 ml of chloroform: isoamyl alcohol (24 : 1) mixture,
- Mix well for 30 min,
- Spin 20 min, 10000 x g, RT,
- Transfer water phase to fresh tube, add same volume of ice cold isopropanol and invert tube ~ 10 times, nucleic acids should become visible,
- Spin 30 min, 10000 x g, RT,
- Discard supernatant, wash pellet with 2 ml 70 % EtOH,
- Discard EtOH, dry pellet and dissolve in 250 µl of 1 x TE (10 mM TrisHCl pH 8.0, 1 mM EDTA pH 8.0),
- Check DNA quality and quantity on 0.8 % agarose gel. (If RNA quantity is several fold less than DNA, RNase treatment is not necessary for DArT applications).