

CIHEAM



Centre
International
de Hautes Etudes
Agronomiques Méditerranéennes

International
Centre for
Advanced
Mediterranean Agronomic Studies

Thèse / Thesis

requis pour
l'obtention du Titre

submitted
for the Degree of

Master of Science

CARACTERISATION MOLECULAIRE PAR
MICROSATELLITES DES ORGES A SIX
RANGS DE LA COLLECTION NUCLEAIRE
ESPAGNOLE

Samia YAHIAOUI

Saragosse, Avril 2003

Institut Agronomique Méditerranéen de
Mediterranean Agronomic Institut of
Zaragoza



3-2



CIHEAM

**Centre International de Hautes Etudes
Agronomiques Méditerranéennes**

*International Centre for Advanced
Mediterranean Agronomic Studies*

Secrétariat Général / General Secretary

11, rue Newton
75116 PARIS

Tel.: (33-1) 53 23 91 00 – Fax: (33-1) 53 23 91 01
e-mail: secretariat@ciheam.org



**Instituts Agronomiques Méditerranéens
Mediterranean Agronomic Institutes**

(IAM)

Bari – Chania – Montpellier – Zaragoza

IAM - Bari

Via Ceglie 9
70010 Valenzano, Bari, Italy
Tel.: (39) 80 78 06 111 – Fax: (39) 80 78 06 206
e-mail: [iamdir@iamb.it](mailto:iamdri@iamb.it)

IAM - Chania

P.O. Box 85
73100 Chania, Crete, Greece
Tel.: (30) 821 81 151 – Fax: (30) 821 81 154
e-mail: alkinoos@maich.gr

IAM - Montpellier

3191, route de Mende – BP 5056
34033 Montpellier Cedex 1, France
Tel.: (33-4) 67 04 60 00 – Fax: (33-4) 67 54 25 27
e-mail: iamm@iamm.fr

IAM - Zaragoza

Apdo. 202
50080 Zaragoza, Spain
Tel.: (34) 76 57 60 13 – Fax: (34) 76 57 63 77
e-mail: iamz@iamz.ciheam.org

1015525
CB 242412

TC-2003-2

CENTRE INTERNATIONAL DE HAUTES ETUDES AGRONOMIQUES MEDITERRANEENNES
INSTITUT AGRONOMIQUE MEDITERRANEEN DE ZARAGOZA

**CARACTERISATION MOLECULAIRE PAR MICROSATELLITES
DES ORGES A SIX RANGS DE LA COLLECTION NUCLEAIRE ESPAGNOLE**

Samia YAHIAOUI

Travail réalisé au Département d'Amélioration et de Production Végétale, Consejo Superior de Investigaciones Científicas (CSIC), Zaragoza, sous l'encadrement des **Drs Ana M^a CASAS** et **Ernesto IGARTUA**,

et soutenu publiquement le 1^{er} avril 2003, devant le jury suivant:

- **Ignacio ROMAGOSA**, Département de Production Végétale et Sciences Forestières, Centro Universitario UdL-IRTA, Lleida,
- **Nicolás JOUVE**, Département de Biologie Cellulaire et de Génétique, Facultad de Ciencias, Universidad de Alcalá de Henares, Madrid,
- **Antonio MARTIN**, Institut d'Agriculture Durable, Consejo Superior de Investigaciones Científicas (CSIC), Córdoba,
- **Pere ARUS**, Unité d'Amélioration Végétale, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Centre de Cabriels, Barcelona,
- **Dunixi GABIÑA**, Directeur Adjoint, Institut Agronomique Méditerranéen de Saragosse.

Q-11.081



Remerciements

A l'issue de ce travail j'aimerais exprimer mes remerciements et ma reconnaissance à toutes les personnes qui ont contribué d'une manière ou d'une autre à le rendre possible.

A mes directeurs de thèse Dr. Ana CASAS et Dr. Ernesto IGARTUA, pour avoir accepté de m'encadrer. Je tiens à leur exprimer ma profonde gratitude et reconnaissance pour leur aide, leurs orientations et conseils et surtout pour la patience qu'ils ont eu envers moi.

A toute l'équipe de l'IAMZ particulièrement, le Directeur Miguel Valls, Emilio Moràn, Ramzi Belkhdja et Antonio del Valle pour avoir confié en moi en m'offrant l'opportunité de réaliser ce master.

A mes collègues de laboratoire et du département de génétique et production végétale de la Station Expérimentale de Aula Dei (CSIC), pour leurs encouragements.

Mes remerciements vont aussi à Farida Dechmi et Ratiba Medjdoub, pour m'avoir soutenu durant les moments difficiles par lesquelles je suis passée.

Je tiens à remercier aussi Hocine pour son soutien, sa disponibilité et son aide.

Comme je remercie tout les camarades du Campus, pour leurs encouragements.

A mes parents

SOMMAIRE

INDEX DES TABLEAUX	iii
INDEX DES FIGURES	v
Resumen	vii
Résumé	ix
Summary	xi
I- INTRODUCTION	3
1-1 L'espèce <i>Hordeum vulgare</i> L.....	3
1-1-1 Taxonomie et cytogénétique.....	3
1-1-2 Origine et domestication.....	4
1-1-3 Amélioration de l'espèce	5
1-2 Méthodes d'évaluation de la diversité génétique	12
1-2-1 Caractères morphologiques	12
1-2-2 Marqueurs biochimiques	13
1-2-3 Marqueurs moléculaires	14
1-2-3-1 Méthodes non basées sur la PCR	14
1-2-3-2 Méthodes basées sur la PCR	15
a- PCR avec amorces arbitraires.....	15
b- PCR avec amorces spécifiques.....	16
1-2-3-3 Comparaison entre les différents marqueurs moléculaires	18
II - MATERIEL ET METHODES	23
2-1 Matériel végétal.....	23
2-2 Microsatellites étudiés	28
2-2-1 Description	28
2-2-2 Conditions d'amplifications et électrophorèse.....	30
2-2-3 Analyse des gels	30
2-3 Traitements statistiques des résultats	30
2-3-1 Similarité et indice de diversité	30
2-3-2 Analyse moléculaire de variance.....	32
2-3-3 Analyses de régression	33
2-3-4 Déséquilibre de Linkage.....	34
III- RESULTATS	37

3-1 Polymorphisme.....	37
3-2- Comparaison entre groupes de matériel établis à priori.	43
3-2-1- Polymorphisme	43
3-2-2 Indices de diversité.....	43
3-2-3 Analyse moléculaire de variance.....	46
3-3 Similarité génétique et analyse du cluster	51
3-3-1 Similarité génétique.....	51
3-3-2 Analyse Cluster	52
3-4 Analyse des coordonnées principales	57
3-4 Analyse des coordonnées principales	59
3-5 Diversité génétique entre les groupes obtenus par le cluster	63
3-5-1 Diversité génétique totale	63
3-5-2 Diversité génétique intra et entre-groupes du cluster	63
3-5-3 Distribution de la diversité entre les deux sous-groupes 2A et 2B	64
3-6 Distribution de la diversité génétique en relation avec les conditions du milieu d'origine	71
3-7 Déséquilibre de linkage ou de liaison	77
IV Discussion	85
4-1 Polymorphisme.....	85
4-2 Diversité	86
4-3 Associations SSRs et données écogéographiques	91
4-4 Déséquilibre de linkage.....	92
V- Conclusions.....	97
Références bibliographiques	101
Annexes	111

INDEX DES TABLEAUX

Tableau 1. Présentation des 148 accessions d'orge à 6 rangs de la collection nucléaire en fonction de leur distribution géographique par région agro-écologique et par province.....	24
Tableau 2. Pedigree et origine des variétés analysés dans cette étude.....	26-27
Tableau 3. Description des 29 microsatellites utilisés dans cette étude, incluant la localisation chromosomique, le motif répétitif et la taille en pb de chacun d'eux.....	28
Tableau 4. Origine des microsatellites.....	29
Tableau 5. Nombre d'allèles et rang de taille détecté par les différents loci; nombre d'individus correspondant à chaque allèle de chaque locus.....	39
Tableau 6. Nombre total d'allèles détectés dans chacun des groupes Espagnol, Européen et CIMMYT-ICARDA, et nombre d'allèles spécifiques de chacun d'eux.....	44
Tableau 7. Diversité et différenciation génétique intra et entre groupes Espagnol et Européen.....	45
Tableau 8. Analyse moléculaire de variance (AMOVA) basée sur la méthode de distance: somme carrée des écarts (SMM); groupe Espagnol vs groupe Européen.....	46
Tableau 9. Analyse moléculaire de variance (AMOVA) basée sur la méthode de distance: nombre d'allèles différents (IAM)); groupe Espagnol vs groupe Européen	47
Tableau 10. Analyse moléculaire de variance locus par locus réalisée selon les deux modèles IAM et SMM (groupe Espagnol vs groupe Européen); présentation des deux indices de fixation statistique F_{st} et R_{st}	48
Tableau 11. Similarité génétique (méthode Nei et Li, 1979), résumé en considérant les différentes combinaisons entre les groupes de germplasmé établis à priori	51
Tableau 12. Diversité génétique:, totale, intra et entre groupes issus du cluster et différenciation génétique	64
Tableau 13. Analyse moléculaire de variance (AMOVA) basée sur la méthode de distance: nombre d'allèles différents (IAM)); sous-groupe 2A vs sous-groupe 2B.....	65

Tableau 14. Analyse moléculaire de variance (AMOVA) basée sur la méthode de distance: somme carrée des écarts (SMM). sous-groupe 2A vs sous-groupe 2B.....	65
Tableau 15. Analyse moléculaire de variance locus par locus réalisée selon les deux méthodes IAM et SMM (sous-groupe A vs sous-groupe B); présentation des deux indices de fixation statistique Fst et Rst.....	66
Tableau 16. Coefficients de détermination établis par les analyses de régressions effectués entre les allèles (variable dépendante) avec les facteurs altitude, latitude et région agro-écologique (variables indépendantes), à un niveau de signification de 0.01. Les chiffres marqués en gras correspondent aux paramètres significatifs dans le modèle combiné.....	72
Tableau 17. Coefficients de détermination fournis par les analyses de régressions effectuées entre les loci (variables dépendantes) avec les facteurs altitude, latitude et région agro-écologique (variables indépendantes) à un niveau de signification de 0.01. Les chiffres marqués en gras correspondent aux paramètres significatifs au niveau du modèle combiné.....	74
Tableau 18. Nombre et pourcentage de paires de loci montrant un déséquilibre de linkage a un seuil de signification de 0.01%, dans les différents groupes (Espagnol, Européen, groupe 2 du cluster) et dans les sous-groupes 2A et 2B.....	77

INDEX DES FIGURES

Figure 1a. Principales variétés d'orge à deux rangs cultivées en régime sec en Espagne (estimation des superficies, DGA 2002).....	9
Figure 1b. Principales variétés d'orge à six rangs cultivées en régime sec en Espagne (estimation des superficies, DGA 2002).....	9
Figure 2a. Polymorphisme fourni par le microsatellite Bmac032 chez 42 accessions d'orge à six rangs de la collection nucléaire.....	41
Figure 2b. Polymorphisme fourni par le microsatellite HVM003 chez 42 accessions d'orge à six rangs de la collection nucléaire.....	41
Figure 3. Figure 3. Représentation graphique des génotypes de 152 accessions espagnoles ordonnées selon leur numéro BNG (à gauche) et 28 variétés européennes (à droite), Les génotypes sont représentés en lignes et les chromosomes sont arrangés en colonnes de 1H-7H, avec les loci SSRs analysés, correspondant à chaque chromosome. L'allèle le plus fréquent de chaque locus est représenté en rouge; les autres allèles sont représentés par d'autres couleurs comme c'est indiqué dans la légende (à droite). L'allèle nul est représenté sans couleur.	49
Figure 4. Dendrogramme résultant d'une analyse cluster UPGMA de 183 génotypes: 152 espagnols, 28 européens, 3 ICARDA-CIMMYT, en utilisant une matrice de similarité calculée selon la méthode Nei et Lei (1979), avec 29 microsatellites. Les différents groupes sont distingués par différentes couleurs. Les noms des génotypes sont colorés selon le type détecté par le STS MWG699 (A- Rouge, D-Bleu; pour plus de détail voir explication page 53).....	55
Figure 5. Photo d'un gel montrant les deux types A et D détectés par le STS MWG699 dans les génotypes à 6 rangs de la collection nucléaire.....	57
Figure 6. Analyse en coordonnées principales réalisée à partir de la matrice de similarité génétique (Nei et et Li, 1979) de 183 génotypes: 152 espagnols, 28 européens et 3 CIMMYT-ICARDA, analysés par 29 microsatellites.....	61
Figure 7. Représentation graphique des génotypes des deux sous-groupes 2A et 2B. Le sous groupe 2A (78 génotypes) est schématisé à gauche alors que et le sous- groupe 2B (57 génotypes) est représenté à droite. Les génotypes sont représenté en lignes. Les chromosomes sont arrangés de 1H-7H en colonnes, avec les loci SSRs analysés, correspondant à chaque chromosome. L'allèle le plus fréquent de chaque locus est représenté en rouge, les autres allèles sont représentés par d'autres couleurs comme c'est indique dans la légende (à droite). L'allèle nul est représenté sans couleur.....	69

Figure 8. Représentation de la distribution des allèles: a) Bmag369-4 (jaune), 5 (gris) et 6 (bleu); b) HVLTPPB/b-2 (jaune) et 3 (bleu); c) HVM062-2 (jaune), 3 (gris) et 4 (bleu); d) HVM040-2 (jaune) et 4 (bleu), selon les paramètres agro-écologiques (latitude et altitude), DIVA-GIS (<i>Hijmans et al.</i> , 2002)	75
Figure 9. Déséquilibre de linkage entre 28 loci polymorphiques, dans le groupe des orges Espagnoles (152 génotypes) à 6 rangs de la collection nucléaire, à un niveau de signification de 0.01.....	79
Figure 10. Déséquilibre de linkage entre 29 loci polymorphiques, dans le groupe des orges Européennes de référence (28 génotypes), à un niveau de signification de 0.01%.....	79
Figure 11. Déséquilibre de linkage entre 28 loci polymorphiques, dans le sous-groupe 2A (78 génotypes), à un niveau de signification de 0.01%.....	80
Figure 12. Déséquilibre de linkage entre 28 loci polymorphiques, dans le sous-groupe 2B (57 génotypes), à un niveau de signification de 0.01%.....	80
Figure 13. Déséquilibre de linkage dans le groupe 2 du cluster (A+B: 135 génotypes), à un niveau de signification de 0.01%.....	81

Resumen

Este trabajo se centra en la caracterización molecular mediante microsatélites y el estudio de la diversidad genética de las entradas de seis carreras de la colección nuclear española de cebada.

Veintinueve microsatélites, distribuidos por los siete cromosomas, han sido empleados para analizar 183 genotipos de cebada: 152 entradas de la colección nuclear, líneas puras procedentes de variedades locales, y 31 entradas de otros orígenes como testigos: 28 cultivares de otros países europeos (que constituyen el conjunto de referencia), y 3 entradas de materiales procedentes de los programas de CIMMYT-ICARDA.

El conjunto de microsatélites empleado detectó un alto grado de polimorfismo general. Asimismo, se identificaron bastantes alelos específicos de grupos de germoplasma, especialmente en las entradas españolas.

Los análisis de grupos (*cluster*) y de coordenadas principales llevados a cabo sobre la matriz de similitudes genéticas, para el conjunto de las entradas, revelaron la singularidad de la mayoría de las entradas españolas, comparadas con el conjunto de referencia. Ambos análisis sugirieron además la existencia de dos grupos principales de entradas españolas. Estos dos grupos mostraron diferencias en distribución geográfica y en frecuencias alélicas para varios marcadores, incluyendo un STS ligado al gen *vrs1* que determina el carácter dos-seis carreras.

Un análisis molecular de varianza mostró diferencias genéticas importantes entre la colección nuclear y el conjunto de referencia y también entre los dos grupos principales de entradas españolas.

La distribución geográfica de muchos microsatélites parece estar influenciada por la latitud, altitud o región agroecológica del lugar de recolección. Este hecho sugiere que los patrones de diversidad observados pueden haber sido causados en parte por fuerzas adaptativas.

Se apreció un desequilibrio de ligamiento importante entre los loci estudiados. Este fue aparentemente mayor en las entradas de la colección nuclear que en el conjunto de referencia. Para las entradas españolas, el elevado desequilibrio de ligamiento se mantiene sólo en uno de los dos grupos principales.

Los patrones de diversidad genética encontrados son de uso directo para los mejoradores, pues permiten tomar decisiones informadas sobre la selección de parentales y de estrategias de cruzamiento para las entradas locales españolas.

Résumé

L'étude est basée essentiellement sur la caractérisation moléculaire et l'analyse de la diversité génétique des orges à six rangs de la collection nucléaire espagnole.

Vingt neuf microsatellites réparties sur l'ensemble des chromosomes, ont été choisis afin de caractériser 183 génotypes. Parmi ces derniers nous comptons 152 génotypes espagnols de la collection nucléaire (lignées pures provenant de variétés locales), et 31 génotypes d'autres origines comme témoins: 28 variétés commerciales européennes et 3 cultivars sélectionnés à partir d'un matériel CIMMYT-ICARDA.

Les microsatellites élus pour cette analyse ont permis la détection d'un polymorphisme élevé dans le matériel analysé. Nous avons noté aussi la présence d'allèles spécifiques dans chacun des groupes de génotypes cités ci dessus et particulièrement dans le matériel de la collection nucléaire.

Les analyses des groupes (cluster) et de coordonnées principales réalisées à partir de la matrice de similarité, pour l'ensemble des génotypes, ont révélé la singularité de la majorité des accessions espagnoles par rapport aux variétés du reste de l'Europe. En plus ces deux analyses ont suggéré l'existence de deux groupes principaux dans le matériel espagnol. Ces deux derniers se différencient quant à leur distribution géographique et par la fréquence allélique de divers marqueurs, incluant un STS lié au gène *vrs1*, qui déterminent le type d'épi, 2 ou 6 rangs.

L'analyse moléculaire de variance a révélé l'existence d'une variabilité génétique importante entre les orges espagnoles et européennes ainsi qu'entre les deux principaux groupes espagnols.

La distribution allélique n'est pas aléatoire, elle est conditionnée par certaines variables éco-géographiques. Certains allèles ont montré de fortes associations avec la latitude, l'altitude et la région agro-écologique, ce qui suggère un rôle important des facteurs adaptatifs dans la distribution de la diversité.

D'autre part, l'évaluation du déséquilibre de linkage, dans chacun des groupes cités ci dessus, a montré que celui-ci est beaucoup plus important dans le groupe espagnol qu'au niveau des variétés de référence. De même, nous avons noté que ce taux de déséquilibre de linkage important dans le germplasm espagnol est particulièrement concentré dans l'un des deux sous-groupes. Les patrons de diversité trouvés peuvent être d'une utilité directe pour l'améliorateur, dans la mesure qu'ils permettent un choix judicieux de parents et de stratégies de croisement pour les variétés traditionnelles espagnoles.

Summary

This study focuses on the characterization with microsatellites of the six-row accessions of the Spanish barley core collection (SBCC), and the analysis of the genetic diversity found.

Twenty-nine microsatellites, more or less equally distributed along the seven chromosomes, were selected to analyse 183 barley genotypes: 152 accessions from the SBCC (inbred lines derived from local landraces), and 31 accessions from other origins as checks: 28 cultivars from other European countries (reference set), and 3 accessions from the CIMMYT-ICARDA pool.

A high level of polymorphism was detected by the microsatellites. There was a number of private alleles in all groups of genotypes, notably among the Spanish accessions.

Cluster and principal coordinates analyses on the similarity matrix for the entire set of genotypes confirmed the distinctiveness of a majority of Spanish accessions, compared to other European origins. Both analyses separated most of the SBCC materials from the checks, but also suggested the presence of two rather distinct main groups in the SBCC accessions. These two groups differed in geographic distribution, and in allelic frequencies for several markers, among them an STS linked to the *vrs1* locus (which determines spike type).

A molecular analysis of variance revealed important genetic differences both between Spanish and other European genotypes, and also between the two main SBCC groups.

The geographic distribution of many microsatellites seemed influenced by latitude, altitude, and agroecological region, suggesting adaptation as one of the driving forces that shaped the patterns of genetic diversity observed.

The analysis of linkage disequilibrium (LD) showed apparent larger LD among Spanish accessions than among the reference set. Also, this remarkable LD was mostly evident in only one of the main Spanish groups.

The patterns of genetic diversity found can be of direct use to breeders, as they will allow a judicious choice of parents and crossing strategies for Spanish landraces.