

ACADEMIE DE MONTPELLIER

ENSAM

ECOLE NATIONALE SUPERIEURE AGRONOMIQUE DE MONTPELLIER

THESE

Présentée à l'Ecole National Supérieure Agronomique de Montpellier

pour obtenir le diplôme de doctorat

Formation Doctorale : Ressources Phytogénétiques et Interactions Biologiques

Ecole Doctorale : Biologie Intégrative

**Impact de la gestion paysanne sur la différenciation génétique et
phénotypique des populations traditionnelles de maïs
des Vallées Centrales de Oaxaca au Mexique**

par

Gaël Pressoir

Soutenue publiquement le 12 septembre 2003 devant le jury composé de :

Alain Charcosset

Rapporteur

Joël Cuguen

Rapporteur

Thierry Robert

Examineur

Julien Berthaud

Directeur de thèse

André Charrier

Président

DEDICACE

J'aimerais remercier les agriculteurs de Oaxaca pour leur participation au projet et surtout pour nous avoir donné le matériel de cette thèse.







Photos J. Berthaud

REMERCIEMENTS

J'aimerais tout d'abord remercier le Laboratoire de Biotechnologie Appliquée, ABC, du Centre International d'Amélioration du Maïs et du Blé, CIMMYT, ainsi que l'équipe DYNADIV de l'Institut de Recherche pour le Développement, IRD, qui m'ont accueilli. Mes remerciements vont tout particulièrement à Julien Berthaud qui m'a encadré.

Je remercie tous ceux qui ont participé de près ou de loin aux comités de thèse ou à la correction des épreuves, en particulier : Julien Berthaud, Alain Charcosset, Brigitte Gouesnard, Estelle Jaligot, Claire Neema, Jean-Louis Pham, David Poland et Joëlle Ronfort, " merci ".

A toutes les personnes du Laboratoire de Biotechnologie Appliquée du CIMMYT où ce travail a été réalisé, " merci ".

A Alfonso Aguirre, Mauricio Bellon, Irma Manuel Rosas et Jorge Mendoza sans qui il n'aurait pas été possible de procéder à la collecte des échantillons ; pas de matériel, pas de thèse, donc " merci ".

A Juan et Martin qui ont passé avec moi des mois à planter, récolter, égrener et mesurer, en bref à produire l'ensemble des données agro-morphologiques (près de 15000 plantes mesurées individuellement pour un grand nombre de caractères quantitatifs) ; pas de données, pas de thèse, donc " merci ".

A André Charrier et Thierry Robert qui ont bien voulu faire partie de mon jury de thèse et Alain Charcosset et Joël Cuguen qui ont accepté d'évaluer le présent document, " merci ".

Comment n'oublier personne ! A tous ceux qui de près ou de loin m'ont aidé, " merci ".

SOMMAIRE

I.	INTRODUCTION	2
II.	LA DIFFÉRENCIATION GÉNÉTIQUE ET PHÉNOTYPIQUE	13
III.	MATÉRIEL	24
IV.	<i>Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico</i>	26
V.	<i>Population structure and strong divergent selection shape phenotypic diversification in maize landraces</i>	39
VI.	<i>Effect of kernel color in structuring genetic diversity in maize landraces</i>	54
VII.	Etude du déséquilibre de liaison et de la différenciation entre populations pour des marqueurs disposés le long du gène <i>Shrunken 2</i>	67
VIII.	Perceptions et pratiques paysannes	74
IX.	DISCUSSION ET CONCLUSIONS	81
X.	BIBLIOGRAPHIE	98

Les chapitres de résultats sont en minuscules

I. INTRODUCTION

L'adaptation et l'évolution d'une espèce telle que le maïs interviennent au sein d'un environnement façonné par l'agriculteur. Ces phénomènes se déroulent selon une échelle de temps courte comparée à celle observée dans le cas d'espèces sauvages. Depuis sa domestication le maïs a connu des changements considérables du point de vue morphologique. Les plus anciens vestiges de la culture du maïs, remontant à 4200 ans avant notre ère, ont été trouvés dans les vallées centrales de Oaxaca au Mexique (Benz 2001; Piperno et Flannery 2001), et présentent des différences importantes avec les variétés modernes. La reconstitution de séries chronologiques à partir des rachis découverts lors de fouilles archéologiques suggère que l'adaptation et l'évolution phénotypique vers le maïs actuellement cultivé se sont opérées de manière graduelle (Wilkes 1989). Les maïs échantillonnés dans la région de Oaxaca sont aussi les plus proches, sur le plan phylogénique, de l'ancêtre sauvage du maïs (Matsuoka *et al.* 2002). De plus, c'est une région où la gestion des populations traditionnelles de maïs par les agriculteurs peut être documentée et où très peu de variétés modernes de maïs ont été adoptées (Bellon *et al.* 2002). Il est donc possible d'y analyser la dynamique de l'évolution d'une plante cultivée dans un contexte d'agriculture traditionnelle en fonction des contraintes biologiques, géographiques, sociologiques et culturelles.

Par rapport aux autres espèces cultivées, le maïs présente une importante variabilité, au niveau génétique et, plus encore, phénotypique (Buckler *et al.* 2001; Matsuoka *et al.* 2002). Par contraste, les mécanismes de création et de maintien de cette variabilité ont rarement été décrits et demeurent largement inconnus. Notamment, les rôles joués par l'agriculteur et par la gestion paysanne dans ces processus restent à déterminer.

Il s'agira au cours de cette thèse d'évaluer l'impact de la gestion paysanne sur la diversité génétique et sa dynamique. Nous faisons l'hypothèse que l'agriculteur de par sa gestion des variétés de maïs et de par son processus de décision (gestion de la démographie, rejet ou acceptation d'une population, sélection à l'intérieur de cette population selon des critères qui lui sont propres ou qu'il partage avec d'autres agriculteurs) est le moteur de l'adaptation et de l'évolution du maïs et particulièrement de l'évolution des caractères phénotypiques.

Cette thèse se fixe donc comme objectif de tester cette hypothèse, en identifiant et analysant les facteurs qui contribuent à la dynamique de création et maintien de la variabilité génétique

et tout particulièrement la variabilité phénotypique au sein de populations soumises à la gestion paysanne traditionnelle dans le cadre d'une région du Mexique.

Ces résultats doivent nous permettre de conduire une réflexion sur la différenciation phénotypique et la caractérisation variétale ainsi que sur les concepts de variété, cultivar et d' "unité de gestion de semence", ainsi que sur les stratégies de conservation dynamique et à la ferme et la valorisation des ressources génétiques.

Evolution phénotypique des plantes cultivées

La transformation des formes sauvages vers celles cultivées a été réalisée par l'homme sur de nombreuses espèces. Elle implique un processus de domestication de durée inégale selon les espèces. Dans le cas des céréales ce processus implique la sélection de mutations d'un nombre limité de gènes (impliqués dans l'égrenage, le tallage et la dormance par exemple, Harlan 1975). Du fait de ces mutations la plante cultivée se retrouve dépendante de l'homme pour assurer sa descendance. On peut alors parler de plante domestiquée. La post-domestication correspond quant-à-elle à l'étape suivante où l'homme n'a plus recours à la cueillette ou sélection d'individus au sein du compartiment sauvage. C'est l'agriculteur qui choisit alors les individus reproducteurs. Les plantes cultivées offrent ainsi un modèle très particulier d'évolution. La post-domestication correspond à une évolution récente, voire ayant toujours cours, et surtout à une évolution rapide sous le contrôle de l'agriculteur.

Pernès (1985) présente la domestication comme un laboratoire de l'évolution des caractères adaptatifs : " Les études de génétique des populations se heurtent, dans la mesure où elles tentent de représenter les processus de l'évolution, aux durées de l'expérimentation nécessaire. Il faudrait exercer des sélections significatives pendant des milliers de générations, répétées dans des conditions et sur du matériel variés. Mais quand l'homme tente de transformer héréditairement les plantes pour qu'elles s'adaptent aux techniques culturales qu'il impose, il réalise avec constance une telle expérience. La domestication des céréales est ainsi d'un intérêt exceptionnel : elle constitue un ensemble d'expérience répétées et prolongées de génétique évolutive ". Ces arguments peuvent être repris pour la phase de post-domestication. Distinguer la domestication de l'évolution adaptative post-domestication nous apparaît important du fait que les mécanismes et gènes impliqués dans l'une ou l'autre ne sont pas nécessairement les mêmes. Nous pouvons donc dans les régions ayant maintenu les

formes traditionnelles d'agriculture analyser les résultats de ces multiples expériences d'évolution tout en observant les conditions de leur déroulement.

Le maïs cultivé présente depuis l'apparition des plus anciennes formes domestiquées une longue période d'évolution et adaptation morphologique (Wellhausen *et al.* 1952). La grande variabilité phénotypique du maïs que l'on peut encore observer dans les champs des agriculteurs dans des régions telles les vallées d'altitude au Mexique, nous offre la possibilité d'étudier l'évolution adaptative post-domestication et le rôle de l'agriculteur dans cette diversification.

1.1 Gestion paysanne dans un agrosystème traditionnel

Le paysage agricole et l'agriculture traditionnelle

L'agriculture traditionnelle au Mexique se démarque de façon importante de l'agriculture moderne. (i) Au sein de l'agrosystème traditionnel de maïs au Mexique, de très nombreuses populations coexistent au sein d'un même environnement ou paysage, (ii) les parcelles de culture sont petites et fragmentées et de plus, (iii) elles se jouxtent fréquemment engendrant ainsi d'importants effets de bordure autorisant ainsi les flux de pollen (Bellon et Brush 1994 ; Louette *et al.* 1997 ; Perales 1999).

Si ce paysage agricole diffère de celui observé dans les pays développés, il en va de même des pratiques paysannes. En effet, (iv) les agriculteurs recyclent leurs semences, c'est à dire qu'à la fin de chaque cycle de culture ils sélectionnent les semences qu'ils planteront lors du cycle suivant (Anderson 1947 ; Louette et Smale 2000 ; Nigh 1976 ; Smale *et al.* 1999 ; Wellhausen *et al.* 1952) et, (v) ils échangent des semences et mélangent parfois des graines de diverses provenances (Aguirre Gómez 1999 ; Louette *et al.* 1997).

Nous détaillerons dans les paragraphes suivants certaines de ces caractéristiques qui nous apparaissent importantes dans le cadre de ce travail.

Populations et métapopulations de maïs

Louette *et al.* (1997) ont montré que la gestion de la diversité se faisait par les agriculteurs au niveau de leurs populations plutôt qu'au niveau d'un ensemble de populations partageant des

La notion de variété traditionnelle ou “ landrace ” recouvre un grand nombre d’acceptions (Zeven 1998). Si les variétés modernes, commerciales, sont bien définies à partir de critères d’homogénéité et de stabilité, il n’en est pas de même des variétés traditionnelles. Ces critères ne sont pas forcément interprétés de façon aussi stricte, voire ne sont pas du tout pris en compte par les agriculteurs. Ceci est peut être l’explication de la grande variabilité de ces variétés en regard des variétés modernes, à la fois au niveau intra-population et entre les populations gérées par différents agriculteurs. Pour parler d’un ensemble de populations paysannes partageant des caractères phénotypiques importants du point de vue de l’agriculteur ou de la gestion paysanne, nous préférons l’utilisation du terme “ **morphotype** ” à celui de “ variété ”

Nous reviendrons sur ces concepts en Discussion (page 85)

caractéristiques morphologiques “ saillantes ” (“ morphotypes ”). Notre travail portera donc sur la caractérisation de ces populations de maïs.

Dans leurs travaux portant sur les pratiques culturelles employées à Cuзалapa au Mexique, Louette *et al* (1997) rapportent que les graines provenant des propres champs des agriculteurs sont utilisées pour réensemencer seulement 53% des parcelles. Ceci souligne l'importance des échanges de semences entre agriculteurs. L'agriculteur ne retenant souvent pour le réensemencement qu'un petit nombre d'épis sur une récolte, les lots de graines utilisés lors du réensemencement sont de taille réduite. Ces lots sont fréquemment hétérogènes et d'origines multiples du fait des échanges et mélanges effectués entre les agriculteurs ou de l'achat de semences nouvelles sur les marchés (Rice *et al.* 1998). Par ailleurs, les variations dans les dates de semis et les effets de bordure entre les parcelles adjacentes permettent l'intervention de flux de pollen (Bellon et Brush 1994; Louette 2000; Louette *et al.* 1997). De plus chaque agriculteur possède ses propres critères de sélection (Louette et Smale 2000).

Dans l'ensemble, ces données font apparaître que la structuration des populations traditionnelles de maïs, que ce soit à l'échelle d'un village ou à l'échelle de la région, peut être assimilée à une métapopulation (Louette 2000). En effet, une métapopulation se définit comme un groupe de sous-populations interconnectées par des flux de gènes et soumis à des extinctions et recolonisations locales (Slatkin et Wade 1978).

L'objectif étant ici de comprendre la dynamique de la diversité dans le cadre de l'agriculture traditionnelle et donc de comprendre l'impact de la gestion paysanne sur le maintien et la création de cette diversité, il apparaît nécessaire, de comprendre et décrire le fonctionnement des métapopulations de maïs à la fois à l'échelle d'une population, de l'ensemble des populations d'un agriculteur, du morphotype tel qu'il est perçu et appréhendé par l'agriculteur (couleur des grains par exemple), du village et enfin de la région.

1.2 Le cadre de l'évolution contemporaine ou post-domestication du maïs

Du fait de leur gestion selon des pratiques culturelles traditionnelles, les populations de maïs des Vallées Centrales de Oaxaca sont confrontées à des événements d'extinction, de fondation et de migration. Ces événements peuvent résulter de flux polliniques, ou encore des décisions prises ou d'échanges de semences réalisés par l'agriculteur. L'action de celui-ci peut s'exercer à différents niveaux de la gestion de sa population. Il peut choisir de poursuivre ou

d'interrompre la culture d'une population donnée, ou d'adopter une nouvelle population éventuellement différente du point de vue morphologique. Le lot de graines provenant de la récolte de l'année, utilisé pour procéder au réensemencement, peut être mélangé aux semences fournies par un autre exploitant. Enfin, la superficie ensemencée peut varier. L'ensemble de ces décisions influe sur la démographie de la population de maïs cultivée, à la fois qualitativement et quantitativement. En second lieu, l'agriculteur est amené à opérer des choix de sélection, en retenant les épis constituant le lot de semences qui donnera la génération suivante. Pour l'essentiel, cette sélection est effectuée postérieurement à la récolte (Aguirre Gómez 1999). Les critères de sélection des épis sont généralement leur taille, leur vigueur et leur état sanitaire, ainsi que l'aspect (couleur, forme) du grain et du rachis (Aguirre Gómez 1999). Ces considérations (extinction, colonisation ou fondation, migration, et sélection possiblement résultantes de l'action et de la gestion paysanne) nous ont amené à adopter des méthodes de génétique des populations.

Les différents processus et forces évolutives en milieu paysan, qu'ils soient d'ordre démographique ou ayant trait à la sélection qu'elle soit pratiquée consciemment ou non par l'agriculteur, auront possiblement des conséquences distinctes sur la variabilité des différentes classes de marqueurs ou des caractères morphologiques. Le présent travail, envisageant de prendre en compte l'ensemble de ces facteurs, se positionne par conséquent à l'interface entre différentes échelles d'investigation. (Figure 1.1)

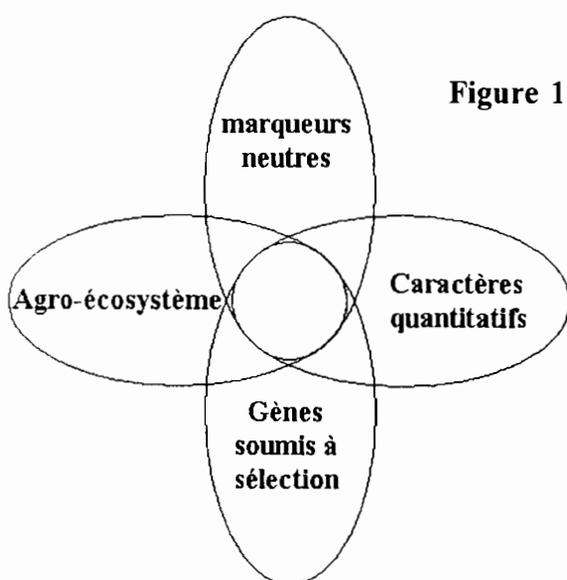


Figure 1.1 – Jeux de données qui seront utilisés pour évaluer la diversité à l'intérieur et entre les populations paysannes.

C'est trop souvent "le point de vue qui fait l'objet" (Ferdinand De Saussure) voire parfois "le point de vue qui est l'objet" (Luis J. Prieto). La diversité est souvent confondue, à tort, avec la mesure qui en est faite, celle-ci étant fonction du parti-pris méthodologique initial.

Différents jeux de données sont disponibles pour l'évaluation de la diversité génétique et l'analyse de ses composantes : les informations issues de l'étude des pedigrees, la caractérisation agro-morphologique, ou encore l'étude des marqueurs moléculaires, que ceux-ci soient neutres ou donnent prise à la sélection. L'utilisation isolée d'un de ces jeux de données fournit une mesure de la variabilité qui découle des événements démographiques et sélectifs qui sont propres aux marqueurs ou caractères considérés. Seule l'utilisation d'une combinaison de ces différentes sources d'informations est à même de nous offrir un aperçu global (sinon exhaustif) des mécanismes intervenant dans le maintien et la création de la variabilité génétique et phénotypique, de la dynamique de la diversité, au sein du système étudié.

1.3 Plan de la thèse

Ce travail se propose de prendre en compte les pratiques paysannes et la diversité phénotypique chez le maïs. En réalisant la synthèse des informations relatives aux processus démographiques et sélectifs au sein des populations étudiées, nous allons tenter de déterminer (i) comment les agriculteurs gèrent et exploitent la variabilité génétique et phénotypique, (ii) le rôle de la sélection paysanne dans les processus d'adaptation du maïs (iii) si le processus d'adaptation et d'évolution phénotypique du maïs est encore en cours. Pour finir, nous proposerons des pistes de travail susceptibles d'apporter des renseignements complémentaires sur l'origine et les bases moléculaires de la variabilité sur laquelle repose l'évolution phénotypique post-domestication du maïs.

Organisation de la variabilité génétique au sein des populations de maïs à grains blancs

Les maïs à grains blancs sont cultivés, à quelques rares exceptions près, par l'ensemble des agriculteurs et correspondent à leur population principale (c'est à dire celle qui couvre l'essentiel de la surface cultivée par chaque agriculteur, Smale *et al.* 1999). Nous allons dans la première partie de la thèse étudier la l'organisation de la variabilité génétique au sein de ces populations. Dans une autre partie, nous nous intéresserons également aux maïs à grains de couleur cultivés par un nombre réduit d'agriculteurs sur des surfaces moindres.

Association s'entend d'une corrélation étroite entre le phénotype et un génotype donné (cartographie d'association : Buckler et Thornsberry 2002 ; Remington *et al.* 2001 ; Thornsberry *et al.* 2001). Un marqueur neutre est donc un marqueur présentant une association suffisamment faible avec le phénotype pour être considéré comme neutre évolutivement.

L'association des marqueurs moléculaires (isozymes, RFLP, AFLP, microsatellites) avec le phénotype est suffisamment faible pour que la dynamique de leur évolution puisse être considérée, sur une échelle de temps réduite, comme indépendante de la sélection. Ainsi, l'étude de ces marqueurs apporte essentiellement des informations quant aux forces démographiques (migration, dérive) auxquelles sont soumises les populations. A de rares exceptions près (Raybould *et al.* 1996), les résultats obtenus à partir des différents types de marqueurs montrent de très bonnes corrélations. Cette tendance a été particulièrement vérifiée dans le cas du maïs, au cours d'études portant sur des lignées (Pejic *et al.* 1998) ou sur des populations (Dubreuil et Charcosset 1998). Des processus évolutifs similaires semblent donc être à l'œuvre à l'intérieur de ces différentes catégories de marqueurs.

Cette neutralité permet l'étude de la diversité, de la structuration et des flux de gènes à l'intérieur de chacune des populations étudiées et entre elles. Aussi, une première partie de notre étude sera consacrée aux conséquences de la gestion paysanne sur la structuration de la diversité de ce type de marqueurs au sein des populations de maïs cultivées par les agriculteurs. Le chapitre IV (*Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico*) s'intéressera plus particulièrement à la structuration intra- et inter-population pour ce type de marqueurs.

Entre les populations, les flux de gènes s'effectuent par le biais des graines (échangées entre les agriculteurs) ou du pollen (dispersion par le vent). Ces deux voies de dissémination procédant selon des modalités très distinctes, il est très probable que la structuration de la diversité ne soit pas la même pour des marqueurs à hérédité biparentale ou pour ceux à hérédité maternelle, tels que des marqueurs chloroplastiques (Birky *et al.* 1989). Cette distinction est d'autant plus cruciale que les échanges de graines, par rapport à la dispersion *via* le pollen, autorisent des flux de gènes entre des populations séparées par des distances très importantes. Il apparaît donc nécessaire, dans le cadre de cette étude, d'estimer les parts respectives de l'échange de semences et de la dissémination pollinique dans la structuration de la diversité génétique, en utilisant une combinaison de marqueurs nucléaires et de marqueurs cytoplasmiques (Ennos 1994). Par ailleurs, il sera intéressant de confronter les données de gestion des populations cultivées et d'échanges de graines, fournies par les agriculteurs, avec la structuration de la diversité ainsi obtenue. L'ensemble des informations générées devra nous aider à déterminer si la gestion paysanne conduit à un isolement génétique des populations cultivées les unes par rapport aux autres et, le cas échéant, à quelle échelle (région, village, lot de semences) ce phénomène est perceptible.

Les réponses apportées à cette première question nous permettront éventuellement de caractériser les mécanismes de différenciation intervenant entre les populations détenus par différents agriculteurs. Nous nous attacherons également à déterminer si une population paysanne correspond à un mélange ou si elle constitue une population homogène. C'est à dire de déterminer s'il existe une structuration interne aux populations paysannes et de caractériser cette structure.

Différenciation phénotypique au sein des populations de maïs à grains blancs

Il n'est pas évident de distinguer si des flux de gènes limités entre des populations sont déterminants dans l'émergence de la différenciation phénotypique, ou si celle-ci peut résulter de processus de sélection divergente dans un contexte où les flux de gènes demeurent importants (Latta et McKay 2002; Merilä et Crnokrak 2001) Autrement dit, la caractérisation d'une différenciation entre des populations au niveau de marqueurs moléculaires reflète-elle celle qui est observée au niveau des caractères phénotypiques ?

Au cours du deuxième volet de cette thèse, nous chercherons à savoir si la différenciation phénotypique entre les populations de maïs est imputable à l'isolement de ces populations, ou si elle résulte de la sélection effectuée par l'agriculteur. Cette question prend tout son poids dans le cadre de la présente étude, où il est envisageable que la sélection paysanne influe sur le maintien et sur la création de la variabilité phénotypique entre les variétés traditionnelles de maïs. Comme il a été évoqué plus haut, les marqueurs moléculaires (marqueurs " neutres ") ne fournissent pas d'information directe sur les forces sélectives s'exerçant sur les populations et sur les divergences génétiques et phénotypiques qui en résultent. Par conséquent, la structuration de la diversité, telle qu'elle est définie en se basant sur des marqueurs considérés comme neutres, ne reflète pas nécessairement la structuration obtenue à partir de caractères quantitatifs donnant prise à la sélection (Lynch *et al.* 1999; McKay et Latta 2002; Podolsky et Holtsford 1995). En résumé, les données de structuration de la diversité génétique obtenues en utilisant les marqueurs neutres sont assimilables à la structuration qui serait obtenue en l'absence de sélection. Dès lors, la divergence qui serait observée entre cette estimation et les valeurs de diversité génétique résultant de la mesure des caractères quantitatifs permettrait de détecter des effets attribuables à la sélection paysanne au niveau de la variabilité phénotypique. De plus, les mécanismes démographiques et sélectifs à l'œuvre chez les populations étudiées se reflètent dans des changements, d'une population à l'autre, de la matrice des variances et covariances génétiques correspondant aux caractères quantitatifs

(Roff 2000). La sélection divergente est susceptible d'affecter la proportionnalité entre ces matrices alors que dérive et migrations ne l'affecte pas (Roff 2000). L'application d'une sélection divergente affectant la proportionnalité entre ces matrices, au contraire des événements de dérive et de migration, il nous sera possible de caractériser les conséquences de l'action paysanne sur la différenciation phénotypique des populations de maïs. Ces différents points feront l'objet du chapitre V (*Population structure and strong divergent selection shape phenotypic diversification in maize landraces*).

Différenciation génétique et phénotypique des populations de maïs à grains noirs

Outre les populations de maïs à grain blanc traitées dans les chapitres IV et V, et qui occupent l'essentiel des surfaces dédiées à cette céréale, les agriculteurs cultivent également d'autres types de maïs, dont ceux produisant des grains noirs. Cette culture est gérée selon des principes très différents de ceux gouvernant la culture des maïs à grains blancs, aussi est-il concevable que cette différence se traduise au niveau de l'organisation de la variabilité à l'intérieur et entre ces deux compartiments. L'intérêt suscité par la culture des maïs à grains noirs étant, entre autres raisons, motivé par la précocité de floraison que les agriculteurs lui attribuent (Bellon *et al.* 2002), nous examinerons le fondement de cette association entre couleur du grain et précocité.

Selon des approches similaires à celles énoncées pour les chapitres IV et V, nous chercherons, respectivement, à caractériser la structuration de la diversité génétique et à détecter d'éventuels mécanismes de différenciation phénotypique au sein des populations de maïs à grain noir et blanc.

1. Ces populations (à grains noirs) obéissent-elles aux mêmes processus démographiques et présentent-elles une organisation de la diversité génétique neutre semblables à celles des maïs à grains blancs ?

La problématique est donc similaire à ce qui a été mentionné pour le chapitre IV

2. La différenciation phénotypique au sein de ces populations à grains noirs répond-elle aux mêmes mécanismes que celle observée chez les maïs à grains blancs ? Quelle est la relation entre les deux ?

Les résultats obtenus seront comparés à ceux générés lors de nos études conduites sur les maïs à grain blanc, et constitueront le chapitre VI (*Effect of kernel color in structuring genetic diversity in maize landraces*).

Déséquilibre de liaison à l'intérieur, et organisation spatiale de la variabilité génétique, des gènes possiblement associés à la variation phénotypique

Des études utilisant des échantillonnages couvrant l'essentiel de l'aire de répartition du maïs ont mis en évidence une décroissance rapide du déséquilibre de liaison à l'intérieur de nombreux gènes, intervenant sur des distances de quelques centaines de paires de bases (Remington *et al.* 2001; Tenaillon *et al.* 2002; Tenaillon *et al.* 2001; Thornsberry *et al.* 2001 ; Wang *et al.* 1999). Nous avons donc tenté d'établir si un tel phénomène est observable à l'échelle, considérablement plus réduite, des Vallées centrales de Oaxaca, et s'il est d'ampleur comparable avec ce qui a été observé lors de ces travaux.

La structuration de la diversité pour des gènes ou locus devrait refléter la sélection à laquelle ils sont soumis. Par conséquent, elle est supposée se démarquer du profil de la diversité obtenu en se basant sur les marqueurs neutres, et nous fournir la matière à une comparaison. Néanmoins, dans la plupart des cas, les caractères quantitatifs sont dépendants d'un nombre élevés de locus distincts. Un caractère agromorphologique est donc susceptible de résulter de l'effet d'un certain nombre de QTLs (McKay et Latta 2002). La prise en compte de cette dernière remarque dans le contexte de notre étude nous amène à nous demander si l'évaluation et la structuration de la diversité aux QTLs d'intérêt agronomique est plus fidèlement représentée par la variation observée au niveau des marqueurs neutres ou au niveau phénotypique. Les implications possibles de nos conclusions vis-à-vis des techniques d'évaluation de la diversité génétique, notamment au niveau des gènes ou locus associés à la variation phénotypique, seront ensuite discutées.

Nous étudierons au sein de nos populations la variation observée au niveau d'un gène intervenant dans le développement du grain (*shrunk2*). Nous présenterons nos conclusions concernant les deux derniers paragraphes dans le chapitre VII (Etude du déséquilibre de liaison et de la différenciation entre populations pour des marqueurs disposés le long du gène *Shrunken 2*).

Pratiques culturelles paysannes et différenciation génétique et phénotypique des populations de maïs

Enfin, nous mettrons à profit le chapitre VIII (Perceptions et pratiques paysannes) pour faire le bilan de l'apport de cette thèse vis-à-vis de la connaissance des pratiques culturelles paysannes. Nous confronterons les résultats obtenus, aux pratiques paysannes d'échanges et

mélanges de semences, à la perception qu'ils ont de la sélection, et aux différentes modalités de gestion et de sélection observées aux cours des enquêtes.

Le choix des outils pour mener à bien cette étude

De par leurs influences multiples sur le maintien des variétés et populations traditionnelles, les choix effectués par l'agriculteur sont susceptibles d'agir sur la répartition spatiale de la diversité. L'étude de la diversité génétique peut être définie comme la mesure de la variation génétique entre des individus ou parmi (variation intra-groupe) et entre (variation inter-groupes) des groupes d'individus ou des populations. Toutefois, en fonction de la méthode qui est choisie pour estimer la diversité génétique, la mesure obtenue peut revêtir des significations bien distinctes. Il importe pour cette raison de se munir d'outils d'évaluation adaptés au problème posé (outils décrits dans le chapitre suivant, chapitre II).

II. LA DIFFÉRENCIATION GÉNÉTIQUE ET PHÉNOTYPIQUE

Les différents outils disponibles pour analyser la différenciation génétique ont fait l'objet de nombreux travaux mettant en jeu la validation expérimentale des modèles théoriques. Une comparaison détaillée de ces outils doit nous permettre d'établir leur pertinence relativement à l'étude que nous nous proposons de mener à bien. "La théorie sans la pratique est impuissante, la pratique sans la théorie est aveugle" (Emmanuel Kant suivi par Henri Bergson).

2.1 Différenciation des populations pour des marqueurs neutres

Les statistiques F de Wright

Chaque agriculteur gère une ou plusieurs populations de maïs. Chacune de ces populations évolue dans des conditions qui lui sont propres et peut acquérir une certaine originalité génétique, caractérisée par des fréquences alléliques, plus ou moins différentes de celles des populations voisines. On assiste ainsi à une différenciation des populations. Cependant le plus souvent elles ne sont pas totalement isolées les unes des autres et elles continuent à échanger des individus par migration. Ce flux migratoire contribue à un brassage génétique qui tend à homogénéiser les populations et à limiter leur différenciation. Des polymorphismes neutres vis-à-vis des forces sélectives et évolutives refléteront l'effet des migrations et de la dérive des populations sur les fréquences alléliques. Nous serons amenés à examiner les conséquences de ces phénomènes sur la différenciation génétique au sein de la région des vallées centrales de Oaxaca.

Il est difficile de réaliser une mesure physique de la dispersion par graines ou pollen de façon correcte au sein de populations parce que l'estimation de cette dispersion est souvent impossible dans des systèmes non-expérimentaux.

Le F_{st} ou indice de fixation, est une mesure de la réduction relative de l'hétérozygotie de ces populations. L'indice de fixation doit donc être compris comme un index de déviation par rapport aux proportions de Hardy-Weinberg. La fréquence attendue des hétérozygotes sous Hardy-Weinberg est alors $(1 - \sum p_i^2)$ ou encore sa valeur non biaisé (Nei 1987) $H_e = 2n(1 - \sum p_i^2)/(2n-1)$, où p est la fréquence d'un allèle donné. Les statistiques F ont été précisées par Wright (1969) qui a proposé une généralisation permettant de caractériser la structure des populations. Nei (1977) a montré comment calculer les statistiques F pour un nombre

arbitraire d'allèles, car elles peuvent être exprimées simplement en fonction des hétérozygoties observées et attendues :

H_o : Hétérozygotie observée en moyenne par individu, sur l'ensemble des sous-populations.

H_S : Hétérozygotie attendue par individu pour une sous-population, en la supposant à aux proportions génotypiques de Hardy-Weinberg.

H_T : Hétérozygotie attendue par individu, en supposant la population globale aux proportions génotypiques de Hardy-Weinberg.

$$\bar{F}_{IS} = \frac{H_S - H_o}{H_S} \quad F_{IT} = \frac{H_T - H_o}{H_T} \quad F_{ST} = \frac{H_T - H_S}{H_T}$$

La statistique F_{st} est également le rapport de la variance observée des fréquences alléliques sur la variance attendue dans une population idéale (dont les caractéristiques comprennent la reproduction aléatoire, aucune sélection et un taux de mutations négligeable).

Le F_{st} correspond à la réduction d'hétérozygotie dans les sous-populations, liée aux différences de fréquences alléliques moyennes. Si toutes les sous-populations ont la même fréquence allélique et sont à l'équilibre, F_{st} est nul. Dans le cas contraire, qui implique que H_T soit plus grand que H_S , le F_{st} sera positif. Ceci indique que les sous-populations ont des fréquences alléliques moyennes différentes. Pour des marqueurs neutres : le F_{st} reflète l'action conjuguée de la dérive, qui différencie les populations, et de la migration, qui tend à les homogénéiser. F_{st} est ainsi un indicateur de la cohésion de l'ensemble des sous-populations considérées. Le F_{is} traduit un déficit (valeur négative) ou un excès (valeur positive) d'homozygotes dans les populations considérées. Et donc le F_{is} est un index de déviation par rapport aux proportions de Hardy-Weinberg à l'intérieur des sous-populations considérées. Ou encore, la statistique F_{is} , comme définie par Wright, est la corrélation moyenne des deux gènes d'un individu par rapport à deux gènes pris au hasard dans une des subdivisions de la population. On peut penser dans un premier temps que F_{is} estime l'effet dû aux croisements non panmictiques dans les sous populations Néanmoins, l'effet Wahlund peut être source de confusion lorsqu'un expérimentateur, croyant étudier une population unique, analyse en fait les données relatives à un échantillon couvrant deux sous-populations qu'il n'a pas distinguées. Il peut alors observer un déficit en hétérozygotes (F_{is} positif) qu'il interprétera par un écart à la panmixie, alors qu'il est dû à la différence de fréquences entre les deux sous-populations. Un F_{is} positif traduit donc soit un écart à la panmixie (croisements entre

apparentés ou homogamie) ou encore un effet Wahlund. Le F_{it} traduit lui la réduction d'hétérozygotie globale entre l'individu et la population globale.

D'une manière générale, les statistiques F sont reliées les unes aux autres par la relation (Wright 1969) :

$$(1 - F_{IT}) = (1 - \overline{F_{IS}})(1 - F_{ST})$$

Cockerham (1969, 1973) a montré que les corrélations entre gènes, similaires aux statistiques F pouvaient être estimées par une analyse de variance des fréquences alléliques. Une telle analyse consiste à décomposer la variabilité génétique totale en diverses composantes de variances qui expriment la proportion de la variance totale attribuée à différents niveaux de subdivision de population, soit : aux différences entre gènes à l'intérieur des individus (σ_w^2), aux différences entre individus à l'intérieur des dèmes (σ_b^2), et enfin aux différences entre dèmes (σ_a^2). Cockerham a aussi montré que les statistiques F pouvaient être estimées par les relations :

$$F_{IS} = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2} ; F_{ST} = \frac{\sigma_a^2}{\sigma_w^2 + \sigma_b^2 + \sigma_a^2} ; F_{IT} = \frac{\sigma_b^2 + \sigma_w^2}{\sigma_w^2 + \sigma_b^2 + \sigma_a^2}$$

Ces estimateurs diffèrent quelque peu des estimateurs de Nei à partir des hétérozygoties observées et attendues, et sont relativement moins biaisés, surtout lorsque le nombre de dèmes et la taille des échantillons sont petits. Nous différencierons ce mode de calcul du précédent en nous référant pour ce dernier à θ ($F_{is}=f$, $F_{st}=\theta$ et $F_{it}=F$, Weir et Cockerham 1984). C'est cette dernière méthode qui sera utilisée dans ce document.

Si toutes les sous-populations sont aux proportions génotypiques de Hardy-Weinberg on a $F_{is}=0$, et donc $F_{st}=F_{it}$. Si toutes les populations sont aux proportions génotypiques de Hardy-Weinberg et ont les mêmes fréquences alléliques, F_{is} et F_{st} sont nuls et F_{it} aussi. La division en sous-populations n'existe plus en fait et la population globale est aux proportions génotypiques de Hardy-Weinberg.

Nous avons choisi de travailler avec les statistiques F de Wright parce que (i) dans des espèces démographiquement comparable le F_{st} est le paramètre génétique le mieux corrélé avec l'aptitude à la dispersion (la corrélation entre le F_{st} et l'aptitude à la dispersion au sein d'espèces animales de tailles efficaces comparable est de 0.72, Bohonak 1999). Pour des

marqueurs neutres, il reflète donc bien l'isolement, les migrations et la dérive au sein de l'ensemble considéré. (ii) Le F_{st} de Wright peut être estimé pour les différentes classes de marqueurs mais aussi pour les caractères quantitatifs (Lynch *et al.* 1999) et permettre ainsi la comparaison des différentes mesures de la diversité génétique (voir paragraphe 2.2).

Calcul du F_{st} pour des marqueurs différant par leur mode de mutation

Les microsatellites sont des marqueurs, d'hérédité biparentale, caractérisés par un polymorphisme du nombre de répétition d'un motif de quelques nucléotides, et sont couramment utilisés pour étudier les patrons de différenciation génétique. De part leur mode de mutation préférentiellement par pas, ces marqueurs sont porteurs de deux types d'information : l'identité allélique et la taille allélique. La différence de taille entre deux allèles dépend du nombre de mutations. Différentes mesures de différenciation génétique exploitent l'une ou l'autre de ces informations. Le F_{st} ou θ ($F_{st}=\theta$, Weir et Cockerham 1984) se basent sur l'identité allélique, alors que R_{st} et $(\delta\mu)^2$ se basent sur les tailles des allèles (tous des analogues du F_{st}). R_{st} et $(\delta\mu)^2$ ne seront pas utilisés parce qu'ils sont inappropriés pour des populations faiblement différenciées (Goldstein et Clark 1995 ; Slatkin 1995) et, de plus, la plupart des marqueurs microsatellites utilisés, et plus généralement les microsatellites du maïs, comportent à la fois des mutations pas à pas et des mutations en dehors du motif répété (Matsuoka *et al.* 2002a).

Calcul et comparaison du F_{st} des marqueurs différant par leur mode d'hérédité

Le F_{st} permet de rendre compte de la part de la variation inter-population sur la variation totale et donc de rendre compte de la différenciation entre populations pour différentes classes de marqueurs mais aussi nous le verrons plus tard dans l'exposé pour des caractères quantitatifs. Pour des marqueurs neutres si le taux de mutation est largement inférieur au taux de migration, le F_{st} reflète essentiellement le nombre de migrants ($N_e m$) et fournit une mesure indirecte des flux de gènes au sein du système étudié. Mais si le F_{st} reflète $N_e m$, la formule $F_{st}=1/(1+4N_e m)$, pour un marqueur d'hérédité biparentale, n'est pas exact surtout si les taux de migration et les effectifs efficaces des populations ne sont pas répartis de façon homogène (Whitlock et McCauley 1999). En effet la formule $F_{st}=1/(1+4N_e m)$ suppose un modèle en île idéal. C'est à dire, que (i) chaque population est composée du même nombre d'individus, N , et échange des gènes avec le même taux m , avec chacune des autres populations (ii) le système est à l'équilibre et les différents paramètres ne varient pas au cours du temps et enfin

(iii) la reproduction est panmictique au sein de chaque population et les générations sont non chevauchantes. Pour ces raisons, si une estimation du F_{st} (et donc de la différenciation des populations) est donnée, nous ne donnerons pas d'estimation de $N_e m$.

Les flux de gènes peuvent être soit par voie pollinique ou encore par voie de graines. Dans le cas d'une plante cultivée dans le cadre de l'agriculture traditionnelle, les flux de graines sont sous le contrôle direct de l'agriculteur. Cet agriculteur peut favoriser ou non les flux sur de grandes distances. Les choix faits par l'agriculteur (échange ou absence d'échange et distance des échanges) sont alors susceptibles d'influer sur la structuration des populations. Ennos (1994) a montré que l'on peut estimer le rapport des flux par voie de graines par rapport aux flux polliniques à partir du F_{st} mesurée à partir de marqueurs d'hérédité nucléaire et à celui mesurée à partir de marqueurs cytoplasmiques.

$$r = \frac{\left(\frac{1}{F_{STn}}\right)(1 + F_{IS}) - 2\left(\frac{1}{F_{STc}}\right)}{\left(\frac{1}{F_{STc}}\right)}$$

où F_{STn} est le F_{st} calculé pour un marqueur à hérédité biparentale et F_{STc} est celui calculé pour un marqueur cytoplasmique.

Le résultat obtenu en utilisant le modèle d'Ennos (1994) peut être biaisée si le système s'écarte d'un modèle en île idéal du fait de flux inégalement répartis dans la métapopulation, du fait de l'isolement par la distance ou d'autres facteurs (Hu et Ennos 1999 ; Oddou-Muratorio *et al.* 2001). Il fournit néanmoins une référence permettant la comparaison de la différenciation entre populations obtenue pour ces différentes classes de marqueurs. Ces deux types de marqueurs (hérédité biparentale et maternelle) sont donc utilisés dans le présent travail.

Isolement par la distance

La différenciation entre populations reflète possiblement plus qu'une simple balance de la dérive et des migrations. En effet, nous avons vu précédemment que les conditions du modèle en île peuvent ne pas être vérifiées et que l'une des causes possibles peut être l'isolement par la distance. On peut s'attendre, si les migrations se font de proche en proche, à ce que les dèmes soient d'autant plus proches génétiquement qu'ils sont proches géographiquement.

Rousset (1997) a montré que, sous le modèle d'isolement par la distance, on s'attend à une relation linéaire entre $F_{st}/(1-F_{st})$ et $\ln(\text{Distance géographique})$. La corrélation entre une matrice de distance génétique et spatiales (entre $F_{st}/(1-F_{st})$ et $\ln(\text{Dist})$, par exemple) peut être testée grâce à un test de Mantel (Sokal et Rohlf 1995).

Ecart à la panmixie au sein d'une " population "

De nombreux auteurs (Barcaccia *et al.* 2003; Brown et Allard 1970; Dubreuil et Charcosset 1998; Garnier-Géré 1992; Kahler *et al.* 1986; Lefort-Buson *et al.* 1991; Salanoubat et Pernes 1986) ont rapporté avoir observé un excès d'homozygotes dans les populations en pollinisation ouverte de maïs. Ce, indistinctement de la famille de marqueurs considérés (isozymes, microsatellites et RFLP). Nous avons fait remarquer précédemment que cette valeur de F_{is} supérieur à zéro peut être due à un écart à la panmixie (homogamie ou croisement consanguin) ou à un effet Wahlund. Si nous sommes amenés à observer une semblable déviation par rapport aux proportions de Hardy-Weinberg, nous testerons laquelle de ces trois hypothèses explique cet écart. Nous verrons comment distinguer ces trois cas de figures dans les deux paragraphes qui suivent.

Homogamie

Quand les individus qui s'accouplent ont un certain degré de ressemblance, on parle alors d'homogamie. Inversement si le choix se fait pour un partenaire dissemblable on parle d'hétérogamie. Les termes de "croisements positivement ou négativement assortis", plus généraux, sont aussi utilisés.

En accord avec cette définition, un test simple de l'homogamie peut être mis en place. On s'attend, dans le cas d'homogamie, à une corrélation négative entre la dissemblance phénotypique et un coefficient d'apparentement. Les méthodes développées pour tester l'autocorrélation spatiale (Hardy et Vekemans 1999) peuvent donc être utilisées. Nous testerons donc la corrélation entre le coefficient d'apparentement de Wright, ρ_{ij} , entre paires d'individus d'une même population (ressemblance pour des marqueurs moléculaires) à la dissemblance phénotypique pour le caractère susceptible d'expliquer l'homogamie.

Croisement entre individus apparentés ou effet Wahlund

Dans le cas où l'homogamie n'expliquerait pas la déviation aux attendus d'Hardy-Weinberg, les deux autres hypothèses doivent être testées. (i) Quand les individus qui s'accouplent ont un certain degré de parenté, on parle alors de consanguinité. Elle peut être induite par des contraintes spatiales provoquant une consanguinité de position entre individus géographiquement proches (par exemple l'autofécondation – on est plus proche de soi-même). (ii) Ce que l'on a cru être une population unique, est en fait un échantillon couvrant deux ou plusieurs sous-populations que l'on n'a pas distinguées.

Overall et Nichols (2001) ont montré que l'on peut distinguer ces deux cas de figures. À partir des probabilités conditionnelles de croisements entre apparentés (l'autofécondation par exemple) ou alors d'un effet Wahlund (différence de fréquence entre les différentes composantes), une fonction de vraisemblance est alors calculée pour départager les deux hypothèses.

2.2 Différenciation phénotypique

Comparaison de la structuration entre marqueurs et caractères quantitatifs

La différenciation entre des populations peut être estimée au moyen de marqueurs neutres ou de caractères quantitatifs. La validité de la comparaison qui peut être établie entre ces deux évaluations a fait l'objet d'un débat opposant Lewontin à Felsenstein et Rogers (Felsenstein 1986; Lewontin 1984; Rogers 1986). Mais, comme suggéré par Felsenstein et Rogers (Felsenstein 1986; Rogers 1986), cette comparaison permet de dégager des profils de différenciation imputables à la sélection au niveau de caractères quantitatifs (McKay et Latta 2002 pour une revue).

Comme mentionné précédemment, le F_{st} de Wright peut être estimé pour les différentes classes de marqueurs mais aussi pour les caractères quantitatifs (Lynch *et al.* 1999) et permettre ainsi la comparaison des différentes mesures de la diversité génétique. Nous utiliserons le terme Q_{st} pour distinguer le F_{st} calculé pour des caractères quantitatifs de celui calculé pour des marqueurs moléculaires. Une propriété intéressante de cette valeur est que, dans le cas de neutralité du caractère et des marqueurs considérés au sein des populations étudiées, on obtient théoriquement une égalité entre les deux mesures : $Q_{st} = F_{st}$ (Lande 1992; Whitlock 1999).

Si on suppose que l'effet des allèles à chaque locus est additif on a (Wright 1951) :

$$\begin{aligned}\sigma_{st}^2 &= 2 F_{st} \sigma_a^2 \\ \sigma_{is}^2 &= (1 + F_{it} - 2 F_{st}) \sigma_a^2 = (1 + F_{is})(1 - F_{st}) \sigma_a^2\end{aligned}$$

On peut en déduire le F_{st} d'un caractère quantitatif (Q_{st}):

$$Q_{st} = \frac{\sigma_{st}^2(1 + F_{is})}{2\sigma_{is}^2 + \sigma_{st}^2(1 + F_{is})}$$

Une telle analyse consiste à estimer la proportion de la variance génétique totale attribuée à différents niveaux de subdivision de populations, soit : la variance génétique, entre individus, intra-populations des valeurs reproductives (σ_{is}^2), la variance génétique inter-populations des valeurs reproductives (σ_{st}^2). En cas de neutralité du caractère étudié, et dans le modèle additif, la valeur attendue de ce F_{st} est la même que celles de marqueurs neutres (Lande, 1992, et voir Yang *et al.*, 1996, pour l'évidence d'une bonne corrélation).

On aura, par exemple, pour des familles de demi-frères, et en admettant que la variance entre famille de demi-frères estime un quart de la variance additive (Falconer and Mackay 1996) :

$$\begin{aligned}\sigma_{st}^2 &= \sigma_{pop}^2 \\ \text{et, } \sigma_{is}^2 &= 4\sigma_{op}^2\end{aligned}$$

où σ_{pop}^2 est la variance entre populations et σ_{op}^2 la variance entre famille de demi-frères

Hendry (2002) fait la remarque que, dans les cas où un faible taux de migration tend vers l'ordre de grandeur du taux de mutation, l'égalité attendue entre le F_{st} et le Q_{st} dans une situation de neutralité n'est pas forcément vérifiée. Pour cette raison, nous avons choisi d'examiner la diversité au niveau de marqueurs microsatellites tri- et tétra-nucléotidiques pour lesquels le taux de mutation est considérablement plus faible que celui de marqueurs dinucléotidiques chez le maïs ($\mu \ll 5.1 \times 10^{-5}$ pour des tri- et tétra-nucléotidiques contre 7.7×10^{-4} pour de dinucléotidiques, Vigouroux *et al.* 2002). Ainsi, nous entendons nous placer dans une situation où le taux de migration est largement supérieur au taux de mutation, afin de ne pas biaiser l'égalité attendue.

Dans le modèle additif, l'épistasie conduit à une sous-estimation du Q_{st} (Whitlock 1999). De ce fait, une valeur du Q_{st} supérieure au F_{st} sera indicative de l'existence d'un processus de divergence morphologique due à la sélection. De même, la dominance, moyennée sur

l'ensemble des locus, ne modifiera vraisemblablement pas l'égalité attendue $F_{st} = Q_{st}$ (Whitlock 1999).

La comparaison du F_{st} calculé pour des marqueurs neutres et celui calculé pour des caractères quantitatifs ou encore pour des polymorphismes soumis à la sélection peut être réalisée. Ce afin de dégager les profils de différenciation imputables à la sélection (sur les caractères quantitatifs ou polymorphisme associés).

Corrélations et covariances entre caractères soumis à la sélection

La sélection s'effectue le plus souvent de manière simultanée sur plusieurs caractères, aussi la réponse à la sélection multivariée peut-elle être envisagée comme une extension du modèle de sélection portant sur un caractère ($R=h^2S$), peut s'écrire : $\Delta\bar{z} = \mathbf{GP}^{-1}\mathbf{S}$, ou $\Delta\bar{z}$ est le vecteur des réponses moyennes, \mathbf{G} est la matrice des variances covariances génétiques, \mathbf{P} la matrice des variances covariances phénotypiques et \mathbf{S} le vecteur des différentiels de sélection (Lande 1979).

La dérive entraîne des changements proportionnels (égalité des matrices ou égalité des composantes principales) alors que la sélection est susceptible d'entraîner des changements non-proportionnels (les matrices ont certaines composantes principales qui sont différentes ou la structure des matrices est sans relation aucune) (Roff 2000 pour une revue).

La comparaison des matrices de variances covariances génétiques renseigne donc sur l'héritabilité multivariée de ces populations (\mathbf{GP}^{-1}) et donc sur leurs différentiels de réponse à un même index de sélection. Elle renseigne également sur l'histoire à la fois démographique et sélectives des populations étudiées.

La gamme des outils disponibles pour l'analyse de la variation entre matrices s'est pendant longtemps limitée à des tests évaluant le degré de significativité de corrélations entre matrices, à l'exemple du test Mantel (Sokal et Rohlf 1995).

Le développement récent de nouvelles démarches analytiques permet désormais de tester le degré de relation existant entre deux matrices. Ces nouvelles méthodes permettent de tester si deux matrices sont égales, proportionnelles, partagent des composantes principales ou alors ont des structures sans aucune relation (Flury 1988; Phillips et Arnold 1999; Roff 2000; Roff 2002).

La comparaison de matrices \mathbf{P} a d'abord été développée par Flury (1988). Le modèle a ensuite été étendu par Phillips et Arnold (1999) aux matrices \mathbf{G} . Cette méthode de maximum de vraisemblance, dite de hiérarchie de Flury, permet de tester le partage de composantes principales, la proportionalité ou l'égalité de deux matrices. Sa programmation est malheureusement laborieuse (Phillips et Arnold 1999; Roff 2002) et les programmes existants (Phillips et Arnold 1999) ne permettent pas les analyses hiérarchiques telles que le test de l'effet village.

Une autre méthode a été développée par Roff (2002) et repose sur une analyse multivariée de variance (approche MANOVA) faisant appel au *Lambda* (Λ) de *Wilk* ou à la *Trace* (T) de *Pillai* pour tester la significativité de la variation entre les matrices \mathbf{G} . La comparaison de la hiérarchie de Flury et de l'approche MANOVA par Roff (2002) montre que si les deux méthodes donnent des résultats comparables, la facilité de programmation de l'approche MANOVA autorise une plus grande flexibilité dans l'utilisation de programmes statistiques standards (tels que SAS, <http://www.sas.com>), et la réalisation de différentes analyses hiérarchiques (par exemple : village, population dans village). Lors de nos comparaisons entre les matrices \mathbf{G} , nous emploierons donc cette approche, en appliquant la méthode du Jackknife (omission d'une famille lors de chaque itération de l'analyse) préconisée par Roff (2002) pour pallier l'existence d'une seule matrice \mathbf{G} pour chaque population.

Si les matrices de variances-covariances (\mathbf{G}) sont proportionnelles, alors les matrices de corrélations génétiques (\mathbf{G}_c) seront égales. Donc nous testerons dans un premier temps l'égalité des matrices \mathbf{G} et si elles diffèrent de façon significative, nous testerons l'égalité des matrices \mathbf{G}_c . Trois situations peuvent être mises en évidence : (i) l'égalité des matrices \mathbf{G} , (ii) les matrices \mathbf{G} sont inégales mais proportionnelles, ce qui suppose une l'égalité des matrices \mathbf{G}_c , et enfin, (iii) les matrices \mathbf{G} sont non-proportionnelles, ce qui suppose l'inégalité à la fois des matrices \mathbf{G} et \mathbf{G}_c .

2.3 Le déséquilibre de liaison

Corrélation entre paires d'allèles au sein des populations

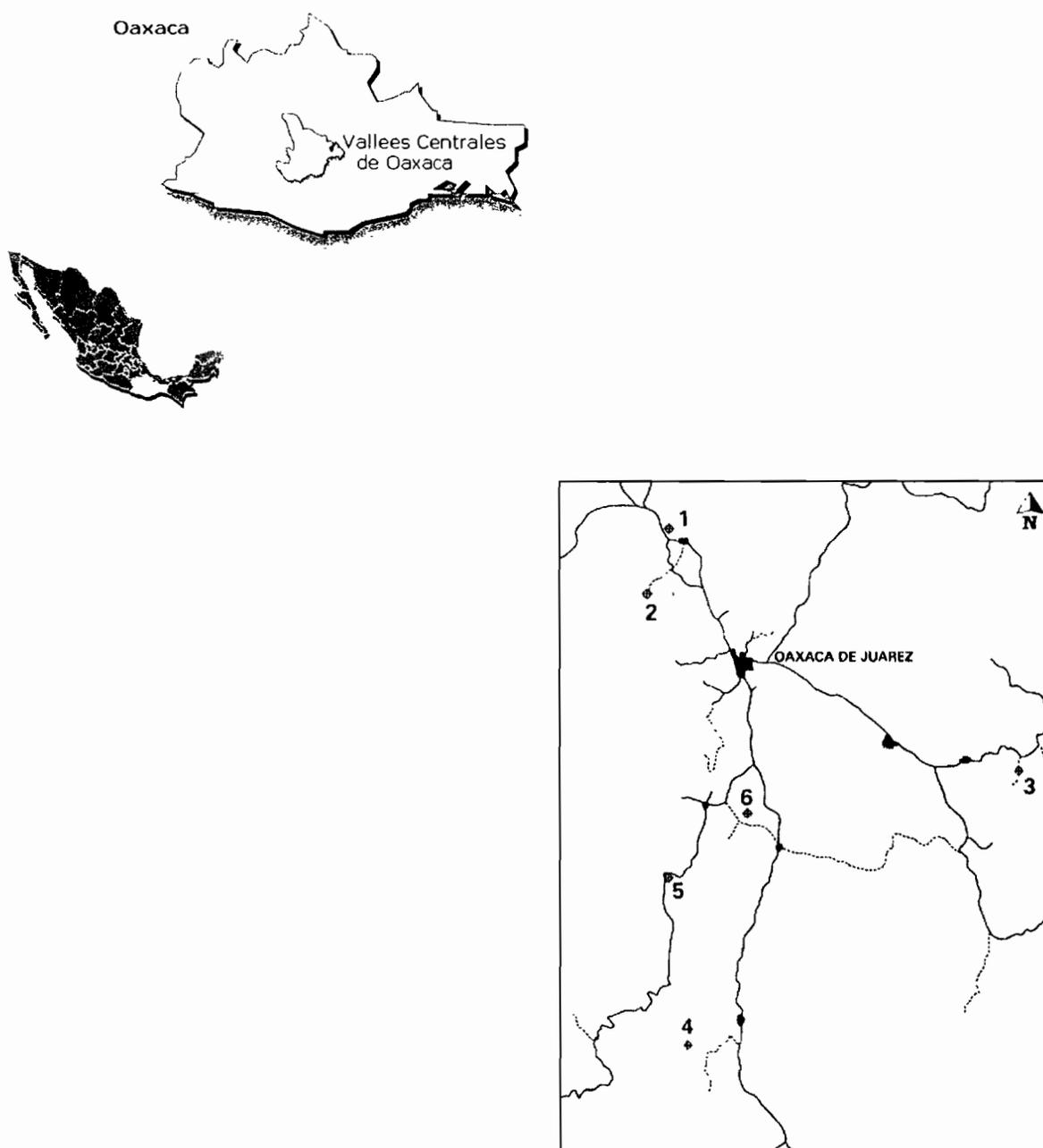
Pour tester la corrélation entre paires d'allèles au sein de populations étudiées, nous utiliserons le R_{ij} de Weir (1979). Sa significativité peut être testée par une statistique du χ^2 (Weir 1979). Cette méthode est néanmoins fortement affectée par la présence d'allèles rares.

Nous lui préférons un test-exact en testant le R_{ij} obtenu par rapport à ceux calculés en permutant les locus entre individus.

Les statistiques D d'Ohta

Ohta (1982) définit cinq statistiques permettant de décomposer la variance du déséquilibre de liaison en composantes intra et inter-populations (D_{is} , D_{st} , D'_{is} , D'_{st} et D_{it}). D_{is} représente la variance du déséquilibre intra-population, D_{st} représente la variance inter-populations sur les fréquences gamétiques attendues, D'_{is} est une composante de la variance entre sous populations sur les associations alléliques observées, D'_{st} est la variance du déséquilibre sur l'ensemble des populations et enfin D_{it} est la variance totale du déséquilibre. On peut à partir de ces statistiques distinguer trois situations : (i) les associations alléliques sont spécifiques d'une population donnée, $D_{is} < D_{st}$ et $D'_{is} > D'_{st}$, (ii) le déséquilibre de liaison est systématique et les associations alléliques sont les mêmes sur l'ensemble des populations, $D_{is} > D_{st}$ et $D'_{is} < D'_{st}$, (iii) et enfin la situation intermédiaire, $D_{is} > D_{st}$ et $D'_{is} > D'_{st}$, qui traduit un déséquilibre inégalement réparti (ne concerne pas l'ensemble des populations), et aussi que les associations alléliques (quand elles sont significatives) ne sont pas les mêmes d'une population à l'autre ($D'_{is} > D'_{st}$).

Figure 3.1 – Localisation de l'étude. La région des Vallées Centrales de Oaxaca dans l'état de Oaxaca au Mexique. Ces six villages sont localisés dans les vallées centrales de Oaxaca au Mexique. Ils comprennent les villages suivant : 1. Huitzo, 1730 m (mètres au-dessus du niveau de la mer) 17°15'N 96°51'W; 2. Mazaltepec, 1700 m 17°06'N 96°52'W; 3. San Lorenzo, 1830 m 16°51'N 96°16'W; 4. Amatengo, 1310 m 16°30'N 96°47'W; 5. Valdeflores, 1447 m 16°45'N 96°49'W; 6. Santa Ana, 1520 m 16°50'N 96°42'W.



III. MATÉRIEL

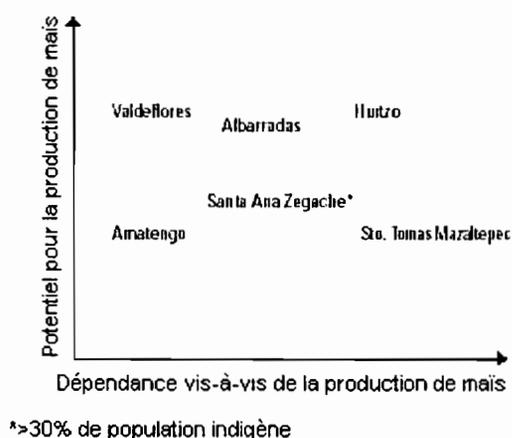
3.1 Choix de la région et des villages pour réaliser l'étude

L'état de Oaxaca (Figure 3.1) est un Etat du Mexique où la culture indienne est encore très forte. Le maïs y constitue la culture prépondérante, dont la finalité principale est l'alimentation humaine. C'est une région où la gestion des variétés traditionnelles de maïs par les agriculteurs peut être documentée et où très peu de variétés modernes de maïs ont été adoptées (Bellon *et al.* 2002). De plus, la culture du maïs revêt ici une importance particulière, de par ses implications agricoles et culturelles (Bellon *et al.* 2002), ce qui a motivé la réalisation d'une importante étude socio-économique au sein des villages sur lesquels portent la présente étude (Bellon *et al.* 2002; Smale *et al.* 1999).

L'ensemble de ces raisons fait apparaître la région de Oaxaca comme un bon modèle d'étude de l'impact de la gestion paysanne sur la dynamique de la diversité génétique chez le maïs.

Ces six villages ont été choisis par l'équipe de socio-économie du CIMMYT pour faire l'objet d'une étude d'une importante étude socio-économique (Bellon *et al.* 2002; Smale *et al.* 1999). Ils ont été choisis afin de maximiser l'hétérogénéité pour des variables socio-économiques (dépendance à l'égard du maïs et appartenance ethnique de la population) et de potentiel de production du maïs (Bellon *et al.* 2002; Smale *et al.* 1999) (Figure 3.2).

Figure 3.2 – Choix des villages pour réaliser cette étude (Smale *et al.* 1999).



La surface moyenne des fermes dans les villages étudiés est entre 3 et 4 ha (Smale *et al.* 1999). Seul Huitzo et Amatengo ont des surfaces d'exploitation significativement plus faibles que les quatre autres villages. Huitzo de l'ensemble des villages est le moins dépendant à l'égard de l'agriculture. Santa Ana Zegache et Mazaltepec présente une importante proportion de leurs populations parlant le zapotèque.

Du fait de l'étude socio-économique préalable à ce travail (Bellon *et al.* 2002; Smale *et al.* 1999) la liste de l'ensemble des agriculteurs cultivant le maïs était disponible pour chacun de ces villages.

Tableau 3.1 Exploitants agricoles par village étudié

Village	Nombre d'exploitants
Huitzo	326
San Tomás - Mazaltepec	253
San Lorenzo	447
Amatengo	146
Valdeflores	320
Santa Ana - Zegache	465

3.2 Collecte du matériel

La distance, à laquelle les variations de fréquences alléliques au sein des populations deviennent indépendantes, est utile pour déterminer l'extension des flux de gènes (mesuré à l'aide du F_{st} , Chapitre II). La variation d'une année à l'autre au sein de chacune de ces populations impose de réaliser l'échantillonnage sur une seule et même génération, " snapshot " à travers la région étudiée.

Le matériel a été collecté entre janvier et février 2000 au cours des premiers mois de cette thèse. Il s'agit de la récolte du cycle planté en 1999. Dix exploitations agricoles ont été sélectionnées aléatoirement dans chacun des six village. 60 exploitations ont donc fait l'objet d'une collecte d'échantillons. Ces exploitations avaient fait l'objet d'une enquête socio-économique préalable. Nous avons néanmoins à l'occasion de la collecte réalisée une enquête complémentaire liée à l'histoire de la population, la gestion des semences et aux échanges de semences réalisés entre agriculteurs. Nous avons collecté l'ensemble des populations distinctes identifiées par l'exploitant. Entre 20 et 40 épis ont été prélevés pour chacune de ces populations.

IV. Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico

Accepted by *Heredity*

Abstract

Assessing the impact of farmer management of maize landraces in the Central Valleys of Oaxaca, Mexico is crucial to an understanding of maize evolution, as it was first domesticated there. In this paper, we report on the impact of traditional farmer management of maize populations in this region in structuring molecular diversity and on the population dynamics of maize landraces. These populations, from a sample of local landraces cultivated by farmers in six villages, show little among-population differentiation ($F_{st}=0.011$). Most surprisingly, there is no isolation by distance and small among-village differentiation ($F_{st}=0.003$). For an outbreeding plant such as maize, one would expect populations to fit Hardy-Weinberg equilibrium, but significant homozygote excess ($F_{is}=0.13$) was found. This homozygote excess shows remarkable interpopulation and interlocus differences. We show that this pattern is related to variation in the mean anthesis-silking interval as well as to the flowering range or heterogeneity in flowering of a given population. A short anthesis-silking interval and high level of heterogeneity in flowering precocity will favor assortative mating. This leads to a locus-dependent population substructure giving an unusual case of Wahlund effect and inbreeding while high levels of seed exchange among farmers prevent population differentiation at both village and regional levels.

Key words

Assortative mating, flowering heterogeneity, F_{st} , inbreeding, maize landraces, population structure

INTRODUCTION

Little research has been directed towards understanding the evolutionary process behind plant evolution in agricultural ecosystems in traditional farming communities. Maize populations, like natural populations, are subject to migration and drift, to both natural selection and farmers' selection, and finally to local extinction and recolonization processes. Therefore, in order to gain a complete idea of the system, it is important to document in detail the genetic dynamics in these farmers' fields and the impact of their various practices.

Maize diversified first in the highlands of Mexico soon after domestication. Matsuoka *et al.* (2002) show that the domestication of maize is based on a unique event, and that maize accessions from the highlands of Oaxaca are genetically the closest to the wild ancestors of maize (*Zea mays* ssp. *parviglumis*). They report that the basal-most maize in a phylogenetic analysis, including maize and its wild relative teosinte, are maize accessions collected in regions close to the Central Valleys of Oaxaca. Furthermore, archeological work (Benz 2001; Piperno and Flannery 2001) has revealed remains of the oldest known maize in the State of Oaxaca, at the Guila Naquitz cave, dating back to 4200 B.C. The Oaxaca region has also been reported to hold a large amount of variation in terms of phenotypic diversity (Bellon *et al.* 2003).

Today in Oaxaca, farmers still cultivate maize populations in a traditional manner. Saving seed from one season to the next is a well-defined practice. In addition, the Central Valleys of Oaxaca show very little presence of, or impact from, modern varieties (Bellon *et al.* 2003; Smale *et al.* 1999), with most of the area still planted in local landraces. Therefore, this region offers unique conditions for the study of the evolutionary processes that are key to maize evolution. The concept of "landrace" is complex (Zeven 1998) and its complete definition remains an issue of contention amongst some parties. For the purposes of this paper, the term landrace refers to a maize population cultivated in a traditional fashion and managed by a single farmer.

Small-scale Mexican farmer management practices are central to the evolution of maize and its diversity. Key practices include the planting of numerous maize populations within a small area. Consequently, even if desired, farmers are incapable of preventing the exchange of pollen between populations (Bellon and Brush 1994; Louette *et al.* 1997). Furthermore, Mexican farmers commonly acquire seed from both local and distant farmers or sources, often

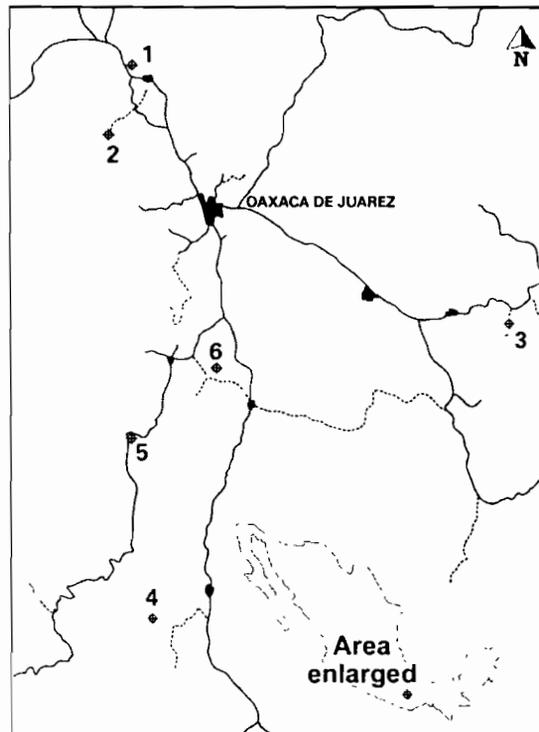


Figure 4.1 – Location of sampled villages in the Central Valleys of Oaxaca around the city of Oaxaca de Juarez (Mexico). Villages are numbered from 1 to 6. Altitude, latitude, and longitude are also given: 1. Huitzo, 1730 masl (meters above sea level) 17°15'N 96°51'W; 2. Mazaltepec, 1700 masl 17°06'N 96°52'W; 3. San Lorenzo, 1830 masl 16°51'N 96°16'W; 4. Amatengo, 1310 masl 16°30'N 96°47'W; 5. Valdeflores, 1447 masl 16°45'N 96°49'W; 6. Santa Ana, 1520 masl 16°50'N 96°42'W.

mixing their seed with seed from other farmers or that purchased from markets (Louette and Smale 2000).

In addition to farmer management, the biology of the species is expected to play a major role in structuring maize populations. Maize is a monoecious species that bears two inflorescence types, the staminate tassel and the pistillate ear. There is generally a delay between male and female flowering, with male flowering generally occurring before female flowering. If the delay between male and female flowering is short enough, it could allow assortative mating to occur. Homozygous excess, and its variation across loci and populations, has been reported for maize open-pollinated populations using both isozymes and RFLP markers (Brown and Allard 1970; Dubreuil and Charcosset 1998; Garnier-Géré 1992; Kahler *et al.* 1986; Lefort-Buson *et al.* 1991; Salanoubat and Pernes 1986). Assortative mating would produce a locus-dependent Wahlund effect (Nevo *et al.* 2000), which was tested in the research reported herein.

In this paper we assess the genetic diversity and population structure of maize landraces from the Central Valleys of Oaxaca. Because 83% of the total maize in this area is white kernel maize (Smale *et al.* 1999), this study focuses on this maize type. We assess key agroecological factors and components of farmer management behind maize population dynamics. We also describe how this genetic diversity is structured, how management practices by farmers have an effect on population differentiation, and effects related to distance and seed exchange among Oaxacan farmers. In addition, we investigate the consequences of some flowering traits that could play a role in assortative mating in maize landrace populations. We use markers specifically linked to the QTLs of flowering traits, and other markers scattered throughout the genome.

MATERIAL AND METHODS

Survey on farmer management

We selected six villages (Figure 1) for this study, because of their contrasting situations, in terms of ethnicity, and their maize production potential, based on an earlier study (Bellon *et al.* 2003; Smale *et al.* 1999). Ten households per village were randomly chosen, giving a total of 60 farmers surveyed. We gathered information on farmer seed management and on seed exchange practices between and within villages.

Material used for the genetic analysis

Of the six studied villages, Santa Ana and Huitzo showed the highest contrast in ethnicity, number of maize populations per farmer, and potential in terms of maize production (Smale *et al.* 1999), and were therefore studied more extensively. A total of 31 populations were assayed, including field evaluation and genotyping. These included nine populations from Huitzo, three from Mazaltepec, three from San Lorenzo, three from Amatengo, three from Valdeflores, and 10 from Santa Ana. Households were selected randomly among the farmers cultivating a population of white kernel maize within each village. Sampling was carried out within a single generation. We randomly selected 20 open-pollinated families for each population and genotyped one individual per family; field evaluation for flowering traits was carried out on 18 of these families.

Simple sequence repeat genotyping

A total of 11 microsatellite markers were assayed. SSR primers were selected from the maizeDB database of public SSRs, and included the following: phi011, phi227562, phi96100, phi101049, phi029, phi093, phi024, phi452693, phi034, phi014, and umc1061. They consist of tri- or tetranucleotide repeats. Markers were selected according to their chromosomal locations, in order to provide for genome-wide coverage, and also by the size of the amplification product, to allow multiplexing on an automated DNA sequencer. Sequences and mapping positions can be downloaded at <http://www.agron.missouri.edu/ssr.html>. Of the 11 microsatellites used for this study, most did not exhibit a stepwise variation. Three of these markers (phi011, phi024, and phi452693) map close to genes or QTLs involved in flowering time or anthesis-silking interval (Gale and Devos 1998; Ribaut *et al.* 1996; Thornsberry *et al.* 2001; Veldboom *et al.* 1994).

Cytoplasm genotyping

All populations were also characterized with chloroplastic markers. We used a polymorphic chloroplast set of primers surrounding a polyA repeat in the *psbK/psbI* intergenic region with the following sequences: zmcp7430-F: CGAAGCTGCTGTAAGTTTTTCG and zmcp7430-R: AAGACTTCTCGGCTCTTATCCA (Provan *et al.* 1999).

Analysis of the among-population genetic structure

Because villages were not randomly selected, the genetic description given is specific to our sample. Overall $F_{st}=\theta$ (Weir and Cockerham 1984) was calculated for the entire set of 31 populations. Jackknifing over populations and loci was used to provide a confidence interval, according to Weir (1996). θ values were estimated using GDA 1.1 software (Lewis and Zaykin 2002), which performs hierarchical F-statistics. Different levels of population subdivision were tested as suggested in Weir (1996). For a random mating population (within sample) or random distribution of individuals (between samples), F-statistics are expected to be null.

A matrix of pairwise $F_{st}/(1-F_{st})$ was estimated as well as a matrix of geographic distances $\ln(Dist)$ (Rousset 1997) between villages (geographic coordinates provided in Figure 1) to test for isolation by distance. A Mantel test (Sokal and Rohlf 1995) was used to test for the independence of the matrices.

The ratio of pollen to seed flow, according to Ennos (1994) is:

$$r = \frac{\left(\frac{1}{F_{STn}}\right)(1 + F_{IS}) - 2\left(\frac{1}{F_{STc}}\right)}{\left(\frac{1}{F_{STc}}\right)}$$

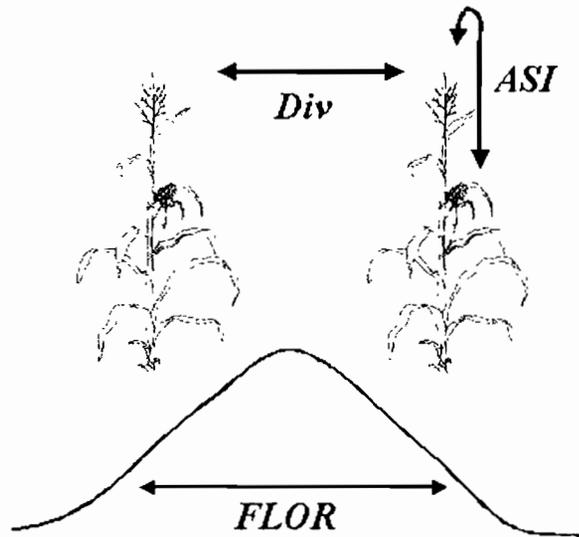
where F_{STn} is the F_{st} calculated for microsatellite markers and F_{STc} is calculated for cytoplasmic markers.

Analysis of the genetic variation within populations

We estimated the expected unbiased heterozygosity (Nei 1987) $H_e=2n(1-\sum_i p_i^2)/(2n-1)$ for microsatellite markers and $H_e=n(1-\sum_i p_i^2)/(n-1)$ for cytoplasmic markers. Homozygote excess was estimated according to Weir and Cockerham (1984) within population and for the whole set of populations. For pairwise linkage disequilibrium among loci, the within-population correlation coefficient R (Weir 1979) was calculated and tested by permuting genotypes within locus within population using the *GENETIX* 4.02 software (Belkhir *et al.* 2001).

Within-population genetic variation in relation to temporality in flowering

In order to characterize populations for flowering, 18 open-pollinated families were sampled for each of the 31 populations. Up to 12 progenies per family were evaluated (giving a total of $31 \times 18 \times 12 = 6888$ plants evaluated). Field layout was a two-replicate design with hierarchical structure (population plots randomly assigned and family plots randomly assigned within populations).



The experiment was carried out at the CIMMYT experimental station at El Batán, Texcoco, Mexico. Days to silking (DS) and days to anthesis (DA) were assessed. Measurements were carried out under well-watered conditions and were averaged over family. Anthesis-silking interval (ASI) results from the operation $ASI = DS - DA$. Flowering range ($FLOR$) was estimated for each population $FLOR = DS_{max} - DS_{min}$ where DS_{max} and DS_{min} are the maximum and minimum DS in a given population. The flowering range within a given population is an estimate of the heterogeneity in flowering within this population.

Genetic structure within populations was assessed by temporal autocorrelation analysis in the same manner as spatial autocorrelation analysis (Hardy and Vekemans 1999). We used an estimate of Wright's coefficient of relationship, ρ_{ij} , between pairs of individuals (Hardy and Vekemans 1999) that corresponds to Moran's I -statistic using individual allele frequencies (Dewey and Heywood 1988). The genotype of each individual (one individual per family) was used as the variable of interest to calculate ρ_{ij} . The family mean flowering time was used in the same way as other authors have used spatial coordinates (Hardy and Vekemans 2001). Therefore, we define as the divergence in flowering time between pairs of families $Div = |DA_1 - DA_2|$ where DA_1 and DA_2 are the DA mean values of family 1 and 2, respectively. Regression analysis of ρ_{ij} on Div was carried out using the *SPAGeDi* 1.0 software (Hardy and Vekemans 2002). The probabilities under the hypothesis that there is no relation between Div and ρ_{ij} , Pr values for $|obs| = exp$, were estimated after 10,000 random permutations of temporal locations. It is equivalent to carrying out a Mantel test.

Table 4.1 Among-population genetic structure

		Within-village differentiation	Among-village differentiation
θ	SSRs all loci	0.011 ^S	0.003 ^S
	Cytoplasmic markers	0.022 ^S	0.007 ^{NS}

^{NS} not significantly different to zero. ^S significantly different to zero. Estimates of among villages and within villages components of $F_{st}(\theta)$ for microsatellites (SSRs) markers and cytoplasmic markers

RESULTS

Farmer management of maize populations

Of the 60 farmers surveyed, six reported having participated in seed exchange between different villages during the last 10 years. Of these six reported cases, four correspond to the foundation of a new population and two to the mixing of seeds of the preexisting population with seeds from another village. Twelve of the 60 farmers also reported that during the last 10 years they have mixed the seed of the preexisting population with seed provided by another farmer from the same village, in order to provide enough seed for the next generation.

Principal coordinate analysis for molecular markers

All tested SSR loci were polymorphic in all populations. Principal coordinate analysis provides little evidence for population differentiation. Principal coordinates one and two together explain less than 6% of the total variation. Projection of the populations over the first two planes shows a uniform distribution and no grouping of populations. Furthermore, all populations show an overlapping of their distribution. Microsatellite polymorphism appears to be continuous and not related to geographic origin of the studied populations.

Among-population genetic structure

F_{st} values obtained both for nuclear and cytoplasmic (chloroplast) markers are indicated in Table 1. All populations, even those separated by up to 100 km, were found to share chloroplast DNA haplotypes. There was no statistical significance for isolation by distance for either microsatellites or cytoplasmic markers. Furthermore, we observed low among-village F_{st} values, significantly lower than among-population F_{st} values. This could be explained by long distance gene flow (i.e., seed exchange between villages). Supporting the predominance of seed versus pollen flow is an estimate of the ratio of pollen to seed flow (Ennos 1994), which is inferior to 1 with $r=0.55$ where $F_{STc}=0.028$ and $F_{STn}=0.011$. To estimate r , we did not consider among-village differentiation; we considered the inbreeding coefficient $F_{is}=0$, as we will show later in this paper that the observed excess of homozygotes does not correspond to consanguineous mating.

Genetic variation within populations

Genetic diversity estimated over all populations is $H_e=0.71$ for microsatellite markers and $H_e=0.49$ for cytoplasmic markers. A significant departure from Hardy-Weinberg equilibrium is observed in almost all populations. Estimates of F_{is} over all populations by loci show significant variation of homozygote excess among loci (Table 2). To test the uniformity of the homozygote excess among loci, we checked the distribution of F_{is} . One empty region (gap) was found. The three loci that map close to genes or QTLs involved in flowering time or anthesis-silking interval (phi011, phi024, and phi452693) showed significantly higher F_{is} values than the others (Table 2).

Table 4.2 Population structure by loci

Locus	Chromosomal		
	location (bin)	f	θ
phi 011	1.09	0.261±0.039	0.015±0.009
phi 227562	1.11	0.059±0.025	0.009±0.004
phi 96100	2.00	0.075±0.029	0.017±0.008
phi 101049	2.08	0.092±0.024	0.003±0.003
phi 029	3.04	0.126±0.046	0.004±0.007
phi 093	4.08	0.143±0.020	0.011±0.008
phi 024	5.01	0.252±0.035	0.004±0.005
phi 452693	6.04	0.347±0.029	0.011±0.008
phi 034	7.02	0.079±0.027	0.010±0.006
phi 014	8.04	0.093±0.037	0.019±0.009
umc 1061	10.06	0.002±0.021	0.020±0.010
overall markers		0.138±0.031	0.011±0.002

Estimates of F_{is} (f) and F_{st} (θ) (not considering among-village differentiation) according to Weir and Cockerham (1984) by loci over all populations and over all markers. 95% confidence interval provided by jackknifing over populations for individual loci and by jackknifing over loci for the overall markers value.

We evaluated the number of significant linkage disequilibria between pairs of loci. They are not higher than expected by chance alone at $p \leq 5\%$ (5.1% positive over 31 populations and 55

Table 4.3 Genetic and flowering variation within populations

Village	Farmer's ID	H_e	H_{obs}	F_{is}	$FLOR$	ASI
Huitzo	107	0.72	0.54	0.25	16.62	3.52
Huitzo	108	0.65	0.52	0.21	17.03	4.40
Huitzo	110	0.69	0.55	0.20	12.92	3.94
Huitzo	115	0.69	0.58	0.17	32.25	3.98
Huitzo	121	0.68	0.55	0.20	16.72	5.80
Huitzo	122	0.69	0.65	0.07	15.88	4.08
Huitzo	129	0.71	0.66	0.07	12.92	3.47
Huitzo	136	0.66	0.59	0.12	17.83	3.03
Huitzo	139	0.71	0.57	0.21	23.50	5.05
Mazaltepec	211	0.70	0.61	0.13	20.75	7.23
Mazaltepec	222	0.69	0.64	0.07	18.33	8.36
Mazaltepec	235	0.68	0.65	0.05	26.48	5.34
San Lorenzo	309	0.69	0.67	0.03	18.17	6.96
San Lorenzo	314	0.66	0.55	0.17	14.72	3.47
San Lorenzo	317	0.67	0.64	0.05	13.33	8.76
Amatengo	405	0.70	0.63	0.10	10.99	4.61
Amatengo	433	0.70	0.61	0.14	12.94	4.31
Amatengo	439	0.71	0.62	0.13	8.17	2.91
Valdeflores	512	0.69	0.61	0.13	9.24	4.52
Valdeflores	533	0.69	0.64	0.07	10.81	4.08
Valdeflores	536	0.70	0.62	0.12	11.67	4.80
Santa Ana	602	0.71	0.63	0.11	10.18	4.68
Santa Ana	606	0.70	0.58	0.17	13.67	4.36
Santa Ana	610	0.72	0.64	0.11	12.92	5.02
Santa Ana	614	0.72	0.61	0.17	12.79	3.07
Santa Ana	619	0.70	0.57	0.18	12.60	4.17
Santa Ana	625	0.70	0.63	0.10	14.00	4.34
Santa Ana	630	0.72	0.57	0.21	15.22	4.63
Santa Ana	633	0.69	0.58	0.16	9.23	3.34
Santa Ana	639	0.71	0.52	0.27	18.92	2.99
Santa Ana	640	0.70	0.61	0.13	10.60	3.97

Expected heterozygosity (H_e), observed heterozygosity (H_{obs}), homozygote excess over all markers (F_{is}), within population heterogeneity in flowering time measured as flowering range ($FLOR$) and anthesis-silking interval (ASI) for 31 populations from six villages in the Central Valleys of Oaxaca. $FLOR$ and ASI are given in days

Table 4.4 Regression of homozygote excess F_{is} on flowering range ($FLOR$) and the mean anthesis-silking interval (ASI)

Parameter	Estimate	Student's t value		Std Error of Estimate
		Parameter=0	Pr > t	
INTERCEPT	0.142441	3.54	0.0016	0.04029086
$FLOR$	0.007635	2.88	0.0079	0.00265083
ASI	-0.02404	-3.71	0.0010	0.00647796

$R^2=0.39$. F_{is} was calculated over all markers.

possible different pairs of loci). Using a Bonferroni correction, no significant linkage disequilibrium is observed.

Within-population genetic variation in relation to temporality in flowering

A summary of genetic variation and variation for flowering traits for all populations is presented in table 3. Two populations show clear evidence of admixture of material of different flowering precocity. Population from farmer 235 (Table 3) shows a bimodal distribution of the flowering time, and the population from farmer 115 (Table 3) has a very wide flowering range compared with the distribution of the flowering range among all populations. These two populations will not be considered for regression analysis.

The within population range in flowering time and anthesis-silking interval differs greatly from one population to the other. The regression of the overall homozygote excess (using all 11 microsatellite markers) on mean anthesis-silking interval and flowering range is highly significant (Figure 2 and Table 4).

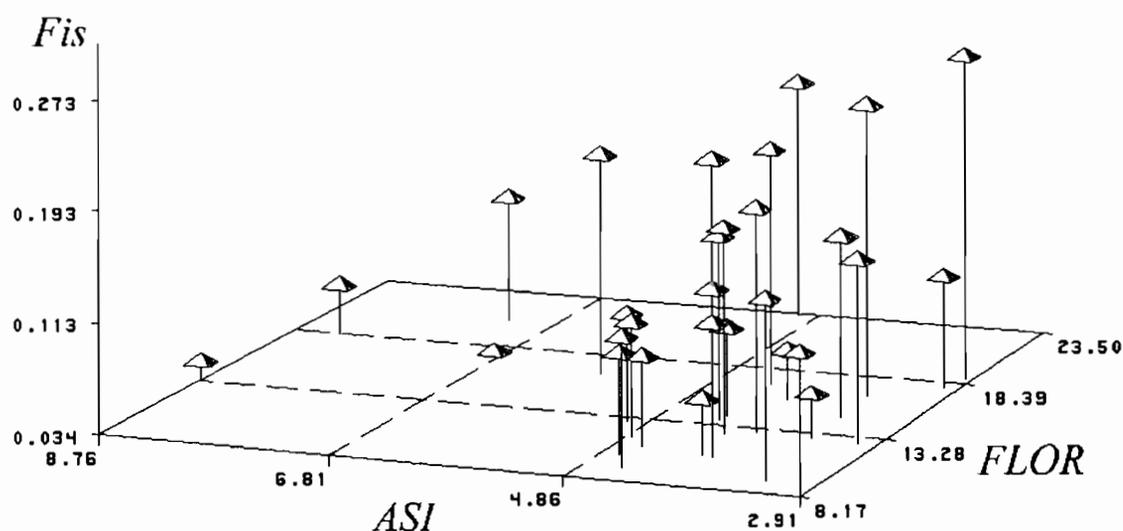
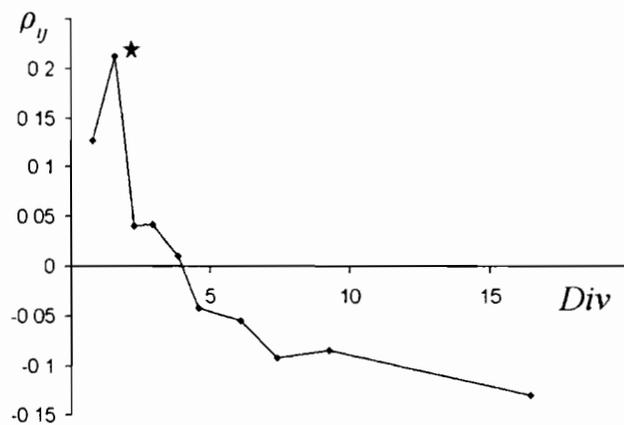


Figure 4.2 Homozygote excess (F_{is}) as a function of the flowering range ($FLOR$) and the anthesis-silking interval (ASI). F_{is} was calculated over all markers.

The population showing the most elevated F_{is} value (639 in Table 3) shows a significant correlation between the flowering distance and relatedness measured as Moran's I -statistic

(temporal autocorrelogram shown in Figure 3). For this population, pairwise estimates of Moran's I -statistic (ρ_{ij}) for the divergence in flowering (Div) class was between 0.75 and 1.56 days, with an average of 1.18 days, which significantly deviates from the expected value obtained by permuting locations as show in figure 3 ($Pr=0.029$, star in Figure 3). In addition, the linear regression of ρ_{ij} on Div is significant ($Pr=0.0018$). Regression slope is $m=-0.0244$ and the intercept, $b=0.1136$. These results suggest a clear case of assortative mating (correlation coefficient, $r=-0.243$) for this population. This population was tested for assortative mating because it is the only population showing a low anthesis-silking interval value, a high heterogeneity in flowering, and an elevated F_{is} value. To conduct the regression, we used only the three markers (described earlier) showing significantly higher F_{is} values than the others.

Figure 4.3 Autocorrelogram showing the temporal segregation of the genotypes (Div is the divergence in flowering time) in function of the relatedness measured as Moran's I -statistic (ρ_{ij}). Given for the population showing the most elevated overall F_{is} value and the lowest ASI value (population from farmer 639 in Table 3).



DISCUSSION

Low level of among-population differentiation

Village and distance do not appear to be determinants in population differentiation. The difference between cytoplasmic and nuclear F_{st} is not as large as expected when compared to other allogamous plants, and even less than that seen for some autogamous grasses (Ennos 1994). Data from the survey on farmer management corroborate the genetic data analysis and indicate a large amount of seed-flow between maize populations within the Central Valleys of Oaxaca. Considerable seed exchange by farmers seems to be common in Mexico (Louette *et al.* 1997). Furthermore, while the proportion of farmers reporting seed flow is high, it could be undervalued as farmers are keen to test new populations (Bellon *et al.* 2003), which

often results in seed mixing as a consequence of storage practices. The maize ears in husks will frequently be kept after harvest in a single pile, regardless of whether they are from the tested population or a farmer's preexisting population. We observed that populations from other villages are often tested, and, although the farmer may not adopt them, seed mixing may occur. Seed flow among farmers allows long distance gene flow within and among villages. These high levels of gene flow ensure the maintenance of high levels of genetic diversity. Levels of diversity are high when compared to those measured with maize accessions representing the entire maize genetic diversity in the Americas (Matsuoka *et al.* 2002). Therefore, a maize landrace should not be considered as a separate entity, but rather as an open genetic system. Furthermore, our results underline the importance of farmers' choices in determining gene flow among maize landrace populations, as seed flow results from farmers' decisions.

To the best of our knowledge, this population study in maize is the first to be conducted based on a small geographic area. In contrast to our research, other maize population studies (Gauthier *et al.* 2002; Sanou *et al.* 1996) looked at larger areas (Europe and Burkina-Faso) and showed much higher population differentiation. We believe more studies are needed on maize populations in Mexico and more generally the Americas, both at the regional and inter-regional scales, to investigate the patterns of population structure. Although distance does not seem to contribute to village isolation within small geographic areas, Matsuoka *et al.*'s (2002) results clearly suggest isolation by distance at a continental scale.

Temporal heterogeneity in allelic frequencies

While there is little among-population differentiation, a large amount of homozygote excess is observed within these populations. Usually, homozygote excess is attributed to consanguineous mating, population substructure, or to an artifact due to factors like null alleles. Enjalbert and David (2000) have inferred the outcrossing rate in wheat using molecular data at various loci. In a more recent study, Overall and Nichols (2001) have shown that it is possible to distinguish consanguinity from population substructure using multilocus genotype data. However, none of the above applies to the situation described in this paper. The biology of the species rejects the simple explanation of inbreeding, and the significant variation of homozygote excess among loci (F_{is} value differing considerably from one locus to another) does not correspond to what would be expected for population admixture. Nonrandom mating has been previously described in open-pollinated maize populations

(Bijlsma *et al.* 1986; Brown and Allard 1970). Kahler *et al.* (1989) have shown that selfing did not contribute significantly to the inbreeding that occurred in the studied population. Here we present an unusual case of Wahlund effect and inbreeding, which is in some ways similar to that described in mole crickets (Nevo *et al.* 2000), and which corresponds to assortative mating. We show that in the populations with the most elevated homozygote excess, there is a significant correlation between relatedness and the mean family divergence in flowering time. It appears that overall F_{is} value depends on both the flowering range and on the anthesis-silking interval of a given population (*FLOR* and *ASI* together explain around 39% of the variation for homozygote excess). The existence of a large anthesis-silking interval has the consequence of preventing assortative mating because of the long delay between male and female flowering. A low anthesis-silking interval and large flowering range of the population will result in a temporal heterogeneity in allelic composition for maternal plants and the pollen pool. Work by Kahler *et al.* (1989) had previously suggested that homozygote excess in an open-pollinated maize population could be the consequence of positive assortative mating resulting from an overlap in the flowering period of plants carrying alleles that are identical by descent from a recent common ancestor.

Marker loci with elevated F_{is} values map close to flowering genes

The three loci showing the most elevated F_{is} values are in regions known to be associated with flowering traits. Phi011 is within the interval between *dwarf8* and *indeterminate1*. While *dwarf8* has been shown to be associated with variation in flowering time (Thornberry *et al.* 2001), *Indeterminate1* is a putative transcriptional regulator of floral transition that is thought to be a major player in controlling flowering time. Phi024 is in a region that is syntenic to the *dwarf8-indeterminate1* region and maps close to *dwarf9*, a possible duplication of *dwarf8* on chromosome 5 (Gale and Devos 1998). Phi452693 maps close to a major anthesis-silking interval QTL (Ribaut *et al.* 1996; Veldboom *et al.* 1994). A large flowering range within a given population will lower the effective population size of a given class of flowering time and therefore lead to inter-class differentiation.

Regarding phenotypic evolution

Understanding population structure and its pattern are crucial to understanding phenotypic evolution. It makes possible association studies with clear assumptions about population structure and its origin. The Central Valleys of Oaxaca offer a unique model for the study of

the impact of farmer management and selection on the phenotypic diversification and evolution of maize because of the large amounts of phenotypic variation (Bellon *et al.* 2003), the unique position of these populations in maize evolution and diversification (Matsuoka *et al.* 2002), and the patterns of population structure in this region described in our results. The observed variation in flowering range and in anthesis-silking interval between populations suggest that the pattern of population structure for these traits could be very different from that described for molecular markers. In a future paper we will describe the pattern of population structure for quantitative traits and the impact of farmer management and selection on population differentiation for these traits, in order to understand the basis of phenotypic evolution in maize.

V. Population structure and strong divergent selection shape phenotypic diversification in maize landraces

A shorter version of this chapter has been accepted by *Heredity*

Abstract

To conserve the long-term selection potential of maize, it is necessary to investigate past and present evolutionary processes that have shaped quantitative trait variation. Understanding the dynamics of quantitative trait evolution is crucial to future crop breeding. We characterized population differentiation of maize landraces from the State of Oaxaca, Mexico for quantitative traits and molecular markers. Q_{st} values were much higher than F_{st} values obtained for molecular markers. While low values of F_{st} (0.011 within-village and 0.003 among-villages) suggest that considerable gene flow occurred among the studied populations, high levels of population differentiation for quantitative traits were observed (i.e., an among-village Q_{st} value of 0.535 for kernel weight). Our results suggest that although quantitative traits appear to be under strong divergent selection, a considerable amount of gene flow occurs among populations. Furthermore, we characterized nonproportional changes in the **G** matrix structure both within and among villages that are consequences of farmer selection. As a consequence of these differences in the **G** matrix structure, the response to multivariate selection will be different from one population to another. Large changes in the **G** matrix structure could indicate that farmers select for genes of major and pleiotropic effect. Farmers' decision and selection strategies have a great impact on phenotypic diversification in maize landraces.

Key words

F_{st} , **G** matrix, maize landraces, population structure, Q_{st} , quantitative traits

INTRODUCTION

One of the most important issues in maize evolution is how to explain the extraordinary morphological and genetic diversity that exists among the maize landraces (Matsuoka *et al.* 2002). This is of great interest to breeders and maize geneticists because understanding maize phenotypic evolution is critical to future maize breeding. As new insights into phenotypic diversity are brought to light by association mapping in maize (Thornsberry *et al.* 2001), elucidating the mechanisms underlying the extraordinary morphological and genetic diversity will allow us to elaborate new strategies for breeding and for screening genebanks for alleles that confer desired traits. To better conserve the current and future adaptive ability of a species, it is necessary to investigate past and present evolutionary processes that shaped quantitative trait variation.

Crop plant evolution occurred and possibly still occurs in an environment that is managed by traditional farming communities. It results in a very patchy environment where a farmer in a limited area, contiguous to other farmers' fields, manages a given landrace. Little research has been aimed at better understanding the evolutionary process behind plant evolution within the agricultural ecosystems of traditional farming communities. Maize landrace populations, like natural populations, are subject to migration and drift, to both natural selection and farmers' selection, and finally to local extinction and recolonization processes. Therefore, to gain a complete picture of the system, it is important to document in detail the genetic dynamics in these farmers' fields and the impact of their various practices. Traditional maize agroecosystems provide a framework to investigate local selection for a number of quantitative traits in a highly subdivided metapopulation with high migration rates among populations. The management practices of small-scale Mexican farmers are key to the evolution of maize and its diversity. Crucial practices include the planting of numerous maize populations within a small area under the management of a single farmer. A maize landrace or population is defined by the farmer in terms of ear characteristics; ear type will be maintained by the farmers through conservative selection in spite of considerable gene flow (Louette *et al.* 1997; Louette and Smale 2000). Today, in Oaxaca, many farmers still cultivate maize in the same way their ancestors have for thousands of years. Seeds for planting are obtained from ears selected by the farmers from the harvest of the previous cycle. On many occasions, seed is also obtained from neighbors or other sources, allowing farmer-controlled seed flow among populations (Smale *et al.* 1999). In addition, the Central Valleys of Oaxaca show very little

presence of or impact from modern varieties (Bellon *et al* 2003; Smale *et al* 1999). Furthermore, previous research suggests that this region possesses a large amount of phenotypic variation (Bellon *et al* 2003), therefore, it offers unique conditions for the study of the evolutionary processes that are fundamental to maize phenotypic evolution and diversification.

Measuring the strength of selection in a farmer's field and with a farmer's selection criteria is not an easy task. Our experience shows us that it is difficult to obtain from farmers a precise description of their selection practices. While all the harvested ears go through the selection process, the farmer and his wife are both involved at different stages throughout the year in selecting the ears for the next cycle. Therefore, a constant strength of selection over time or across farms is unlikely.

To measure the effects of farmer selection and on-farm management on phenotypic differentiation, we prefer to use methods that provide indirect evidence of selection. When referring to "selection," we include as "farmer selection", both direct selection by the farmers and indirect selection, resulting either from farmers' practices or from natural selection. The comparison of population differentiation for quantitative traits as opposed to that of molecular markers (McKay and Latta 2002 for a review) and the comparison of genetic variance and covariance matrices across populations (Roff 2000 for a review) are used.

After fierce debate over the comparison of variation of gene frequencies among populations with variation of quantitative traits (Felsenstein 1986; Lewontin 1984; Rogers 1986), such comparison has recently allowed evolutionary and ecological geneticists to gain insight into the process of phenotypic evolution (McKay and Latta 2002 for a review). As suggested by Rogers (1986) and Felsenstein (1986), over time this comparison has proven its value in providing valuable information about phenotypic evolution in natural populations, although Lewontin (1984) considered it to be meaningless. Wright standardized index F_{st} , was shown to be the same for neutral additive quantitative characters as for a neutral marker (Lande 1992). Therefore, a value of F_{st} for quantitative trait (Q_{st}) greater than F_{st} for neutral loci is evidence of spatially divergent evolution by natural selection (Whitlock 1999). The use of the comparison of Q_{st} and F_{st} for molecular markers provides a tool allowing the comparison of patterns of population structure and differentiation resulting from farmers' selection.

As genetic progress is not based on single trait improvement, and yield is a complex trait with many components, farmer selection could affect the covariance among traits of agronomical

interest. Multivariate selection response depends on the genetic variance-covariance matrix (G matrix) structure, which ultimately determines the outcome of long-term selection. The sign and strength of the correlation between traits of agronomic interest in maize will affect the outcome of multitrait selection. Comparison of G matrix structure also yields information on past history, selection, and drift processes of populations. Interest in G matrix comparison has been discussed by Roff (2000). In this study, we investigate whether farmers' selection of certain quantitative traits permits a differentiation of populations and whether this results in changes in the G matrix structure. Therefore, in support of the previously mentioned comparison of population differentiation for molecular markers and quantitative traits, G matrix comparison also provides evidence of selection in farmers' fields and the consequences of this selection on the correlations and covariations among traits.

Our goal in this paper is to describe the impact of farmers' selection on phenotypic differentiation and diversification in the Central Valleys of Oaxaca. As 83% of the total maize in the Central Valleys is white kernel maize (Smale *et al.* 1999), this study focuses on this maize type. We compare the patterns of population differentiation for quantitative traits to that previously described for molecular markers (see companion paper). This will provide a better understanding of the evolutionary dynamics and also of the adaptive ability of traditional maize landraces.

MATERIAL AND METHODS

Material used for the genetic analysis

A total of 31 populations were assayed in the field evaluation and by genotyping. The material used in this study has been described elsewhere (see companion paper). All maize landraces used in this study predominantly showed characteristics of the Bolita race as described by Wellhausen *et al.* (1952).

Simple sequence repeat genotyping

A total of 11 microsatellites markers were used for analysis (see companion paper). SSR primers were selected from the maizeDB database of public SSRs, and included the following: phi011, phi227562, phi96100, phi101049, phi029, phi093, phi024, phi452693, phi034, phi014, and umc1061. Markers were selected according to their chromosomal

locations, in order to provide for genome-wide coverage, and also by the size of the amplification product, to allow multiplexing on an automated DNA sequencer.

Characterizing populations for quantitative traits

Eighteen open-pollinated families were sampled for each of the 31 populations. The field layout was a two-replicate design with hierarchical structure (population plots randomly assigned and family plots randomly assigned within populations). The experiment was carried out at the CIMMYT experiment station at El Batán, Texcoco, Mexico. Up to 12 open-pollinated progenies per family were evaluated (giving a total of $31 \times 18 \times 12 = 6888$ plants evaluated) for plant development, ear development, and kernel traits. These included days to silking (*DS*), days to anthesis (*DA*), anthesis-silking interval (*ASI*), plant height (*PH*), ear height (*EH*), ear length (*EL*), ear width (*EW_i*), ear weight (*EW*), row number (*RN*), kernel count (*KC*), kernel count per rows (*KCR*), cob width (*CW_i*), overall grain weight (*GW*), kernel weight (*KW*), and kernel thickness (*KT*). All measurements, except for days to silking and days to anthesis, were made at plant maturity.

Principal component analysis (PCA)

PCA was performed using SAS software (SAS Online documentation at <http://www.sas.com>) over the whole data set and included all measured traits *EL*, *EW_i*, *EW*, *RN*, *KC*, *KCR*, *CW_i*, *GW*, *KW*, *KT*. To produce graphical output for ear and kernel traits, principal component scores were averaged over family in order to show variation between populations. Principal component scores were averaged over family and then over population to show variation between villages.

Computing θ

Overall $F_{st} = \theta$ (Weir and Cockerham 1984) was calculated for the entire set of 31 populations. Jackknifing over populations and loci was used to provide a confidence interval according to Weir (1996). A matrix of pairwise F_{st} among populations was also calculated. F_{st} values were estimated using GDA 1.1 (Lewis and Zaykin 2002), which performs hierarchical F-statistics. Different levels of population subdivision were tested as suggested in Weir (1996).

Analysis of variation for quantitative traits

Analysis of variance and calculation of variance components were carried out using the following model:

$$P_{ijk} = \mu + Pop_i + HS(Pop)_{ij} + \epsilon_{ijk}$$

Where P_{ijk} is the phenotypic value of the j^{th} family in the i^{th} population; μ is the overall mean; Pop_i the i^{th} population effect; $HS(Pop)_{ij}$ the j^{th} family effect within the i^{th} population; ϵ_{ijk} the residual representing variability within families.

Assuming that allelic effects at each locus are additive, we have (Wright 1951):

$$\begin{aligned}\sigma_{st}^2 &= 2 F_{st} \sigma_a^2 \\ \sigma_{is}^2 &= (1 + F_{it} - 2 F_{st}) \sigma_a^2 = (1 + F_{is})(1 - F_{st}) \sigma_a^2\end{aligned}$$

From which it can be deduced that F_{st} for quantitative traits (Q_{st}):

$$Q_{st} = \frac{\sigma_{st}^2(1 + F_{is})}{2\sigma_{is}^2 + \sigma_{st}^2(1 + F_{is})}$$

Assuming the variance component for open-pollinated (half-sib) families to estimate one-quarter of the additive genetic variance (Falconer and Mackay 1996), we have:

$$\begin{aligned}\sigma_{st}^2 &= \sigma_{pop}^2 \\ \text{and, } \sigma_{is}^2 &= 4\sigma_{op}^2\end{aligned}$$

Therefore, in the case of assortative mating or maternal effect, we will underestimate Q_{st} values as a result of overestimating σ_{is}^2 . Even with assortative mating or maternal effect, a value of Q_{st} greater than F_{st} is evidence of spatially divergent evolution by selection.

Intraclass correlations for maternal open-pollinated were estimated as follows:

$$t_{op} = \frac{\sigma_{op}^2}{\sigma_z^2}$$

Where σ_{op}^2 is the variance among open-pollinated families and σ_z^2 the total phenotypic variance within populations.

$4t_{op}$ provides the best estimate of heritability where there is no assortative mating or maternal effect.

Genetic correlations between pairs of traits (X and Y) were estimated over the entire experiment:

$$r_A = \frac{Cov_{op}(XY)}{\sqrt{\sigma_{op}^2(X)\sigma_{op}^2(Y)}}$$

Genetic variance covariance matrices (**G** matrices) as well as genetic correlation matrices, were calculated for each population.

Because the villages were chosen to maximize heterogeneity for socio-economic and maize production potential variables (Bellon *et al.* 2002; Smale *et al.* 1999), a fixed effect model was used to estimate variance and covariance components. All variance and covariance components were estimated by the method of moments. Mean squares were estimated using Proc GLM type III sums of square using the SAS software (SAS Online documentation at <http://www.sas.com>).

Comparison of population structure for neutral markers and quantitative traits

In the first step, F_{st} and Q_{st} were evaluated for all populations. Confidence intervals were calculated through jackknifing over populations for quantitative traits and over populations and loci for microsatellite markers. A Mantel test (Sokal and Rohlf 1995) was used to test for the independence of the pairwise F_{st} versus Q_{st} matrix. Pairwise F_{st} and Q_{st} matrices were used as genetic distance matrices to draw trees using the UPGMA clustering analysis.

Comparison of genetic G matrices

Matrices were compared using a multivariate analysis of variance (MANOVA) approach as described by Roff (2002). *Wilks' lambda* (Λ) as well as *Pillais' trace* (T) were examined to test for significance of village and population within village effects. Roff's method (2002) was chosen over Flury's (Flury 1988; Phillips and Arnold 1999) because it allows for nested comparisons and can easily be computed on standard statistical packages. Furthermore, maximum-likelihood methods like the Flury method can be very computer-intensive (Roff 2002), especially for very large data sets. Multivariate analysis of variance was carried out with a fixed effect model using a Proc GLM MANOVA statement under SAS software (SAS Online documentation at <http://www.sas.com>).

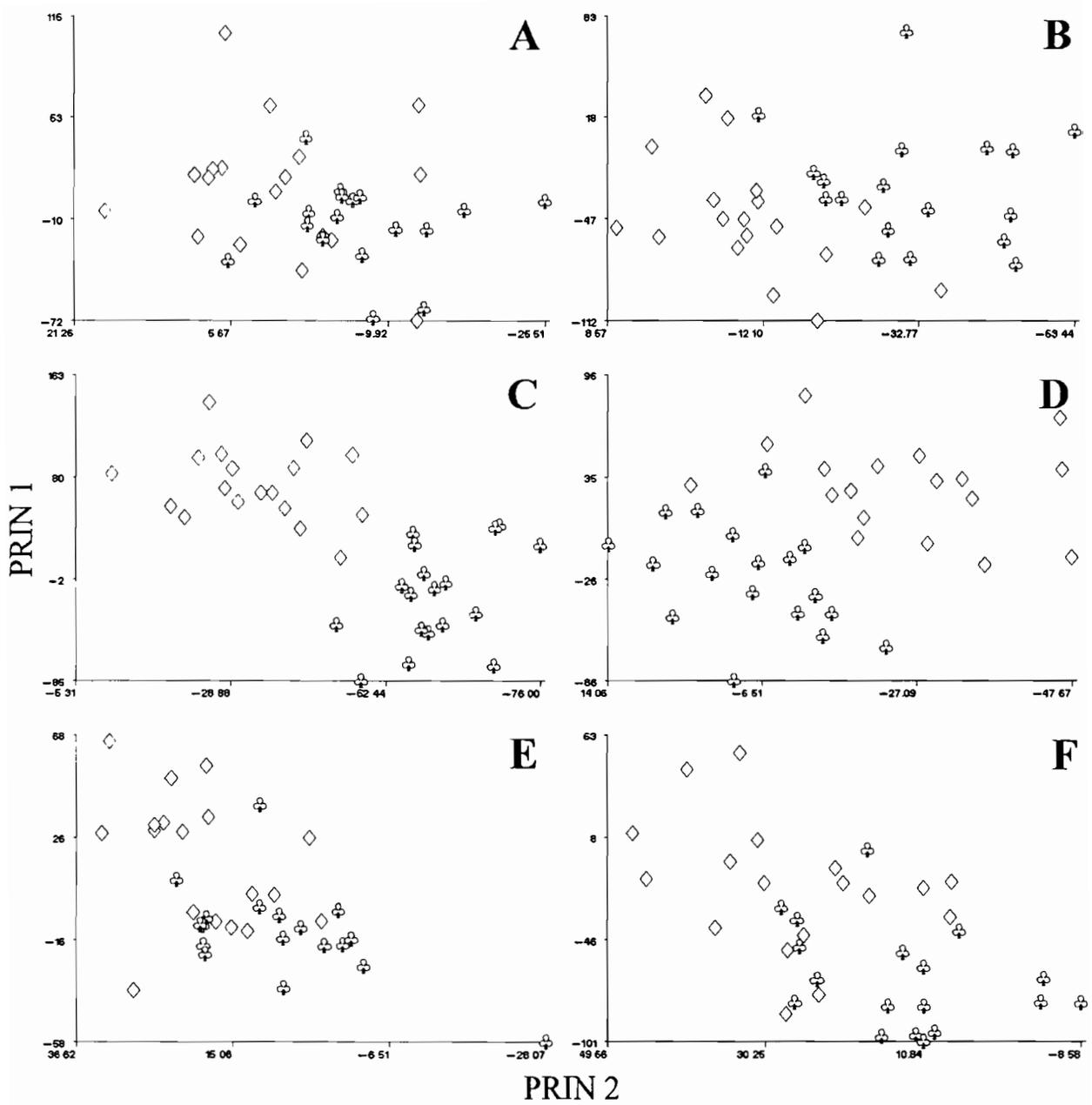


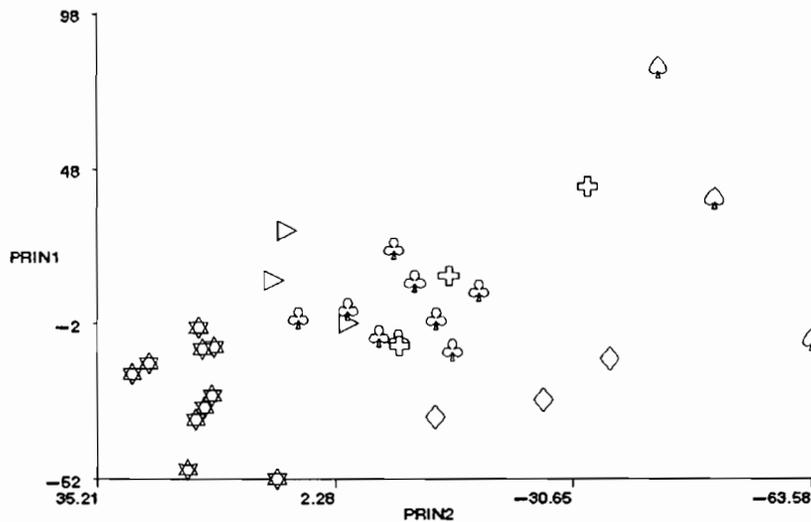
Figure 5.2 Principal Component Analysis for all measured ear and kernel traits was carried out on raw data set and then the principal components were averaged over families. Data are shown for couples of maize populations belonging to the same village. **A.** Populations 110 and 122 from Huitzo; **B.** 222 and 235 from Mazaltepec; **C.** 314 and 317 from San Lorenzo; **D.** 433 and 439 from Amatengo; **E.** 533 and 536 from Valdeflores; and **F.** 610 and 614 from Santa Ana. Populations were chosen because they showed little or no overlapping although they belonged to the same village.

RESULTS

Principal component analysis for quantitative traits

Principal component analysis conducted on ear and kernel traits clearly shows a grouping of population by village. The distribution is not uniform along the first principal components and we observe a grouping of populations by geographic (village) origin. In Figure 1 we give the mean value over populations of the principal components. Furthermore, we observe that some populations show little or sometimes no overlapping of their distribution, even though they were taken from the same village (one example from each village is given in Figure 2).

Figure 5.1 Principal Component Analysis for all measured ear and kernel traits on maize populations from the Oaxaca Central Valleys.



Principal component analysis was carried out on the raw data set and then the principal components were averaged over families and then over populations. Populations from six different villages include populations from Santa-Ana Zegache (star), Valdeflores (flag), Huitzo (club), Mazaltepec (diamond), Amatengo (cross), and San-Lorenzo (spade). Principal component 3 allows for better separation of Huitzo, Valdeflores, and Amatengo.

Among-population genetic structure for molecular markers

All tested SSRs loci were polymorphic in all populations. F_{st} values obtained for molecular markers are 0.011 within villages, 0.003 among villages, and 0.011 not considering the village level of hierarchy. There was no statistically significant evidence for isolation by distance. We observed low among-villages F_{st} values, significantly lower than within-village values.

Population structure using molecular markers has been described in more detail elsewhere (see companion paper).

Population structure for quantitative traits

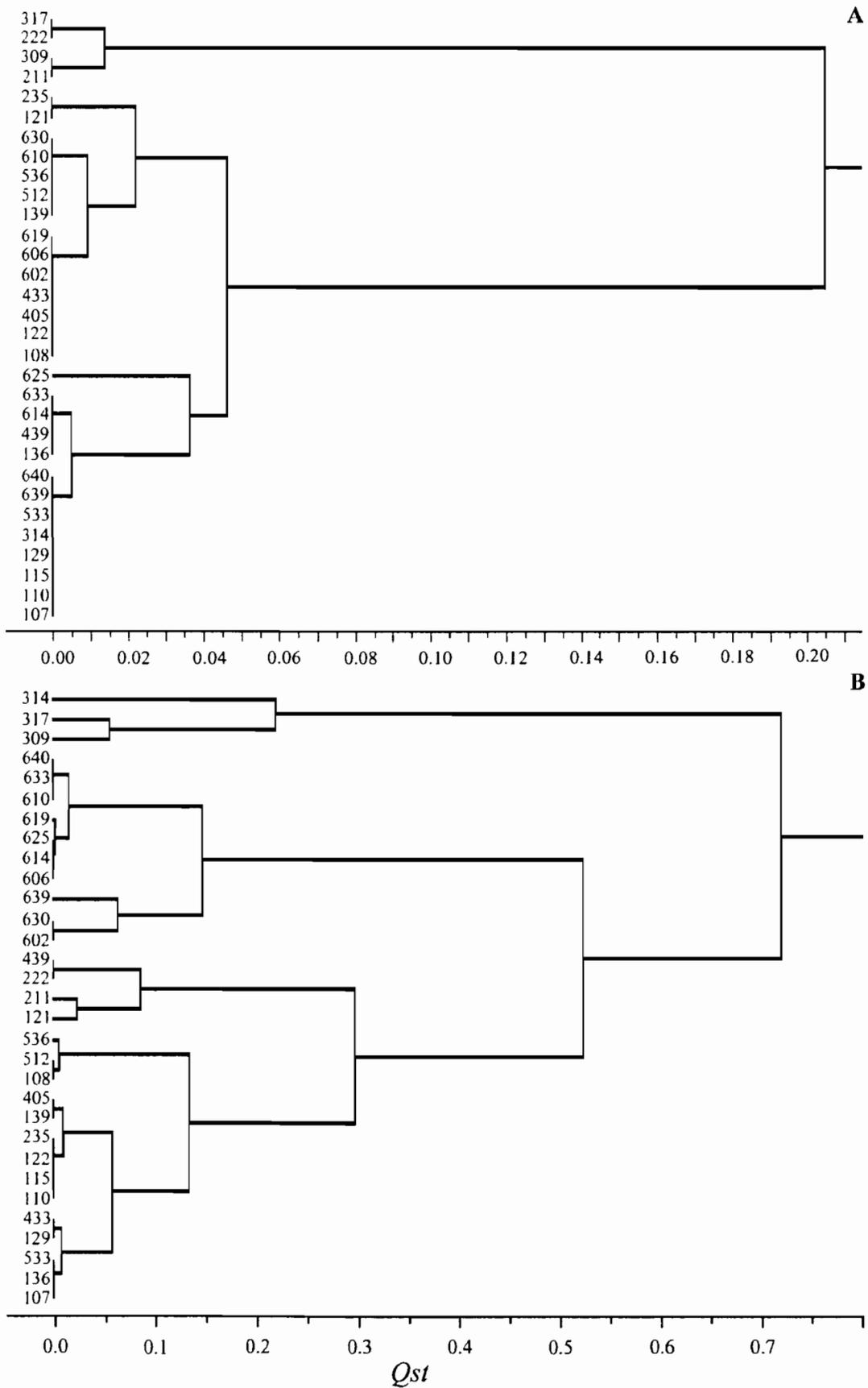
Despite considerable gene flow among populations and the resulting low level of population differentiation at these loci, considerable population differentiation occurs for quantitative traits. F_{st} values for quantitative traits (Q_{st}) are significantly higher than those observed for molecular markers (Table 1). Strong among-villages structure for a number of ear and kernel traits are observed (e.g., the among villages Q_{st} value for kernel weight is 0.535). Furthermore, the patterns of population structure differ greatly from one trait to another.

Table 5.1 Genetic variation among and within populations

		Population differentiation ^a			
		$4t_{op}$	Within villages	Among villages	Among populations
ear traits					
ear weight	(EW)	0.286	0.077	0.176	0.233
ear width	(EW _i)	0.474	0.090	0.064	0.151
overall grain weight	(GW)	0.274	0.071	0.099	0.159
number of grain per ear	(NG)	0.326	0.055	0.198	0.221
row number	(RN)	0.438	0.074	0.276	0.316
number of grain per row	(NGR)	0.286	0.031	0.066	0.081
kernel trait					
kernel thickness	(KT)	0.292	0.086	0.376	0.421
kernel weight	(KW)	0.510	0.040	0.535	0.523
cob traits					
cob weight	(CW)	0.659	0.053	0.314	0.327
cob width	(CW _i)	0.646	0.099	0.232	0.307
flowering traits					
days to anthesis	(DA)	1.488	0.056	0.085	0.130
days to silking	(DS)	1.258	0.065	0.103	0.154
anthesis silking interval	(ASI)	0.447	0.040	0.046	0.076
morphological traits					
plant height	(PH)	1.084	0.051	0.012	0.062
ear height	(EH)	1.115	0.029	0.040	0.060
Molecular markers					
SSRs all loci			0.011	0.003	0.011

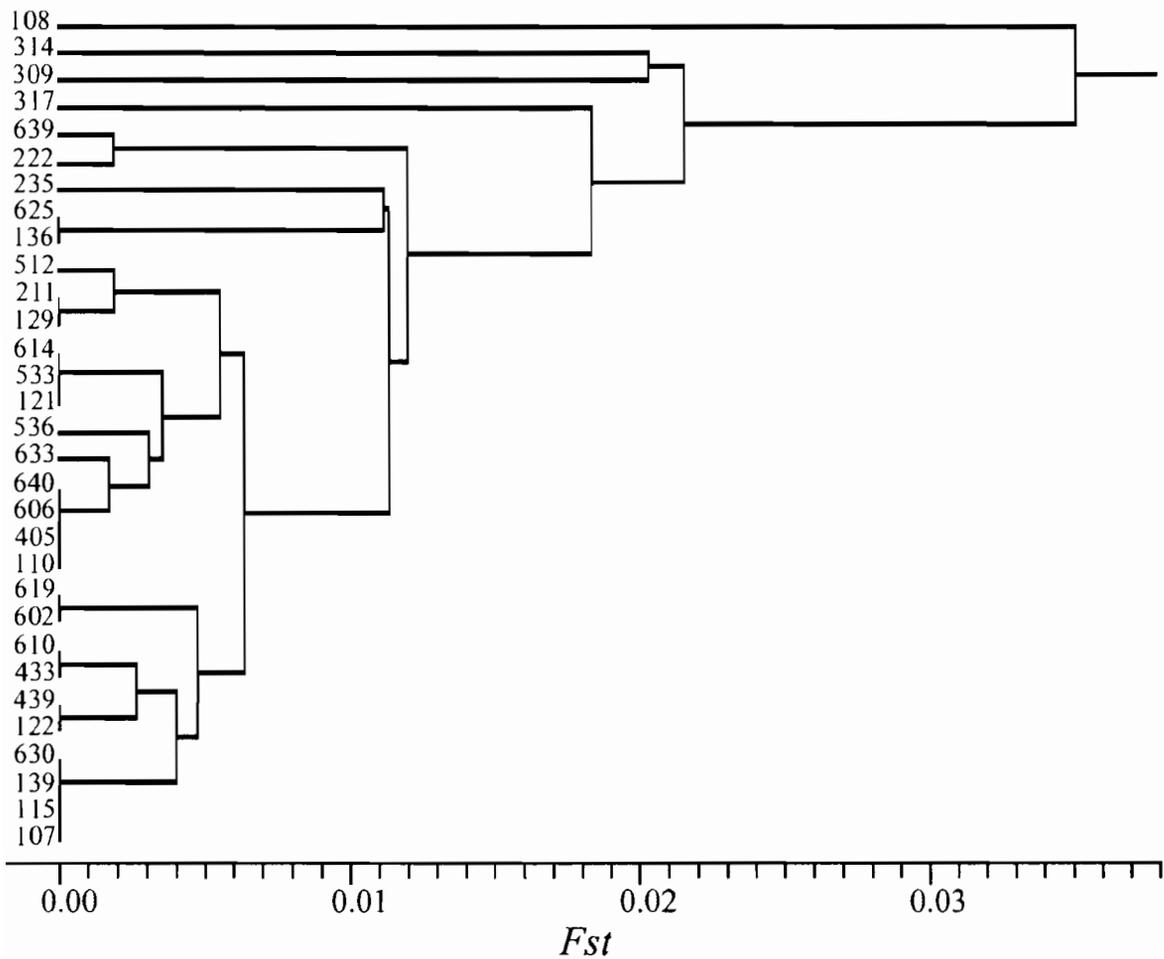
^a population differentiation: $F_{st}=\theta$ for molecular markers and Q_{st} for quantitative traits.

Figure 5.3 UPGMA tree using pairwise Q_{st} as genetic distance measures. **A.** for anthesis-silking interval (*ASI*) and **B.** for kernel weight (*KW*).



For traits showing the lowest within-village or among-villages Q_{st} values, a study of the pairwise Q_{st} matrix or trees built by UPGMA using Q_{st} as a distance, shows that a number of populations are considerably differentiated from the others. Examples are given in Figure 3 for anthesis-silking interval and kernel weight. For all traits, we observe high pairwise Q_{st} values in comparison to the pairwise F_{st} values measured for molecular markers, which show considerably less population differentiation (Figure 4).

Figure 5.4 UPGMA tree using pairwise $F_{st}=\theta$ as genetic distance measures.



In addition to Q_{st} being significantly higher than the F_{st} value for molecular markers, a Mantel test shows no significant correlation between pairwise Q_{st} and F_{st} values for any of the traits considered.

Heritability

Within population heritability estimates measured over the entire experiment are given in Table 1. We observe that four times the intraclass correlation ($4t_{op}$) for maternal open-

pollinated families clearly overestimates heritabilities for flowering traits. This could be the consequence of assortative mating for flowering traits. Assortative mating in maize landrace populations has been treated elsewhere (see companion paper).

Comparison of G matrices

The **G** matrix was calculated for each population. We examined variation of the **G** matrix, both at the population and village levels, following the method described by Roff (2002). The nested MANOVA shows a highly significant effect of population within village as well as among villages (Table 2). The test is highly significant ($p < 0.0001$) for both village and population within village for the variance covariance matrix as well as the correlation matrix, which indicates that nonproportional changes affect the structure of the variance covariance matrix. Results show that changes in matrix proportionality occurred both within- and among-villages. Element by element comparisons show a significant ($p < 0.001$) effect of population within-village as well as for village for the variance and covariance of all the studied traits. All traits examined in this study are either direct or indirect yield components and show Q_{st} values higher than F_{st} . Furthermore, we observe considerable changes in the **G** matrix structure, including cases where high correlations that are present in some villages or populations are low and no longer significant in others. No significant changes in the sign of correlations were noted.

Table 5.2 Variation in the **G** matrix over population and village

	Pillais' T^1	Approx. F	Num DF	Den DF	$p < F$
Population(Village)	24.519	217	2625	11175	<0.0001
Village	4.937	319	525	2135	<0.0001

¹Probabilities are the same for Wilks' Λ

DISCUSSION

Comparison of F_{st} values for molecular markers and quantitative traits

Some authors have raised concerns about the effects of epistasis and dominance on expected value of Q_{st} (Lynch *et al.* 1999). In the case of epistasis, Whitlock (1999) has shown that additive-by-additive epistasis would decrease the expectation of Q_{st} . Whether dominance will

increase or decrease the expected Q_{st} deserves further study. However, it is suggested by Whitlock (1999), that, averaged over a uniform distribution of allele frequencies, the contribution of dominance to Q_{st} approaches zero. Therefore, a value of Q_{st} greater than F_{st} for neutral loci is evidence for spatially divergent evolution by selection (Whitlock 1999).

Considering dominance in our case, except for yield and directly related traits such as ear weight and overall grain weight that show considerable dominance variance, all the other studied traits show low ratios of dominance to additive genetic (Hallauer and Miranda 1988).

Comparative population structure for molecular markers and quantitative traits

The Q_{st} values measured for quantitative traits are significantly higher than F_{st} values obtained for molecular markers, as expected in the case of diversifying selection. In this case, molecular markers do not provide any information about the different breeding potential of these populations. Most studies comparing population structure for molecular markers to that of quantitative traits have described situations in which $Q_{st} \geq F_{st}$ (McKay and Latta 2002) for a review). Their findings could be attributed to a bias because most of the authors looked at cases of diversifying selection. For the rare cases in which homogeneous selection was investigated, authors have found in some plant species that $Q_{st} < F_{st}$ (Petit *et al.* 2001). This suggests that in no case can F_{st} be considered as a conservative estimate of Q_{st} .

Selection in spite of gene flow or gene flow in spite of selection

Farmers have been shown to practice conservative selection on their maize populations based on a limited number of traits (Louette and Smale 2000). Our survey shows the same practice in the Oaxaca Central Valleys as farmers reportedly select a given ear type over generations. The characterization of high gene flow among these populations (see companion paper) and the strong divergent selection characterized in this study clearly show differentiation between populations for quantitative traits despite considerable gene flow and, reciprocally, considerable gene flow despite strong divergent selection by Oaxacan farmers. The development of differentiation between populations for quantitative traits despite considerable gene flow will occur if the number of migrants is large enough ($Nm > 1$) and when the migrant proportion is low enough ($m < s$) (McKay and Latta for a review). In other words, selection at a given loci will have little effect at other loci, unless they are closely linked (Barton and Bengtsson 1986).

Selection and changes in the G matrix structure

Under farmer management, maize landrace populations are subject to significant changes in their **G** matrices at both the within-village and among-villages levels of hierarchy. Nonproportional changes in the structure of the **G** matrix are associated with selection while drift would be responsible for proportional changes (Roff 2000). Therefore, our result showing nonproportional changes in the **G** matrix structure supports the result obtained by the comparison of the population differentiation for molecular markers to that for quantitative traits.

Populations with different **G** matrices are expected to respond very differently to multivariate selection. Therefore, population and villages with different **G** matrices will have different breeding potential.

Possible impacts of genes of major effect on quantitative trait variation

One of the key questions is whether variation at genes of interest is better estimated by variation at neutral loci or by variation for quantitative traits. McKay and Latta (2002) suggest that in the case of a polygenic trait, variation at QTLs is probably better reflected by variation at neutral markers than by the one seen for the quantitative trait. A condition to that statement is that trait value should result from the variation of many loci. However, genes or QTLs of major effect have been shown in numerous studies in maize and other species to account for a large part of the phenotypic variation. Almost half of the traits have a QTL accounting for at least 20% of the total phenotypic variation (Lynch and Walsh 1998 for a review). Agrawal *et al* (2001) have recently suggested that genes of major effect could dramatically alter the **G** matrix structure and therefore significantly alter the outcome of multivariate selection. The considerable changes that we observe in the **G** matrix structure both within-village and among-villages could be explained by farmer selection on genes of major pleiotropic effect. These changes in the **G** matrix structure and strong differentiation for quantitative traits among populations that are little differentiated for neutral markers could be used as a diagnostic to identify landraces to be screened for new allelic variation.

Implications for the management of genetic resources

Habitat fragmentation, resulting from the abandonment of traditional farming, could lead to gene flow erosion among traditional maize populations (Berthaud *et al.* 2001). Gene flow

erosion could increase the rate of extinction (Stockwell *et al.* 2003) among populations that are under strong selection pressure. A potential corrective measure that uses current strategies would take advantage of the absence of strong population structure at neutral markers and then favor the establishment of artificial gene flow. However, in our case, the population structure for quantitative traits is considerably higher than that observed for molecular markers. As a consequence, both sets of criteria (neutral markers and quantitative traits) should be taken into account when developing genetic resource management strategies. At the same time, determining a management unit for these strategies should also be based on the observed population structure, which is mainly based on phenotypic variation. Because for most traits much higher values were observed for Q_{st} (among-villages) than for Q_{st} (within-village), the village should serve as the unit for managing genetic resources.

Maize racial classification

Wellhausen *et al.* (1952) classified Mexican maize populations into major races. It is a taxonomic approach based on morphological traits. Many authors have used racial considerations to compare the diversity among maize landraces. All the landraces used in this study predominantly display the characteristics of the Bolita race as described by Wellhausen *et al.* (1952). However, we note the presence in a few landraces of some traits that would be typical of races such as Pepitilla or Zapalote chico (Bellon *et al.* 2002) though they never encompassed all the landraces of a given village. For this reason we will define Bolita as a complex rather than a race.

Our results strongly suggest that the morphological characteristics of a given landrace are either maintained or created by an active process resulting from farmers' and villages' selection criteria. Phenotypical resemblance or lack of resemblance cannot be considered as only reflecting historical processes of seed flow as assumed in the work of Wellhausen *et al.* (1952). It could be the product of either historical or convergent and divergent selection processes.

Farmer management results in phenotypic diversification in maize landraces

We report in this paper on the effects of farmers' management on population structure and differentiation for phenotypic traits as opposed to that seen for molecular markers. Not only do we observe considerably higher population differentiation for quantitative traits than for

molecular markers, but we also observe highly significant changes in the **G** matrix structure that imply changes of covariation and correlation among traits between different villages and also within villages. Both of these are responses to farmers' selection and management (indirect selection) practices. Although, to the best of our knowledge, it is the first time these techniques have been used for a crop species, they are increasingly used with wild species (McKay and Latta 2002 for a review) for the study of "contemporary evolution" (Stockwell *et al.* 2003). The integration of surveys of neutral and selected variation is therefore advisable for the study of contemporary evolution and before building genetic resources management strategies.

VI. Effect of kernel color in structuring genetic diversity in maize landraces

(Work in progress)

Abstract

This paper reports that phenotypic differentiation in maize populations from the Central Valleys of Oaxaca, Mexico, is an active and ongoing process involving a number of traditional management and selection practices by farmers. White and black kernel maize populations are managed differently by farmers and we show that these differences lead to differential organization of the genetic variation. Black kernel maize populations are little differentiated from white kernel maize populations, both for microsatellite and cytoplasmic markers. Nevertheless, they show a higher impact from drift because of smaller effective population size. Our research showed that less seed flow, frequent extinction and colonization events, and strong founder effects led to much higher differentiation among black maize populations than among white populations, especially for a maternally inherited marker. Regarding phenotypic differentiation, the populations are more differentiated for quantitative traits than for microsatellite markers (Q_{st} values are much higher than the F_{st} values). Strong multivariate divergent selection ($Q_{st} \gg F_{st}$ and nonproportional changes of the **G** matrix) takes place, and includes both indirect and direct selection by farmers. Black kernel maize populations show greater variability and high levels of population differentiation for flowering traits. These result from an unusual selection process: some farmers practice indirect selection for earliness by sowing late in the season while others maintain the early type by a population acceptance or rejection process.

Key words

F_{st} , **G** matrix, maize landraces, population structure, Q_{st} , quantitative traits

INTRODUCTION

In the Central Valleys of Oaxaca, Mexico, farmers manage maize populations with different kernel colors in distinct ways. Subjected to very different farmer management practices, these populations may show very different patterns of population differentiation for the different classes of markers and traits.

In a recent article, the authors described the patterns of population structure for neutral markers in maize landraces from the Central Valleys of Oaxaca, Mexico (Chapter IV). That article focused on white kernel maize because it represents 83% of the total maize grown in the area. The importance of farmer controlled seed flow was shown as it relates to the maintenance of genetic diversity in the Central Valleys of Oaxaca. Another article (Chapter V) described the population differentiation of white maize for quantitative traits, as opposed to the differentiation measured for molecular markers. We indicated how this difference is related to strong divergent selection that leads to phenotypic differentiation among white kernel maize populations, especially for ear and kernel traits.

Black kernel maize populations, like other colored kernel populations, are grown in much smaller sized plots than white kernel maize populations (0.17 ha versus 2.5 on average, Smale et al. 1999). Population sizes are consequently smaller for these populations and, in the Central Valley of Oaxaca, they face more frequent extinction and recolonization events than do white kernel maize populations. They are also less subject to the mixing of seeds from different sources, a common practice in white maize (Mendoza 2000). Population sizes, migration rates, and extinction and recolonization processes are expected to vary greatly between white and black kernel maize populations. Furthermore, farmers often report planting a newly acquired population (colonization event) over a very small plot for seed multiplication purposes. This practice probably only requires a very limited number of ears or reproductive female parents for founding the new black kernel maize population. Another reason the authors chose to work on black kernel maize populations was because Oaxacan farmers usually associate it with early maturity (Smale et al. 1999), which could also favor different management practices. As black kernel maize populations are managed differently than white, we investigated whether different farmer practices result in different spatial organization of the genetic and phenotypic variation than that observed for white kernel maize.

In addition to black kernel maize, colored populations also include purple, pinto, and yellow kernel types. While the yellow color is produced by an accumulation of carotenoids and is genetically controlled by the Y1 gene, purple and black kernel colors are produced by the accumulation of anthocyanin pigments in the aleurone layer. This trait is under the control of several genes (Coe et al., 1988). Among these, the C1 gene codes for a transcription factor required for the pigment synthesis. This gene shows dominant and recessive “colored aleurone” and dominant and recessive “colorless aleurone” alleles. The pinto maize type displays a wide range of kernel color phenotypes and includes individual ears with white, purple, and black kernels. Interestingly, it is usually derived from a white and black or purple hybridization.

In this paper, we assess the genetic diversity and population structure of white and black kernel maize populations from the Central Valleys of Oaxaca. Key agroecological factors and components of farmer management behind maize population dynamics are assessed. A description is provided about how this genetic diversity is structured and how management practices by farmers influence population differentiation. And, we compare the patterns of population differentiation for quantitative traits to those measured for molecular markers. Particular attention is given to flowering traits.

MATERIAL AND METHODS

Materials

We selected six villages for this study (same as in Chapter IV). Ten households per village were randomly chosen, giving a total of 60 surveyed farmers. Of these, six farmers cultivated a black kernel maize population and provided us with the corresponding maize ears. The white kernel maize populations were the same as those described in our earlier studies (Chapters IV and V). Thirty-one white kernel maize populations and the six black kernel populations were included in the molecular analysis. A subset of 12 white kernel maize populations was selected to cover the six previously mentioned villages and the genetic and phenotypic variation (among- and within-populations) present in white kernel maize (phenotypic variation reported in a previous article, Chapter V). These, along with the six black kernel populations, were phenotypically evaluated.

Genotyping

A total of 11 microsatellite markers were used for analysis (see companion paper). SSR primers were selected from the maizeDB database of public SSRs and included the following: phi011, phi227562, phi96100, phi101049, phi029, phi093, phi024, phi452693, phi034, phi014, and umc1061. Markers were selected according to their chromosomal locations, in order to provide for genome-wide coverage, and also by the size of the amplification product, to allow multiplexing on an automated DNA sequencer. All populations were also characterized with a chloroplastic marker. We used a polymorphic chloroplast set of primers surrounding a polyA repeat in the *psbK/psbI* intergenic region with the following sequences: zmc7430-F: CGAAGCTGCTGTAAGTTTTCG and zmc7430-R: AAGACTTCTCGGCTCTTATCCA (Provan et al. 1999).

Characterizing populations for quantitative traits

Eighteen open pollinated families were sampled for each population. Field layout was a two-replicate design with hierarchical structure (population plots randomly assigned and family plots randomly assigned within populations). The experiment was carried out at the CIMMYT experiment station at El Batán, Texcoco, Mexico, during the 2001 summer cycle and at the CIMMYT experiment station at Tlatizapán, State of Morelos, Mexico, during the 2001-2002 winter cycle. Up to 12 open pollinated progenies per family (18 families per population) were evaluated in each of the experiments (giving a total of $2 \times 18 \times 18 \times 12 = 7,776$ plants evaluated) for ear, kernel, and plant development traits. These included ear length (*EL*), ear width (*EW_i*), ear weight (*EW*), row number (*RN*), kernel count (*KC*), kernel count per rows (*KCR*), cob width (*CW_i*), cob weight (*CW*), overall grains weight (*GW*), kernel weight (*KW*), kernel width (*KW_i*), kernel depth (*KD*), kernel thickness (*KT*), days to silking (*DS*), days to anthesis (*DA*), anthesis-silking interval (*ASI*), plant height (*PH*) and ear height (*EH*). All measurements, except for days to silking and days to anthesis, were made at plant maturity. *DS*, *DA*, *ASI*, *PH* and *EH* were only measured at the El Batán location.

Data analysis

We estimated the expected unbiased heterozygosity (Nei 1987), $H_e = 2n(1 - \sum p_i^2) / (2n - 1)$, for microsatellite markers and, $H_e = n(1 - \sum p_i^2) / (n - 1)$, for cytoplasmic markers. The average number of alleles per locus for microsatellite markers and number of alleles for cytoplasmic

markers are also reported. A non-parametric method, the Wilcoxon-Mann-Whitney test ($p < 0.05$), based on the ranks of the observations, is used for testing if the heterozygosity is different in white and black kernel maize populations.

Overall F_{st} (θ , Weir and Cockerham 1984) was calculated for the entire set of 31 populations. θ values were estimated using GDA 1.1 software (Lewis and Zaykin 2002), which performs hierarchical F -statistics.

The apparent ratio of pollen to seed flow, according to Ennos (1994) is:

$$r = \frac{\left(\frac{1}{F_{STn}}\right)(1 + F_{IS}) - 2\left(\frac{1}{F_{STc}}\right)}{\left(\frac{1}{F_{STc}}\right)}$$

where F_{STn} is the F_{st} calculated for microsatellites markers and F_{STc} is calculated for cytoplasmic markers.

F_{st} for quantitative traits (Q_{st}) is (described with more details elsewhere, chapter V):

$$Q_{st} = \frac{\sigma_{st}^2(1 + F_{is})}{2\sigma_{is}^2 + \sigma_{st}^2(1 + F_{is})}$$

where, assuming the variance component for open-pollinated (half-sib) families to estimate one-quarter of the additive genetic variance (Falconer and Mackay 1996), we have $\sigma_{st}^2 = \sigma_{pop}^2$ and $\sigma_{is}^2 = 4\sigma_{op}^2$. Therefore, in the case of assortative mating or maternal effect, we will underestimate Q_{st} values as a result of overestimating σ_{is}^2 . Even with assortative mating or maternal effect, a value of Q_{st} greater than F_{st} is evidence of spatially divergent evolution by selection (Whitlock 1999).

Genetic correlations between pairs of trait (X and Y) were estimated over the entire experiment:

$$r_A = \frac{Cov_{op}(XY)}{\sqrt{\sigma_{op}^2(X)\sigma_{op}^2(Y)}}$$

Table 6.1 – Genetic diversity within black and white kernel maize populations

Village	Kernel color	Farmer's ID	SSRs		Cytoplasmic markers	
			H_e ¹	Number of allele/locus	H_e	Number of allele
Huitzo	white	107	0.70	6.09	0.19	2
Huitzo	white	108	0.63	5.09	0.42	3
Huitzo	white	110	0.67	6.00	0.44	2
Huitzo	white	115	0.68	5.82	0.52	2
Huitzo	white	121	0.67	6.00	0.47	3
Huitzo	white	122	0.68	6.09	0.27	2
Huitzo	white	129	0.69	5.36	0.51	2
Huitzo	white	136	0.65	5.36	0.39	2
Huitzo	white	139	0.70	5.91	0.51	2
Mazaltepec	white	211	0.68	5.91	0.19	2
Mazaltepec	white	222	0.67	6.18	0.48	2
Mazaltepec	white	235	0.66	6.00	0.51	3
San Lorenzo	white	309	0.67	5.91	0.69	3
San Lorenzo	white	314	0.64	5.27	0.47	3
San Lorenzo	white	317	0.65	5.55	0.52	2
Amatengo	white	405	0.68	6.00	0.57	3
Amatengo	white	433	0.68	5.82	0.64	3
Amatengo	white	439	0.69	5.91	0.66	3
Valdeflores	white	512	0.67	5.27	0.53	2
Valdeflores	white	533	0.67	5.82	0.52	2
Valdeflores	white	536	0.68	5.82	0.41	2
Santa Ana	white	602	0.69	5.55	0.51	2
Santa Ana	white	606	0.68	5.55	0.49	3
Santa Ana	white	610	0.70	5.91	0.34	2
Santa Ana	white	614	0.70	6.36	0.56	3
Santa Ana	white	619	0.68	5.64	0.51	2
Santa Ana	white	625	0.68	5.91	0.56	3
Santa Ana	white	630	0.70	6.09	0.51	3
Santa Ana	white	633	0.67	5.82	0.42	3
Santa Ana	white	639	0.69	5.82	0.62	3
Santa Ana	white	640	0.68	5.82	0.44	2
Huitzo	black	129	0.64	5.27	0.52	2
Mazaltepec	black	216	0.66	6.09	0.27	2
Amatengo	black	424	0.63	5.55	0.20	2
Santa Ana	black	614	0.67	5.82	0.52	2
Santa Ana	black	619	0.67	5.73	0.27	2
Santa Ana	black	630	0.69	6.00	0.00	1

¹ Les valeurs de H_e sont plus faibles que celles rapportées dans le chapitre IV : Du fait de la difficulté de distinguer certains allèles (identification), ils ont été ici considérés comme identiques. D'où un nombre inférieur d'allèles et un H_e plus faible.

Genetic variance covariance matrices (**G** matrices) as well as genetic correlation matrices were calculated for each population in each of the two environments (using the following traits: *EL*, *EW_i*, *EW*, *RN*, *KC*, *KCR*, *CW_i*, *CW*, *GW*, *KW*, *KW_i*, *KD* and *KT*).

All variance and covariance components were estimated by the method of moments. Mean squares were estimated using Proc GLM type III sums of square using the SAS software (SAS Online documentation at <http://www.sas.com>).

G matrices were compared using a multivariate analysis of variance (MANOVA) approach as described by Roff (2002). *Wilks' lambda* (Λ) and *Pillais' trace* (*T*) were examined to test for significance of population, color, environment, and their interactions effects. Roffs' method (2002) was chosen over Flury's (Flury 1988; Phillips and Arnold 1999) because it allows for nested comparisons and can be easily computed on standard statistical packages. Furthermore, maximum-likelihood methods, such as the Flury method, can be very computer-intensive (Roff 2002), especially for very large data sets. Multivariate analysis of variance was carried out using a Proc GLM MANOVA statement under SAS software (SAS Online documentation at <http://www.sas.com>).

RESULTS

Genetic variation within populations

Genetic diversity estimated over all populations is $H_e=0.71$ for microsatellite markers and $H_e=0.49$ for cytoplasmic markers. Black kernel maize populations show reduced levels of diversity when compared to the whites kernel maize populations (Table 6.1). The Wilcoxon-Mann-Whitney two sample test shows significant differences for H_e between the two compartments, both for microsatellite ($p<0.01$) and cytoplasmic ($p<0.05$) markers.

Among-population genetic structure

Principal coordinate analysis using data for microsatellite provides little evidence for population differentiation. Principal co-ordinates one and two together explain less than 5% of the total variation. Projection of the populations over the first two planes shows a uniform distribution and no grouping of populations. In addition, all populations overlap in their distribution. Kernel color does not appear to differentiate populations or maize types. F_{st} values obtained both for nuclear and cytoplasmic (chloroplast) markers are indicated in

Table 6.2. Black and white kernel maize populations share the same chloroplastic DNA haplotypes. The differentiation between black and white populations for chloroplastic markers is $F_{st}=0.02$. Nevertheless, population differentiation among black kernel maize populations is $F_{st}=0.449$ for this maternally inherited marker. It contrasts sharply with the value observed for markers with biparental inheritance. Among black kernel maize populations F_{st} is also higher than that observed for white kernel maize, but the difference is smaller than that seen for maternally inherited markers.

Table 6.2. Among-population genetic structure using molecular markers

		Among white pop.	Among Black pop.	Between white & Black pop.
θ	SSRs all loci	0.011	0.030	0.004
	Cytoplasmic markers	0.028	0.449	0.020
r		0.55	13	3

θ is the Wright F_{st} for molecular markers among or between groups (Weir 1996)

r is the apparent pollen- to seed-flow ratio (Ennos 1994)

Pairwise F_{st} values show that if black kernel populations are little differentiated from the whites as a group, individual black kernel populations show higher F_{st} , in paired comparisons, both with white and black populations (Table 6.3).

Table 6.3 – Pairwise comparison using Wright F_{st} .

	θ	Range	Average	Number of comparisons
Among white kernel maize populations		-0.013 – 0.067	0.011	465
Between black and white kernel populations		-0.009 – 0.063	0.023	186
Among black kernel maize populations		-0.004 – 0.061	0.030	15

Data provided using microsatellite markers (SSR) at all loci. $F_{st} = \theta$ (Weir 1996)

Table 6.4
Genetic variation among populations

Traits		El Batan			Tlaltizapan			Across sites		
		Q_{Wst}	Q_{Bst}	Q_{Cst}	Q_{Wst}	Q_{Bst}	Q_{Cst}	Q_{Wst}	Q_{Bst}	Q_{Cst}
ear development										
ear length	<i>EL</i>	0.161	0.117	-0,025	0.360	0.144	-0,019	0.202	0.202	-0,023
ear width	<i>EWi</i>	0.284	0.041	-0,015	0.194	0.023	-0,016	0.296	0.039	-0,019
ear weight	<i>EW</i>	0.295	0.006	-0,023	0.230	0.085	-0,021	0.321	0.055	-0,033
row number	<i>RN</i>	0.362	0.020	0,047	0.420	0.060	0,050	0.431	0.037	0,055
kernel count	<i>KC</i>	0.263	0.016	0,061	0.184	0.060	0,128	0.264	0.053	0,116
kernel count per row	<i>KCR</i>	0.067	0.029	0,018	0.117	0.043	0,086	0.099	0.054	0,064
cob width	<i>CWi</i>	0.308	0.193	-0,009	0.368	0.204	-0,036	0.397	0.221	-0,035
cob weight	<i>CW</i>	0.382	0.065	-0,033	0.318	0.123	-0,041	0.408	0.115	-0,045
overall grain weight	<i>GW</i>	0.247	0.000	-0,015	0.149	0.034	-0,013	0.240	0.028	-0,021
kernel development										
kernel weight	<i>KW</i>	0.605	0.090	0,110	0.474	0.302	0,042	0.636	0.223	0,084
kernel width	<i>KWi</i>	0.697	0.069	0,114	0.724	0.308	0,045	0.732	0.181	0,079
kernel depth	<i>KD</i>	0.204	0.100	-0,012	0.288	0.110	-0,005	0.251	0.136	-0,005
kernel thickness	<i>KT</i>	0.514	0.254	0,183	0.406	0.320	0,119	0.535	0.320	0,163
plant development										
days to anthesis	<i>DA</i>	0.236	0.352	0,221						
days to silking	<i>DS</i>	0.258	0.317	0,221						
anthesis-silking interval	<i>ASI</i>	0.075	0.006	0,039						
plant height	<i>PH</i>	0.136	0.205	0,235						
ear height	<i>EH</i>	0.137	0.262	0,178						

Q_{Wst} is the Wright F_{st} for quantitative traits among white kernel maize populations.

Q_{Bst} is the Wright F_{st} for quantitative traits among black kernel maize populations.

Q_{Cst} is the Wright F_{st} for quantitative traits between white and black kernel maize populations.

Population structure for quantitative traits

As shown in Table 6.4, F_{st} values for quantitative traits (Q_{st}) are, for most traits, conserved across environments. Q_{st} values among white kernel maize populations are similar to those measured over the 31 genotyped white kernel populations in a previous evaluation (Chapter V).

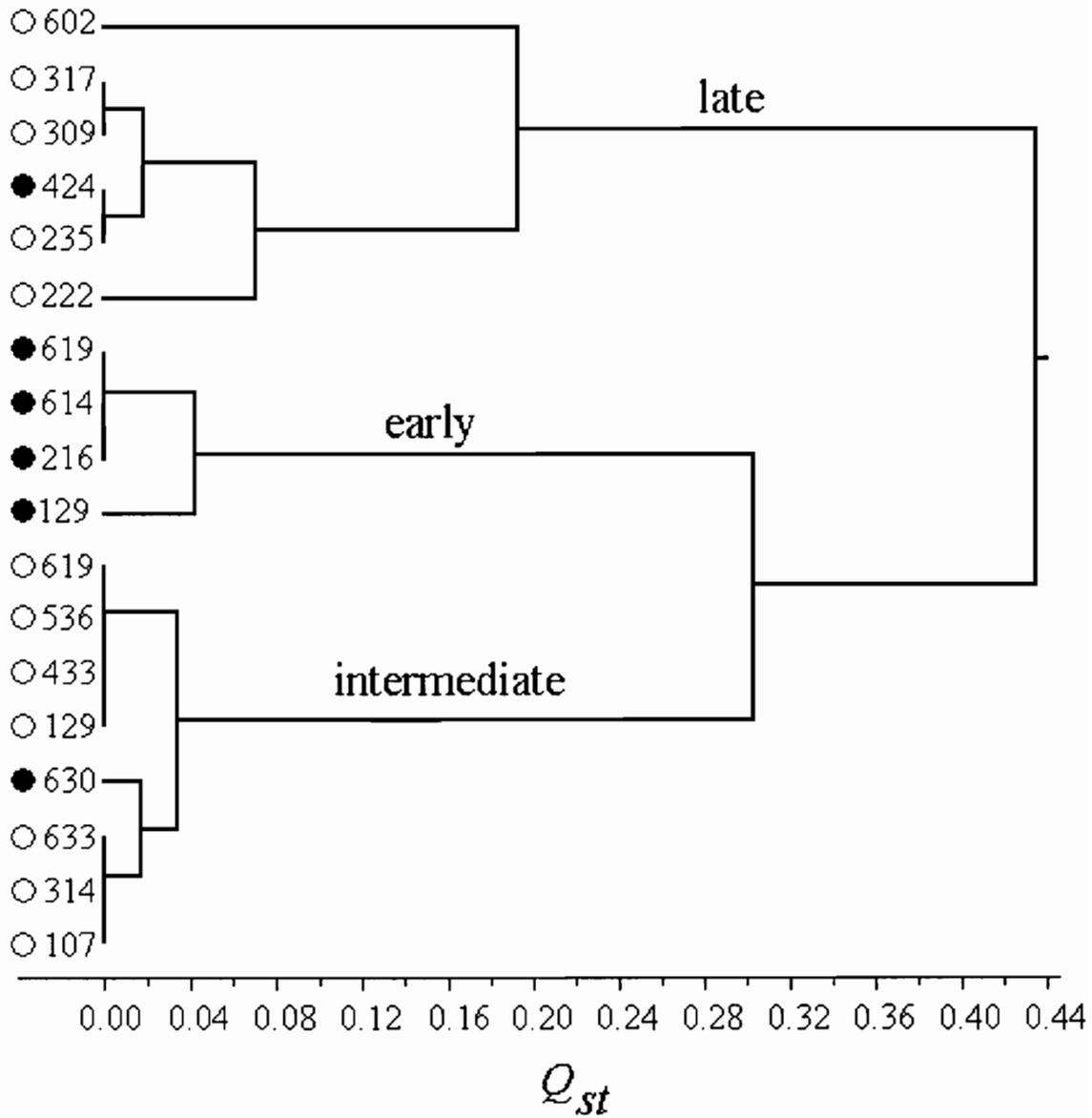
Q_{st} values are considerably higher than the F_{st} value measured with SSR markers. For all ear and kernel traits, the level of population differentiation is higher among white populations than among black. Nevertheless, there is little phenotypic differentiation between white and black kernel maize populations. The differentiation between white and black populations for kernel weight and kernel width can be attributed to the absence of a black kernel maize population from San Lorenzo in our sampling (non-white maize included mainly pinto types in this village). Indeed, these two traits strongly differentiate this village from the others (Chapter V).

The **G** matrix was calculated for each population, and variation of the **G** matrix was examined at both the population and kernel color levels, following the method described by Roff (2002). The nested MANOVA shows a highly significant effect ($p < 0.0001$) of population within maize type and between maize types. $G \times E$ is also significant ($p < 0.0001$) both for maize kernel color type and population. The tests are highly significant for the variance covariance matrix and the correlation matrix, which indicates that nonproportional changes affect the structure of the variance covariance matrix.

Population differentiation for flowering earliness

Results show that black kernel maize populations have a distinct pattern of population differentiation compared to white kernel maize populations (Table 6.4). This is especially true for flowering-related traits. We see a high differentiation between black and white kernel maize populations for development traits including days to silking (DS), days to anthesis (DA), plant height (PH), and ear height (EH). These traits show very high genetic correlation both within- and across-populations (tall plants are late maturing).

Figure 6.3 UPGMA tree using pairwise Q_{st} as genetic distance measures.



Data provided for days to anthesis. ● Black kernel maize. ○ White kernel maize.

As seen in Figure 6.3, black kernel maize populations cover the entire spectrum of flowering earliness. They include early maturing, intermediate, and late types. This contrasts with white kernel maize that do not show early maturing populations. The absence of early maturing white kernel maize populations was also seen over a much larger sample of white kernel maize populations in a previous study (Chapter V). Strong population differentiation is seen for this trait among black kernel maize populations ($Q_{st}=0.352$ among blacks, $Q_{st}=0.236$ among white populations, and $Q_{st}=0.221$ between black and white populations for days to anthesis).

Comparison of Q_{st} for ear and vegetative descriptors

Q_{st} values for ear and kernel descriptors tend to be higher among white populations than among black. The opposite is seen for vegetative descriptors (Figure 6.2). Results of the comparison related to ear and kernel descriptors are to be taken with caution because white kernel maize types were selected in the six villages to cover the whole range of phenotypic differentiation present with that color characteristic.

Figure 6.2 Comparison of Q_{st} values among white kernel maize populations (Q_{Wst}) and black kernel maize populations (Q_{Bst})

Ear and kernel descriptors: \diamond
 Vegetative descriptors: \blacklozenge

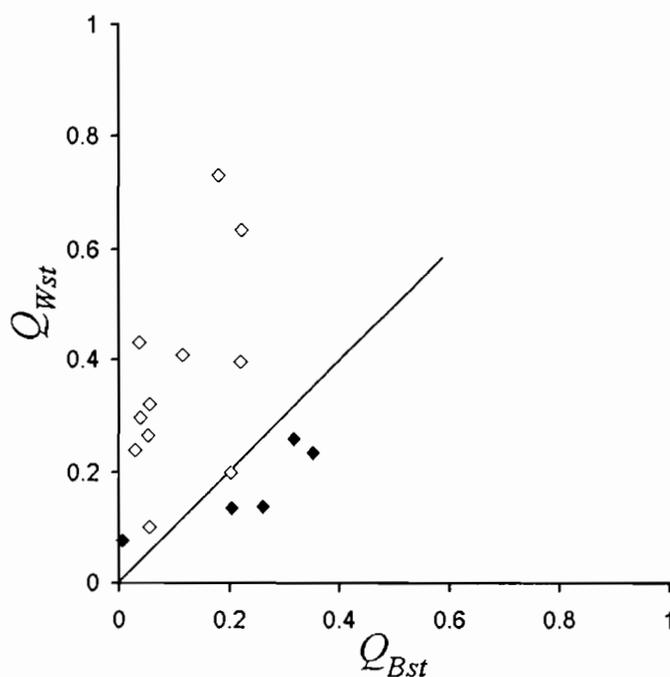


Table 6.5 – Within-population genetic and phenotypic variation for Days to Anthesis

Farmer's ID	Kernel color	Mean over families			σ^2_{op}	σ^2_{tot}	t_{op}
		mean	min	max			
107	white	83.029	75.833	88.818	8.296	21.590	0.384
129	white	84.123	81.333	88.182	2.873	17.563	0.164
222	white	94.176	87.727	99.417	7.640	22.588	0.338
235	white	89.843	83.083	98.417	19.937	34.030	0.586
309	white	91.601	85.250	97.000	6.052	21.652	0.280
314	white	82.554	79.091	86.636	1.944	20.359	0.095
317	white	91.941	87.083	98.000	7.426	28.160	0.264
433	white	83.867	80.000	89.889	4.461	16.754	0.266
536	white	83.800	78.917	88.500	7.170	20.616	0.348
602	white	87.098	83.667	89.750	1.860	16.535	0.112
619	white	84.566	80.167	90.667	7.176	18.284	0.392
633	white	82.518	73.667	87.250	6.902	17.836	0.387
129	black	75.286	71.750	85.083	8.812	19.318	0.456
216	black	77.235	74.000	84.636	6.746	13.725	0.492
424	black	89.661	82.900	94.500	8.563	24.604	0.348
614	black	77.627	73.417	82.077	4.734	15.148	0.313
619	black	77.833	73.583	80.273	1.911	12.835	0.149
630	black	81.169	76.455	85.750	5.735	17.241	0.333

Where σ^2_{op} is the among-family variance (open pollinated or half-sibs families), σ^2_{tot} is the total phenotypic variance, t_{op} is the intraclass correlation or the fraction of phenotypic variance attributable to the differences among open-pollinated families. $4t_{op}$ provides the best estimate of heritability where there is no assortative mating or maternal effect.

Within population variation for flowering traits

Aside from the population mean, the within-population genetic variance varies greatly from one population to another (Table 6.5). One population (the white kernel maize population from Farmer 235) even shows a bimodal distribution, which suggests the recent admixture of populations with different flowering earliness. $4t_{op}$ is above 1 in most of the populations considered. This can be explained by assortative mating, which increases the resemblance among individuals of a given open pollinated family. Flowering-related assortative mating has already been characterized in a previous article for the white kernel populations from the Central Valleys of Oaxaca (Chapter IV). As a result, the accumulation of favorable earliness alleles is unlikely to be uniform across a given population, even for an “early” population, because of the inner-population structure associated with this trait.

DISCUSSION

Gene flow among and between black and white kernel maize populations

There are many differences in farmers’ management practices in black kernel maize populations, including, (i) smaller populations, (ii) less seed mixing and more frequent extinction and subsequent recolonization, and (iii) recolonization events, sometimes involving a very limited number of female founders (Mendoza 2000; Smale et al. 1999). These differences have consequences for the spatial organization of the genetic variation at neutral markers. Indeed, F_{st} is much higher among black kernel maize populations, both for biparentally and maternally inherited markers, than what is observed in white kernel maize populations (F_{st} for maternally inherited markers is very high among black kernel maize populations with $F_{st}=0.49$). In addition, in black kernel maize populations, we observe reduced levels of genetic diversity (H_e lower in black kernel maize populations than in white). Strong fluctuations of the frequencies of maternally inherited haplotypes and reduced levels of genetic diversity are expected in cases of founder effects, which is compatible with the above observations of farmers’ practices. Black kernel maize populations seems to have considerably less seed-flow among them than what is seen in the white kernel maize populations (Chapter IV). In addition, they could face a strong founder effect because they are adopted by a farmer who is increasing the population from a very limited number of ears or female parents.

The very little population differentiation between the two groups (black and white populations) poses the question of possible seed flow between them. Two hypotheses for this can be made. First, they could diverge from a common ancestral population without any bottleneck. Second, there is a significant amount of seed flow between black and white populations. While the first hypothesis is unlikely in the light of farmer practices (seed multiplication from a very limited number of ears after adopting a new population), some farmer practices favor the second hypothesis. Indeed, farmers reported the emergence of populations of the pinto type from crosses between white and black kernel maize populations, while some reported the “extraction” of either a pure white or black type from the pinto populations. These practices would allow seed flow between black and white populations. The pinto type would then appear as a “seed-flow bridge” between them. In addition, the genetic control of kernel color at the *C1* locus shows both dominant and recessive “colorless aleurone” as well as dominant and recessive “colored aleurone” alleles allowing gene flows to go unnoticed by the farmers (for both pollen and seed flow).

Kernel color does not appear to be a barrier to gene flow for the studied populations of the Central Valleys of Oaxaca, Mexico.

Divergent selection for flowering earliness among black kernel maize populations

High population differentiation is seen among white maize populations for ear and kernel traits and among black maize populations for vegetative traits. These Q_{st} values are higher than the F_{st} values estimated for the same sets of populations. Therefore, divergent selection takes place in farmers’ fields as a value of Q_{st} greater than F_{st} for neutral loci is evidence for spatially divergent evolution by selection (Whitlock 1999). In addition, nonproportional changes in the **G** matrix structure are also the result of a divergent selection process (Roff 2000).

In this article, we see evidence of strong divergent selection for flowering-related traits in black kernel maize populations. As mentioned earlier, the lower population differentiation for ear and kernel traits among black maize than among white maize populations is to be taken with caution because it could be partly linked to the sampling. Rather, we will emphasize the difference between black and white populations for vegetative traits.

How do farmers select differently and for different traits in black and white populations? What do we know about farmers’ practices in these two groups with special regards to

flowering traits? While farmers reported cases of late planting for black kernel maize populations, none were reported for white maize. Thus, the white kernel maize populations being planted early in the season will take advantage of the full rainy season. By contrast, black kernel maize populations are subject to various planting dates. Populations sowed late in the season will have a shorter rainfed period. Late planting will generate selection pressure for early maturity. But not all early black kernel maize populations are planted late. For the farmer, late planting of an early maturing population provides the security of harvesting “something,” even if the main white kernel maize population suffers drought at an early stage of development. This approach also allows farmers to sow black maize in areas that were not planted to white maize earlier in the season, thereby taking full advantage of all of his or her arable land. Planting an early population early in the season provides the farmer with a harvest before the white maize matures. Therefore, both practices provide economic incentives for the farmers.. Farmers practice selection by acceptance and rejection while testing a new population. This testing is, as previously mentioned, is often performed in a very small plot, and allows for seed increase if the farmer decides to adopt the population. The farmer bases his or her decision to adopt on many more traits than usually used to select ears for seed production in a typical cycle (Mendoza 2000). In contrast to a typical selection, he or she will evaluate the newly acquired population for flowering precocity.

In a previous article, the authors described strong divergent selection in white kernel maize populations for ear and kernel traits (chapter V). This divergent selection could easily be related to farmers’ post-harvest selection of maize ears used to plant the next cycle. Farmers do not select for maturity or flowering earliness in the fields. Nevertheless, we observe that black kernel maize populations face strong divergent selection for flowering earliness. This results from an unusual selection procedure: some farmers practice indirect selection for earliness by sowing late in the season while others maintain the early type by a population acceptance or rejection process.

Conclusions

Through a series of articles (this chapter and Chapters IV and V), we see how farmer management allows maize populations from the Central Valleys of Oaxaca to be highly differentiated phenotypically, in spite of high migration rates among them. This phenotypic differentiation is the result of a strong divergent selection process promoted either directly by the farmers or occurring in farmers’ fields. Multivariate differential divergent selection for all

the considered traits of agronomic importance was observed, with significant changes in the genetic correlations among them. Divergent selection occurs for ear and kernel traits in white kernel maize populations (Chapter V), and flowering-related traits in black kernel maize populations (this chapter). The traditional maize agroecosystem appears to be a very patchy and heterogeneous environment with very different selection pressures occurring from one patch to the next, which allows phenotypic diversification under farmers' divergent selection pressures.

Today's maize geneticists and breeders have at their disposal a set of long-term evolutionary genetic experiments carried out by small-scale farmers throughout post-domestication era of maize. The remarkable genetic and phenotypic diversity of maize should be viewed in light of this contemporary evolution. Our work on maize populations from the Central Valleys of Oaxaca, Mexico, in the same region where maize was domesticated (Benz 2001; Piperno and Flannery 2001) and in the heart of the species' center of diversity (Matsuoka et al. 2002), strongly suggests that its phenotypic diversity is likely to reflect a post-domestication history of farmers' (past and present) divergent selection. As the organization of genetic and phenotypic variation tells a tale of these divergent selection processes, identifying population and germplasm material for the identification of alleles of agronomic interest should consider its patterns.

VII. Etude du déséquilibre de liaison et de la différenciation entre populations pour des marqueurs disposés le long du gène *Shrunken 2*

INTRODUCTION

La nature du déséquilibre de liaison entre un polymorphisme donné et un polymorphisme responsable de la variation phénotypique impose une contrainte majeure aux études de génétique fonctionnelle. En effet, il est indispensable de disposer de marqueurs étroitement associés au phénotype afin d'obtenir des renseignements sur l'évolution phénotypique et fonctionnelle d'une espèce. Le déséquilibre de liaison dépend essentiellement des phénomènes de dérive, de sélection, et d'évènements de recombinaison. Il a été démontré que ces derniers interviennent de manière préférentielle à l'intérieur des gènes (Fu et al 2002, Thuriaux 1977). Chez le maïs plus spécifiquement, la décroissance rapide du déséquilibre de liaison à l'intérieur des gènes serait la résultante de l'importance de l'effectif efficace de cette espèce tout au long de son évolution et d'un taux considérable de recombinaisons intragéniques (Remington et al 2001, Tenaillon et al 2001). Cette décroissance rapide du déséquilibre de liaison, vérifiée dans l'ensemble du génome du maïs, permet d'établir des corrélations très fines entre génotype et phénotype (Remington et al 2001, Thornsberry et al 2001).

Le maïs est un modèle pour la recherche portant sur l'évolution phénotypique (Doebley 1998) ainsi que pour l'étude de l'évolution du génome chez les plantes (Gaut et al 2000). L'association du phénotype et du génotype y est possible avec une très grande résolution (Remington et al 2001, Thornsberry et al 2001). Cette forte résolution ne peut être atteinte que s'il y a une décroissance rapide du déséquilibre de liaison. Toutefois, il faut remarquer que cette chute du déséquilibre de liaison a été observée à partir d'échantillonnages couvrant la majorité de l'aire de répartition du maïs (Remington *et al.* 2001; Tenaillon *et al.* 2001 et 2002; Thornsberry *et al.* 2001 ; Wang *et al.* 1999). Il nous a paru intéressant de vérifier si un tel phénomène est observable à l'échelle, considérablement plus réduite, des Vallées centrales de Oaxaca, et s'il est d'ampleur comparable avec ce qui a été obtenu lors des travaux antérieurs.

L'ensemble des populations de maïs d'une région telle Oaxaca est un système non cloisonné en terme de flux de gènes, assimilable, de par sa structure, à une métapopulation (Chapitre IV). Les populations traditionnelles de maïs, comme décrit au cours de chapitres précédents (Chapitres V et VI), sont soumises à une importante sélection divergente. Sélection portant de façon différentielle sur les différents caractères du grain et de l'épi (chapitre V) ainsi que la

floraison (chapitre VI). De la sélection divergente résulte que la différenciation entre populations pour les caractères quantitatifs est supérieure à celle de marqueurs moléculaires neutres ($Q_{st} \gg F_{st}$). Qu'en est-il pour des polymorphismes possiblement associés à la variation phénotypique donnant prise à la sélection ? Un tel polymorphisme serait soumis non seulement à la dérive et à la migration mais aussi à la sélection portant sur le caractère correspondant. Quelle est alors la conséquence de cette sélection sur les profils de distribution de la variabilité des polymorphismes de ce gène ? La structuration de la diversité pour des gènes ou locus devrait refléter la sélection à laquelle ils sont soumis. Par conséquent, elle est supposée se démarquer du profil de la diversité obtenu en se basant sur les marqueurs neutres, et nous fournir la matière à une comparaison. Néanmoins, dans la plupart des cas, les caractères quantitatifs sont dépendants d'un nombre élevé de locus distincts. Un caractère agromorphologique est donc susceptible de résulter de l'effet d'un certain nombre de QTLs (McKay et Latta 2002). La prise en compte de cette dernière remarque dans le contexte de notre étude nous amène à nous demander si l'évaluation et la structuration de la diversité aux QTLs d'intérêt agronomique est plus fidèlement représentée par la variation observée au niveau des marqueurs neutres ou au niveau phénotypique. Les implications des techniques d'évaluation de la diversité génétique, notamment au niveau des gènes ou locus associés à la variation phénotypique, seront discutées ; Ceci, dans le but de comprendre les mécanismes ayant façonné la variabilité phénotypique (variabilité pour les QTLs et les caractères quantitatifs) et rechercher l'origine de celle-ci.

Le gène *Shrunken2* (*Sh2*) est exclusivement exprimé dans l'albumen du grain de maïs. Il code pour la grande sous-unité de l'adénosine-diphosphoglucose-pyrophosphorylase ou AGPase, qui est l'enzyme catalysant la première étape de l'unique voie de synthèse des amidons (Nelson et Pan, 1995 pour revue). Cette position cruciale confère potentiellement à *Shrunken2*, qui est lié à un QTL majeur, à fort effet additif, influant sur les concentrations en protéines et en huile du grain (Goldman et al., 1993), un fort impact sur les caractéristiques du grain. De fait, des expériences de mutagenèse par insertion au niveau de l'exon 16 ont abouti à une augmentation du poids du grain (Giroux et al. 1996). Cet exon code pour la région permettant la dimérisation de l'enzyme (la petite sous-unité étant codée par le gène *Brittle2* interagissant avec la protéine codée par *Shrunken2* au niveau de la partie codée par l'exon 16, Greene et Hannah 1998a, Greene et Hannah 1998b), aussi est-il vraisemblable que des polymorphismes localisés dans cette région puissent être associés à des variations phénotypiques. De même, on peut penser que des polymorphismes des régions promotrices et

Tableau 7.1 – Déséquilibre de liaison par paire de loci

Population	<i>Sh2-1 et Sh2-2</i>		<i>Sh2-1 et Sh2-3</i>		<i>Sh2-2 et Sh2-3</i>	
	<i>R_{ij}</i>	<i>Pr</i>	<i>R_{ij}</i>	<i>Pr</i>	<i>R_{ij}</i>	<i>Pr</i>
107	0.294	0.118	0.257	0.244	0.308	0.196
108	0.104	0.865	0.065	1.000	0.274	0.165
110	0.543	0.002	0.478	0.013	0.300	0.111
115	0.226	0.373	0.239	0.343	0.192	0.508
121	0.423	0.052	0.112	0.674	0.120	0.745
122	0.238	0.320	0.253	0.306	0.386	0.020
129	0.124	0.710	0.186	0.626	0.308	0.088
136	0.163	0.610	0.195	0.534	0.201	0.468
139	0.049	0.978	0.415	0.039	0.196	0.504
211	0.111	0.800	0.180	0.466	0.285	0.235
222	0.402	0.022	0.484	0.018	0.462	0.015
235	0.230	0.362	0.421	0.139	0.356	0.108
309	0.273	0.144	0.429	0.004	0.266	0.173
314	0.465	0.057	0.254	0.311	0.288	0.265
317	0.452	0.026	0.495	0.006	0.262	0.201
405	0.445	0.017	0.215	0.472	0.189	0.546
433	0.178	0.607	0.184	0.547	0.273	0.242
439	0.272	0.180	0.082	0.918	0.192	0.498
512	0.421	0.113	0.214	0.510	0.211	0.455
533	0.117	0.839	0.231	0.420	0.287	0.254
536	0.329	0.086	0.165	0.627	0.355	0.102
602	0.149	0.765	0.139	0.690	0.467	0.016
606	0.298	0.105	0.258	0.308	0.207	0.543
610	0.220	0.397	0.271	0.217	0.169	0.620
614	0.228	0.368	0.144	0.719	0.245	0.269
619	0.119	0.934	0.203	0.463	0.322	0.070
625	0.107	0.892	0.270	0.191	0.341	0.064
630	0.147	0.707	0.059	0.859	0.263	0.235
633	0.195	0.477	0.041	0.941	0.416	0.058
639	0.264	0.255	0.099	0.943	0.228	0.276
640	0.087	0.873	0.198	0.449	0.352	0.110

Déséquilibre de liaison

Les valeurs des corrélations intra-populations par paire d'allèles (R_{ij} ; i allèle considéré au locus 1, j au locus 2) selon Weir et Cockerham (1977) ont été calculées pour chaque paire de locus. Un test par permutation à l'intérieur de chaque population (test de la valeur réelle par rapport aux valeurs des permutations) a été réalisé en utilisant le logiciel *GENETIX* 4.02 (Belkhir et al. 2001). Nous avons également calculé les composantes d'Ohta (statistiques D d'Ohta) du déséquilibre de liaison (Ohta 1982) qui permet de calculer les composantes intra- et inter- populations de la variance du déséquilibre (D).

RESULTATS

Déséquilibre de liaison

L'étude des différentes paires de marqueurs le long du gène *Shrunken2* nous a conduit à mettre en évidence 12 déséquilibres de liaisons significatifs sur 93 observations réalisées (Tableau 7.1). Par contraste, les résultats de l'étude des marqueurs SSRs n'avaient pas montré une telle proportion de déséquilibre de liaison significatifs (chapitre IV).

Néanmoins, nous observons que la répartition de ces déséquilibres significatifs entre les différentes populations est hétérogène (Tableau 7.1). On peut remarquer à ce titre que la population 222 se distingue fortement des autres par un déséquilibre significatif pour tous les couples de marqueurs. Dans la plupart des populations, le déséquilibre de liaison n'est pas significativement différent de zéro.

De plus, malgré le fait que dans certaines populations le déséquilibre de liaison est significatif, les valeurs observées sont faibles. Sur l'ensemble des populations, les coefficients de corrélation entre les paires d'allèles sont encore inférieurs aux coefficients calculés à l'intérieur de chaque population. Sur l'ensemble des populations, nous obtenons ainsi $R_{ij}=0.13$ entre *Sh2-1*, *Sh2-2*, $R_{ij}=0.096$ entre *Sh2-1* et *Sh2-3*, et $R_{ij}=0.144$ entre *Sh2-2* et *Sh2-3*.

Tableau 7.3 – Structuration des populations

Locus	localisation		
	chromose (bin)	f	θ
Marqueurs microsatellites (SSR)			
phi 011	1.09	0.261±0.039	0.015±0.009
phi 227562	1.11	0.059±0.025	0.009±0.004
phi 96100	2.00	0.075±0.029	0.017±0.008
phi 101049	2.08	0.092±0.024	0.003±0.003
phi 029	3.04	0.126±0.046	0.004±0.007
phi 093	4.08	0.143±0.020	0.011±0.008
phi 024	5.01	0.252±0.035	0.004±0.005
phi 452693	6.04	0.347±0.029	0.011±0.008
phi 034	7.02	0.079±0.027	0.010±0.006
phi 014	8.04	0.093±0.037	0.019±0.009
umc 1061	10.06	0.002±0.021	0.020±0.010
SSRs (tous)		0.138±0.031	0.011±0.002
Marqueurs <i>Sh2</i>			
<i>Sh2</i> -1	3.09	0.248±0.033	0.021±0.012
<i>Sh2</i> -2	3.09	-0.002±0.035	-0.003±0.005
<i>Sh2</i> -3	3.09	0.163±0.039	0.005±0.008

Estimateur du F_{is} , f , et du F_{st} , θ_{st} , [Weir et Cockerham 1984]. Intervalle de confiance à 95% obtenu par la procédure du jackknife sur les populations pour chaque locus et jackknife sur les loci pour les valeurs sur l'ensemble des marqueurs microsatellites.

Composantes d'Ohta de la variance du déséquilibre de liaison

Les valeurs des composantes d'Ohta calculées pour nos marqueurs le long de *Shrunken2* sont semblables à celles obtenues pour les couples de marqueurs SSRs (Tableau 7.2). Trois conclusions peuvent être tirées de ces résultats ($D'_{is} > D'_{st}$ et $D_{is} > D_{st}$, pour tous les marqueurs microsatellites et *Shrunken2*). (i) La valeur du déséquilibre de liaison est faible (D_{it}). (ii) La composante intra-population du déséquilibre de liaison est supérieure à la composante inter-population ($D_{is} > D_{st}$). (iii) Mais les associations alléliques ne sont pas les mêmes d'une population à l'autre ($D'_{is} > D'_{st}$). Cette dernière remarque est très importante pour *Shrunken2* ; si, comme vu précédemment, le déséquilibre peut être important localement (dans certaines populations, e.g. 222), il n'y a pas d'association allélique systématique.

Tableau 7.2 – Composantes d'Ohta du déséquilibre de liaison

Locus	D_{IS}	D'_{IS}	D_{ST}	D'_{ST}	D_{IT}
Marqueurs SSRs *					
étendue variation	0.017-0.045	0.08-0.124	0.009-0.022	0.000-0.003	0.082-0.126
moyenne	0.032	0.100	0.017	0.001	0.102
Marqueurs <i>Sh2</i>					
<i>Sh2</i> -1 et <i>Sh2</i> -2	0.022	0.095	0.019	0.004	0.099
<i>Sh2</i> -1 et <i>Sh2</i> -3	0.023	0.098	0.020	0.009	0.107
<i>Sh2</i> -2 et <i>Sh2</i> -3	0.025	0.064	0.013	0.016	0.080

* Moyenne et étendue de la variation pour les marqueurs SSRs calculés pour 55 couples (11 locus)

Statistiques de Wright (F_{st} et F_{is})

Les valeurs du F_{st} et du F_{is} présentent des différences significatives entre les différents marqueurs étudiés le long de *Sh2* (Tableau 7.3). Ainsi, le marqueur *Sh2*-2 se trouve dans une situation proche des proportions de Hardy-Weinberg (F_{is} non significativement différent de zéro) et ne présente pas de différenciation significative entre populations (F_{st} non significativement différent de zéro). Par contraste, le marqueur *Sh2*-1 présente le F_{st} le plus élevé de tous les marqueurs étudiés et une valeur F_{is} élevée.

DISCUSSION

Les populations traditionnelles de maïs cultivées dans les vallées centrales de Oaxaca constituent un échantillonnage de choix pour une étude à grande résolution de l'association entre génotype et phénotype. En effet, ce système présente simultanément (i) une diversité au niveau des marqueurs neutres comparable à celle observée à l'échelle de l'espèce (chapitre IV, Matsuoka *et al.* 2002), (ii) une absence de cloisonnement vis-à-vis des flux de gènes (chapitre IV), (iii) une importante différenciation phénotypique (Q_{st} élevés, chapitres V et VI) et (iv) une chute du déséquilibre de liaison intra-génique sur quelques centaines de paires de bases (ce chapitre)

Le déséquilibre de liaison est encore perceptible surtout pour quelques populations mais les valeurs prises par ce déséquilibre (D composantes d'Ohta) ou par les corrélations entre paires de locus (R_{ij}) sur des distances de quelques centaines de paires de bases permettent d'envisager des études d'association entre phénotype et génotype avec une grande résolution que ce soit à l'échelle des populations ou de la région. L'étude d'échantillons couvrant l'ensemble de l'aire de répartition du maïs entraîne une structuration liée à l'isolement (Vigouroux *et al.* 2003). L'étude de modèles, tels que les populations des vallées centrales de Oaxaca, permet de s'affranchir de la structuration associée à cet isolement tout en maximisant la variabilité phénotypique (Q_{st} élevé, chapitres V et VI).

Ce type de modèle est également à même de fournir des renseignements sur les relations entre la répartition de la diversité génétique observée au niveau de marqueurs neutres, celle observée au niveau des caractères phénotypique, et celle liée aux polymorphismes associés à cette variation phénotypique (relation F_{st} , Q_{st} et le F_{st} mesuré pour des polymorphismes associés au phénotype ou QTL_{st}). Il devrait ainsi être possible d'avoir un aperçu des mécanismes, à l'échelle moléculaire, ayant façonné la répartition de la variabilité phénotypique.

La structuration des populations pour des marqueurs répartis le long de *Sh2* ne permet pas de conclure de façon claire à une sélection qui porterait sur les polymorphismes caractérisés sur ce gène. Pour *Shrunken2*, nous ne savons pas si les polymorphismes que nous avons caractérisés sont associés ou non à la variation phénotypique. Pour réaliser ce genre d'étude, il aurait fallu choisir préalablement des polymorphismes, que l'on sait, associés à la variabilité phénotypique et d'autres non-associés. Des polymorphismes associés à de la variabilité

phénotypique ont été caractérisé pour *Shrunken2* (résultats non publié à ce jour). Toutefois, ils sont associés à d'importantes interactions génotype par environnement (Ed Buckler, communication personnelle). Ceci nous amène à conclure que le gène *Sh2* ne constitue pas le support le plus favorable aux travaux que nous nous proposons de mener.

Nous avons vu dans le chapitre IV que l'homogamie liée à la précocité de floraison exerce une influence importante sur un marqueur localisé dans l'intervalle entre les gènes *indeterminate1* et *dwarf8*. Or, une association a été mise en évidence entre des polymorphismes le long du gène *dwarf8* et la variation phénotypique pour la précocité (Thornsberry *et al.* 2001). De plus, le gène *Indeterminate1*, qui code pour un facteur de transcription possiblement impliqué dans la transition florale, pourrait être un QTL important du contrôle de la floraison. A l'appui de cette hypothèse, un polymorphisme d'insertion/délétion du gène *indeterminate1* a été associé à une importante différence entre les dates de floraison (Thornsberry *et al.* 2002 et communication personnelle). Néanmoins ce travail n'a pu faire l'objet d'un article car les auteurs n'ont pu discriminer la part associée au phénotype de celle liée à la structuration pour ces marqueurs au sein de l'échantillon, à l'image de ce qui avait été réalisée sur *dwarf8* (Buckler et Thornsberry 2002; Thornsberry *et al.* 2001). La présence éventuelle de ce polymorphisme sur *Indeterminate1* parmi les populations étudiées dans les vallées d'Oaxaca, où nous ne notons pas de différenciation importante pour des marqueurs neutres (chapitre IV et VI), nous permettrait de trancher cette question. Le cas échéant, nous disposerions alors d'un support adéquat pour étudier l'association entre ce polymorphisme, la précocité et l'homogamie au sein des populations de maïs blanc, et pour évaluer l'impact de la sélection exercée par l'agriculteur. De même, nous serions à même d'étudier la différenciation pour ce polymorphisme au sein des populations de maïs noir, qui montrent une importante diversité au niveau de la précocité de floraison (chapitre VI).

VIII. PERCEPTIONS ET PRATIQUES PAYSANNES

Le projet pluridisciplinaire “ Oaxaca ” conduit par le CIMMYT comprend, outre la présente thèse, des études ciblant les aspects génétiques, sociologiques et économiques relatifs à la culture du maïs dans cette région. Il est difficile de savoir avec exactitude ce que fait chaque agriculteur car les informations recueillies sont souvent partielles et limitées par les difficultés de compréhension avec l’enquêteur. Ce chapitre fera le bilan de l’apport de cette thèse vis-à-vis de la connaissance des pratiques culturelles paysannes. Nous confronterons les résultats obtenus, aux pratiques paysannes d’échanges et mélanges de semences, à la perception qu’ils ont de la sélection, et aux différentes modalités de gestion et de sélection observées aux cours des enquêtes. Nous utiliserons les résultats de notre propre enquête ainsi que ceux de nos différents collaborateurs du projet.

8.1 Flux de gènes induits par les pratiques paysannes

Au cours du chapitre IV, nous avons constaté une faible différenciation entre populations pour des marqueurs microsatellites. Ce résultat se vérifie à la fois pour des marqueurs nucléaires et pour des marqueurs cytoplasmiques, ce qui suggérerait d’important flux de semences. De quelle manière les pratiques paysannes permettent-elles d’expliquer la faible différenciation observée pour ces deux classes de marqueurs ?

En supplément des épis qu’ils sélectionnent à partir de leur récolte de l’année pour effectuer le réensemencement de l’année suivante, les agriculteurs ont recours à des sources de semences extérieures. Celles-ci peuvent être situées à l’intérieur du même village (un agriculteur d’Huitzo: “ *Compraria con los vecinos a quien he vendido* – Je l’achèterais à mes voisins à qui j’ai déjà vendu de la semence ”) ou dans des villages proches (“ *Lo volveria a comprar en Yogoloxi con familiares de mi esposo*. – Je le rachèterais à Yogoloxi à des parents de mon mari ”, nous a déclaré une agricultrice d’Amatengo). Enfin, des semences d’origines plus lointaines peuvent être achetées sur les marchés, comme en témoigne un exploitant de Santa Ana Zegache: “ *Hace 3 años conseguí semilla parte en Ocotlan y parte con mi tío. Soy de Cotla y me es difícil de conseguir maíz aquí en Sta Ana Zegache*. – Il y a trois ans je me suis procuré une partie de la semence sur le marché d’Ocotlan et une autre avec mon oncle. Je suis originaire de Cotla et il m’est difficile d’obtenir de la semence ici ”.

La mise en culture, sur une même exploitation, de populations différentes, peut résulter, à terme, dans des mélanges de semences. Ceux-ci peuvent être le fruit d'une décision de l'agriculteur (propos recueillis à Valdeflores: “ *Empeze con esta variedad en el 1998 porque hubo sequia y perdi mi semilla. De hecho ese año compre dos tipos que sembré por separado. Pero en 1999 decidí revolver los dos.* – J’ai commencé à cultiver cette variété en 1998 parce que j’avais perdu mon maïs à cause de la sécheresse. Cette année là j’ai acheté deux maïs que j’ai semé séparément. Mais en 1999 j’ai décidé de les mélanger”). Dans d'autres cas, les graines de différentes origines sont mélangées pour des raisons pratiques, comme chez ces agriculteurs de San Lorenzo partageant le même champ: “ *Agregue de la semilla del Sr A. porque sembré a media y el mediero compro semilla con el Sr A.* – J’ai mélangé ma semence à celle de Mr A. parce que je partage mon champ avec un autre agriculteur qui lui a acheté de la semence”. Enfin, il est fréquent qu'une nouvelle population de maïs soit cultivée à titre d'essai, et que les graines correspondantes soient mélangées à celles des populations "résidentes" lors de la récolte, comme l'indique un agriculteur d'Amatengo “ *No me gusto, pero ya se mezclo con el mio en la cosecha* – Il (ce maïs) ne m’a pas plu ; Mais il s’est mélangé au mien à la récolte”. De tels mélanges sont rendus possible par l'habitude qu'ont les paysans d'emmagasiner l'ensemble de leur récolte sous la forme d'un monticule unique, avant de procéder au choix des grains utilisés lors du réensemencement. Ce qui n’apparaît pas dans les enquêtes, c’est que près de la moitié des échanges de semences entre agriculteurs se fait pour évaluer une population différente de la leur (Badstue, communication personnelle).

Parmi les exploitants que nous avons interrogés, 12 (soit 20%) ont ainsi affirmé avoir réalisé des mélanges entre leurs graines avec celles obtenues chez un agriculteur du même village, et 6 (soit 10%) ont participé à des échanges de semences entre différents villages. Dans ce dernier groupe, deux des agriculteurs ont procédé à un mélange de graines, et les quatre autres ont mis en culture de nouvelles populations à partir de la semence obtenue dans d’autres villages (fondation).

Cependant, une enquête antérieure réalisée sur les 6 mêmes villages avait révélé que près de la moitié des agriculteurs avaient eu recours à des mélanges de semences (Mendoza 2000). La divergence entre ces résultats et les nôtres peut être expliquée par deux facteurs. En premier lieu, l’enquête de Mendoza a été réalisée en 1998, c'est-à-dire après que la sécheresse prolongée survenue en 1997 ait considérablement affecté la production de nombreux paysans. La conséquence prévisible d'un tel phénomène est l'emploi massif de mélanges de graines, faute de récolte suffisante pour effectuer le réensemencement. En-dehors de l'augmentation

ponctuelle et conjoncturelle de cette pratique, il est toutefois possible que les chiffres obtenus à la suite de notre enquête en sous-estiment l'importance. En effet, nous nous sommes basés sur les informations fournies par les agriculteurs eux-mêmes à l'époque où nous avons procédé à l'échantillonnage de leurs populations. Par contraste, Mendoza (2000) avait eu l'occasion de suivre l'ensemble des événements se déroulant entre la récolte et les semences suivantes.

L'ensemble des informations recueillies va dans le sens d'une pratique généralisée des échanges de semences. Selon la source de graines externe utilisée par l'agriculteur, ces échanges permettent des flux de gènes entre populations plus ou moins éloignées. Ceci explique la faible différenciation mesurée à partir des marqueurs moléculaires à hérédité nucléaire et chloroplastique (Chapitre IV).

8.2 Compartimentation des populations en fonction de la couleur du grain

Au cours du chapitre VI, nous avons constaté que les deux "compartiments" représentés par les populations de maïs à grains blancs et noirs, respectivement, sont faiblement différenciés au niveau des marqueurs moléculaires. Ce résultat se vérifie à la fois pour des marqueurs nucléaires et pour des marqueurs cytoplasmiques, ce qui semble indiquer que des flux de gènes se produisent régulièrement entre ces deux ensembles, par voie de semence et par voie de pollen, respectivement. Par ailleurs, nos résultats indiquent que la différenciation entre les populations de maïs noir est plus importante que celle observée entre les populations de maïs blanc. De quelle manière les pratiques paysannes permettent-elles de justifier cette différence entre maïs blanc et noir, tout en autorisant les flux de gènes entre eux ?

La gestion de la culture des populations de maïs noir se distingue de celle des maïs blancs sur plusieurs points (Smale *et al.* 1999). En premier lieu, la surface consacrée aux maïs colorés est généralement plus réduite que celle occupée par le maïs blanc: en moyenne 0.17 ha contre 2.5, respectivement, soit un rapport de 1 pour 15, chez les paysans cultivant les deux types de maïs. D'autre part, les maïs de couleur font l'objet de moins de mélanges de graines de la part des agriculteurs. La conjonction de la petite taille des parcelles dédiées et de la moindre importance accordée à ces types de maïs, et enfin, de la difficulté à se procurer les semences correspondantes (excepté dans le village de Santa Ana Zegache), résulte dans une fréquence élevée des processus d'extinction de ces populations, par rapport à ce qui est trouvé chez le maïs blanc. De nombreux événements de fondation ou de recolonisation sont en revanche

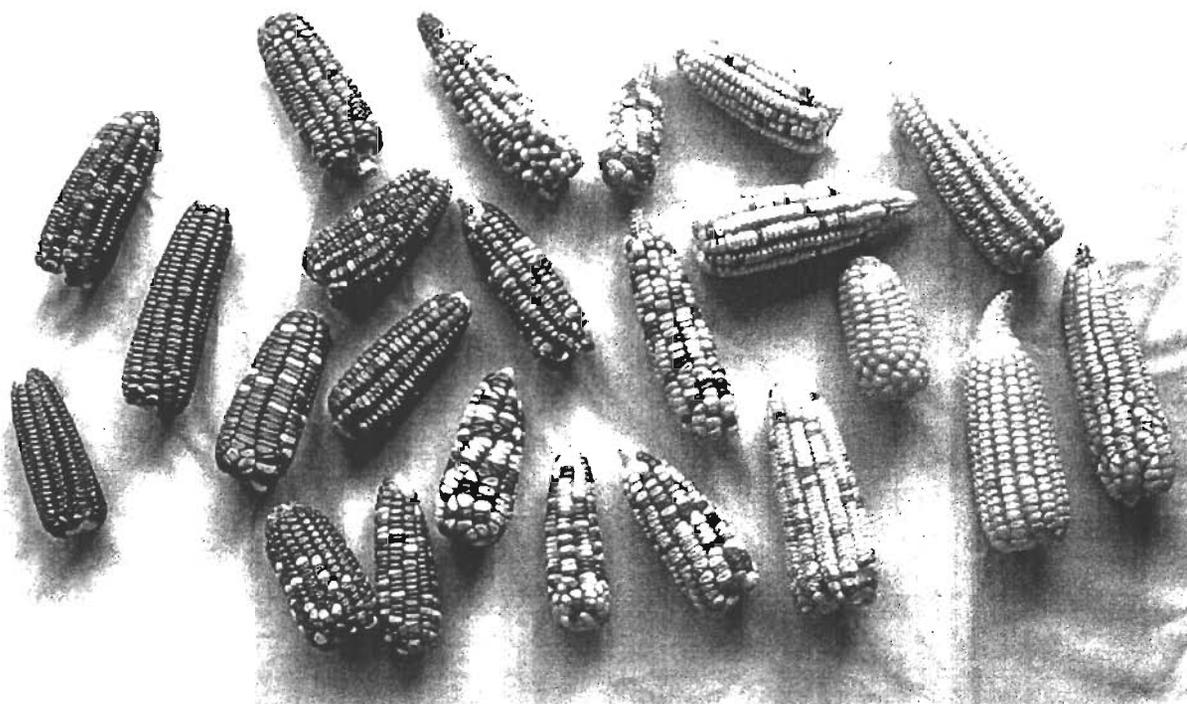
rendus possibles par le biais d'échanges de graines ou par l'ensemencement, à titre d'essai, d'une petite fraction d'une parcelle à partir de quelques épis de couleur. Ce dernier cas de figure, lorsqu'il débouche sur l'adoption de la population de maïs coloré, est susceptible de provoquer une importante dérive, notamment au niveau des marqueurs à hérédité maternelle (chapitre VI).

Des différences de comportement ont été constatées entre les agriculteurs confrontés à la présence de grains de couleur sur des épis de maïs blanc. Alors que certains d'entre eux éliminent de tels épis, d'autres nous ont déclaré les utiliser, grains noirs compris, pour le réensemencement (Figure 8.1). Il en résulte qu'un des individus issu des grains noirs est susceptible, à la récolte suivante, d'être sélectionné comme faisant partie de la population de maïs noir. Par suite, les monticules constitués par les agriculteurs regroupent non seulement un mélange des deux types d'épis, mais également la gamme des phénotypes intermédiaires de grain.

Figure 8.1 – L'agriculteur n'élimine pas systématiquement les épis comportant des grains d'autres couleurs. Les épis provenant de différentes populations cultivées dans la même exploitation sont fréquemment conservés sous la forme d'un unique monticule.



Le cas du maïs pinto est particulièrement intéressant à cet égard. Ces populations se caractérisent en effet par des épis présentant simultanément plusieurs couleurs de grain (blancs et/ou jaunes, noirs et pourpres). La coexistence de plusieurs couleurs de grain au sein d'une population est permise par le déterminisme complexe de la couleur de la couche à aleurone. Par la pratique des mélanges de graines entre maïs de couleur et leur culture au voisinage d'autres types de maïs, certains agriculteurs ont ainsi assisté à l'émergence d'un maïs pinto à partir de populations blanches ou jaunes. Un agriculteur de Santa Ana Zegache nous dit “ *Mi esposa sembró unos surcos de maiz negro pero la regañe porque mancha al otro maiz.* – Ma femme a semé quelques lignes de maïs noirs. Je l'ai grondé parce que cela tachait l'autre ”. Un autre, à Santa Ana Zegache, déclare “ *No me gustan los de color porque se*



Noir

Pinto

Blanc

revuelven mucho y hay que ponerlos aparte – Je n’aime pas les maïs de couleur parce qu’ils se mélangent beaucoup (avec les autres) et il faut les isoler”.

Le maïs pinto est cultivé en proportions variables dans tous les villages (Smale *et al.* 1999) ; à San Lorenzo, la surface qu’il occupe (environ 30%) est plus importantes que celle observée dans d’autres villages. Les témoignages obtenus auprès des exploitants de ce village nous ont permis de confirmer que le maïs pinto constitue un support aux flux de semences se déroulant entre les maïs de couleurs différentes. Ils rendent compte des nombreuses possibilités d’extraction par sélection réalisables à partir du maïs pinto: “*El maiz negrito viene de la semilla pinta que compre y como me gusto el color, trato de conservarlo, escogiendo mazorcas de color negro cuando saco semilla* – Le maïs noir vient de maïs pinto que j’ai acheté. Comme la couleur m’a plu, je l’ai conservé en choisissant les maïs noirs lors de la sélection de semence”. “*Lo seleccione del maiz pinto que tenia antes, porque no se vendia el pinto* – Je l’ai sélectionné [mon maïs blanc] d’un maïs pinto que j’avais avant parce que le pinto ne se vend plus”. “*Mi papa sembraba del pinto pero despues fui seleccionando el puro blanco* – Mon père semait du pinto mais j’ai après sélectionné le blanc pur”.

Les populations de maïs à grains noirs sont gérés différemment des maïs blancs notamment vis-à-vis des tailles de populations et des événements d’extinction ; cela, explique la réduction de la diversité observée au sein des populations de maïs de couleur (H_e des populations de maïs noir inférieures à celles observées chez les blancs, chapitre VI). Néanmoins les populations de maïs blanc, pinto et noir participent d’un continuum autorisant les flux de gènes et notamment de graines au sein des populations considérées (faible différenciation entre blancs et noirs, chapitre VI).

8.3 Perceptions paysannes de la variabilité phénotypique

Variabilité intra-population

Nous traiterons ici de l'exemple de la variabilité phénotypique pour un caractère dont nous avons vu qu’il fait l’objet d’une sélection de la part des agriculteurs: la précocité de floraison (Chapitres IV et VI). Comme nous l’avons vu précédemment, chaque population présente une palette différente, et plus ou moins diversifiée, de la variabilité existant pour un caractère donné. En ce qui concerne la précocité de floraison, cette hétérogénéité est extrêmement variable d’une population à l’autre. Néanmoins, il est rarement question de cette variabilité

intra-population dans les propos des agriculteurs, à l'exception notable d'un paysan de San Lorenzo qui nous a déclaré être en mesure de distinguer plusieurs sous-populations, au sein de la variété qu'il cultive, en fonction des décalages qu'il observe dans leurs dates de floraison respectives.

Les agriculteurs n'ont pas forcément conscience de la variabilité de leur population pour un caractère donné, et ce, même pour un caractère montrant une structuration et hétérogénéité intra-population aussi marqué que la précocité de floraison (Chapitres IV et VI).

Variabilité entre populations d'un même village

Dans la région de Oaxaca, nous avons observé que la variabilité phénotypique inter-populations est plus importante entre des villages différents qu'à l'intérieur d'un même village. Toutefois, le chapitre V nous a donné l'occasion de démontrer qu'à cette dernière échelle, une différenciation phénotypique est bien détectable. Tous les agriculteurs d'un village n'ont cependant pas conscience de ces différences: “ *Todos tenemos lo mismo* – on a tous la même chose ” disent certains, “ *Compraria con cualquiera porque todos siembran de esta semilla* – J'achèterais du maïs à n'importe qui, vu que tout le monde sème le même ”, nous affirme un agriculteur d'Amatengo, alors qu'un de ses voisins déclare: “ *Cada quien tiene el suyo. No se si consigua el mismo* – chacun a le sien, je ne sais pas si je retrouverais le même ”.

De même que pour la variabilité intra-population vue précédemment, les différences entre populations d'un même village ne sont pas perçues par tous les agriculteurs.

Flux de semences à longue distance et variabilité phénotypique inter-villages

Les résultats exposés au cours du chapitre V mettent en évidence une importante différenciation phénotypique entre des populations de maïs cultivées dans des villages différents, perceptible au niveau de caractères quantitatifs. Parallèlement, la différenciation mesurée au niveau des marqueurs moléculaires demeure faible, plus particulièrement au niveau des marqueurs à hérédité nucléaire. Ce dernier résultat suggère la pratique d'échanges de graines entre villages, qui représentent ainsi un flux de gènes prépondérant par rapport au pollen pour ces longues distances.

Les agriculteurs ont conscience des différences phénotypique et adaptatives existant entre leur population et celles cultivées dans d'autres villages, et certains choisissent de privilégier les échanges de semences avec les villages cultivant un morphotype de maïs ressemblant à celui

qu'ils cultivent déjà, ainsi ce paysan de Santa Ana Zegache: “ *Para conseguir semilla de este tipo buscaria en Zegache, Santiago o Sn Martin* – pour obtenir ce type de maïs il faudrait qu'il cherche à Santa Ana Zegache, Santiago ou San Martin”. D'autres saisissent au contraire l'opportunité de ces échanges à longue distance pour expérimenter des variétés très différentes de la leur, qui viendront s'y mélanger, voire la remplacer: “ *Hace 15 años, ella queria de otro maiz, por eso conseguí un criollo bolita, el cual, aunque la mazorca este mas chiquita si rinde. En el mercado de Ocotlan vio el maiz con granos grandes y boludos y compro un almud, porque ademas le dijeron que aguanta la sequedad. En un principio lo sembraba aparte, pero despues en la cosecha se le fue revolviendo, y tiene aproximadamente 10 años que siembro un revuelto o una mezcla.* – Il y a 15 ans elle [sa femme] a voulu un autre type de maïs. C'est pour cette raison que je me suis procuré un Bolita. Bien que l'épis soit plus petit il possède un bon rendement. Sur le marché de Ocotlan j'ai vu le maïs aux gros grains ronds, j'en ai acheté quatre kilos. En plus il m'ont dit qu'il résistait à la sécheresse. Au début je l'ai semé séparément. Mais petit à petit ils se sont mélangés à la récolte et depuis dix ans je sème ce mélange ”, nous a raconté un agriculteur d'Amatengo.

Bien qu'ils aient conscience des différences morphologiques entre populations de villages différents, certains agriculteurs réalisent tout de même des introductions et mélanges qu'ils soient pratiqués volontairement ou non. Ceci permet donc d'expliquer la maintenance de flux de semences à longue distance malgré l'existence d'une importante différenciation phénotypique (Chapitres IV et V).

8.4 Conclusions

Il est difficile d'obtenir des données quantitatives sur les échanges de semences et sur la sélection paysanne à partir des enquêtes. En effet il n'y a pas forcément adéquation entre les actions des agriculteurs et la perception qu'ils en ont. Toutefois, ces enquêtes nous ont permis de conforter nos résultats.

IX DISCUSSION ET CONCLUSIONS

Nous allons dans un premier temps résumer nos résultats, et rappeler comment ils nous renseignent sur les conséquences de la gestion paysanne sur la diversité génétique et phénotypique et sa répartition. Ensuite, nous discuterons ces données à la lumière des études socio-économiques menées sur les pratiques culturelles des exploitants, afin d'en saisir les implications quant à la conservation et à la mise en valeur des ressources génétiques.

La stratégie multi-approches que nous nous sommes proposés de suivre dans le cadre de cette thèse nous a permis d'appréhender les différents aspects revêtus par la gestion paysanne des populations de maïs dans les Vallées Centrales de Oaxaca au Mexique. L'utilisation de marqueurs neutres nous a donné un aperçu des événements de migration et dérive entrant dans l'historique de cette culture. Par ailleurs, les conséquences de la sélection en milieu paysan ont été rendues perceptibles par l'analyse de la diversité au niveau de caractères quantitatifs. En prise directe avec les informations collectées sur cet agro-écosystème, les résultats que nous avons exposés sont à même de nous faire comprendre par quels mécanismes l'environnement et le mode de gestion entourant la culture de ces populations traditionnelles de maïs façonnent la diversité de celles-ci.

9.1. Répartition et gestion de la variabilité génétique des populations traditionnelles de maïs dans les Vallées Centrales d'Oaxaca

Un système non cloisonné en terme de flux de gènes

Nous avons montré précédemment que les populations que nous avons étudiées possèdent une diversité faiblement structurée au niveau des marqueurs “ neutres ”, ne donnant pas prise à la sélection. Ce résultat a été vérifié aussi bien pour des marqueurs à hérédité biparentale que pour des marqueurs à hérédité maternelle. L'importance des échanges de semences entre agriculteurs, déjà soulignée à la suite des enquêtes effectuées auprès de ceux-ci, trouve donc sa confirmation dans nos travaux (Chapitres IV et VI). L'existence d'échanges de graines entre agriculteurs avait déjà été mise en évidence lors de travaux antérieurs ciblant le village de Cuzalapa dans l'état de Jalisco au Mexique (Louette 2000; Louette *et al.* 1997; Louette et Smale 2000). Les conclusions de ces études, partagées par celles que nous exposons ici, contribuent à faire apparaître les flux de gènes, qu'ils soient dépendants (graines) ou non

(pollen) de l'action de l'agriculteur, comme une part intégrante du système d'agriculture traditionnelle au Mexique.

Notre étude est, à notre connaissance, la première qui se soit attachée à détailler la structuration, entre agriculteurs, des populations de maïs mexicaines (ou lots de semences) au moyen de marqueurs moléculaires. En l'absence de précédent, nous sommes contraints, pour tenter d'évaluer la portée de nos résultats, de procéder par analogie avec des travaux portant sur des systèmes, de culture du maïs, très différents du nôtre. Ainsi, Gauthier *et al.* (2002) montrent une différenciation plus élevée entre les populations traditionnelles de maïs en Europe ($G_{st}=0.31$), par rapport à ce que nous avons mesuré sur nos populations ($F_{st}=0.011$). De même, les populations traditionnelles de maïs du Burkina-Faso montrent une structuration plus importante ($F_{st}=0.22$; Sanou *et al.* 1996) que celle que nous avons observée. Il faut toutefois remarquer que ces études concernent des aires géographiques à la fois plus étendues et plus fragmentées par rapport aux Vallées Centrales d'Oaxaca, qui sont traversées par de nombreuses routes d'échange utilisées depuis des temps reculés. Nos données sont davantage comparables aux valeurs trouvées au niveau d'une vallée des Alpes italiennes ($F_{st}=0.066$; Barcaccia *et al.* 2003). Il a été démontré que, dans l'ensemble du continent américain, la différenciation des populations de maïs au niveau des marqueurs neutres est principalement due à l'isolement par la distance, en-dehors de quelques événements récents de migration (Vigouroux *et al.* 2003). Pour cette raison, la répartition de la variabilité génétique du maïs semble être continue dans les Amériques, cette hypothèse étant étayée par l'absence apparente de cas de " goulot d'étranglement " (Matsuoka *et al.* 2002). Par contraste, un effet fondateur fort exercé par des introductions multiples a favorisé la différenciation entre les populations de maïs européennes (Dubreuil *et al.* 2003 et communication personnelle), avec pour conséquence une répartition discrète de la variabilité génétique. La résultante de ces diverses observations est que, sur le continent américain, seule la prise en compte d'aires géographiques considérables, dépassant de loin celles de la présente étude, permettrait l'observation de phénomènes de différenciation aussi prononcés que ceux constatés en Europe ou en Afrique. Par contre, au sein de systèmes de petite taille, comme dans le cas des vallées d'Oaxaca ou de la vallée alpine étudiée par Barcaccia *et al.* (2003), des F_{st} faibles et une absence d'isolement par la distance sont observés.

Distribution de la diversité au niveau des caractères quantitatifs

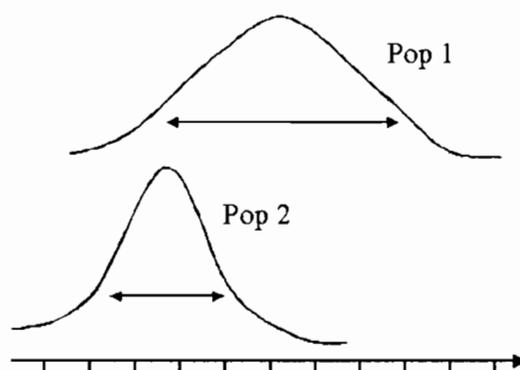
Nos travaux ont permis de montrer que, pour la plupart des caractères dont nous avons mesuré la diversité, on obtient des valeurs telles que $Q_{st} \gg F_{st}$. Cette inégalité a été surtout trouvée pour les caractéristiques morphologiques du grain et de l'épi (chapitre V et VI) ainsi que pour les caractères liés à la floraison (chapitre VI). Ce résultat est révélateur des influences multiples exercées par les pratiques culturales paysannes sur la diversité phénotypique au sein des vallées centrales d'Oaxaca. Les effets dus à l'action de l'agriculteur sont ainsi perceptibles à la suite de processus de sélection, que celle-ci soit directe (chapitre V et VI) ou indirecte par le biais d'un mode de gestion spécifique au type de maïs cultivé (en jouant sur les dates de semis à l'exemple des maïs à grains noirs, chapitre VI).

De plus, les différences observées entre les Q_{st} correspondant à des caractères distincts illustrent le fait que chacun des caractères quantitatifs étudiés possède son propre profil de distribution de la diversité phénotypique. Nous avons également montré au cours de notre travail que la variabilité observée pour un caractère quantitatif donné au sein d'une population particulière (son hétérogénéité) est propre à cette dernière (hétérogénéité liée à la floraison, chapitres IV et VI).

Figure 9.1 – Répartition de la variabilités intra- et inter-populations pour un caractère quantitatif.

La variabilité intra-population (ou hétérogénéité) pour un caractère donné n'est pas également répartie d'une population à l'autre.

La part de variabilité inter-population (ou Q_{st}) est fonction du caractère considéré.



Enfin, la matrice de variances covariances (matrice **G**) et la matrice de corrélations génétiques (matrice **G_c**) subissent des changements non-proportionnels, qui sont significatifs non seulement entre villages différents mais également à l'intérieur d'un village donné (chapitre V). Or, les changements attendus à la suite d'un phénomène de dérive seraient proportionnels (égalité des matrices ou alors égalité des composantes principales), la sélection entraînant par contraste des changements non-proportionnels (certaines des composantes principales sont différentes d'une matrice à l'autre, ou la structure des matrices est sans relation aucune; Roff 2000). Ces conclusions impliqueraient donc que, à l'échelle des villages étudiés de Oaxaca, la

différenciation phénotypique entre les populations de maïs résulterait de la sélection en dépit de flux de gènes importants (chapitres V et VI).

9.2 Gestion paysannes de la diversité, populations traditionnelles et variétés de maïs

Nous avons montré que la sélection exercée par les paysans des vallées centrales d'Oaxaca sur leurs populations traditionnelles de maïs parvient à induire une différenciation phénotypique hautement significative. Ce phénomène est d'autant plus remarquable qu'il se manifeste en concurrence avec d'importants flux de gènes entre les populations cultivées. Nous allons détailler ici quelles sont les décisions prises par l'agriculteur qui entraînent ce processus de différenciation phénotypique. Nous, nous attacherons également à établir quelles sont les parts respectives de la sélection conservatrice (l'agriculteur maintient un type déterminé) et de la sélection divergente (l'agriculteur modifie les caractéristiques de sa population) dans l'émergence de cette différenciation.

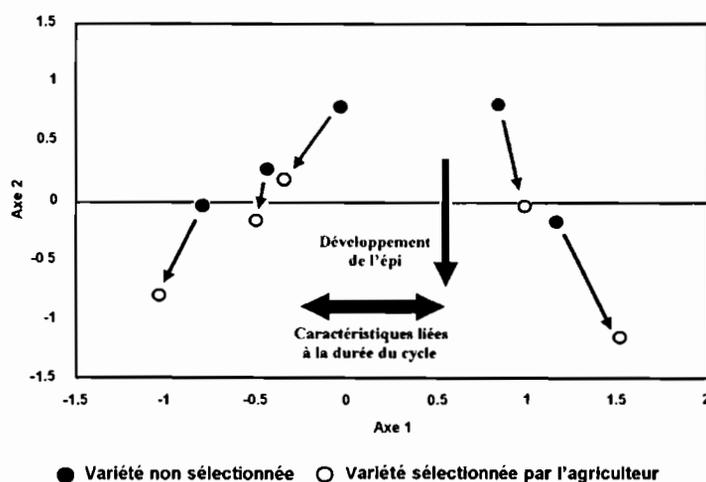
Perceptions, critères, et modalités de la sélection paysanne

Ainsi que nous l'avons déjà évoqué, l'action de l'agriculteur s'effectue d'une part directement, au travers du choix des épis et des grains utilisés pour le réensemencement de sa parcelle, et d'autre part indirectement, en modulant la date du semis en fonction de la précocité de floraison de la population.

Les caractères de l'épi et du grain recherchés par l'agriculteur, et qui lui tiennent lieu de critères de sélection, sont extrêmement relatifs et qualitatifs. A titre d'exemple, il n'existe aucun calibrage des tailles de grains existantes, aussi la définition d'une population à "gros" grains est-elle purement subjective et soumise à des variations d'un exploitant à l'autre. Ceci se reflète dans les entretiens réalisés auprès de ces derniers, certains affirmant posséder et maintenir le même matériel que les voisins, d'autres, au contraire, ayant conscience des différences. Une différence de perception entraîne ainsi une différence dans la pression sélective appliquée à chaque cycle.

Toutefois, nos résultats vont dans le sens d'un plus grande ressemblance entre les populations cultivées au sein du même village, où ces critères de sélection déclarés sont les mêmes, qu'entre villages différents. Cette tendance est particulièrement vérifiée dans le cas des caractères de l'épi (chapitre V).

Figure 9.2 – Figure tirée de Louette et Smale (Louette and Smale 2000). ACP sur des caractères de l'épis et des caractères végétatif liés à la précocité indiquant l'influence de la sélection effectuée selon l'habitude de l'agriculteur, *versus* le comportement de la même population en l'absence de sélection (tirage aléatoire des épis pour le réensemencement).



Le pouvoir de décision de l'agriculteur se traduit également par le choix des populations qu'il adopte ou rejette, ce qui équivaut à des événements de colonisation et migration ou extinction, intervenant dans la gestion de sa population. Le poids exercé par de telles décisions est particulièrement sensible lors de la sélection pour la précocité de floraison, où l'agriculteur n'hésite pas à changer de population dès que celle qu'il a adoptée ne lui donne plus satisfaction pour ce caractère. Ce principe de gestion pragmatique de sa culture par l'agriculteur peut être étendu à l'ensemble des caractères sur lesquels il applique sa sélection: une population est abandonnée dès lors qu'elle ne remplit plus les critères (de taille de l'épi ou du grain, de précocité, de rendement, de qualité de la tortilla) qui l'ont faite retenir initialement.

Lors de travaux antérieurs conduits sur un système de culture comparable de maïs (Louette 2000; Louette *et al.* 1997; Louette et Smale 2000), il a été observé que les caractéristiques morphologiques propres à chaque population sont globalement maintenues au cours du temps, alors même que s'exercent des flux de gènes entre populations qui s'y opposent. Les auteurs ont par suite conclu à l'application d'une sélection conservatrice par les agriculteurs.

Or, comme nous l'avons montré plus haut, si nos propres résultats vont bien dans le sens d'un maintien globale d'un morphotype dans un village donné, nous montrons également que les différences de perception de la part des agriculteurs produisent une part variable de divergence des critères de sélection entre eux. Si nos travaux rendent effectivement compte d'une structuration par village et de l'existence d'un morphotype caractéristique de chaque village, nous mettons également en évidence une différenciation phénotypique résultant de la sélection paysanne à l'intérieur même du village (différenciation phénotypique intra-village et différences significatives entre les matrices **G** au sein d'un même village).

Davantage qu'une contradiction des conclusions de Louette *et al.*, notre étude apporte une nuance: fondamentalement il y a bien une conservation au niveau des caractères recherchés par l'exploitant pour une population donnée, mais cette conservation prend une forme différente (elle est plus ou moins fidèle, ou au contraire biaisée) en fonction de l'agriculteur. C'est cette part de subjectivité, laissée libre par l'absence de standardisation du système de sélection, qui autorise l'émergence d'une différenciation phénotypique.

Nous arrivons donc à ce qui pourrait apparaître comme une contradiction ; les agriculteurs différencient leurs populations tout en les conservant !

Différenciation phénotypique et caractérisation du morphotype village

Comme nous l'avons vu au cours de cette étude, la différenciation phénotypique est perceptible au sein de la population cultivée par un agriculteur, lequel conserve imparfaitement le morphotype de maïs qui a cours dans son village. Cependant, nos résultats ne présument en rien de la manière dont la variabilité phénotypique est capable d'évoluer sur le long terme à l'échelle du village.

Certains indices nous offrent néanmoins un aperçu sur le déroulement probable de cette évolution. De nombreux agriculteurs interrogés dans le village d'Amatengo ont été incités à changer de population cultivée (extinction suivie de colonisation) à la suite de l'adoption d'une nouvelle mesure comme base de prix de vente du maïs. Alors que l'ancienne référence était fondée sur le volume de grain, conditionnant une sélection pour ce caractère, le nouveau standard a été établi en fonction du poids. En conséquence de cette pression économique, les exploitants se sont tournés vers des morphotypes possédant un grain plus dense. Cet exemple illustre un cas de figure où le changement du morphotype privilégié au sein d'un village peut se produire très rapidement, grâce à l'adoption d'un nouveau morphotype au détriment de l'ancien et par le biais de la sélection et d'échanges de semences à l'intérieur du village ou entre villages.

Délimitation des lots de semence, variétés et cultivars en système paysan

En ce qui concerne la différenciation phénotypique résultant de la sélection paysanne, on peut se demander à partir de quel stade on peut considérer qu'une population a divergé de la variété initiale au point de représenter une nouvelle variété. Les données que nous avons rassemblées montrent en effet une très forte structuration de la variabilité phénotypique, détectable à de multiples niveaux (agriculteur, village, couleur du grain, précocité, région). Dans le but d'établir une unité de gestion des ressources génétique *in situ*, on peut donc être tenté de regrouper des populations en variétés définies sur la base de caractères phénotypiques.

S'il est relativement aisé de définir un lot de semence comme étant l'ensemble des épis qui serviront pour le semis du cycle suivant (Louette 2000; Louette *et al.* 1997; Louette et Smale 2000), il est beaucoup plus difficile de définir cultivars et variétés dans le cadre du système paysan sur lequel porte notre étude.

Louette (2000) définit la variété ou le cultivar comme un groupe de lots de semences portant le même nom et considérés comme formant un groupe homogène selon des critères morphologiques. On voit bien comment cette définition est susceptible de laisser libre cours à la perception qu'ont les agriculteurs de leur population.

Selon le point de vue de l'obteneur ou du sélectionneur, la "variété végétale" s'entend comme un ensemble de végétaux globalement similaires de par leurs caractères et qui peuvent être reproduits ou multipliés sans modification significative à l'issue de chaque cycle. Cette définition correspond donc à une population de plantes dont les caractéristiques essentielles sont homogènes entre les individus et stables au cours du temps (convention UPOV, 1991).

Or, par opposition avec les définitions précédentes, dans le système d'agriculture traditionnelle que nous avons étudié, nous avons affaire à un groupe de populations hétérogènes, chacune pouvant être amenée à évoluer dans le temps en fonction de la perception qu'ont les agriculteurs de ses caractéristiques. L'apport permanent représenté par des flux de gènes d'origines diverses (pollen, échanges de graines) et le continuum de la variation phénotypique du maïs dans l'ensemble de la région rendent difficile la délimitation de variétés (groupe homogène de populations). Les groupes de populations locales (tels qu'identifiés par les agriculteurs) constituent un système non cloisonné en terme de flux de gènes, assimilable, de par sa structure, à une métapopulation. En cela, nos conclusions rejoignent celles de Louette (Louette 2000; Louette *et al.* 1997; Louette et Smale 2000).

Dans les vallées centrales d'Oaxaca, la "variété traditionnelle" de maïs n'est pas une entité immuable dans le temps, dont la variabilité génétique et phénotypique serait clairement délimitée. Il s'agit plutôt d'un ensemble de populations que leurs caractéristiques phénotypiques, telles qu'elles sont perçues par les agriculteurs, rattachent au même "morphotype".

La diversité phénotypique chez le maïs et le concept de race

Le concept de race chez le maïs a été initialement introduit par Anderson et Cutler (1942), puis développé par Wellhausen *et al* (1952). Par analogie avec ce qui est observé lors de croisements entre races “pures”, ces derniers auteurs proposent en effet que les différents morphotypes existants chez le maïs proviennent des intercroisements entre des races de maïs primitives ou plus anciennes. Dans cette hypothèse, une race “nouvelle” issue d'un tel croisement possèdera par conséquent des caractéristiques morphologiques intermédiaires par rapport à celles des deux races parentales. Réciproquement, l'autofécondation d'individus appartenant à cette race intermédiaire régénère les morphotypes parentaux chez une partie de la descendance. L'occurrence de ces deux types d'évènements a été jusqu'ici considérée comme la preuve de la validité de l'hypothèse de départ. La proposition de Wellhausen *et al.* (1952), appuyée par Anderson (1953) est donc que la différenciation phénotypique du maïs reflète, *via* les hybridations et les migrations, l'historique des flux de gènes entre les différentes races. Par conséquent, toute possibilité de convergence évolutive ou de différenciation phénotypique en dépit de flux de gènes importants sont exclus par ce modèle. Si la répartition de la diversité phénotypique ne reflète que les hybridations et les migrations comme postulé par Wellhausen *et al.* (1952), alors on s'attend à avoir une bonne corrélation entre la répartition de la diversité phénotypique et la répartition de la diversité génétique mesurée à l'aide de marqueurs neutres. Or, à l'échelle des villages étudiés de Oaxaca, la différenciation phénotypique entre les populations de maïs résulterait de la sélection en dépit de flux de gènes importants. Nous nous situons par conséquent en opposition au modèle de Wellhausen *et al.* (1952) décrit plus haut.

Le maïs des vallées centrales de Oaxaca appartient à la race Bolita, qui est considérée comme récente selon le schéma d'évolution découlant de la proposition de Wellhausen *et al.* (1952). Cette race possède en effet des caractéristiques morphologiques qui semblent dériver de celles de races apparemment plus anciennes, et son importante diversité phénotypique est cohérente selon Wellhausen avec une émergence récente. Cette hypothèse est infirmée par des travaux récents. Les analyses entreprises par Matsuoka (2002) au moyen de marqueurs moléculaires placent en effet cette race à la base de l'arbre phylogénique du maïs. De fait, la diversité que nous avons observée à l'intérieur de la race Bolita permet de penser que par sélection divergente on pourrait extraire des morphotypes s'apparentant à des races différentes. Le modèle de l'évolution du maïs avancé par Wellhausen *et al.* regarderait donc à tort ces morphotypes comme étant d'origine postérieure à Bolita, car il semble tenir compte des seuls processus démographiques (dérive et migrations) au détriment des effets de la sélection.

9.4 Evaluation, conservation et valorisation des ressources génétiques

A. Organisation spatiale de la diversité neutre, de la variabilité des caractères quantitatifs sélectionnés et de la variation allélique des gènes responsable de cette variation phénotypique

Relation entre la diversité génétique neutre et la diversité phénotypique

Différents systèmes chez le maïs (lignées, populations) ont été étudiés afin de mettre en évidence une éventuelle relation entre des polymorphismes détectés au niveau de marqueurs (enzymatiques ou moléculaires) neutres, et les performances agronomiques. En d'autres termes, l'enjeu consiste à déterminer la relation entre le F_{st} , la différenciation ou une distance génétique mesurée à l'aide de marqueurs neutres, et le Q_{st} , ou la différenciation phénotypique. Les différents travaux engagés, chez le maïs, ont généré des conclusions très contrastées: alors que Smith *et al.* (1990) concluent à l'absence de corrélation significative, Dillmann *et al.* (1997) aboutissent à un résultat opposé.

La question revient à se demander si la différenciation phénotypique ne peut avoir lieu que dans une situation où les flux de gènes sont limités, c'est à dire si les populations sont suffisamment isolées, ou si l'hypothèse de Wellhausen *et al.* (1952, voir encadré) est vérifiée. Réciproquement, la présence de flux de gènes importants est censée interdire toute différenciation phénotypique ; Lucchin *et al.* (2003) justifient ainsi l'absence de divergence phénotypique entre les populations traditionnelles de maïs des Alpes italiennes.

Vérifier l'existence d'une telle association est d'une importance cruciale dans le contexte de notre étude s'intéressant à l'impact de la sélection paysanne sur la différenciation phénotypique du maïs. A savoir, d'importantes pressions de sélections autorisent-elles la différenciation phénotypique en dépit de flux de gènes important ?

Peu d'études ont été réalisées sur la confrontation entre la structuration de la diversité évaluée à partir des marqueurs moléculaires et celle obtenue à partir des caractères quantitatifs. Ces travaux ont exclusivement traité du cas d'espèces sauvages (McKay et Latta 2002 pour une revue), et jusqu'à présent aucune espèce cultivée n'avait fait l'objet de telles comparaisons.

Dans le cadre de ces expérimentations, on a observé que, chez la plupart des espèces allogames, des phénomènes de différenciation peuvent se produire malgré la présence de flux

de gènes considérables entre les populations (Kremer *et al.* 1997; Morgan *et al.* 2001; Podolsky et Holtsford 1995; Prout et Barker 1993; Yang *et al.* 1996). Des mécanismes de sélection divergente peuvent alors être mis en évidence. C'est le cas de notre modèle d'étude où se déroulent de tels flux de gènes et où une intensité de sélection importante est appliquée par les agriculteurs, autorisant ainsi la différenciation phénotypique ($Nm > 1$ et $m < s$, Chapitre V et VI).

La situation est très différente pour les espèces pratiquant des formes d'inbreeding, qu'il s'agisse d'autogamie stricte ou de croisements préférentiels entre individus apparentés, ainsi que pour celles qui sont soumises à des forces d'isolement et de dérive. Dans de tels cas de figure, les effets d'une sélection divergente ne pourront pas être distingués de ceux provoqués par la mise en jeu conjointe de faibles flux de gènes et d'un phénomène de dérive, aboutissant à une différenciation des populations (Bonnin *et al.* 1996; Kuittinen *et al.* 1997; Long et Singh 1995; Lynch *et al.* 1999; Spitze 1993). En revanche, la mise en évidence d'un processus de sélection convergente est alors possible (Petit *et al.* 2001)

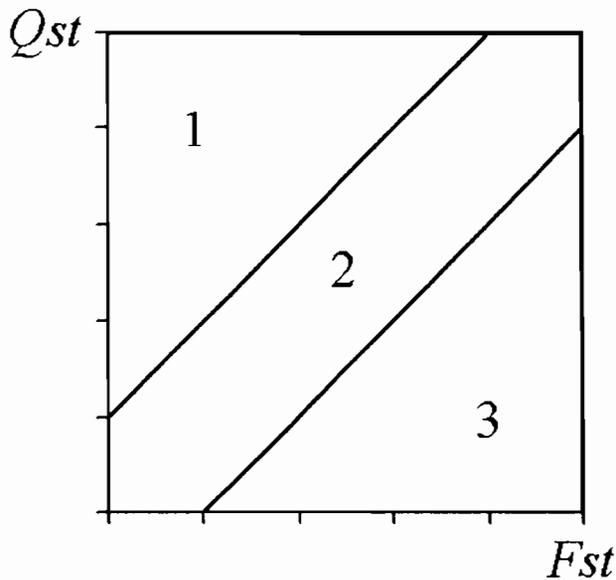


Figure 9.3 – Comparaisons Q_{st} et F_{st} . Trois situations peuvent être observées. Les situations de sélection divergente ou convergente peuvent être mises en évidence (situations 1 et 3)

1. $Q_{st} > F_{st}$,
2. $Q_{st} \approx F_{st}$
3. $Q_{st} < F_{st}$

Merilä et Crnokrak (2001) ont émis l'hypothèse que le Q_{st} est corrélé de manière positive au F_{st} et que l'isolement des populations est indispensable à la différenciation phénotypique. Latta et Mc Kay (2002) soulignent néanmoins la possibilité que des processus de sélection débouchent sur une différenciation phénotypique, et ce, malgré des flux de gènes importants entre populations. Dans un tel cas de figure, représenté par nos propres travaux au cours desquels nous avons établi que $Q_{st} \gg F_{st}$, il n'est pas possible d'établir un lien systématique

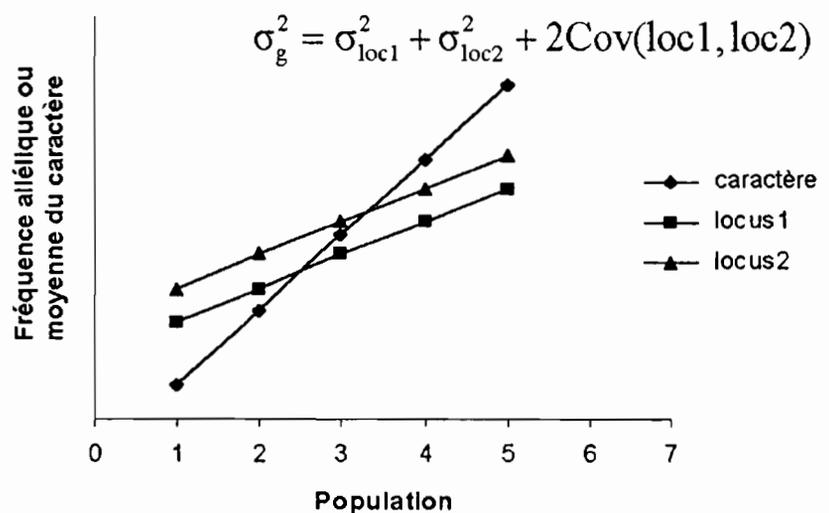
QTL (ou **QTLs**) : Quantitative Trait Loci or Polymorphism. Nous l'utiliserons ici pour un polymorphisme ou gène associé à la variation phénotypique

QTL_{st} : Par analogie, F_{st} calculé pour un polymorphisme associé à la variation phénotypique

SNP : Single Nucleotide Polymorphism (Polymorphisme de substitution nucléotidique)

IDP : Insertion-Deletion Polymorphism (Polymorphisme d'insertion-délétion)

Figure 9.4 – La variabilité phénotypique ne reflète pas nécessairement la variation observée aux QTLs régissant ces caractères. Figure tirée de McKay et Latta (2002)



entre le Q_{st} et le F_{st} . Des conclusions similaires sont supportée par l'étude de Petit *et al.* (2001), menée sur deux espèces ou groupe de populations pour lesquels ces auteurs obtiennent au contraire $Q_{st} \ll F_{st}$. Ces résultats corroborent l'hypothèse émise par McKay et Latta (2002), selon laquelle F_{st} et Q_{st} ne sont pas corrélés chez l'ensemble des espèces.

En résumé, le F_{st} ou la différenciation mesurée à l'aide de marqueurs neutres n'est pas un estimateur du Q_{st} ou de la différenciation phénotypique. La différenciation phénotypique peut donc intervenir en dépit de flux de gènes important (sélection divergente avec $Nm > 1$ et $s > m$), de même que des convergences adaptatives peuvent être observées, indépendamment de l'importance des flux de gènes.

De façon plus particulière, pour le maïs, Vigouroux *et al.* (2003) concluent que l'isolement par la distance est le principal mécanisme de différenciation des populations de maïs dans les Amériques au niveau des marqueurs neutres. On peut dire, en accord avec ces travaux, que la ressemblance mesurée à l'aide de marqueurs moléculaires reflète ainsi la proximité géographique et non la ressemblance phénotypique.

Estimer la répartition de la variabilité génétique aux gènes responsables de la variation phénotypique

La relation entre la divergence estimée à l'aide de marqueurs neutres et celles estimées à partir de caractères quantitatifs, étant établie, nous nous interrogeons sur la relation entre la diversité allélique des gènes responsables de la variation phénotypique (ou QTL) et la variabilité phénotypique. La relation directe entre les deux est loin d'être évidente. La structuration des populations étudiées au niveau de ces QTLs dépend du profil de distribution de la variabilité génétique pour ces caractères, du nombre de gènes et de leurs effets, et d'éventuels effets pléiotropiques des gènes correspondants. Ces différents facteurs exercent chacun leur influence dans la structuration des populations de maïs au niveau des gènes responsables de la variation phénotypique.

Si les caractères quantitatifs sont gouvernés par un grand nombre de gènes à faible effet, McKay et Latta (2002), montrent que l'organisation spatiale de la variabilité observée aux QTLs ne reflète pas nécessairement celle du caractère considéré. Ces auteurs suggèrent que le F_{st} mesuré à partir de marqueurs neutres reflète mieux celui mesuré pour ces QTLs (QTL_{st}) et serait un meilleur estimateur du QTL_{st} que le Q_{st} . Toutefois, aucune donnée expérimentale n'a encore permis de valider cette hypothèse.

Déséquilibre de liaison et association

La chute rapide, sur quelques centaines de paires de bases, du déséquilibre de liaison dans les gènes, a pour conséquence que seuls les polymorphismes responsables de la variation phénotypique montrent une association forte avec le phénotype (corrélation entre le génotype et le phénotype). Il en résultera que certains polymorphismes situés, au sein du même gène, à quelques centaines de paires de base d'un polymorphisme associé à la variation phénotypique, présenteront une association suffisamment faible avec le phénotype pour que la dynamique de leur évolution puisse être considérée comme indépendante de la sélection.

On peut donc très bien imaginer avoir des marqueurs, situés à quelques centaines de paires de bases les uns des autres, qui montrent différents profils d'organisation spatiale de la diversité. D'où, l'intérêt d'étudier à la fois les associations entre polymorphismes et phénotype et les profils d'organisation spatiale de la diversité génétique correspondants (QTL_{st} et Q_{st}).

Polymorphismes associés à la variation phénotypique et déséquilibres de liaison

La chute abrupte du déséquilibre de liaison à l'intérieur de nombreux gènes du maïs est un phénomène abondamment documenté (Remington *et al.* 2001; Tenaillon *et al.* 2002; Tenaillon *et al.* 2001; Thornsberry *et al.* 2001 ; Wang *et al.* 1999). Ces différents travaux ont en commun d'être basés sur l'étude d'échantillons de maïs couvrant une large proportion de l'aire de répartition de l'espèce ou tout du moins l'essentiel. Bien que notre travail ait ciblé une aire géographique considérablement plus réduite, nous avons été en mesure d'observer le long du gène *sh2* une diminution comparable du déséquilibre de liaison (chapitre VII).

Van Tienderen *et al.* (2002) ont récemment préconisé l'utilisation des polymorphismes ou marqueurs situés dans les gènes associés aux caractères d'intérêt pour évaluer la diversité génétique. A la lumière de nos résultats, il est évident que cette suggestion se heurte, chez le maïs, à une contre-indication majeure. Du fait de la chute rapide du déséquilibre de liaison, un tel marqueur ne reflétera pas nécessairement la structuration de la diversité pour les polymorphismes responsables de la variation phénotypique au sein du même gène (voir encadré).

Nous proposons, pour contourner cette difficulté, l'étude des polymorphismes SNP ou IDP associés à la variation phénotypique (Buckler et Thornsberry 2002). L'établissement des profils d'organisation de la diversité associés à ces marqueurs doit permettre de déterminer la relation entre F_{st} , Q_{st} et QTL_{st} .

Les populations de maïs comme modèles d'étude de la différenciation phénotypique

L'importante diversité phénotypique disponible chez le maïs (Buckler *et al.* 2001) offre une grande variété de situations permettant d'étudier *in situ* les différents aspects que l'évolution phénotypique est susceptible d'adopter.

Les populations de maïs cultivées de manière traditionnelle dans les vallées centrales d'Oaxaca, comme nous l'avons vu au cours de cette étude, sont faiblement structurées au niveau des marqueurs ne donnant pas prise à la sélection (chapitre IV). Elles constituent ainsi un référentiel de choix pour examiner l'association entre le polymorphisme le long d'un gène donné et la variabilité phénotypique. Jusqu'ici, seule l'utilisation d'échantillonnages considérablement plus larges avait permis de réaliser de telles expérimentations (Buckler et Thornsberry 2002; Thornsberry *et al.* 2001). En outre, contrairement à ces travaux, les

populations que nous avons examiné ici autoriseraient l'approfondissement des questions demeurées en suspens quant à l'existence de relations entre le F_{st} , le Q_{st} et le QTL_{st} et l'évaluation de la part imputable à la sélection paysanne à l'image de ce qui a été réalisé ici (chapitres V et VI).

Afin d'être en mesure de statuer sur la capacité des différents estimateurs de la variabilité génétique à refléter la diversité phénotypique, il est nécessaire d'apporter un complément à notre étude. Pour cela, nous nous proposons d'examiner le comportement de populations de maïs placées dans des situations différentes de celles de la région de Oaxaca. Par exemple, les populations de maïs se rattachant aux types ("races" pour Wellhausen *et al.* 1952) Cacahuacintle et Salpor sont à la fois phénotypiquement très similaires et géographiquement distantes. En revanche, les populations de type Palomero Toluqueño, qui partagent l'aire de répartition de populations de type Cacahuacintle, s'en distinguent fortement par la morphologie; en outre, Matsuoka *et al.* (2002) ont montré leur proximité phylogénique. La comparaison entre ces deux situations (Palomero Toluqueño/Cacahuacintle et Cacahuacintle/Salpor), illustrant les inégalités extrêmes $Q_{st} \ll F_{st}$ et $Q_{st} \gg F_{st}$, respectivement, nous permettrait probablement d'obtenir un aperçu des relations entre diversité aux marqueurs neutres et diversité phénotypique.

De plus, la ressemblance entre Cacahuacintle et Salpor est explicable soit par un événement de migration ancien précédant un processus de sélection conservatrice, soit par l'application dans ces deux régions d'une sélection convergente. Au niveau des marqueurs moléculaires, cette convergence évolutive peut reposer sur la sélection de polymorphismes alléliques relatifs à des gènes identiques, ou sur la sélection de gènes différents. L'utilisation des polymorphismes SNP et IDP et l'examen de leur association avec la variation phénotypique peut nous permettre, en combinaison avec une approche de génétique des populations, d'identifier les protagonistes et les mécanismes de cette convergence évolutive.

Enfin, nous projetons d'effectuer, en parallèle, des estimations de la différenciation au moyen des différentes classes de marqueurs dans le but de relier l'organisation spatiale de la variation aux QTLs (QTL_{st}) avec celles mesurées à partir de marqueurs neutres (F_{st}) ou des caractères phénotypiques (Q_{st}). En utilisant ces mesures conjointement à une approche de cartographie d'association, il deviendra alors possible de déterminer en quoi la sélection paysanne influe sur la répartition des polymorphismes associés à la variation phénotypique.

Le cF_{st} : une mesure de différenciation tenant compte des associations alléliques

Le cF_{st} est l'extension du modèle ANOVA développé par Weir (1996) ou :

$$X_{ij} = \mu + P_i + G_{ij}$$

X_{ij} est la valeur observée pour le $j^{\text{ème}}$ haplotype de la $i^{\text{ème}}$ sous-population, μ est la fréquence de l'allèle A, P_i est l'effet population et G_{ij} est l'effet des haplotypes dans les populations.

On a alors :

$$F_{st} = \frac{MSP - MSG}{MSP + (n-1)MSG}$$

MSP est le carré moyen entre les populations, MSG est le carré moyen entre individus à l'intérieur des populations et n le nombre moyen d'haplotypes à l'intérieur des populations.

On peut en dériver la matrice des statistiques F multi-alléliques (Long 1986, Yang et Yeh 1993 et Kremer *et al.* 1997) :

$$F_{st} = [MSP + (n-1)MSG]^{-1/2} (MSP - MSG) [MSP + (n-1)MSG]^{-1/2}$$

Dans le cas biallélique on a :

$$cF_{st} = \lambda_{\max}\{F_{st}\} = TR\{F_{st}\}$$

Dans les cas où l'on aurait plus de deux allèles, on considérera, comme suggéré par Yang et Yeh (1993) et par Kremer *et al.* (1997), que :

$$cF_{st} = \lambda_{\max}\{F_{st}\}$$

où λ_{\max} est la valeur maximum prise par le vecteur des valeurs propres et TR la trace de la matrice.

Si les associations alléliques ne sont pas les même d'une population à l'autre :

$$cF_{st} > F_{st}$$

le F_{st} étant alors calculé selon la méthode décrite par Weir (1996) ou celle décrite par Long (1986) qui, à l'inverse de Kremer, n'a pas considéré les associations alléliques (covariance entre allèles) dans son F_{st} multi-allélique :

$$F_{st} = \frac{TR\{Diag[MSP - MSG]\}}{TR\{Diag[MSP + (n-1)MSG]\}}$$

Comprendre la répartition au sein des populations ou collections de maïs de la variation allélique des gènes d'intérêt agronomique, variabilité sur laquelle repose la sélection paysanne.

Les recombinaisons entre gènes liés ont été évoquées comme hypothèse pour expliquer la stagnation et le brusque redémarrage du progrès génétique au sein des programmes de sélection récurrente sur le long terme (exposition d'une variabilité masquée par des linkages négatifs). Sachant que les recombinaisons ont essentiellement lieu au sein même des gènes (Fu *et al.* 2002, Thuriaux 1977), la possibilité de création d'un nouvel allèle par recombinaison entre deux polymorphismes associés à la variation phénotypique est envisageable. On peut aussi penser que ce serait le mécanisme principal générant la variabilité génétique sur laquelle repose la sélection paysanne et *in fine* la diversité et la différenciation phénotypique. En effet, les quatre principales hypothèses pouvant être avancées sur l'origine de la variabilité phénotypique sont (i) la mutation qui permet l'émergence de nouveaux allèles, (ii) la recombinaison intra-génique qui permet également la création de nouveaux allèles, (iii) la recombinaison entre gènes qui expose une variabilité masquée par des linkages négatifs et (iv) l'épistasie qui peut générer de la variance du fait des changements causés par la sélection sur le fond génétique.

Pour des polymorphismes non soumis à sélection, le déséquilibre de liaison est contrôlé par les flux de gènes, la dérive et le taux de recombinaison (tout comme dérive et flux de gènes contrôlent la valeur du F_{st}). Kremer *et al.* (1997) ont développé une mesure de différenciation tenant compte du déséquilibre de liaison (cF_{st} , Kremer *et al.* 1997, voir encadré). Quel est l'attendu du cF_{st} en cas de neutralité pour des polymorphismes physiquement liés ? Quelle est la conséquence de la sélection s'exerçant sur ces différents polymorphismes ? Quelle est la relation du F_{st} , du cF_{st} , du QTL_{st} , du $cQTL_{st}$, du Q_{st} des caractères correspondant à ces polymorphismes (QTLs) et enfin de la structure de la matrice \mathbf{G} , dans une situation de sélection divergente ? En d'autres termes, comment est distribuée la variation allélique des gènes et polymorphismes d'intérêt agronomique soumis ou, ayant été soumis dans le cas des collections, à la sélection paysanne ?

Une telle étude est réalisable chez le maïs du fait (i) des avancées de la cartographie d'association et de l'identification des polymorphismes associés à la variation phénotypiques (Buckler et Thornsberry 2002; Thornsberry *et al.* 2001), (ii) de l'existence de populations présentant des caractéristiques (dérive et flux de gènes) permettant à la fois une très faible

différenciation inter-population pour les marqueurs neutres et (iii) une rapide décroissance du déséquilibre de liaison (décroissance du déséquilibre qui permet la cartographie d'association avec une forte résolution) tout en présentant (iv) une forte différenciation phénotypique qui résulte de la sélection (Sélection divergente et diversifiante). Nous avons montré que ces conditions sont remplies pour les populations traditionnelles des vallées centrales de Oaxaca au Mexique.

B. Stratégies de conservation et de valorisation des ressources génétiques

Pour établir des stratégies de conservation ou d'exploitation pour les populations de maïs, il est indispensable de connaître les profils de distribution respectifs de la diversité moléculaire et de la diversité phénotypique (chapitre V). Les relations existant entre elles ont été modélées à la fois par le parcours évolutif de ces populations, et par les choix effectués par les agriculteurs. Par conséquent, la gestion des ressources génétiques du maïs ne saurait se dispenser de la prise en compte de ces mécanismes, qui influent sur le maintien et la création de la variabilité.

Conservation *in situ*

Lors de nos travaux, nous avons été confrontés, dans un premier temps, aux particularités de la structure des populations traditionnelles de maïs dans les vallées centrales d'Oaxaca. Comme nous l'avons vu, l'importance des flux de gènes apparente ce système à une métapopulation (Louette 2000). La part variable d'hétérogénéité au sein de chacune de ces populations, et l'existence de profils de distribution de la diversité propres à chaque caractère, indique que les populations ou groupes de populations sont soumises au sein de cette métapopulation à une importante sélection divergente.

D'autre part, en comparant nos données aux déclarations des agriculteurs, nous avons acquis la certitude que la sélection exercée par ces derniers résulte en une conservation non rigoureuse des traits " saillants " d'un morphotype. Ce phénomène découle lui-même de l'absence de normalisation des critères de sélection entre les agriculteurs, qui laisse chacun libre de gérer sa population en fonction de sa perception et de l'évolution de ses besoins au cours du temps. Les agriculteurs sont le moteur de la différenciation phénotypique.

Il ressort de ces différentes observations que les deux caractéristiques principales du système que nous avons étudié sont son absence d'homogénéité et son instabilité. Aucune fraction

représentative ne peut donc en être isolée afin de constituer l'unité de gestion des ressources génétiques. Nous proposons de sortir de ce schéma réducteur et de considérer ce système comme un ensemble, autrement dit, de considérer les populations de maïs dans toute la complexité de leur environnement. Cette démarche implique donc de prendre en compte la diversité des niches adaptatives formées par les différentes exploitations paysannes, afin de préserver la dynamique des flux de gènes et des processus sélectifs entretenant la diversité chez ces populations (chapitre V). Il s'agit donc moins d'une conservation de la diversité *in situ*, que d'une gestion de celle-ci, au-travers de la perpétuation des mécanismes qui la génèrent.

Les réflexions qui précèdent introduisent la notion de seuil critique pour la préservation de la diversité existante: une proportion " suffisante " de populations traditionnelles de maïs doit être cultivée selon les méthodes paysannes. Dans l'hypothèse où ces deux conditions ne seraient pas remplies, la fragmentation de l'habitat aurait pour conséquence l'isolement, voire, au terme de l'érosion des flux de gènes, l'extinction des populations traditionnelles de maïs (Chapitre V ; Berthaud *et al.* 2001)

Le facteur humain étant un facteur prépondérant dans le maintien et dans la création de la diversité, cette préservation suppose que les agriculteurs aient une incitation (économique ou culturelle) à une telle conservation des pratiques traditionnelles. Il incombe désormais à l'étude socio-économique qui constitue le pendant de cette thèse (Bellon *et al.* 2002; Smale *et al.* 1999) de déterminer quelles sont, ou pourraient être à l'avenir, les motivations des agriculteurs en faveur de la poursuite de la culture des populations traditionnelles sur l'ensemble d'un village ou d'un ensemble de villages contiguës. Toutefois, cette préservation devient plus difficile car dans les vallées centrales de Oaxaca, les agriculteurs âgés sont nombreux, si bien que la pérennité du système n'est pas assurée. L'immigration des jeunes vers les villes ou vers les Etats-Unis hypothèque l'avenir de l'agriculture traditionnelle. Le maintien dans certaines vallées italiennes de la culture de populations traditionnelles est à ce titre encourageant (maintien en dépit d'une agriculture largement mécanisée et reposant sur l'achat de semences hybrides par la majorité agriculteurs européens). Ce maintien repose essentiellement sur l'existence d'un marché-niche permettant des prix plus élevés. Ceci est possible en raison d'utilisations culinaires particulières du maïs (et donc des caractéristiques particulières du grain) ainsi que d'un attachement culturel à l'agriculture traditionnelle (Barcaccia *et al.* 2003, Lucchin *et al.* 2003). Ce sont des caractéristiques que l'on retrouve aussi dans de nombreuses régions du Mexique et à Oaxaca en particulier.

Palomero Toluqueño et Cacahuacintle

Dans la région où ces deux “ types ” morphologiques se côtoient, il serait faux d’imaginer que les populations appartiennent soit à l’un ou à l’autre. Nombreuses sont les populations de types intermédiaire et aussi les populations qui présentent une importante variabilité intra-population pour les caractères morphologiques définissant les deux idéotypes.



Image tirée de Arqueologia Mexicana Vol 5 Num 25

La meilleure combinaison allélique pour un gène donné

(i) La chute du déséquilibre de liaison à l’intérieur d’un gène donné et (ii) l’existence de plusieurs polymorphismes associés à la variation phénotypique au sein de ce gène, font qu’il pourrait être utile d’identifier les allèles regroupant l’ensemble de polymorphismes favorables.

D’où l’intérêt d’identifier à la fois les associations entre le génotype et le phénotype et les profils d’organisation de la diversité pour ces polymorphismes responsables de la variation phénotypique.

Exploitation des ressources génétiques

Une évaluation satisfaisante de la diversité génétique au niveau des populations traditionnelles de maïs, dans le but de leur valorisation en amélioration, doit permettre (i) de choisir au mieux les populations devant être introduites dans des schémas de sélection ou de conservation afin de maximiser la diversité génétique (ii) d'identifier les populations qui donneront la meilleure réponse à la sélection, (iii) d'identifier les populations présentant les meilleures performances agronomiques dans différents systèmes de test (iv) d'identifier les sources de gènes d'intérêt pour introgression dans le "germplasm" amélioré.

Dans cette optique, il est indispensable de s'interroger simultanément sur l'identification des sources possibles d'allèles d'intérêt, et sur les méthodes à mettre en œuvre dans le but d'en tirer parti. Nous proposons ici une stratégie possible, sans pour autant exclure qu'il en existe d'autres.

Les récents travaux de cartographie d'association ont permis de mettre en évidence de nouveaux polymorphismes (SNP et IDP) liés à la variation phénotypique chez le maïs, en s'appuyant sur la décroissance rapide du déséquilibre de liaison à l'intérieur des gènes (Buckler et Thornsberry 2002; Thornsberry *et al.* 2001). L'étude de marqueurs neutres répartis sur l'ensemble du génome peut permettre, en appliquant la méthode décrite par Pritchard *et al.* (2000), de discriminer la part associée au phénotype de celle liée à la structuration au sein de l'échantillon pour ces marqueurs (Buckler et Thornsberry 2002; Thornsberry *et al.* 2001).

Le complexe Bolita, étudié ici, ou, la région regroupant des populations de maïs de type Palomero Toluqueño et Cacahuacintle (ainsi que les morphotypes intermédiaires, voir encadré), sont potentiellement des systèmes propices à l'identification des SNPs et IDPs associés à la variation phénotypique, et de plus, ils permettent de s'affranchir de la structuration associée à l'isolement tout en maximisant la variabilité phénotypique. Ceci dans des échantillons qui maximisent aussi la diversité génétique. En effet, la diversité observée au niveau de marqueurs neutres dans ces régions représente l'essentiel de la diversité observée à l'échelle de l'espèce toute entière (Matsuoka *et al.* 2002). La délimitation, parmi les populations traditionnelles de maïs, du meilleur échantillonnage sur lequel appliquer la cartographie d'association nous offrirait, à terme, une opportunité de sélectionner la meilleure combinaison allélique (voir encadré) pour un gène donné (Buckler et Thornsberry 2002).

9.5 Conclusions

Le défi auquel nous sommes aujourd'hui confrontés est de comprendre l'évolution phénotypique et adaptative d'une espèce. Notre travail amène un certain nombre de réponses sur l'évolution et l'adaptation des populations de maïs soumises à la sélection paysanne.

Ce travail nous a permis de confirmer que le processus d'adaptation et d'évolution phénotypique du maïs est encore en cours. Nous avons pu déterminer en partie comment les agriculteurs gèrent et exploitent la variabilité phénotypique et le rôle de la sélection en milieu paysan dans les processus d'adaptation du maïs. Notre travail nous a permis de montrer que la sélection paysanne est le moteur de la différenciation phénotypique et altère profondément la structure de la matrice **G** (chapitres V et VI). Nous avons vu que cette dernière observation suggère que la sélection porte sur des gènes à effets majeurs et pléiotropiques. Il reste à comprendre la répartition au sein des populations ou collections de maïs de la variation allélique des gènes d'intérêt agronomique et son origine.

Si la sélection paysanne entraîne une divergence phénotypique de leurs populations, qu'en est-il à l'échelle moléculaire ? Comment l'adaptation et l'évolution phénotypique rapide d'une espèce comme le maïs se résoud-elle à l'échelle moléculaire ? L'étude sur *shrunken2* (chapitre VII) si elle a apporté certains éléments intéressants sur le déséquilibre de liaison et la différenciation entre populations pour des polymorphismes caractérisés le long de ce gène, les conclusions que nous tirons de ce travail ne sauraient être généralisées. D'autres gènes et polymorphismes fonctionnels (polymorphismes associés à la variation phénotypique) sont à même de présenter des situations très différentes. L'identification préalable de ces polymorphismes accompagnée de l'étude de l'organisation spatiale de leurs variations alléliques ainsi que de la caractérisation des déséquilibres de liaison entre eux, abordées lors des paragraphes précédents, permettraient de comprendre l'impact de la gestion et sélection paysanne sur la variation allélique associée au phénotype. Le maïs, du fait de l'existence d'un grand nombre de populations contiguës (échanges de semences ou flux pollinique) tout au long de son évolution, et de son mode de reproduction essentiellement allogame, présente une chute rapide du déséquilibre de liaison, et ce, même à une échelle aussi réduite que celle des vallées centrales de Oaxaca. Nous pouvons ainsi identifier les polymorphismes responsables de la variation phénotypique et caractériser la différenciation, entre et au sein des populations, associée à cette variation allélique.

BIBLIOGRAPHIE

- Agrawal, A. F., E. D. Brodie, 3rd and L. H. Rieseberg, 2001 Possible consequences of genes of major effect: transient changes in the G-matrix. *Genetica* **112-113**: 33-43.
- Aguirre Gómez, J. A., 1999 Análisis Regional de la Diversidad del Maíz en el Sureste de Guanajuato, pp. 78, Thesis, *Facultad de Ciencias*. Universidad Nacional Autónoma de México, Mexico, D.F.
- Anderson, E., 1947. Field Studies of Guatemalan Maize. *Annals of the Missouri Botanical Garden* **34**: 433-467.
- Anderson, E., 1953 Introgressive hybridization. *Biol. Rev.* **28**: 280-307.
- Anderson, E., and H. C. Cutler, 1942 Races of *Zea mays*: I. Their recognition and classification. *Annals of the Missouri Botanical Garden* **32**: 297-322.
- Barcaccia, G., M. Lucchin and P. Parrini, 2003 Characterization of a flint mays (*Zea mays* L. convar mays) Italien landrace: II. Genetic diversity and relatedness assessed by SSR and Inter-SSR molecular markers. *Genet Res and Crop Evol* **50**: 253-271.
- Barton, N., and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising populations. *Heredity* **57**: 357-376.
- Belkhir, K. P., P. Borsa, L. Chikhi, N. Raufaste and F. Bonhomme, 2001 GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations, pp. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.
- Bellon, M. R., J. Berthaud, M. Smale, J. A. Aguirre, S. Taba *et al.*, 2003 Participatory landrace selection for on-farm conservation: An exemple from the central valleys of Oaxaca, Mexico. *Genet Res and Crop Evol* **50**: 401-416
- Bellon, M. R., and S. B. Brush, 1994 Keepers of maize in Chiapas, Mexico. *Economic Botany* **48**: 196-209.
- Benz, B. F., 2001 Archaeological evidence of teosinte domestication from Guila Naquitz, Oaxaca. *Proc. Natl. Acad. Sci. USA* **98**: 2104-2106.
- Berthaud, J., J. C. Clément, L. Emperaire, D. Louette, F. Pinton *et al.*, 2001 The role of local level gene flow in enhancing and maintaining genetic diversity., pp. 81-103 in *Broadening the genetic base of crop production.*, edited by H. D. Cooper, C. Spillane and T. Hodgkin. CABI Publishing in association with FAO and IPGRI, Rome, Italy.
- Bijlsma, R., R. W. Allard and A. L. Kahler, 1986 Nonrandom mating in an open-pollinated maize population. *Genetics* **112**: 669-680.

- Birky, C. W., Jr., P. Fuerst and T. Maruyama, 1989 Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**: 613-627.
- Bonnin, I., J. M. Prosperi and I. Olivieri, 1996 Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): a comparative analysis of population structure. *Genetics* **143**: 1795-1805.
- Brown, A. H. D., and R. W. Allard, 1970 Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. *Genetics* **66**: 133-145.
- Buckler, E. S., and J. M. Thornsberry, 2002 Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol* **5**: 107-111.
- Buckler, E. S., J. M. Thornsberry and S. Kresovich, 2001 Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77**: 213-218.
- Cockerham, C. C., 1969 Variance of gene frequencies. *Evolution* **23**: 72-84.
- Cockerham, C. C., 1973 Analysis of gene frequencies. *Genetics* **74**: 679-700.
- Cockerham, C. C., and B. S. Weir, 1987 Correlations, descent measures: drift with migration and mutation. *Proc. Natl. Acad. Sci. USA* **84**: 8512-8514.
- Coe, E. H., M. G. Neuffer and D. Hoisington, 1988 The genetics of corn, pp. 81-258 in *Maize Breeding and Genetics*, edited by D. B. Walden. Wiley & Sons, New-York.
- Crnokrak, P., and J. Merilä, 2002 Genetic population divergence: markers and traits. *Trends Ecol. Evol.* **17**: 501.
- Devos, K. M., and M. D. Gale, 1997 Comparative genetics in the grasses. *Plant Mol. Biol.* **35**: 3-15.
- Dewey, S. E., and J. S. Heywood, 1988 Spatial genetic structure in a population of *Psychotria nervosa*. I. Distribution of genotypes. *Evolution* **42**: 834-838.
- Dillmann, C., A. Bar-Hen, D. Guérin, A. Charcosset and A. Murigneux, 1997 Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. *Theor. Appl. Genet.* **95**: 92-102.
- Doebley, J., 1998 Maize as a model system for investigating the molecular basis of morphological evolution in plants. *Symp Soc Exp Biol* **51**: 127-132.
- Doebley, J., and L. Lukens, 1998 Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**: 1075-1082.
- Dubreuil, P., and A. Charcosset, 1998 Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. *Theor. Appl. Genet.* **96**: 577-587.

- Dubreuil, P., C. Rebourg, M. Warburton, M. Chastanet, B. Gouesnard *et al.*, 2003 Use of DNA pooling to assess diversity within and among maize populations. Application to the investigation of maize introduction into Europe., pp. in *Maize genetics conference*, Lake Geneva, WI.
- Enjalbert, J., and J. L. David, 2000 Inferring recent outcrossing rates using multilocus individual heterozygosity: application to evolving wheat populations. *Genetics* **156**: 1973-1982.
- Ennos, R. A., 1994 Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250-259.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to quantitative genetics*. Longman, New-York.
- Felsenstein, J., 1986 Population differences in quantitative characters and gene frequencies: a comment on papers by Lewontin and Rogers. *Am. Nat.* **127**: 731-732.
- Flury, B., 1988 *Common principal components and related multivariate models*. J. Wiley, New-York.
- Fu, H., Z. Zheng and H. K. Dooner, 2002 Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proc Natl Acad Sci USA* **99**: 1082-1087.
- Gale, M. D., and K. M. Devos, 1998 Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**: 1971-1974.
- Garnier-Géré, P., 1992 Contribution à l'étude de la variabilité inter- et intra-population chez le maïs (*Zea mays* L.): valorisation d'informations agromorphologiques et enzymatiques., thèse. Institut National Agronomique Paris-Grignon, Paris.
- Gaut, B. S., M. Le Thierry d'Ennequin, A. S. Peek and M. C. Sawkins, 2000 Maize as a model for the evolution of plant nuclear genomes. *Proc Natl Acad Sci USA* **97**: 7008-7015.
- Gauthier, P., B. Gouesnard, J. Dallard, R. Redaelli, C. Rebourg *et al.*, 2002 RFLP diversity and relationships among traditional European maize populations. *Theor Appl Genet* **105**: 91-99.
- Giroux, M. J., J. Shaw, G. Barry, B. G. Cobb, T. Greene *et al.*, 1996 A single mutation that increases maize seed weight. *Proc. Natl. Acad. Sci. USA* **93**: 5824-5829.
- Goldman, I., T. R. Rocheford and J. W. Dudley, 1993 Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. *Theor Appl Genet* **87**: 217-224.

- Greene, T. W., and L. C. Hannah, 1998a Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions. *Proc. Natl. Acad. Sci. USA* **95**: 13342-13347.
- Greene, T. W., and L. C. Hannah, 1998b Maize endosperm ADP-glucose pyrophosphorylase SHRUNKEN2 and BRITTLE2 subunit interactions. *Plant Cell* **10**: 1295-1306.
- Hallauer, A. R., and J. B. Miranda, 1988 *Quantitative genetics in maize breeding*. Iowa State University press, Ames.
- Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman *et al.*, 1996 Evolution of Anthocyanin Biosynthesis in Maize Kernels: The Role of Regulatory and Enzymatic Loci. *Genetics* **143**: 1395-1407.
- Hardy, O. J., and X. Vekemans, 1999 Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145-154.
- Hardy, O. J., and X. Vekemans, 2001 Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution* **55**: 943-954.
- Hardy, O. J., and X. Vekemans, 2002 SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618-620.
- Hendry, A. P., 2002 $Q_{st} \approx F_{st}$? *Trends Ecol. Evol.* **17**: 502.
- Kahler, A. L., A. R. Hallauer and G. C.O., 1986 Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. *Theor Appl Genet* **72**: 592-601.
- Kahler, A. L., D. V. Shaw and R. W. Allard, 1989 Nonrandom mating on tasseled and detasseled plants in an open pollinated population of maize. *Maydica* **34**: 15-21.
- Kremer, A., A. Zanetto and A. Ducouso, 1997 Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics* **145**: 1229-1241.
- Kuittinen, H., A. Mattila and O. Savolainen, 1997 Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity* **79**: 144-152.
- Lande, R., 1979 Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution* **33**: 402-416.
- Lande, R., 1992 Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381-389.
- Latta, R. G., and J. K. McKay, 2002 Genetic population divergence: markers and traits. *Trends Ecol. Evol.* **17**: 501-502.

- Lefort-Buson, M., V. Lavergne, J. J. Daudin, A. Charcosset, S. J.P. *et al.*, 1991 Genetic variability among populations of maize germplasm. 2. Enzymatic polymorphism and its relationship to quantitative trait diversity. *Maydica* **36**: 237-246.
- Lewis, P. O., and D. Zaykin, 2002 GDA 1.1 for Windows, pp. Departement of Ecology and Evolutionary Biology, The University of Conecticut.
- Lewontin, R. C., 1984 Detecting population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* **123**: 115-124.
- Long, A. D., and R. S. Singh, 1995 Molecules versus morphology: the detection of selection acting on morphological characters along a cline in *Drosophila melanogaster*. *Heredity* **74**: 569-589.
- Louette, D., 2000 Traditional management of seed and genetic diversity: what is a landrace?, pp. 109-142 in *Genes in the field. On-Farm conservation of crop diversity.*, edited by S. B. Brush. IPGRI, Rome, Italy.
- Louette, D., A. Charrier and J. Berthaud, 1997 In situ conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. *Economic Botany* **51**: 20-38.
- Louette, D., and M. Smale, 2000 Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. *Euphytica* **113**: 25-41.
- Lucchin, M., G. Barcaccia and P. Parrini, 2003 Characterization of a flint mays (*Zea mays* L. convar mays) Italien landrace: I. Morphological-phenological and agronomic traits. *Genet Res and Crop Evol* **50**: 315-327.
- Lynch, M., M. Pfrender, K. Spitze, N. Lehman, D. Allen *et al.*, 1999 The quantitative and molecular genetic architecture of a subdivided species: *Daphnia pulex*. *Evolution* **53**: 100-110.
- Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quatitative traits*. Sinauer, Sunderland, MA.
- Matsuoka, Y., S. E. Mitchell, S. Kresovitch, M. M. Goodman and J. Doebley, 2002a Microsatellites in *Zea* - variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* **104**: 436-450.
- Matsuoka, Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler *et al.*, 2002b A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **99**: 6080-6084.
- McKay, J. K., and R. G. Latta, 2002 Adaptative population divergence: markers, QTL and traits. *Trends Ecol. Evol.* **17**: 285-291.

- Mendoza Gonzales J., 2000 Participacion de la mujer en la seleccion de semilla de maiz en los valles centrales de Oaxaca, pp. 57 Thesis, *Facultad de Estudios Superiores de Cuatitlan*. Universidad Nacional Autónoma de México, Cuatitlan Izcalli, Edo. Mex, Mexico.
- Merilä, J., and P. Crnokrak, 2001 Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* **14**: 892-903.
- Morgan, K. K., J. Hicks, K. Spitze, L. Latta, M. E. Pfreder *et al.*, 2001 Patterns of genetic architecture for life-history traits and molecular markers in a subdivided species. *Evolution* **55**: 1753-1761.
- Nei, M., 1987 *Molecular and evolutionary genetics*. Columbia Univ. Press., New York.
- Nelson, O. E., and D. Pan, 1995 Starch synthesis in maize endosperms. *Annu Rev Plant Physiol* **46**: 475-496.
- Nevo, E., A. Beiles, A. B. Korol, Y. I. Robin, T. Pavlicek *et al.*, 2000 Extraordinary multilocus genetic organization in mole crickets, Gryllotalpidae. *Evolution* **54**: 586-605.
- Nigh, R. B., 1976. Evolutionary Ecology of Maya Agriculture in Highland Chiapas, México. Unpublished Ph. D. dissertation Stanford University. Ann Arbor: University Microfilms International.
- Oddou-Muratorio, S. R., J. Petit, B. Le Guerroue, D. Guesnet and B. Demesure, 2001 Pollen-versus seed-mediated gene flow in a scattered forest tree species. *Evolution* **55**: 1123-1135.
- Ohta T, 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc Natl Acad Sci USA* **79**: 1940-1944.
- Overall, A. D., and R. A. Nichols, 2001 A method for distinguishing consanguinity and population substructure using multilocus genotype data. *Mol. Biol. Evol.* **18**: 2048-2056.
- Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, P. Castiglioni *et al.*, 1998 Comparative analysis of genetic similarity among maize imbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theor. Appl. Genet.* **97**: 1248-1255.
- Perales-Rivera, H. R., 1998. Conservation and evolution of maize in Amecameca and Cuautla Valleys of Mexico. Ph.D. thesis, University of California, Davis. University Microfilms International, Ann Arbor.
- Petit, C., H. Fréville, A. Mignot, B. Colas, M. Riba *et al.*, 2001 Gene flow and local adaptation in two endemic plant species. *Biological Conservation* **100**: 21-34.

- Phillips, P. C., and S. J. Arnold, 1999 Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**: 1506-1515.
- Piperno, D. R., and K. V. Flannery, 2001 The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc. Natl. Acad. Sci. USA* **98**: 2101-2103.
- Podolsky, R. H., and T. P. Holtsford, 1995 Population structure of morphological traits in *Clarkia dudleyana*. I. Comparison of FST between allozymes and morphological traits. *Genetics* **140**: 733-744.
- Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.
- Prout, T., and J. S. Barker, 1993 F statistics in *Drosophila buzzatii*: selection, population size and inbreeding. *Genetics* **134**: 369-375.
- Provan, J., P. Lawrence, G. Young, F. Wright, R. Bird *et al.*, 1999 Analysis of the genus *Zea* (Poaceae) using polymorphic chloroplast simple sequence repeats. *Plant. Syst. Evol.* **218**: 245-256.
- Raybould, A. F., J. Goudet, R. J. Mogg, C. J. Gliddon and A. J. Gray, 1996 Genetic structure of a linear population of *Beta vulgaris* ssp. *maritima* (sea beet) revealed by isozyme and RFLP analysis. *Heredity* **76**: 111-117.
- Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**: 11479-11484.
- Ribaut, J.-M., D. A. Hoisington, J. A. Deutsch, C. Jiang and D. Gonzalez-de-Leon, 1996 Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. *Theor. Appl. Genet.* **92**: 905-914.
- Rice, E. B., M. Smale and J. L. Blanco, 1998 Farmers Use of Improved Seed Selection Practices in Mexican Maize: Evidence and Issues from the Sierra de Santa Marta. *World Development* **26**: 1625-1640.
- Roff, D., 2000 The evolution of the G matrix: selection or drift? *Heredity* **84**: 135-142.
- Roff, D., 2002 Comparing G matrices: a MANOVA approach. *Evolution* **56**: 1286-1291.
- Rogers, A. R., 1986 Population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* **127**: 729-730.
- Rousset, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219-1228.

- Salanoubat, M., and J. Pernes, 1986 Enzyme polymorphisms within and between European maize populations. *Maydica* **31**: 269-278.
- Sanou, J., B. Gouesnard and A. Charrier, 1996 Isozyme variability in West African maize cultivars (*Zea mays* L.). *Maydica* **42**: 1-11.
- Slatkin, M., and M. J. Wade, 1978 Group selection on a quantitative character. *Science* **75**: 3531-3534.
- Smale, M., J. A. Aguirre, M. R. Bellon, J. Mendoza and I. Manuel Rosas, 1999 *Farmer management of maize diversity in the central valleys of Oaxaca, Mexico: CIMMYT/INIFAP 1998 baseline socioeconomic survey. CIMMYT Economics working paper 99-09*. International Maize and Wheat Improvement Center (CIMMYT), Mexico, D.F.
- Smith, O. S., J. S. C. Smith, S. L. Bowen and R. A. Tenborg, 1990 Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, heterosis and RFLPs. *Theor. Appl. Genet.* **80**: 833-840.
- Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. W.H. Freeman and company, New-York.
- Spitze, K., 1993 Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* **135**: 367-374.
- Stockwell, C. A., A. P. Hendry and M. T. Kinnison, 2003 Contemporary evolution meets conservation biology. *Trends Ecol. Evol.* **18**: 94-101.
- Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. Doebley *et al.*, 2002 Patterns of Diversity and Recombination Along Chromosome 1 of Maize (*Zea mays* ssp. *mays* L.). *Genetics* **162**: 1401-1413.
- Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* **98**: 9161-9166.
- Thomsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen *et al.*, 2001 Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**: 286-289.
- Thomsberry, J. M., M. M. Goodman, J. F. Doebley and E. Buckler, 2002 Analysis of putative maize flowering time genes using association tests, pp. in *Maize Genetics Conference*, Orlando, Fla.
- Thuriaux, P., 1977 Is recombination confined to structural genes on the eukaryotic genome? *Nature* **268**: 460-462.

- van Tienderen, P. H., A. A. de Haan, C. G. van der Linden and B. Vosman, 2002 Biodiversity assessment using markers or ecologically important traits. *Trends Ecol. Evol.* **17**: 577-582.
- Veldboom, L. R., M. Lee and W. L. Woodman, 1994 Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL of morphological traits. *Theor. Appl. Genet.* **88**: 7-16.
- Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis *et al.*, 2002 Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol* **19**: 1251-1260.
- Vigouroux, Y., Y. Matsuoka, M. M. Goodman, G. J. Sanchez and J. Doebley, 2003 Population structure and gene diversity of American maize landraces, pp. in *Maize genetics conference*, Lake Geneva, WI.
- Wang, R. L., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999 The limits of selection during maize domestication. *Nature* **398**: 236-239.
- Weir, B. S., 1979 Inferences about linkage disequilibrium. *Biometrics*: 235-254.
- Weir, B. S., 1996 *Genetic data analysis II*. Sinauer Associates, Inc., Sunderland.
- Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.
- Wellhausen, E., J. Roberts, L. M. Roberts and E. Hernández, 1952 *Races of maize in Mexico: Their origin, characteristics, and distribution*. The Bussey Institution, Harvard University, Cambridge, Massachusetts.
- Whitlock, M. C., 1999 Neutral additive genetic variance in a metapopulation. *Genet. Res.* **74**: 215-221.
- Wilkes, G., 1989 Maize: domestication, racial evolution, and spread, pp. 440-455 in *Foraging and farming: the evolution of plant exploitation*, edited by D. R. Harris and G. C. Hillman. Unwin Hyman, London.
- Wright, S., 1951 The genetic structure of populations. *Ann. Eugenics* **15**: 323-354.
- Wright, S., 1965 The interpretation of population structure by F-statistics with special regards to systems of mating. *Evolution* **19**: 395-420.
- Wright, S., 1969 *Evolution and the Genetics of Populations, vol. 2, The Theory of Gene Frequencies*. University of Chicago Press, Chicago.
- Yang, R. C., F. C. Yeh and A. D. Yanchuk, 1996 A comparison of isozyme and quantitative genetic variation in *Pinus contorta* ssp. *latifolia* by FST. *Genetics* **142**: 1045-1052.
- Zeven, A. C., 1998 Landraces: a review of definitions and classifications. *Euphytica* **104**: 127-139.

Impact of farmers' management on genetic and phenotypic differentiation of traditional maize landraces in the Central Valleys of Oaxaca, Mexico

Abstract

To understand the evolution of maize, it is crucial to assess the impact of management of maize landraces by farmers living in the Central Valleys of Oaxaca, Mexico, where maize was first domesticated. In this thesis, I report how traditional farmer's management of maize populations affects genetic diversity, with special regard to the dynamics of quantitative trait evolution.

The molecular diversity seen in these populations is similar to that observed at the whole species level. These populations show little within-population differentiation ($F_{st}=0.011$) and no isolation by distance. This results from considerable seed exchanges among farmers. However, we observed an unusual within-population Wahlund effect resulting from assortative mating.

While gene flow occurred among the studied populations, high levels of population differentiation for quantitative traits were observed ($Q_{st} \gg F_{st}$). I show that phenotypic diversification is the result of divergent selection by farmers in spite of predominantly conservative selection. This diversification acts differently depending on the traits and maize types considered. Different types of maize are managed differently and I show that divergent selection is not based on the same traits in populations with distinct kernel color. Finally, I show that linkage disequilibria fall dramatically within a few hundred base pairs as observed at the whole species level.

I conclude that farmers' decisions and selection strategies have a great impact on adaptation and phenotypic diversification in maize landraces. The discussion focuses on the relation between the spatial organization of (i) phenotypic variability, (ii) allelic variation responsible for this phenotypic variation, (iii) and on intragenic linkage disequilibria. These different levels of expression of the genetic diversity have to be taken into account in order to establish and elaborate strategies for conserving and using maize genetic resources.

Keywords

Adaptation and phenotypic differentiation, assortative mating, F_{st} , \mathbf{G} matrices, intragenic linkage disequilibria, Mexico, Q_{st}

Résumé

Cette étude a pour cadre les vallées centrales de Oaxaca au Mexique, au cœur du centre de diversité et de la zone de domestication du maïs. Son objectif est d'évaluer comment les agriculteurs façonnent leurs populations traditionnelles et d'identifier les facteurs qui interviennent dans leur évolution.

La diversité au niveau des marqueurs neutres au sein de ces populations est comparable à celle observée à l'échelle de l'espèce. La faible différenciation observée ($F_{st}=0,011$) et l'absence d'isolement par la distance résultent de flux considérables de semences véhiculées par les agriculteurs. Nous constatons tout de même une structuration intra-population associée à l'homogamie pour la floraison.

On observe néanmoins une importante différenciation phénotypique ($Q_{st} \gg F_{st}$). Chaque caractère quantitatif possède un profil de distribution de la variabilité génétique qui lui est propre. Les populations étudiées sont soumises à une sélection divergente s'exerçant de façon indépendante sur de nombreux caractères en fonction du type de maïs considéré ; en effet, cette sélection ne porte pas sur les mêmes caractères sur des populations présentant des grains de couleurs différentes (gérées selon des principes distincts).

Enfin, à l'échelle des vallées étudiées, on constate une chute du déséquilibre de liaison intra-génique sur quelques centaines de paires de bases, à l'image de ce qui est observé pour des échantillons couvrant l'essentiel de l'aire de répartition de l'espèce.

L'agriculteur est le moteur de l'adaptation et de la diversification phénotypique au sein de ces populations. La discussion porte sur la relation entre l'organisation spatiale (i) de la variabilité phénotypique, (ii) de la variation allélique des gènes responsable de cette variation phénotypique, (iii) et du déséquilibre de liaison intra-génique. Cette structuration doit être prise en compte dans l'élaboration de stratégies de conservation et de valorisation des ressources génétiques.

Mots clés

Adaptation et différenciation phénotypique, Déséquilibre de liaison intra-génique, F_{st} , Homogamie, Matrices G , Mexique, Populations traditionnelles de maïs, Q_{st}