Rapport de Master 2 Recherche
Science de l'Univers, Environnement, Ecologie
Spécialité Ecologie, Biodiversité, Evolution
Parcours biologie de la conservation
Année scolaire 2008-2009

Evaluation des risques liés à l'introduction de colza transgénique : étude de faisabilité d'une analyse de la structure génétique du paysage sur la base des fréquences alléliques des marqueurs microsatellites

Présenté le 18 juin 2009 par Yann POTTIER

Responsables : Sébastien OLLIER et Jane LECOMTE
Unité Ecologie, Systématique et Evolution, UMR 8079
Université Paris-Sud 11 – CNRS - AgroParisTech
Département Biodiversité, Systématique et Evolution,
Equipe "Flux de transgènes et Diversité dans les Agro-écosystèmes"

Remerciement

Je tiens à remercier dans un premier temps Sébastien Ollier pour son soutient tout le long du stage, ses explications claires et pour m’avoir sorti de l’embarras plus d’une fois.

Je remercie également Jane Lecomte et Diane Bailleul pour les sympathiques journées champêtres et pour leurs réponses à mes questions.

Sommaire

I. Introduction 4

II. Matériels et méthodes 5

1. Le colza 5

2. Présentation du site 5

3. Présentation des données 6

4. Assignations à partir des variétés 9

5. Analyses des données 10
1 Etude des variétés 11
2 Etude des champs 11
3 Etude des populations férales 13

III. Résultats 15

1 Etude des variétés 15

2 Etude des champs 19
1 Assignation des champs 19
2 Typologies des champs 20

3 Etude des populations férales 23

IV. Discussion 26

Bibliographie 30

Annexes

Annexe n°1 : Liste des variétés avec leurs nombres de génotypes.
Annexe n°2 : Assignation par vraisemblance.
Annexe n°3 : Tableau indiquant le nombre de populations génotypées sur chaque route ou chemin en 2004 et 2005.
Annexe n° 4 : Assignation des champs, doublets d'histogrammes.
Annexe 5 : Etude des populations de 2005 sur la route 5 par ACOM, typologie des marqueurs en fonction de la typologie de référence.

I. Introduction

Les surfaces cultivées avec des organismes génétiquement modifiés ont représenté 125 millions d'hectares dans le monde en 2008. 25 pays ont désormais recours à des cultures biotechnologiques, alors qu'ils n'étaient que six en 1996. Des impacts écologiques liés à la culture de ces plantes ont été identifiés. Il s'agit de la modification du potentiel invasif des plantes associée à l'expression d'un transgène, pouvant conduire à des modifications de la composition des communautés végétales et animales et à l'érosion de la biodiversité (Kelly et al., 2006). La question des flux de gènes au sein des paysages agricoles s'avère au coeur de ces impacts.

Ce stage fait parti d'une étude s'inscrivant dans une ANR intitulée "flux de gènes et impact sur la biodiversité" (http://www.ese.u-psud.fr/spip.php?article24) qui vise à comprendre les conséquences potentielles de la mise en culture de plantes génétiquement modifiées sur les flux de gènes et la biodiversité à l'échelle du paysage. Pour cela il est nécessaire de mieux comprendre comment s'organise et évolue la variabilité génétique à cette échelle. Bien que cette problématique concerne toutes les plantes cultivées, cette étude porte sur le colza car cette plante a une forte capacité à se maintenir dans l'environnement après avoir été cultivée, ce qui pourrait lui permettre de servir de réservoir de gènes à partir desquels un transgène pourrait se propager. Ce risque de propagation est bien évidemment lié à la structure du paysage et à son évolution dans le temps. Des travaux ont été effectués sur cette problématique, notamment des modèles démographiques sur la dynamique et la dispersion des populations de colza (Garmier A. ; 2006). Ils ont montré que les populations férales proviennent des champs et du transport des graines, que les banques de graines (graines dans le sol) ont une plus grandes importance dans le maintient des populations férales que l'auto-recrutement. Par ailleurs, des données génétiques ont été récoltées afin de compléter les données démographiques mais n'ont été que partiellement étudiées dans la thèse de Sandrine Pivard qui s'est limitée aux assignations et à la description de la composition variétale du paysage et son évolution (Pivard S. ; 2006).

Mon travail s'inscrit dans la continuité de ces travaux et vise à approfondir l'étude de ces données génétiques. Dans un premier temps l'objectif est d'évaluer la variabilité génétique afin de comprendre en quoi elle est susceptible de conditionner la propagation d'un gène dans un environnement particulier à l'échelle du paysage. Des données génétiques, consistant en un typage microsatellites de graines et de feuilles prélevées dans des champs de colza, mais également dans les repousses de colza que l'on retrouve le long des routes et des

chemins, ont été recueillies entre 2002 et 2005. L'objectif de ce stage est de tester la faisabilité d'une analyse sur la base des fréquences alléliques des marqueurs microsatellites. Dans la mesure où elles seraient concluantes, les recherches viseront in fine à fournir des outils d'aide à la décision et à la gestion des flux de transgènes dans le paysage, ainsi qu'à leur modélisation.

II. Matériels et méthodes

II.1. Le colza

Le colza (*Brassica napus*) est une plante issue d'un croisement spontané entre un chou (*Brassica oleracea*) et une navette (*Brassica rapa*). L'origine de cet hybride n'est pas encore élucidée. L'hybride aurait été sélectionné ensuite sous deux formes : le colza pour son huile et le rutabaga pour ses racines. De nos jours, le colza est cultivé principalement pour son huile alimentaire, mais aussi plus récemment pour la production de biocarburant. Son cycle dure dix mois. Le semis a lieu en septembre jusqu'à la récolte en juillet en passant par une phase de floraison en avril. *Brassica napus* s'hybride avec un nombre certain d'apparentées, souvent adventices des cultures, en sympatrie dans certaines régions, au sein du genre *Brassica* (principalement avec *Brassica Rapa)* ou même avec d'autres genres de Brassicaceae. Les rétrocroisement se produisent spontanément entre les hybrides pouvant entraîner l'introgression (Jørgensen et Andersen, 1994 ; Hansen *et al.*, 2001), et de la transgénèse (Frello et al., 1995 ; Lefol et al., 1995 ; Mikkelsen et al., 1996 ; Chèvre et al., 1997).

II.2. Présentation du site

La zone d'étude est centrée sur un silo de collecte de graines et englobe la commune de Selommes, ainsi qu'une partie des communes de Rhodon, Baigneux, Villetrun, Perigny, Villemardy et Villegrimont pour un total de 42km^2 dans le département du Loir-et-cher. Le plan d'occupation des sols a été numérisé à l'INRA d'Avignon puis corrigé avec un logiciel de SIG. La zone d'étude est essentiellement composée de terres agricoles (Figure II.2.).

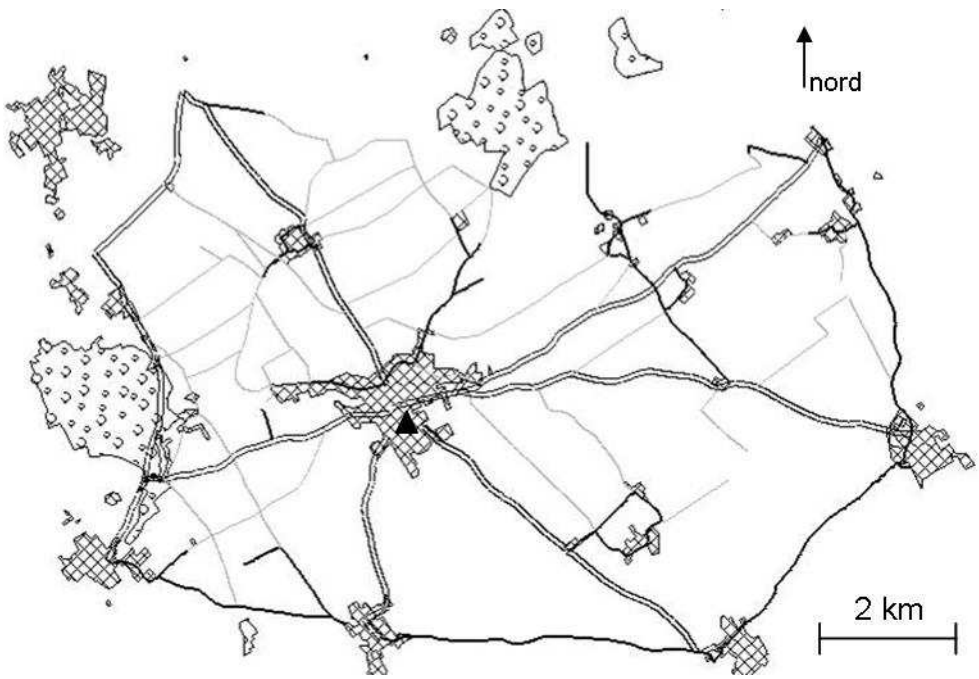


Figure II.2: Réseaux routiers de Selommes, composé de chemins (fines lignes grises), routes à une voie (fines lignes noires) et de routes à deux vois (lignes doubles). L'emplacement du silo (triangle noir). Les polygones quadrillés représentent les villages et les autres polygones représentent des forêts.

II.3. Présentation des données.

Le colza présent sur la zone étudiée se trouve sous plusieurs "formes" :

- en champs de colza
- en repousses dans un champ autre que du colza ou dans une jachère. Le colza provient ici de graines produites par le champ de l'année précédente ou d'années antérieures (banque de graines)
- en repousses sur les bords de routes ou de chemins (populations férales). Le colza provient ici de graines provenant d'un champ contigu de l'année précédente ou d'une année antérieure (banque de graine), ou bien de graines apportées par le transport, ou de graines provenant de l'autorecrutement

Les données génétiques consistent en un typage microsatellites de graines et de feuilles prélevées entre 2002 et 2005 dans des champs de colza mais également dans les repousses de colza que l'on retrouve le long des routes et des chemins. La totalité des voies de transport de la zone est suivie bisannuellement, sauf lorsque les conditions météorologiques rendent certaines voies impraticables.

Jusqu'en 2003, le suivi des populations férales consistait à prélever en juillet des graines matures (une grosse poignée de siliques) plante à plante sur au maximum 10 individus de chaque population présente à cette période ou la totalité des individus si leur nombre est

inférieur ou égale à 10. A partir de 2003, des feuilles ont commencé à être prélevées, selon les mêmes effectifs, sur une partie des populations présentes en juillet 2003 et sur la totalité en juillet 2004. Réalisant qu'une partie non négligeable des populations férales échappait aux prélèvements car disparues (probablement fauchées ou traitées chimiquement) avant juillet, la dernière évolution méthodologique a consisté à effectuer les prélèvements de feuilles à floraison en 2005.

Compte tenu du nombre de populations présentes dans la zone et de leurs effectifs, il est impossible d'enregistrer par GPS la position de chaque plante férale individuellement. Donc sont considérées dans le même ensemble de coordonnées GPS toutes les plantes séparées de moins de 10 mètres les unes des autres. Les populations sont donc définies géographiquement, comme un « continuum » de plantes. Il est important de noter que les populations férales à Selommes sont linéaires et suivent le réseau de routes et chemins, du fait que la quasi-totalité des espaces « marginaux » sont linéaires et qu'il n'y a pas d'espaces entre champs.

Les populations sont repérées depuis une voiture roulant à vitesse très réduite (5 à 20 km/h) sur le coté droit de la route ou du chemin, par l'observation attentive à plusieurs personnes des bordures et du contenu des champs. Chaque côté est parcouru individuellement. Lorsque des plantes sont détectées, l'enregistrement de points GPS est démarré au plus près de la première plante et se poursuit au rythme d'un point par seconde jusqu'à la dernière plante visible. Un membre de l'équipe descend et vérifie l'absence de plantes dans les 10 m suivants.

En ce qui concerne les champs, dix prélèvements sont fait pour chacun en deux endroits depuis 2003, en 2002 la variété des champs a été définie par une enquête.

données 2002 :		
	champ	pop
G	0	2544
F	0	0
données 2003 :		
	Champ	pop
G	2026	1363
F	0	0
données 2004 :		
	Champ	pop
F	687	168
G	832	1436
données 2005 :		
	champ	pop
G	0	0
F	894	1400

Tableau II.3 : Présentation des effectifs pour lesquels des données génétiques sont disponibles; (G : graine, F : feuille) prélevée sur des plantes en champs (champ), sur des populations férales (pop).

Les données disponibles sont très hétérogènes suivant les années (Tableau II.3.) :

- en 2002 toutes les graines proviennent de populations de bord de routes
- en 2002 et 2003 seules des graines ont été récoltées
- en 2004 des graines et des feuilles ont été récoltées
- en 2005 seules des feuilles ont été récoltées

Les graines et les feuilles ne fournissent pas la même information. En effet l'information génétique contenue dans la feuille provient de la plante mais l'information génétique contenue dans la graine provient en parti de la plante et en parti d'une plante dont on ne connaît pas la localisation en cas d'allofécondation. On considère que le taux d'autofécondation est de 63%.

Les données sur lesquelles j'ai travaillé (2002 à 2005) étaient disponibles dès le début de mon stage, mais j'ai participé à la campagne de prélèvement des données pour l'année en cours, ce qui m'a permis de mieux appréhender la complexité du paysage agricole et la réalité du terrain.

II.4. Assignations à partir des variétés.

L'objectif de l'étude est de caractériser les flux de gènes entre champs et populations férales et évaluer leur maintien dans le temps et l'espace. Une première approche afin de suivre le flux génétique est d'assigner les plantes des populations et des champs à leur variété pour pouvoir faire le lien entre elles. Pour cela, nous comparons le génotype des éléments génotypés (graines ou feuilles) avec les génotypes de références d'au moins 30 graines par variété génotypé et ce pour 49 variétés connues pour avoir été cultivées dans la région de Selommes. Les graines ont été obtenues chez les semenciers et au Cétiom (centre technique interprofessionnel des oléagineux métropolitains). Chacune de ces variétés contient un ou plusieurs génotypes différents (annexe n° 1).

A partir des données microsatellites, des assignations ont été faites grâce à un programme réalisé et exécuté sous R (Ihaka et al., 1996) par Diane Bailleul en thèse actuellement. L'assignation consiste à associer une variété à un génotype donné. Cela est fait en comparant les sites microsatellites des échantillons aux fréquences alléliques des variétés par vraisemblance. (Des résultats de l'assignation seront utilisés dans ce travail mais l'assignation elle-même n'étant pas le sujet de mon stage les explications seront développées en annexe 2).

Dans un premier temps les graines des variétés que l'on nomme graines de référence ont été réassignées aux variétés pour vérifier la cohérence des assignations. Cela donne des résultats corrects (génotype de la graine assigné correctement à sa variété) pour 95 % des génotypes (Communication personnelle de Diane). Cependant la méthode des assignations est limitée par le brassement génétique dû à la reproduction sexuée qui fait perdre de la précision aux assignations à chaque génération mais aussi par la proximité génétique des variétés qui dérivent souvent les une des autres, ce qui est problématique pour étudier les flux et le maintient de gènes.

II.5. Analyses des données

Pour décrire la structure génétique et la modéliser à partir des fréquences alléliques, nous ne pouvons pas utiliser les modèles de génétique des populations classiques, car les mécanismes utilisés dans ces modèles (mutation, sélection, dérive …) et les hypothèses d'Hardy Weinberg ne sont pas vérifiées du fait de la nature des processus en jeux dans un tel agro-écosystème. Les compositions alléliques des champs sont par exemple contrôlées par l'homme et non issus de la reproduction des populations présentes sur le site lors de la génération précédente. Pour les populations hors champs, les compositions alléliques sont fortement influencées par la migration constante d'allèles provenant des champs.

C'est pourquoi on souhaite mener une analyse exploratoire sur la base des fréquences alléliques des marqueurs, et l'objectif de mon stage est de tester la faisabilité de cette approche en comparant les résultats avec ceux de obtenus par assignation, et en essayant de retrouver des structures génétiques attendues. Les analyses de données sont faites avec le logiciel et les packages ade4 (Dray S. et al., 2007) et adegenet (Jombart T., 2008).

Dans un premier temps nous transformons les données génétiques en fréquences alléliques. On obtient alors un tableau donnant les fréquences alléliques de chaque allèle de chaque marqueur pour l'ensemble des individus génotypés. L'analyse de ces tableaux a pour objectif de mettre en évidence la diversité génétique exprimée sur l'ensemble des marqueurs moléculaires, et de voir si cette diversité présente une structuration révélatrice de processus écologiques intéressants. Elle vise notamment à mettre en évidence des différences ou des ressemblances en termes de compositions des champs et compositions des populations férales en tenant compte de leur proximité spatiale et temporelle.

Une des méthodes efficace et couramment utilisée pour visualiser ces différences est l'analyse en composantes principales (ACP) du tableau des fréquences alléliques où les individus-lignes correspondent aux graines ou aux feuilles, et les variables-colonnes correspondent aux différents allèles des différents marqueurs (Hotelling, 1933 ; Menozzi,1978 ; Shim S et al., 2000). Mais cette méthode mélange les informations de tous les marqueurs sans que l'on sache si ces informations concordent entre elles, on ne sait donc pas si la typologie obtenue est partagée par l'ensemble des marqueurs, ou consiste en une simple juxtaposition de structures individuelles.

C'est pourquoi l'information apportée par chaque marqueur est ensuite visualisée avec une ACP par marqueur avant de coordonner les représentations des différents marqueurs et de

rechercher une typologie consensuelle. L'analyse de co-inertie multiple (ACOM) (Chessel et al., 1996), déjà utilisée en écologie (Bady et al., 2004 ; Hedde et al., 2005), permet d'extraire l'information commune à l'ensemble des marqueurs (Laloë et al., 2007). Ces deux analyses multivariées (ACP et ACOM) sont réalisées sur trois types de tableaux : celui donnant les fréquences alléliques des graines des variétés de références ; celui donnant les fréquences alléliques des graines/feuilles des champs ; celui donnant les fréquences alléliques des populations férales.

II.5.1 Etude des variétés.

Dans un premier temps nous faisons une ACP globale, puis une ACP par marqueur et une ACOM, sur les données microsatellites des variétés pour tester la capacité de discrimination de ces méthodes. En effet, si les variétés ne se distinguent pas les unes des autres lors de ces analyses, l'étude ne sera guère envisageable par une approche basée sur les fréquences alléliques.

II.5.2 Etude des champs.

Dans la partie "étude des champs" nous allons comparer l'assignation à l'approche par fréquences alléliques sur les données de l'année 2004. Pour l'étude des champs nous disposons non seulement de l'information génétique transformée en tableau de fréquences alléliques mais également de plusieurs fichiers donnant l'information spatiale relative au positionnement des champs :

- les fichiers shapes contenant les identifiants des prélèvements et leur localisation spatiale,
- les fichiers shapes contenant les polygones représentant les champs dans l'espace,
- un fichier avec la correspondance entre les identifiants des prélèvements et les polygones représentant les champs.

Un travail important a consisté à mettre en forme ces données afin de faire correspondre les identifiants des prélèvements fait en bordures de champs avec les champs eux-mêmes (Figure II.5.2.1). Par exemple, en rendant utilisable par le logiciel R le fichier qui fait la correspondance entre les identifiants des prélèvements et les polygones représentant les champs. Ou encore, pour corriger les erreurs tel les identifiants des prélèvements qui ne

correspondaient à aucun champ (48) ou à l'opposer ceux qui correspondaient à plusieurs champs.

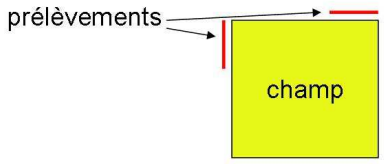


figure II.5.2.1 : illustration de la localisation des prélèvements d'un champs et du polygone représentant la surface du champs.

En ce qui concerne l'assignation, elle se fait au niveau des individus. Hors ici, ce qui nous intéresse est d'assigner un champs, c'est-à-dire un ensemble d'individus censés être de la même variété. Pour faire cette assignation à l'échelle d'un champ, nous avons développé une fonction graphique qui permet de visualiser les résultats issus des assignations individuelles en terme de composition d'un champ sous la forme de deux histogrammes (figure III.2.2).

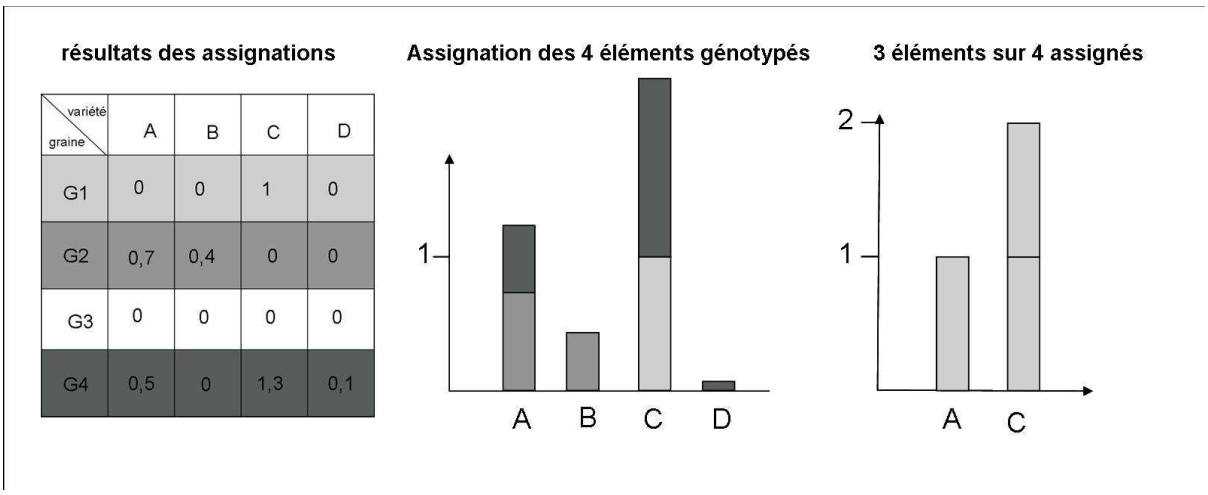


Figure II.5.2.2 : exemple de ce que contient un doublet d'histogramme présentant la composition d'un champ.

-à gauche : tableau présentant les résultats des assignations de graines. Les lignes sont les graines assignées, les colonnes sont les variétés auxquelles les graines sont assignées. (le tableau est présenté ici pour la compréhension mais il est absent des figures présentées dans les résultats ; les valeurs sont purement arbitraires)

- au milieu : un histogramme représentant les assignations des éléments génotypés (graine ou feuille) pour chaque champ. En abscisse les variétés et en ordonnée la

vraisemblance de l'assignation. Chaque élément génotypé est représenté par un niveau de gris.

- à droite : un autre histogramme présentant le résultat de l'assignation de chaque élément génotypé du champ, le nombre d'éléments génotypés par variété est en ordonné, les variétés sont en abscisse.

L'assignation est faite en observant les histogrammes de chaque champ et en notant la variété à laquelle nous décidons d'assigner le champ. Parfois, quand la composition est trop hétérogène, il est impossible d'assigner une variété à un champ.

Pour observer si l'approche par fréquences alléliques permet une bonne discrimination (champs de même variété regroupés ensemble sur la typologie) et comme nous voulons comparer cette méthode a l'assignation, nous faisons la typologie avec les données des champs qui ont pu être assigné aux quatre principales variétés (AVO, PLN, CNY et BJO).
Nous faisons une ACP des éléments génotypés que l'on visualise de deux manières :

-en affichant sur le plan factoriel de l'ACP tous les éléments génotypés étiquetés en fonction de la variété auquel leur champ est assigné (figure III.2.2.1).

-en affichant sur le plan factoriel de l'ACP les centres de gravités des éléments assignés par champs liées par un trait à l'étiquette de leur assignation (figure III.2.2.2).

Pour observer la capacité de discrimination de chacun des marqueurs nous faisons une ACP pour chaque marqueur des éléments génotypés en plaçant sur le plan factoriel le centre de gravité des champs assignés à AVO, PLN, CNY et BJO (figure III.2.2.3). Puis nous finissons l'analyse par une ACOM.

II.5.3 Etude des populations férales.

Lors de l'acquisition des données le critère qui définit un ensemble de plants de colza de bord de route comme une entité population est qu'il n'y ait pas d'interruption de dix mètres ou plus entre les colzas. Ce critère crée, dans les données, des populations de longueurs importantes qui ne sont pas cohérentes du point de vue fonctionnelle et qui rend par ailleurs difficile un suivi pluriannuel. En effet, si entre deux populations étant génétiquement différentes (par exemple si elles se situent en bordure de champs de variété différentes) se trouvent des individus isolés mais étant séparés de moins de 10m, les deux populations et les

individus isolés seront notés et prélevés comme étant une seule population et leurs données génétiques ne seront pas distincts.

Afin de pouvoir étudier l'évolution de la composition génétique des populations dans le temps, les populations sont projetées sur des segments de trois mètres. Pour cela un fichier de représentation spatial des segments a été créé, contenant 84 402 segments de trois mètres couvrant le réseau de routes et de chemins de la zone étudiée. Puis les populations sont projetées sur les segments qu'elles recouvrent avec les informations identifiant la route, la direction (pour connaître de quel côté de la route se situe la population), et les segments couvrant le deux extrémités de la population.

Exemple d'étude d'une route.

On fait l'étude des populations de bordures de deux manières. Dans un premier temps, on étudie l'évolution temporelle entre deux populations se situant au même endroit d'une année sur l'autre, puis dans un deuxième temps on étudie le lien spatial entre les populations présentent le long d'une route.

-évolution temporelle :

A partir de la table indiquant la présence de populations sur chaque route et chemin suivant les années (Annexe n°3), on sélectionne une route où des populations sont génotypées plusieurs années. On observe si elles se recouvrent ou se chevauchent entre 2004 et 2005 pour faire un suivi de l'évolution génétiques de ces populations et observer si un tel suivi est faisable par l'approche par fréquences alléliques.

Pour cela, après avoir sélectionné une population étant présente en 2004 et en 2005 et remarqué que l'on dispose des données génétiques du champ la bordant en 2004 (Figure III.3.2), on fait une ACP des fréquences alléliques des éléments génotypés de ces populations (Figure III.3.3).

-lien spatial :

Le lien entre les populations bordant la route 5 est étudié pour l'année 2005. Dans un premier temps on fait une ACP des fréquences alléliques des éléments génotypés des populations se trouvant de chaque côté de la route (Figure III.3.4) dans le but d'observer une structuration ou l'absence de structuration sur le plan factoriel. Dans un deuxième temps, on visualise les résultats des deux premiers axes de l'ACP sur la carte représentant la route 5 (Figure III.3.5) pour observer la présence ou l'absence de structuration spatiale.

. III. Résultats

III.1 Etude des variétés.

Typologies des graines de variétés en fonction de leurs fréquences alléliques :

Dans l'ensemble du rapport, les histogrammes représentant la part d'information relative (valeur propre) apportée par chaque axe dans les ACP ou ACOM des données microsatellites seront présentés en petit sur leur figure du plan factoriel.

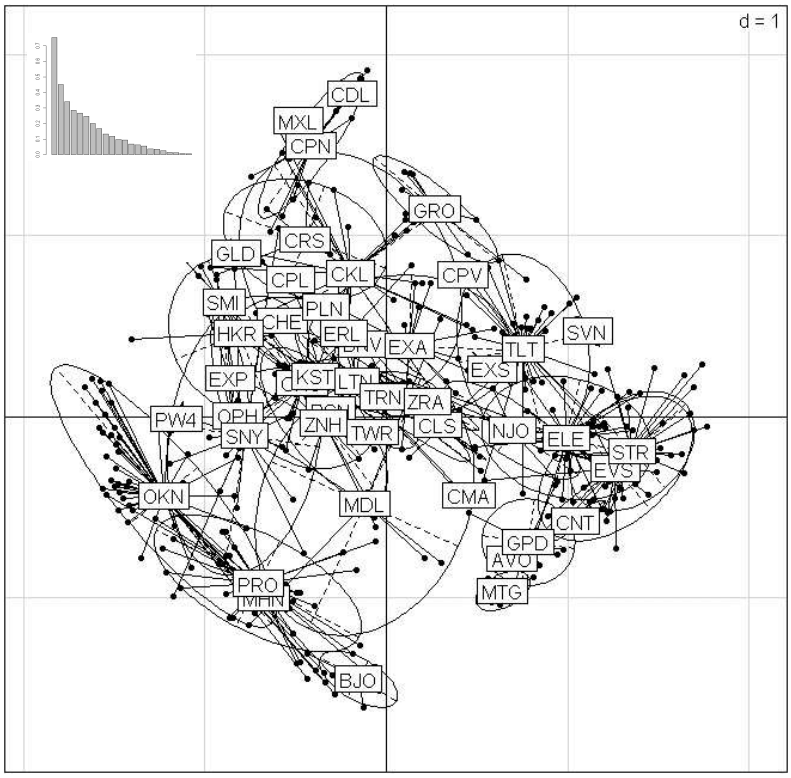


Figure III.1.1 Plan factoriel de l'ACP des données microsatellites des graines de référence des variétés, les ellipses de confiance délimitent la zone où se situe 95% des éléments génotypés de chaque variété. Les variétés sont étiquetées par leur diminutif (voir annexe 1 pour le lien nom/diminutif).

A partir de la figure III.1.1 on peut dire que les éléments génotypés sont discriminables en fonction de leur variété. Les deux premiers axes expliquent 35,6% de la variabilité. La forme de l'histogramme montre qu'il existe une structure. Cette structure reflète en partie les différences en termes de variétés. Etant donné le nombre de variétés étudié ici et le fait que certaines soient très proches génétiquement il est normal que des variétés soient proche sur le plan factoriel. Si la zone étudiée est cultivée principalement avec des variétés qui sont très proches en termes de fréquence alléliques, on ne discriminiera pas la variabilité génétique avec l'approche par les fréquences.

Pour observer le rôle des marqueurs dans la discrimination de l'ACP et quels marqueurs sont les plus discriminants on fait une ACP par marqueur.

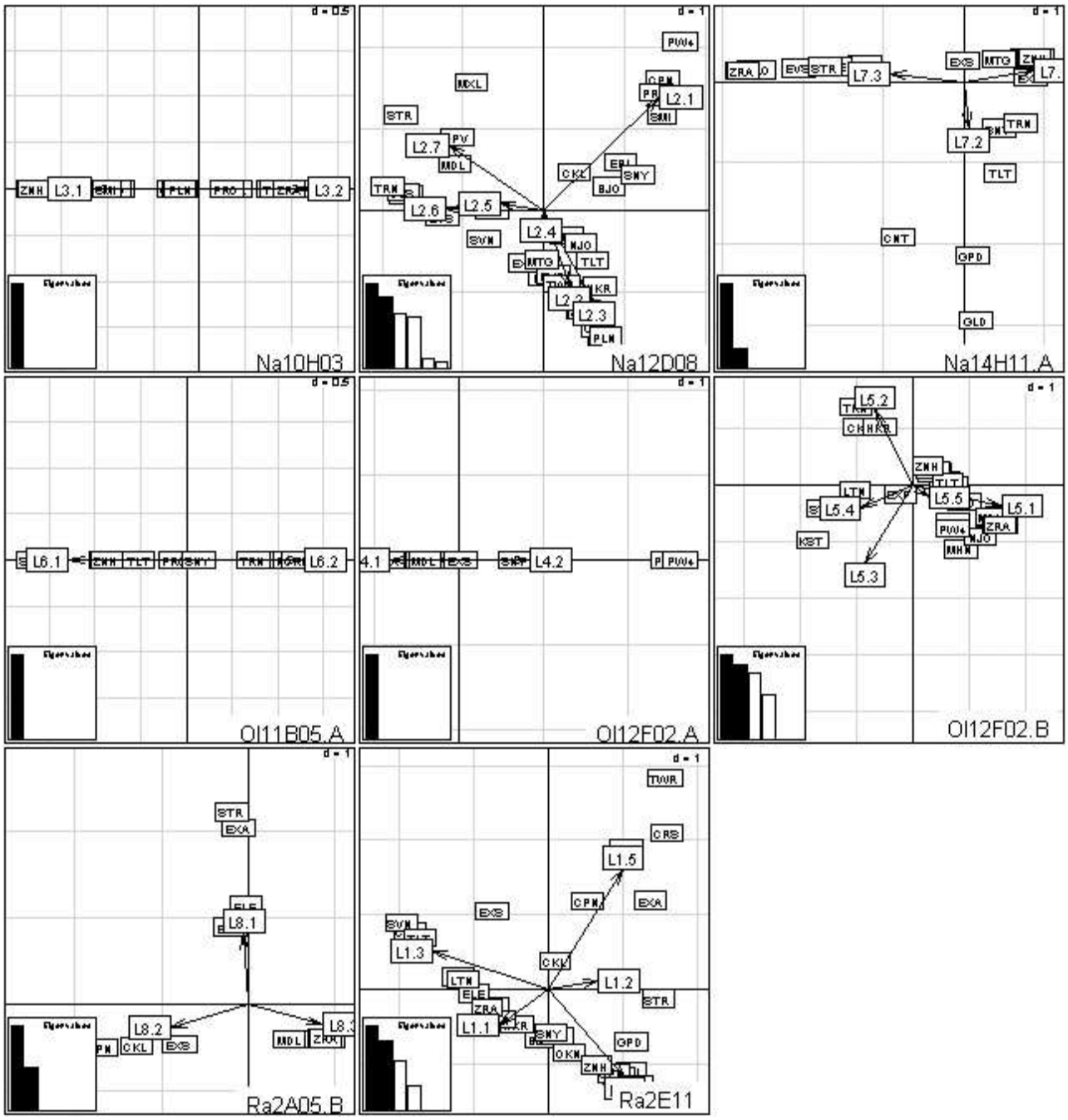


Figure III.1.2 Plans factoriels des l'ACP des données microsatellites des graines de référence des variétés (une ACP par marqueur). Les histogrammes des valeurs propres des l'analyses sont fournis en bas à gauche de chaque plan factoriel.

Exemples d'informations apportées par l'ACP par marqueur: les marqueurs Na10H03, Ol11B05.A et Ol12F02.A discriminent les variétés pour seulement 2 allèles en fonction de la fréquence de la présence de ces allèles dans la variété. L'allèle Ra2A05.B discrimine les variétés en trois groupes distincts : il est plus polymorphe.

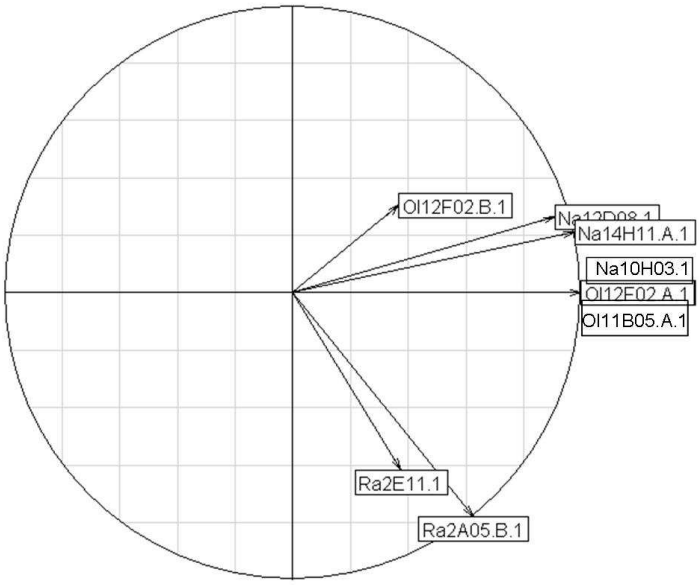


Figure III.1.3 les premiers axes de co-inertie multiple (ACOM) des données microsatellites des graines de référence des variétés.

Les premiers axes montrent si les marqueurs apportent une information différente ou proche. L'information apportée par les marqueurs est très proche à l'exception de Ra2E11 et Ra2A05.B qui sont séparés des autres. L'ACOM va construire une typologie de référence en privilégiant la famille de marqueurs donnant des typologies semblables.

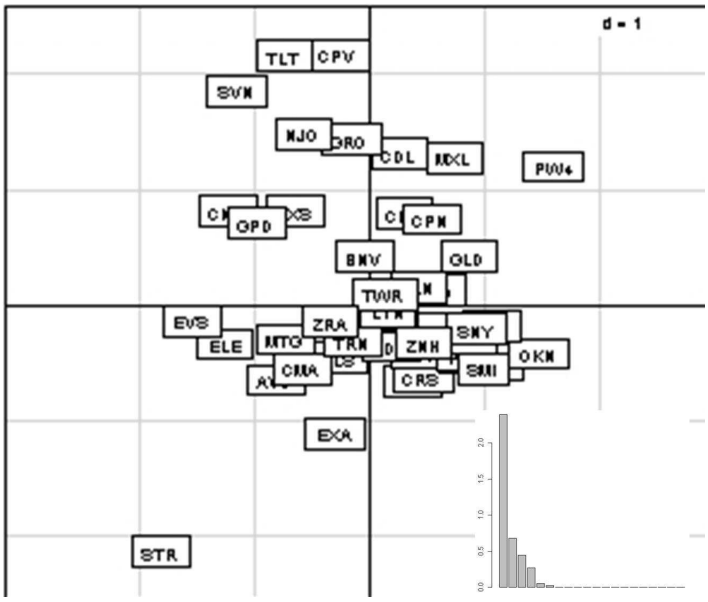


Figure III.1.4 : Typologie de référence de l'analyse de co-inertie multiple (ACOM) des tableaux de fréquences alléliques des huit marqueurs. L'histogramme des valeurs propres de l'analyse est en bas à droite.

Dans ce cas l'ACOM ne permet pas une meilleure discrimination des variétés que l'ACP mais elle présente une typologie plus cohérente avec les typologies apportées par chaque marqueur car elle trouve une combinaison des marqueurs qui maximise la typologie (Laloë et al., 2007) (figure III.1.5). Les deux premiers axes expliquent 82,55% de la variabilité. On le voit bien sur l'histogramme représentant la part d'information relative

(valeurs propres). Les deux premiers axes contiennent une plus grande partie de la variabilité que ceux de l'ACP (Figure III.1.1).

III.2 Etude des champs.

III.2.1 Assignation des champs

Etant donné le grand nombre de champs à assigner seuls quelques exemples caractéristiques de ce que l'on obtient sont montrés ici.

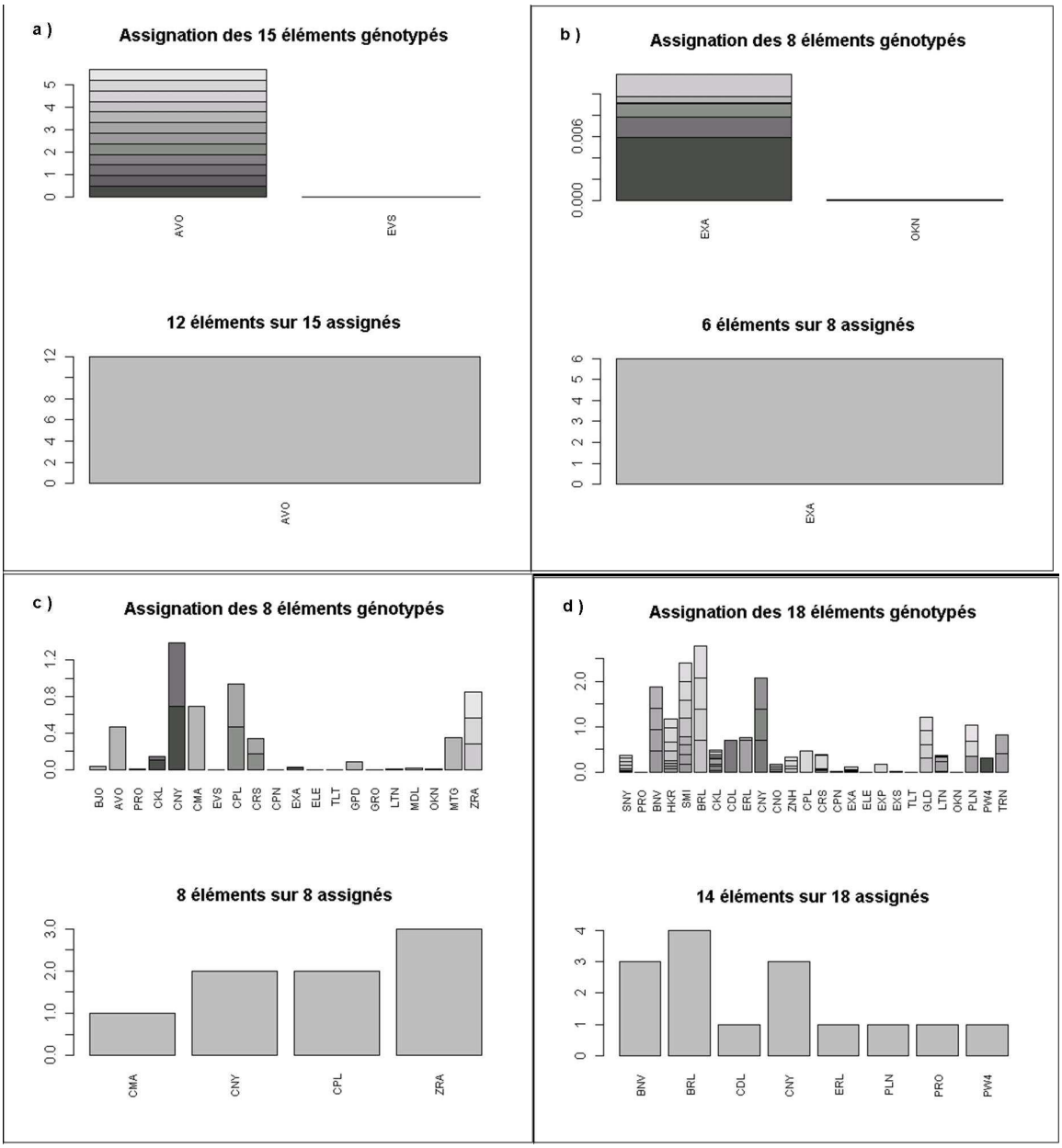


figure III.2.1 : exemples de doublets d'histogrammes représentant la composition d'un champ. L'histogramme du haut présente en abscisse les variétés et en ordonné la vraisemblance de l'assignation des éléments génotypés. Chaque élément génotypé est représenté par un niveau de gris. L'histogramme du bas présente les résultats de l'assignation de chacun des éléments génotypés du champ.

La figure III.2.1.a. montre un cas où l'assignation du champ à une variété est évidente, 12 des 15 éléments génotypés de ce champ sont assignés à la même variété, avec une vraisemblance suffisamment élevée.

La figure III.2.1.b montre un cas où l'assignation du champ à une variété semble évidente, en effet 6 des 8 éléments génotypé de ce champ sont assignés à la même variété mais la vraisemblance de ces assignations est très faible pour certains éléments génotypés. Cela peut être du à la pondération qui pondère la probabilité qu'un élément génotypé soit assigné à une variété lorsque cet élément correspond à un génotype qui est peu fréquent dans la variété.

La figure III.2.1.c montre un cas où l'assignation des éléments génotypés est claire sauf pour un élément qui est assigné (histogramme du haut) à 4 variétés (AVO, CMA, GPD et MTG), mais l'assignation du champ n'est pas faisable car il n'y a pas de variété majoritaire.

La figure III.2.1.d montre un cas où l'assignation du champ à une variété n'est pas claire, en effet chaque élément génotypé de ce champ est assigné à plusieurs variétés (histogramme du haut). Et ce champ contient au minimum 8 variétés.

III.2.2 Typologies des champs.

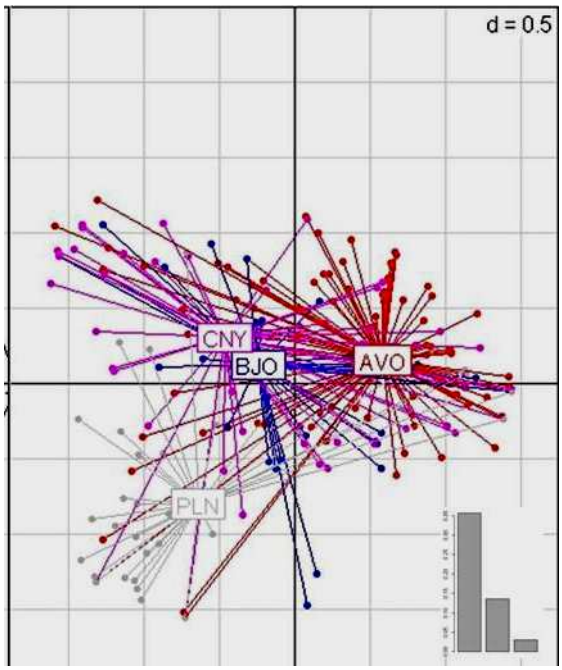


figure III.2.2.1 : Plan factoriel de l'ACP pour les champs assignés à AVO, PLN, CNY et BJO, les éléments génotypés sont colorés en fonction de l'assignation de leur champ.

AVO et PLN sont bien différenciées, mais CNY et BJO sont peu différenciés avec l'approche par fréquences alléliques car il y a une forte variabilité individuelle et elles doivent avoir des allèles communs. En effet le bruit réduit lors de l'assignation des champs est important (cf figure III.2.1.d, il y a maximum 4 graines qui appartiennent à la même variété), cela explique la dispersion des éléments génotypés de chacun des champs sur le plan factoriel car les fréquences alléliques des éléments génotypés ne correspondent pas tous à la variété auquel le champ a été assigné.

Pour réduire la variabilité individuelle on regarde ce que donne la typologie sur la base des champs (figure II.2.4), au lieu de représenter les individus, on représente les champs par le centre de gravité des individus qui le compose :

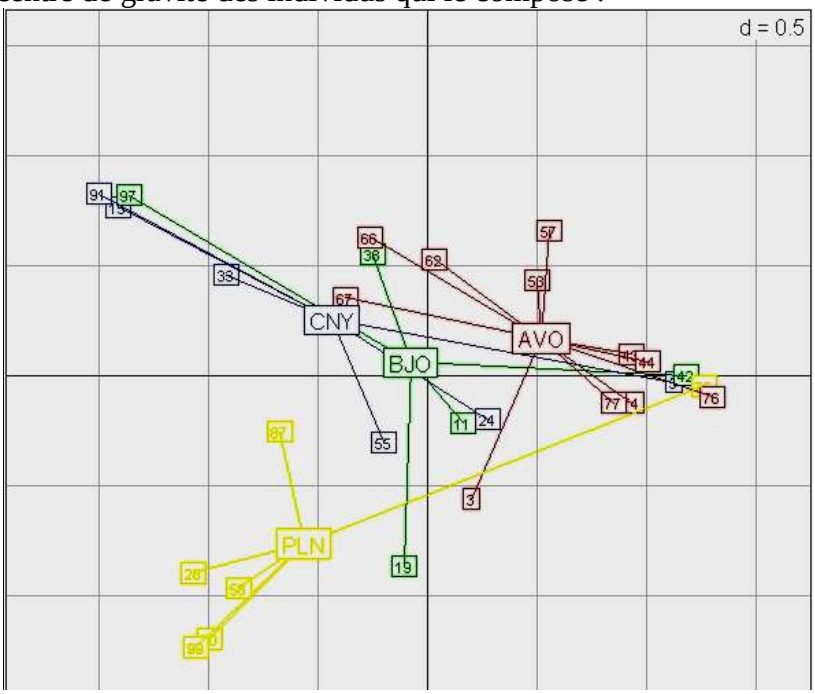


Figure II.2.2.2 Typologie ACP sur la base des champs.

La dispersion est plus faible lorsque l'on fait la représentation sur la base des centres de gravités des champs. A l'exception d'un seul, les champs assignés à PLN sont bien groupés et discriminés, les champs assignés à AVO sont groupés mais peu discriminés des champs assignés à BJO et à CNY qui sont très dispersés sur les deux premiers axes de l'ACP.

figure III.2.2.3 : Plans factoriels des ACP pour chaque marqueur pour les champs assignés à AVO, PLN, CNY et BJO.

Certains marqueurs discriminent bien les centres de gravités des champs assignés à une variété (Na14H11.A, Ol12F02.B, et Ra2E11). C'est le problème de l'ACP, elle mélange l'information de tous les marqueurs alors que certains discriminent moins bien que d'autre, et on ne peux que relativement faire un choix car les bon marqueur pour certaines variétés ne seront pas forcement de bons marqueurs lorsque l'on étudie d'autres variétés.

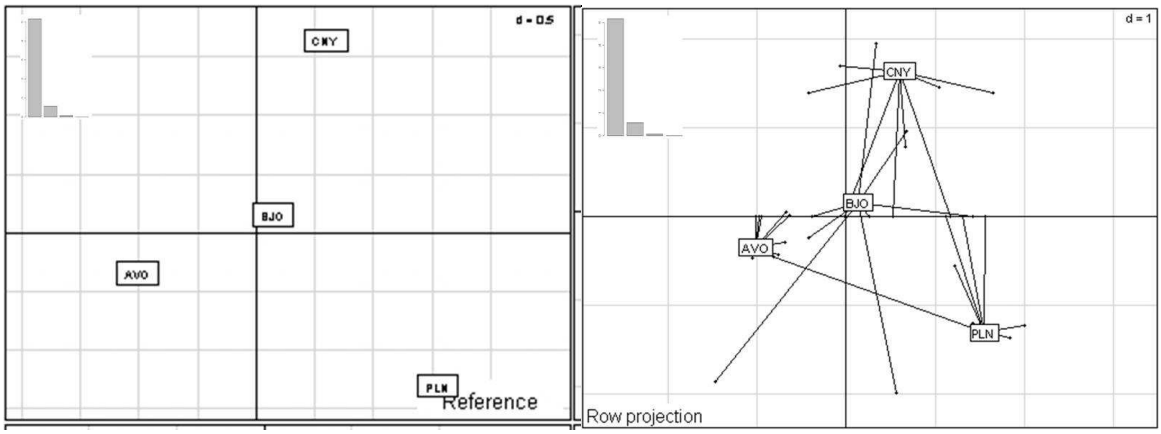


Figure II.2.2.4 plan factoriel de l'ACOM sur la base des champs.

L'ACOM permet une bonne discrimination des centres de gravité des champs tout en contenant presque la totalité de l'information sur les deux premiers axes (histogramme des valeurs propres en haut à gauche). Les champs AVO, CNY et PLN sont bien discriminés les uns des autres mais les champs de BJO sont très dispersés.

III.3 Etude des populations férales.

Exemple d'étude sur la route 5.

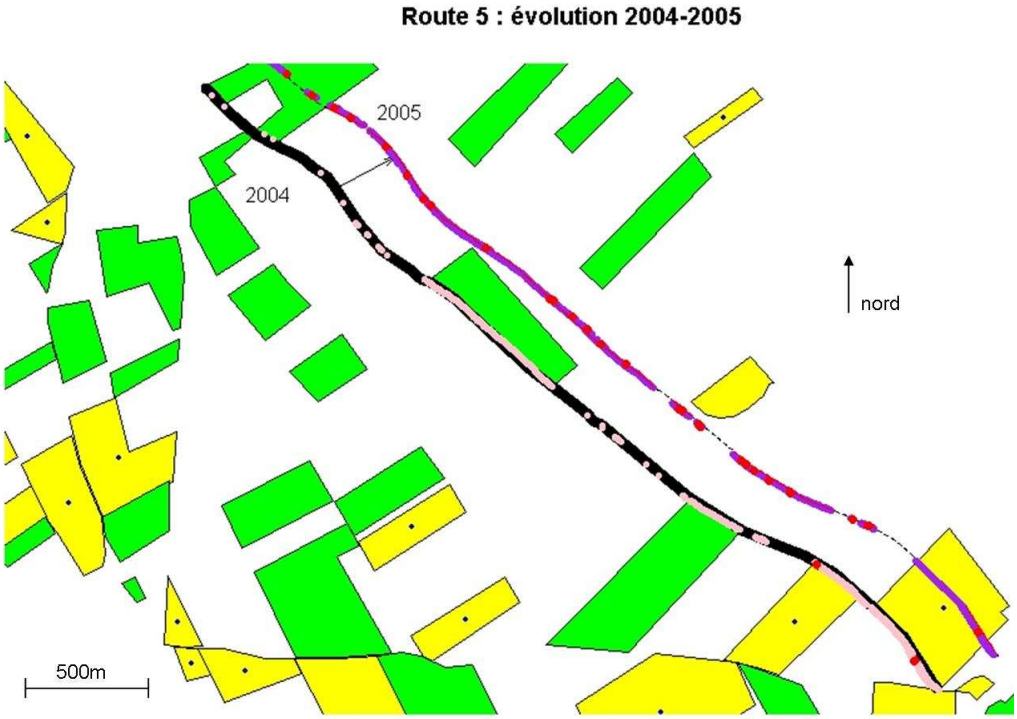


Figure III.3.1 : carte de la zone entourant la route 5. La flèche noire représente la translation effectuée pour que les informations de la route 5 en 2005 ne recouvrent pas les informations de 2004.
Polygones jaunes : champs de colza présents en 2004 ; Polygones verts : champs de colza présents en 2003 ; ligne noire : route 5 ; points roses pales: populations de bord de routes présents en 2004 ; violets : populations de bord de routes présents en 2005 ; points rouges : populations de bord de route génotypées en 2004 et 2005; points bleus (noirs) : champs de colza génotypés en 2004.

Les populations de bord de route sont présentes sur presque tout le long de la route 5 en 2005 en effectifs important, tandis qu'en 2004 elles sont principalement présentes le long de l'emplacement des champs de 2003.

Nous avons observé une seule population se recouvrant et étant génotypées en 2004 et 2005 sur l'ensemble de la zone d'étude :

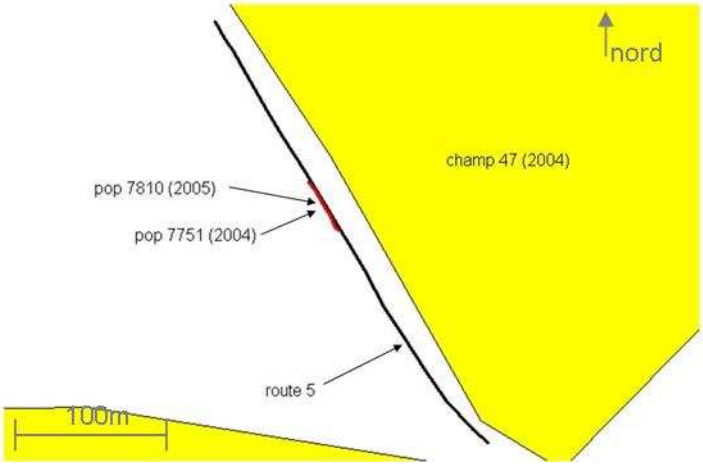


Figure III.3.2 : carte de la zone entourant les segments communs des populations de bord de route 7751 en 2004 et 7810 en 2005 toutes deux génotypées. Elles se situent sur la route 5 en bordure du champ 47 cultivé en 2004 lui aussi génotypé.

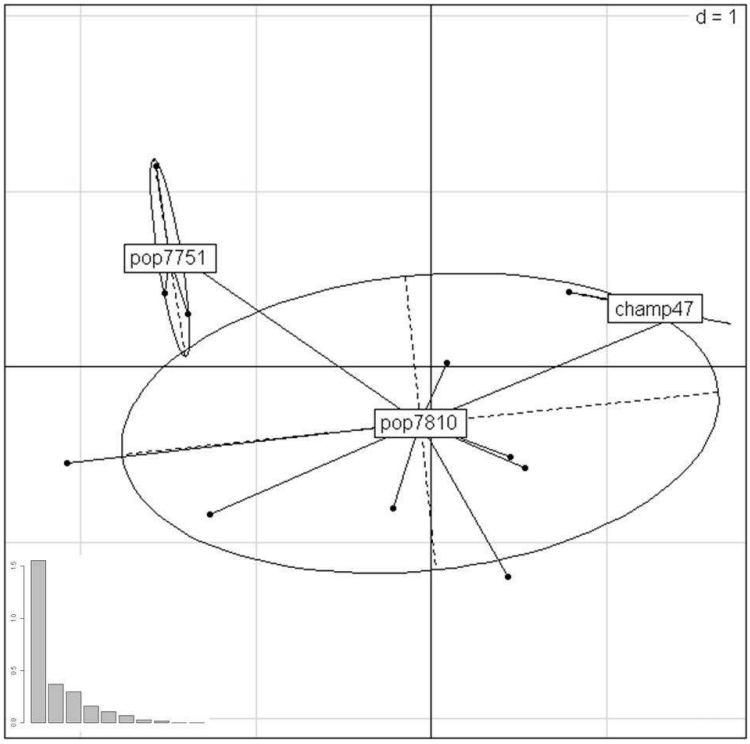


Figure III.3.3 : plan factoriel de l'ACP des fréquences alléliques des populations de bord de route 7751 et 7810, et du champ 47.

Sur le premier axe de l'ACP les éléments génotypés de la population 7810 en 2005 correspondent à un mélange entre les éléments génotypés du champ 47 et de la population 7751 qui le bordait en 2004.

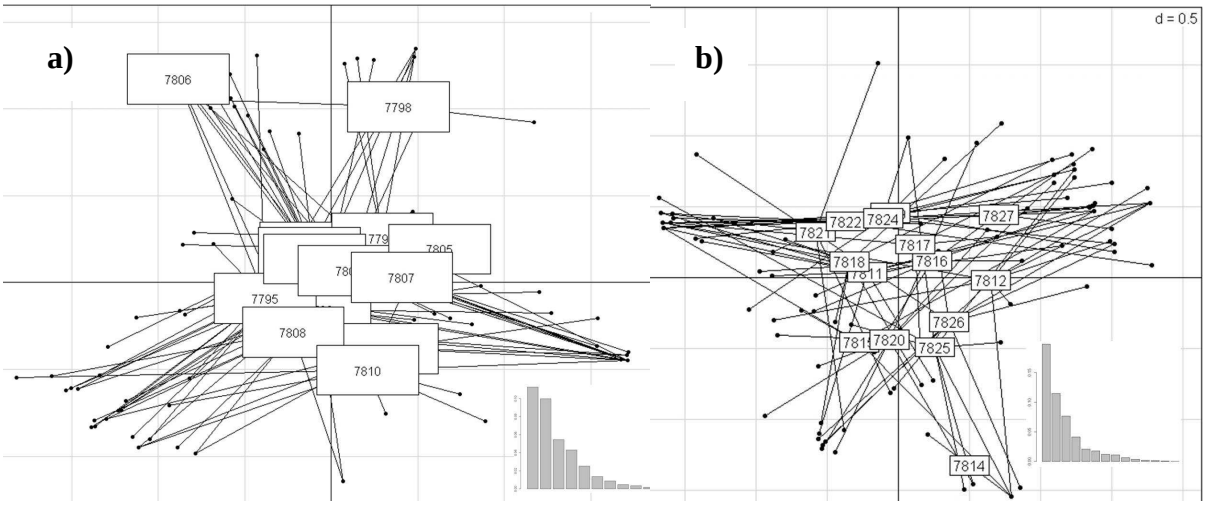


Figure III.3.4 : plan factoriel des ACP inter population des fréquences alléliques des populations de bord de route présentent en 2005 sur la route 5 (a. populations sur la bordure à droite en direction du sud, b. populations sur la bordure à droite en direction du nord).

On n'observe pas de structurations particulières des populations présentent en 2005 en bordure la route 5 sur les plans factoriels.

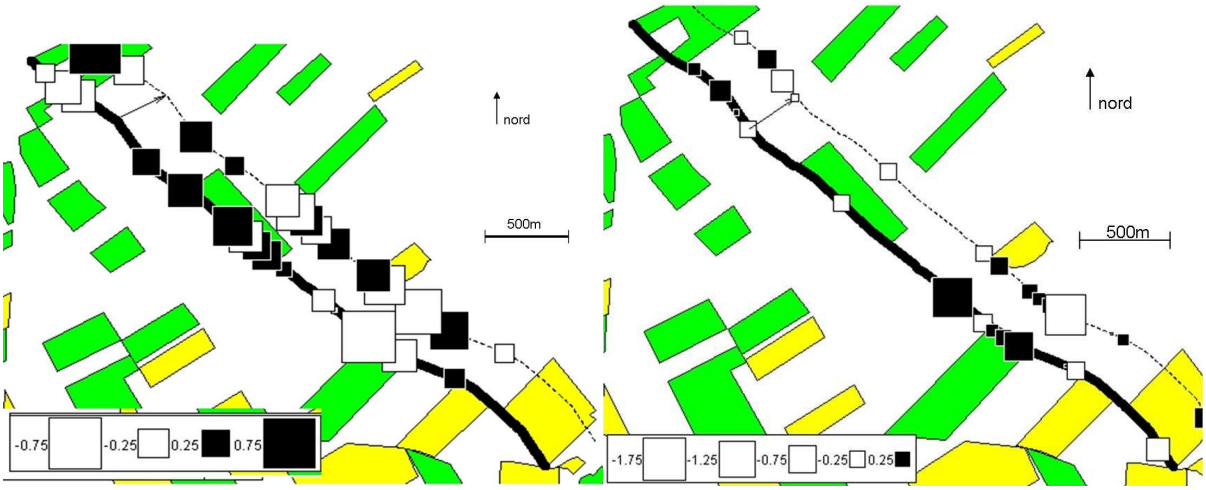


Figure III.3.5 : présentation dans l'espace des résultats des axes 1 et 2 de l'ACP inter population des fréquences alléliques des populations de bord de route présentent en 2005 sur la route 5 ;
carte du haut : populations sur la bordure à droite en direction du nord ;
carte du bas : populations sur la bordure à droite en direction du sud ;
Polygones jaunes : champs de colza présents en 2004 ; Polygones verts : champs de colza présents en 2003 ;
ligne noire : route 5 .

La présentation sur la carte des résultats de l'ACP des fréquences alléliques des populations en bordure de la route 5 montre que les valeurs proches sont regroupées.

IV. Discussion

Pertinence de l'analyse des données sous forme de fréquences alléliques.

Le but de la typologie par fréquences alléliques est de discriminer suffisamment la variabilité génétique du paysage pour pouvoir l'analyser et étudier son évolution afin de comprendre ses mécanismes dans une structuration de type agricole.

Nous avons utilisé des outils méthodologiques pour représenter l'information des données de fréquences alléliques de marqueurs moléculaires :

-L'ACP (Hotelling, 1933 ; Menozzi ,1978) opère une réduction dimensionnelle nécessaire à la visualisation de l'information apportée par l'ensemble des marqueurs, ou par chacun d'eux.

-L'ACOM (Chessel et al., 1996) permet d'extraire les structures de populations qui sont communes à l'ensemble des marqueurs. On obtient une typologie de référence à laquelle les marqueurs participent à des degrés différents. Cette participation peut constituer une base pour évaluer l'apport d'un marqueur en terme de typologie dans le but de construire une typologie plus discriminante. Dans ce cas un marqueur utile sera un marqueur séparant bien des groupes de populations.

L'étude des graines de références (partie III.1) nous a permis de savoir que l'on pouvait utiliser l'ACP (Figure III.1.1) et l'ACOM (Figure III.1.4) pour visualiser l'information. L'ACP par marqueur (Figure III.1.2) nous montre que les marqueurs peuvent être utilisés pour discriminer les variétés. En effet sur les figures III.1.1 et III.1.4 certaines variétés sont très proches mais la discrimination est correcte aux vues du nombre de données affichées et de la proximité génétique de certaines variétés. La projection des premiers axes (Figure III.1.3) montre que deux marqueurs, Ra2E11 et Ra2A05.B, discriminent différemment des autres, de ce fait il y a une perte d'information lors de l'ACOM qui prend en compte les marqueurs donnant une typologie semblable.

L'étude des champs de l'année 2004 (partie III.2) montre que l'assignation quand elle est possible élimine le bruit en simplifiant l'information, et permet de travailler avec des données claires. Mais on a observé les limites du traitement des données par assignation quand on assigne une variété pour chaque champ. En effet l'assignation n'est possible que pour 48 des 93 champs (si on a un niveau élevé de tolérance pour le bruit, sinon on peut en assigner moins). L'absence d'uniformité des données génétiques d'un champs peut être due au fait que l'agriculteur ait semé plusieurs variétés ; ou qu'il y ait des repousses issues des

champs d'années précédentes ; ou encore que les prélèvements aient été fait près du bord du champ et que des graines provenant du transport ou des populations de bordures aient germé dans le champ ; ou bien qu'il y ait plus de génotypes dans la variété du champ que de génotype dans les graines utilisées comme référence pour cette variété.

La difficulté à assigner les champs à une variété est révélatrice de la complexité de la structure génétique d'un paysage agricole qui est pourtant fortement structuré. Contrairement à ce que l'on pensait les champs ne sont pas homogènes. Cela va augmenter la difficulté prévue pour comprendre ce qui se passe en terme de flux génétique vu l'hétérogénéité de certains champs.

L'assignation n'est pas faisable pour les populations de bord de route étant donné la variabilité des génotypes présents et du nombre de générations potentielles entre les variétés et les plantes présentes dans ces populations. En effet, un certain nombre de graine ou feuille sont assignables à plusieurs variétés et certaines graines ou feuilles ne sont pas assignées du tout. Par ailleurs, comme on ne connaît pas forcément toutes les variétés de référence. Si on travaille avec les résultats des assignations, on perd une partie de l'information génétique.

Contrairement a l'assignation, l'analyse des données sous forme de fréquences alléliques maintient le bruit (figure III.2.2.1) ce qui permet de travailler avec un haut degré d'information car le maintien du bruit est due au maintient de l'information qui est perdue lors de l'assignation. En effet à part quelques exceptions les champs ne sont pas composés de variété pure (annexe 4). Le paysage génétique est plus compliqué qu'on ne le soupçonnait et l'analyse des données génétiques au niveau du paysage va être plus compliquée qu'on ne le pensait.

Mais on peut observer avec l'étude du suivie de populations d'une année sur l'autre (partie III.3.1), que lorsque que l'on travail à une échelle plus petite, la typologie par fréquences alléliques est un outil efficace pour obtenir une bonne observation permettant une interprétation des données (Figure III.3.3). En effet sur cette figure le lien, entre les deux populations consécutives au même endroit deux années de suite et le champ présent la première année, est clair.

Sur les plans factoriels des ACP des populations présentent en 2005 en bordure la route 5 (figure III.3.4) on n'observe pas de structurations particulières des populations. Par contre lorsque l'on représente les résultats des deux premiers axes des ACP sur la carte de la route 5 (figure III.3.5), on observe une structuration spatiale des populations. Il serait intéressant des comparer ces populations avec les champs qui bordent la route sur plusieurs

années pour comprendre avec précision la part de leurs influences sur les populations de bords de routes.

L'approche par assignation au niveau des ensembles de population n'est pas réalisable. L'approche par fréquences alléliques est encore améliorable, on n'a pas trouvé d'approche optimale car le problème de l'ACOM est que lorsque l'on a deux groupes de marqueurs (figure III.1.3) cette méthode privilégie l'un d'eux et perd l'information apportée par l'autre. Il faudrait pouvoir trouver une méthode qui permette de combiner les marqueurs qui fonctionnent ensembles.

Le suivit de l'évolution du paysage génétique est complexe, du fait de la grande superficie étudiée cela demanderait un investissement important pour permettre la précision nécessaire à une étude génétique telle que l'on veut la faire. L'échantillonnage tel qu'il est fait actuellement ne permet pas de travailler avec cette précision. En effet sur l'ensemble des données il y a seulement deux populations génotypées qui se chevauchent entre 2004 et 2005 et peu de routes ou de chemins pour lesquels on dispose de populations de bords de route génotypées pendant ces deux années. De ce fait peu de données sont disponibles pour faire un suivi de l'évolution génétique des populations. Car cela a un coût important de faire génotyper des échantillons pour huit marqueurs.

Ce qui serait intéressant, si on en avait les moyens, serait de faire cultiver sur une parcelle un colza contenant un allèle qui ne serait pas présent dans le paysage. Il serait facilement discriminable et le suivi de son évolution dans le paysage serait alors beaucoup plus facile et précis qu'une étude génétique globale.

La variété YUDAL est un bon exemple, si elle n'était pas mâle stérile cette variété serait un bon candidat, elle n'est pas cultivée et donc absente du paysage en dehors de quelques plantes utilisées pour des études par capture de pollen, pendant l'étude (pas présenté dans les résultats) on a remarqué qu'un de ses allèle est fortement discriminé par le marqueur Na12D08 (Figure IV.1).

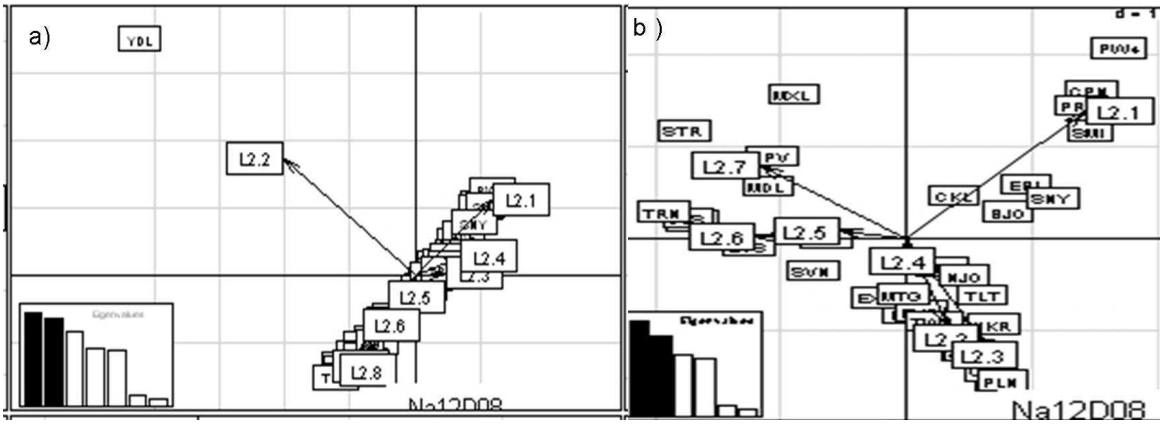


Figure IV.1 : ACP pour le marqueur Na12D08, avec a) ou sans b) les graines de références de la variété YUDAL (YDL).

Ce marqueur est fortement discriminant pour l'allèle L2.2 qui n'est actuellement pas présent dans le paysage génétique, mais qui est présent dans cette variété. Un allèle non présent dans le paysage et aussi discriminé que celui-ci permettrai, avec l'acquisition des données tel qu'elle est actuellement effectuée (partie II.3), de faire une étude précise et réaliste de la capacité de maintient et de dispersion de gènes provenant d'une source de colza génétiquement modifiée. Cette approche pourrait permettre un suivi plus exhaustif car il n'y aurait alors plus qu'à génotyper un seul marqueur.

Bibliographie

Bady P., Dolédec S., Dumont B., Fruget J.-F. (2004) Multiple co-inertia analysis: a tool for assessing synchrony in the temporal variability of aquatic communities, *Comptes Rendus Biologies* **327** 29-36.

Chessel, D.; Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* **44** 35-60

Chevre A. M., Barret P., Eber F., Dupuy P., Brun H., Tanguy X. et Renard M. (1997) Selection of stable Brassica napus B-juncea recombinant lines resistant to blackleg (Leptosphaeria maculans) .1. Identification of molecular markers, chromosomal and genomic origin of the introgression *Theoretical and Applied Genetics* **95** 1104-1111.

Dray S. and Dufour A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22** 1-20.

Frello S., Hansen K. R., Jensen J. et Jorgensen R. B. (1995) Inheritance of rapeseed (*Brassica napus*)-specific RAPD markers and a transgene in the cross *B-juncea* x (*B-juncea* x *B-napus*). *Theoretical and Applied Genetics* **91** 236-241.

Garmier A. (2006) Dynamique et dispersion d'une espèce cultivée échapée des champs : le cas du colza. Université Paris XI, Orsay.

Hansen S. B., Siegismund H. R. et Jørgensen R. B. (2001) Introgression between oilseed rape (Brassica napus L.) and its weedy relative B. rapa L. in a natural population. *Genetics Resources Crop Evolutio.* **48** 621-627.

Hedde M., Lavelle P., Joffre R., Jiménez J.J., Decaëns T. (2005) Specific functional signature in soil macro-invertebrate biostructures. *Functional Ecology* **19** 785-793.

Hotelling H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24** 498-520.

Ihaka R. et Gentleman R. (1996) R: a language for data analysis and graphics. *Journal of computational and graphical statistics* **5** 299-314

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

Jørgensen R. B. et Andersen B. (1994) Spontaneous hybridization between oilseed rape (*Brassica napus*) and weedy *B. campestris* (Brassicaceae): A risk of growing genetically modified oilseed rape. *American Journal of Botany* **81** 1620-1626.

Kelly CK , Bowler M , Breden F (2006) An analytical model assessing the potential threat to natural habitats from insect resistance transgenes: continuous transgene input. *biology letters* **2** 293-297

Menozzi P., Piazza A. (1978) Cavalli-Sforza L.L., Synthetic maps of human gene frequencies in Europeans, *Science* **201** 786-792.

Mikkelsen T. R., Jensen J. et Jørgensen R. B. (1996) Inheritance of oilseed rape (*Brassica napus*) RAPD markers in a backcross progeny with *Brassica campestris*. *Theoretical & Applied Genetics* **92**: 492-497.

Laloë D., Jombart T., Dofour A.B., Moazami-Goudarzi K. (2007) Consensus genetic structuring and typological value of markers using multiple co-inertia analysis. *Genet. Sel. Evol.* **39** 545–567

Lefol E., Danielou V., Darmency H., Boucher F., Maillet J. et Renard M. (1995) Gene dispersal from transgenic crops. I. Growth of interspecific hybrids between oilseed rape and the wild hoary mustard. *Journal of Applied Ecology* **32**: 803-808.

Pivard S. (2006) Processus écologiques impliqués dans la présence d'une plante cultivée dans les milieux naturels et semi-naturels d'un agroécosystème : Cas des populations férales de colza (Brassica napus L.). Université Paris XI, Orsay.

Shim S. I. and Jørgensen R. B. (2000) Genetic structure in cultivated and wild carrots (Daucus carota L.) revealed by AFLP analysis. *Theoretical and Applied Genetics* **101** 227-233

Annexe n°1 : Liste des variétés avec leurs nombres de génotypes.

abréviation	nom de la variété	nombre de génotypes différents	abréviation	nom de la variété	nombre de génotypes différents
ALR	aligator	2	GPD	gaspard	7
AVO	aviso	2	GRO	geronimo	8
BJO	banjo	7	HKR	hektor	12
BNV	bienvenu	2	KST	kosto	3
BRL	bristol	1	LTN	lutin	18
BSN	boston	2	MDL	madrigal	5
CDL	cadillac	1	MHN	mohican	2
CHE	cheyenne	3	MTG	montego	4
CKL	cocktail	30	MXL	maxol	3
CLS	columbus	5	NJO	navajo	4
CMA	campala	2	OKN	orkan	24
CNO	cando	16	OPH	olphi	4
CNT	contact	6	PLN	pollen	12
CNY	canary	1	PRO	pronto	30
CPL	capitol	2	PW4	pr45w4	3
CPN	captain	7	SMI	samourai	7
CPV	cap-vert	4	SNY	synergy	11
CRS	carolus	5	STR	star	20
ELE	elite	23	SVN	savanna	1
ERL	eurol	3	TLT	talent	22
EVS	elvis	21	TRN	tradition	5
EXA	extra	9	TWR	twister	1
EXP	expert	3	ZNH	zenith	28
EXS	express	3	ZRA	zeruca	3

Annexe n°2 : Assignation par vraisemblance.

La vraisemblance qu'un génotype G_i puisse être le descendant, par autofécondation ou allofécondation intravariétale, d'une variété *v* s'écrit de la manière suivante:

$$L(G_i, v) = \sum_{gv} f_{gv} \left[\overbrace{sT\left(G_i / M_{gv}, M_{gv}\right)}^{\text{Autofécondation mère } M_{gv}} + \overbrace{(1-s)\sum_{gv'} f_{gv'} T\left(G_i / M_{gv}, F_{gv'}\right)}^{\text{Allofécondation intra-variétale mère } M_{gv} \text{ - père } F_{gv'}} \right]$$

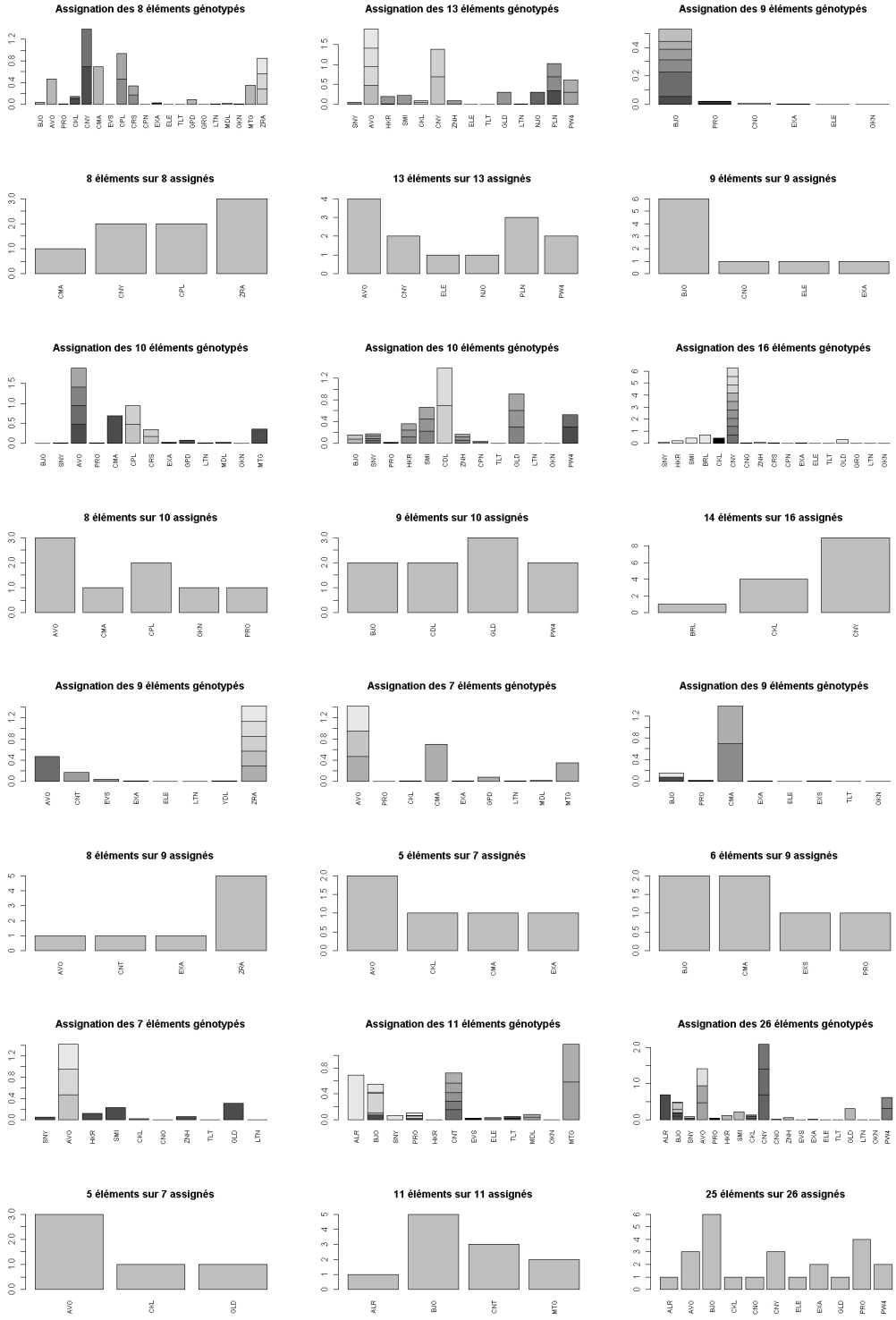
où f_{gv} et $f_{gv'}$ sont les fréquences de génotypes *gv* et *gv'* de la variété *v*, *s* est le taux d'autofécondation et $T(G_i/M_{gv}, F_{gv'})$ est la probabilité qu'une mère de génotype *gv* et qu'un père de génotype *gv'* aient engendré un descendant de génotype G_i.

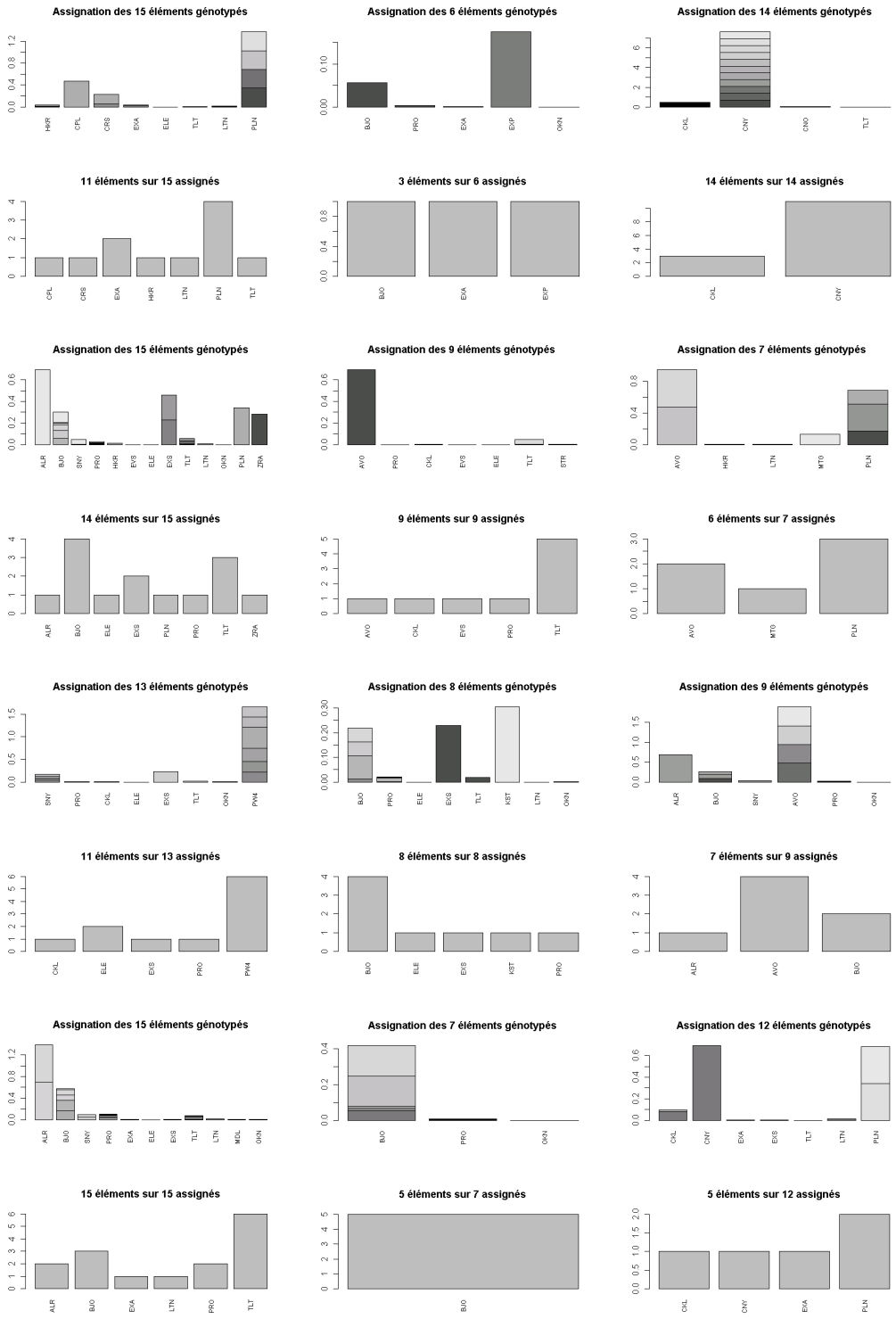
Annexe n°3 : Tableau indiquant le nombre de populations génotypées sur chaque route ou chemin en 2004 et 2005.

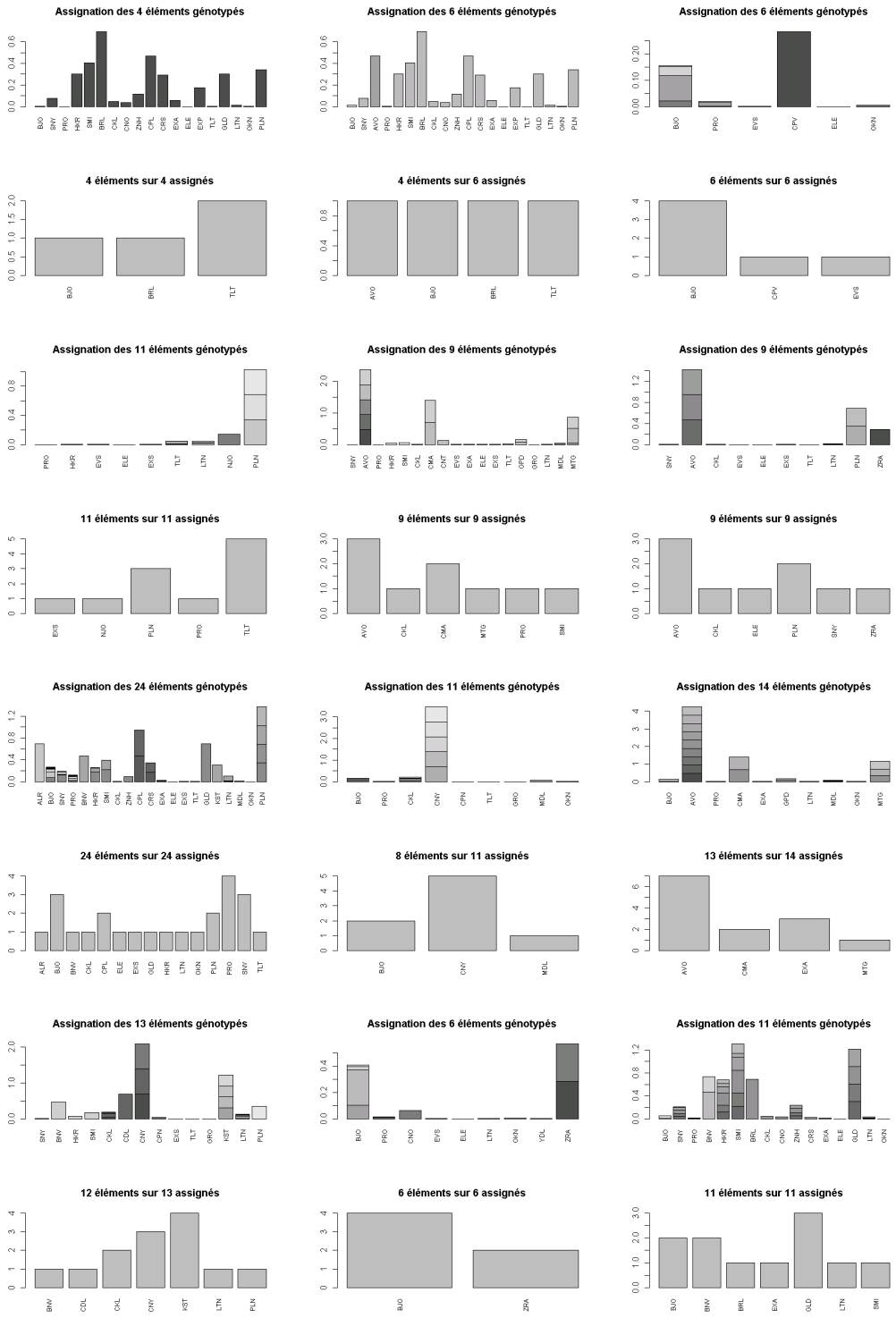
id_route	2004	2005
100	0	4
11	0	1
114	4	3
115	0	1
116	0	1
118	14	0
121	16	2
122	3	17
132	0	10
134	1	4
136	3	6
140	0	3
141	1	3
142	0	3
143	0	5
145	0	3
18	1	0
185	0	3
187	0	3
188	0	10
191	0	3
198	0	1
199	0	1
2	6	0
20	1	0
203	1	1
214	2	0
215	0	5
228	0	3
248	1	0
249	4	0
250	0	1
256	1	0
274	0	1

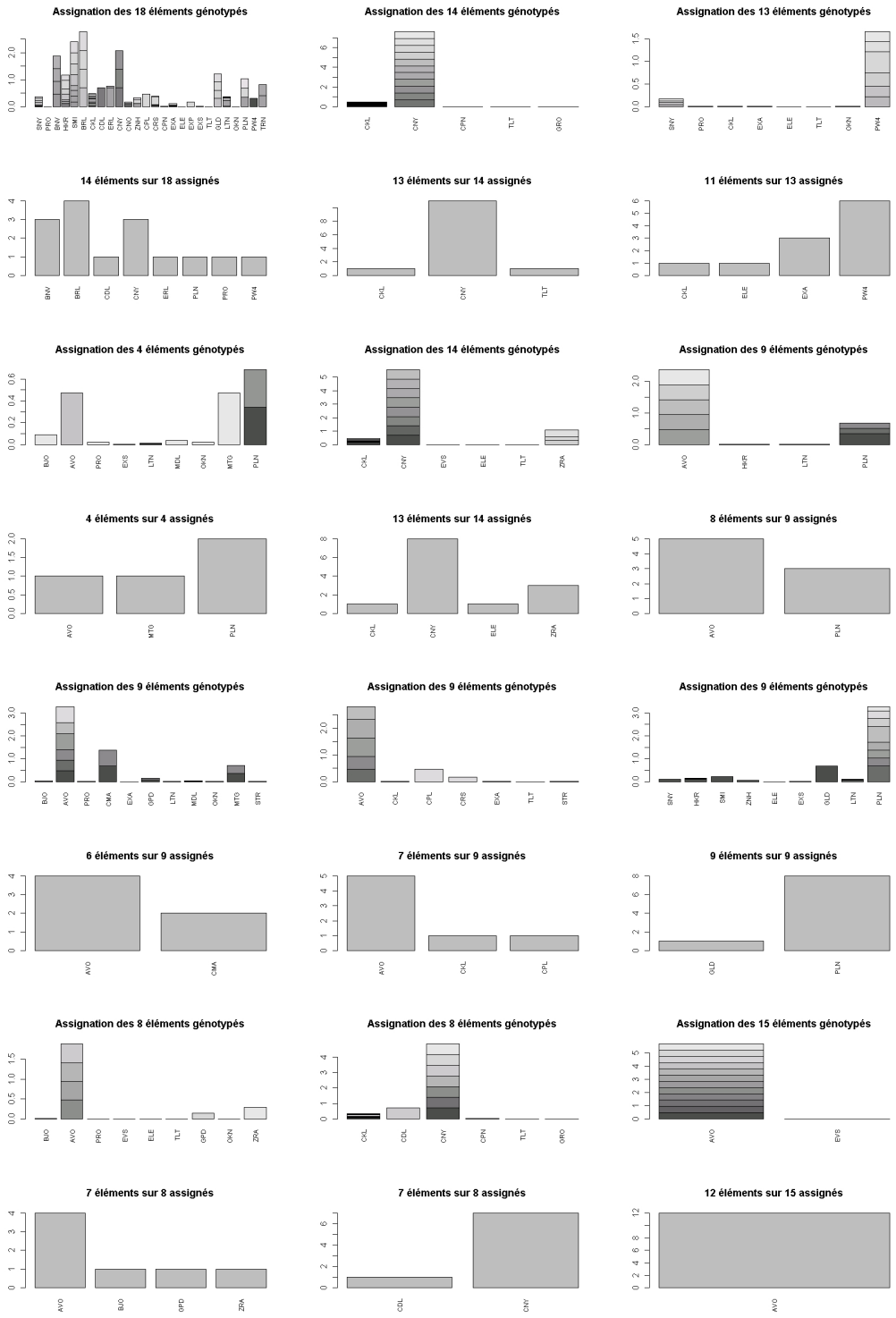
id_route	2004	2005
281	0	1
288	1	7
296	1	0
3	5	16
302	4	0
303	0	3
304	0	3
306	0	2
307	1	0
313	0	1
317	0	1
403	0	6
471	0	7
5	3	29
50	1	0
53	4	0
54	0	1
57	2	0
6	2	0
64	2	0
65	0	5
67	0	4
68	2	0
69	4	0
7	0	4
70	3	4
71	0	11
73	3	5
78	0	6
8	0	5
81	0	2
82	0	8
84	0	1
87	0	2

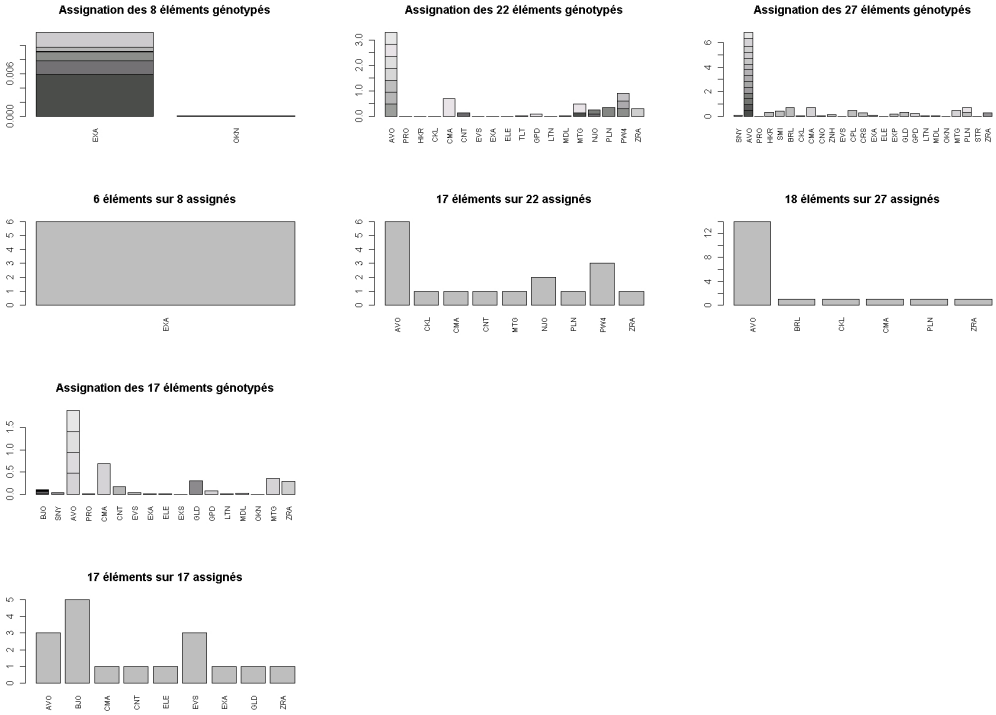
Annexe n° 4 : Assignation des champs, doublets d'histogrammes.



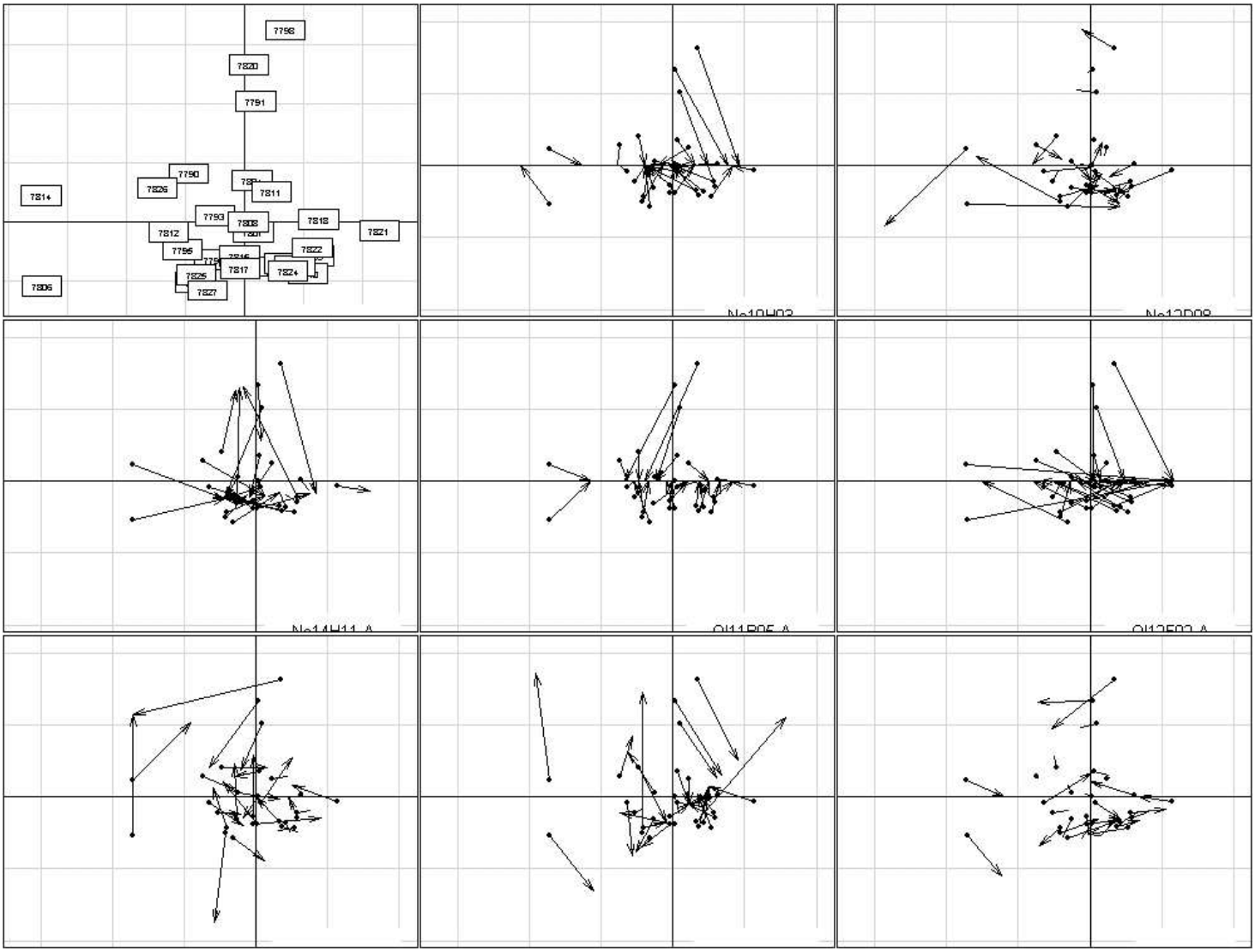

Assignation des 15 éléments génotypés
HKR CPL CRS EXA ELE TLT LTN PLN
Assignation des 6 éléments génotypés
BUO PRO EXA EXP ORN
Assignation des 14 éléments génotypés
CHL CNY CNO TLT
11 éléments sur 15 assignés
CPL CRS EXA HKR LTN PLN TLT
3 éléments sur 6 assignés
BUO EXA EXP
14 éléments sur 14 assignés
CHL CNY
Assignation des 15 éléments génotypés
ALR BUO SNY PRO HKR EVS ELE EXS TLT LTN ORN PLN ZRA
Assignation des 9 éléments génotypés
AVO PRO CHL EVS ELE TLT STR
Assignation des 7 éléments génotypés
AVO HKR LTN MTO PLN
14 éléments sur 15 assignés
ALR BUO ELE EXS PLN PRO TLT ZRA
9 éléments sur 9 assignés
AVO CHL EVS PRO TLT
6 éléments sur 7 assignés
AVO MTO PLN
Assignation des 13 éléments génotypés
SNY PRO CHL ELE EXS TLT ORN PW4
Assignation des 8 éléments génotypés
BUO PRO ELE EXS TLT KST LTN ORN
Assignation des 9 éléments génotypés
ALR BUO SNY AVO PRO ORN
11 éléments sur 13 assignés
CHL ELE EXS PRO PW4
8 éléments sur 8 assignés
BUO ELE EXS KST PRO
7 éléments sur 9 assignés
ALR AVO BUO
Assignation des 15 éléments génotypés
ALR BUO SNY PRO EXA ELE EXS TLT LTN MOL ORN
Assignation des 7 éléments génotypés
BUO PRO ORN
Assignation des 12 éléments génotypés
CHL CNY EXA EXS TLT LTN PLN
15 éléments sur 15 assignés
ALR BUO EXA LTN PRO TLT
5 éléments sur 7 assignés
BUO
5 éléments sur 12 assignés
CHL CNY EXA PLN


Assignation des 4 éléments génotypés
Assignation des 6 éléments génotypés
Assignation des 6 éléments génotypés
4 éléments sur 4 assignés
4 éléments sur 6 assignés
6 éléments sur 6 assignés
Assignation des 11 éléments génotypés
Assignation des 9 éléments génotypés
Assignation des 9 éléments génotypés
11 éléments sur 11 assignés
9 éléments sur 9 assignés
9 éléments sur 9 assignés
Assignation des 24 éléments génotypés
Assignation des 11 éléments génotypés
Assignation des 14 éléments génotypés
24 éléments sur 24 assignés
8 éléments sur 11 assignés
13 éléments sur 14 assignés
Assignation des 13 éléments génotypés
Assignation des 6 éléments génotypés
Assignation des 11 éléments génotypés
12 éléments sur 13 assignés
6 éléments sur 6 assignés
11 éléments sur 11 assignés


Assignation des 18 éléments génotypés
Assignation des 14 éléments génotypés
Assignation des 13 éléments génotypés
14 éléments sur 18 assignés
13 éléments sur 14 assignés
11 éléments sur 13 assignés
Assignation des 4 éléments génotypés
Assignation des 14 éléments génotypés
Assignation des 9 éléments génotypés
4 éléments sur 4 assignés
13 éléments sur 14 assignés
8 éléments sur 9 assignés
Assignation des 9 éléments génotypés
Assignation des 9 éléments génotypés
Assignation des 9 éléments génotypés
6 éléments sur 9 assignés
7 éléments sur 9 assignés
9 éléments sur 9 assignés
Assignation des 8 éléments génotypés
Assignation des 8 éléments génotypés
Assignation des 15 éléments génotypés
7 éléments sur 8 assignés
7 éléments sur 8 assignés
12 éléments sur 15 assignés


Assignation des 8 éléments génotypés
0.000
0.005
EVA
OKN
Assignation des 22 éléments génotypés
0.0
1.0
2.0
3.0
AVO
PRO
HIVR
CHL
CMA
CNT
EVS
EVA
ELE
TLT
GPD
LTN
MDL
MTG
NJO
PLN
PW4
ZPA
Assignation des 27 éléments génotypés
0
2
4
6
SNY
AVO
PRO
HIVR
SMI
BRL
CHL
CMA
CNO
ZNH
EVS
CPL
CRS
EVA
ELE
EVP
GLD
GPD
LTN
MDL
OKN
MTG
PLN
STR
ZPA
6 éléments sur 8 assignés
0
1
2
3
4
5
6
EVA
17 éléments sur 22 assignés
0
1
2
3
4
5
6
AVO
CHL
CMA
CNT
MTG
NJO
PLN
PW4
ZPA
18 éléments sur 27 assignés
0
4
8
12
AVO
BRL
CHL
CMA
PLN
ZPA
Assignation des 17 éléments génotypés
0.0
0.5
1.0
1.5
BJO
SNY
AVO
PRO
CMA
CNT
EVS
EVA
ELE
EVS
GLD
GPD
LTN
MDL
OKN
MTG
ZPA
17 éléments sur 17 assignés
0
1
2
3
4
5
AVO
BJO
CMA
CNT
ELE
EVS
EVA
GLD
ZPA

Annexe 5 : Etude des populations de 2005 sur la route 5 par ACOM, typologie des marqueurs en fonction de la typologie de référence.



Les points représentent la typologie de référence, et les flèches désignent l'emplacement où le marqueur positionne les populations

Résumé

L’objectif de ce travail est d’évaluer la faisabilité d’une analyse de l’évolution de la structure génétique de populations de colza à l’échelle du paysage, sur la base des fréquences alléliques de marqueurs microsatellites. Cette étude vise plus généralement à comprendre les conséquences potentielles de la mise en culture de plantes génétiquement modifiées sur les flux de gènes et la biodiversité à l’échelle du paysage.

Pour cela nous avons travaillé sur des données génétiques consistant en un typage par huit marqueurs microsatellites de graines et de feuilles prélevées entre 2002 et 2005 dans des champs de colza mais également dans les repousses de colza que l’on retrouve le long des routes et des chemins autour de la commune de Selommes.

Les méthodes utilisées sont purement descriptives, multivariées, et relèvent du champ de l’analyse de données. Les tableaux de fréquences alléliques sont étudiés par l’analyse en composantes principales (ACP), analyse que l’on compare à l'analyse de co-inertie multiple (ACOM) qui permet d'identifier dans quelle mesure les typologies obtenues sur chaque marqueur sont cohérentes entre elle.

Le travail effectué permet de conclure que l’on peut travailler avec les fréquences alléliques à une échelle fine en utilisant ces méthodes d’analyse. L’observation de plusieurs résultats cohérents avec les processus sous-jacents suggère que la perte d’information qu’entraîne le passage en fréquences n’est pas rédhibitoire pour étudier les flux de gènes par cette approche.

Summary

The aim of this work is to evaluate the evolution of the genetic structure of population of colza on a landscape scale, based on allelic frequencies of microsatellite markers. This study more generally aims at understanding the potential consequences on gene flows and biodiversity of cultivating genetically modified plants.

Accordingly, we worked with genetic data consisting of a typing by eight microsatellite markers of seeds and of plant samples taken between 2002 and 2005 in fields and roads' edge of colza collected around the city of Selommes.

The methods used are purely descriptive, multivariate and derived from field analysis of data. The tables of allelic frequencies are studied by using a classical Principal Component Analysis (PCA), and a multiple co-inertia analysis (MCOA) which allows identification of the coherence between typologies obtained by different markers.

Our results make it possible to conclude that we can work with the allelic frequencies on a fine scale by using such methods of analysis. The observation of several results coherent with the subjacent processes suggests that the loss of information that the changeover in frequencies involves is not fatal to studying gene flow under this approach.