

MEMOIRE DE FIN D'ETUDES

Présenté pour l'obtention du DIPLÔME D'INGENIEUR AGRONOME

**Spécialisation : Amélioration des Plantes et Ingénierie végétale Méditerranéennes Et
Tropicales (APIMET)**

Etude du déterminisme génétique de l'alternance du Pommier



Par **Yan Holtz**

Année de soutenance : 2012

Organisme d'accueil : INRA Montpellier

MEMOIRE DE FIN D'ETUDES

Présenté pour l'obtention du DIPLÔME D'INGENIEUR AGRONOME

**Spécialisation : Amélioration des Plantes et Ingénierie végétale
Méditerranéennes Et Tropicales (APIMET)**

Etude du déterminisme génétique de l'alternance du Pommier

Par

Yan Holtz

**Mémoire préparé sous la direction de :
Jean-Jacques KELNER**

**Organisme d'accueil :
INRA Montpellier**

Présenté le : 20/09/2012

**Devant le Jury :
-P. Berthomieu
-A. Pellegrino
-A. Dolomez**

**Maître de stage :
Evelyne Costes**

Remerciements

En tout premier lieu, je tiens à remercier Evelyne Costes, ma maître de stage, pour le temps qu'elle a consacré à mon travail, la qualité de ses conseils et la confiance qu'elle m'a accordée.

Je tiens ensuite à remercier toutes les personnes qui ont participées au bon déroulement de ce stage : Sébastien, Gilbert, Roland, Sylvain et les élèves de 2A pour leur aide lors du phénotypage ; Jean-Baptiste pour son aide indispensable en statistique ; François Laurens pour l'accueil qu'il m'a réservé au sein de l'INRA d'Angers ; et toute l'équipe de l'INRA Montpellier de manière générale avec qui il a été un réel plaisir de cohabiter pendant ces 6 mois.

Un grand merci à Chapi chapon, Mademoiselle Henry et à Manu : la profondeur inaltérable de vos yeux a transformé cette expérience en un réel plaisir.

Finalement, ce stage clôturant mon cursus au sein de Montpellier SupAgro, je tiens ici à remercier toutes les personnes qui m'ont accompagné lors de ces 4 incroyables années, notamment mes parents que je remercie infiniment.

Sommaire

| | |
|--|----|
| Introduction..... | 1 |
| Revue Bibliographique | 2 |
| 1. Importance économique du Pommier | 2 |
| 2. Taxonomie, histoire et cytogénétique | 2 |
| 3. Objectifs de sélection | 3 |
| 4. Floraison du Pommier | 3 |
| 5. Le phénomène d’alternance | 5 |
| 5.1. Définitions | 5 |
| 5.2. Effets de l’alternance et moyens mis en œuvre pour les éviter | 6 |
| 5.3. Causes de l’alternance..... | 6 |
| 5.4. Quantification..... | 7 |
| Objectif du stage et démarche..... | 9 |
| Matériels et méthodes | 10 |
| 1. Matériel végétal | 10 |
| 2. Protocole de phénotypage..... | 11 |
| 2.1. Phénotypage des séquences de floraison (échelle locale)..... | 11 |
| 2.2. Phénotypage de la quantité totale d’inflorescences de l’arbre | 14 |
| 3. Modélisation de l’alternance | 14 |
| 3.1. Modélisation de la quantité annuelle d’inflorescences par génotype..... | 14 |
| 3.2. Modélisation des probabilités de transition à l’échelle locale..... | 16 |
| 3.3. Modélisation de la synchronisation des branches | 19 |
| 3.4. Validation des Index : Clustering de génotype | 19 |
| 4. Création d’une carte génétique consensus et recherche QTL | 20 |
| 4.1. Carte consensus..... | 20 |
| 4.2. Recherche QTL..... | 21 |
| Résultats | 23 |
| 1. Sélection de génotypes remarquables..... | 23 |
| 2. Modélisation de la quantité d’inflorescence..... | 24 |
| 2.1. Illustration du modèle de tendance | 24 |

| | | |
|------|--|------------------------------------|
| 2.2. | Présentation des Index..... | 25 |
| 2.3. | Corrélation entre index. | 26 |
| 2.4. | Capacité des Index à décrire la régularité des génotypes..... | 27 |
| 3. | Modélisation des probabilités de transition (échelle locale)..... | 28 |
| 3.1. | Choix de l'ordre maximal de la chaîne de Markov | 28 |
| 3.2. | Adaptation des modèles aux familles | 29 |
| 3.3. | Résultat du modèle de probabilité de transition | 30 |
| 3.4. | Capacité des index à décrire les génotypes | 30 |
| 4. | Classification des génotypes | 32 |
| 5. | Corrélation entre alternance à échelle locale et globale. | 35 |
| 5.1. | Corrélation des indices | 35 |
| 5.2. | Classification des génotypes..... | 36 |
| 5.3. | Comparaison des 2 classifications. | 38 |
| 6. | Analyse QTL | 39 |
| | Discussion | 42 |
| 1. | Possibilité d'appliquer le protocole de phénotypage à large échelle ou sur d'autres espèces | 42 |
| 2. | Perfectionnement des indices..... | 43 |
| 2.1. | Sensibilité de l'interaction mémoire génotype à la longueur des séquences..... | 43 |
| 2.2. | Interaction génotype x année pour les génotypes en opposition de phase par rapport au reste de la population | 44 |
| 3. | Estimation de la corrélation entre échelle globale et échelle locale | 45 |
| 4. | Recherche QTL..... | 46 |
| | Conclusion et Perspective | 48 |
| | ABSTRACT | Erreur ! Signet non défini. |

Table des figures

| | |
|--|----|
| FIGURE 1 : MARQUEURS MORPHOLOGIQUES DE CROISSANCE : UNE BOURSE EST REPRESENTEE A GAUCHE (A) AVEC UNE POUSSE DE BOURSE (B), UNE CICATRICE D'ARRET DE CROISSANCE EST REPRESENTEE A DROITE (C) (PHOTO BAPTISTE GUITTON, YAN HOLTZ). | 4 |
| FIGURE 2 : REPRESENTATION D'UNE INFLORESCENCE DE POMME. | 5 |
| FIGURE 3 REPRESENTATION DU POURCENTAGE DE BOURSES PAR ANNEE POUR DES GENOTYPES ALTERNANTS (A) ET REGULIERS (B ET C). LE POURCENTAGE A L'ECHELLE DE L'ARBRE EST REPRESENTE EN NOIR, LES BRANCHES 1, 2 ET 3 EN BLEU, ORANGE ET ROUGE. | 7 |
| FIGURE 4 : POPULATIONS ETUDIEES ET LEUR LOCALISATION DANS 3 PARCELLES D'ETUDES. | 10 |
| FIGURE 5 : ANALYSE DE LA SEQUENCE D'UN RAMEAU COURT. | 12 |
| FIGURE 6 : REPRESENTATION DES SEQUENCES OBSERVEES POUR LES FAMILLES STARKIMSON X GRANNY (GAUCHE) ET X3263 X BELRENE (DROITE) | 12 |
| FIGURE 7 : REPRESENTATION DES SEQUENCES OBSERVEES POUR LES FAMILLES HIVW, N ET P. | 13 |
| FIGURE 8 : DEVELOPPEMENT DE LA MEMOIRE 00 A L'ORDRE 3 (A GAUCHE) OU A L'ORDRE 2 (A DROITE) | 18 |
| FIGURE 9 : REPRESENTATION DU GL 14 POUR LES CARTES GENETIQUES DES 4 POPULATIONS (A DROITE) ET DE LA CARTE CONSENSUS (A GAUCHE). LES MARQUEURS PONTS SONT REPRESENTES PAR DES TRAITS BLEUS. | 21 |
| FIGURE 10 : REPRESENTATION GRAPHIQUE DE LA PROPORTION DE BOURGEONS FLORAUX EN FONCTION DU TEMPS POUR UN GENOTYPE REGULIER SYNCHRONISE (A/), REGULIER DESYNCHRONISE (B/), ALTERNANT (C/), ET IRRÉGULIER (D/). LES 3 BRANCHES SONT REPRESENTES UNE A UNE RESPECTIVEMENT EN ROUGE, VERT ET VIOLET PUIS DE FAÇON REGROUPEE EN NOIR. | 24 |
| FIGURE 11 (A) : REPRESENTATION GRAPHIQUE DE LA PRODUCTION ANNUELLE D'INFLORESCENCES POUR LE GENOTYPE 1 DE LA FAMILLE STARKIMSON X GRANNY (ECHELLE LOCALE). LA DROITE DE TENDANCE MODELISEE PAR LE MODELE DE TENDANCE EST REPRESENTEE. | 25 |
| FIGURE 12 (B) : REPRESENTATION DES RESIDUS DU MODELE DE TENDANCE DES ANNEES N (ORDONNEE) EN FONCTION DES RESIDUS DES ANNEES N-1 (ABSCISSE). | 25 |
| FIGURE 14 : HISTOGRAMME DE LA FREQUENCE DES OCCURRENCES DE MEMOIRE 1111 AU SEIN DES SEQUENCES DE LA POPULATION X3263 X BELRENE. | 28 |
| FIGURE 15 : REPRESENTATION DE L'ENTROPIE EN FONCTION DE L'INTERACTION GENOTYPE*MEMOIRE 1111 POUR TOUS LES GENOTYPES DE LA FAMILLE X3263 X BELRENE. | 28 |
| FIGURE 16 : REPRESENTATION DE L'EFFET D'INTERACTION GENOTYPE * MEMOIRE 01 (ORDRE 2) (A) ET 0101 (ORDRE 4) (B) EN FONCTION DU BBI RESIDUEL NORMALISE POUR LA FAMILLE X3263 X BELRENE. | 29 |
| FIGURE 17 : VALEUR DES EFFETS FIXES DES MODELES DE PROBABILITE DE TRANSITION. | 30 |
| FIGURE 18 : REPRESENTATION GRAPHIQUE DE L'ENTROPIE DES GENOTYPES DE LA FAMILLE X3253 BELRENE EN FONCTION DE LEUR BLUP D'INTERACTION GENOTYPE X MEMOIRE 00 (A), GENOTYPE MEMOIRE 10 (B), GENOTYPE MEMOIRE 01 (C), GENOTYPE MEMOIRE 11 (D). | 31 |
| FIGURE 19 : REPRESENTATION GRAPHIQUE DES GENOTYPES DE LA FAMILLE X3253 X BELRENE EN FONCTION DE LEUR BLUP D'INTERACTION GENOTYPE * ANNEE 2010 ET DE LEUR ENTROPIE. GENOTYPES REGULIERS SYNCHRONISES EN BLEU, LES REGULIERS DESYNCHRONISES EN VIOLET, LES ALTERNANTS EN ROUGE ET LES IRRÉGULIERS EN ORANGE. | 32 |
| FIGURE 21 : REPRESENTATION DES VALEURS D'ENTROPIE, DE BBI RESIDUEL NORMALISE ET DE COEFFICIENT D'AUTO CORRELATION POUR LES 4 CLUSTERS : REGULIERS SYNCHRONISES (BLEU), REGULIERS DESYNCHRONISES (VIOLET), ALTERNANTS (ROUGE) ET IRRÉGULIERS (ORANGE). | 34 |
| FIGURE 22 : RELATION ENTRE LES BBI RESIDUELS NORMALISES (A) ET ENTRE LES COEFFICIENTS D'AUTO CORRELATION (B) CALCULES A ECHELLE GLOBALE ET LOCALE. | 35 |
| FIGURE 23 : QUANTITE DE BOURGEONS FLORAUX PAR ANNEE MESUREE A L'ECHELLE GLOBALE (A), PUIS SUR UN ECHANTILLON DE L'ARBRE (ECHELLE LOCALE (B)) | 36 |
| FIGURE 24 : BOX PLOT DES 4 CLUSTERS POUR LES 3 VARIABLES UTILISEES. L'ECHELLE LOCALE EST REPRESENTEE A GAUCHE, L'ECHELLE GLOBALE A DROITE. LE CLUSTER BLEU REPRESENTE LES GENOTYPES REGULIERS SYNCHRONISES, LE VIOLET LES REGULIERS DESYNCHRONISES, LE ROUGE LES ALTERNANTS, LE ORANGE LES IRRÉGULIERS. | 37 |

| | |
|--|----|
| FIGURE 25 : REPRESENTATION DES GROUPES DE LIAISONS 1 ET 10 DE LA CARTE CONSENSUS. LES QTLs DETECTES SONT REPRESENTES PAR DES BANDES DE COULEURS. LA TAILLE DE CHAQUE BANDE REPRESENT L'INTERVALLE DE CONFIANCE DU QTL, LA COULEUR LA VARIABLE EN QUESTION, LE TEXTE LA FAMILLE CORRESPONDANTE. (FIGURE OBTENUE SOUS METAQTL) | 41 |
| FIGURE 26 : PROFIL DE PRODUCTION DU GENOTYPE 39 DE LA POPULATION STARKIMSON X GRANNY. | 43 |
| FIGURE 27 : PROFIL DE PRODUCTION DU GENOTYPE 195 DE LA POPULATION X3263 X BELRENE..... | 44 |

Table des tableaux

| | |
|--|----|
| TABEAU 1 : EFFECTIF D'ARBRES ET DE GENOTYPES PHENOTYPES POUR CHAQUE FAMILLE D'ETUDE | 11 |
| TABEAU 2 : BILAN DE LA PROCEDURE DE PHENOTYPAGE : NOMBRE D'INDIVIDUS DE SEQUENCES, D'EVENEMENTS ET FLORAUX POUR CHAQUE POPULATION..... | 13 |
| TABEAU 3 : PRESENTATION DES INDEX RELATIFS A LA QUANTITE D'INFLORESCENCES POUR LA POPULATION STARKIMSON X GRANNY (A), X3263 X BELRENE (B) ET LES POPULATIONS HIVW N ET P CONFONDUES (C)..... | 26 |
| TABEAU 5 : REPARTITION DES INDIVIDUS REMARQUABLES AU SEIN DES 4 CLUSTERS..... | 34 |
| TABEAU 6 : VARIABLES UTILISEES AUX DEUX ECHELLES LORS DE LA CLASSIFICATION HIERARCHIQUE ASCENDANTE | 37 |
| TABEAU 7 : EFFECTIF DE CHAQUE CLUSTER POUR LES 2 ECHELLES | 38 |
| TABEAU 8 : EFFECTIF PASSANT D'UN CLUSTER A UN AUTRE LORS DU CHANGEMENT D'ECHELLE | 38 |
| TABEAU 9 : QTLs DETECTES PAR MQM MAPPING POUR LES VARIABLES RELATIVES A L'ALTERNANCE CONCERNANT LES POPULATIONS STARKIMSON X GRANNY (A), X3263 X BELRENE (B) ET HIVW (C). CHAQUE TABLEAU DONNE LA VARIABLE, LE GROUPE DE LIAISON SUR LEQUEL LE QTL A ETE DETECTE (LG), LA POSITION DU QTL SUR CE LG, LE COFACTEUR UTILISE POUR LA DETECTION, LE LOD SCORE ET LA PROPORTION DE VARIANCE GENETIQUE EXPLIQUE (R^2). LES EFFETS ADDITIFS FEMELLES (Af) ET MALES (AM) AINSI QUE L'EFFET DE DOMINANCE (D) SONT AUSSI INDiques..... | 39 |

Introduction

L'irrégularité de production chez les arbres fruitiers est un paramètre agronomique d'importance majeure en raison de l'influence qu'elle peut avoir sur la qualité des fruits, et de l'impact environnemental et économique que son contrôle engendre (éclaircissage notamment). L'intégration de paramètres de régularité dans les schémas de sélection doit donc devenir une priorité. Cependant, la sélection de génotypes réguliers s'avère difficile en raison de la complexité du trait. Ainsi, le développement d'outils permettant une évaluation fiable et précoce du phénomène d'alternance est nécessaire.

La présente étude est basée sur 5 populations de pommiers hybrides localisées dans les stations expérimentales des INRA de Montpellier et d'Angers. Le phénotypage a été réalisé à plusieurs échelles : globale (considérant l'arbre dans sa totalité), puis locale (considérant seulement un échantillon de l'arbre). Les arbres ont été observés après la phase juvénile, afin de mettre en place des méthodes permettant une analyse de l'alternance précoce, pour pouvoir raccourcir les cycles de sélection.

Le premier objectif a été de créer des indices permettant d'évaluer efficacement l'irrégularité de production. Ces indices doivent permettre de discriminer les différents profils de production existant au sein des populations.

Une fois ces indices développés, il a été possible de vérifier si l'analyse d'un échantillon de l'arbre seulement reflète fidèlement l'alternance de l'arbre entier. En effet, les protocoles de phénotypage concernant la régularité de production sont très lourds. Travailler sur un sous échantillon de l'arbre seulement permettrait de réduire cette phase de travail.

Finalement, le déterminisme génétique de l'alternance a été étudié en mettant en relation la variation de ces indices de régularité avec les données génotypiques des individus. Cette analyse a été réalisée de manière préalable population par population, avant de mettre les résultats obtenus en commun, de façon à rechercher des zones QTL communes aux différentes populations.

Revue Bibliographique

1. Importance économique du Pommier

La pomme est le troisième fruit le plus consommé dans le monde (après la banane et le citron). En 2009, la production mondiale a atteint 71.28 million de tonnes pour 4.9 millions d'hectares récoltés, ce qui représente 30 milliards de dollars (FAO, 2011). Même si la surface mondiale cultivée est en diminution, la production mondiale subit une augmentation ces 25 dernières années.

La Chine est le premier producteur mondial de pomme à l'heure actuelle, avec 32 millions de tonnes produites en 2009 (FAO, 2011). La France est classée au 7^{ème} rang avec 1.95 millions de tonnes produites en 2009 (FAO, 2011) sur 51 000 hectares.

Il existe des centaines de cultivars différents, mais seulement 20 d'entre eux ont une importance commerciale. Parmi les principaux on nommera « Golden delicious », « Delicious », « Gala », « Fuji » et « Granny Smith »

2. Taxonomie, histoire et cytogénétique

La pomme cultivée est un hybride interspécifique de type *Malus x Domestica* (Korban & Skirvin, 1984). C'est une espèce faisant partie des Rosacées. Les espèces du genre *Malus* sont présentes généralement en Europe, Asie et Amérique du nord.

A l'heure actuelle, c'est *Malus sieversii* qui est reconnu comme étant l'ancêtre de la pomme domestiquée (Juniper & Mabberley, 2006). Il existe également de nombreuses similarités entre le génome de la pomme cultivée et de *Malus sylvestris* (Coart et al., 2006).

La majorité des pommiers cultivés sont des diploïdes de type $2n=34$ (Brown, 1992). Le nombre de paires de chromosomes est donc très élevé (17) en comparaison avec les autres sous familles des Rosaceae qui en possèdent généralement entre 7 et 9. Deux principales hypothèses viennent expliquer ce phénomène : il s'est produit des événements d'autoploïdisation suivie de diploïdisation et d'aneuploïdisation (Evans & Campbell, 2002), ou des événement d'alloploïdisation (Chevreau et al., 1985, Phipps et al., 1991, Velasco et al., 2010).

3. Objectifs de sélection

L'objectif principal des projets de sélection concernant les pommiers porte sur une combinaison entre la qualité du fruit (Gardiner et al., 2007) et la résistance aux maladies (Janick et al., 1996 ; Laurens, 1999). La qualité du fruit concerne son apparence (couleur, taille, forme, ...) et ses qualités organoleptiques. La capacité de conservation du fruit est aussi un paramètre important. Concernant la résistance aux maladies, elle concerne principalement la tavelure et le mildiou.

Sélectionner des génotypes qui soient précoces, productifs et réguliers avec le minimum de travail concernant la conduite de la parcelle est le dernier axe de sélection.

4. Floraison du Pommier

Le pommier est une espèce allogame, du fait d'une auto incompatibilité pollinique de type gamétophytique (Lespinasse, 1992). Ainsi, les cultivars de pommier sont hautement hétérozygotes (Chevreau et al, 1985).

Tous les ans, chaque méristème subit un arrêt de croissance à une période qui dépend de son type : court, moyen ou long (Costes et al., 2006). La portion du rameau située entre 2 arrêts de croissance successifs est nommée unité de croissance. La croissance reprend au printemps, selon 2 cas de figure possibles :

-Si le méristème subit une induction florale, il se transformera en bourgeon floral qui donnera une inflorescence, elle-même responsable de la production de fruits. Chez le pommier, l'induction florale a lieu principalement au niveau des bourgeons terminaux. Le diamètre de l'axe portant l'inflorescence augmente lors du développement du fruit. Cette partie est nommée bourse (Pratt, 1988). C'est un marqueur morphologique de choix (figure 1.a) : la présence d'une inflorescence en cet endroit sera détectable durant toute la vie de l'arbre grâce à la présence de cette cicatrice. Une pousse de bourse (Figure 1.b) peut être produite à partir du méristème axillaire de la bourse (Luckwill, 1970).

-Si le méristème ne subit pas d'induction florale, le bourgeon reste végétatif et donne naissance à un nouveau rameau (cas du bourgeon axillaire) ou à une nouvelle unité de croissance dans le prolongement du rameau existant (cas du bourgeon terminal). L'interruption de croissance donne naissance à une marque caractéristique observable sur la branche (figure 1.c).

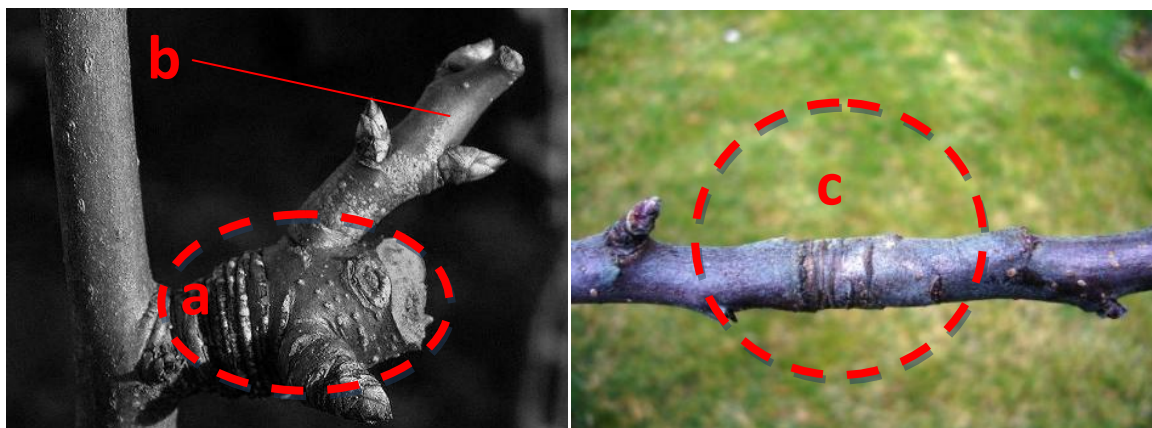


Figure 1 : marqueurs morphologiques de croissance : une bourse est représentée à gauche (a) avec une pousse de bourse (b), une cicatrice d'arrêt de croissance est représentée à droite (c) (photo Baptiste Guitton, Yan Holtz).

Ainsi, en observant un rameau, il est possible de déterminer quel type d'unité de croissance (végétative ou florale) a eu lieu chaque année. Si une bourse est présente, c'est qu'il y a eu inflorescence et sympodisation (développement d'une pousse de bourse). Dans le cas contraire, c'est qu'il y a eu bourgeon végétatif. Il est donc possible de retracer l'historique complet d'une branche en retraçant les séquences d'évènements le long des axes.

Il existe une période de la vie de l'arbre durant laquelle l'induction florale est impossible, même dans des conditions favorables. Cette période est appelé phase juvénile, elle dure de la germination jusqu'à la première fleur (Goldschmidt & Samach, 2004). Chez les plantes pérennes et en particulier chez le pommier, la phase juvénile est longue : entre 4 et 8 ans (Hackett, 1985). Ce paramètre est à prendre en compte dans l'étude de l'alternance : il faudra attendre plusieurs années pour que la phase mature de l'arbre soit atteinte et que la régularité de production puisse être évaluée.

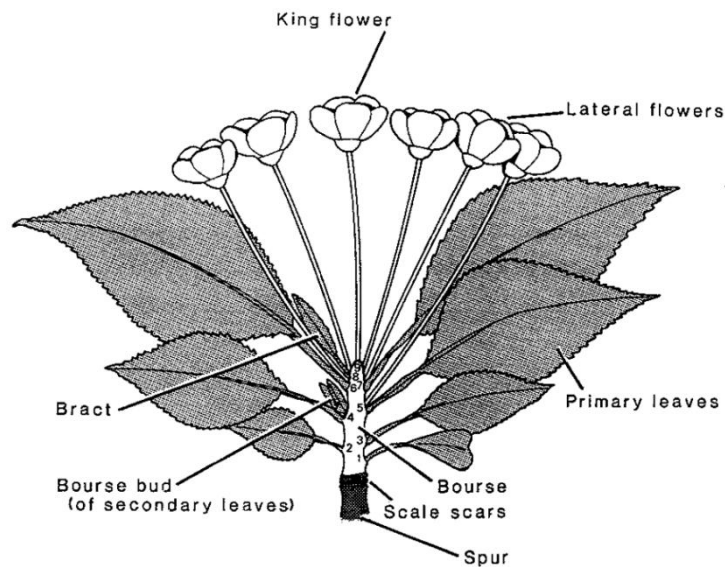


Figure 2 : Représentation d'une inflorescence de pomme.

Les inflorescences donneront ensuite naissance à un ou plusieurs fruits. La chute physiologique a lieu durant les premières phases de développement. Elle consiste en une chute naturelle d'une partie ou de la totalité des fruits présents. Cette chute permet à la plante de contrôler la quantité finale de fruits qu'elle développe, selon les réserves nutritionnelles dont elle dispose.

Il existe une forte corrélation entre la quantité d'inflorescences et la production en fruits. Ainsi, la quantité d'inflorescences est une variable permettant d'analyser la régularité de production d'un arbre.

5. Le phénomène d'alternance

5.1. Définitions

L'irrégularité de production est un comportement qui se réfère à des arbres ayant une quantité irrégulière de charge en fruit au cours d'années consécutives. Les années dites « ON » présentent une forte production avec beaucoup de fruits de petites tailles, les années « OFF » présentent une très faible production (Monselise & Goldschmidt, 1982). L'alternance est un cas particulier de l'irrégularité pour lequel on passe successivement d'une année « ON » à une année « OFF ».

Les irrégularités de production peuvent avoir lieu à différentes échelles : champ, groupe d'arbres, arbre et même au sein des branches d'un même arbre. L'alternance peut parfois être expliquée par des causes environnementales (stress, gel etc..). Cependant, le décalage de phase entre certains arbres voisins prouve que d'autres causes entrent en jeu (Monselise and Goldschmidt, 1982).

5.2. Effets de l'alternance et moyens mis en œuvre pour les éviter

Le phénomène d'alternance a un impact important sur la quantité et la qualité des fruits récoltés chaque année. Il faut ajouter à cela certains problèmes agronomiques tels que les casses de branches en raison de leur surcharge, l'augmentation de la sensibilité aux maladies et au froid. A l'heure actuelle ce phénomène est contrôlé par la mise en place de pratiques culturales permettant d'atténuer ses effets, notamment l'éclaircissage (mécanique ou chimique), la taille et l'utilisation de porte greffe. Cependant ces techniques demandent beaucoup de temps et de moyens économiques et ont un impact sur l'environnement (éclaircissage chimique). De plus, les nouvelles directives européennes tendent à interdire de plus en plus de produits chimiques utilisés actuellement pour l'éclaircissage.

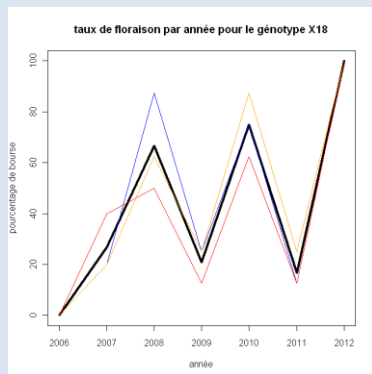
Une solution économiquement et environnementalement durable, serait l'utilisation de variétés présentant une floraison régulière au cours des années. Certaines variétés sont déjà reconnues pour posséder ces caractéristiques, c'est le cas de Granny Smith par exemple (Lauri et al., 1995).

5.3. Causes de l'alternance.

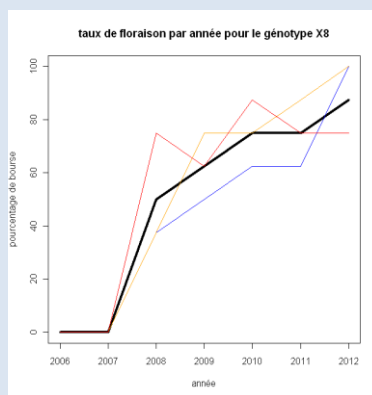
La principale hypothèse permettant d'expliquer le phénomène d'alternance considère que les fruits d'une année n ont un effet sur la floraison de l'année $n+1$. Chez le pommier, l'initiation florale de l'année $n+1$ a lieu durant le développement du fruit et des graines de l'année n . Beaucoup d'hypothèses ont été soulevées, selon lesquelles il y aurait une inhibition hormonale de l'initiation florale par le fruit en développement (Jonkers, 1979). La présence d'un fruit a un effet négatif sur le développement d'inflorescence sur le méristème le plus proche, notamment suite à la production de gibbérélines par ses graines (Nielsen et Dennis, 2000).

De plus, au sein des pommiers cultivés, certains cultivars sont connus comme étant réguliers, d'autres comme alternants. Ceci montre que la régularité possède aussi des origines génétiques (Monselise & Goldschmidt, 1982). Il existe par contre une forte interaction avec l'environnement. De ce fait certains cultivars ont pu être considérés comme alternants par certains auteurs, et réguliers par d'autres dans des environnements différents. C'est le cas des cultivars « Golden Delicious » et « Yellow Transparent » considérés comme régulier aux Etats Unis (Westwood, 1978) et irréguliers aux pays bas (Jonkers, 1979).

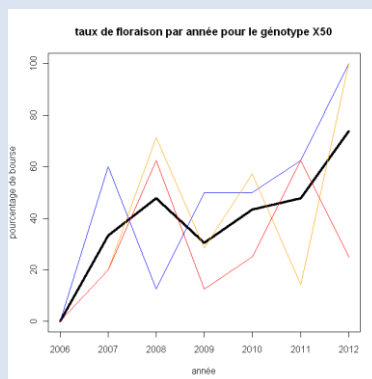
Exemple : Importance du paramètre de synchronisation dans l'évaluation de l'alternance d'un génotype.



a.



b.



c.

Figure 3 Représentation du pourcentage de bourses par année pour des génotypes alternants (a) et réguliers (b et c). Le pourcentage à l'échelle de l'arbre est représenté en noir, les branches 1, 2 et 3 en bleu, orange et rouge.

L'alternance à échelle globale peut être masquée par une désynchronisation à l'échelle locale : un génotype à production stable peut donc présenter un ensemble de branches à production stable (Fig 1.b), ou bien un ensemble de branches à production alternantes mais désynchronisées

5.4. Quantification.

La première difficulté rencontrée lors de l'étude de l'alternance est de trouver une manière de la quantifier. Plusieurs approches ont été proposées pour quantifier la régularité de production, mais aucune d'entre elles n'ont été utilisées dans des schémas de sélection jusqu'à présent.

Un index permettant d'estimer la déviation de production sur une succession d'années à été développé (Hoblyn et al., 1936). Cet index a par la suite été renommé BBI : « Biennial Bearing Index » (Wilcox, 1944).

Le BBI est l'indice communément utilisé pour quantifier l'alternance d'un arbre. Il est exprimé comme suit :

$$BBI = \frac{1}{T_{g,r} - 1} \sum_{t=2}^{T_{g,r}} \frac{|Y_{g,r,t} - Y_{g,r,t-1}|}{|Y_{g,r,t-1}| + |Y_{g,r,t}|},$$

$Y_{g,r,t}$: production (Yield) pour la réplique r du génotype g à l'année t .

$T_{g,r}$: nombre d'années de mesure pour ce même individu.

Cette équation quantifie la fluctuation de la production entre chaque année. Plus il y aura de différence entre les productions successives, plus le BBI sera élevé. Dans le cas où la production serait exactement la même à chacune des années, le BBI sera égal à zéro.

Cependant, elle ne prend pas en compte l'augmentation normale de cette variable au cours du temps lorsqu'il s'agit d'arbre sortant de leur phase juvénile. Ainsi, un arbre parfaitement régulier mais sortant progressivement de sa phase juvénile aura un BBI différent de 0.

De plus, d'autres mécanismes tels que la désynchronisation des branches viennent complexifier le phénomène d'irrégularité. En effet, une irrégularité peut être complètement masquée par une

désynchronisation de branches alternantes (Lauri et *al.*, 1995). Ce cas est illustré dans la figure 3.

Finalement, l'alternance se caractérise par une irrégularité de production mais d'autres paramètres tels que la synchronisation des branches doivent être pris en compte.

La thèse de Baptiste Guitton intitulée « Genetic control of biennial bearing in apple » s'est intéressée à ces questions, en se basant principalement sur une population ayant pour parents « Starkimson » et « Granny » (125 génotypes). Certains index ont été développés permettant effectivement d'apprécier le caractère alternant des individus en phase d'augmentation de production. Cependant, cette étude se base sur une seule population et développe chaque index à une échelle spécifique.

Il existe donc encore un champ d'étude important concernant l'analyse de l'alternance du pommier, des méthodes pour l'estimer et de son déterminisme génétique.

Objectif du stage et démarche

Ce stage s'inscrit dans la poursuite des études engagées lors de la thèse de B. Guitton (Guitton, 2011). Il a pour objectif d'améliorer les outils permettant la quantification de l'alternance afin d'accélérer les processus d'amélioration variétale pour ce caractère. Il pourrait s'agir de prédicteurs morphologiques juvéniles ou de marqueurs moléculaires permettant de détecter des allèles favorables. Ainsi, l'étude que j'ai conduite a eu pour objectif d'analyser ses causes, son fonctionnement et son déterminisme génétique sur plusieurs populations de pommiers en ségrégation.

L'étude suivante se base sur 5 populations, soit 716 génotypes. La taille du jeu de données obtenu permet de répondre à 3 problématiques emboîtées :

- (i) Evaluer les profils de production des génotypes étudiés. Ceci n'est possible que si des indices permettant de décrire efficacement les génotypes peuvent être développés. Ces indices, couplés, doivent permettre de discriminer chaque individu par rapport à ses caractéristiques de régularité.
- (ii) Déterminer si un échantillon de l'arbre peut être suffisant pour appréhender l'alternance d'un génotype. Ceci pourrait permettre d'alléger significativement les processus de phénotypage. Il est donc nécessaire d'évaluer l'influence qu'a l'échantillonnage sur l'évaluation de la régularité de production.
- (iii) Analyser le déterminisme génétique de la régularité de floraison. La détection de régions chromosomiques d'intérêt (QTL) comporte plusieurs intérêts. Elle permet d'une part de déterminer s'il existe des zones d'importance communes entre populations et d'autre part d'ouvrir des pistes de recherche sur la compréhension des mécanismes moléculaires sous jacents. A plus long terme, la détection de QTL permettrait de raccourcir les cycles de sélection par la sélection assistée par marqueurs.

Matériels et méthodes

1. Matériel végétal

L'étude a été réalisée sur 5 populations situées sur 3 parcelles différentes (2 à Montpellier, 3 à Angers). Ces populations sont toutes reliées par certains parents communs. Les relations d'apparentement sont fournies en annexe (1).

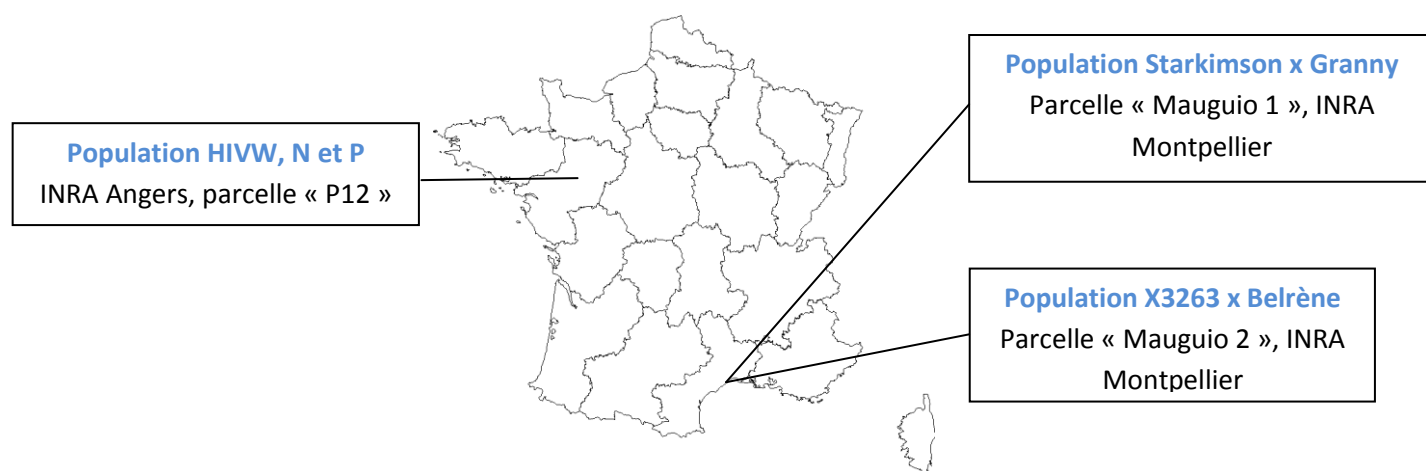


Figure 4 : Populations étudiées et leur localisation dans 3 parcelles d'études

Les parcelles de Montpellier présentent des arbres à développement libre : pas de taille ni d'éclaircissage. Les arbres sont répartis de manière aléatoire dans les différents rangs de la parcelle. La distance entre chaque rang est de 6m, la distance entre chaque arbre de 2m. La population Starkimson x Granny a été plantée en 2004, elle est composée de 125 génotypes répétés 2 fois, soit 250 arbres au total. La population X3263 x Belrène a été plantée en 2005, elle est composée de 277 génotypes dont une cinquantaine a été répété 2 fois, soit 333 arbres au total.

Sur la parcelle d'Angers le mode de conduite culturale est le même pour les trois populations présentes : il correspond à des conditions agronomiques standards dans des vergers de production. Ainsi, les arbres sont taillés chaque année. De plus, les multiples passages de tracteur entre les rangées cassent toutes les branches dirigées vers le milieu des allées. Les arbres sont âgés : ils ont été plantés en 1992.

| Parcelle | Nb. arbres. phénotypés | Nb. Génot. phénotypés | Parcelle | Lieu |
|---------------------|---------------------------|--------------------------|-----------|-------------|
| Starkimson x Granny | 228 | 119 | Mauguio 1 | Montpellier |
| X3263 x Belrène | 333 | 277 | Mauguio 2 | Montpellier |
| HIVW | 183 | 170 | P12 | Angers |
| N | 103 | 99 | P12 | Angers |
| P | 52 | 51 | P12 | Angers |
| Total | 899 | 716 | | |

Tableau 1 : Effectif d'arbres et de génotypes phénotypés pour chaque famille d'étude

2. Protocole de phénotypage.

2.1. Phénotypage des séquences de floraison (échelle locale)

L'objectif est de relever les événements végétatifs et floraux qui ont eu lieu le long d'un rameau, vu comme une séquence d'étude. La séquence chronologique est relevée en partant de 2012 et en remontant année par année. Chaque année, la présence de bourse ou de trace de bourgeon végétatif est notée. La présence de bourse est représentée par le chiffre 1, l'absence par le chiffre 0 (figure 5).

Les relevés s'effectuent sur un échantillon de l'arbre. Ainsi, seulement certaines branches ont été phénotypées. Une branche est définie comme la partie d'un arbre qui dérive du tronc. Ces branches sont choisies de manière à obtenir des séquences les plus longues et les plus lisibles possibles sur l'arbre. Au sein de chaque branche, plusieurs rameaux sont choisis eux aussi aléatoirement, il peut s'agir de rameaux courts comme de rameaux longs. Un rameau est défini comme la partie d'un arbre qui part d'une branche.

Entre 2 événements on note aussi la taille de la pousse les séparant : « C » pour court (moins de 5 cm), « M » pour moyen (entre 5 et 15 cm) et « L » pour long (>15 cm) .

Un exemple est donné pour la séquence ci-dessous :



Figure 5 : Analyse de la séquence d'un rameau court

Les modes de conduite et caractéristiques de chaque parcelle étant différents, les protocoles de phénotypage ont été légèrement adaptés. Le type ainsi que le nombre de séquences étudiées n'est pas toujours le même.

- *Parcelles de l'INRA Montpellier (populations Starkimson x Granny et X3263 x Belrène)*

Pour chaque arbre, 2 ou 3 branches ont été étudiées (pour Starkimson x Granny et X3263 x Belrène respectivement). Dans le cas de Starkimson x Granny les séquences ont été décrites sur 2 rameaux par unité de croissance de la branche principale. Pour X3263 x Belrène les relevés ont été effectués pour 4 rameaux partant de l'unité de croissance 2006 et 3 rameaux de l'unité 2007, ainsi que pour la branche principale elle-même.

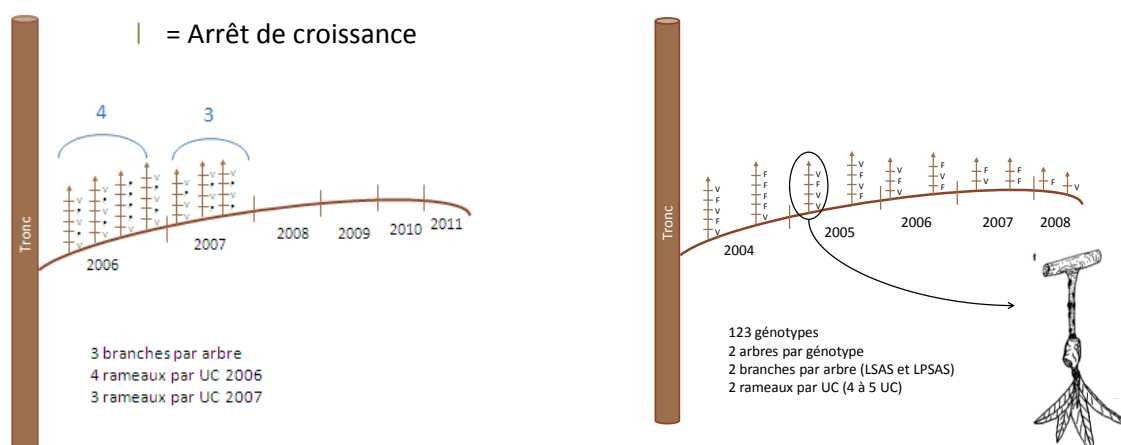


Figure 6 : Représentation des séquences observées pour les familles Starkimson x Granny (gauche) et X3263 x Belrène (droite)

- Parcelle « P12 », INRA Angers.

Cette parcelle comporte des arbres plus âgés donc plus grands. La majorité des branches étant cassées ou taillées, il est la plupart du temps impossible de remonter des séquences jusqu'au tronc. De plus, une mousse verte étant présente sur les branches, la lecture est difficile. Ainsi, le protocole a été légèrement modifié pour s'adapter à ces contraintes. Le nombre de branches a été fixé à 2, à raison de 5 rameaux par branches, peu importe l'unité de croissance sur laquelle elles sont situées. Seules les séquences d'une longueur de 5 années successives au moins sont sélectionnées. Dans les cas où le nombre de séquences disponibles sur les deux branches est inférieur à 8, une troisième branche est étudiée.

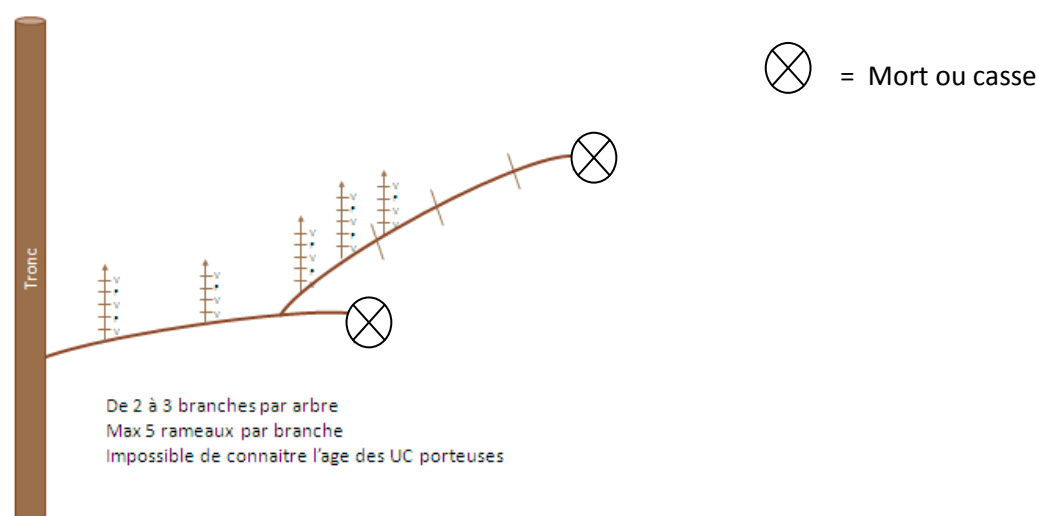


Figure 7 : représentation des séquences observées pour les familles HIVW, N et P.

Les informations concernant les séquences relevées et leurs caractéristiques pour chaque population sont récapitulées dans le tableau 2.

| Famille | Nb. arbres. phénotypés | Nb. Séquences tot. | Nb. Séquences moyen / Ind | Nb. Evènements Veg | Nb bourse | Total | % moy bourse | Nb année moy / séquence |
|-------------|---------------------------|--------------------------|---------------------------------|--------------------------|--------------|-------|-----------------|----------------------------|
| S x Granny | 228 | 4036 | 18 | 4975 | 3184 | 8159 | 61 | 3 |
| X x Belrène | 333 | 6134 | 18 | 23099 | 19493 | 42592 | 54 | 7 |
| HIVW | 183 | 1487 | 8 | 2983 | 6153 | 9136 | 33 | 6 |
| N | 103 | 895 | 9 | 1556 | 4150 | 5706 | 27 | 6 |
| P | 52 | 423 | 8 | 858 | 1770 | 2628 | 33 | 6 |
| Total | 899 | 12975 | - | 33471 | 34750 | 68221 | - | - |

Tableau 2 : Bilan de la procédure de phénotypage : nombre d'individus de séquences, d'évènements et floraux pour chaque population.

2.2. Phénotypage de la quantité totale d'inflorescences de l'arbre

Ces données issues de la thèse de Baptiste Guitton (Guitton, 2011) ne sont disponibles que pour la famille Starkinson x Granny. Le nombre d'inflorescences a été compté pour l'arbre dans sa totalité pendant 5 années successives (de 2005 à 2009).

3. Modélisation de l'alternance

3.1. Modélisation de la quantité annuelle d'inflorescences par génotype

La quantité d'inflorescences peut être relevée chaque année à l'échelle de l'arbre (échelle globale), ou à l'échelle d'un échantillon de l'arbre (échelle locale). Dans les 2 cas la modélisation suit la même méthode dont les bases ont été développées dans la thèse de Baptiste Guitton (2011). Afin de dissocier l'évolution de production au cours du temps de l'alternance, un modèle de tendance basé sur un modèle linéaire mixte est construit. Il modélise la quantité d'inflorescences au cours du temps. Il est proposé d'évaluer l'irrégularité comme l'amplitude des résidus autour de ce modèle.

Dans un deuxième temps, une étude de la structure des résidus est menée. Elle permettra de distinguer des profils irréguliers de profils alternants.

- *Modèle de tendance*

La variation du nombre d'inflorescences au cours du temps est représentée à l'aide du modèle linéaire mixte suivant :

$$Y_{g,r,t} = \beta + \beta_g + \zeta_{g,r} + (\alpha + \alpha_g + \xi_{g,r})t + \varepsilon_{g,r,t}$$

$Y_{g,r,t}$ représente la production de la réplication r du génotype g à l'année t . $\beta + \alpha t$ représente la tendance générale, traitée en temps qu'effet fixe. $\zeta_{g,r}$ et $\xi_{g,r}$ représentent la tendance spécifique à la répétition r du génotype g , considérés comme effet aléatoire. Les effets du génotype g , considérés comme fixes, sont définis par β_g et α_g .

- *Le BBI résiduel normalisé*

Le BBI résiduel normalisé est un index qui découle du BBI dont le calcul est exposé page 15.

Cet indice étudie la fluctuation de la quantité de bourse entre chaque année mais ne prend pas en compte l'augmentation de cette variable au cours du temps lors de la sortie de la phase juvénile. Afin

de prendre en compte cette augmentation, on va s'intéresser non plus à la production de chaque année, mais à l'écart entre la production observée et celle estimée par le modèle de tendance présenté plus haut.

L'utilisation des résidus, (normalisés par la production moyenne du génotype afin de pouvoir comparer les génotypes entre eux) donne finalement l'expression BBI Résiduel Normalisé :

$$\text{BBI_res_norm} = \frac{\sum_r \sum_{t=2}^{T_{g,r}} |\hat{\varepsilon}_{g,r,t} - \hat{\varepsilon}_{g,r,t-1}| / \sum_r (T_{g,r} - 1)}{\sum_r \sum_{t=1}^{T_{g,r}} Y_{g,r,t} / \sum_r T_{g,r}}.$$

Cet index permet de mesurer l'amplitude des déviations autour d'une droite de tendance. Cependant, il ne permet pas de faire la distinction entre des profils alternants et irréguliers. La distinction des deux sera permise par l'analyse de la dépendance des résidus successifs.

- *Le coefficient d'auto-corrélation*

L'alternance est un phénomène associé à une dépendance entre les événements successifs. Une production forte une année ($\hat{\varepsilon}_{g,r,t} > 0$) sera suivie d'une production faible l'année suivante ($\hat{\varepsilon}_{g,r,t} < 0$), le coefficient de corrélation entre deux résidus successif sera donc négatif.

Cette dépendance entre résidus successifs peut être caractérisée en utilisant un modèle autorégressif de premier ordre (AR1) décrit ainsi :

$$\varepsilon_{g,r,t} = (\gamma + \gamma_g + \gamma_{g,r})\varepsilon_{g,r,t-1} + u_{g,r,t}$$

$\varepsilon_{g,r,t}$ représente les résidus du modèle de tendance décrit précédemment : c'est-à-dire l'écart entre la quantité d'inflorescences mesurée et celle prédite par le modèle de tendance. γ représente le coefficient moyen du modèle AR1, γ_g le coefficient spécifique au génotype g pour ce modèle AR1, $\gamma_{g,r}$ le coefficient correspondant à la réplication r du génotype g (effet fixe). $u_{g,r,t}$ est le résidu du modèle pour la réplication r du génotype g à l'année t. Le modèle n'a pas d'ordonnée à l'origine puisque par construction les résidus sont centrés sur 0.

Ainsi, un génotype régulier aura des valeurs de $\gamma + \gamma_g$ proches de 0. Il en sera de même pour les génotypes irréguliers. Les génotypes alternants, eux, auront des valeurs fortes.

3.2. Modélisation des probabilités de transition à l'échelle locale

- *Développement du modèle*

Les modèles de Markov sont des outils mathématiques appropriés pour appréhender les dépendances entre évènements successifs. Ils consistent à estimer une variable à un temps t en fonction des M valeurs prises précédemment. Ces M valeurs définissent la mémoire m , mémoire utilisée pour prédire la valeur courante. Dans notre cas, la variable est la présence d'un bourgeon floral ($F_{g,r,t,m,l} = 1$) ou son absence ($F_{g,r,t,m,l} = 0$) qui dépend de g le génotype, r la répétition, t l'année et m la mémoire. Par exemple $m = 10$ correspond à une floraison à $t=t-2$ (2 ans auparavant), mais pas à $t=t-1$ (année précédente). La probabilité de $F=1$ doit donc être modélisée en fonction de tous ces paramètres.

Les séquences étudiées varient d'une longueur de 2 à 8 évènements. Ceci donne donc un nombre de mémoires m possibles gigantesque (2^7), et on peut supposer que l'évènement de l'année $n-8$ aura peu ou pas d'influence sur celui de l'année n . Ainsi, il a fallu déterminer quel est le nombre d'années optimal à prendre en compte pour prédire la probabilité de floraison.

La probabilité d'avoir une floraison ($F_{g,r,t,m,l} = 1$) suit une loi de Bernouilli. Ainsi, le nombre d'évènements de floraison par séquence suit une loi Binomiale de type

$$N_i \sim B(n_i, \pi_i)$$

avec N_i le nombre d'évènements de floraison, n_i la taille de la séquence et π_i la probabilité de ($F_{g,r,t,m,l} = 1$). Le modèle utilisé pour analyser une variable qui suit une loi binomiale est le modèle généralisé. La variable d'entrée n'est pas la quantité d'inflorescences mais le logarithme de la proportion d'inflorescences. Finalement, le modèle d'étude est décrit comme suit :

$$\log \left(\frac{P(F_{g,r,t,m,l} = 1)}{P(F_{g,r,t,m,l} = 0)} \right) = \Omega + \mu_m + \alpha_t + \beta_{g,m} + \sum_{g,t} + \gamma_r$$

Il s'agit d'un modèle linéaire mixte généralisé dans lequel Ω est un intercepte fixé, μ_m est l'effet fixe de la mémoire m , α_t est l'effet fixe de l'année t , $\beta_{g,m}$ est l'effet aléatoire d'interaction entre le génotype g et la mémoire m , $\sum_{g,t}$ est l'effet aléatoire d'interaction entre le génotype g et l'année t et γ_r est un effet aléatoire indépendant dépendant de la réplication.

Les paramètres du modèle ont été estimés par maximum de vraisemblance et les estimations des effets aléatoires (BLUPs : best linear unbiased prediction) ont été obtenus en utilisant les packages lme4 (Bates et al., 2011) de R version 2.8.

2 estimations sont intéressantes pour notre étude :

- $\beta_{g,m}$ permet d'estimer la probabilité d'avoir une floraison suite à une mémoire m pour un génotype g en particulier. Ainsi, un génotype alternant aura tendance à donner un bourgeon végétatif après une mémoire 01 (BLUP négatif), contrairement au génotype régulier qui lui donnera un bourgeon floral plus souvent (BLUP positif). Certaines mémoires devraient donc permettre de discriminer les génotypes alternants des réguliers.
- $\sum g, t$ permet d'évaluer la capacité qu'a un génotype à réagir à l'environnement. Les années dites « OFF », les individus alternant ont une probabilité très faible de donner un bourgeon floral (BLUP négatif), à l'inverse des réguliers qui eux donneront des bourgeons floraux (BLUP positif).

- *Adaptation du modèle à la famille*

Le modèle a été adapté aux caractéristiques de la famille d'étude. En effet, selon la structure de la famille et le processus de phénotypage correspondant, certains effets ont du être supprimés ou ajoutés. Ainsi, dans le cas où des répétitions existent, un effet individu a été envisagé. Pour les populations de la parcelle d'Angers, il a été envisagé de les étudier séparément puis dans un modèle commun avec un effet famille. Le choix des modèles a été réalisé par comparaison de modèles emboîtés et par l'examen des indices AIC (Akaike's information criterion) et BIC (Bayesian Information Criterion).

- *Formation des mémoires d'ordres variables*

Il existe une large gamme d'années mémoire à disposition. Une chaîne de Markov à ordre fixe prendra en compte toutes les mémoires réalisables pour un nombre d'année donné. Ainsi, pour des séquences de n années, on aura 2^n combinaisons possibles. Par exemple, une chaîne de Markov d'ordre 2 consiste à étudier les effets des années $n-1$ et $n-2$ sur la probabilité de floraison à l'année n . Il existe 4 mémoires différentes dans ce cas : 00, 10, 01, 11. Le modèle estimera la probabilité d'avoir floraison suite à deux années sans floraison (mémoire 00), etc.

Pour cette étude une chaîne de Markov à ordre variable a été utilisée (Guédon et Costes, 2012) : les mémoires peuvent avoir des longueurs comprises entre 1 et un ordre maximal. Une mémoire n'est développée à l'ordre suivant que si l'année ajoutée modifie la probabilité d'obtenir une bourse l'année n .

Reprenons l'exemple de la population X3263 x Belrène à l'ordre 2 : les mémoires étudiées sont 00, 01, 10, 11. Dans le cas d'une chaîne de Markov d'ordre variable maximum 3, si la probabilité d'avoir

une bourse après l'évènement 100 est significativement différente de celle après une mémoire 000 (Fig 8), alors la mémoire 00 sera développée en 100 et 000. Sinon, elle restera à l'ordre 2 (Fig 8).

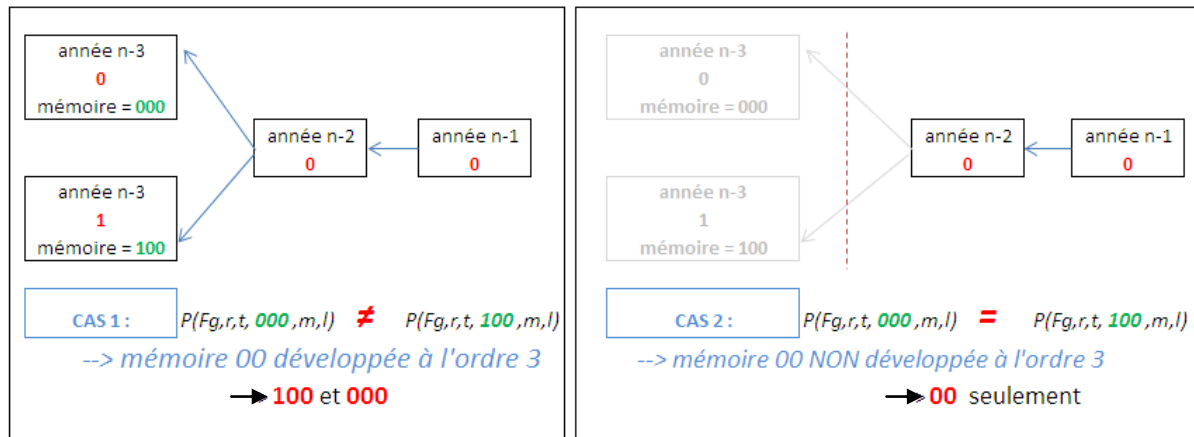


Figure 8 : développement de la mémoire 00 à l'ordre 3 (à gauche) ou à l'ordre 2 (à droite)

Ce même raisonnement peut être poussé à un ordre maximal m . A noter cependant que plus l'ordre maximal augmente, plus la quantité de données analysables diminue. En effet, pour une séquence de longueur de t années, le nombre d'évènements analysables sera seulement de $t-m$, puisque les m premières années seront utiles uniquement pour estimer l'année $m+1$. Ainsi, seules les mémoires allant de 2 à 4 années ont été envisagées.

• Choix de l'ordre maximal de la chaîne de Markov

Pour chacune des 5 populations un choix été réalisé entre des modèles d'ordre variable 2, 3 ou 4. Le choix a été réalisé selon deux critères :

- Le nombre de séquences de tailles suffisantes pour l'analyse. En effet, prendre en compte une mémoire d'ordre m signifie utiliser les m premières années de la séquence uniquement comme mémoire. Ces années ne serviront plus à estimer des probabilités de floraison. De plus, ceci diminue le nombre d'occurrences de chacune des mémoires, puisque le nombre de mémoires différentes augmente. Ainsi, si l'ordre est trop élevé, certains types de mémoire deviendront inexistantes chez une partie des individus, empêchant alors une comparaison avec le reste de la population. De cette façon, augmenter l'ordre des mémoires amène à estimer des probabilités de floraisons sur des quantités de données moindres.

Par exemple, le modèle d'ordre maximal 4 est caractérisé par les mémoires 0000, 0001, 0010, 0011 etc. Ainsi, le BLUP de l'effet d'interaction entre mémoire 1111 et génotype représente la probabilité qu'a un génotype d'avoir un bourgeon floral après avoir eu des bourgeons floraux durant 4 années

consécutives. Dans le cas où cet évènement n'aurait jamais eu lieu, le BLUP aura une valeur égale à 0. Le nombre d'individus ayant un BLUP égal à 0 permettra de déterminer si ces mémoire sont utilisables ou non en terme d'effectif.

- La capacité à discriminer les génotypes par rapport à leurs caractéristiques de production. La mémoire maximale la plus cohérente sera celle qui discriminera au mieux les différentes classes de génotypes. Si des génotypes à caractéristiques clairement distinctes se retrouvent avec des valeurs d'interaction génotypes x mémoire semblables, l'ordre en question n'aura pas d'intérêt. De plus, les résultats obtenus avec une mémoire m doivent être corrélés avec d'autres index représentant l'alternance, comme par exemple le BBI résiduel normalisé.

3.3. Modélisation de la synchronisation des branches

Comme exposé précédemment, la synchronisation entre chaque branche est un paramètre nécessaire à prendre en compte pour décrire le profil d'un génotype. L'indice proposé pour évaluer ce paramètre est l'entropie. Elle est calculée comme suit :

$$Ent_{g,r} = - \frac{1}{\sum_t n_{g,r,t}} \sum_t \left(n_{g,r,t,0} \log \frac{n_{g,r,t,0}}{n_{g,r,t}} + n_{g,r,t,1} \log \frac{n_{g,r,t,1}}{n_{g,r,t}} \right)$$

Ici $n_{g,r,t}$ représente le nombre d'évènements de floraison pour la réplication r du génotype g à l'année t. $N_{r,t,0}$ représente le nombre de bourgeons végétatifs pour ce même individu lors de cette même année, $N_{r,t,1}$ de même pour les bourgeons floraux.

La valeur d'entropie est comprise entre 0 et log 2, les valeurs les plus fortes représentant les génotypes les plus désynchronisés.

3.4. Validation des Index : Clustering de génotype

Afin de valider la pertinence des index proposés, ils ont été utilisés pour classer les génotypes par classification hiérarchique ascendante, selon la méthode de Ward. Cette classification est faite sous R à l'aide du package Cluster.

La répartition des individus au sein des classes ainsi obtenues donne des informations quant à la capacité des index à discriminer les génotypes selon leur profil de régularité. L'objectif est ici de déterminer un jeu d'index qui sera capable de discriminer efficacement les profils de régularité.

4. Création d'une carte génétique consensus et recherche QTL

4.1. Carte consensus

Afin de mettre en commun les QTLs trouvés dans chaque population séparément, il est nécessaire de les projeter sur une même carte. La première étape consiste donc à obtenir une carte consensus des 5 populations.

La carte consensus a été réalisée par le logiciel « MetaQTL », logiciel en libre accès développé par l'INRA (Veyrieras et al. 2005). MetaQTL travaille avec les marqueurs communs de chaque carte. Il utilise ces derniers pour faire le lien entre les différentes cartes génétiques initiales. Il devient alors possible de placer la totalité des marqueurs sur une seule et même carte. Dans le cas où des marqueurs ont des ordres différents d'une carte à une autre, les marqueurs en question sont supprimés.

Les cartes génétiques de chaque famille ont été développées par le passé, ce sont elles qui ont permis la création de la carte consensus. 4 cartes ont été utilisées correspondant aux 5 populations :

- carte génétique consensus des parents Granny et Starkinson (Segura et al., 2009) , 177 marqueurs au total. (carte_A)
- carte génétique consensus des parents X3263 et Belrène (Celton et al., 2011) , 210 marqueurs (carte_B)
- carte génétique consensus de la famille HIVW 292 marqueurs (carte_C)
- carte génétique consensus des familles N et P 89 marqueurs (carte_D)

La carte consensus comporte 534 marqueurs, le groupe de liaison 14 est donné à titre d'exemple à gauche de la figure 9.

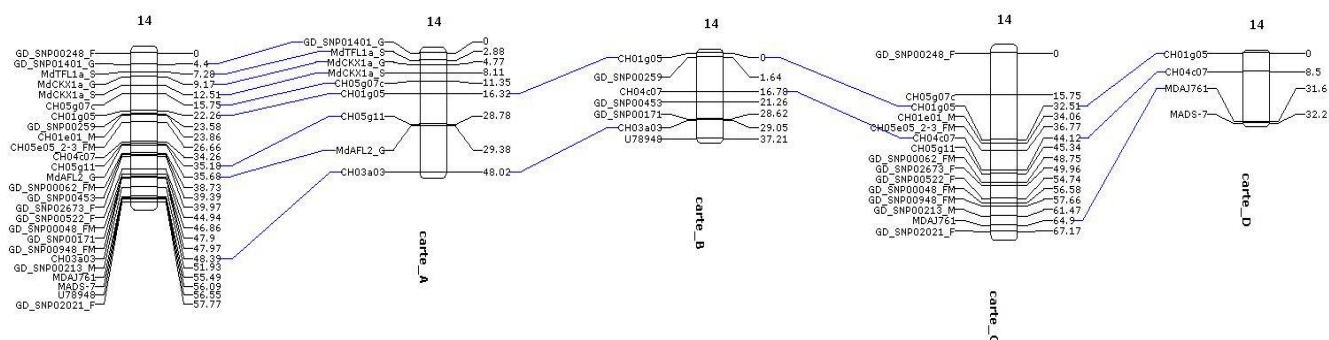


Figure 9 : représentation du GL 14 pour les cartes génétiques des 4 populations (à droite) et de la carte consensus (à gauche). Les marqueurs ponts sont représentés par des traits bleus.

4.2. Recherche QTL.

La recherche de QTL est tout d'abord réalisée famille par famille. Les QTLs détectés sont ensuite mis en relation sur la carte consensus obtenue à l'aide du logiciel MétaQTL.

- Démarche pratiquée

Le principe de la détection de QTLs repose sur une association entre un phénotype observé et un polymorphisme allélique révélé par un marqueur génétique. Plusieurs méthodes de détection de QTL existent, les plus utilisées étant la recherche simple marqueur, « l'interval mapping » (IM) et le « multiple QTL mapping » (MQM). La recherche QTL réalisée durant mon stage a utilisé chacune de ces méthodes. Elle a été réalisée à l'aide du logiciel MapQTL 6 (Van Ooiejn, 2004) qui permet d'effectuer chaque étape de raisonnement : tests non paramétrique, calcul des seuils de significativité, IM, sélection automatique de cofacteurs et MQM.

La méthode utilisée pour déterminer les seuils de significativité des QTLs repose sur des tests de permutation. Elle a l'avantage de ne pas dépendre de la distribution des données. Le nombre de permutations utilisé est 1000. La valeur seuil est déterminée avec un risque de 5%. Un QTL est validé à partir du moment où son LOD est supérieur à cette valeur seuil. L'ensemble des LOD seuils est donné en annexe.

L'analyse simple marqueur est ensuite réalisée sous forme d'un test non paramétrique (test de Kruskal-Wallis). La présence d'un QTL est testée sur chaque locus un à un. La présence d'un QTL au niveau d'un locus se traduit par une large différence de valeur du trait pour chaque classe allélique. Par contre, la détection d'un QTL présent entre 2 marqueurs espacés est impossible.

L'IM intègre les informations de la carte génétique dans l'analyse, afin de détecter la présence de QTL entre chaque marqueur, ce qui augmente la puissance de détection. Un LOD est calculé pour chaque position de la carte génétique. Toutes les régions qui présentent un LOD supérieur au LOD seuil sont des potentiels QTLs.

Le MQM est une méthode développée indépendamment par Jansen (1993) et Zeng (1993). Elle permet de s'affranchir de l'effet des QTLs précédemment détectés par IM en utilisant les marqueurs les plus proches de ces QTLs comme cofacteur. Elle permet de s'affranchir des effets d'interaction entre QTLs.

- Caractérisation des QTLs

Les QTLs détectés sont caractérisés par plusieurs paramètres. L'intervalle de confiance d'un QTL est situé entre les positions où sont observés les valeurs de LOD égale au LOD maximum du QTL plus ou moins 1. Un QTL est ensuite caractérisé par son importance relative sur la valeur du trait, c'est-à-dire la proportion de la variation phénotypique totale qui est expliquée par le QTL (R^2). Dans le cas où plusieurs QTLs sont détectés pour un même trait, la proportion de la variation phénotypique totale expliquée par l'ensemble de ces QTLs est calculée (R^2 global). Pour ce faire, un modèle linéaire est mise en place sous R, dans lequel les effets de chaque marqueur ainsi que leurs interactions sont pris en compte. Enfin, les valeurs moyennes du trait pour chaque classe allélique sont comparées et permettent d'estimer les effets d'additivité mâle et femelle, ainsi que les effets de dominance. Les effets d'additivité mâle sont donnés par la relation $A_m = ((\mu_{ac} + \mu_{bc}) - (\mu_{ad} + \mu_{bd}))/4$. Ici μ_{ac} est la moyenne du trait pour la classe allélique ac. Les effets d'additivité femelle valent $A_f = ((\mu_{ac} + \mu_{ad}) - (\mu_{bc} + \mu_{bd}))/4$. Les effets de dominance sont donnés par la relation $D = ((\mu_{ac} + \mu_{bd}) - (\mu_{ad} + \mu_{bc}))/4$.

Résultats

1. Sélection de géotypes remarquables

Les profils de production de la totalité des individus ont été tracés. La représentation de la quantité d'inflorescences par année (par branches séparées puis en les regroupant) permet de distinguer les différents profils de régularité présents au sein des 5 populations. Ces profils peuvent être regroupés en 4 classes :

- Les réguliers synchronisés présentent une production stable, à l'échelle de l'arbre ainsi qu'à l'échelle de chaque branche. Chaque branche suit le même profil de production.

- Les réguliers désynchronisés possèdent des branches à comportement irrégulier. Cependant, la quantité d'inflorescences totale des branches réunies présente elle un profil régulier.

- Les alternants produisent successivement beaucoup puis peu d'inflorescences.

- Les irréguliers présentent une production désorganisée : la quantité d'inflorescence varie d'une année à l'autre, mais sans suivre un comportement alternant.

Parmi la totalité des géotypes, certains ont été choisis graphiquement comme représentatifs de chaque type de comportement. Ces géotypes remarquables présentent clairement les caractéristiques de chaque profil. A titre d'exemple, quatre d'entre eux sont représentés dans la figure 10 :

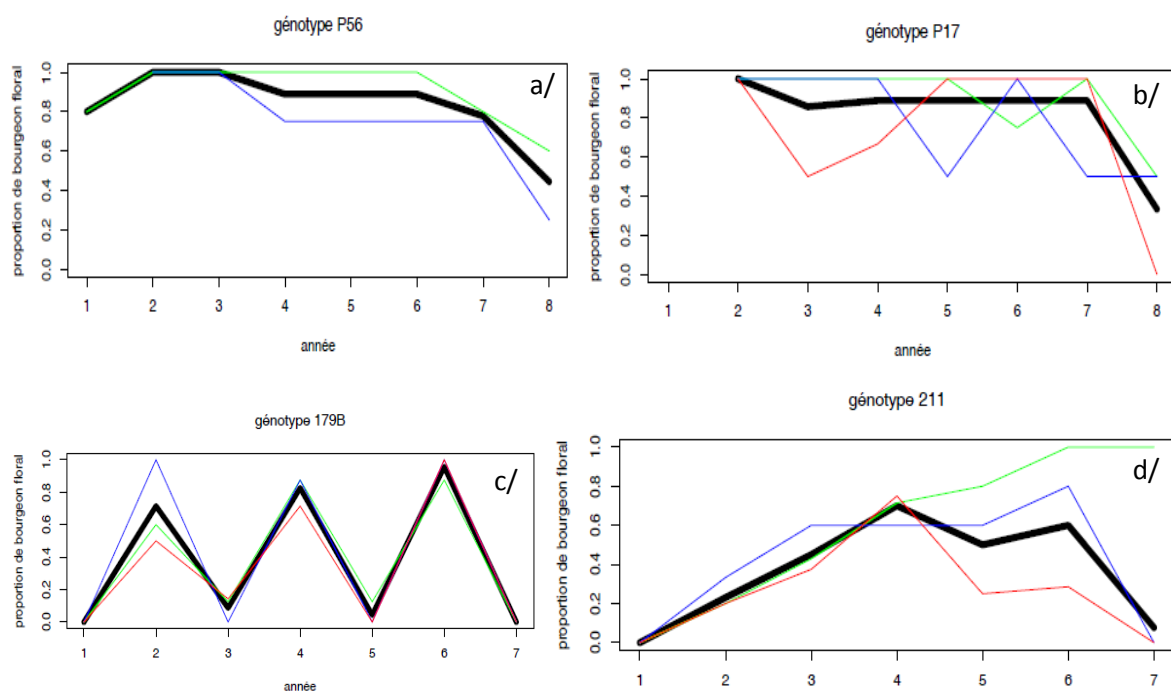


Figure 10 : Représentation graphique de la proportion de bourgeons floraux en fonction du temps pour un génotype régulier synchronisé (a/), régulier désynchronisé (b/), alternant (c/), et irrégulier (d/). Les 3 branches sont représentées une à une respectivement en rouge, vert et violet puis de façon regroupée en noir.

Neuf génotypes clairement irréguliers ont été sélectionnés, 12 réguliers synchronisés, 8 régulier désynchronisés et 8 alternants. Ces génotypes ont été choisis parmi les 5 populations. Ils peuvent servir de vérification, en effet, leurs index doivent présenter des valeurs cohérentes avec leur type de profil.

2. Modélisation de la quantité d'inflorescence

Les indices basés sur la quantité d'inflorescences peuvent être calculés à l'échelle locale comme à l'échelle globale. Dans le premier cas, la quantité d'inflorescences annuelle de l'échantillon sélectionné est modélisée. Dans le deuxième cas, c'est la production d'inflorescences de l'arbre entier qui est prise en compte.

2.1. Illustration du modèle de tendance

Le modèle de tendance permet de modéliser l'augmentation de la production en fleur au cours du temps. Les valeurs estimées par le modèle de tendance sont données par la relation :

$$\hat{Y}_{g,r,t} = \beta + \beta_g + \beta_{g,r} + (\alpha + \alpha_g + \alpha_{g,r} + E[\xi_{g,r} | Y = y])t$$

Ici $Y_{g,r,t}$ représente la quantité d'inflorescences estimée par le modèle de tendance pour la réplication r du génotype g l'année t . Un exemple est donné en figure 11 pour un génotype alternant.

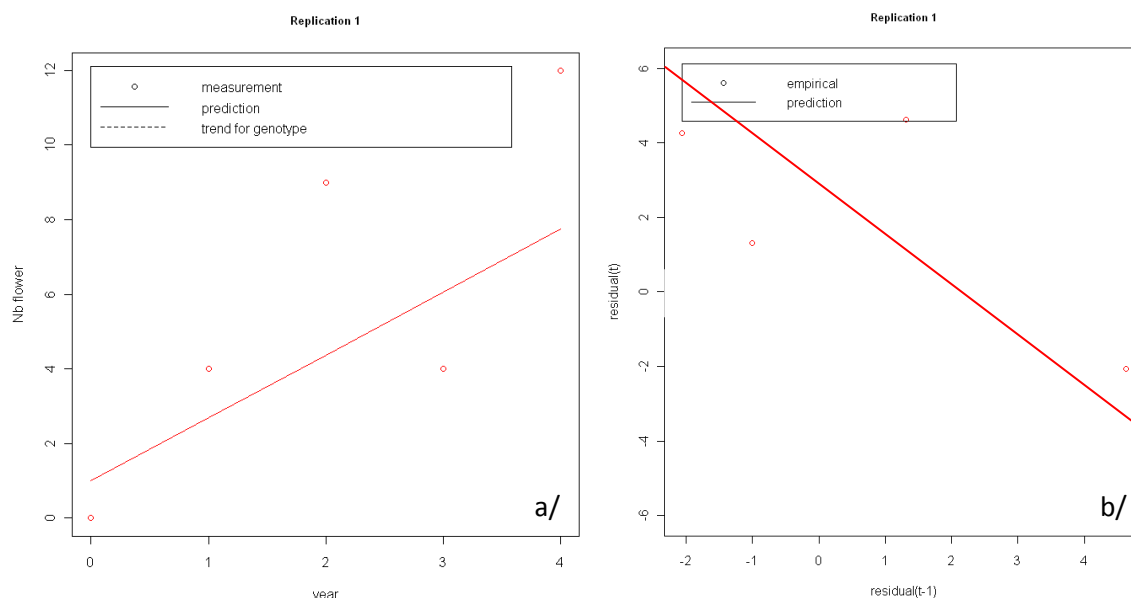


Figure 11 (a) : Représentation graphique de la production annuelle d'inflorescences pour le génotype 1 de la famille Starkinson x Granny (échelle locale). La droite de tendance modélisée par le modèle de tendance est représentée.

Figure 12 (b) : Représentation des résidus du modèle de tendance des années n (ordonnée) en fonction des résidus des années $n-1$ (abscisse).

Les résidus du modèle de tendance sont compris entre -15,5 et 15, ils sont directement utilisés pour calculer le BBI.res.norm. Dans le cas de ce génotype alternant, les valeurs observées sont très différentes des valeurs prédites (figure 11). Le BBI.res.norm est par conséquent élevé (BBI.res.norm = 0.97).

Le modèle auto régressif de premier ordre (AR1) utilise ensuite les résidus de ce modèle et analyse leur structure. Pour le génotype 1 présenté, la quantité d'inflorescence est successivement supérieure puis inférieure aux valeurs prédites. La modélisation des résidus de l'année n en fonction de l'année $n-1$ donne donc une pente négative (figure 12). Le coefficient d'auto corrélation est en effet égal à -0,448.

2.2. Présentation des Index

Le modèle de tendance développé précédemment a permis de calculer le BBI.res.norm ainsi que le coefficient d'auto corrélation. Les principales caractéristiques de ces index sont données dans le tableau 3, ainsi que celles concernant l'entropie.

Tableau 3 : présentation des index relatifs à la quantité d'inflorescences pour la population Starkinson x Granny (a), X3263 x Belrène (b) et les populations HIVW N et P confondues (c).

| | variable | min | max | moyenne | variance |
|----|-----------------|-------|------|---------|----------|
| a. | BBI.res.norm | 0,26 | 2,76 | 1,16 | 0,42 |
| | Coeff. Autocor. | -1,34 | 1,54 | -0,27 | 0,18 |
| | Entropie | 0 | 0,63 | 0,3 | 0,02 |
| | variable | min | max | moyenne | variance |
| b. | BBI.res.norm | 0,15 | 2 | 0,92 | 0,19 |
| | Coeff. Autocor. | -1,34 | 1,08 | -0,39 | 0,25 |
| | Entropie | 0 | 0,66 | 0,45 | 0,02 |
| | variable | min | max | moyenne | variance |
| c. | BBI.res.norm | 0,08 | 2,15 | 0,63 | 0,15 |
| | Coeff. Autocor. | -1,06 | 0,59 | -0,36 | 0,17 |
| | Entropie | 0 | 0,65 | 0,34 | 0,02 |

Le BBI.res.norm est compris entre 0.08 et 2.76. Sa valeur représente l'amplitude des résidus du modèle de tendance. Plus sa valeur est élevée, moins le génotype est régulier. Les valeurs de BBI.res.norm sont très proches d'une population à l'autre. Le coefficient d'auto corrélation est compris entre -1,34 et 1,54. Une valeur proche de -1 représente une corrélation négative entre résidus successifs, et donc un génotype de type alternant. L'entropie est comprise entre 0 et 0.6, les valeurs élevées représentant des génotypes ayant des branches très désynchronisées. Les valeurs des coefficients d'auto corrélation et d'entropie sont très proches d'une population à une autre.

2.3. Corrélation entre index.

Il existe une corrélation négative (coefficient de corrélation = -0,61) entre le BBI et le coefficient d'auto corrélation (tableau 4). Cette valeur correspond à la description des index : un génotype alternant possèdera un BBI élevé et un coefficient d'auto corrélation proche de -1.

Néanmoins, la corrélation entre la synchronisation (Entropie) des branches et la régularité (BBI et coefficient d'autocorrélation) est faible (Tableau 4). C'est pourquoi la prise en compte de la synchronisation des branches est importante : un génotype régulier peut contenir des branches régulières synchronisées où des branches irrégulières désynchronisées.

Tableau 4 : coefficients de corrélation de Pearson de chaque index modélisant la quantité d'inflorescence toutes familles confondues.

| | Entropie | | |
|------------------------|----------|--------------|------------------------|
| Entropie | 1 | BBI res norm | |
| BBI res norm | -0,37 | 1 | coeff. Autocorrélation |
| coeff. Autocorrélation | 0,19 | -0,61 | 1 |

2.4. Capacité des Index à décrire la régularité des génotypes

Le BBI résiduel normalisé, le coefficient d'auto corrélation et l'entropie doivent permettre de distinguer les différents profils de régularité. Ainsi, les individus remarquables sélectionnés précédemment doivent avoir des valeurs très distinctes selon leur type de profil.

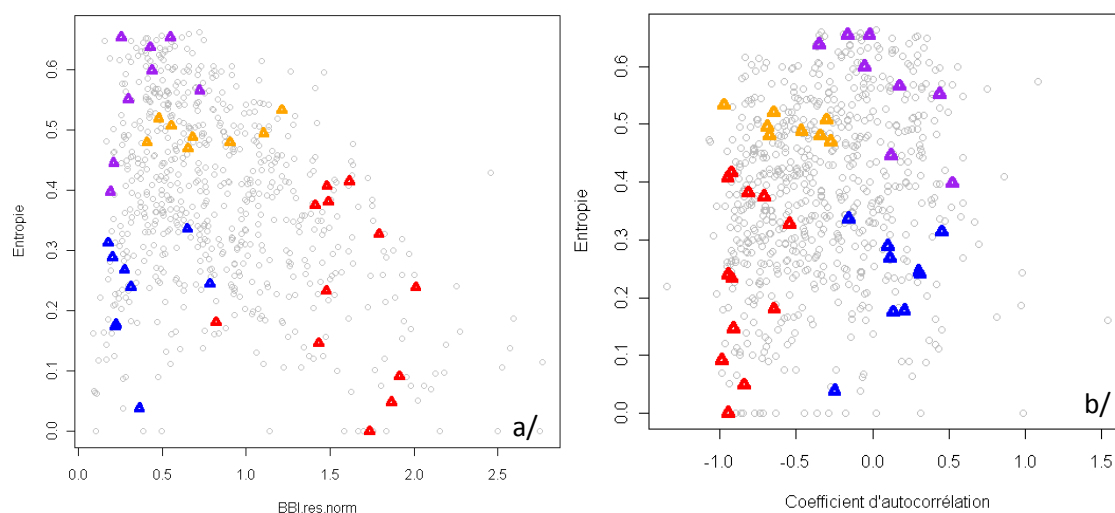


Figure 13 : Représentation graphique de l'entropie des génotypes de toutes les familles en fonction de leur BBI résiduel normalisé (a/) et de leur coefficient d'auto corrélation (b/). Les génotypes réguliers synchronisés sont indiqués en bleu, les réguliers désynchronisés en violet, les alternants en rouge et les irréguliers en orange.

La figure 13 est une représentation graphique des génotypes selon ces 2 index et l'entropie. Concernant le BBI.res.norm, les génotypes alternants (en rouge) se retrouvent effectivement avec des valeurs plus fortes que les réguliers (en bleu et violet). Les génotypes irréguliers se placent avec

des valeurs intermédiaires. L'entropie permet bien de discriminer les génotypes réguliers synchronisés qui ont une valeur faible (bleu) des réguliers désynchronisés (violet) qui ont une valeur forte. De plus, les arbres alternants sont plus synchrones que les irréguliers. Les mêmes conclusions peuvent être tirées concernant le coefficient d'auto corrélation : les génotypes remarquables alternants se situent avec des valeurs proches de -1, les autres profils étant plus proches de 0.

3. Modélisation des probabilités de transition (échelle locale)

3.1. Choix de l'ordre maximal de la chaîne de Markov

Les mémoires d'ordre 3 et 4 se sont révélées trop élevées pour que leurs effets puissent être estimés avec les données disponibles. En effet, la proportion de génotypes (exemple Fig 14) ayant eu au moins une fois 4 bourgeons floraux successifs (mémoire 1111) dans leur séquence est très faible (<35%).

Ainsi, la majorité des génotypes se retrouvent avec une estimation de leur interaction mémoire x génotype égale à 0 (exemple Figure 15), ce qui ne donne aucun renseignement concernant leur profil de production. Ce type de mémoire ne permettra donc pas de discriminer les génotypes entre eux et ne pourra par conséquent pas être utilisé.

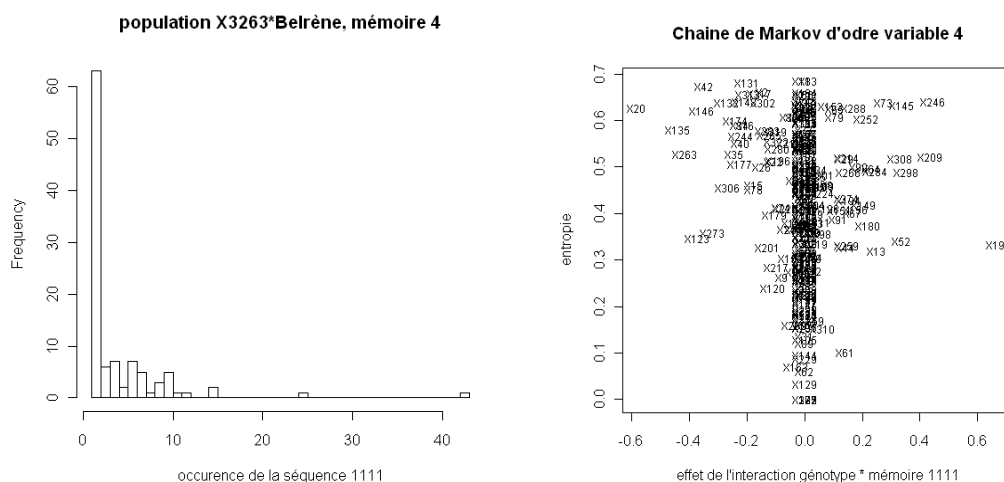


Figure 14 : Histogramme de la fréquence des occurrences de mémoire 1111 au sein des séquences de la population X3263 x Belrène

Figure 15 : Représentation de l'entropie en fonction de l'interaction génotype*mémoire 1111 pour tous les génotypes de la famille X3263 x Belrène.

La corrélation entre l'interaction génotypes * mémoire et le BBI résiduel normalisé (Figure 16) a été observée pour les mémoires d'ordre 2,3 et 4. Une corrélation est présente pour la mémoire d'ordre 2 (figure 16.a) : plus les individus sont alternants plus leur BBI est fort et plus leur effet d'interaction avec la mémoire 01 est faible (forte chance d'avoir un bourgeon végétatif après une bourse). Il n'existe pas de corrélation entre le BBI.res.norm et la mémoire d'ordre 4 (Figure 16.b). L'utilisation de mémoires d'ordre 2 est donc plus cohérente.

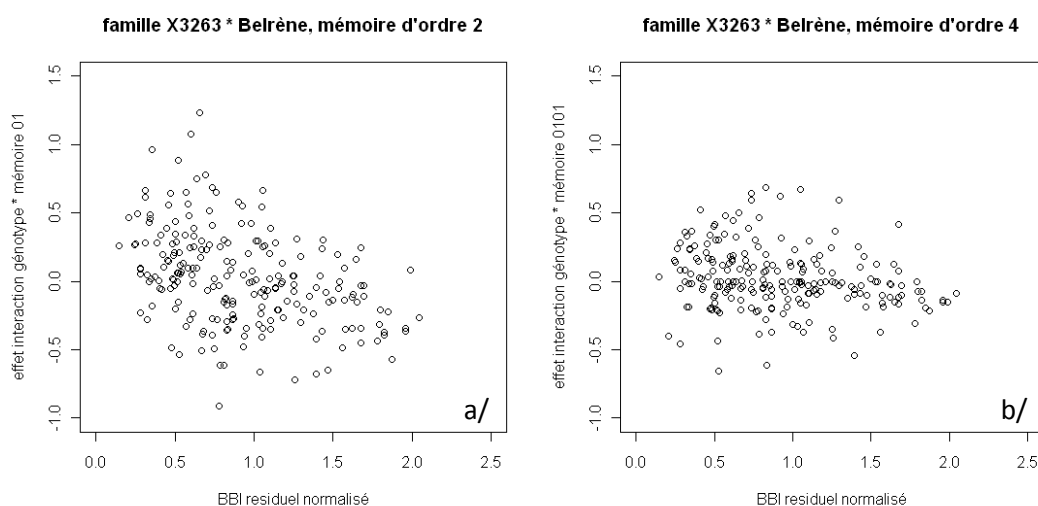


Figure 16 : représentation de l'effet d'interaction génotype * mémoire 01 (ordre 2) (a) et 0101 (ordre 4) (b) en fonction du BBI résiduel normalisé pour la famille X3263 x Belrène.

Les résultats obtenus avec l'ordre 3 sont similaires à ceux obtenus avec l'ordre 4. Compte tenus de toutes ces observations, la mémoire d'ordre 2 est la plus adaptée pour discriminer les profils de régularité des génotypes.

3.2. Adaptation des modèles aux familles

Concernant la population Starkimson x Granny, le modèle intégrant un effet individu s'avère plus performant qu'un modèle sans effet individu (AIC = 3198 contre AIC = 3236). C'est également le cas pour la population X3263 x Belrène (AIC = 26677 contre AIC = 27725).

Les familles HIVW, N et P ont été traitées ensemble, en ajoutant un effet famille fixe au sein du modèle. En effet, les effectifs des populations N et P étaient trop faibles pour être traitées séparément.

3.3. Résultat du modèle de probabilité de transition

Les résultats des modèles utilisés sont présentés dans la figure 17. L'intercept représente l'effet de la mémoire 00 pour la première année estimée. Les effets fixes de la mémoire 01 sont généralement les plus faibles de tous les effets fixes relatifs à la mémoire. Les génotypes alternants donnent effectivement très peu d'inflorescences après ce type de mémoire. Le raisonnement inverse est valable pour l'effet de la mémoire 10. Les effets fixes des années ne sont pas les mêmes selon les familles. La famille X3263 x Belrène présente une année OFF en 2010, les populations d'Angers principalement en 2012.

| | | Intercept | | Effet fixe : mémoire | | | Effet fixe : année | | | | |
|---------------------|------------|-----------|---------|----------------------|--------|-------|--------------------|--------|--------|-------|--|
| Famille | paramètre | γ | θ01 | θ10 | θ11 | 2008 | 2009 | 2010 | 2011 | 2012 | |
| X3263 x Belrène | estimation | -0,05 | -0,66 | 1,03 | -0,23 | - | 0,06 | -0,29 | 0,01 | 0,64 | |
| | p value | 0,68 | 1e-16 | 2E-16 | 0.0097 | - | 0.6651 | 0.0544 | 0.9633 | 1e-05 | |
| Starkimson x Granny | estimation | -2,58 | 0,64 | 3,03 | 1,22 | 3,3 | 1,06 | - | - | - | |
| | p value | 2E-16 | 0.00314 | 2E-16 | 1e-06 | 1e-16 | 0.00106 | - | - | - | |
| HIVW, N ,P | estimation | 1,21 | -0,14 | 1,61 | 0,7 | -0,12 | 0,02 | -1,74 | -0,42 | -2,68 | |
| | p value | 1E-12 | 0,28 | 1E-16 | 1e-08 | 0,58 | 0,89 | 1E-16 | 0,03 | 1E-16 | |

Figure 17 : valeur des effets fixes des modèles de probabilité de transition.

3.4. Capacité des index à décrire les génotypes

La figure 18 représente l'entropie en fonction des valeurs d'interaction génotype x mémoire pour chaque type de mémoire. La totalité des individus de la population X3263 x Belrène sont représentés. Les triangles de couleurs représentent les individus remarquables sélectionnés auparavant.

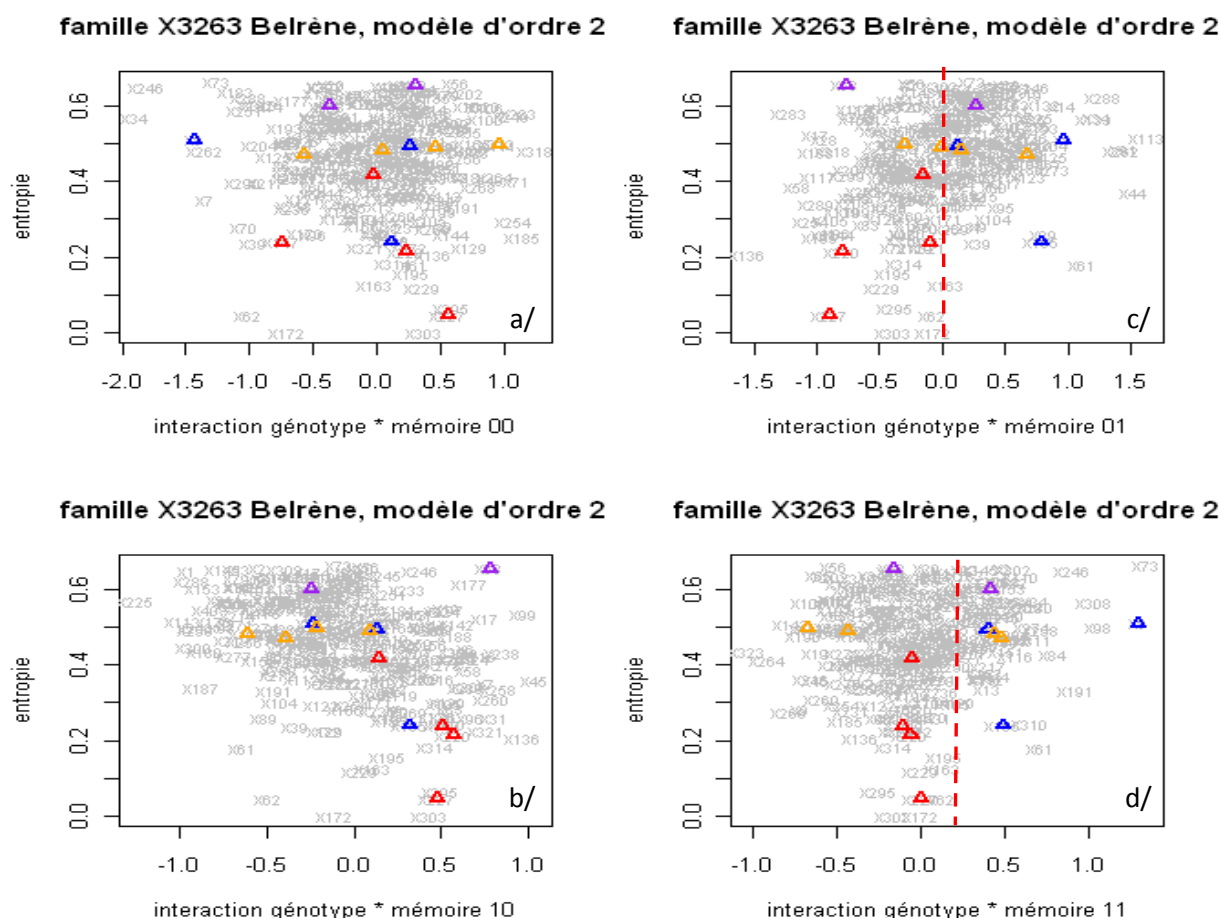


Figure 18 : Représentation graphique de l'entropie des génotypes de la famille X3253 Belrène en fonction de leur BLUP d'interaction génotype x mémoire 00 (a), génotype mémoire 10 (b), génotype mémoire 01 (c), génotype mémoire 11 (d). Les génotypes réguliers synchronisés sont indiqués en bleu, les réguliers désynchronisés en violet, les alternants en rouge et les irréguliers en orange.

Les mémoires permettant de discriminer les génotypes en fonction de leur alternance sont les mémoires 01 et 11. En effet, un génotype alternant aura tendance à donner un bourgeon végétatif suite à une succession bourgeon végétatif bourgeon floral. Il aura donc un Blup faible. Réciproquement, un génotype régulier aura tendance à donner des bourgeons floraux de manière générale (Blup positif). Les graphiques précédents (Figure 18) permettent de discriminer ces génotypes remarquables : les alternants sont systématiquement situés à l'opposé des réguliers synchronisés. Seuls les irréguliers semblent assez mal répartis.

A contrario, les mémoires 00 et 10 ne permettent pas de discriminer les génotypes (génotypes remarquables mélangés indépendamment de leur profil). Après une succession bourgeon floral bourgeon végétatif, un génotype alternant donnera à nouveau un bourgeon floral, il en va de même

pour un individu régulier qui a tendance tout le temps produire. Il est donc logique que ces mémoires ne permettent pas de discriminer les différents profils de production.

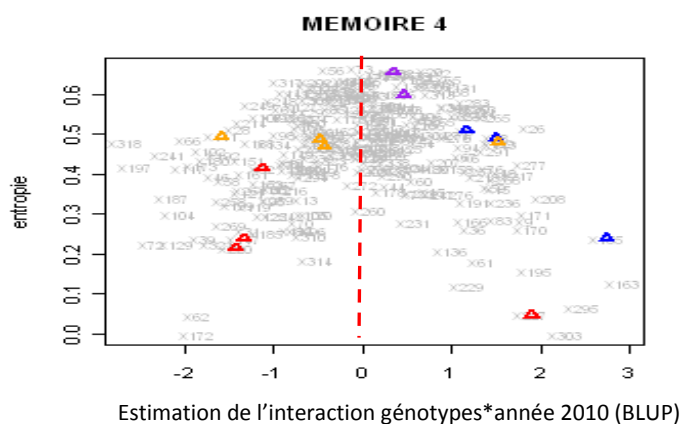


Figure 19 : Représentation graphique des génotypes de la famille X3253 x Belrène en fonction de leur BLUP d'interaction génotype * année 2010 et de leur entropie. Génotypes réguliers synchronisés en bleu, les réguliers désynchronisés en violet, les alternants en rouge et les irréguliers en orange.

En ce qui concerne l'interaction entre génotypes et années, ce sont les années dites « OFF » qui sont intéressantes (Exemple de l'année 2010 figure 19). En effet, lors d'une année OFF, la majorité des génotypes alternants donnent très peu d'inflorescences. De ce fait, ils présentent un BLUP très faible (Inférieur à -1). Une discrimination avec les génotypes réguliers est alors possible.

Dans la majorité des cas (Exemple figure 19), les individus réguliers sont opposés aux génotypes irréguliers, que ce soit pour leurs interactions génotype*mémoire ou pour leurs interactions génotypes*années. De plus, les génotypes désynchronisés ont systématiquement une entropie plus forte.

4. Classification des génotypes

La classification hiérarchique ascendante a permis de répartir l'ensemble des génotypes en 4 classes. Chaque classe représente un type de profil (alternant, régulier synchronisé, régulier désynchronisé et irrégulier). Cette classification s'effectue à partir d'un jeu de variables. Une multitude de solutions ont été testées, utilisant un nombre variable d'indices parmi tous ceux développés jusqu'à présent.

Finalement, les index retenus pour maximiser la cohérence de la classification des génotypes sont l'entropie, le BBI résiduel normalisé et le coefficient d'auto corrélation. Les résultats de cette classification sont présentés ci-dessous. Pour que cette classification soit satisfaisante, les 4 classes

doivent être clairement séparées par les 3 index, et les individus remarquables doivent être classés dans leur profil correspondant.

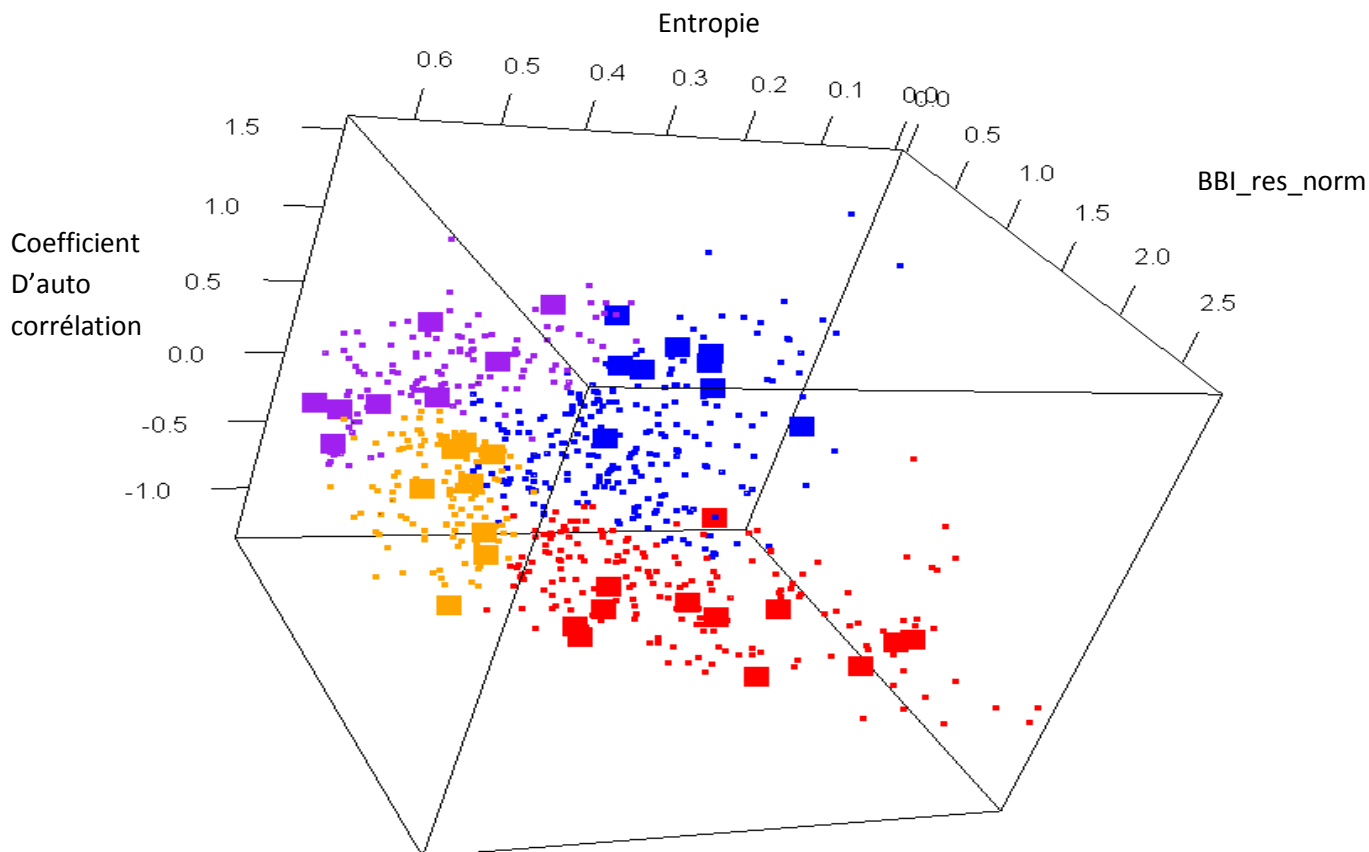


Figure 20: Représentation graphique de l'ensemble des génotypes selon 3 variables : entropie (x), coefficient d'auto corrélation (y) et BBI résiduel normalisé (z). Les couleurs représentent les 4 clusters d'individus obtenus par classification hiérarchique ascendante. Les carrés représentent les individus sélectionnés comme remarquables de chaque classe. les génotypes réguliers synchronisés sont représentés en bleu, les réguliers désynchronisés en violet, les alternants en rouge et les irréguliers en orange.

La représentation en 3 dimensions des génotypes (figure 20) met en avant une claire distinction des 4 clusters. La combinaison des 3 index utilisés permet donc de discriminer efficacement les différents profils de régularité de production.

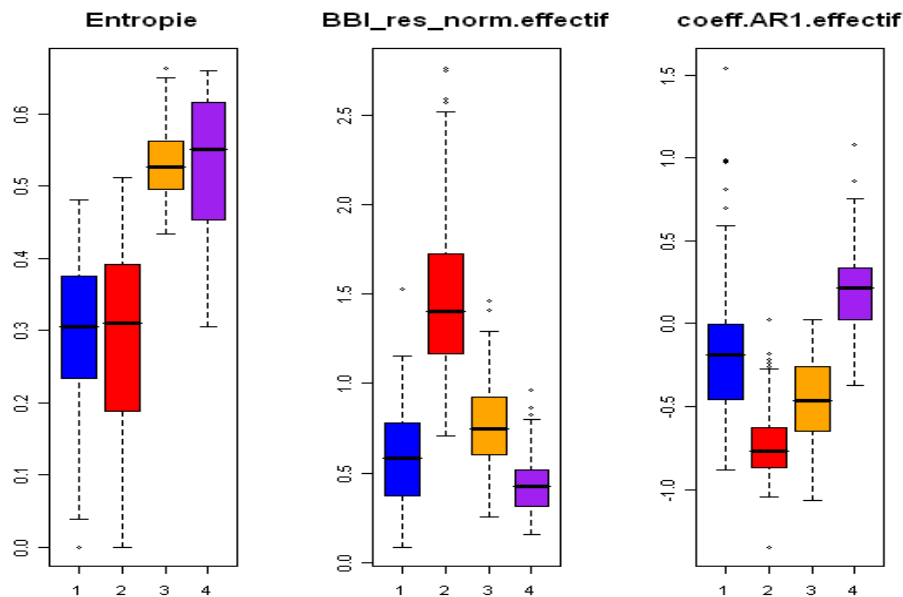


Figure 21 : Représentation des valeurs d'entropie, de BBI résiduel normalisé et de coefficient d'auto corrélation pour les 4 clusters : réguliers synchronisés (bleu), réguliers désynchronisés (violet), alternants (rouge) et irréguliers (orange).

La figure 21 représente les valeurs moyennes de chaque variable par classe. Concernant l'entropie, on observe une séparation entre les génotypes réguliers synchronisés (bleu) et alternants (rouge) par rapport aux autres. L'entropie étant la variable représentant la synchronisation des branches ce résultat est cohérent. Les génotypes irréguliers sont effectivement synchronisés : une désynchronisation de leur branche ne leur permettrait pas d'obtenir une variation si importante de production d'une année sur l'autre. Le BBI résiduel normalisé et le coefficient d'auto corrélation permettent de faire une distinction entre les génotypes réguliers, irréguliers et alternants.

Tableau 5 : Répartition des individus remarquables au sein des 4 clusters.

| | classe 1 | classe 2 | classe 3 | classe 4 |
|-------------------------|----------|----------|----------|----------|
| irrégulier | 8 | 0 | 0 | 1 |
| régulier synchronisé | 1 | 11 | 0 | 0 |
| régulier désynchronisés | 0 | 0 | 8 | 0 |
| alternants | 0 | 0 | 0 | 8 |

Concernant les génotypes remarquables, la majorité d'entre eux se situent dans la classe qui leur correspond (tableau 5). Par exemple, le génotype 82 de la famille X3263 x Belrène qui présente clairement une production alternante se retrouve dans le cluster 2 qui regroupe les génotypes alternants, et ce grâce à la valeur de ses 3 index. Ainsi, sur 8 génotypes sélectionnés comme

clairement irrégulier, 8 se retrouvent dans la même classe (classe 1). Il en est de même avec les génotypes remarquables réguliers synchronisés pour lesquels seul un génotype ne se trouve pas dans la bonne classe. La totalité des génotypes remarquables alternants et réguliers désynchronisés se regroupent de manière conforme (classe 4 et 3 respectivement).

Ainsi, les classes étant clairement distinctes entre elles et permettant de placer correctement les individus remarquables dans leurs classes respectives, on peut estimer que les indices de modélisation sont pertinents. Ils permettent d'appréhender la régularité de production d'un génotype.

5. Corrélation entre alternance à échelle locale et globale.

Les index développés ci-dessus étant à présent validés, ils peuvent servir à comparer les profils de régularité estimés à 2 échelles : locale (échantillon de l'arbre) et globale (arbre dans sa totalité).

5.1. Corrélation des indices

Les indices relatifs à la quantité d'inflorescence (BBI résiduel normalisé et coefficient d'auto corrélation) ont été calculés aux deux échelles pour la population Starkimson x Granny.

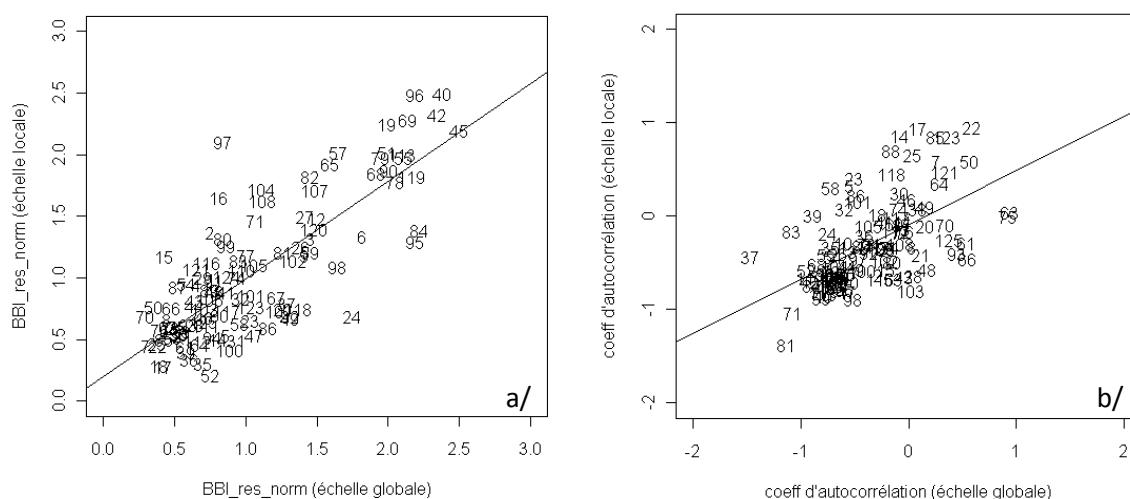


Figure 22 : Relation entre les BBI résiduels normalisés (a) et entre les coefficients d'auto corrélation (b) calculés à échelle globale et locale.

Les coefficients de corrélation (Figure 22) entre le BBI aux échelles locales et globales et entre le coefficient d'auto corrélation à ces deux échelles sont respectivement de 0.7921 et 0.5615.

L'utilisation d'un échantillon de l'arbre seulement entraîne donc une perte de 20 à 45% de l'information contenue dans l'arbre. Globalement, il est donc possible d'approximer le comportement d'un génotype en phénotypant seulement un échantillon de celui-ci.

Il existe parfois des différences de comportement entre échelle globale et locale. C'est dans ce genre de cas (exemple du génotype B097 figure 23) que l'échantillonnage ne permet pas de représenter le comportement général de l'arbre.

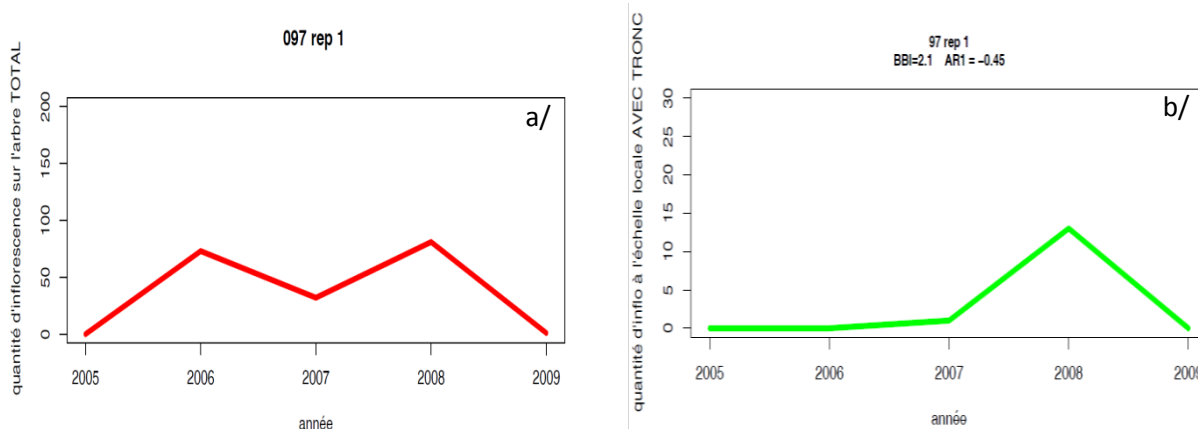


Figure 23 : Quantité de bourgeons floraux par année mesurée à l'échelle globale (a), puis sur un échantillon de l'arbre (échelle locale (b))

5.2. Classification des génotypes

Une classification hiérarchique ascendante a été effectuée aux deux échelles, séparant à chaque fois la totalité des génotypes en 4 classes, supposées représenter les 4 types de profils de production. Dans le cas où l'échelle locale suffirait à estimer l'échelle globale, chaque génotype devrait se placer dans la même classe dans les 2 cas.

Les variables utilisées pour réaliser la classification sont à nouveau le BBI.res.norm, le coefficient d'auto corrélation et l'entropie. L'entropie est la même dans les 2 cas. Elle ne peut en effet être calculée qu'à l'échelle locale. Leurs ordres d'importance lors de la classification sont inscrits dans le tableau 7. Il apparaît que les classes ont été formées à partir de variables différentes.

Tableau 6 : variables utilisées aux deux échelles lors de la classification hiérarchique ascendante

| | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--------|----------------------|--|--|--|--|
| | profil correspondant | régulier synchronisé | irrégulier | alternant | régulier désynchronisé |
| local | variables utilisées | entropie BBI_res_norm | entropie coeff d'auto corrélation | BBI_res_norm entropie coeff d'auto corrélation | coeff d'auto corrélation BBI_res_norm |
| global | variables utilisées | BBI_res_norm entropie coeff d'auto corrélation | BBI_res_norm entropie coeff d'auto corrélation | coeff d'auto corrélation entropie BBI_res_norm | entropie BBI_res_norm |

Ainsi, à l'échelle locale, les individus régulier synchronisés sont discriminés principalement par leur valeur d'entropie, contrairement à l'échelle globale où c'est le BBI résiduel normalisé qui a le plus d'influence.

Pour chaque classification, les valeurs moyennes des 3 variables ont permis d'estimer quel type de profil était représenté par chaque classe.

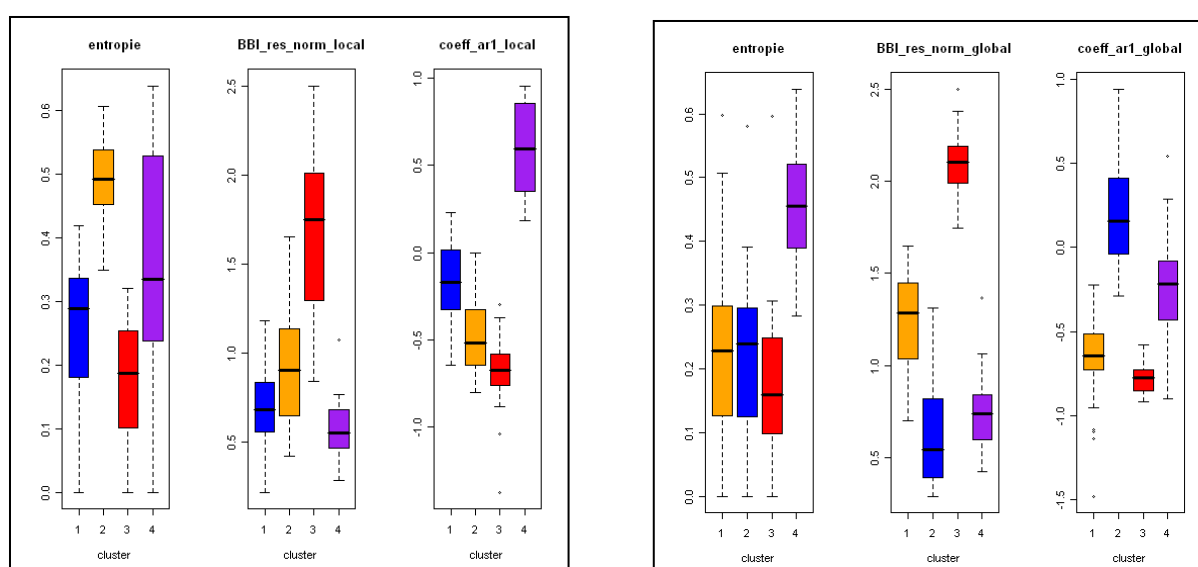


Figure 24 : box plot des 4 clusters pour les 3 variables utilisées. L'échelle locale est représentée à gauche, l'échelle globale à droite. Le cluster bleu représente les génotypes réguliers synchronisés, le violet les réguliers désynchronisés, le rouge les alternants, le orange les irréguliers.

Les valeurs des variables de chaque classe sont très proches pour chaque échelle. Elles correspondent à celles décrites précédemment (4.).

Tableau 7 : effectif de chaque cluster pour les 2 échelles

| | Régulier synchronisés | irrégulier | alternant | Régulier désynchronisés |
|-----------------|-----------------------|------------|-----------|-------------------------|
| Echelle locale | 34 | 28 | 36 | 17 |
| Echelle globale | 23 | 29 | 18 | 45 |

Cependant, les effectifs de chaque cluster sont très différents lorsque l'on passe d'une échelle à l'autre (tableau 6). Par exemple, 17 génotypes appartiennent à la classe des réguliers désynchronisés à l'échelle locale, contre 45 à l'échelle globale.

5.3. Comparaison des 2 classifications.

Le tableau 8 donne le nombre de génotypes passant d'une classe à l'autre lors du changement d'échelle. Ainsi, sur les 23 génotypes considérés comme régulier synchronisés à l'échelle globale, 13 ont été considérés comme réguliers synchronisés à l'échelle locale, 1 irrégulier, 1 alternant et 8 réguliers désynchronisés.

Tableau 8 : effectif passant d'un cluster à un autre lors du changement d'échelle

| | | local | | | |
|--------|-------------------------|-----------------------|-------------------------|------------|-----------|
| | | Régulier synchronisés | Régulier désynchronisés | irrégulier | alternant |
| global | Régulier synchronisés | 13 | 8 | 1 | 1 |
| | Régulier désynchronisés | 13 | 8 | 23 | 1 |
| | irrégulier | 7 | 1 | 3 | 18 |
| | alternant | 1 | 0 | 1 | 16 |

Globalement, les 2 classifications donnent donc des résultats assez différents. Par exemple, la majorité des individus irréguliers de l'échelle globale ont été considérés alternant à l'échelle locale. Cependant, il n'existe quasiment pas de confusion entre les génotypes alternants et les réguliers. Par exemple, dans la quasi-totalité des cas, un individu alternant sera classé alternant ou irrégulier. Il n'y a presque jamais de confusion avec les classes des génotypes réguliers. De même, un individu régulier synchronisé est dans la majorité des cas reclassé régulier synchronisé ou régulier désynchronisé. Il est très rarement classé avec les irréguliers ou les alternants.

6. Analyse QTL

Au total, 8 QTLs ont pu être identifiés dans 3 familles sur 5, présentés dans le tableau 7. Les populations N et P n'ont pas permis de détecter de QTL. En effet, le nombre d'arbres phénotypés n'était pas suffisant et leur cartes génétiques n'étaient pas saturées. La variabilité génétique expliquée par les QTLs détectés varie de 6.2 à 17.7 %. Le tableau 9 présente l'intégralité des QTLs avec leurs principales caractéristiques.

| | Variable | Group | Position | Locus | LOD | R ² | Af | Am | D |
|----|--------------|-------|----------|----------------|------|----------------|-------|-------|-------|
| a. | BBI.res.norm | LG8 | 12,829 | CH01e12_G | 4,36 | 12,4 | -0,20 | -0,10 | -0,15 |
| | | LG1 | 46,089 | MdBFTa_SG | 3,1 | 8,6 | -0,01 | 0,19 | -0,07 |
| | | LG15 | 87,121 | CH02c09_SG | 3,13 | 8,7 | 0,08 | 0,18 | -0,06 |
| | .01 | LG8 | 19,322 | Hi04b12_S | 7,16 | 17,7 | 0,11 | 0,09 | 0,28 |
| | | LG1 | 37,64 | AG04_S | 4,61 | 10,8 | 0,04 | -0,24 | 0,01 |
| | | LG15 | 58,884 | CH02d11_SG | 3,6 | 8,2 | -0,10 | -0,10 | 0,17 |
| | .00 | LG11 | 26,141 | CH02d12_SG | 5,3 | 13,1 | -0,23 | 0,00 | 0,08 |
| | | | | | | | | | |
| | Variable | LG | Position | Locus | LOD | R ² | Af | Am | D |
| b. | .01 | LG01 | 40,219 | GD_SNP01772 | 3,8 | 6,6 | 0,15 | 0,05 | -0,03 |
| | BBI_res_norm | LG10 | 18,821 | CH2C11A | 4,08 | 6,7 | 0,10 | -0,12 | 0,02 |
| | | LG10 | 47,703 | GD_SNP0257 | 3,78 | 6,2 | 0,02 | 0,14 | 0,03 |
| | | | | | | | | | |
| | Variable | Group | Position | Locus | LOD | R ² | Af | Am | D |
| c. | Entropie | LG06 | 2,494 | GD_SNP01996_FM | 4,16 | 9,1 | -0,03 | -0,02 | 0,04 |
| | BBI_res_norm | LG03 | 11,195 | GD_SNP00882_F | 3,63 | 8,2 | 0,05 | 0,13 | 0,02 |
| | | LG10 | 10,228 | CH01b12_FM | 3,58 | 8 | -0,02 | 0,06 | 0,10 |
| | | LG09 | 5,336 | GD_SNP00514_M | 3,05 | 6,8 | -0,08 | -0,08 | 0,00 |
| | AR1 | LG06 | 17,604 | CH03d07_FM | 4,33 | 8,9 | -0,02 | -0,12 | 0,02 |
| | | LG09 | 3,68 | GD_SNP02013_F | 4,29 | 8,8 | 0,09 | 0,01 | 0,07 |
| | | LG10 | 12,228 | CH01b12_FM | 4,19 | 8,5 | 0,01 | -0,08 | -0,09 |
| | | LG03 | 20,948 | CH03g07_FM | 3,42 | 6,9 | -0,07 | -0,10 | -0,01 |

Tableau 9 : QTLs détectés par MQM mapping pour les variables relatives à l'alternance concernant les populations Starkimson x Granny (a), X3263 x Belrène (b) et HIVW (c). Chaque tableau donne la variable, le groupe de liaison sur lequel le QTL a été détecté (LG), la position du QTL sur ce LG, le cofacteur utilisé pour la détection, le LOD score et la proportion de variance génétique expliqué (R²). Les effets additifs femelles (Af) et mâles (Am) ainsi que l'effet de dominance (D) sont aussi indiqués.

Au total, 7 QTLs ont été détectés pour la population Starkimson x Granny. Trois d'entre eux concernent le BBI.res.norm, sur 3 groupes de liaison différents (8, 1 et 15), avec des LOD variant de 3,1 à 4,36. Ces QTLs regroupés expliquent 27% de la variabilité phénotypique totale (R² global). Ces

mêmes QTLs sont retrouvés pour l'interaction mémoire 01 x génotype, avec des LOD compris entre 3,6 et 7,16 et un R^2 global de 0.10. Une dernière zone chromosomique est détectée concernant l'interaction mémoire 00 x génotype, sur le chromosome 11, expliquant 13,1% de la variabilité phénotypique totale.

Le QTL du groupe de liaison 1 est retrouvé chez la population X3263 x Belrène, pour l'interaction mémoire 01 x génotype. Il présente un LOD score de 3,8 et explique 6,6% de la variabilité totale. Deux QTL sont présent concernant le BBI.res.norm, tous deux situés sur le groupe de liaison 10, aux positions 18,8 et 47,7, avec des LOD de 4,08 et 3,78 respectivement.

Le premier QTL de ce groupe de liaison est retrouvé chez la population HIVW, pour le BBI.res.norm (LOD de 3,58). Deux autres QTLs ont été détectés pour ce caractère, sur les groupes de liaisons 3 et 9 (LOD de 3,63 et 3,05). Ces 3 régions sont également détectées pour le coefficient d'auto corrélation, avec des LOD toujours supérieurs à 4. Le groupe de liaison 6 présente un QTL pour le coefficient d'auto corrélation ainsi que pour l'entropie (LOD de 4,33 et 4,16 respectivement). Le pourcentage de la variabilité phénotypique totale expliqué par ces QTLs est de respectivement 14% et 35% pour le BBI résiduel normalisé et le coefficient d'auto corrélation.

En résumé, sur les 8 zones chromosomiques d'intérêt, 7 sont présentes pour deux variables au moins. De plus, Deux QTLs ont été détectés sur 2 populations (groupes de liaison 1 et 10). Ces 2 groupes de liaison sont représentés dans la figure 25.

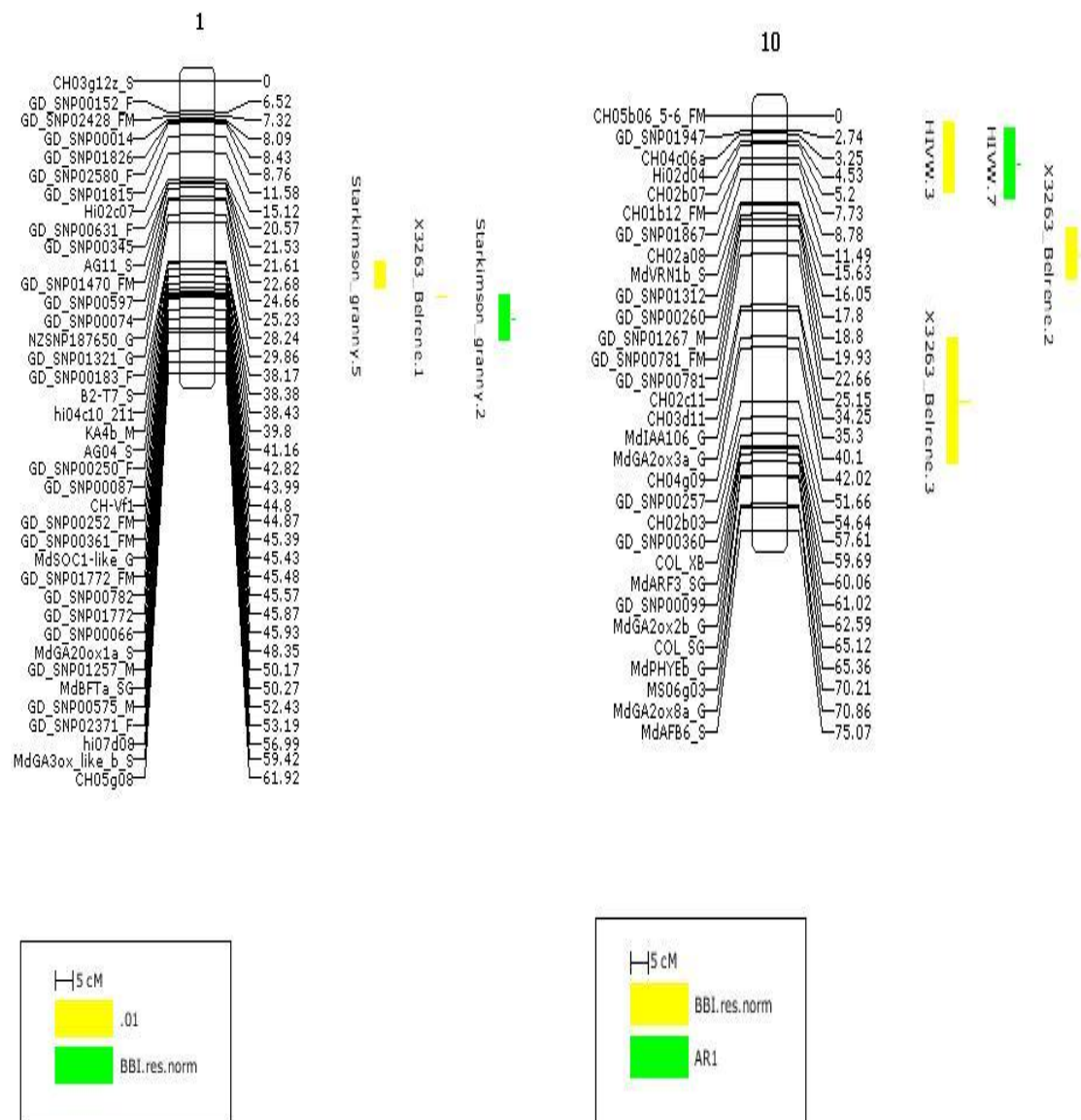


Figure 25 : représentation des groupes de liaisons 1 et 10 de la carte consensus. Les QTLs détectés sont représentés par des bandes de couleurs. La taille de chaque bande représente l'intervalle de confiance du QTL, la couleur la variable en question, le texte la famille correspondante. (Figure obtenue sous MétaQTL)

Discussion

1. Possibilité d'appliquer le protocole de phénotypage à large échelle ou sur d'autres espèces

L'alternance est un phénomène généralement étudié sous l'angle de la fluctuation de la production en fruits au cours du temps (Cilas et al., 2011). Pourtant, cette étude se base sur la quantité d'inflorescences. Ce choix résulte du fait que la quantité d'inflorescence est moins sujette aux effets environnementaux que certains caractères en relation avec les fruits. Par exemple, des phénomènes tels que la charge en fruits ou la chute de fruits due au climat ou aux maladies dépendent énormément de paramètres environnementaux (Guitton et al., 2011). Ainsi, il paraît plus cohérent de s'intéresser à la quantité d'inflorescences qui reflète mieux le déterminisme génétique de l'alternance.

Le protocole de phénotypage à échelle locale a consisté en une observation rétrospective des événements floraux. Ce mode d'observation n'est possible que dans le cas où les événements de floraison sont visibles à posteriori. Ainsi, ce phénotypage rétrospectif sera adaptable à toutes espèces fruitières présentant les mêmes caractéristiques. C'est le cas de la poire (Huet, 1972) ou de la noix (Sabatier et al., 1998). Ce sera par contre plus difficile chez des espèces à floraison latérale tel que celles appartenant au genre *Prunus* (Fournier et al., 1994 ; Costes et al., 2006), puisqu'il n'existe pas de marqueurs morphologiques permettant une analyse rétrospective des événements de floraison.

Les indices caractérisant les profils de production ont été calculés dans diverses conditions agronomiques. Les arbres des parcelles d'Angers sont taillés régulièrement, la conduite de ces parcelles est fidèle celle mise en place couramment. Les parcelles de Mauguio, en revanche, se développent librement sans taille. De plus, l'âge des arbres étudiés varie énormément selon la parcelle, la taille des séquences relevées varie donc aussi. Finalement, les protocoles de phénotypage étant adaptés à chaque population, la taille et le nombre de séquence par arbre présentent eux aussi des modalités différentes. Les indices développés permettent de discriminer les profils de production dans tous ces cas de figure. Ils sont donc robustes et peuvent être utilisés dans une multitude de cas de figure.

2. Perfectionnement des indices

Les indices développés permettent de caractériser efficacement le profil de régularité d'un génotype. Cependant, certains profils particuliers peuvent parfois être mal évalués en raison de subtilités échappant au calcul des indices.

2.1. Sensibilité de l'interaction mémoire génotype à la longueur des séquences.

Les indices prenant en compte les séquences dans leur ensemble (BBI.res.norm, entropie etc.) semblent mieux rendre compte de la régularité d'un individu que ceux basés sur les probabilités de transition. L'interaction entre une mémoire et un génotype représente la probabilité d'avoir une inflorescence après une séquence donnée. Ainsi, dans le cas où cette mémoire n'apparaîtrait qu'une seule fois à l'échelle du génotype, la probabilité sera représentée par une seule occurrence, ce qui amène à une estimation peu représentative du profil du génotype.

Par exemple, le génotype numéro 39 de la famille Starkimson x Granny présente un profil très régulier (figure 26), et a été phénotypé sur 28 séquences au total. Ce génotype sort de sa phase juvénile en 2007, il ne produit pas d'inflorescence avant. Ensuite, il augmente progressivement et régulièrement sa quantité d'inflorescence tous les ans. Ce génotype présente un profil de production régulier.

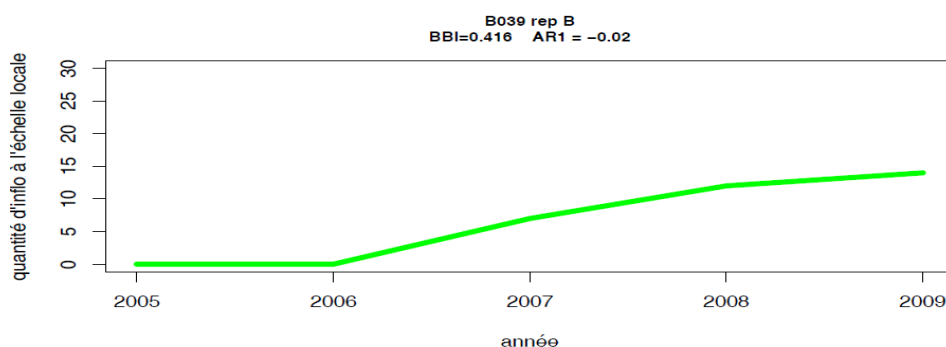


Figure 26 : profil de production du génotype 39 de la population Starkimson x Granny.

L'interaction génotype mémoire 01 discrimine bien les réguliers des alternants : les alternants sont bien en année OFF après une année ON (interaction <0), les réguliers restent quand à eux en production une fois sortis de leur phase juvénile (interaction >0). Le génotype 39 devrait donc se retrouver dans ce deuxième cas.

Cependant, l'estimation de son interaction mémoire 01 génotype est de -0,046, ce qui est une valeur relativement faible, légèrement en dessous de la moyenne de la population et qui correspond plutôt à un génotype alternant.

L'explication de cette mauvaise estimation réside dans le nombre de fois où la mémoire 01 apparaît. Sur les 28 séquences du génotype 39, l'occurrence bourgeon végétatif suivi de bourse (01) n'est présente que 2 fois. Dans les 2 cas, elle donnera suite à un bourgeon végétatif. Ainsi, ce génotype sera mal représenté par cette variable calculée sur un nombre d'occurrences trop faible.

Les variables d'interaction génotype mémoire sont donc très sensibles à la longueur des mémoires et des séquences prises en compte. Il faut donc être très vigilant lors de leur utilisation, notamment sur des populations telles que Starkimson x Granny pour laquelle les séquences sont très courtes.

2.2. Interaction génotype x année pour les génotypes en opposition de phase par rapport au reste de la population

Les années dites OFF et ON sont déterminées par le comportement général de la population. Ainsi, sur la parcelle X3263 x Belrène par exemple, les individus alternants sont dans leur majorité OFF en 2010 et ON en 2011. Une interaction génotype x année 2010 faible révèle une faible production d'inflorescences lors de l'année 2010. Le génotype en question est donc probablement alternant, en phase avec la population.

Cependant, certains génotypes peuvent se retrouver en décalage avec la population. C'est le cas du génotype 195 qui présente un profil clairement alternant, mais avec une année ON en 2010. L'estimation de l'interaction génotype année 2010 est donc très élevée : 1,95.

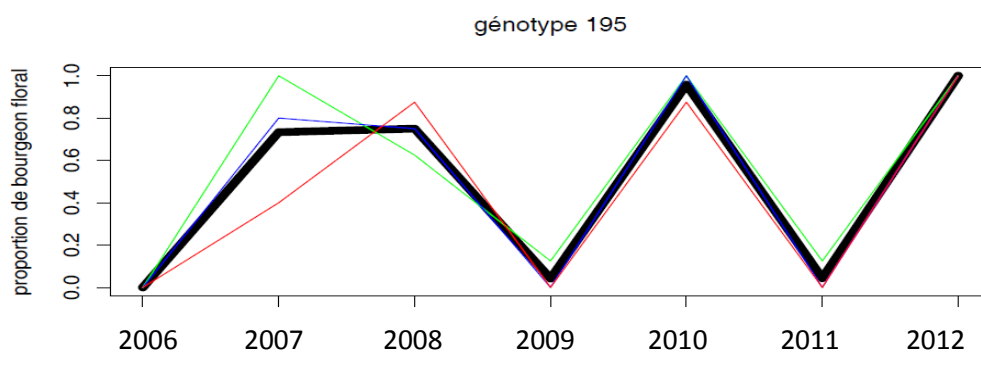


Figure 27 : profil de production du génotype 195 de la population X3263 x Belrène

L'utilisation de cet index doit donc être réalisée avec précaution. Il fonctionne dans la majorité des cas, mais appréhende mal le comportement des génotypes qui sont en opposition de phase avec le reste de la population.

3. Estimation de la corrélation entre échelle globale et échelle locale

Jusqu'à présent, la régularité de production a toujours été évaluée de manière distincte selon l'échelle à laquelle l'étude se produisait. Dans le cas de l'échelle globale, la quantité d'inflorescence par année était modélisée, donnant lieu au BBI et au coefficient d'auto corrélation (Guitton et al., 2011). Dans le cas de l'échelle locale, les probabilités de transitions donnaient naissance à des variables tels que les interactions génotypes*mémoires et génotype*années (Guitton et al., 2011).

Cette étude homogénéise la méthode d'estimation entre les 2 échelles. Les séquences relevées sur un échantillon de l'arbre sont utilisées pour estimer la quantité d'inflorescences chaque année. Ainsi, les mêmes index modélisant la quantité d'inflorescence par année peuvent être mis en place aux 2 échelles.

Les indices de régularité sont relativement bien corrélés entre les 2 échelles. Il devrait donc être possible d'estimer la régularité de production d'un génotype à partir d'un sous ensemble de ses branches.

Suite à la classification hiérarchique ascendante effectuée aux deux échelles, la majorité des génotypes se retrouvent classés dans le profil qui leur correspond : les irréguliers se retrouvent dans la majorité des cas dans les clusters irréguliers (irréguliers et alternants) et les réguliers sont principalement retrouvés dans les clusters réguliers synchronisés et désynchronisés.

Cependant, beaucoup d'erreurs ont lieu lors de la répartition des génotypes. Les indices de régularité étant corrélés, ces différences peuvent provenir de la méthode de clusterisation.

La classification n'accordant pas la même importance aux variables selon l'échelle, elle ne peut pas donner des résultats très proches. De même, on constate que les effectifs attribués à chacune des classes sont très différents dans les 2 cas. (Tableau 5).

Dans ces conditions il est peu probable d'obtenir une répartition homogène des génotypes selon si l'on considère l'arbre entier ou seulement un échantillon. La méthode utilisée pour séparer les génotypes n'est peut être pas la plus adéquate. Sachant quelles variables sont capables de

discriminer chaque caractéristique, une classification au cours de laquelle l'intérêt des variables serait contrôlé est à envisager.

Finalement, l'utilisation d'un échantillon de l'arbre s'avère suffisante dans le cas d'un schéma de sélection. En effet, les méthodes de sélection classiques consistent à rejeter les individus ne présentant clairement pas le trait attendu. L'utilisation des indices calculés à l'échelle locale ne classe jamais un génotype régulier à l'échelle globale parmi les alternants (figure 8). Ainsi, le génotype ne sera pas supprimé du schéma de sélection. L'échantillonnage n'amène jamais à supprimer des génotypes réguliers.

4. Recherche QTL

La détection de QTL est la première étape dans le développement d'outils permettant la sélection assistée par marqueurs. Elle ouvre aussi des perspectives concernant la connaissance du contrôle moléculaire de l'alternance puisqu'elle permettra par la suite d'obtenir des informations concernant des gènes candidats, des ARNs messagers et des facteurs de transcriptions contenus en lien avec cette région chromosomique.

La recherche de QTL s'est faite sur les indices développés dans ce rapport. L'obtention de plusieurs QTLs dont certains co-localisent entre populations montre à nouveau que ces indices sont bien des variables qualitatives adéquates pour l'estimation de l'alternance du pommier.

Le BBI résiduel normalisé est la seule variable pour laquelle des QTLs ont été détectés au sein des trois populations sans exception. Un seul de ces QTLs est commun à 2 populations, il est situé sur le groupe de liaison 10 et concerne les familles X3263 x Belrène et HIVW (LOD de 4,08 et 3,58 respectivement). Cependant, les QTLs concernant ce trait co-localisent parfois avec des QTLs qui concernent d'autres traits, à savoir l'interaction génotype x mémoire 01 pour Starkimson x Granny et le coefficient d'auto corrélation pour la population HIVW.

La quasi totalité des zones QTLs sont communes à plusieurs variables ou plusieurs population. Ceci apporte de la crédibilité à leur rôle concernant le déterminisme de l'alternance du pommier, ainsi que sur l'efficacité des index développés. Cependant, les LOD obtenus ainsi que les pourcentages de variances expliqués par ces variables sont relativement faibles. Le déterminisme génétique des paramètres de régularité du pommier semblent donc très polygénique. Il s'agirait d'un contrôle impliquant une multitude de locus à faibles influences.

Deux régions chromosomiques principales ressortent de cette étude comme ayant un rôle majeur dans le contrôle de l'alternance du pommier. Il s'agit des QTLs des groupes de liaison 1 et 10. Ces deux régions co-localisent non seulement chez plusieurs populations mais en plus pour plusieurs traits. De plus, elles ont déjà été identifiées comme QTL pour la texture de la pomme (Laurens F, Lahaye M, 2009).

Le QTL de la région 1 a de plus été détecté lors d'autres études (Guitton et al., 2011). Il co-localise avec des QTLs liés à la quantité d'inflorescence d'une année et à la quantité de fruits produits l'année d'avant. Ce QTL serait donc lié à une relation antagoniste entre production de fruit et développement d'inflorescences l'année suivante.

Le QTL du groupe de liaison 10 a lui aussi été détecté concernant l'évolution de la fermeté élasticité du fruit (Laurens F, Lahaye M, 2009).

Conclusion et Perspective

L'alternance est un caractère complexe pouvant être caractérisé par plusieurs descripteurs. Ce n'est donc pas un caractère quantifiable directement. Il nécessite la mise en place de plusieurs indicateurs permettant de transformer des données observées sur plusieurs années en une ou plusieurs variables représentant le génotype dans son ensemble.

Différents index permettant de caractériser la régularité d'un profil de production ont été proposés. Il a été démontré qu'ils possèdent la capacité de discriminer des génotypes selon leur tendance à alterner, et ce dès la sortie de la phase juvénile. De plus, ces indicateurs se sont révélés utilisables dans des conditions agronomiques diverses, et pour des arbres ayant des âges différents.

Ces indices peuvent donc servir de base à toutes études concernant la régularité de production.

En ce qui concerne la capacité d'un sous échantillon de l'arbre à le représenter, il a été démontré qu'une corrélation forte existe entre les 2 échelles. Cependant, des améliorations restent à faire concernant les méthodes de classification utilisées pour répartir les arbres dans chacun des profils d'alternance. D'autres méthodes devront donc être testées, notamment la classification supervisée.

L'utilisation des index a ensuite pu permettre une étude du déterminisme génétique de ce caractère. Certaines régions chromosomiques semblent ressortir fréquemment, malgré que sur les 5 populations étudiées, 2 ne présentaient pas des effectifs suffisant à une analyse QTL de qualité. Les 5 familles ayant des parents en commun, une prise en compte de la matrice d'apparentement devrait permettre d'affiner cette recherche de QTLs

BIBLIOGRAPHIE

- Chevreau E, Lespinasse Y, Gallet M. 1985.** Inheritance of pollen enzymes and polyploid origin of apple (*Malus × domestica* Borkh.). *Theoretical and Applied Genetics* **71**: 268-277.
- Cilas C, Montagnon C, Bar-Hen A. 2011.** Yield stability in clones of *Coffea canephora* in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Genetics & Genomes* **7**: 421-429.
- Coart E, Van Glabeke S, Loose MD, Larsen AS, Roldán-Ruiz I. 2006.** Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology* **15**: 2171-2182.
- Costes E, Lauri P-É, Regnard J-L. 2006.** Analyzing fruit tree architecture. Implication for tree management and fruit production. *Horticultural Reviews* **32**: 1-61.
- Evans RC, Campbell CS. 2002.** The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American Journal of Botany* **89**: 1478 -1484.
- FAO. 2011.** Food and Agriculture Organization of the United Nations. *FAO Statistics Division*. URL: <http://faostat.fao.org>. Page consulted the 18/08/2011.
- Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA. 2007.** Apple. Genome mapping and molecular breeding in plants. Berlin Heidelberg (New York, USA): Kole C, 1-62.
- Goldschmidt EE, Samach A. 2004.** Aspects of flowering in fruit trees. *Acta Horticulturae* **653**.
- Guillon B. 2011.** *Genetic control of biennial bearing in apple*. Montpellier, France: INRA (institut national de la recherche agronomique).
- Hackett WP. 1985.** Juvenility, maturation and rejuvenation in woody plants. *Horticultural Reviews* **7**: 109-115.
- Hoblyn TN, Grubb NH, Painter AC, Wates BL. 1936.** Studies in biennial bearing. *International Journal of Pomology and Horticulture Science* **14**: 39-76.
- Huet J. 1972.** Etude des effets des feuilles et des fruits sur l'induction florale des brachyblastes du Poirier. *Physiologie Végétale* **10**: 529-545.
- Janick J, Cummins JM, Brown SK, Hemmat M. 1996.** Apples. Fruit Breed, Volume I: Tree and Tropical Fruits. New York, USA: Janick J and Moore JN, 1-77.

- Jonkers H. 1979.** Biennial bearing in apple and pear: A literature survey. *Scientia Horticulturae* **11**: 303-307.
- Juniper BE, Mabberley DJ. 2006.** *The Story of the Apple*. Portland, Oregon, USA: Timber Press (OR).
- Korban S, Skirvin RM. 1984.** Nomenclature of the cultivate apple. *HortScience* **19**: 177-180.
- Laurens F. 1999.** Review of the current apple breeding programmes in the world: objectives for scion cultivar improvement. *Acta Horticulturae* **484**: 163-170.
- Lauri P-É, Térouanne E, Lespinasse J-M, Regnard J-L, Kelner J-J. 1995.** Genotypic differences in the axillary bud growth and fruiting pattern of apple fruiting branches over several years - an approach to regulation of fruit bearing. *Scientia Horticulturae* **64**: 265-281.
- Lespinasse Yves. 1992.** Le pommier. In: Gallais A. and Bannerot H. ed. *Amélioration des espèces végétales cultivées, objectifs et critères de sélection*. Paris, France: INRA, 579-594.
- Luckwill LC. 1970.** The control of growth and fruitfulness of apple trees. Acad. Press. In: *Physiology of Tree Crops*. London, UK: Luckwill LC, Cutting CV, 237-254.
- Monselise SP, Goldschmidt EE. 1982.** Alternate Bearing in Fruit Trees. *Horticultural Reviews* **4**: 128-173.
- Neilsen JC, Dennis FG. 2000.** Effects of seed number, fruit removal, bourse shoot length and crop density on flowering in 'Spencer Seedless' apple. *Acta Horticulturae* **527**: 137-146.
- Phipps JB, Robertson KR, Rohrer JR, Smith PG. 1991.** Origins and evolution of the subfamily Maloideae (*Rosaceae*). *Systematic Botany*: 303-332.
- Pratt C. 1988.** Apple flower and fruit: morphology and anatomy. *Horticultural Reviews* **10**: 273-308.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troglio M, Pruss D, et al. 2010.** The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature Genetics* **42**: 833-839.
- Wilcox J. 1944.** Some factors affecting apple yields in the Okanagan Valley. 1. tree size, tree vigor, biennial bearing, and distance of planting. *Scientific Agriculture* **25**: 189.

VII ANNEXES

Annexe 1 : Relation d'apparentement entre les populations

Annexe 2 : Script R : Calcul de l'entropie

Annexe 3 : Script R : Calcul des interactions génotypes x mémoire

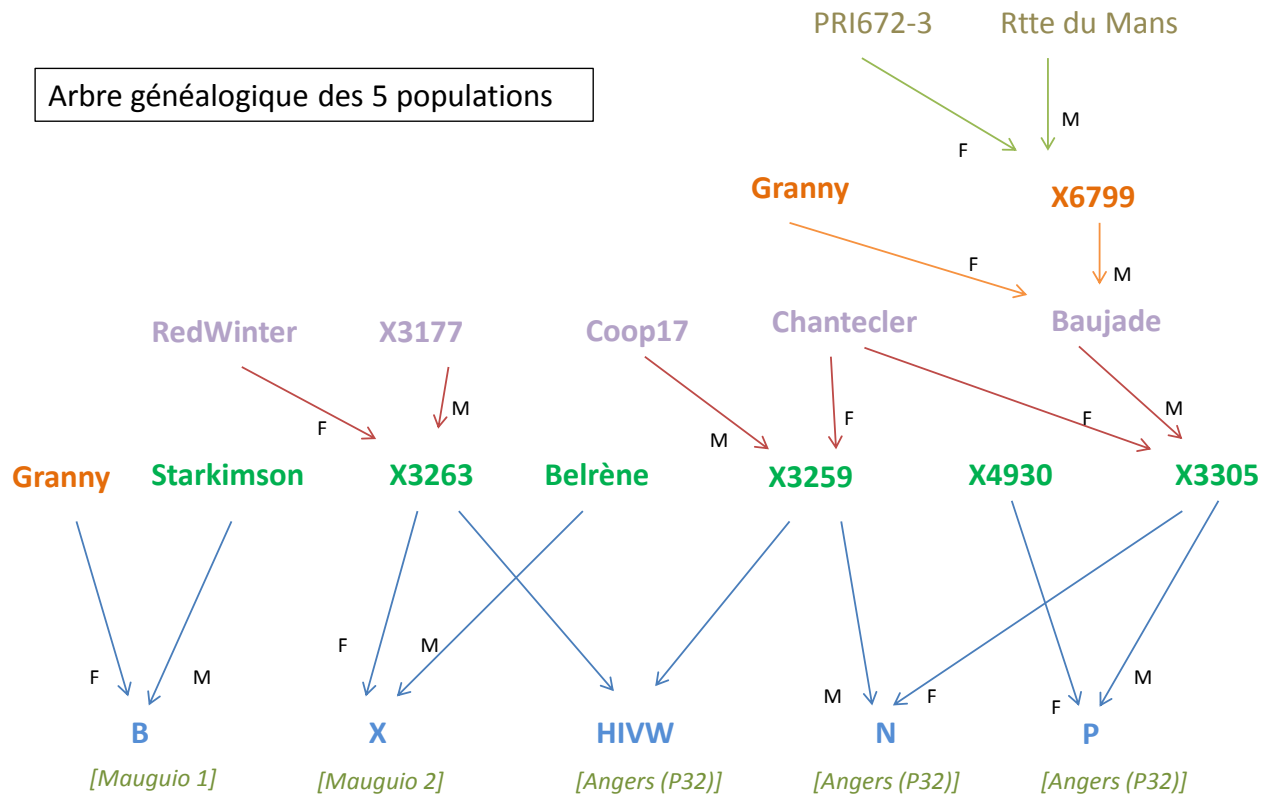
Annexe 4 : Script R : Calcul du BBI résiduel normalisé

Annexe 5 : Script R : calcul du coefficient d'auto corrélation

Annexe 6 : LOD seuil pour chaque variable et chaque groupe de liaison.

Annexe 1 : Relation d'apparentement entre les populations

Arbre généalogique des 5 populations



M = parent utilisé en temps que Mâle dans le croisement
 F = parent utilisé en temps que Femelle dans le croisement

Annexe 2 : Script R : Calcul de l'entropie

```
calcul.synchronisme.entropie.genotype.individus.moy = function(T)
{
  # Calcul du taux de synchronisme de floraison des PA
  # genotype par genotype et individu par individu
  dates = levels(T$AnneeFloraison)
  nb.dates = length(dates)
  genotypes = levels(factor(T$Genotype))
  nb.genotypes = length(genotypes)
  tab.dates = rep(0, nb.dates)
  freq.dates = rep(0, nb.dates)
  entropie.moy = rep(0, nb.genotypes)
  entropie.ind = NULL
  res = NULL
  for (g in 1:nb.genotypes) {
    T.i = T[T$Genotype==genotypes[g],]
    individus = levels(factor(T.i$Individu))
    nb.individus = length(individus)
    for (i in 1:nb.individus) {
      indiv = individus[i]
      if (sum((T.i$Genotype == genotypes[g]) & (T.i$Individu == indiv)) > 0) {
        e = 0
        for (t in 1:nb.dates) {
          selected = (T.i$AnneeFloraison == dates[t]) & (T.i$Genotype == genotypes[g]) &
(T.i$Individu == indiv)
          tab.dates[t] = sum(selected)
          if (tab.dates[t] > 0) {
            freq.dates[t] = mean(T.i$Floraison[selected], na.rm=TRUE)
            if ((freq.dates[t] > 0) & (freq.dates[t] < 1)){
              e = e - tab.dates[t] * (freq.dates[t] * log(freq.dates[t]) +
                (1-freq.dates[t]) * log((1-freq.dates[t]))))
            }
          }
        }
        e = e / sum(tab.dates)
        entropie.ind = rbind(entropie.ind, c(indiv, genotypes[g], e))
      }
    }
  }
  return(entropie.moy)
}
```

```

        entropie.moy[g] = entropie.moy[g] + e / nb.individus}}
res1 = data.frame(genotypes, entropie.moy)
colnames(res1) = c("Genotype", "Entropie")
res2 = data.frame(entropie.ind[,1], entropie.ind[,2], as.numeric(entropie.ind[,3]))
colnames(res2) = c("Individu", "Genotype", "Entropie")
res$moy.gen = res1
res$ind = res2
res
}

```

Annexe 3 : Script R : Calcul des interactions génotypes x mémoire

```
# effets fixes : Memoire, AnneeFloraion.
# interactions Memoire | genotype et AnneeFloraion | genotype (avec effet individuel)
modl.annefloraion.memoire.interaction.gen.ind = glmer(Floraion ~ Memoire + AnneeFloraion + (1
| Genotype:Memoire) + (1 | Genotype:AnneeFloraion) + (1 | Individu), data = mauguio2.d2, family
= binomial)
# AIC 26677 BIC 26766
modl = modl.annefloraion.memoire.interaction.gen.ind
# matrice pour caracteriser chaque genotype : effet du parent, effet de l'annee
l.m = levels(data2$Memoire)
l.a = levels(data2$AnneeFloraion)
r.gm = ranef(modl)$`Genotype:Memoire`
n.gm = row.names(r.gm)
r.ga = ranef(modl)$`Genotype:AnneeFloraion`
n.ga = row.names(r.ga)

n.g <- length(levels(factor(mauguio2.d2$Genotype)))

m.g <- matrix(0, n.g, 1)
m.g2 <- matrix(0, n.g, length(l.m) + length(l.a))

m.g[,1] <- levels(factor(mauguio2.d2$Genotype))

m.g <- data.frame(m.g,m.g2)

names(m.g) <- c("Genotype", l.m, l.a)
# g:m : m croit en 1er
# ajout effets aleatoires memoire
for (i in 1:length(n.gm)) {
  name.i <- n.gm[i] # nom de l'interaction
  split <- strsplit(name.i,":")
  genotype <- split[[1]][1]
  memoire <- split[[1]][2]
```

```

commande = paste("m.g[m.g$Genotype==genotype, ]$",memoire,"" = r.gm[i,1]", sep="")
eval(parse(text=commande))}

# ajout effets aleatoires AnneeFloraison
for (i in 1:length(n.ga)) {
  name.i <- n.ga[i] # nom de l'interaction
  split <- strsplit(name.i,":")
  genotype <- split[[1]][1]
  annee <- split[[1]][2]
  commande = paste("m.g[m.g$Genotype==genotype, ]$",annee,"" = r.ga[i,1]", sep="")
  eval(parse(text=commande))}

```


Annexe 4 : Script R : Calcul du BBI résiduel normalisé

```
calcul.BBI <- function(Vec, Norm.Vec = NULL) {  
  # calcul du BBI d'un vecteur constitue d'une sequence Vec  
  # et d'une sequence de normalisation Norm.Vec  
  n <- length(Vec)  
  if (is.null(Norm.Vec)) {  
    Norm <- abs(Vec[1:(n-1)]) + abs(Vec[2:n]) }  
  else {  
    Norm <- abs(Norm.Vec) }  
  Norm[Norm == 0] <- 1 # la valeur de Norm ne sera pas vraiment utilisee dans ce cas  
  # car la difference de Vec est nulle  
  BBI.res <- sum((abs(Vec[2:n] - Vec[1:(n-1)])) / Norm) / (n-1)  
  return(BBI.res) }  
  
tables.BBI.var <- function(BP = B) {  
  # calcul de differentes variantes du BBI :  
  # 1) BBI sur les residus normalises par la production sur 2 annees a chaque pas de temps  
  # 2) BBI sur les residus normalises par la production predite a chaque pas de temps  
  # 3) variance residuelle des genotypes  
  # normalisation des BBI en calculant le BBI par individu  
  tb <- NULL  
  for (g in 1:125) {  
    N <- BP[BP[,1]==g,] # donnees pour le genotype g  
    if (dim(N)[1] > 0) { # le genotype existe  
      r <- rep(0, 4) # valeurs du genotype + 3 indices  
      r2 <- rep(0, 2) # valeur de l'indice 1 pour les 2 replications  
      r3 <- rep(0, 2) # indice 2, etc.  
      r4 <- rep(0, 2)  
      r5 <- rep(0, 2)  
      r[1] <- g  
      t1 <- 0
```

```

t2 <- 0
if (N[1,2]==1000+g) { # 1ere replication existe
  t1 <- sum(N$R == 1000+g) # effectif replication 1
  # calcul du BBI sur les residus
  # normalisation par la production
  s1 <- N$Res[N$R==1000+g]
  n1 <- N$Y_inf[N$R==1000+g]
  n1 <- n1[2:t1]
  i1 <- n1==0
  s1[2:t1][i1] <- s1[1:(t1-1)][i1]
  n1[i1] <- 0
  r2[1] <- calcul.BBI(s1, n1)
  # normalisation par la prediction
  s2 <- N$Res[N$R==1000+g]
  n2 <- N$Ypred[N$R==1000+g]
  n2 <- n2[2:t1]
  i2 <- n2==0
  s2[2:t1][i2] <- s2[2:(t1-1)][i2]
  r3[1] <- calcul.BBI(s2, abs(n2))
  # ecart-type residuel
  r4[1] <- sqrt((t1-1)*var(N$Res)/t1)
  # coefficient de variation residuelle empirique
  r5[1] <- r4[1] / mean(N$Y_inf[N$R==1000+g])}
if (t1 != dim(N)[1]) { # 2eme replication existe
  t2 <- sum(N$R == 2000+g) # effectif replication 2
  # calcul du BBI sur les residus
  # normalisation par la production
  s1 <- N$Res[N$R==2000+g]
  n1 <- N$Y_inf[N$R==2000+g]
  n1 <- n1[2:t2]
  i1 <- n1==0
  s1[2:t2][i1] <- s1[1:(t2-1)][i1]
  n1[i1] <- 0
  r2[2] <- calcul.BBI(s1, n1)
  # normalisation par la prediction

```

```

s2 <- N$Res[N$R==2000+g]
n2 <- N$Ypred[N$R==2000+g]
n2 <- n2[2:t2]
i2 <- n2==0
s2[2:t2][i2] <- s2[2:(t2-1)][i2]
r3[2] <- calcul.BBI(s2, abs(n2))
# ecart-type residuel
r4[2] <- sqrt((t2-1)*var(N$Res)/t2)
# coefficient de variation residuelle empirique
r5[2] <- r4[2] / mean(N$Y_inf[N$R==2000+g])
# moyennes des indices ponderees par les effectifs
r[2] <- (((t1-1) * r2[1]) + ((t2-1) * r2[2])) / (max((t1 - 1), 0) + max((t2 - 1), 0))
r[3] <- (((t1-1) * r3[1]) + ((t2-1) * r3[2])) / (max((t1 - 1), 0) + max((t2 - 1), 0))
r[4] <- ((t1 * r4[1]) + (t2 * r4[2])) / (max(t1, 0) + max(t2, 0))
r[5] <- ((t1 * r5[1]) + (t2 * r5[2])) / (max(t1, 0) + max(t2, 0))
tb <- rbind(tb, r)}
return(tb)}

```

Annexe 5 : Script R : calcul du coefficient d'auto corrélation

```
# V) MODELE AUTOREGRESSIF SUR LES RESIDUS ET ESTIMATION #####

# V.1) effet fixe

# On travaille sur les residus. On veut estimer un modele autoregressif d'ordre 1
# Pour cela on va estimer un modele lineaire sur les residus en prenant comme variable explicative le
residu precedent
# On veut  $E_{grt} = a_g * E_{gr}(t-1) + u_{grt}$ 

#Calcule du nbr d'année de mesure
D <- cbind(D, D[,1], D[,1])
names(D)[9] <- "T.max"
names(D)[10] <- "T.min"

# calcul du nombre de residus pour chaque replication
for (i in gen.list) {
  D1 = D[D$R==1000+i,]
  D2 = D[D$R==2000+i,]
  if (dim(D1)[1] > 0) {
    D$T.min[D$R==1000+i] = min(D1$T)
    D$T.max[D$R==1000+i] = max(D1$T) }
  if (dim(D2)[1] > 0) {
    D$T.min[D$R==2000+i] = min(D2$T)
    D$T.max[D$R==2000+i] = max(D2$T)}}

# Préparation du tableau de données pour modèle AR1
Res1 = D[D$T!=D$T.min,]$Res      # Res1 est le vecteur colonne qui comprend les residus de l'annee
1 a Tmax pour tous les genotypes g
Res0 = D[D$T!=D$T.max,]$Res      # Res0 est le vecteur colonne qui comprend les residus de
l'annee 0 a Tmax-1
l<-length(Res1)
```

```
M<-matrix(0,l,n.g)          # M est la matrice des covariables; elle contient les vecteur
colonne [E_g10,...,E_g14,E_g20,...,E_g24],
                             # pour chaque genotype g, qui sont disposes en diagonale; avec des zeros
partout ailleurs.
```

```
IM <- A[D$T!=D$T.max,1:2]    # IM est la matrice des genotypes et replications
```

```
debut = 1
```

```
for (g in gen.list) {
  M.g = D[D$G==g,] # donnees genotype g
  if (dim(M.g)[1] > 0) {
    # suppression des dernieres observations
    # de chaque replication
    M.g = M.g[M.g$T != M.g$T.max,]
    fin = debut + dim(M.g)[1]
    M[debut:(fin-1),g] = Res0[debut:(fin-1)]
    debut = fin}}

```

```
DM <- data.frame(cbind(IM,M, Res0))
```

```
# bidouillage : je donne le nom "Ag" a chaque colonne de la matrice M
```

```
for (g in gen.list) {
  commande<-paste("A",toString(g),"<-M[,",toString(g),"]",sep="")
  eval(parse(text=commande))
  # deux premieres colonnes de DM : G et R
  commande<-paste("names(DM)[",toString(g+2),"]<-'A",toString(g),"",sep="")
  eval(parse(text=commande))}

```

```
#TABLEAU DE DONNEES POUR AR1
```

```
#edit(DM)          ## On estime comme un modele lineaire particulier (formes de Res1, M, Beta
particulieres) mais c'est bien un modele AR1 (Auto-Regressif d'ordre 1) !!
```

```
# Modèle 1
```

```

# AR1<-lm(Res1~-
1+A1+A2+A3+A4+A5+A6+A7+A8+A9+A10+A11+A12+A13+A14+A15+A16+A17+A18+A19+A20+A21+A
22+A23+A24+A25+A26+A27+A28+A29+A30+A31+A32+A34+A35+A36+A37+A38+A39+A40+A41+A42+
A43+A44+A45+A46+A47+A48+A49+A50+A51+A52+A53+A54+A55+A57+A58+A59+A60+A61+A62+A6
3+A64+A65+A66+A67+A68+A69+A70+A71+A73+A74+A75+A76+A77+A78+A79+A80+A81+A82+A83+
A84+A85+A86+A87+A88+A89+A90+A91+A92+A93+A94+A95+A96+A97+A98+A99+A100+A101+A102
+A103+A104+A105+A106+A107+A108+A110+A111+A112+A113+A114+A115+A116+A117+A118+A11
9+A120+A121+A122+A123+A124+A125,data=DM)
variante.res = FALSE
commande <- "AR1<-lm(Res1~-1+ "
for (i in 1:(length(gen.list)-1)) {
  commande <- paste(commande, "A", gen.list[i], "+", sep="")
commande <- paste(commande, "A", gen.list[i+1], ",data=DM)", sep="")
eval(parse(text=commande))

```

Annexe 6 : LOD seuil pour chaque variable et chaque groupe de liaison.

| variable | Starkimson x Granny | | X3263 x Belrène | | HIVW | |
|-----------------------------------|---------------------|-----------|-----------------|----------|-------|----------|
| | GL | LOD seuil | Group | Interval | Group | Interval |
| Interaction génotype x mémoire 01 | GW | 4,2 | 1 | 2,7 | GW | 4,3 |
| | LG1 | 2,6 | 2 | 2,8 | LG01 | 2,9 |
| | LG10 | 2,8 | 3 | 2,7 | LG02 | 3 |
| | LG11 | 2,7 | 4 | 2,5 | LG03 | 2,8 |
| | LG12 | 2,7 | 5 | 2,9 | LG04 | 2,5 |
| | LG13 | 2,8 | 6 | 2,6 | LG05 | 2,7 |
| | LG13 | 2,9 | 7 | 2,4 | LG06 | 2,9 |
| | LG14 | 2,6 | 8 | 2,4 | LG07 | 2,7 |
| | LG15 | 2,9 | 9 | 2,7 | LG08 | 2,8 |
| | LG17 | 2,7 | 10 | 2,9 | LG09 | 2,7 |
| | LG2 | 2,8 | 11 | 2,7 | LG10 | 2,7 |
| | LG3 | 2,6 | 12 | 2,7 | LG11 | 2,8 |
| | LG4 | 2,7 | 13 | 2,6 | LG12 | 2,7 |
| | LG5 | 2,7 | 14 | 2,5 | LG13 | 2,9 |
| | LG6 | 2,6 | 15 | 3 | LG14 | 2,7 |
| | LG7 | 2,6 | 16 | 2,7 | LG15 | 2,8 |
| | LG8 | 2,7 | 17 | 2,8 | LG16 | 2,7 |
| | LG9 | 2,6 | GW | 4,2 | LG17 | 2,9 |
| Interaction génotype x mémoire 00 | GW | 4,1 | 1 | 2,7 | GW | 4,2 |
| | LG1 | 2,6 | 2 | 2,8 | LG01 | 2,9 |
| | LG10 | 2,7 | 3 | 2,8 | LG02 | 3,1 |
| | LG11 | 2,6 | 4 | 2,6 | LG03 | 3 |
| | LG12 | 2,6 | 5 | 2,9 | LG04 | 2,6 |
| | LG13 | 2,8 | 6 | 2,7 | LG05 | 2,9 |
| | LG14 | 2,5 | 7 | 2,5 | LG06 | 2,8 |
| | LG15 | 2,9 | 8 | 2,5 | LG07 | 2,6 |
| | LG16 | 2,6 | 9 | 2,8 | LG08 | 2,9 |

| | | | | | | |
|-----------------------------------|------|-----|----|-----|------|-----|
| | LG17 | 2,6 | 10 | 2,9 | LG09 | 2,7 |
| | LG2 | 2,7 | 11 | 2,7 | LG10 | 2,7 |
| | LG3 | 2,7 | 12 | 2,5 | LG11 | 2,7 |
| | LG4 | 2,6 | 13 | 2,7 | LG12 | 2,8 |
| | LG5 | 2,6 | 14 | 2,4 | LG13 | 2,8 |
| | LG6 | 2,6 | 15 | 2,9 | LG14 | 2,7 |
| | LG7 | 2,6 | 16 | 2,6 | LG15 | 2,9 |
| | LG8 | 2,7 | 17 | 2,8 | LG16 | 2,7 |
| | LG9 | 2,6 | GW | 4,2 | LG17 | 2,9 |
| Interaction génotype x mémoire 11 | GW | 4 | 1 | 2,7 | GW | 4,2 |
| | LG1 | 2,6 | 2 | 2,8 | LG01 | 2,8 |
| | LG10 | 2,6 | 3 | 2,9 | LG02 | 2,9 |
| | LG11 | 2,6 | 4 | 2,5 | LG03 | 2,8 |
| | LG12 | 2,5 | 5 | 2,8 | LG04 | 2,5 |
| | LG13 | 2,8 | 6 | 2,6 | LG05 | 2,9 |
| | LG14 | 2,5 | 7 | 2,5 | LG06 | 2,9 |
| | LG15 | 2,8 | 8 | 2,5 | LG07 | 2,7 |
| | LG16 | 2,6 | 9 | 2,7 | LG08 | 2,8 |
| | LG17 | 2,5 | 11 | 2,6 | LG09 | 2,6 |
| | LG2 | 2,6 | 12 | 2,5 | LG10 | 2,8 |
| | LG3 | 2,6 | 13 | 2,7 | LG11 | 2,9 |
| | LG4 | 2,7 | 14 | 2,6 | LG12 | 2,8 |
| | LG5 | 2,5 | 15 | 2,8 | LG13 | 2,8 |
| | LG6 | 2,6 | 16 | 2,7 | LG14 | 2,6 |
| | LG7 | 2,6 | 10 | 2,8 | LG15 | 3 |
| | LG8 | 2,5 | GW | 4,1 | LG16 | 2,7 |
| | LG9 | 2,4 | 17 | 2,9 | LG17 | 2,8 |
| Entropie | GW | 4,2 | 1 | 2,7 | GW | 4,1 |
| | LG1 | 2,7 | 2 | 2,7 | LG01 | 3 |
| | LG10 | 2,8 | 3 | 2,9 | LG02 | 2,9 |
| | LG11 | 2,6 | 4 | 2,4 | LG03 | 2,9 |
| | LG12 | 2,5 | 5 | 2,8 | LG04 | 2,5 |
| | LG13 | 2,8 | 6 | 2,6 | LG05 | 2,8 |

| | | | | | | |
|--|------|-----|----|-----|------|-----|
| | LG14 | 2,6 | 7 | 2,6 | LG06 | 2,8 |
| | LG15 | 2,9 | 8 | 2,3 | LG07 | 2,6 |
| | LG16 | 2,7 | 9 | 2,7 | LG08 | 2,7 |
| | LG17 | 2,5 | 11 | 2,6 | LG09 | 2,6 |
| | LG2 | 2,7 | 12 | 2,7 | LG10 | 2,8 |
| | LG3 | 2,7 | 14 | 2,4 | LG11 | 2,8 |
| | LG5 | 2,6 | 15 | 2,8 | LG12 | 2,8 |
| | LG6 | 2,7 | 16 | 2,7 | LG13 | 2,8 |
| | LG7 | 2,7 | 17 | 2,7 | LG14 | 2,7 |
| | LG8 | 2,6 | GW | 4,1 | LG15 | 2,9 |
| | LG9 | 2,6 | 13 | 2,6 | LG16 | 2,7 |
| | LG4 | 2,7 | 10 | 2,9 | LG17 | 2,9 |

ABSTRACT

Biennial bearing is defined as the irregular crop load of a tree over consecutive years. The main assumption underlining biennial bearing is that the fruit load of a given year inhibits flower formation for the following year. This phenomenon generates major agronomic problems for fruit species such as apple, by reducing the production during “OFF” years and fruit quality during “ON” years, while increasing orchard management costs, especially for fruit thinning. A strategy to attenuate biennial bearing is to develop new varieties with regular production.

My study used 5 apple segregating populations issued from crosses between contrasting parents for architectural and flowering features. Phenotyping of the populations was achieved at local scale by observing the succession of floral and vegetative meristems in terminal position of shoots. The number of flowering occurrences over years was modeled by a mixed model. A new biennial bearing index was proposed, based on deviation around flowering occurrences trend over years. Auto-regression between residuals was investigated to represent the dependencies between consecutive years. Flowering probability was also modeled with a high-order Markov model with mixed effects. This provided different estimators of biennial bearing that were used to classify the genotypes and identify different alternation profiles.

Comparing the collected local scale data with data collected at the whole tree scale, it was shown that an estimation of the alternation profile of the genotypes can be done with an acceptable precision, with a sample of the trees only. That should permit to lighten the phenotyping protocol.

Finally, the descriptors were used for QTL detection and a meta-analysis was performed over the 5 populations. Even though 2 populations were too small for QTL detection, 8 genomic regions were involved in the biennial bearing variation. All of them were detected for several descriptors. Two of them co-localized for different populations. These results suggest a polygenic control of biennial bearing and opens new perspective for exploring the underlying genetic mechanisms.

Keywords : *Malus x domestica*, biennial bearing, modeling, genetic determinism.