



HAL
open science

Génomique fonctionnelle et marqueurs de qualité chez l'abricot

Jérôme Grimplet

► **To cite this version:**

Jérôme Grimplet. Génomique fonctionnelle et marqueurs de qualité chez l'abricot. Alimentation et Nutrition. Institut National Polytechnique (Toulouse), 2004. Français. NNT : 2004INPT015A . tel-04628295

HAL Id: tel-04628295

<https://ut3-toulouseinp.hal.science/tel-04628295>

Submitted on 28 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre :

THÈSE

présentée

pour obtenir

LE TITRE DE DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE

École doctorale : Sciences Ecologiques, Vétérinaires, Agronomiques & Bioingénieries

Spécialité : Qualité et Sécurité des Aliments

Par Mr Jérôme GRIMPLET

Génomique fonctionnelle et marqueurs de qualité chez l'abricot

Soutenue le 8 décembre 2004 devant le jury composé de :

Mr.	Guy ALBAGNAC, DR, INRA Montpellier	Président
MM.	Jean-Claude PECH, PR, ENSA Toulouse	Directeur de thèse
	Christophe PLOMION, DR, INRA Bordeaux	Rapporteur
	Andrea SCHUBERT, PR, Université Turin	Rapporteur
	Jean-Marc AUDERGON, IR, INRA Avignon	Membre
	Charles ROMIEU, CR, INRA Montpellier	Membre

Résumé

En lien avec la sélection assistée par marqueurs de la qualité des fruits chez l'abricotier (*Prunus armeniaca*), 71 marqueurs moléculaires impliqués dans la régulation hormonale et dans le contrôle de l'acidité, du taux de sucre, de la texture, de la biosynthèse des arômes, et des pigments, ont été mis en évidence par étude conjointe de l'évolution du transcriptome et du protéome au cours de la maturation et entre génotypes contrastés.

Une base de données a été construite contenant 5200 unigènes formés à partir de 18000 EST, fonctionnellement annotés et regroupés par familles multigéniques. Des lames de microarray ont été construites sur lesquelles 1800 unigènes pertinents ont été déposés sous forme d'oligonucléotides 50 mères définis dans la région 3' non codante des transcrits. Celles-ci ont permis de mener l'étude de l'expression des gènes au cours du développement de la variété Bergeron et au travers différentes variétés contrastées pour la vitesse de maturation, la production d'éthylène et la couleur. L'évolution de l'abondance des protéines totales de Bergeron a été étudiée au cours de son développement par électrophorèse bidimensionnelle. Pour faciliter l'identification des spots protéiques, une base de données de protéines végétales chimériques incluant les mutations spécifiques du genre *Prunus* au sein des séquences complètes de leurs meilleurs orthologues à été construite à partir des données EST.

Abstract

In order to improve apricot (*Prunus armeniaca*) quality with markers assisted selection, 71 molecular markers involved in hormonal regulation and control of acidity and sugar content, texture, flavours and pigments biosynthesis were highlighted by transcriptomics and proteomics comparison of ripening stages and contrasted genotypes.

Thus a database, containing 5200 unigenes extracted from 18000 functionally annotated and multigenic families gathered EST, was designed. Microarrays slides were built, and 1800 relevant unigenes were spotted as specific 50mers oligonucleotides designed in the 3' untranslated region of transcripts. Gene expression study was performed on Bergeron cultivar development and between contrasted cultivars for ethylene outburst and color. Total proteins abundance during Bergeron development was analyzed on two-dimensional gel electrophoresis. For improving identification, a plant proteins database including mutations from conceptually translated expressed sequence tags from *Prunus* into best matches, was designed

Mots clés : Abricot, Transcriptomique, EST, Protéomique, Bioinformatique, Génomique fonctionnelle haut débit.

Remerciements

Ce travail a été réalisé à l'Institut National de la Recherche Agronomique, dans l'équipe Biologie Intégrative de la Vigne et du Raisin dirigé par Monsieur Albagnac. Je tiens à le remercier pour m'avoir accueilli.

Je tiens à remercier Messieurs Christophe Plomion et Andrea Schubert pour avoir accepté de juger ce travail et de m'avoir fait l'honneur de participer au jury en tant que rapporteurs.

Je remercie également Messieurs Jean-Claude Pech, Jean-Marc Audergon, Guy Albagnac et Charles Romieu de m'avoir fait l'honneur d'examiner ce travail.

Je remercie Nancy Terrier et Charles Romieu de m'avoir encadré pendant ces trois années, pour tout ce qu'ils m'ont apporté et pour leur amitié.

Je tiens à remercier l'ensemble des personnes qui ont collaboré à ce travail:

- A l'UR GAFL, Jean-Marc Audergon, Patrick Lambert et Jean-Paul Bouchet.
- A l'UMR SQPOV, Sylvie Bureau, Maggy Grotte, Maryse Reich et Barbara Gouble, pour les mesures physiques ; Isabelle Marty et Line Tichit pour les PCR Quantitatives.
- Les personnes qui m'ont apporté leur aide précieuse en protéomique, François Xavier Sauvage, François Chevalier, Nicolas Sommerer, Emmanuelle Demey, José Walter Gaspar et Anne-Laure Gancel.

Je remercie Didier Mbéguié-Mbéguié et Bernard Fils-Lycaon qui ont amorcé ce travail et Christophe Rothan qui a accepté de participer aux différents comités de thèses

Je remercie les départements CEPIA et GAP d'avoir accepté de financer ces travaux, et plus particulièrement Anne Marie Chèvre pour son encadrement auprès des thésards.

Un grand merci à tout ceux qui ont accepté de relire ce manuscrit, Nancy, Charles, Agnès, FX, Sylvie, Barbara, Maggie, Jean-Marc, Catherine, Sébastien et particulièrement Martine.

J'adresse mes sincères remerciements à tous les membres de l'équipe BIVR, Agnès pour m'avoir supporté dans son bureau pendant trois ans, Martine, Clotilde, Sandrine, Lucie, Catherine, Eliette, Jean-Luc, François, Philippe, José, et Didier, ainsi qu'à tout le personnel de l'UMR SPO. Merci aussi à l'équipe FT de m'avoir accueilli dans leurs locaux pendant la rédaction.

Je tiens à remercier Jean Emmanuel et Gil pour m'avoir donné envie de faire ce métier, ainsi que Guy Albagnac pour m'avoir fait confiance à la fin de mon DEA.

Je remercie Patrick pour ses dépannages en informatique et pour m'avoir hébergé, au début de ma thèse. Coucou à toute la famille.

Je remercie tout ceux qui prendront le courage de lire ce manuscrit, j'espère qu'il sera souvent en haut de la pile.

Ma dernière pensée va à mes parents, mon frère, et aux potes, de Montpellier des Ardennes et d'ailleurs.

PUBLICATIONS EN PREMIER AUTEUR

GRIMPLET J., GASPAR J.W., GANCEL A.L., SAUVAGE F.X., ROMIEU C. Including mutations from conceptually translated expressed sequence tags into orthologous proteins improves the preliminary assignment of peptide mass fingerprints on non model genomes. *Proteomics* (Accepté)

GRIMPLET J., ROMIEU C., SAUVAGE F.X., LAMBERT P., AUDERGON J.M., TERRIER N. (2004). Transcriptomics and Proteomics Tools towards Ripening Markers for Assisted Selection in Apricot. *Acta hort.* (Sous presse)

GRIMPLET J., ROMIEU C., AUDERGON J-M., ALBAGNAC G., LAMBERT P., BOUCHET J-P., TERRIER N. (2003). High throughput detection of isogenes among 13000 EST from Apricot fruit (*Prunus armeniaca*). Soumission prévue fin 2004.

Publications prévues (titres non définitifs)

Grimplet et al. : Etude du développement de l'abricot au cours de sa maturation par approche protéomique

Grimplet et al. : Isolation de marqueurs de qualité de l'abricot par l'analyse de l'expression des gènes par microarray

COMMUNICATIONS ORALES DANS DES CONGRÈS NATIONAUX OU INTERNATIONAUX SUR MES TRAVAUX DE THÈSE

GRIMPLET J., ROMIEU C., SAUVAGE F.X., TERRIER N., LAMBERT P., AUDERGON J.M., DIRLEWANGER E., COSSON P., LE DANTEC L., MOING A. (2004). EST in *Prunus* genus (Apricot and Peach). International Rosacea Genome Mapping Conference. Clemson University 22-24 May.

GRIMPLET J., ROMIEU C., SAUVAGE F.X., LAMBERT P., AUDERGON J.M. TERRIER N. (2003). Transcriptomics and Proteomics Tools towards Ripening Markers for Assisted Selection in Apricot. Eucarpia Symposium on Fruit Breeding and Genetics. Angers 1-5 september.

GRIMPLET J., ROMIEU C., TERRIER N., BUREAU S., GROTTÉ M., GOUBLE B., MARTY I., REICH M., AUDERGON J-M., BOUCHET J-P., LAMBERT P. (2002) Bioanalyse de 7000 ESTs d'Abricot. VI^{ème} rencontre du groupe Biologie Moléculaire des Ligneux. Avignon 22-24 octobre.

COMMUNICATIONS AU COURS DE REUNIONS DE GROUPES DE TRAVAIL

GRIMPLET J. Génomique fonctionnelle de la maturation de l'abricot. Session « Analyse du transcriptome », dans le cadre de l'action transversale LIGNOME. Nancy 29-31 janvier 2002.

GRIMPLET J. Intégration des données d'expression - Paramétrage de Blast pour détecter les isogènes - Utilisation de CAP3 et phrap. Réunion de Bio-Informatique du programme Lignome. Bordeaux 27-28 Juin 2002.

GRIMPLET J. Identification de gènes et de protéines potentiellement impliqués dans les critères de qualité chez l'abricot. Rencontres des Doctorants en Bioinformatique. Ivry sur Seine 22-23 Septembre 2003.

GRIMPLET J. Identification de marqueurs de qualité par approche de génomique fonctionnelle haut débit chez l'abricot. Journées Jeunes Chercheurs du Département de Génétique et Amélioration des Plantes. Montpellier 22-23 Avril 2004.

PUBLICATIONS EN COLLABORATION

TERRIER N., GLISSANT D., GRIMPLET J., BARRIEU F., ABBAL P., COUTURE C., AGEORGES A., ATANASSOVA R., LÉON C., RENAUDIN J.P., DEDALDECHAMP F., ROMIEU C., DELROT S., AND HAMDI S. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera L.*) development. *Planta* (soumis).

SARRY J.-E., GRIMPLET J., VALLIER M. J., MONDOLOT-COSSON L., ANDARY C., GÜNATA Z., & ROMIEU C. (2003). Purification and characterization of cell-wall glycoside hydrolase from grapevine (*Vitis vinifera L.*) berry skins. *Phytochemistry* (en révision).

COMMUNICATION ORALES SUR DES TRAVAUX EN COLLABORATION

TERRIER N., GLISSANT D., GRIMPLET J., BARRIEU F., ABBAL P., COUTURE C., AGEORGES A., ATANASSOVA R., LÉON C., RENAUDIN J.P., DEDALDECHAMP F., DELROT S., HAMDI S. AND ROMIEU C. (2004) Isogene Specific Oligo Arrays Reveal Multifaceted Changes in Gene Expression During Grape Berry (*Vitis vinifera L.*) Development. 7th International symposium on Grapevine Physiology and Biotechnology, Davis, June 21-25 2004

GASPAR J.W., GRIMPLET J., SAUVAGE F.X., DUCAMP M.N., ROMIEU C. (2004) Apport de la protéomique à l'étude du développement de la mangue (*Mangifera indica*). VII ème rencontre du groupe Biologie Moléculaire des ligneux. Clermont Ferrand 11-13 mai

COUTURE C., TERRIER N., GLISSANT D., ABBAL P., AGEORGES A., ATANASSOVA R., DEDALDECHAMP F., GRIMPLET J., BARRIEU F., DELROT S., ROMIEU C., HAMDI S. (2003) Development and maturation of grape berries : transcriptome analysis for high throughput studies. 7th International Symposium of Enology of Bordeaux June 19, 20 and 21

TERRIER N., COUTURE C., GLISSANT D., ATANASSOVA R., ABBAL P., AGEORGES A., DEDALDECHAMP F., GRIMPLET J., DELROT S., HAMDI S., ROMIEU C., BARRIEU F. (2003) Transcriptome Analysis of Developing Grape Berries. Séminaire 2003 Génoplatte, Poitiers, 18-20 mars 2003 (Poster, oral communication)

ABBAL P., AGEORGES A., ALBAGNAC G., ATANASSOVA R., BARRIEU F., COUTURE C., DEDALDECHAMPS F., DELROT S., GRIMPLET J., HAMDI S., ROMIEU C., TERRIER N. (2003) Isogene specific probes for high throughput studies on grape berry development. Plant & Animal Genomes XI Conference. San Diego 11-15 january.

Table des matières

I INTRODUCTION.....	16
II ETUDE BIBLIOGRAPHIQUE.....	19
II.1 Caractéristiques physiologiques de l'abricotier	20
II.1.1 Conditions climatiques de culture.....	20
II.1.2 L'arbre.....	21
II.1.3 Anatomie du fruit.....	21
II.1.4 Composition et valeur nutritionnelle du fruit mûr.....	22
II.1.5 Le développement du fruit.....	22
II.2 Caractéristiques géographiques et génétiques de l'abricotier.....	23
II.2.1 Dispersion de l'abricotier.....	23
II.2.2 Diversité génétique.....	24
II.3 La maturation des fruits.....	25
II.4 Les critères de qualité liés à la maturation.....	26
II.4.1 La perte de fermeté.....	26
II.4.1.1 Biochimie de la paroi cellulaire.....	28
II.4.1.1.1 Polygalacturonases.....	28
II.4.1.1.2 Pectine méthylestérases.....	29
II.4.1.1.3 Pectate lyases.....	29
II.4.1.1.4 Xyloglucan-endotransglycosylases.....	30
II.4.1.1.5 Expansines.....	30
II.4.1.1.6 β -glucanases.....	30
II.4.2 Couleur.....	31
II.4.2.1 Dégradation des chlorophylles.....	31
II.4.2.2 Biosynthèse des caroténoïdes.....	31
II.4.2.3 Dégradation des caroténoïdes et biosynthèse de l'acide abscissique (ABA).....	33
II.5 Arômes.....	34
II.6 Biosynthèse des Hormones.....	35
II.6.1 Ethylène.....	35
II.6.2 Auxines.....	35
II.6.3 Gibbérellines.....	36
II.6.4 Méthyl jasmonate.....	36
II.7 Régulation hormonale au cours du développement du fruit.....	37
II.7.1 Voie de perception et signalisation de l'éthylène chez Arabidopsis.....	37
II.7.2 Voie de signalisation de l'éthylène chez les fruits.....	40
II.7.3 Effet de l'éthylène sur les critères de qualité des fruits.....	41
II.7.4 Effet de l'ABA sur les critères de qualité des fruits.....	42
II.7.5 Effet de l'auxine sur les critères de qualité des fruits.....	43
II.7.6 Effet de la gibbérelline sur les critères de qualité des fruits.....	43
II.7.7 Effet du jasmonate sur les critères de qualité des fruits.....	44
II.8 Outils de génomique fonctionnelle.....	44
II.8.1 Les EST (Expressed Sequence Tag).....	44
II.8.2 Techniques d'étude de l'évolution du transcriptome.....	46

II.8.2.1 Techniques courantes et utilisation chez les fruits.....	46
II.8.2.2 Les microarrays.....	47
II.8.3 Protéomique.....	48
III Matériel & Méthodes.....	49
III.1 Matériel végétal.....	50
III.1.1 Récolte et échantillonnage des fruits.....	50
III.1.1.1 Fruits utilisés pour la construction des banques ADNc.....	50
III.1.1.2 Fruits utilisés pour les expérimentations microarray, protéomique et PCR quantitative.....	50
III.1.2 Mesures des paramètres physico chimiques et physiologiques des fruits.....	51
III.1.2.1 Mesure de la fermeté.....	51
III.1.2.2 Mesure de la couleur.....	51
III.1.2.3 Mesure du dégagement éthylénique.....	52
III.1.2.4 Analyses biochimiques.....	52
III.1.2.4.1 Indice réfractométrique et acidité titrable.....	52
III.1.2.4.2 Dosage des sucres et des anions organiques.....	52
III.1.2.4.3 Dosage des pigments.....	52
III.1.2.4.4 Analyse des arômes.....	53
III.2 Méthodes de biochimie relatives à la protéomique.....	53
III.2.1 Électrophorèse bidimensionnelle.....	53
III.2.1.1 Préparation des extraits de protéines totales.....	54
III.2.1.2 Première dimension IEF.....	54
III.2.1.3 Deuxième dimension.....	55
III.2.1.4 Analyse d'image.....	56
III.2.2 Spectrométrie de masse.....	56
III.2.2.1 Préparation des spots protéiques.....	57
III.2.2.1.1 Excision et lavage des spots.....	57
III.2.2.1.2 Digestion trypsique.....	57
III.2.2.1.3 Extraction des peptides.....	58
III.2.2.1.4 Elimination des sels.....	58
III.2.2.2 Analyse par spectrométrie de masse MALDI-TOF (MS) et MALDI-TOF-TOF (MS/MS).....	58
III.2.2.3 Identification des peptides.....	58
III.3 Méthodes de biologie moléculaire.....	59
III.3.1 Préparation des banques d'ADNc	59
III.3.2 Production des séquences d'EST.....	60
III.3.3 Extraction des ARN.....	60
III.3.4 Préparation d'ADN plasmidique.....	61
III.3.5 Test d'hybridation/ southern blot.....	61
III.3.6 Synthèse du premier brin d'ADNc pour PCR quantitative.....	62
III.3.7 PCR quantitative.....	62
III.3.8 Synthèse des ARN antisens pour hybridation sur microarray.....	63
III.3.9 Hybridation sur lame de microarray, lavages et acquisition de l'image.....	65
III.3.9.1 Hybridation sur lame de microarray et lavages.....	65
III.3.9.2 Acquisition des images.....	65
III.4 Méthodes de bioinformatique.....	66

III.4.1	Édition des séquences.....	66
III.4.1.1	Base calling.....	66
III.4.1.2	Élimination du vecteur et des séquences courtes.....	66
III.4.1.3	Concaténation des séquences 3' et 5' issues du même clone.....	66
III.4.1.4	Assemblage des séquences redondantes.....	67
III.4.1.5	Calcul électronique du niveau d'expression des gènes (Northern électronique4)....	67
III.4.1.6	Détection des isogènes et des chimères.....	68
III.4.1.7	Annotation fonctionnelle.....	68
III.4.1.7.1	Procédure automatique.....	68
III.4.1.7.2	Procédure manuelle.....	69
III.4.1.8	Calcul du non codant.....	69
III.4.2	Construction de la base de données.....	70
III.4.3	Conception des microarrays. Design des oligonucléotides.....	70
III.4.4	Protocole expérimental d'hybridation.....	71
III.4.5	Analyse de l'image et traitement des données.....	72
III.4.5.1	Normalisation des données.....	72
III.4.5.2	Analyse statistique des données.....	73
III.4.5.2.1	ANOVA.....	73
III.4.5.2.2	Profil d'expression.....	74
III.4.6	Support bioinformatique à l'analyse protéomique.....	74
III.4.7	Lexique des termes de bioinformatique.....	75
IV	Résultats.....	77
IV.1	Caractérisation physiologique des fruits.....	78
IV.1.1	Émission d'éthylène.....	78
IV.1.2	Sucres.....	79
IV.1.3	Poids et fermeté.....	79
IV.1.4	Couleur.....	80
IV.1.5	Acidité.....	81
IV.1.6	Profils aromatiques.....	82
IV.2	Obtention et traitement des EST.....	83
IV.2.1	Caractérisation des banques d'ADNc.....	83
IV.2.2	Étude comparative des stratégies de séquençage en 3' et en 5'.....	84
IV.2.2.1	Critères de séquençage.....	84
IV.2.2.1.1	Nombre de séquences par boîte.....	84
IV.2.2.1.2	Taille moyenne des séquences brutes.....	85
IV.2.2.1.3	Comparaison de la taille des séquences avec les données de la bibliographie.....	85
IV.2.2.1.4	Répartition de la taille des séquences.....	86
IV.2.2.1.5	Taille des séquences supérieures à 250 pb.....	87
IV.2.2.1.6	Séquences exploitables.....	87
IV.2.2.2	Caractéristiques des séquences.....	87
IV.2.2.2.1	Partie non codante des séquences identifiées en 3'.....	87
IV.2.2.2.2	Séquences supérieures à 250 pb identifiées.....	88
IV.2.2.2.3	Séquences identifiées indépendamment de leur longueur.....	88
IV.2.2.2.4	Score des séquences identifiées.....	89
IV.2.2.2.5	Redondance : nombre de clones redondants.....	89
IV.2.2.2.6	Conclusion.....	90
IV.2.3	Traitement des EST.....	91

IV.2.3.1	Production du jeux d'unigènes.....	91
IV.2.3.1.1	Séquençage.....	92
IV.2.3.1.2	Base calling.....	93
IV.2.3.1.3	Élimination du vecteur et des séquences courtes.....	93
IV.2.3.1.4	Concaténation des extrémités 3' et 5'.....	94
IV.2.3.1.5	Assemblage.....	95
IV.2.3.2	Informations relatives aux EST.....	96
IV.2.3.2.1	Analyses de la redondance des EST.....	96
IV.2.3.2.2	Redondance intra et inter-banque.....	97
IV.2.3.2.3	Homologie par rapport aux bases de données publiques.....	98
IV.2.3.2.4	Distribution des unigènes en catégories fonctionnelles.....	98
IV.2.3.2.5	Analyse des familles multigéniques.....	101
IV.2.3.2.6	Test d'hybridation.....	103
IV.2.3.2.7	Estimation de l'expression des gènes pendant le développement.....	104
IV.2.3.2.8	Ajustement des données de 1996 et de 2002 pour les PCR quantitatives.....	107
IV.2.3.2.9	Validation de l'expression des gènes par PCR quantitative.....	107
IV.3	Etude du transcriptome par microarray.....	108
IV.3.1	Ajustement des stades de développement entre les différentes variétés.....	108
IV.3.2	Conception des lames et analyse statistique des résultats.....	109
IV.3.2.1	Conception des lames.....	109
IV.3.2.1.1	Design des oligonucléotides.....	109
IV.3.2.1.2	Analyse de l'hybridation des lames.....	111
IV.3.2.2	Analyse statistique de l'expression des gènes.....	112
IV.3.2.2.1	ACP.....	112
IV.3.2.2.2	ANOVA globale.....	113
IV.3.2.2.3	ANOVA locale.....	114
IV.3.3	Évolution des niveaux d'expression des transcrits par microarray.....	116
IV.3.3.1	Gènes régulés au cours du développement.....	116
IV.3.3.1.1	Catégorie fonctionnelle des gènes différentiellement exprimés pendant le développement.....	116
IV.3.3.1.2	Profils d'expression globaux.....	118
IV.3.3.1.3	Cluster 1.....	119
IV.3.3.1.4	Cluster 2.....	120
IV.3.3.1.5	Cluster 3.....	121
IV.3.3.1.6	Cluster 4.....	122
IV.3.3.1.7	Cluster 5.....	122
IV.3.3.1.8	Cluster 6.....	123
IV.3.3.2	Gènes différentiellement exprimés entre les variétés.....	125
IV.3.3.2.1	Cluster 1.....	126
IV.3.3.2.2	Cluster 2.....	127
IV.3.3.2.3	Cluster 3.....	128
IV.3.3.2.4	Cluster 4.....	128
IV.3.3.2.5	Cluster 5.....	129
IV.3.3.2.6	Cluster 6.....	129
IV.3.3.2.7	Cluster 7.....	130
IV.3.3.2.8	Cluster 8.....	131
IV.3.3.2.9	Cluster 9.....	131
IV.3.3.2.10	Cluster 10.....	132
IV.3.3.2.11	Cluster 11.....	132
IV.3.3.2.12	Cluster 12.....	133
IV.3.3.2.13	Cluster 13.....	133
IV.3.3.2.14	Cluster 14.....	134

IV.3.3.2.15 Cluster 15.....	134
IV.3.3.2.16 Cluster 16.....	135
IV.4 Aide à l'identification des protéines.....	136
IV.4.1 Résumé.....	137
IV.4.2 Introduction.....	138
IV.4.3 Matériel et Méthodes.....	139
IV.4.3.1 Matériel biologique et production des empreintes massiques peptidiques.....	139
IV.4.3.2 Assemblage des EST.....	139
IV.4.3.3 Construction de la base de données chimériques.....	140
IV.4.3.4 Identification des PMF.....	141
IV.4.4 Résultats et Discussion.....	141
IV.4.4.1 Conception de la base de données de chimères entre EST et protéines orthologues... 141	
IV.4.4.2 Analyse par MALDI-TOF des spots issus des gels 2D.....	142
IV.4.4.3 Illustration et validation des peptides prédis par MS/MS.....	145
IV.4.4.4 Valeur ajoutée de la procédure par rapport aux approches EST et inter-espèces. .	148
IV.4.5 Conclusion.....	150
IV.5 Étude du protéome.....	152
IV.5.1 Introduction.....	152
IV.5.2 Mise au point du protocole d'extraction des protéines totales d'abricot.....	152
IV.5.2.1 Evaluation des protocoles existants.....	152
IV.5.2.2 Mise au point du protocole.....	153
IV.5.3 Analyse des gels, estimation de la reproductibilité et de l'abondance des protéines.....	154
IV.5.4 Les protéines identifiées.....	158
IV.5.4.1 Les protéines impliquées dans la photosynthèse.....	160
IV.5.4.2 Protéines impliquées dans le métabolisme des sucres.....	161
IV.5.4.3 Protéines impliquées dans le grandissement cellulaire et la perte de fermeté.....	162
IV.5.4.4 Protéines impliquées dans le métabolisme et la réponse aux hormones.....	162
IV.5.4.5 Protéines impliquées dans le métabolisme de l'azote	163
IV.5.4.6 Protéines impliquées dans la défense vis à vis du stress oxydatif.....	164
IV.5.4.7 Protéines diverses.....	164
IV.5.4.8 Protéines non identifiées.....	166
V Discussion.....	167
V.1 Éthylène.....	168
V.1.1 Biosynthèse de l'éthylène.....	168
V.1.1.1 Amino-cyclopropane-1-carboxylate synthase (ACS).....	168
V.1.1.2 Amino-cyclopropane-1-carboxylate oxydase (ACO).....	169
V.1.1.3 La protéine E8.....	169
V.1.1.4 Autres membres de la famille multigénique des dioxygénases.....	169
V.1.1.5 S-adenosylméthionine synthétase (SAMS).....	170
V.1.1.6 Protéine kinases calcium dépendantes	171
V.1.2 Transduction du signal éthylène.....	171
V.1.2.1 Récepteurs d'éthylène.....	171
V.1.2.2 Facteurs de réponse à l'éthylène (ERF).....	172

V.1.3	Autres gènes potentiellement reliés au signal éthylène.....	176
V.1.4	Conclusion.....	177
V.2	Métabolisme, régulation et perception des autres hormones.....	178
V.2.1	Auxine	178
V.2.1.1	AUX/IAA.....	178
V.2.1.2	Facteurs de réponse à l'auxine.....	180
V.2.1.3	Autres gènes régulés par l'auxine.....	181
V.2.1.4	Conclusion.....	181
V.2.2	Cytokines.....	182
V.2.3	Méthyl jasmonate.....	182
V.2.3.1	Biosynthèse	182
V.2.3.2	Transduction du signal jasmonate.....	184
V.2.4	Acide gibbérellique	184
V.2.4.1	Biosynthèse.....	184
V.2.4.2	Régulation du signal GA.....	185
V.3	Biosynthèse des caroténoïdes et de l'acide abscissique.....	186
V.3.1	Biosynthèse des caroténoïdes.....	186
V.3.2	Dégradation des caroténoïdes / biosynthèse de l'ABA.....	188
V.3.3	Gènes potentiellement régulés par l'ABA.....	191
V.4	Métabolisme secondaire (hors caroténoïdes).....	192
V.4.1	Flavonoïdes.....	192
V.4.1.1	Métabolisme général des flavonoïdes.....	192
V.4.1.2	Biosynthèse des anthocyanes.....	194
V.4.2	Arômes.....	195
V.5	La perte de fermeté.....	196
V.5.1	Dépolymérisation des pectines.....	197
V.5.1.1	Pectine méthylestérase (PME).....	197
V.5.1.2	Pectate lyase.....	198
V.5.1.3	Polygalacturonase.....	198
V.5.2	Dépolymérisation du réseau d'hémicellulose.....	199
V.5.2.1	Xylogucan-endotransglycosylases (XET).....	199
V.5.2.2	Expansine.....	200
V.5.2.3	Endo β 1,4 glucanases.....	201
V.5.3	Grandissement cellulaire.....	201
V.5.4	Conclusion sur la perte de fermeté.....	202
V.6	Métabolisme du carbone.....	203
V.6.1	Photosynthèse.....	203
V.6.2	Métabolisme du saccharose.....	205
V.6.2.1	Saccharose synthase.....	206
V.6.2.2	UDP-Glucose pyrophosphorylase.....	207
V.6.2.3	Xylulose kinase.....	208
V.6.2.4	Autres gènes.....	208
V.6.3	Métabolisme du sorbitol.....	209
V.6.4	Métabolisme des acides, transport des ions.....	210

V.7 Autres gènes/protéines.....	211
V.7.1 Protéines identifiées.....	211
V.7.1.1 Antioxydation.....	211
V.7.1.2 Métabolisme de l'azote	212
V.7.2 Unigènes correspondant à des protéines inconnues ou non identifiées	213
VI Bilan, Conclusions et Perspectives.....	214
VI.1 Bilan sur les outils mis en place.....	215
VI.2 Bilan sur les candidats mis en évidence.....	217
VI.2.1 Dosage de composés.....	217
VI.2.2 Mesure de l'expression des gènes dans des variétés de phénotypes extrêmes.....	219
VI.2.3 Mesure de l'expression des gènes après récolte ou après application d'hormones exogènes.....	220
VI.2.4 Détermination de la fonction des protéines inconnues.....	220
VI.3 Poursuite du travail sur les « candidats qualité ».....	222
VII Références Bibliographiques.....	224
VIII Annexe.....	247
VIII.1 Annexe A.....	248
VIII.2 Annexe B.....	249
VIII.3 Annexe C.....	250
VIII.3.1 Installation de la base de données.....	250
VIII.3.2 Utilisation de la base de données.....	250

Abréviation

1-MCP 1-méthycyclopropène
ABA Acide abscissique
ACC acide 1-aminocyclopropane-1-carboxylique
ACO ACC Oxydase
ACP Analyse en Composantes Principales
ACS ACC synthases
ADN Acide Désoxyribonucléique
AFLP Amplified Fragment Length Polymorphisms
ANOVA ANalysis Of VAriance
AOS Allène oxyde synthase
ARF Auxin Response Factor
ARN Acide Ribonucléique
ARR Arabidopsis Response Regulator
ASR Abscissic Stress Related
BLAST Basic Local Aligment Search Tool
CDKP Calcium Dependant protein kinase
CHAPS Sodium 1-heptanesulfonate
CTR Constitutive Triple Response
DEPC Diethyl pyrocarbonate
DOXP 1-Deoxy-D-Xylulose-5-Phosphate
DREB Drought response element binding
DTT Dithiothreitol
EBF Ethylene Binding Factor
EDTA Ethylenedinitro Tetraacetic Acid
Electrophorèse 2D: électrophorèse sur gels bidimensionnels
empreintes massiques peptidiques (PMF)
ER Ethylene Response
EREBP Ethylene Response Element Binding Protein
ERF Ethylene Response Factor
ERS Ethylene Response Sensor
EST Expressed Sequences Tag
ETR Ethylene Receptor
FBPase Fructose bisphosphate aldolase
G3PDH Glycéraldéhyde-3-phosphate déshydrogénase
GA Gibberellic acid
GGPP GéranylGéranyl Diphosphate
HSP Heat Shock Protein
HPLC High Pressure Liquid Chromatography
IAA Indol-acétic acid
IEF Isoelectric Focalisation
INN Isogènes niveau nucléotidique,
INP Isogènes niveau protéique,
IPP Inositol Pyro Phosphate
JAA Jour Après anthèse
JMT Jasmonate carboxyl méthyltransférase
LEA protéine abondante à la fin de l'embryogénèse
LOX Lipooxygénase
MALDI-TOF (MS) Matrix assisted laser desorption/ionization time of flag
MES 2-merc ap- toethanesulfonate

MTA 5' -méthythioadenosine
NCED 9-cis-époxy-caroténoïde dioxygénase
nrEST Non redondant EST
OPDA 12-oxo-phytodiénoïque
Pb paire de bases
PCR Polymerase Chain Reaction
PG Polygalacturonase
PM Poids Moléculaire
PME Pectine Méthylestérase
PMEI Inhibiteurs de pectine méthylestérase
PMF Peptide Mass Fingerprint
PMSF Phenylmethanesulfonyl Fluoride
P/p poids/poids
PPO Polyphénol Oxydase
P/v poids/volume
PVPP Polyvinylpolypyrrolidone
QTL Quantitative Trait Locus
Rubisco Ribulose 1,5-bisphosphate carboxylase
SAGE Seial Analysis of Genes Expression
SAM S-adenosylméthionine
SAMS S-adenosylméthionine synthétase
SDS Sodium Dodecyl Sulfate
SDS-PAGE Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SOD Superoxyde dismutase
SOTA Self Organizing Tree Algorithm
SSH Supression Soustractive Hibridization
SUS Sucrose Synthase
TCA Trichloroacetic acid
TE TRIS-EDTA
TEMED N'-Tetraméthy-1,2-ethanediamine
TPI Triosephosphate isomérase
UCK UMP/CMP Kinase
UPS Universal Stress Protein
UTR Untranslated region
XET Xyloglucane Endo Transglycosylases
ZEP Zéaxanthine époxydase

I INTRODUCTION

La qualité organoleptique des fruits est une des principales préoccupations des consommateurs, des producteurs et des distributeurs. L'évolution des fruits des *Prunus* cultivés (principalement abricotier, pêcher) subit une forte accélération au cours de la maturation et la rapidité d'évolution post récolte du fruit mûr s'avère particulièrement problématique chez l'abricot. En conséquence, la connaissance des mécanismes impliqués dans l'élaboration de la qualité des fruits de ces espèces ligneuses à temps de sélection long, est de première importance. De plus, l'abricotier est l'une des espèces fruitières cultivées qui présente la variabilité génétique la plus large : les métabolites primaires varient dans un rapport de 1 à 20 (5 fois plus que la tomate) alors que le dégagement éthylénique peut évoluer dans un rapport de 1 à 100.

Les travaux développés sur l'abricot ont déjà permis de mettre en évidence certains marqueurs moléculaires des différentes étapes du développement du fruit (Thèse Didier Mbéguié-Mbéguié, 2000). Au début de cette thèse (2001), seulement 204 séquences nucléotidiques étaient connues chez les *Prunus*, dont 45 chez l'abricotier. Dorénavant, différentes stratégies d'analyse du génome, du transcriptome et du protéome offrent des perspectives pour décrypter les voies métaboliques liées à la qualité organoleptique des fruits, et plus particulièrement celles qui interviennent dans le métabolisme des sucres, acides, pigments, composés phénoliques, précurseurs d'arômes et polymères pariétaux. La stratégie retenue a été de décrire de façon substantielle l'expression du génome de l'abricotier, préalablement au criblage des gènes dont l'expression varie en liaison avec l'évolution de la qualité du péricarpe.

Cette thèse s'inscrit dans le cadre du projet LIGNOME qui vise à mettre en place les outils de génomique fonctionnelle chez les espèces ligneuses d'intérêt agronomique et d'ouvrir la voie à la sélection variétale assistée par marqueurs. Elle correspond aux premières étapes du projet abricotier (Fig. 1). L'objectif finalisé de la thèse est de pouvoir identifier, à travers le génome de l'abricotier, les gènes potentiellement marqueurs de qualité du fruit, suivant le double crible du développement, et de certaines différences variétales. Cette analyse devrait permettre de mieux comprendre le processus de maturation de l'abricot au niveau de la (co) régulation des différentes voies métaboliques, de la différenciation et spécificité isogénique de l'organe fruit et des voies de signalisation les régissant.

Afin de réaliser cet objectif, il est donc nécessaire de :

1) Mettre en place les outils de génomique fonctionnelle chez l'abricotier :

- construire une banque d'EST d'abricot à partir des banques d'ADNc existantes correspondant aux stades clés de la maturation pour augmenter les chances d'obtenir des cibles d'intérêt ;

- caractériser ces banques en identifiant le plus précisément possible la fonction des gènes les constituants ;

- mettre en place un support permettant de pouvoir étudier à haut débit (lames de microarray) l'expression des gènes ;

- développer un moyen de vérifier si ces gènes sont effectivement traduits dans les cellules et s'ils sont sujets ou non à des régulations post-transcriptionnelles ;

2) Etudier la régulation des gènes dans le contexte de la maturation, ainsi qu'entre des variétés de phénotypes contrastés. Ceci pour identifier les candidats potentiellement impliqués dans l'évolution des critères de qualité en faisant l'hypothèse que les transcrits/protéines, dont l'abondance varie simultanément avec un trait du phénotype, sont des facteurs limitants de ce caractère.

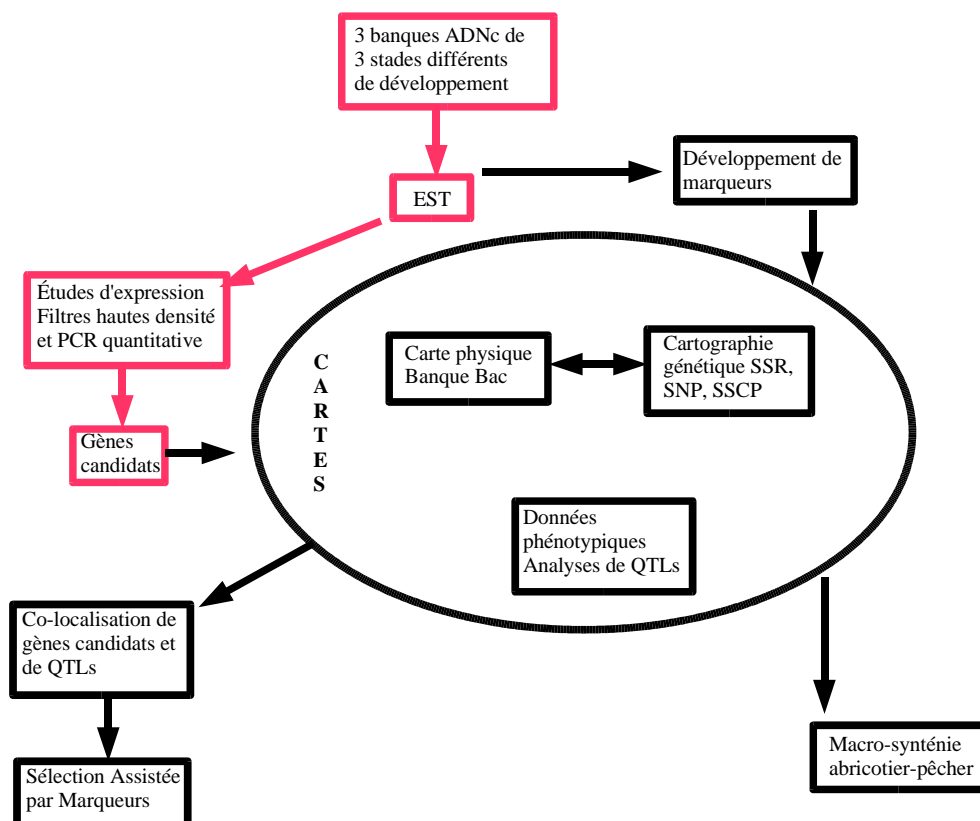


Figure 1 : Organigramme du projet LIGNOME pour l'abricotier. En rouge, les parties effectuées au cours de cette thèse.

II ETUDE BIBLIOGRAPHIQUE

L'abricot est le 20^{ème} fruit cultivé en terme de volume avec 2,708 millions de tonnes produites par an. Les principaux producteurs sont dans l'ordre la Turquie, l'Iran, le Pakistan, la France et l'Espagne.

La faible capacité de conservation de l'abricot a des incidences économiques importantes, du fait de la limitation de la commercialisation des fruits aux régions proches de celles de production et durant une période relativement courte. Ainsi, la quasi totalité des échanges sur le marché de l'abricot se situe dans les pays méditerranéens, pendant la période estivale et les exportations sont extrêmement limitées.

II.1 Caractéristiques physiologiques de l'abricotier

II.1.1 Conditions climatiques de culture

La production mondiale est principalement située dans une large bande comprise entre 25° et 45° de latitude nord, souvent en association avec les cultures de pêches, nectarines et prunes. Cependant, l'abricot n'est pas aussi adaptable aux conditions climatiques que ces autres fruits à noyau. Une variété qui produit correctement dans une région est souvent improductive dans une autre. L'abricotier requiert un climat méditerranéen chaud, a besoin d'hivers froids et humides pour arrêter sa dormance et d'étés chauds et secs. L'abricotier fleurit tôt, ce qui l'expose aux gels de printemps dans la plupart des régions où il est cultivé. Les fruits ont tendance à éclater par temps humide. L'arbre est assez résistant à la sécheresse, mais nécessite une irrigation supplémentaire pour que les fruits atteignent leur potentiel de rendement maximal.

Les variétés commerciales sont toutes greffées. L'abricotier est adapté aux sols dont les pH sont compris entre 6 et 8. L'espèce est assez tolérante aux conditions alcalines mais est très sensible aux concentrations élevées en sel.

II.1.2 L'arbre

La taille de l'arbre peut atteindre entre 10 et 15 mètres, mais en culture la taille est maintenue inférieure à 3,5 m. Les feuilles sont caduques. Les fleurs qui apparaissent avant les feuilles sont blanches ou roses, avec 5 sépales, 5 pétales réguliers et plusieurs étamines. Les feuilles sont lisses, grandes et arrondies avec les bords dentelés et un apex en pointe. Le pétiole, de couleur tendant vers le rouge, mesure de 1 à 3 centimètres.

II.1.3 Anatomie du fruit

Le fruit de l'abricotier est une drupe, c'est à dire un fruit simple charnu à noyau qui dérive d'un ovaire infère à un carpelle situé dans le conceptacle caduque au sommet duquel sont fixées les pièces florales.

La partie externe du péricarpe (mésocarpe et épicarpe) est charnue et comestible (Fig. 2). La partie interne (endocarpe) est lignifiée (noyau) ; cette partie entoure et protège la graine. On observe à la base du fruit la cicatrice du pédoncule floral et au sommet le point de chute du style. Le sillon que l'on observe sur un côté du fruit représente la suture carpellaire qui s'étend de l'attache du pédoncule à l'apex. Le fruit provient donc d'un seul carpelle, dans lequel une seule graine (parfois deux) se développe(nt).

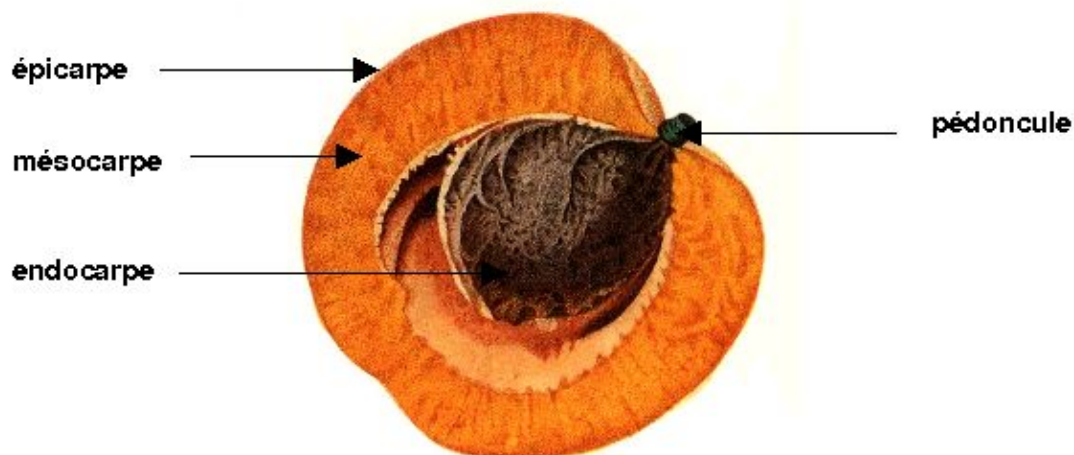


Figure 2 : Schéma simplifié d'une coupe longitudinale d'abricot à maturité.

Le mésocarpe est un tissu majoritairement parenchymateux qui devient mou lorsque le fruit est mûr ; il est fortement vascularisé. Chez les fruits mûrs, le mésocarpe et l'endocarpe sont séparés par une cavité périnucléaire. Le noyau, dans la majorité des variétés est donc

libre ou faiblement adhérent, d'où la classification en drupe de ce fruit. Pour certaines variétés, cependant, le noyau est très adhérent.

II.1.4 Composition et valeur nutritionnelle du fruit mûr

La chair contient environ 85 % d'eau pour les fruits mûrs (Wills, 1987). Les principaux constituants retrouvés dans la vacuole sont les sucres et les acides organiques (malate et citrate). Le saccharose est la forme majoritaire d'hydrate de carbone (60 %) chez l'abricot. Le glucose (25 %), le fructose (6 %) et le sorbitol (7 %) sont en quantité plus minoritaires. La concentration en sorbitol est relativement faible, dans les autres rosacées, comme la pêche et la pomme, il est transporté en quantité via le phloème vers le fruit où il est probablement métabolisé en sucres réducteurs (Genard et Souty, 1996 ; Gao *et al.*, 2003b). L'abricot ne contient pas d'amidon. Les deux anions organiques majoritaires sont les acides citrique et malique, dont les concentrations et le rapport sont extrêmement variables en fonction des variétés. On retrouve aussi de l'acide quinique en faible quantité. La plupart des variétés d'abricot est riche en carotènes (majoritairement β carotène) ce qui leur confère une bonne activité provitaminique A et leur couleur de fond orangée. Les anthocyanes sont responsables des taches rouges en sur-impression sur la pellicule. Chez les Double Rouge, elles peuvent pratiquement recouvrir toute la pellicule.

Il est cependant difficile et vain de tirer un profil type de l'abricot tant la diversité phénotypique est élevée chez cette espèce.

II.1.5 Le développement du fruit

Après la floraison, l'accroissement de la taille des fruits suit une courbe en forme de « double sigmoïde » découpée en 3 phases (Fig. 3) :

- La première phase (I) de la courbe de croissance résulte principalement de l'élargissement de toutes les parties de l'ovaire à l'exception de l'endosperme et de l'embryon. L'allure du début de cette partie de la courbe est exponentielle car on assiste à la fois à la multiplication des cellules et à leur élargissement. La phase de multiplication cellulaire est relativement courte (1 à 2 semaine).

- Dans la seconde phase (II) la croissance ralentit. Elle est consacrée à la lignification de l'endocarpe et la croissance est confinée à l'endosperme et à l'embryon.
- Durant la troisième phase (III). Les cellules du mésocarpe retournent en croissance active. Elles vont se charger en assimilats et en eau, la taille et le poids du fruit vont augmenter ; c'est au cours de cette phase qu'a lieu la maturation et que le mésocarpe devient comestible.

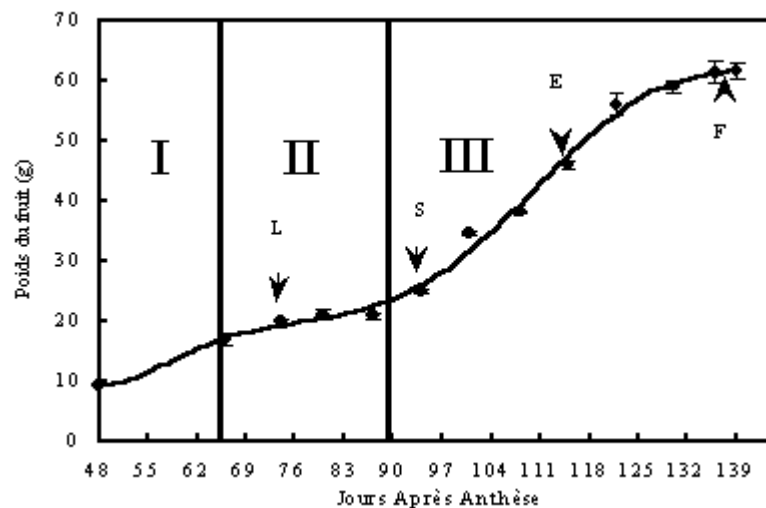


Figure 3 : Courbe de croissance de l'abricot (variété Bergeron) sur l'arbre. (d'après Mbéguié-Mbéguié, 2000). L : début de lignification du noyau ; S : début d'accumulation du saccharose ; E : début d'émission autocatalytique d'éthylène et de la maturation du fruit F : Fruit physiologiquement mûr.

II.2 Caractéristiques géographiques et génétiques de l'abricotier

II.2.1 Dispersion de l'abricotier

L'abricotier est originaire des régions montagneuses du nord et du nord ouest de la Chine dans le secteur de la grande muraille. Il y est cultivé depuis environ 4000 ans. Il existe des centres d'origine secondaire possibles dans la région autonome du Xinjiang et en Russie orientale (Vavilov, 1949). Au cours des siècles suivants, des graines ont été introduites en Asie centrale (Arménie, Perse). L'abricotier a été introduit au sud de l'Europe (Grèce) au cours des conquêtes d'Alexandre Le Grand pendant le 4^{ème} siècle avant JC. Il est arrivé en Italie au 1^{er} siècle après JC, en Angleterre en 1542 et aux États Unis pendant le 19^{ème} siècle.

L'abricotier a été introduit en France à travers deux routes différentes. Les premières variétés, originaires d'Arménie et d'Afrique du Nord ont été apportées vers l'an 1000 par les arabes dans le sud de la France. Puis, 440 ans plus tard, des variétés plus adaptées aux régions septentrionales provenant de Hongrie et d'Europe centrale ont fait leur apparition (Mehlenbacher *et al.*, 1990 ; Faust *et al.*, 1998).

II.2.2 Diversité génétique

Il existe, en fonction des classifications, entre 3 et 10 espèces d'abricotier, toutes diploïdes à $2n=16$ chromosomes (Mehlenbacher *et al.*, 1990). Communément, 4 espèces sont reconnues : *Prunus armeniaca*, *Prunus mandshurica*, *Prunus sibirica* et *Prunus mume*. Sur la base de critères morphologiques et de descriptions pomologiques, la plupart des variétés a été classée dans l'espèce *Prunus armeniaca*. D'après leur adaptation écologique, Kostina (1969) a classé les différentes variétés de *Prunus armeniaca* en quatre sous groupes géographiques : le Centre asiatique, l'Irano-Caucasien, l'Européen et le Dzungar-Zailij. Récemment, les groupes Nord chinois et Est chinois ont été ajoutés à cette classification (Layne *et al.*, 1996). A l'intérieur de ces groupes, on retrouve des caractéristiques groupe-spécifiques comme le type d'arbre, de fruits, la période de dormance, l'auto-incompatibilité et la résistance aux maladies.

Les études de diversité génétique conduites au cours des dernières années ont permis de tirer d'importantes conclusions sur la diversité de l'abricotier et en particulier sur le groupe européen. La diversité entre les variétés européennes et nord américaines est si élevée que les variétés américaines ne peuvent pas être considérées comme dérivant du groupe européen uniquement (Badenes *et al.*, 1996 ; Hormaza 2002). Les variétés européennes et plus particulièrement les espagnoles présentent peu de diversité génétique entre elles (de Vincente *et al.*, 1998).

Récemment, Hagen *et al.* (2002) en utilisant des marqueurs AFLP ont mis en évidence les relations génétiques entre *Prunus armeniaca* et les variétés proches. Ainsi, la distance génétique avec d'autres espèces (dont *Prunus mume*) a pu être établie et quatre groupes de la variété *Prunus armeniaca* ont été mis en évidence en utilisant une palette représentative des variétés eurasiennes. A l'intérieur de ces 4 groupes, les variétés sont originaires des mêmes régions et possèdent des propriétés agronomiques spécifiques et homogènes. Il a été observé une plus grande diversité génétique chez les variétés orientales. En utilisant des marqueurs

SSR sur 74 accessions appartenant aux 6 groupes de *Prunus armeniaca* définis par Layne *et al.* (1996). Zhebentyayeva *et al.* (2003) ont mis en évidence des groupes qui ne sont pas totalement corrélés à l'origine géographique des cultivars. Ceci reflète l'histoire complexe de la domestication de l'abricotier, en particulier dans le groupe d'Asie centrale, où les sous groupes sont fortement distants ; les sous groupes Kopet-Dag et Zeravshan semblent être proches du groupe Dzungar-Zailij, tandis que le sous groupe Fergana est proche des variétés irano-caucasiennes et européennes. Leurs observations rejoignent celles de Kostina (1931) selon qui la vallée de Fergana serait la région la plus représentative et la plus diversifiée. Les ressources génétiques de cette région ont eu une influence considérable sur la dispersion des abricotiers à travers le monde.

En ce qui concerne plus spécifiquement la variabilité phénotypique du fruit, 400 variétés ont été caractérisées (Audergon *et al.*, 1991) et une forte variabilité a été observée pour l'ensemble des critères influençant la qualité du fruit (calibre, couleur, fermeté, sucres, acides organiques, vitesse de maturation, profils aromatiques).

II.3 La maturation des fruits

Classiquement, la maturation est définie comme étant la phase de développement des fruits qui s'étend de la véraison (début du virage de la teinte) à la maturité. Elle débute au cours de la troisième phase de croissance et se termine alors que la sénescence est déjà engagée. La finalité de la maturation est de rendre les fruits attractifs pour promouvoir leur consommation par des organismes qui vont faciliter la libération et la dispersion des graines. Cette opération s'effectue via une modification des tissus comestibles (le péricarpe et le mésocarpe chez l'abricot) (Adams-Phillips *et al.*, 2004a).

Les fruits peuvent être divisés en 2 groupes contrastés de mécanismes de maturation. Les fruits climactériques (abricot, tomate pomme, banane) présentent une explosion de la synthèse d'éthylène autocatalytique et une augmentation transitoire de leur activité respiratoire pendant la maturation, contrairement aux fruits non climactériques (fraise, raisin, *Citrus*) (Yang et Hoffman, 1984). Chez les fruits non climactériques, l'application d'éthylène exogène ne permet pas d'enclencher la maturation, bien que certaines voies métaboliques puissent y répondre partiellement (Alonso *et al.*, 1995).

Chez l'abricotier, comme chez les autres espèces climactériques, la vitesse d'évolution sur l'arbre est corrélée avec le dégagement d'éthylène. Quelques jours avant ce dégagement, le fruit est déjà apte à mûrir après récolte s'il est soumis à de l'éthylène exogène. En croissance sur l'arbre dès que le dégagement d'éthylène se produit, le fruit évolue rapidement vers la maturité.

Les spécificités biochimiques et physiologiques de la maturation varient en fonction des espèces, cependant, celles-ci incluent généralement le changement de couleur, de texture, de flaveur, d'arôme, de contenu nutritionnel et la susceptibilité aux pathogènes évoluent de façon convergente même chez des fruits qui ne dérivent pas des mêmes assises cellulaires de la fleur (Seymour *et al.*, 1993). La maturation est influencée par des facteurs internes et externes, parmi lesquels les hormones qui régulent l'expression des gènes durant le développement, la lumière (Alba *et al.*, 2000) et la température. Mais jusqu'à récemment, les avancées significatives dans la compréhension des mécanismes moléculaires ont été focalisées sur le rôle et la régulation de la biosynthèse de l'éthylène (Theologis, 1992). Il co-existe cependant, à la fois chez les fruits climactériques et non climactériques, des voies métaboliques éthylène-dépendantes et indépendantes (Lelièvre *et al.*, 1997). Bien que leur rôle et mode d'action soient actuellement moins bien caractérisés que pour l'éthylène, la plupart des phytohormones, comme l'acide abscissique (Jiang *et al.*, 2000), l'auxine (Visser *et al.*, 1996), l'acide gibbérellique (Choi *et al.*, 2004), l'acide jasmonique (Creelman et Mullet, 1997) influence la maturation des fruits climactériques.

II.4 Les critères de qualité liés à la maturation

II.4.1 La perte de fermeté

La perte de fermeté induite au cours de la maturation chez les fruits est le principal facteur limitant la conservation et la durée de vie des fruits après la récolte (Jimenez-Bermudez *et al.*, 2002) ainsi que les possibilités d'attaque de pathogène. D'un point de vue général, le ramollissement des fruits est dû à la perte d'adhésion intercellulaire, ce mécanisme intervient dans la plupart des organes et à différentes périodes de la vie des plantes (Roberts *et al.*, 2002). Il permet aux racines d'émerger des graines pendant la germination, aux

cotylédons et aux feuilles de croître, aux fruits de devenir comestibles et l'abscission du pédoncule (Fig. 4). Bien que dans l'ensemble de ces cas les processus biochimiques qui consistent à métaboliser la paroi cellulaire soient comparables, la nature du signal qui induit ces changements est différent afin que chaque processus soit enclenché au moment et dans l'organe opportuns.

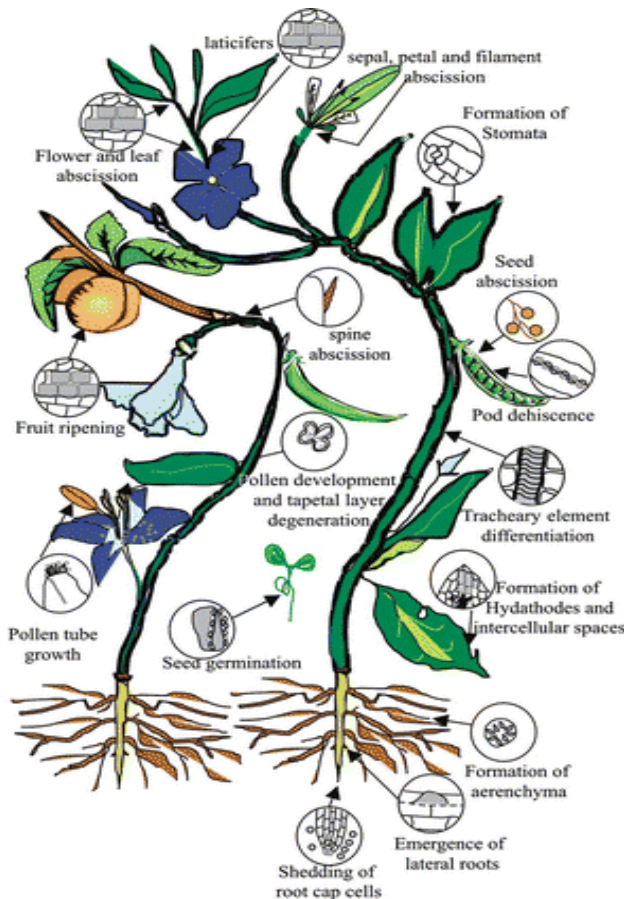


Figure 4 : Les sites où a lieu la séparation des cellules dans les plantes (Roberts *et al.*, 2002).

Dans certains cas, la dégradation de la paroi peut être extrêmement localisée comme pour l'abscission qui est limitée à une couche de cellules (abscission zone, AZ, González-Carranza *et al.*, 1998), alors que dans le fruit le ramollissement affecte l'intégralité des parties comestibles. De par la diversité des organes dans lesquels elles interviennent, les enzymes liées à la dégradation de la paroi sont extrêmement multigéniques, ce qui est en rapport avec la diversité des organes dans lesquels elles interviennent, ou avec des spécificités fonctionnelles mal connues. On compte par exemple chez *Arabidopsis* 38 expansines différentes (Li *et al.*, 2002), au moins 12 endo-1,4- β -D-glucanases (Del Campillo, 1999), 52 polygalacturonases, 20 pectate lyases et 79 pectine estérases (The Arabidopsis Initiative, 2000).

Dans le cadre de l'étude du ramollissement au cours du développement du fruit, il est donc nécessaire de pouvoir identifier d'une part, les profils d'expression des différentes isoformes, et d'autre part celles qui sont fruit-spécifiques.

II.4.1.1 Biochimie de la paroi cellulaire

Cette matrice extracellulaire entourant chaque cellule végétale est composée de 90 % de polysaccharides (pectine, cellulose et hemicellulose chez les fruits) et de 10 % de glycoprotéines. Ces différents constituants forment un réseau complexe de macromolécules responsable des propriétés mécaniques et informationnelles de la paroi cellulaire. Cet exosquelette joue un rôle critique dans le contrôle de la pression de turgescence, qui s'oppose à l'appel d'eau liée à la pression osmotique vacuolaire. L'action de ces deux mécanismes permet de contrôler le grandissement cellulaire, suivant un dynamisme permanent, fortement perturbé lors de l'initiation de la seconde phase de croissance.

La synthèse des polysaccharides de la paroi cellulaire végétale résulte d'un processus complexe nécessitant l'intervention de nombreuses voies biochimiques différentes. A titre d'exemple on estime que la synthèse des pectines, un des composants de la paroi, nécessiterait à elle seule la contribution d'au moins 53 enzymes (Mohnen, 1999). Tous les polysaccharides pariétaux sont synthétisés à partir de sucres diphosphonucléosidiques eux même provenant d'hexoses phosphates par l'intermédiaire de nucléosides diphosphates pyrophosphorylases. Plusieurs enzymes telles que des épiméras, déshydrogénases, déshydratases et décarboxylases sont impliquées dans l'interconversion des sucres diphosphonucléosidiques. La polymérisation des polysaccharides est quant à elle réalisée par des enzymes membranaires qui transfèrent un ose d'un sucre diphosphonucléosidique sur une molécule donnée (glycosyltransférase ou polysaccharide synthase) (Ridley *et al.*, 2001). De la même manière, les réaménagements, ou la dégradation des polysaccharides pariétaux est réalisée par de nombreuses enzymes modifiant et/ou hydrolysant les principaux composants de la paroi. Les gènes codant pour ces enzymes appartiennent aux familles des glucanases, polygalacturonases, pectine méthylestérases, pectate lyases, extensines, xyloglucan endotransglycosylases et expansines.

II.4.1.1.1 Polygalacturonases

Les polygalacturonases (PG) hydrolysent la liaison glycosidique de polymères d'acide galacturonique (comme les pectines). Dans le cadre de la perte de fermeté lors de la

maturation des fruits, la polygalacturonase est une des enzymes qui a été la plus étudiée. Ainsi chez la tomate, l'observation d'une forte activité endo-PG dans les fruits mûrs a conduit à l'hypothèse que les PG jouent un rôle important dans la perte de fermeté (Giovannoni *et al.*, 1989). Cependant, la caractérisation fonctionnelle de transgènes chez lesquels l'activité polygalacturonase était réduite de 99 % n'a pas permis de mettre en évidence d'impact notable sur la texture du fruit (Giovannoni *et al.*, 1989, Smith *et al.*, 1990). Cependant ces mêmes lignées transgéniques ont été utilisées pour démontrer le rôle de la polygalacturonase dans l'augmentation de la susceptibilité envers les pathogènes (Kramer *et al.*, 1992). La spécificité d'expression des polygalacturonases a été étudiée chez le melon. Parmi les 6 transcrits étudiés, 3 s'expriment dans les fruits (Hadfield *et al.*, 1998). Chez la pêche, le niveau d'expression d'une polygalacturonase dans différents cultivars est corrélé avec leur fermeté (Lester *et al.*, 1994). De plus elle est uniquement induite lorsque que le fruit mûrit.

II.4.1.1.2 Pectine méthylestérases

Les pectine méthylestérases (PME) interviennent dans la perte de fermeté en déméthylant les pectines, favorisant ainsi la susceptibilité des ces dernières aux polygalacturonases (Micheli, 2001). Comme les polygalacturonases, l'activité PME est induite en début de maturation du fruit chez la tomate (Tieman *et al.*, 1992), et l'isogène dont l'expression a été étudiée chez la pêche apparaît aussi au début la maturation (Trainotti *et al.*, 2003).

II.4.1.1.3 Pectate lyases

Les pectate lyases catalysent le clivage des liaisons entre les résidus galacturonosyls des pectines dé-estérifiées (Carpita et Gibeaut, 1993). Elles n'agissent pas sur les pectines non dé-estérifiées par les PME. Chez la pêche, deux isogènes de pectate lyase sont induits au cours de la maturation (Trainotti *et al.*, 2003). La suppression d'une pectate lyase dans des fraises transgéniques aboutit à l'obtention de fruits significativement plus fermes (Jiménez-Bermudez *et al.*, 2002). L'activité pectate lyase n'a pu être mesurée « in planta » que récemment (Marin-Rodriguez *et al.*, 2003) et il a fallu attendre les programmes EST pour pouvoir isoler un nombre important de transcrits (potentiellement 30 transcrits différents sur le site du TIGR pour la tomate, www.tigr.org).

II.4.1.1.4 Xyloglucan-endotransglycosylases

Les xyloglucan-endotransglycosylases (XET) sont capables de cliver une chaîne de xyloglucanes (hémicellulose) et de lier l'extrémité réductrice néo-formée à l'extrémité non-réductrice d'un substrat donneur (Fry *et al.*, 1992). Cette réaction permet l'expansion de la paroi des végétaux par transglycosylation d'une molécule de xyloglucane nouvellement sécrétée afin qu'elle s'intègre au réseau de xyloglucane-cellulose (Thompson et Fry, 1997). Les XET sont donc impliquées dans les mécanismes de synthèse, croissance et dégradation de la paroi (Fry *et al.*, 1992 ; Nishitani et Tominaga, 1992). Dans les fruits, les XET jouent potentiellement un rôle dans le grandissement cellulaire (Campbell et Braam, 1999) et le ramollissement au cours de la maturation (Rose et Bennett, 1999). Le profil d'expression de certaines XET est lié à la perte de fermeté chez le kiwi (Schroder *et al.*, 1998) et la tomate (Maclachlan et Brady, 1994). Il est aussi possible que certaines XET soient aptes à utiliser les pectines comme substrat accepteur afin de créer des liaisons xyloglucan-pectine (Thompson *et al.*, 2000).

II.4.1.1.5 Expansines

Les expansines affectent la dépolymérisation des hémicelluloses (McQueen-Mason et Cosgrove, 1994) par une activité non hydrolytique, probablement en accroissant l'accessibilité du substrat pour les autres enzymes (Powell *et al.*, 2003). Chez la poire, les expansines appartiennent à une famille multigénique dont les membres sont régulés différemment au cours de la maturation (Hiwasa *et al.*, 2003a). Rose *et al.* (1997) ont mis en évidence leur rôle dans la perte de fermeté. La suppression de l'isogène LeExp1 chez la tomate permet d'obtenir des fruits significativement plus fermes (Powell *et al.*, 2003). Chez l'abricot deux isogènes ont été identifiés et leur niveau d'expression augmente pendant la maturation (Mbéguié-Mbéguié *et al.*, 2002).

II.4.1.1.6 β -glucanases

Les endo β -1-4-glucanases (hémicellulases) hydrolysent la liaison glycosidique à l'intérieur des chaînes d'hémicellulose. Deux hémicellulases dont le niveau d'expression augmente pendant le développement ont été réprimées par une construction antisens chez la tomate, sans impact sur la perte de fermeté, mais l'abscission des fruits a été inhibée (Brumell *et al.*, 1999). Il convient de préciser que l'activité enzymatique correspondante n'a pas été dosée dans ces études et que la famille est fortement multigénique.

II.4.2 Couleur

Deux mécanismes participent au changement de couleur pendant la maturation : la dégradation des chlorophylles et la synthèse de pigments colorés, majoritairement des caroténoïdes chez l'abricot. On retrouve également des anthocyanes qui sont les pigments majoritaires dans la pellicule de certaines variétés (Double Rouge).

II.4.2.1 Dégradation des chlorophylles

La dégradation des chlorophylles, sous contrôle de l'éthylène pendant la maturation (Trebitsh *et al.*, 1993), est réalisée par la voie métabolique du phéophorbide *a* conduisant à des composés incolores dans les feuilles sénescents (Matile *et al.*, 1999). L'ensemble des composés intermédiaires a été retrouvé dans les fruits de chérimolier (Almela *et al.*, 2000), indiquant que cette voie métabolique est aussi présente dans les fruits. Deux enzymes de cette voie ont été étudiées pendant la maturation, la chlorophyllase *a* chez les *Citrus* (Trebitsh *et al.*, 1993) et la Mg-déchelatase chez la fraise (Costa *et al.*, 2002). Leurs activités sont corrélées avec la dégradation des chlorophylles.

II.4.2.2 Biosynthèse des caroténoïdes

Les caroténoïdes sont une famille de molécules synthétisées dans les plastes et retrouvées chez tous les organismes photosynthétiques. Ils incluent les carotènes, comme le lycopène et le β -carotène, et les xanthophylles comme la lutéine. Chez les variétés communes d'abricots comme Bergeron, le caroténoïde majoritaire dans le fruit est le β -carotène, principal pigment impliqué dans la couleur orangée du fruit. Les caroténoïdes sont aussi les précurseurs de l'acide abscissique (Rock et Zeevart, 1991).

Dans les chloroplastes, les caroténoïdes jouent un rôle vital dans le réseau de pigments-protéines « light-harvesting complex » responsable de l'absorption de la lumière pendant la photosynthèse (Peter et Thornber, 1991). Ce complexe lie les chlorophylles et certains caroténoïdes, principalement la lutéine, la néoxanthine et la violaxanthine (Snyder *et al.*, 2004). Dans les chromoplastes, les caroténoïdes sont impliqués dans la protection contre l'oxydation (Bouvier *et al.*, 1998).

Les plantes synthétisent les caroténoïdes via la voie métabolique du 1-déoxy-D-xylulose-5-phosphate (DOXP) préférentiellement à la voie de l'acide mévalonique (Schwender *et al.*, 1996). Ces deux voies conduisant à la formation d'inositol pyro-phosphate (IPP), la voie de l'acide mévalonique est impliquée uniquement dans la synthèse des stérols, sesquiterpénoïdes

et triterpénoïdes dans le cytosol, tandis que la voie DOXP conduit à la formation d'isoprénoïdes dans les plastes. L'IPP est isomérisé en diméthylallyl diphosphate, qui est le substrat actif nécessaire à la formation de géranylgeranyl diphosphate (GGPP), précurseur du premier caroténoïde, le phytoène. L'activité IPP isomérase est retrouvée dans le cytosol et dans les plastes et deux gènes codant pour l'IPP isomérase sont identifiés chez les plantes (Sun *et al.*, 1998). Une seule enzyme chromoplastique (GGPP synthase, Fig. 5) catalyse la formation de GGPP à partir d'IPP et de DMAPP ; cependant elle est multigénique chez *Arabidopsis* et au moins 5 isogènes sont connus et exprimés dans différents tissus (Okada *et al.*, 2000). La condensation de deux molécules de GGPP pour former le phytoène est réalisée par la phytoène synthase. La tomate contient deux gènes *Psy-1* et *Psy-2*. Le premier est spécifique de la maturation du fruit et le second est prédominant dans les fruits verts et n'a pas de rôle apparemment direct avec la caroténogénèse pendant la maturation (Fraser *et al.*, 1999). Une mutation ou une construction antisens de *Psy-1* engendre des fruits jaunes et l'absence de caroténoïdes (Bird *et al.*, 1991). Deux enzymes membranaires, la phytoène désaturase et la ζ -carotène désaturase convertissent le phytoène en lycopène via le ζ -carotène. Le récent clonage d'une carotène isomérase (Isaacson *et al.*, 2002) a permis d'élucider le mécanisme d'isomérisation cis-trans intervenant pendant ces étapes. La cyclisation du lycopène permet d'obtenir une série de caroténoïdes qui possèdent un ou deux cycles. La lycopène β -cyclase (LCY-B/CRTL-B) catalyse une réaction en deux étapes qui conduit au β -carotène et la lycopène ϵ -cyclase (LCY-E/CRTL-E) conduit au δ -carotène et au ϵ -carotène. L' α -carotène, le précurseur de la lutéine est formé par l'action des deux enzymes. Le genre *Lycopersicon* possède deux β -lycopène cyclases LCY-B et CYC-B, uniquement retrouvées dans les chromoplastes, LCY-B est chloroplastique (Ronen *et al.*, 2000). Les xanthophylles sont formées par oxygénation des carotènes en ajoutant des groupements hydroxy, époxy, aldéhyde, carboxy ou méthoxy. L'hydroxylation est effectuée par deux types d'enzymes, dont une spécifique du β -carotène, la β -carotène hydroxylase (Sun *et al.*, 1996). Le β -carotène est converti en zéaxanthine via la β -cryptoxanthine par cette enzyme. On compte deux β -carotène hydroxylases chez la tomate, une spécifique des fruits verts et une spécifique de la fleur (Ronen *et al.*, non publié, cité par Hirschberg, 2001).

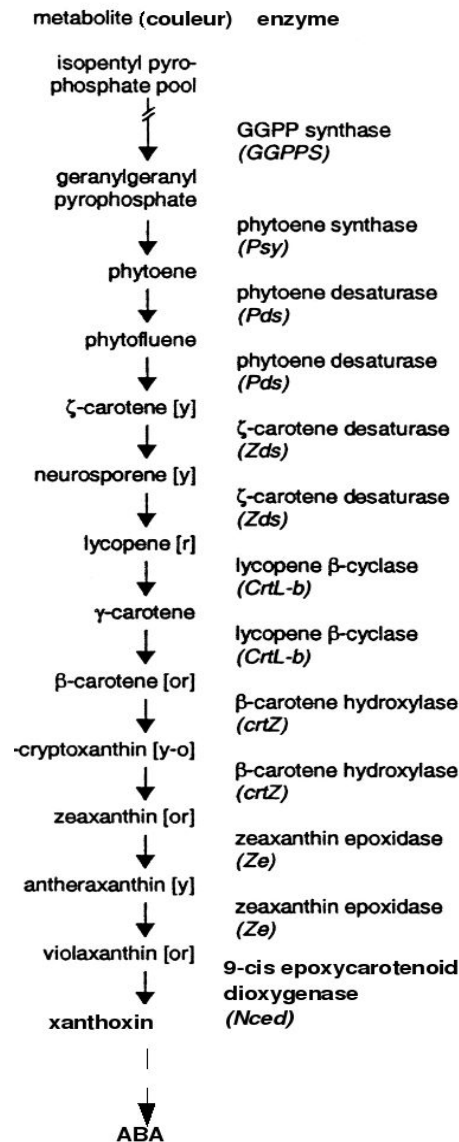


Figure 5 : Schéma simplifié de la voie de biosynthèse et de dégradation des caroténoïdes.

II.4.2.3 Dégradation des caroténoïdes et biosynthèse de l'acide abscissique (ABA)

La première étape spécifique de la voie de synthèse de l'acide abscissique est la conversion de la zéaxanthine en trans-violaxanthine par une époxydation en deux étapes. L'enzyme qui catalyse cette réaction est la zéaxanthine époxydase (ZEP) (Marin *et al.*, 1996). L'enzyme impliquée dans la conversion de la trans-violaxanthine en 9-cis-violaxanthine ou 9'-cis-néoxanthine n'a pas encore été isolée. L'étape suivante, le clivage oxydatif de la 9-cis-violaxanthine et/ou 9'-cis-néoxanthine pour obtenir la xanthoxine, est catalysée par la 9-cis-époxy-caroténoïde dioxygénase (NCED) (Schwartz *et al.*, 1997). Bien que la suppression d'une seule isoforme soit critique pour stopper la synthèse de l'ABA chez la tomate (Burbidge *et al.*,

1999), la NCED est multigénique chez toutes les espèces où elle a été observée, par exemple dans l'avocat (Chernys et Zeevaart, 2000) et *Arabidopsis* (Iuchi *et al.*, 2001 ; Schwartz *et al.*, 2001).

La conversion de la xanthoxine en acide abscissique se déroule dans le cytosol, le mécanisme n'est pas encore connu mais 3 voies métaboliques ont été proposées, via l'aldéhyde abscissique, l'acide xanthoxique ou l'alcool abscissique (Seo et Koshiba, 2002).

II.5 Arômes

La première étude majeure réalisée sur les composés volatiles de l'abricot a été réalisée par Tang et Jennings en 1967. Ces auteurs ont identifié, par chromatographie en phase gazeuse et spectroscopie infrarouge, plusieurs terpènes et lactones. Environ 80 composés volatils ont été identifiés par la suite dans l'abricot, appartenant à différentes familles chimiques telles que les hydrocarbures, les alcools, les aldéhydes, les cétones, les esters, les lactones... (Guichard et Souty, 1988 ; Takeoka *et al.*, 1990) et provenant de diverses voies de biosynthèse : dégradation des acides gras, dégradation des caroténoïdes, métabolisme des acides aminés issus de la voie des dérivés shikimiques.... L'olfactométrie a permis de mettre en évidence que la présence du benzaldéhyde, du linalool, du 4-terpinéol et de l' α -terpinéol pouvait expliquer les odeurs fruitées et florales et les lactones les notes de fond dans la variété Rouge du Roussillon (Chairote *et al.*, 1981). Dans la variété Blenheim, en comparant les concentrations des différents constituants volatils de l'abricot par rapport à leur seuil de détection, Takeoka *et al.* (1990) ont montré que les constituants majeurs de l'arôme sont : la β -ionone, le linalool, la γ -décylactone, le β -cyclocitral, le phénylacétaldéhyde et la γ -octalactone. L'évolution au cours du développement des principaux composés d'arôme a été étudiée chez une variété américaine P305-175 et il a été montré que parmi les familles de composés étudiées (cétones, alcools, aldéhydes, lactones, composés hétérocycliques), seules les lactones apparaissent après le stade vert mature (Gomez et Ledbetter, 1997). Dans la même étude, une augmentation de la plupart des esters a été observée chez la prune avec, en particulier, une forte concentration en méthylsalicylate.

II.6 Biosynthèse des Hormones

La biosynthèse de l'acide abscissique est présentée précédemment dans le paragraphe catabolisme des caroténoïdes (II.4.2.3).

II.6.1 Ethylène

L'éthylène est essentiel pour le développement normal des plantes, leur croissance et leur survie. Cette hormone intervient dans les signalétiques sous-jacentes à la germination des graines, au développement des fleurs et des fruits, dans l'induction de certains mécanismes de défense et en interaction avec de nombreuses hormones végétales.

L'éthylène est produit dans la plupart des tissus végétaux (Yang et Hoffman, 1984). Sa biosynthèse commence à partir de la *S*-adénylméthionine (SAM), le précurseur ubiquiste des méthyations dans de nombreuses autres voies métaboliques, ce qui permet d'expliquer qu'elle soit retrouvée en abondance dans les tissus végétaux. L'ACC synthase convertit la SAM en 1-aminocyclopropane-1-carboxylique acide (ACC) et 5'-méthylthioadénosine (MTA). Il a été suggéré que cette étape serait limitante pour la biosynthèse de l'éthylène (Yang et Hoffman, 1984). L'ACC synthase est codée par une famille multigénique et la transcription des différentes formes est induite par différentes conditions environnementales ou physiologiques (Theologis, 1992). On compte par exemple chez la tomate 9 isoformes (Vogel *et al.*, 1998).

L'étape suivante est la conversion de l'ACC en éthylène par l'ACC oxydase, enzyme présente dans la plupart des tissus en faible quantité relative. Comme pour l'ACC synthase, plusieurs isoformes d'ACC oxydase ont été identifiées et sont activées dans des conditions physiologiques différentes.

II.6.2 Auxines

La principale auxine retrouvée chez les plantes est l'acide indole-acétique (AIA). Elle coexiste avec d'autres auxines dont les rôles physiologiques restent obscurs (Normanly *et al.*, 1995). L'AIA est synthétisé à partir du tryptophane via trois voies possibles (l'acide indole-3-pyruvique, la tryptamine, ou l'indole-3-acetonitrile). Il est possible que la même plante utilise

des voies différentes en fonction de son stade de développement (Michalczuk *et al.*, 1992). La biosynthèse de l'AIA peut aussi être réalisée via une voie tryptophane indépendante ; certains mutants de maïs ne produisent pas de tryptophane mais accumulent 50 fois plus d'AIA que les types sauvages (Wright *et al.*, 1992).

II.6.3 Gibbérellines

La biosynthèse des gibbérellines (GA) peut être divisée en trois étapes en fonction de leur localisation cellulaire pour aboutir aux composés actifs chez les plantes supérieures (GA1, GA3, GA4 et GA7). La première étape est la cyclisation du géranyl-géranyl diphosphate (GGPP, qui est aussi le précurseur du phytoène) en ent-copalyl PP par la ent-copalyl diphosphate synthase et la conversion en ent-kaurène par la ent-kaurène synthase (Duncan et West, 1981). Cette étape a lieu dans les plastes, mais il semble que le GGPP soit synthétisé à partir de la voie du mévalonate contrairement au GGPP destiné à la biosynthèse des caroténoïdes (Hedden, non publié, cité par Hedden et Phillips, 2000). La seconde étape est réalisée dans la membrane du réticulum endoplasmique par une série d'oxydations catalysées par des mono-oxygénases cytochrome P450-dépendantes pour aboutir à la formation du GA12 qui peut être oxydé pour former les GA53. Ces deux composés sont les précurseurs de toutes les autres GA. La troisième étape est la formation des GA bioactives ; elle est catalysée par deux dioxygénases oxoglutarate dépendantes utilisant le fer comme cofacteur. La première enzyme, la GA20 oxydase, est multifonctionnelle et catalyse en plusieurs étapes la formation du GA9 (à partir du GA12) et du GA20 (à partir du GA53) (Lange *et al.*, 1994). Les GA bioactives GA4 et GA1 sont formées à partir de GA9 et GA20 respectivement, par l'action d'une GA3 β -hydroxylase. Dans certaines espèces, les GA9 et GA20 sont aussi converties en GA7 et GA3. A la fin de la voie métabolique, une troisième dioxygénase, la GA-2-oxydase désactive les GA par une 2 β -hydroxylation.

II.6.4 Méthyl jasmonate

Le jasmonate est synthétisé dans les plantes via la voie métabolique des octadécanoïdes et il est similaire aux prostaglandines animales au niveau de sa structure et de sa biosynthèse. La première partie de sa synthèse est réalisée dans les chloroplastes. Après libération de la

membrane plasmique par une phospholipase, l'acide α -linoléinique est oxygéné par une lipoxygénase pour former l'acide 13(S)-hydroxy linoléinique. Ce dernier est converti en acide 12-oxo-phytodiénoïque (OPDA) par l'allène oxyde synthase et l'allène oxyde cyclase. Il est généralement admis que l'OPDA est transporté dans les péroxysomes, où il est réduit par l'OPDA réductase et oxydé pour former le jasmonate (Stintzi et Browse, 2000). La jasmonate carboxyl méthyltransférase (JMT) cytosolique produit finalement le méthyl jasmonate volatil (Seo *et al.*, 2001).

II.7 Régulation hormonale au cours du développement du fruit

Les hormones végétales sont des molécules produites par les plantes et qui influencent leur profil de développement. Elles sont retrouvées dans la plupart des cellules et tissus. La fonction des hormones est de contrôler les événements en envoyant un message chimique à la cellule afin de réaliser une action ; elles peuvent inhiber ou promouvoir l'activité cellulaire.

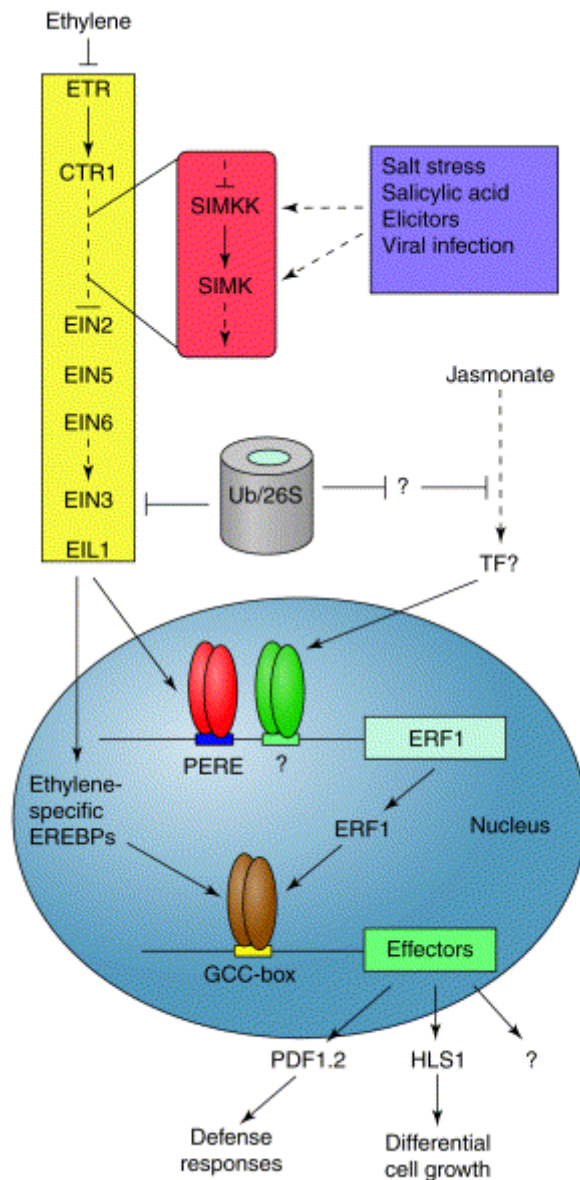
La plupart des hormones ont des effets multiples et agissent à faible concentration. Elles sont impliquées dans la division, l'élongation et la différenciation des cellules.

Les effets des hormones dépendent de la localisation et de la concentration de l'une par rapport aux autres dans un tissu spécifique. Elles fonctionnent souvent en coopération les unes avec les autres et en réponse aux *stimuli* environnementaux.

Nous allons nous focaliser ici, sur leurs modes de régulation et de fonctionnement et sur leurs actions pendant la maturation, en particulier vis à vis des critères de qualité du fruit.

II.7.1 Voie de perception et signalisation de l'éthylène chez Arabidopsis

Le décryptage des étapes impliquées dans la perception de l'éthylène et la transduction du signal a évolué très rapidement et aboutit à une vision de plus en plus complète chez *Arabidopsis* (Guo et Ecker, 2004 ; Fig. 6).



Current Opinion in Plant Biology

Figure 6 : Modèle pour la voie de transduction du signal éthylène et la régulation de l'expression des gènes (Guo et Ecker, 2004).

Chez *Arabidopsis*, l'éthylène peut être perçu par 5 récepteurs regroupés en 2 types. ETR1 et ERS1 (Type 1) possèdent un domaine histidine kinase fonctionnel tandis que pour ETR2, ERS2 et EIN4 (Type 2) ce domaine n'est pas fonctionnel. Cependant ce domaine ne semble pas requis pour la transmission du signal (Wang *et al.*, 2003). L'activité histidine kinase est probablement conservée pour d'autres fonctions impliquées dans la localisation et la stabilité de la protéine ou dans une régulation plus fine du signal. Le domaine histidine kinase dégénéré des récepteurs de type 2 se comporte probablement en tant que ser/thr kinases comme cela a pu être observé chez le tabac (Xie *et al.*, 2003). Les composés phosphorylables intervenant en aval des récepteurs à activité ser/thr sont à l'heure actuelle inconnus. Il est

supposé qu'ils interviennent dans la voie de transduction du signal, puisque ces récepteurs sont fonctionnels (Xie *et al.*, 2002 ; Tieman *et al.*, 2000).

ETR1 a été récemment localisé sur les membranes du réticulum endoplasmique (Chen *et al.*, 2002). En absence d'éthylène les récepteurs sont liés à une protéine CTR1 (Constitutive Triple Response 1). CTR1 est composé d'un domaine N-terminal de fonction inconnue mais qui est nécessaire à l'interaction avec le récepteur (Huang *et al.*, 2003). Le domaine C terminal est homologue aux MAPKKK Ralf. Bien que non requis pour interagir avec CTR1, les domaines histidine kinase des récepteurs de type 1 ont une meilleure affinité avec CTR1 que les domaines ser/thr kinase des récepteurs de type 2 (Gao *et al.*, 2003a). Chez les mutants dont la fonction CTR1 est perdue, une activation constitutive de la réponse à l'éthylène est observée. CRT1 exercerait donc un contrôle négatif lorsqu'il est lié au récepteur et serait inactif quand il en est dissocié (Kieber *et al.*, 1993 ; Huang *et al.*, 2003). L'activité kinase de CTR1 a été identifiée récemment (Huang *et al.*, 2003) et une cascade de MAP-Kinase a été décrite comme étant impliquée dans la réponse à l'éthylène en aval de CTR1 ; une MAPKK (SIMKK) active une MAPK (MAPK6), toutes deux sont induites par l'éthylène (Ouaked *et al.*, 2003). Le lien entre SIMKK et CTR1 reste cependant à démontrer de façon formelle. Ces derniers auteurs ont proposé cette hypothèse en observant le même phénotype en surexprimant SIMKK et en supprimant CTR1.

Le rôle de la protéine potentiellement suivante dans la cascade de transduction du signal, EIN2 est à l'heure actuelle obscur. Les données de génétique indiquent que EIN2 régulerait une étape essentielle de la propagation du signal, entre CTR1 et les facteurs de transcription de type EIN3/EIL (Roman *et al.*, 1995). EIN2 est situé en aval de MAPK6, voire indépendant, puisque les mutants *ein2* ont une activité MAPK6 équivalente au sauvage (Ouaked *et al.*, 2003). Bien que la protéine ait été purifiée (Alonso *et al.*, 1999), ses modes d'activation et d'action sont toujours méconnus. D'après les observations faites sur les mutants *ein2*, il semblerait que la protéine soit un point de convergence dans les voies de signalisation de l'éthylène et de l'acide jasmonique (Alonso *et al.*, 1999). EIN2 agirait sur une série de facteurs de transcription, EIN3 et la famille EIL3 (Solano *et al.*, 1998). Cependant aucun gène EIN3/EIL3 identifié n'a été observé comme étant régulé au niveau du transcriptome par l'éthylène, et EIN3/EIL3 doit subir des modifications post-traductionnelles par protéolyse via 2 protéines F-box (EBF1 et EBF2) (Guo et Ecker, 2003 ; Yanagisawa *et al.*, 2003).

Les homodimères de EIN3/EIL3 se lient sur la région promotrice du facteur de transcription ERF1 (Solano *et al.*, 1998). ERF1 appartient à une famille multigénique

nommée EREBPS intervenant dans un large éventail de mécanismes de régulation (Riechmann et Meyerowitz, 1998). La protéine ERF1 induite va ensuite se lier à la région promotrice appelée GCC box des gènes de réponse à l'éthylène (Fujimoto *et al.*, 2000). ERF1 est aussi impliqué dans la transmission du signal de l'acide jasmonique (Lorenzo *et al.*, 2003) et la région promotrice contenant la GCC box d'un gène de défense *PDF1.2* répond à l'acide jasmonique (Brown *et al.*, 2003). Ce mécanisme n'est pas clarifié.

II.7.2 Voie de signalisation de l'éthylène chez les fruits

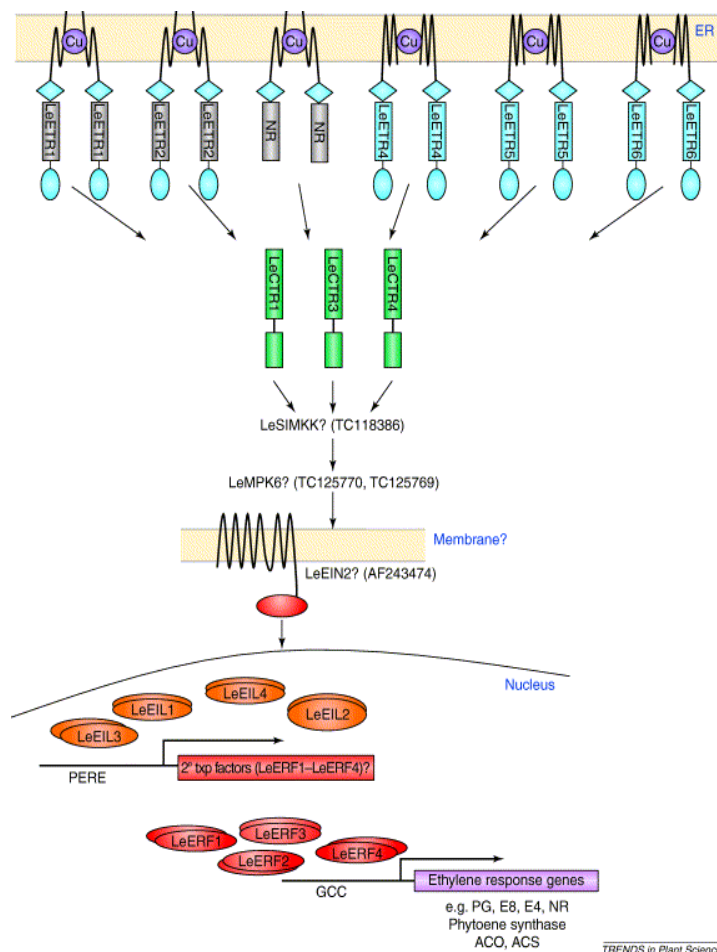


Figure 7 : Perception de l'éthylène et transduction du signal dans la tomate (d'après Adams-Phillips *et al.*, 2004a).

L'ensemble des gènes de la voie de transduction du signal éthylène identifié chez la tomate est présenté en figure 7. Parmi les récepteurs de l'éthylène, NR LeETR5 et LeETR4 sont fortement induits pendant la maturation (Payton *et al.*, 1996 ; Tieman et Klee, 1999). D'après les observations faites sur le mutant *nr* (never-ripe) les récepteurs agissent par

régulation négative de la voie de transduction comme cela a été décrit chez *Arabidopsis* (Klee, 2002). Contrairement à *Arabidopsis*, CTR est multigénique chez la tomate puisque 4 CTR ont été étudiés (Adams-Phillips *et al.*, 2004b). La spécificité des membres de la famille multigénique n'a pas encore été identifiée, mais la multigénicité n'est pas spécifique à la tomate, plusieurs CTR sont retrouvés chez l'orge et le riz (Adams-Phillips *et al.*, 2004b). LeSIMKK, LeMPK6 et LeEIN2 ont été trouvés dans les bases de données EST mais n'ont pas encore été caractérisés. Quatre ERF ont été identifiés et LeERF2 a une expression reliée à la maturation du fruit (Tournier *et al.*, 2003).

II.7.3 Effet de l'éthylène sur les critères de qualité des fruits

Les effets de l'éthylène sur la fermeté de l'abricot ont été mesurés et le blocage des récepteurs d'éthylène par le 1-méthylcyclopropène (1-MCP) retarde le ramollissement (Botondi *et al.*, 2003). Le traitement modifie également l'activité PME chez certaines variétés (Botondi *et al.*, 2003). Le traitement par l'éthylène inhibe la transcription d'une PME fruit-spécifique chez la fraise qui est pourtant un fruit non climactérique ; le traitement à l'auxine l'induit (Castillejo *et al.*, 2004).

- L'éthylène régule la transcription de 2 gènes de PG chez la poire mais n'a pas d'effet sur la transcription de 2 hémicellulases (Hiwasa *et al.*, 2003a).

- Une seule protéine a été identifiée comme étant à la fois impliquée dans la réponse à l'éthylène et dans la perte de fermeté au cours de la maturation (Lu *et al.*, 2001). Elle présente une homologie avec la classe de petites GTPases rab11/YPT3. Son expression est réduite dans le mutant de tomate *nr*, indiquant sa possible liaison avec l'éthylène. Les fruits des transformants antisens sont significativement plus fermes que les sauvages et les activités PG et PME sont diminuées. Les fruits présentent aussi des stries jaunes indiquant un possible effet sur la couleur. Son rôle exact reste à déterminer, même si l'analogie avec le règne animal suggère que l'enzyme puisse être impliquée dans le transport des protéines.

L'éthylène semble être impliqué dans 2 mécanismes liés au changement de couleur pendant la maturation : la dégradation des chlorophylles et la synthèse des pigments caroténoïdes et anthocyanes.

La dégradation des chlorophylles par la chlorophyllase *a* est sous contrôle de l'éthylène (Trebtsch *et al.*, 1993). D'autres hormones sont impliquées dans la régulation de la

chlorophyllase, comme le méthyljasmonate, inducteur chez *Arabidopsis* (Tsuchiya *et al.*, 1999). Chez l'orange, non climactérique, l'éthylène induit aussi l'expression de la chlorophyllase tandis que la gibbérelline en réduit les effets (Jacob-Wilk *et al.*, 1999).

Selon la nature du fruit, la synthèse des caroténoïdes semble mettre en jeu des voies métaboliques dépendantes ou indépendantes de l'éthylène. Chez le melon, la synthèse de caroténoïdes précède le dégagement d'éthylène (Karvouni *et al.*, 1995) et l'expression antisens de l'ACC oxydase ne modifie pas la composition en caroténoïdes (Guis *et al.*, 1997). Dans le cas de la tomate, les résultats restent contradictoires : certains auteurs ont montré que le blocage de la synthèse d'éthylène induit une sévère inhibition de la biosynthèse des caroténoïdes, chez les transformant antisens de l'ACC oxydase (Murray *et al.*, 1993 ; Gray *et al.*, 1994), alors que ce n'est pas le cas dans d'autres expériences (Rombaldi *et al.*, 1996). Dans le mutant *cnr* (Colorless Never Ripe) jaune et qui produit peu d'éthylène (Thompson *et al.*, 1999), la phytoène synthase est absente à maturité, ce qui suppose une possible régulation par l'éthylène. Chez les fruits non climactériques, l'éthylène pourrait induire la transcription de gènes codant pour l'alcool déshydrogénase (Tesnière *et al.*, 2004) et des gènes liés à la biosynthèse des anthocyanes (El-Kereamy *et al.*, 2003).

II.7.4 Effet de l'ABA sur les critères de qualité des fruits

L'acide abscissique jouerait un rôle dans l'induction de la maturation à la fois chez les fruits climactériques et non climactériques (Palejwala *et al.*, 1985 ; Vendrell, 1985). L'effet de l'ABA sur les paramètres de la maturation serait, en partie, relié à l'éthylène (Riov *et al.*, 1990). Dans la banane, le traitement des fruits par l'ABA induit la synthèse d'éthylène et la maturation, et les fruits traités au 1-MCP, inhibiteur des récepteurs de l'éthylène, ne répondent plus à l'ABA ; cependant le traitement à l'ABA juste avant celui au 1-MCP permet aux fruits de mûrir (Jiang *et al.*, 2000). Il semble que l'ABA joue un rôle de coordinateur dans la crise climactérique.

Les effets de l'ABA sur la maturation sont donc principalement à relier à ceux de l'éthylène. Les interactions dans la transduction des signaux éthylène et ABA n'ont pas encore été étudiées chez les fruits, mais chez *Arabidopsis* l'ABA intervient sur la régulation des gènes de transduction du signal éthylénique EIN2 et ETR1 (Beaudoin *et al.*, 2000). Parmi les enzymes de la voie de synthèse de l'ABA, seul le rôle limitant de la NCED a été mis en

évidence dans sa régulation (Tan *et al.*, 1997), y compris dans le fruit (Chernys et Zeevaart, 2000).

II.7.5 Effet de l'auxine sur les critères de qualité des fruits

Un certain nombre de travaux rapportent l'incidence de l'auxine sur l'expression des gènes/protéines reliés à la perte de fermeté des fruits, principalement chez les fruits non climactériques comme la fraise. L'auxine réprime les endo-1-4- β glucanases (Harpster *et al.*, 1998 ; Trainotti *et al.*, 1999), des β -galactosidases (Trainotti *et al.*, 2001) et pectates lyases (Benitez-Burraco *et al.*, 2003), mais induit les pectine estérases (Castillejo *et al.*, 2004) et ne semble pas influencer l'expression de l'expansine homologue à celle déjà identifiée chez l'abricot (Civello *et al.*, 1999 ; Mbéguié-Mbéguié *et al.*, 2002). Il semblerait que chez la fraise, l'auxine réprime la plupart des gènes liés à la maturation, en particulier ceux qui interviennent dans la biosynthèse des anthocyanes pendant les stades pré-matures (Manning, 1998). L'application d'auxine exogène retarde la maturation du raisin (Davies *et al.*, 1997) ainsi celle des fruits climactériques, comme la tomate (Cohen *et al.*, 1996). Plusieurs études rapportent un contrôle positif de l'auxine sur l'expression de certaines ACS dans les organes végétatifs et les fruits jeunes de melon (Ishiki *et al.*, 2000) et de tomate (Yip *et al.*, 1992 ; Balbi et Lomax, 2003), mais cette hormone n'aurait que peu d'influence sur les ACS induites au pic climactérique (Balbi et Lomax, 2003). Chez la pêche il semble que l'auxine induise l'élargissement des cellules (Ohmiya, 2000).

II.7.6 Effet de la gibbérelline sur les critères de qualité des fruits

La gibbérelline (GA) joue un rôle dans la maturation des fruits en inhibant la dégradation des chlorophylles dans les *Citrus* (Trebitch *et al.*, 1993), la synthèse des caroténoïdes dans le kaki (Gross *et al.*, 1984) et des anthocyanes dans la pomme (Awad et Jager, 2002). Au niveau moléculaire, il semblerait que la gibbérelline inhibe l'expression de la phytoène synthase dans la clémentine (Alos *et al.*, 2003) et contrecarre l'effet de l'éthylène qui induit la transcription de la chlorophyllase dans l'orange (Jacob-Wilk *et al.*, 1999).

II.7.7 Effet du jasmonate sur les critères de qualité des fruits

Le jasmonate est relativement abondant dans les tissus reproductifs et pourrait jouer un rôle dans la maturation des fruits. Bien que la présence de jasmonate et de ses dérivés volatiles soit impliquée dans l'attraction des insectes pour la dissémination du pollen, il peut moduler d'autres aspects du développement du fruit, comme la composition en caroténoïdes. Le jasmonate stimule la maturation chez la pomme et la tomate par l'activation de l'ACC oxydase et par conséquent induit la production d'éthylène (Czapski et Saniewski, 1992). L'application de jasmonate exogène inhibe l'accumulation de lycopène et stimule celle de β -carotène chez la tomate (Saniewski et Czapski, 1983). Le jasmonate augmente la synthèse d'anthocyanes dans des cultures de cellules de vigne (Curtin *et al.*, 2003).

II.8 Outils de génomique fonctionnelle

La génomique fonctionnelle a pour but d'établir un lien entre l'expression des gènes et les fonctions (ou dysfonctions) dans les cellules/tissus/organes. Les techniques haut débit se caractérisent par des méthodes expérimentales combinées à des analyses statistiques et informatiques des résultats. La stratégie fondamentale est de pouvoir étendre le champ d'investigation de l'étude d'un seul gène ou protéine à l'étude (en théorie) de tous les gènes ou protéines de manière systématique (Hieter et Boguski, 1997). Largement répandues en médecine et sur les espèces modèles, les approches haut débit font dorénavant leur apparition pour étudier les espèces d'intérêt agronomique et en particuliers, les fruits et leur développement.

Cet effort est particulièrement visible au niveau du séquençage d'EST, dont l'accélération laisse présager d'une explosion prochaine des résultats internationaux sur le transcriptome de nombreuses espèces fruitières. Le séquençage du génome du peuplier vient tout juste d'être achevé.

II.8.1 Les EST (Expressed Sequence Tag)

La technique de séquençage d'EST est apparue au début des années 90 chez l'humain (Adams *et al.*, 1991) et a ensuite été appliquée à un grand nombre d'espèces. On dénombre

actuellement, plus de 22 millions d'EST sur 740 espèces au NCBI (dbEST, 02/07/04, Boguski *et al.*, 1993). Les EST sont des séquences générées automatiquement et en simple lecture, à partir de banques d'ADNc (molécule d'ADN rétrotranscrite à partir d'une population d'ARNm). Les banques d'ADNc contiennent des dizaines de milliers de clones représentant une photographie instantanée de l'expression des gènes d'un tissu dans une condition physiologique précise. Le faible relatif coût du séquençage d'EST en fait le moyen le plus attractif pour avoir une description du transcriptome. Le séquençage d'EST à partir de l'extrémité 5' des ADNc clonés directionnellement a initialement été favorisé. En effet, les séquences 5' sont à même de contenir plus de région codante que celles obtenues à partir de l'extrémité 3' qui contient une partie significative de région non traduite (UTR). A l'heure actuelle, le séquençage à partir de l'extrémité 3' est également pris en compte car il permet d'obtenir plus de séquences uniques et peut être utilisé pour distinguer les gènes paralogues (Rudd, 2003).

Trois espèces fruitières ont fait l'objet d'un séquençage massif, la tomate (150000 séquences), la vigne (140000 séquences) et depuis le printemps 2004, la pomme (120000 séquences). Ces espèces occupent respectivement les 27, 29 et 32^{ème} places tous eucaryotes confondus, en notant cependant que la tomate n'est que la 12^{ème} espèce végétale en nombre d'EST. La caractérisation des banques d'EST de tomate a été réalisée par Van der Hoeven *et al.* (2002) dans un contexte global sans focalisation sur le fruit et la maturation. Pour la vigne l'étude de la dynamique du transcriptome pendant la maturation par l'approche EST a été réalisée sur un sous ensemble des séquences (Terrier *et al.*, 2001a).

Avec l'avancée de la biologie moléculaire computationnelle et des biostatistiques, il est possible d'exploiter et d'analyser à grande échelle des séquences d'EST efficacement et exhaustivement (c'est à dire le Northern électronique, Ewing *et al.*, 1999 ; Schmitt *et al.*, 1999).

II.8.2 Techniques d'étude de l'évolution du transcriptome

II.8.2.1 Techniques courantes et utilisation chez les fruits

Une variété de méthodes a été développée pour quantifier l'abondance des ARNm dans les tissus végétaux. Bien que les méthodes classiques et éprouvées du Northern blot et de la PCR quantitative soient sensibles et permettent une quantification précises de transcrits spécifiques, elle ne sont pas adaptées pour des criblages systématiques.

Le criblage différentiel permet de déterminer les gènes spécifiques d'un tissu par rapport à un autre, par hybridation de banques d'ADNc, et de ne séquencer que les transcrits soumis à une régulation transcriptionnelle. Il est cependant peu résolutif et ne permet de mettre en évidence que les transcrits abondamment exprimés. Cette technique a tout de même permis d'identifier des gènes différentiellement exprimés au cours du développement du raisin (Davies et Robinson, 2000) et de la fraise (Nam *et al.*, 1999).

Le differential display (Liang et Pardee, 1992) utilise une amplification par PCR peu stringente et avec des amorces combinatoires pour amplifier et visualiser sur gel des populations plus grandes d'ARNm. Cependant cette technique est peu quantitative et on retrouve une fraction importante de faux positifs. Cette technique a été utilisée chez la framboise pour identifier des gènes régulés pendant le développement (Jones *et al.*, 2000) et chez la tomate pour mettre en évidence les gènes régulés par l'éthylène (Zegzouti *et al.*, 1999).

La technique AFLP (Bachem *et al.*, 1996), variante du differential display a été utilisée chez la fraise pour identifier des gènes régulés pendant le développement (Martelli *et al.*, 2003).

La technique du SSH (Diatchenko *et al.*, 1996) permet d'identifier les gènes différentiellement exprimés entre deux tissus en éliminant par hybridation soustractive les transcrits communs. En théorie, le SSH permet de mettre en évidence des transcrits faiblement exprimés. Chez les plantes, il a principalement été utilisé pour les études d'interaction hôte / pathogène comme récemment chez le cacao (Verica *et al.*, 2004).

L'analyse en série de l'expression des gènes (SAGE) est une technique exhaustive qui combine le differential display et le séquençage d'ADNc (Velculescu *et al.*, 1995). Elle présente l'avantage d'être quantitative. Cependant cette technique est laborieuse et nécessite une bonne connaissance préalable du génome. Cette technique n'a donc pas encore été développée chez les fruits.

II.8.2.2 Les microarrays

Les microarrays (Schena *et al.*, 1995) tirent l'avantage des collections existantes d'EST et de séquences génomiques, pour comparer simultanément l'expression des gènes dans deux populations d'ARN d'origines différentes. Les lames de microarray sont composées d'une collection de séquences d'ADN (produits PCR, ADNc ou oligonucléotides), cataloguées et déposées sur une lame, généralement de verre. Les progrès technologiques effectués en terme de miniaturisation permettent d'étudier l'ensemble du transcriptome connu d'un organisme sur des surfaces de quelques centimètres carrés, contrairement aux macroarrays de plus faible densité.

Il existe deux types de microarray, les microarrays utilisant comme sonde¹ des ADNc et ceux utilisant des oligonucléotides.

Sur les lames à ADNc, des produits PCR sont fixés sur les lames de microarray. Cette technique présente des avantages par rapport au type oligonucléotide, en particulier le coût relativement faible de fabrication et la possibilité d'utiliser les lames d'une espèce sur une autre (Wang *et al.*, 2004).

Les microarrays à oligonucléotides sont composées de fragments courts d'ADN (couramment, entre 20 et 70 nucléotides) synthétisés *in vitro* puis greffés sur les lames, sauf pour la technologie Agilent où les oligonucléotides sont directement synthétisés sur la lame. Dans le cas de la technologie Affymetrix, entre 11 et 20 différents oligonucléotides 25mères par gène sont synthétisés *in situ* sur les lames. Il n'est donc pas nécessaire de disposer de banque d'ADNc pour pouvoir construire les lames. Le principal avantage de cette technologie est que les sondes peuvent être définies pour être spécifiques de la séquence d'un gène unique, afin de minimiser les risques d'hybridation croisée entre des gènes de la même famille multigénique. Évidemment en contrepartie, les lames ne sont utilisables que sur l'espèce concernée. La principale limitation dans le développement de microarrays de type oligonucléotide reste le coût.

L'étude du développement des fruits a pour l'instant été réalisée sur un nombre relativement restreint d'espèces et de gènes : chez la poire (1364 transcrits redondants, Fonseca *et al.*, 2004), le cacao (1380 transcrits uniques, Jones *et al.*, 2002) et la fraise (1700 transcrits redondants, Aharoni *et al.*, 2002) sur des lames à ADNc, ainsi que sur des lames à oligonucléotides chez la vigne (Terrier *et al.*, 2004, non publié).

¹D'après la nomenclature internationalement reconnue, la cible désigne les acides nucléiques à identifier (dans le cas des microarrays, elle est libre) et la sonde les acides nucléiques fixés et connus (Phimister, 1999).

II.8.3 Protéomique

Une stratégie alternative et complémentaire à l'étude du transcriptome des organismes est de cibler le lien moléculaire suivant dans la chaîne allant des gènes au phénotype : les protéines.

Le terme de protéomique recouvre l'ensemble des techniques d'étude des protéines (structure, interactions protéines/protéines, fonction et condition de biosynthèse). La technologie la plus répandue pour étudier les conditions dans lesquelles les protéines sont synthétisées est l'analyse par électrophorèse sur gels bidimensionnels (2D). Cette technique permet de séparer les protéines suivant leur point isoélectrique et leur poids moléculaire. Tout comme la technique de microarray, l'électrophorèse 2D peut être qualifiée de haut débit, puisqu'elle permet de visualiser environ les mille protéines les plus abondantes d'un tissu ou organite. Le couplage à une analyse par spectrométrie de masse permet d'identifier les protéines détectées. L'électrophorèse 2D pour l'étude à grande échelle est à l'heure actuelle couramment utilisée chez les espèces végétales depuis le milieu des années 90 (Kamo *et al.*, 1995). Cependant elle reste encore très marginale chez les fruits, puisqu'ici seuls les travaux récents sur le raisin (Sarry *et al.*, 2004) et la tomate (Saravanan et Rose, 2004) présentent des gels où au moins 100 spots sont détectables. Pour l'instant aucune étude sur le développement des fruits n'a été publiée. Il existe deux freins majeurs à ce type d'étude chez les fruits : (i) Cet organe est particulièrement contraignant puisqu'il contient une quantité conséquente de polysaccharides chargés solubles (les pectines) qui s'agrègent avec les protéines chargées positivement et interfèrent pendant la migration. (ii) Jusqu'à récemment, peu d'espèces fruitières ont bénéficié d'un séquençage massif et peu de protéines complètes sont connues, ce qui rend difficile l'identification des spots. Les EST représentent une alternative correcte pour l'identification des protéines, mais insuffisante (Gallardo *et al.*, 2003). Ce point sera abordé plus en détail dans le chapitre IV.4.

III Matériel & Méthodes

III.1 Matériel végétal

La récolte et les caractérisations physico-chimiques ont été effectuées par le plateau technique de caractérisation de la qualité des fruits de l'UMR SQPOV à Avignon.

III.1.1 Récolte et échantillonnage des fruits

III.1.1.1 Fruits utilisés pour la construction des banques ADNc

Les banques d'ADNc utilisées pour le séquençage d'EST ont été réalisées à partir de la variété Bergeron. Les abricots, cultivés en plein champ, ont été prélevés en 1996 dans le verger expérimental de l'INRA à Gotheron (Drôme). Les fruits ont été récoltés à 89, 107 et 117 jours après l'anthèse. Des lots de 4 à 30 fruits ont été constitués par distribution aléatoire. Après dosage du dégagement éthylénique et mesure de la fermeté, les fruits ont été immédiatement dénoyautés, découpés en petits morceaux, congelés dans l'azote liquide et conservés à -80 °C.

III.1.1.2 Fruits utilisés pour les expérimentations microarray, protéomique et PCR quantitative

Les expériences de microarray ont été réalisées à partir de 3 variétés, Bergeron, Monique, et une variété blanche issue d'un croisement entre Bergeron et Badami appelée Blanc.

Les abricots, cultivés en plein champ, ont été prélevés en 2002 dans le verger expérimental de l'INRA à Gotheron (Drôme). Les fruits des variétés Bergeron et Blanc ont été récoltés à 53, 65, 92, 107, 109, 114, 121, et 128 jours après anthèse. Les fruits des variétés Blanc ont été récoltés à 53, 65, 92, 99, 107, 114, 121, et 128 jours après anthèse. Les fruits de la variété Monique ont été récoltés à 53, 65, 72, 79, 86, 92, 94, 99 jours après anthèse.

III.1.2 Mesures des paramètres physico chimiques et physiologiques des fruits

Dès réception des prélèvements, 40 fruits sains sont sélectionnés et leur fermeté est déterminée par des tests de compression. Deux fois 6 fruits dont la valeur de fermeté est située de part et d'autre de la valeur médiane du lot sont retenus pour les analyses physiologiques et biochimiques. Les mesures physiologiques (dégagement d'éthylène, mesure de la fermeté...) sont réalisées sur chacun des 12 fruits retenus, immédiatement après la mesure de la couleur. Après ces mesures, les fruits sont dénoyautés, découpés en petits morceaux, congelés dans l'azote liquide et conservés à -80°C. Des sous-lots sont constitués à partir de ces 12 fruits, par classification hiérarchique en retenant les paramètres physiques et physiologiques les plus discriminants afin d'obtenir des lots homogènes. Toutes les mesures biochimiques sont réalisées ultérieurement sur des aliquotes de ces sous-lots pour chaque prélèvement de la campagne 2002. Les mesures sur les abricots de la campagne 1996 ont aussi été effectuées, mais certains protocoles ont été révisés (fermeté,...). Les anciens protocoles ne sont pas décrits ici.

III.1.2.1 Mesure de la fermeté

La fermeté est déterminée au moyen du pénétromètre PENELAUP (Abbal et Planton, 1990) d'après le protocole de Grotte *et al.* (2001). La mesure consiste à déterminer la pression (en Pa) nécessaire pour comprimer le fruit de 3 % de sa hauteur équatoriale entre deux plaques parallèles.

III.1.2.2 Mesure de la couleur

La couleur de chaque fruit est évaluée à l'aide d'un chromamètre Minolta CM-1000R (Minolta, Ramsey, N.J.), suivant l'espace colorimétrique L*a*b* (CIE, 1976). Les valeurs de L*, a* et b* correspondent respectivement à la clarté et aux axes vert-rouge et jaune-bleu.

III.1.2.3 Mesure du dégagement éthylénique

Le dosage est réalisé selon le protocole décrit par Chambroy *et al.* (1995). Les fruits sont placés individuellement dans des bocaux de volume connu et confinés pendant 1h30. L'éthylène émis à 23°C est dosé par chromatographie en phase gazeuse. Le CPG(IGC 121FL) est muni d'une colonne Porapak Q (3 m, gaz vecteur azote) à 110°C et est équipé d'un détecteur à ionisation de flamme. La production d'éthylène est calculée grâce à un étalon externe (10 ppm dans N₂) et exprimée en nmole par kilogramme de fruit et par heure.

III.1.2.4 Analyses biochimiques

Les analyses biochimiques sont effectuées sur une poudre homogène (broyeur à billes, azote liquide).

III.1.2.4.1 Indice réfractométrique et acidité titrable

L'indice réfractométrique est déterminé à l'aide d'un réfractomètre digital (PR-101, ATAGO). Il représente la matière sèche soluble présente dans le jus et est corrélé à la teneur en sucres. Les concentrations sont exprimées en % Brix à 20°C.

L'acidité titrable est obtenue par titration d'une aliquote de mésocarpe par de la soude 0,1 N jusqu'à pH 8,1 à l'aide d'un pHmètre. Les résultats sont exprimés en milliequivalent (méq) par g de matière fraîche.

III.1.2.4.2 Dosage des sucres et des anions organiques

Les différents sucres (glucose, fructose et saccharose) et les anions organiques malate et citrate sont déterminés par des dosages enzymatiques (Roche, Mannheim, Germany) à l'aide d'un analyseur multiparamètre automatique (BM-Hitachi 704). Les sucres sont exprimés en mg par g de matière fraîche et les acides en méq par g de matière fraîche.

III.1.2.4.3 Dosage des pigments

L'extraction des chlorophylles est réalisée en mélangeant 5 g de poudre dans 20 ml d'acétone à 80 % pendant 1 min. L'homogénat est centrifugé (10 min, 7000g, 5 °C). La teneur en chlorophylles totales est estimée à partir de l'absorbance lue à 665 nm.

Les carotènes sont extraits en homogénéisant 10 g de poudre dans 50 ml d'un mélange hexane / acétone / éthanol (2/1/1, v/v/v). 250 µl d'une solution de β-apo-8'-caroténal utilisé comme étalon interne sont ajoutés. L'ensemble est agité à l'abri de la lumière pendant 30 min,

puis filtré sur coton de verre. La phase aqueuse inférieure est éliminée par 3 lavages successifs à l'eau (20 ml). La phase hexanique est séchée et évaporée à 30°C sous vide partiel jusqu'à l'obtention d'un résidu sec. Le résidu est dissous dans 5 ml d'acétone puis l'extrait ainsi obtenu est analysé par HPLC couplée à une barrette de diodes (colonne Vydac 201TP54 RPC₁₈, éluant isocratique acétonitrile/méthanol/dichlorométhane (60/38/2, v/v/v), débit 1 ml/min, longueurs d'ondes : 290 (phytoène), 350 (phytofluène), 400 (z-carotène), 450 (β-carotène), 470 nm (étalon interne)).

III.1.2.4.4 Analyse des arômes

Cinquante grammes de morceau de péricarpe sont broyés et homogénéisés dans 50 ml d'eau et 30 µg de 4-nonanol (étalon interne) sont ajoutés. L'ensemble est centrifugé (30 min, 8000 g, 4 °C) et filtré sur toile de nylon. Après 3 extractions successives avec 10 ml de dichlorométhane, la phase organique est concentrée sur colonne de Vigreux puis sous flux d'azote jusqu'à 500 µl. Un µl est injecté sur un chromatographe en phase gazeuse (colonne capillaire CP-WAX ; gaz vecteur : hélium ; débit : 1,7 ml/min) couplé à un spectromètre de masse (Shimadzu QP2010). Les spectres majoritaires sont comparés avec ceux de la banque de spectres NIST (National Institute of Standard and Technologies) et les analyses statistiques sont réalisées avec les logiciels SYSTAT et Excel STAT.

III.2 Méthodes de biochimie relatives à la protéomique

III.2.1 Électrophorèse bidimensionnelle

L'électrophorèse bidimensionnelle décrite par O'Farrell (1975) permet de séparer les protéines après leur extraction des tissus, dans un gel de polyacrylamide selon 2 dimensions. La première dimension sépare les protéines en fonction de leur point isoélectrique (pI) grâce à un gradient de pH formé par des ampholines chargées. Cette étape est appelée focalisation isoélectrique (IEF). La seconde dimension permet la séparation des protéines en fonction de leur masse moléculaire par une électrophorèse SDS-PAGE.

III.2.1.1 Préparation des extraits de protéines totales

Le protocole utilisé est une adaptation de celui de Sarry *et al.* (2004), par modification du tampon d'extraction, permettant de travailler sur du matériel végétal congelé.

Cinq grammes de poudre congelée à -80 °C sont homogénéisés dans 15 ml de tampon contenant : MES 20 mM, méta-bisulfite 50 mM, EDTA 5 mM, K₂SO₄ 200 mM, DTT 20 mM, β-mercaptoéthanol 28 mM et PMSF 1 mM.

Pour chaque stade de développement, le pH est normalisé à 6,1 par de la soude. Après centrifugation (4000 g, 30 min, 4 °C), le surnageant est précipité avec du TCA 12,5 % (p/v) pendant 2 h à 4 °C et centrifugé (10000 g, 30 min, 4 °C). Le culot est ensuite lavé deux fois avec de l'acétone et une fois avec de l'éthanol, additionnés tous deux de 30 mM de β-mercaptoéthanol. Le culot est remis en suspension dans 2 ml du tampon de solubilisation d'IEF contenant : thiourée 2 M, urée 7 M, CHAPS 4 % (p/v), TritonX-100 0,5 % (p/v), ampholines 0,2 % (p/v) pH 3-10 (Amersham, Buckinghamshire, UK) et DTT 65 mM. L'échantillon est ensuite congelé et décongelé pour faciliter la solubilisation des protéines, puis centrifugé à 200000 g pendant 45 min pour éliminer les polymères insolubles (Rabilloud, 1999). La concentration en protéines est déterminée suivant la méthode de Bradford (1976) en utilisant la sérum albumine bovine comme étalon.

III.2.1.2 Première dimension IEF

La première dimension est une isoélectrofocalisation avec un gradient non linéaire d'immobilines (pH 3-10) fixées sur le gel du strip 18 cm (Amersham, Buckinghamshire, UK).

Une réhydratation passive du gel est effectuée en présence de l'échantillon protéique (350 µl de tampon contenant 700 µg de protéines) pendant une nuit. Après la réhydratation du gel, les protéines migrent et focalisent sous 3500 volt jusqu'à l'équilibre (90000 volt.heures au minimum) en utilisant une cuve Multiphor II (LKB, Amersham, Buckinghamshire, UK) à 18 °C. La migration est effectuée selon le programme suivant : 1 h de 0 à 300 V ; 1 h à 300 V ; 1h de 300 V à 600 V ; 1 h à 600 V ; 3 h de 600 V à 3500 V ; 27 h à 3500 V. Après cette étape les strips peuvent être conservés à -20 °C.

Les strips sont ensuite équilibrés avec du tampon d'équilibration (Tris-HCl 50 mM, pH 8,8 ; Urée 6 M, Glycerol 30 % (p/v), SDS 2 % (p/v)) contenant 2 % (p/v) de DTT pour la réduction des ponts disulfures pendant 20 min sous agitation. Pour l'alkylation des groupement carboxyméthyl des cystéines, le même tampon est utilisé avec 2,5 % (p/v) d'iodoacétamide à la place du DTT, l'agitation est effectuée pendant 20 min. Ces étapes

permettent, en réduisant les ponts disulfures, d'améliorer la migration en deuxième dimension et l'efficacité de la digestion par la trypsine.

III.2.1.3 Deuxième dimension

La deuxième dimension permet de séparer les protéines ayant une masse moléculaire comprise entre environ 15 kDa et 120 kDa. Le gel (24 cm x 20 cm) est coulé entre deux plaques de verre espacées d'1,5 mm. Le gel est composé d'acrylamide 11 % (p/v), bisacrylamide 0,3 % (p/v), Tris-HCl 375 mM, pH 8,2, SDS 1 % (p/v), TEMED 0,05 % (p/v), persulfate d'ammonium 0,05 % (p/v) (Hames et Rickwood, 1990). Les strips sont ensuite déposés au sommet des gels de deuxième dimension et scellés avec une solution d'agarose (agarose 0,5 % (p/v), agarose à bas point de fusion 0,5 % (p/v), SDS 0,2 % (p/v), Tris 25 mM, trace de bleu de bromophénol).

Les gels sont placés dans un tampon d'électrophorèse constitué de Tris 25 mM, glycine 0,193 M et SDS 0,1 % (p/v). La migration est effectuée selon le programme suivant : 2 h à 40 V, 3 h à 70 V et 11 h à 110 V dans une cuve d'électrophorèse ISODALT (Amersham, Buckinghamshire, UK). Les protéines sont colorées au bleu colloïdal d'après le protocole de Neuhoff *et al.* (1988) modifié. Elles sont tout d'abord fixées en immergeant les gels dans un tampon de fixation (éthanol 50 % (v/v), acide phosphorique 1,7 % (v/v)) pendant 2 h. Les gels sont ensuite lavés 3 fois une demi heure dans de l'eau, puis incubés une heure dans le tampon d'incubation (méthanol 34 % (v/v), sulfate d'ammonium 17 % (p/v) et acide phosphorique (v/v) 1,7 %), pour faire précipiter les protéines dans le gel. Les gels sont enfin colorés dans le tampon de coloration (méthanol 34 % (v/v), sulfate d'ammonium 17 % (p/v), acide phosphorique 1,7 % (p/v), bleu brillant de Coomassie G250 (Serva, Heidelberg, Allemagne) 0,05 % (p/v)) pendant 2 à 3 jours, jusqu'à saturation de la coloration.

Les gels sont lavés à l'eau avant numérisation et conservés ensachés dans une solution d'acide acétique 5 % (p/v) à +4 °C.

Pour chaque stade de développement, au moins trois réplicats ont été effectués. Afin de ne pas induire un biais d'abondance des protéines lié au protocole, les extractions et les séparations électrophorétiques pour chaque stade n'ont jamais été réalisées simultanément. Les expériences ont été réalisées par série de 8 extractions/électrophorèses correspondant à des stades différents.

III.2.1.4 Analyse d'image

Les images des gels ont été digitalisées à 300 dpi avec un densitomètre GS 710 (Bio-Rad, Hercules, CA, USA). Elles sont analysées avec le logiciel Progenesis (Nonlinear Dynamics, Newcastle upon Tyne, UK).

La détection des spots est automatique, mais la délimitation de leur contour peut se rectifier manuellement dans le cas de spots rapprochés. Le logiciel déforme l'image de chaque gel de manière à pouvoir superposer l'ensemble des spots et ainsi comparer les gels.

Pour l'analyse quantitative des spots, les volumes des spots sont déterminés d'après leur aire et l'intensité de chaque pixel de la surface. Les volumes sont normalisés en divisant le volume de chaque spot par le volume total des spots sur le gel. L'abondance des spots est donc exprimée en pourcentage du volume total des spots du même gel.

Les trois répliques de chaque stade sont tout d'abord comparées entre elles pour obtenir un gel moyen comprenant l'ensemble des spots visibles d'un même stade, l'écart type est conservé. Les gels moyens de chaque stade sont ensuite comparés pour obtenir le profil de chaque spot aux cours du développement du fruit. Dans un premier temps les spots ont été considérés comme évoluant s'il existait un facteur de variation entre deux stades, au moins supérieur à deux, et si la valeur de l'abondance moyenne du plus faible spot plus son écart type ne soit pas supérieure à celle du plus fort moins son écart type. Ensuite un test de Student est appliqué entre les deux stades dont les intensités sont les plus éloignées. Les spots ayant une P value au moins supérieure à 90 % (90 % de chance que les stades soient significativement différents) sont conservés pour l'analyse au MALDI-TOF. Le test de Student n'a donc pas été effectué directement sur tous les spots afin d'accélérer la procédure.

III.2.2 Spectrométrie de masse

Les analyses de spectrométrie de masse ont pour but de réaliser l'identification des protéines correspondant aux spots présents sur les gels d'électrophorèse 2D. Deux techniques différentes ont été utilisées. La première technique, le MALDI-TOF (MS) (Désorption-Ionisation Laser Assistée par Matrice - Temps de Vol), consiste à déterminer les masses des peptides obtenus après hydrolyse d'une protéine par la trypsine. La seconde, le MALDI-TOF-TOF (MS/MS) consiste à fragmenter un peptide isolé lors de la première analyse et d'obtenir sa carte massique. A l'issue de ces analyses, des spectres de masse sont obtenus pour chaque

protéine (MS) ou fragment peptidique (MS/MS) représentant l'ensemble des masses (exprimées en masse sur charge, m/z) correspondant aux peptides présents dans une protéine (MS) ou aux différents fragments d'un peptide (MS/MS). Par exemple pour le peptide GAK, on peut théoriquement retrouver les masses correspondant aux acides aminés G, A, K et celles des peptides GA, AK et GAK.

III.2.2.1 Préparation des spots protéiques

Afin d'éviter une contamination par les kératines humaines, des précautions doivent être prises au cours de l'excision et de la digestion (gants, masque, bonnet).

Le protocole présenté ici est celui utilisé pour fournir les échantillons sur la plate-forme protéomique du CCIPE (CNRS/INSERM, Montpellier) pour les analyses par MALDI-TOF-TOF (MS/MS). Les échantillons fournis à la plate-forme de l'INRA de Montpellier pour les analyses par MALDI-TOF (MS) sont uniquement excisés des gels et lavés, les étapes suivantes étant robotisées sur place.

III.2.2.1.1 Excision et lavage des spots

Les spots protéiques sélectionnés pour l'identification, sont excisés des gels en utilisant comme emporte-pièce un cône de pipette de 1 ml tronqué de 1 à 2 mm de diamètre. Chaque morceau de gel est placé dans un tube de 1,5 ml auquel on rajoute 200 μ l d'eau, pour conservation. Afin de faciliter l'élimination du bleu de Coomassie, les morceaux de gels sont fragmentés sur une lame de verre avec un scalpel (taille finale des fragments $< 1 \text{ mm}^3$). Le bleu de Coomassie est ensuite éliminé par 5 cycles successifs de réhydratation (NH_4HCO_3 10 mM 10 min) / déshydratation (acétonitrile pur 10 min). Les morceaux de gels décolorés sont séchés au speed vac.

III.2.2.1.2 Digestion trypsique

La solution de trypsine (trypsine 0,015 mg/ml (Sequencing grade, Promega, Charbonnières, France), NH_4HCO_3 85 mM, CaCl_2 4,25 mM) est préparée extemporanément à 0 °C. Dix μ l sont ajoutés sur les morceaux de gel puis incubés sur de la glace pendant 45 min. Dix μ l de NH_4HCO_3 100 mM, CaCl_2 5 mM, sont ajoutés pour recouvrir les gels (environ 10 μ l). La digestion enzymatique est ensuite réalisée à 25 °C pendant 16 h.

III.2.2.1.3 Extraction des peptides

Pour extraire du morceau de gel, les peptides digérés qui ne sont théoriquement pas exsudés, 2 cycles de réhydratation / déshydratation sont réalisés comme lors de l'étape de décoloration (chapitre III.2.2.1.1) en regroupant tous les surnageants. Vingt μl d'acide formique 5 % (v/v) sont ajoutés et mis à incuber à température ambiante pendant 10 min. Vingt μl d'acétonitrile sont ajoutés et le surnageant est rassemblé avec les surnageants précédents, cette étape est répétée une fois. L'ensemble des surnageants est évaporé au speed vac.

III.2.2.1.4 Elimination des sels

Afin d'éliminer les composants pouvant interférer dans l'analyse par spectrométrie de masse, les peptides sont dessalés et concentrés jusqu'à un volume final de 3 μl avec un kit ZipTip (Millipore, Bedford MA, USA) suivant le protocole du fabricant.

III.2.2.2 Analyse par spectrométrie de masse MALDI-TOF (MS) et MALDI-TOF-TOF (MS/MS)

Les analyses MS ont été effectuées à la plate-forme protéomique de l'INRA de Montpellier avec un spectromètre de masse BIFLEX III (Brucker Daltonics, Bremen, Allemagne). Les protéines non identifiées ont été re-analysées par MS sur la plate-forme protéomique CNRS/INSERM de Montpellier. Si l'identification n'était toujours pas réalisée, une analyse par MS/MS est effectuée. L'appareil utilisé est un ULTRAFLEX (Brucker Daltonics, Bremen, Allemagne).

Les spectres sont calibrés en utilisant les pics à m/z 842,510 et m/z 2211,105, correspondant à 2 pics issus de l'autolyse de la trypsine. L'assignation des masses monoisotopiques aux différents fragments est réalisée en utilisant le logiciel SNAP (Brucker Daltonics, Bremen, Allemagne) et vérifiée manuellement. Au CNRS, elle a été réalisée par le personnel technique de la plate-forme.

III.2.2.3 Identification des peptides

L'identification des peptides est réalisée en utilisant le moteur de recherche MASCOT (www.matrixsciences.com) dont une version résidente est disponible sur le campus INRA. L'interrogation a été effectuée sur une base de données dont les spécificités sont détaillées dans le chapitre IV.4. Les paramètres suivants ont été utilisés pour l'identification :

- une tolérance d'erreur de 100 ppm ;
- un clivage manqué accepté ;
- au moins quatre peptides retrouvés dans la protéine ;
- un minimum de 15 % de la séquence de la protéine identifié ;
- les carbamidiméthylations des cystéines et l'oxydation des méthionines, histidines et tryptophanes sont tolérées ;
- la probabilité d'identification due au hasard doit être inférieure à 0,05 (Pappin *et al.*, 1993).

III.3 Méthodes de biologie moléculaire

III.3.1 Préparation des banques d'ADNc

La préparation des banques d'ADNc a été réalisée au cours de la thèse de Didier Mbéguié-Mbéguié (2000).

Les ARN totaux ont été extraits de la pulpe des fruits de la variété Bergeron d'après la méthode modifiée de Wan et Wilkins (1994) (Mbéguié-Mbéguié *et al.*, 1999). Sept grammes de tissus congelés dans l'azote liquide sont mélangés avec 0,46 g d'acide diéthylthiocarbamique. Le mélange est homogénéisé (10 min à 65 °C) dans 12 ml de tampon d'extraction préchauffé (glycine 0,2 M, acide borique 0,2 M, et NaCl 0,6 M, pH 9,6) ainsi que du SDS 5 %, du Nonidet P-40 1 %, 1 g de PVPP (pré-homogénéisé dans le tampon d'extraction) et 5 % (v/v) de β -mercaptoéthanol. Douze ml de phénol (pH 8,0, équilibré avec du Tris-HCl, température ambiante) sont ensuite ajoutés à l'extrait, qui a été homogénéisé à 65 °C pendant 10 minutes supplémentaires. Après centrifugation de l'échantillon (20 min, 13000 g), la phase aqueuse est amenée à une concentration finale de 160 mM de chlorure de potassium et laissée à précipiter une nuit avec de l'éthanol à -20 °C, puis purifiée sur cellulose CC41 (Whatman, Kent, UK) d'après la méthode de Fils-Lycaon *et al.* (1996). Les ARN riches en poly-(A+) sont purifiés avec le polyAtract mRNA Isolation System (Promega, Madison, USA). Les banques d'ADNc sont construites à l'aide du kit « Zap-cDNA synthesis®Kit » (Stratagene, La Jolla, USA) selon le protocole décrit par le fabricant à partir de 5 μ g d'ARNm. Les ADNc double brin synthétisés sont ensuite clonés de manière unidirectionnelle dans le

vecteur UNIZAP-XR aux sites de restriction EcoR1-Xho1. Les vecteurs sont encapsidés *in vitro*. Les plasmides sont excisés avec le système Exassist (Stratagene, La Jolla, USA).

III.3.2 Production des séquences d'EST

Le repiquage des colonies, la purification des plasmides et le séquençage des inserts d'ADNc ont été réalisés par un prestataire de services (MWG-Biotech, Ebersberg, Allemagne). Les amorces de séquençage T3 ou T7 furent utilisées pour pouvoir lire à partir de l'extrémité 5' ou 3' des ADNc respectivement.

III.3.3 Extraction des ARN

L'extraction des ARN totaux de mésocarpe et épicarpe d'abricot pour l'étude de microarray et PCR quantitative est réalisée selon le protocole de Tesnière et Vayda (1991). La vaisselle et l'eau bi-distillée utilisées sont traitées au DEPC et autoclavées. Pour chaque échantillon, 5 g de poudre stockée à -80 °C sont décongelés en présence de 20 ml de tampon d'extraction (Tris 200 mM, LiCl 300 mM, EDTA 10 mM, déoxycholate de sodium 1 % (p/v), SDS 1,5 % (p/v), ATA 1 mM, thiourée 5 mM, DTT 10 mM et NP-40 1 % (v/v), le pH est ajusté à 8,5 avant autoclave). Après centrifugation à 11000 g (30 min, 4 °C), les surnageants sont filtrés sur du tissu de type Miracloth. Après ajout de CsCl (0,2 g.ml⁻¹), ils sont déposés sur un coussin de 10 ml composé de CsCl 5,7 M, Tris-HCl 10 mM, EDTA 10 mM. L'ultracentrifugation est effectuée durant 24 h à 25000 g à 20 °C. Les culots d'ARN totaux présents au fond du tube sont solubilisés dans 1 ml d'eau traitée au DEPC. Les ARN sont précipités par 0,2 volume de LiCl 3 M et 2,5 volumes d'éthanol absolu pendant 1 h à -80 °C. Les culots recueillis après centrifugation à 11000 g (4 °C, 30 min) sont lavés trois fois par 500 µl d'acétate de sodium 3 M, pH 5,5, et centrifugé à 12000 g (4 °C, 15 min). Ils sont ensuite lavés avec 500 µl d'éthanol 80 % (v/v) afin d'éliminer les sels. Les surnageants sont éliminés et les culots séchés sous vide. Les ARN sont remis en suspension dans de l'eau DEPC.

III.3.4 Préparation d'ADN plasmidique

Les minipréparations d'ADN plasmidique pour les analyses rapides de plasmides issus de clones bactériens sont réalisées par le protocole de lyse alcaline décrit par Sambrook *et al.* (1989). Pour cette technique, les bactéries sont cultivées dans un petit volume de milieu sélectif (LB (Bactotryptone, extrait de levure, et NaCl) + ampicilline 0,1 % (v/v)), dont une aliquote de culture (1,5 ml) est centrifugé (12000 g, 5 min). Le culot bactérien est repris dans 100 µl d'une solution de SDS 1 % (p/v) et NaOH 0,2 M. L'ADN bactérien, de haut poids moléculaire, et les protéines sont éliminés par précipitation par 150 µl d'acétate de potassium 5 M pH 4,8. Une extraction au phénol/chloroforme/alcool isoamylique (25/24/1, v/v/v) permet d'éliminer les protéines résiduelles. L'ADN plasmidique, contenu dans la phase aqueuse, est précipité par ajout de 2 volumes d'éthanol. Après centrifugation (12000 g, 15 min), le culot est séché et repris dans 30 µl de tampon TE contenant de la RNase (50 µg.ml⁻¹).

III.3.5 Test d'hybridation/ southern blot

Après extraction, les plasmides sont digérés par XhoI et EcoRI puis séparés par électrophorèse sur un gel d'agarose à bas point de fusion, Nusieve 0,8 % (FMC BioProducts, Rockland, USA). Les inserts sont purifiés avec le kit JETSORB (Genomed, Löhne, Allemagne). Deux ng d'insert ainsi purifiés sont séparés par électrophorèse sur gel d'agarose 1 % puis transférés sur membrane Hybond-N (Amersham, Buckinghamshire, UK) pendant 16 heures. Après transfert, les ADN sont fixés à la membrane à 80 °C pendant 2 heures.

Quarante ng d'insert sont marqués au hasard avec de l'³²PdCTP à l'aide du DNA-labeling system (Megaprime, Amersham, Buckinghamshire, UK). La préhybridation (4 h) et l'hybridation sont réalisées à 65 °C dans 5 ml de SSC 5 X, Denhardt 5 X, SDS 0,5 % (p/v) et 0,1 mg/ml d'ADN de sperme de saumon. Après une nuit d'hybridation avec les sondes, les membranes sont lavées deux fois pendant 20 min dans du SSC 2 X et SDS 0,1 % (p/v) à 65 °C (condition de haute stringence) et puis lavées pendant 5 min avec du SSC 0,1 X à température ambiante.

Après lavage, les membranes sont exposées sur un film Kodak XAR-5 à -80 °C. Le film est révélé pendant 5 min dans le bain de révélation (KODAK, LX24), rincées 30 secondes dans de l'eau et 2 à 4 min dans le bain de fixation (KODAK AL4).

La quantification des signaux obtenus sur les membranes d'hybridation est effectuée à l'aide d'un phospho-imageur (STORM, Amersham, Buckinghamshire, UK).

III.3.6 Synthèse du premier brin d'ADNc pour PCR quantitative

Les ARN totaux sont purifiés et traités avec le kit high pure RNA tissue (Roche, Mannheim, Allemagne). Deux μg d'ARN et 1 μg d'amorce random hexamer sont dénaturés à 65 °C pendant 10 min et immédiatement refroidis sur glace. Le premier brin d'ADNc est synthétisé dans un mélange réactionnel de 40 μl contenant du DTT 10 mM, dNTPs 1 mM, 100 U d'Extend Reverse Transcriptase (Roche, Mannheim, Allemagne) et du tampon Extend Reverse Transcriptase 1 X à 42 °C pendant 1 h. La réaction est arrêtée par dénaturation de l'enzyme à 95 °C pendant 5 min puis refroidie sur de la glace.

III.3.7 PCR quantitative

Les amorces ont été définies avec le logiciel Primer3 (Rozen et Skaletsky, 2000), en imposant les critères suivants :

- température d'accrochage de 60°C \pm 3 ;
- longueur de l'amorce de 20 pb \pm 2 ;
- taille de l'amplicon de 200 pb \pm 50 ;
- pourcentage de GC inférieur à 60 %.

Les séquences sens et antisens des différentes amorces sont reportées dans le tableau 1. Des amorces spécifiques d'ARN ribosomique 25S ont été utilisés comme contrôle endogène.

La PCR quantitative a été réalisée avec le kit QuantiTech SYBR Green PCR kit (Roche, Mannheim, Allemagne) et en utilisant l'appareil PCR Roche Light Cycler pour la mesure et le logiciel Light Cycler Data Analysis Software pour la quantification. Pour chaque réaction PCR, le Ct est mesuré, c'est à dire le nombre de cycle au delà duquel la quantité d'ADN amplifié dépasse le seuil fixé par l'opérateur. Une dilution en série du stade où à priori chaque transcrit est le plus abondant est effectuée et les Ct sont mesurés sur ces dilutions afin de déterminer l'efficacité de la réaction. Si l'efficacité est de 100 %, alors la quantité d'ADN double à chaque cycle de PCR, et l'équation $ct = f(\log[\text{concentration de la matrice}])$ doit donc être une droite de pente $-1/(\log(\text{efficacité}))$, sinon la sous-estimation est corrigée en

conséquence. Chaque réaction est effectuée en triplicat et la spécificité du produit de l'amplification est vérifiée par courbe de fusion. Les niveaux d'expression relatifs des transcrits sont calculés et normalisés par comparaison à l'abondance des ARN ribosomiques 25S.

Tableau 1 : Liste des séquences des amorces utilisées pour les expériences de PCR quantitative.

Nom	ID	Amorce sens	Amorce antisens	Taille de l'amplicon (en pb)
XET	PA17TC1	5'-GCAGTTTGTGGACTTG ATTG-3'	5'-TTAATCGTACATTGTTT GGTACAAC-3'	181
UDP-GP	PA91TC1	GTCCTGGAATGAATCGGAG A	TTATGCTCTGCCTTGGCA TT'	190
HSP18	PA12TC3	CTCAACTCCACCTCCCCTT T	CAGAACGAGAAACCGTA CATGA	162
Riche en Proline	PA1TC49	GCTCTCAATTTTAGCACT TGG	GGCAAGTAGCCTACCAT CCA	176
Inconnue	PA19TC3	ACAAGGATGTCAAGGGCA AG	GGACTGGAAAGCAACAG AACAA	176
Liant l'ADN	PA32TC1	TTTATTCGGGTCCTGTTTG G	TCAGTCCTCAAGGACAG CAA	173
Non identifiée	PA38TC1	AGCTGCTCATCATGATCTC AA	GCCCACTTTCCTTGTTCA AA	222
Non identifiée	PA3TC4	AGGGTTTGAGTCTTCGCTT AC	AAAACCCATACGACGCT GAA	150
Expansine	PA5TC5	GGGGAGAGAGGAGAAGGA GA	AATGGACGATTTCTGCG AAC	163
PMEI	PA5TC7	CAGAGCAAGGCATGAACG AC	CATGATTCAAAACACCA TCACA	180
Non identifiée	PA62TC1	TGCAACACATAACATGGA GGA	ATACCATTCCCACCAACC AA	225
Catalase	PA6TC3	CCGGTCGTATGTGACGATT T	CTGTACGAGCTTTTGCCA CA	159

III.3.8 Synthèse des ARN antisens pour hybridation sur microarray

L'amplification linéaire des ARN antisens et l'incorporation des fluorophores Cy3 et Cy5 sont réalisées à partir d'une version modifiée du protocole de transcription *in vitro* d'Eberwine *et al.* (1992) qui a déjà été utilisée et validée (Baugh *et al.*, 2001 ; Puskas *et al.*,

2002 ; Religio *et al.*, 2002). Les ARN totaux sont rétrotranscrits avec le kit cDNA Synthesis System (Roche, Mannheim, Allemagne). Vingt µg d'ARN sont incubés avec 20 µM d'amorce Oligo[(dT)₂₄T7promoteur]₆₅ et 0,6 ng d'ARNm de luciférase à 70 °C pendant 10 min puis placés sur de la glace. La synthèse du premier brin est ensuite réalisée dans 40 µl contenant du tampon reverse transcriptase 1 X, DTT 0,01 M, 50 U de reverse transcriptase AMV (Roche, Mannheim, Allemagne), 25 U d'inhibiteur de RNase et chaque dNTP à 1 mM. La solution est incubée à 42 °C pendant 60 min et placée sur de la glace. Pour la synthèse du second brin, sont rajoutés : du tampon second brin 1 X, 1,5 mM de chaque dNTP et le mix d'enzyme second brin 1 X. La solution est incubée 2 h à 16 °C puis placée sur de la glace. Vingt unités d'ADN T4 polymérase sont ajoutées (incubation 16 °C, 5 min). La réaction est arrêtée avec de l'EDTA 20 mM, pH 8. Quinze unités de RNAse I sont ensuite ajoutées (incubation 37 °C, 30 min) puis 3 U de Proteinase K (incubation 37 °C, 30 min). L'ADNc double brin est purifié avec le kit High Pure RNA Tissue Kit (Roche, Mannheim, Allemagne) suivant les instructions du fabricant (à l'exception de l'utilisation de la DNase) et dosé au spectrophotomètre. L'amplification linéaire à partir du promoteur T7 et l'incorporation simultanée d'UTP fluorescents (Amersham, Buckinghamshire, UK) sont réalisées avec le kit Mega Script T7 Kit (Ambion, Cambridgeshire, UK). La réaction est réalisée à partir de 200 ng d'ADNc séchés au speed vac dans une solution contenant : le tampon de réaction 1 X, ATP 7,5 mM, CTP 7,5 mM, GTP 7,5 mM, UTP 5 mM, le mix d'enzyme contenant la T7 polymérase 1 X et 1,25 mM de Cy3-UTP ou Cy5-UTP (incubation 37 °C, 16 h, à l'abri de la lumière). Les ARN marqués sont ensuite purifiés avec le kit High Pure RNA Tissue Kit (Roche, Mannheim, Allemagne) suivant les instructions du fabricant (à l'exception de l'utilisation de l'ADNase) et dosés au spectrophotomètre. Le taux d'incorporation des UTP fluorescents est évalué en mesurant la D.O. à 550 et 650 nm respectivement pour Cy3 et Cy5. Il correspond à la quantité de nucléotides marqué ($DO_{550} * Vol_{éch} * dilution / 0,15$) divisée par la quantité totale de nucléotides ($DO_{260} * Vol_{éch} * dilution * 40$).

III.3.9 Hybridation sur lame de microarray, lavages et acquisition de l'image

III.3.9.1 Hybridation sur lame de microarray et lavages

Dix µg de chaque ARN marqué avec les fluorophores (Cy3 et Cy5) sont rassemblés dans un volume de 48 µl et incubés à 94 °C pendant 15 min dans du tampon de fragmentation 1 X (Tris-Acétate 20 mM, pH 8,1, KOAc 50 mM, MgOAc 15 mM) pour fragmenter les ARN. Les ARN sont purifiés et concentrés dans des concentrateurs Microcon (Millipore, Bedford, USA) à 9000 g pendant 15 min. Les ARN sont séchés et repris dans 240 µl de tampon d'hybridation (Na-phosphate 50 mM pH 8,0, formamide 50 %, SSC 6 X, Denhardt 5 X, SDS 0,5 %, MWG Biotech, Ebersberg, Allemagne). L'échantillon est déposé par goutte sur la lame et recouvert d'une lamelle, de manière à le mettre en contact avec l'ensemble de la surface de lame. Celle-ci est ensuite placée dans un four à hybridation pendant 20 h à 42 °C.

Après hybridation, la lame est lavée dans des bains successifs sous agitation :

- 5 min avec une solution SSC 2 X, SDS 0,1 % (30 °C) ;
- 5 min avec une solution SSC 1 X (30 °C) ;
- 5 min avec une solution SSC 0,5 X (30 °C) ;
- 5 min avec une solution SSC 0,2 X (30 °C).

Toutes ces solutions sont préparées avec de l'eau DEPC et filtrées (0,45 µm)

La lame est ensuite séchée par centrifugation à 500 g pendant 2 min.

III.3.9.2 Acquisition des images

L'acquisition de la fluorescence des lames est réalisée à l'aide du scanner GenPix 4000B (Axon Instruments, Union City, USA) et la quantification de l'intensité des spots et des bruits de fond locaux dans les longueurs d'onde à 635 (Cy5) et 532 (Cy3) nm est effectuée sur un ordinateur équipé du logiciel GenPix Pro 3. Les photomultiplicateurs sont réglés de façon à ce que les signaux soient les plus élevés possibles sans qu'aucun spot n'atteigne des valeurs saturantes et que les signaux soient de même ordre de grandeur pour les deux fluorochromes.

III.4 Méthodes de bioinformatique

Les opérations de bioinformatique ont été réalisées sur microordinateur PIII, 667MHz, 250Mo sous les diverses versions du système d'exploitation MandrakeLinux de 8.1 à 9.2.

III.4.1 Édition des séquences

III.4.1.1 Base calling

Les fichiers contenant les données brutes de fluorescence en sortie de séquenceurs (trace files, nomfichier.abi) ont été analysés avec le logiciel Phred afin d'attribuer (i) les noms de bases, « base calling »¹(lexique chapitre III.4.7) et (ii) une probabilité d'erreur de lecture (Ewing *et al.*, 1998), indépendamment du traitement effectué par MWG (Ebersberg, Allemagne). Pour simplifier les notations, une échelle logarithmique est utilisée, suivant la définition $\text{score Phred}^2 = -10 \text{ Log}(p)$, où p représente la probabilité d'erreur d'annotation (Ewing et Green, 1998). Pour chaque séquence, les bases avec un score Phred inférieur à 20 sont rejetées, ce qui correspond à 1% de risque d'erreur d'annotation. Le score minimum de 20 correspond au standard international de qualité d'EST nécessaires pour être soumis dans GenBank.

III.4.1.2 Élimination du vecteur et des séquences courtes

Les séquences de vecteur résiduelles ont été repérées par homologie à la séquence du vecteur pBluescript SK-, puis éliminées avec le logiciel Cross Match (Green, 1994, non publié).

Celles inférieures à 100 pb sont ensuite rejetées. Les séquences restantes sont considérées comme valables si leur e-value³ de BLAST sur elle-même est supérieur à $3e^{-20}$. Cette e-value correspond à un alignement minimal de 100 pb, sans tenir compte des régions de faible complexité, comme la queue poly-A. Ce seuil a été fixé en fonction de l'impossibilité de construire des oligonucléotides 50 mer en deçà de cette longueur de séquence.

III.4.1.3 Concaténation des séquences 3' et 5' issues du même clone

La concaténation éventuelle des séquences 3' et 5' issues d'un même clone a été réalisée à l'aide du script Perl Concaten35 (Tab. 2, regroupant les logiciels créés au cours de cette

thèse). Les séquences EST 3' et 5' sont comparées les unes aux autres à l'aide du logiciel BLASTN (E-value > $1e^{-5}$) (Altschul *et al.*, 1990), puis les informations issues du BLASTN sont mises sous forme de tableau avec le logiciel MuSeqBox (Xing et Brendel, 2001). Concaten35 utilise les positions du début et de fin du recouvrement pour assembler les séquences 3' et 5' possédant une partie commune.

III.4.1.4 Assemblage des séquences redondantes

Les séquences redondantes sont assemblées avec le logiciel TGICl (Perteau *et al.*, 2003). Ce logiciel réalise l'assemblage en deux temps : une première étape de pré-assemblage (pré-clustering) est réalisée par Megablast pour regrouper les EST potentiellement identiques. Le Megablast est une version optimisée de BLAST pour aligner des séquences presque identiques (Zhang *et al.*, 2000).

L'assemblage est ensuite réalisé avec le logiciel CAP3 (Huang et Madan, 1999) sur chaque groupe d'EST précédemment individualisés. Le pré-clustering par Megablast permet de réduire les erreurs d'assemblage, et de réaliser l'assemblage d'un grand nombre de séquences sur des configurations d'ordinateur modestes, par rapport à un assemblage direct avec CAP3. CAP3 a été utilisé avec les paramètres modifiés $p=93$ (pourcentage d'identité) et $a=100$ (longueur minimale de recouvrement). Cette valeur de p est la plus couramment utilisée pour l'assemblage d'EST, par exemple pour la construction des TIGR gene indices (<http://www.tigr.org/tdb/tgi/>). Elle est proche de la valeur la plus stringente d'acceptation compte tenu de la probabilité d'erreur d'annotation (compte tenu du score Phred, on peut trouver 2% de bases divergentes entre deux séquences, soit $p=98$, mais les erreurs de séquençage d'origine biochimique sont également à prendre en compte). Puisque nous n'avons pas considéré les séquences plus courtes que 100 pb, la longueur minimale de recouvrement est plus importante que celle utilisée au TIGR (40 pb).

III.4.1.5 Calcul électronique du niveau d'expression des gènes (Northern électronique⁴)

Pour éliminer le biais dû à des séquences ne se recouvrant pas, mais issues du même transcrit (ce qui est assez fréquent avec les séquences 5', du fait de l'arrêt prématuré de la reverse transcriptase, des structures en épingles ou bien de l'existence réelle d'ARN tronqués), seules les séquences 3' ont été dénombrées pour le calcul électronique du niveau d'expression des gènes. Les copies d'EST sont dénombré pour chaque banque et chaque contig à partir du fichier de sortie .ACE de CAP3. Le logiciel développé à cette occasion (proc_ace_for_ace,

Tab. 2) est décrit plus loin car il possède d'autres fonctionnalités relatives à la base de données (Chapitre III.4.2). Le traitement statistique du Northern électronique a été réalisé comme décrit par Schmitt *et al.* (1999). La méthode consiste à fixer un niveau de signficativité du différentiel d'expression d'un gène entre 2 banques par la méthode des tables de contingence et la loi du chi². Trois classes de niveau de probabilité et d'expression ont été obtenues :

- classe 1 : Pvalue > 99,1 % et un ratio d'expression >10.
- classe 2 : Pvalue > 99 % et un ratio d'expression >5.
- classe 3 : Pvalue > 95 % et un ratio d'expression >2.

III.4.1.6 Détection des isogènes et des chimères

Afin de détecter les isogènes de 1^{ère} catégorie, ceux similaires dès la comparaison de leur séquence nucléique, les unigènes sont comparés à eux-même par BLASTN en acceptant un seuil $1e^{-20}$, pour l'Evalue. Le fichier de résultat de blast est analysé avec le logiciel Family (Abbal, non publié) pour former les jeux de séquences potentiellement isogènes. Chaque jeu d'isogènes est ensuite analysé avec le logiciel CLUSTALX (Jeanmougin *et al.*, 1998), et les séquences potentiellement chimériques sont repérées *de visu* au sein des alignements multiples. Le cas le plus fréquent, aisément détectable, est le décrochage complet des homologies dans la partie 5', alors que l'alignement du 3' non codant était parfait. La deuxième classe d'isogènes, similaires d'après leurs séquences protéiques est obtenue en utilisant un tBLASTX au lieu de BLASTN. Évidemment, les isogènes de la première catégorie sont toujours regroupés dans la seconde.

III.4.1.7 Annotation fonctionnelle

III.4.1.7.1 Procédure automatique

L'ensemble des séquences d'*Arabidopsis thaliana* ainsi que leurs annotations fonctionnelles respectives ont été téléchargées depuis le site du Munich Information Center for Protein Sequences (MIPS, <ftp://ftpmips.gsf.de/cress/arabiprot/>, version v300403). Les unigènes d'abricot sont comparés aux protéines d'*Arabidopsis* (blastX, $evalue = 1e^{-10}$). Les catégories fonctionnelles de chaque unigène sont ensuite attribuées, d'après le catalogue des fonctions du MIPS, en reportant l'annotation des séquences d'*Arabidopsis* à leur homologue chez l'abricot. Un logiciel (autoanot, Tab. 2) a été développé pour automatiser cette procédure, il analyse directement une sortie de blast à l'aide des fonctionnalités bioperl (Stajich *et al.*, 2002) et le fichier d'attribution des fonction des protéines d'*Arabidopsis*. Un

fichier compatible AceDB est généré et peut être facilement parcouru à l'aide de la base de données.

III.4.1.7.2 Procédure manuelle

Afin d'attribuer une fonction aux unigènes ne présentant pas d'homologie avec les protéines d'*Arabidopsis thaliana*, une procédure similaire a été appliquée en effectuant un blastX (e-value $1e^{-10}$) de ces unigènes sur la base de données NCBI (GenBank CDS translation+PDB+Swissprot+PIR+PRF, mise à jour Aug. 16, 2003. 1,031,459 séquences dans la base de données). Les fonctions des unigènes sans homologue chez *Arabidopsis thaliana* sont déterminées à partir de ce fichier et les catégories fonctionnelles sont déterminées par recherche bibliographique.

III.4.1.8 Calcul du non codant

Lorsque les protéines partielles déduites des EST séquencées en 3' d'abricot présentent des homologies avec des protéines des bases de données publiques, il est possible de calculer la longueur de la partie 3' non codante.

```
bga_001_a01.plca_from_15_to_663 663 bp
length = 649

>ref|NP_193946.1| (NC_003075) putative protein [Arabidopsis thaliana]
length = 302 = L
score = 197
expect = = 5e-50
Frame = -1
Query: 649 ESIIKKSECPKELEPEDRSTSVHVNLIIRRNEECPEPVKRRVPPFQGVGRTLGSSTPVASEX 470
Sbjct: 145 DSIRKSECPKELEPVDKRAPVHVNLMRRDEKCEPEKLVSFQGVGRTLGGASSSTASSQ 204
Query: 469 XXXXXXXXXXXXPSVGLVVDETLPSTSVQLRLADGTRLP AHFNHYHHTISDIRAFIDTSRP 290
Sbjct: 205 SNLTDVAAVQSPLQ-SLVVDETLPSTSIQLRLADGTRMVAKFNNHHTVNDIRGFIEFSRP 263
Query: 289 EGPRSYQLQIMGFPPKLLNDPTQTIEQAGLANSVVIQK 176 = Q
Sbjct: 264 GNPNNYTLQVMGFPPKPLTDPSQTIEQAGLASSVVIQK 301 = S
```

Figure 8 : Exemple d'une sortie de blastX d'EST d'abricot comparée à une protéine probable d'*Arabidopsis thaliana*. Q correspond au début de l'alignement sur l'EST et S à la fin de l'alignement sur la protéine comparée.

On définit Delta comme étant le nombre d'acides aminés manquant dans notre séquence (ou « Query ») pour que l'alignement se termine au codon stop du meilleur homologue (ou « Subject »). $\Delta = L - S$, où L = nombre total d'acides aminés et S position du dernier acide aminé identique dans l'alignement, ces deux valeurs étant obtenues à partir de la protéine homologue. Dans l'exemple (Fig. 8), $\Delta = 302 - 301 = 1$. Q correspond au début de l'alignement du « Query » sur le « Subject ». Si delta n'est pas nul, mais pas trop grand (moins de 10), il est raisonnable de supposer que le non codant aura une longueur égale à $Q-1 - 3x$ (Delta). Le facteur 3 vient du fait que Query est compté en pb, et Subject en acides aminés. Dans l'exemple, le non codant a une longueur de $176 - 1 - 3x(1) = 172$ pb.

III.4.2 Construction de la base de données

Toutes les données générées par les différentes procédures présentées précédemment ont été regroupées dans une base de données développée à partir de l'architecture AceDB (Durbin et Thierry-Mieg, 1991). Une boîte à outils a donc été créée pour formater toutes ces données au format Ace (Tab. 2). Tous les logiciels traitant une sortie de blast utilisent les modules bioperl. L'architecture de base d'AceDB a été modifiée pour pouvoir intégrer toutes les informations. La plupart des logiciels sert à extraire des informations déjà présentes, seuls autoannot (chapitre III.4.1.7.1) et proc_ace exécutent des opérations plus complexes. A partir du fichier de sortie de CAP3 (nom_de_la_banque.ace), proc_ace recherche la banque d'ADNc d'origine de chaque séquence à l'intérieur d'un contig d'après son nom, pour visualiser l'expression électronique des unigènes. Les positions des séquences à l'intérieur de chaque contig sont recherchées afin de générer une représentation graphique de l'assemblage et, enfin, les séquences des contigs et des séquences affiliées sont formatées. Ce logiciel est pour l'instant spécifique au projet abricotier et nécessite quelques modifications pour être applicable sur d'autres projets.

Le marquage des séquences contenant des chimères a été effectué avec une version modifiée du logiciel Nmultigenic (Tab. 2).

Cette base de données demeure, faute de temps, axée sur la fonctionnalité au détriment de la convivialité. Cependant la formation d'un novice pour effectuer des recherches simples ne prend pas plus d'une heure. Elle est utilisable sous Windows en téléchargeant la version windows d'aceDB (<http://www.acedb.org/WinAce/CurrentWinAce.shtml>). Elle est incluse sur

le CD en annexe, il est nécessaire de télécharger AceDB à l'adresse suivante : <http://www.acedb.org/>.

III.4.3 Conception des microarrays. Design des oligonucléotides.

Mille huit cents unigènes ont été sélectionnés pour la synthèse des oligonucléotides greffés sur les lames de microarray. Les unigènes uniquement séquencés en 5' et contenant moins de 100 pb dans la région non codante n'ont pas été pris en compte dans le choix des séquences. Les gènes ayant présenté des profils d'expression pertinents lors du développement, en protéomique ou en Northern électronique ainsi que ceux issus des séquences déposées par Didier Mbéguié-Mbéguié au NCBI ont été sélectionnés. Les gènes impliqués dans le métabolisme de la paroi, le métabolisme secondaire, la transduction des signaux, les facteurs de transcription, les protéines de transport et ceux impliqués dans la réponse aux stress environnementaux ont été sélectionnés également. Pour compléter jusqu'à 1800 séquences, des gènes de fonction inconnue ou sans homologues (abricot spécifiques) ont été ajoutés. Les gènes inconnus retenus devaient avoir une longueur minimale de 500 pb afin d'exclure la possibilité que l'absence d'homologues dans les bases de données ne soit pas lié à leur petite longueur, ils devaient également contenir une séquence polyT à leur extrémité pour pouvoir disposer d'une région non codante. Un premier design a été réalisé avec le logiciel Oligoarray2 (Rouillard, 2003) pour s'assurer que la définition d'une sonde spécifique soit possible pour chaque isogène. Les unigènes pour lesquels il était impossible de construire une sonde spécifique ont été éliminés et remplacés par des séquences inconnues. Le design final des oligonucléotides a été réalisé par MWG Biotech (Ebersberg, Allemagne) avec pour consigne de construire les sondes le plus en aval possible (coté 3') et de ne jamais se situer en amont des 400 dernières bases.

MWG Biotech (Ebersberg, Allemagne) a ensuite réalisé la synthèse et le dépôt des oligonucléotides de 50 pb sur les lames de microarray. Les oligonucléotides 5' aminomodifiés ont été déposés en double de façon non adjacente sur les lames.

III.4.4 Protocole expérimental d'hybridation

Pour l'analyse de l'expression des gènes au cours du développement chez la variété Bergeron, un design en boucle a été utilisé afin de minimiser la variance sur la totalité de l'expérience (Yang et Speed, 2002) (Fig. 9). Pour l'analyse des comparaisons des différentes variétés le protocole expérimental est une adaptation du design expérimental adapté aux expériences avec une matrice 2 par 3 facteurs de variation (deux stades, trois variétés) (Glonek et Solomon, 2004) avec une plus grande importance donnée aux différences variétales par rapport à leur évolution temporelle. Chaque expérience a été reproduite par inversion des fluorophores (dye-swap) et l'expression de chaque gène a été mesurée en double sur chaque lame.

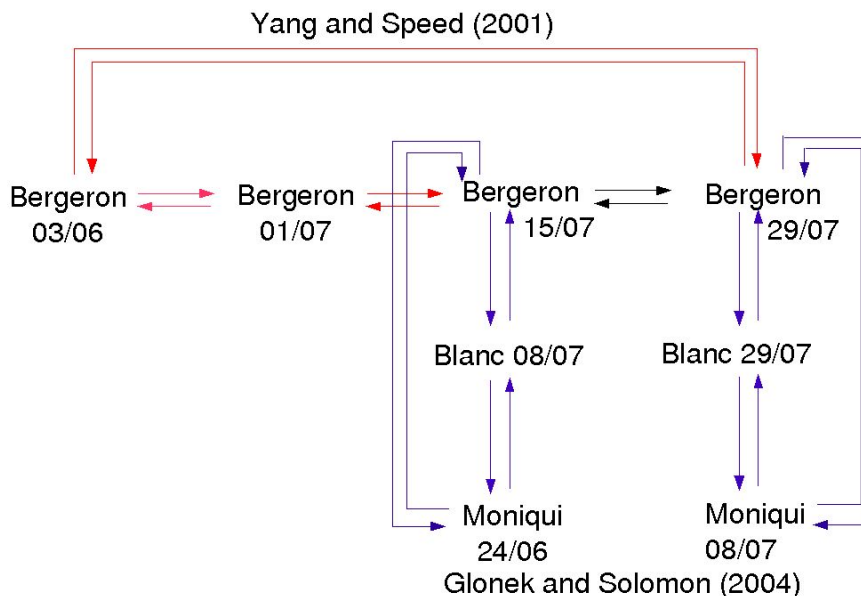


Figure 9 : Schéma du design expérimental utilisé pour l'expérience de suivi du développement (rouge) d'après Yang et Speed (2001) adapté à un suivi cinétique et de celui utilisé pour l'expérience de comparaison de variétés adapté d'après Glonek et Solomon (2004). Les flèches noires correspondent aux lames de microarray utilisées dans les deux expériences.

III.4.5 Analyse de l'image et traitement des données

III.4.5.1 Normalisation des données

Les deux fluorochromes n'ont pas le même encombrement stérique, et n'ont donc pas la même efficacité d'incorporation, ce qui génère un biais dans les valeurs d'intensités obtenues. Il est donc nécessaire d'effectuer une normalisation des données brutes pour corriger ce biais.

On estime que les moyennes des intensités aux 2 longueurs d'ondes doivent être identiques. D'un point de vue biologique, cela revient à supposer qu'en globalité l'expression des gènes est identique dans les deux situations.

La normalisation des données a été réalisée à l'aide du logiciel MIDAS inclus dans le package TM4 (Saeed *et al.*, 2003). Au préalable, le bruit de fond autour de chaque spot est soustrait de la médiane de l'intensité de chaque spot et les intensités des spots où la médiane est inférieure à 2 fois le bruit de fond sont ajustées à 0.

La normalisation est effectuée avec la méthode de régression linéaire locale (LOWESS, Cleveland, 1979) adaptée en haut débit pour les microarrays (Yang *et al.*, 2001). C'est actuellement la méthode la plus répandue pour la normalisation des données de microarray. Cette méthode part de l'hypothèse que le biais sur chaque spot est dépendant de l'intensité du spot. Après normalisation, le rapport corrigé $\log(\text{rouge}/\text{vert})$ est égal au ratio initial $\log_2(R/V) - c(A)$, où $c(A)$ est la valeur locale de la courbe de régression linéaire de $\log_2(R/V)$ vs $\log_{10}(R-V)$. Cette opération revient à centrer sur zéro, mais au lieu d'utiliser la moyenne des intensités de l'ensemble des spots pour tous les spots, on utilise la moyenne des intensités des spots d'intensité voisine pour chaque spot.

L'écart type de chaque bloc de la lame (32 blocs de 60 gènes) est ensuite normalisé en supposant qu'il est identique pour tous. Ceci a pour effet d'éviter d'avoir des différences d'expression liées à la position sur la lame.

III.4.5.2 Analyse statistique des données

Les données ont été traitées par deux analyses indépendantes. La première analyse permet de déterminer par la méthode de l'ANOVA les gènes dont la différence d'expression est significative lorsque l'on fait varier 2 paramètres. La seconde permet d'obtenir une valeur unique d'intensité pour chaque condition (stade de développement ou variété) afin d'examiner les profils d'expression.

III.4.5.2.1 ANOVA

Les données ont tout d'abord été analysées par la méthode statistique ANOVA (Fisher, 1954) pour déterminer les gènes dont le niveau d'expression entre 2 conditions (stade de développement ou variété) est significativement différent. Cette analyse a été réalisée avec le logiciel geneANOVA (Didier *et al.*, 2002) à partir des données normalisées, centrées et réduites. Pour chaque gène le seuil de significativité a été fixé à une p-value de 10^{-3} (99,9 %

de chance de rejeter l'hypothèse « le gène ne varie pas en fonction de la condition »). Une variance minimale a aussi été fixée pour éviter de considérer les gènes significativement différenciellement exprimés, mais ne présentant que peu d'écart entre les conditions (gènes très reproductibles mais peu régulés).

La p-value est indicatrice de l'évidence de la différence de niveau d'expression, en lien avec le paramètre étudié uniquement, et non avec les autres facteurs de variabilité (réplicat sur lame, dye-swap, lame et stade de développement dans l'analyse des variétés). C'est aussi à l'heure actuelle l'une des méthodes les plus répandues pour l'analyse des données provenant des puces à ADN.

Dans cette analyse les gènes pour lesquels plus d'une valeur est manquante ont été éliminés.

III.4.5.2.2 Profil d'expression

Le but de cette analyse est de générer les profils d'expression de chaque gène, puis d'associer les profils d'expression similaires (clustering) des gènes significativement variants d'après l'ANOVA et de déterminer le sens de leur variation (induit ou réprimé pendant le développement ; spécifique de quelle(s) variété(s)).

La comparaison des réplicats à l'intérieur de chaque lame est effectuée avec un logiciel réalisé pour l'occasion (Mareplicate) : les intensités des spots dont les réplicats varient de $\pm 30\%$ sont écartés. Les réplicats acceptés sont moyennés.

L'analyse du « dye swap » est effectuée par le logiciel MIDAS. La valeur du log du ratio croisé est calculée pour chaque spot : $A_g = \log_2(R1 \cdot R2 / G1 \cdot G2)$, où R1 est l'intensité du spot à 635 nm sur la première lame, R2 est l'intensité du spot à 635 nm sur la seconde lame, G1 est l'intensité du spot à 532 nm dans la première lame et G2 est l'intensité du spot dans la seconde lame à 532 nm. Si les expériences sont reproductibles pour le dye swap, alors R1 est proche de G2 et R2 est proche de G1, donc A_g est proche de 0. La valeur de la moyenne et la déviation standard de l'ensemble des A est calculée (A_m). Les spots dont la valeur de A_g est supérieure ou inférieure à $A_m \pm (2 \text{ fois la déviation standard})$ sont écartés. Les spots dont la variation est acceptée sont moyennés.

L'intensité d'expression a été fixée pour un stade choisi arbitrairement. L'intensité d'expression aux autres stades a été calculée en tenant compte des protocoles expérimentaux et des ratios d'expression entre les différentes conditions. La moyenne d'expression est calculée pour chaque gène et le ratio d'induction par rapport à la moyenne a été utilisé pour le

regroupement par profil d'expression, réalisé à l'aide du logiciel Mev inclus dans le package TM4 (Saeed *et al.*, 2003).

III.4.6 Support bioinformatique à l'analyse protéomique

Afin d'intégrer les données issues du séquençage d'EST, et pour améliorer l'identification des empreintes massiques, une procédure a été mise au point afin de former des protéines complètes mais chimériques comprenant les traductions d'EST d'abricotier complétées par des séquences potentiellement homologues aux régions non séquencées. Un logiciel, PT-HP assembler (pour Partial Transcript – Homologous Protein) a été mis au point pour réaliser cette opération. La section matériel et méthode correspondant à ce travail est incluse dans la publication présentée au chapitre IV.4.

III.4.7 Lexique des termes de bioinformatique

1 Base calling : le Base calling est l'étape de transformation des spectres produits par le séquenceur en séquence d'acide nucléique. La précision pour chaque base est déterminée par le score Phred.

2 Score Phred : le score Phred détermine la qualité de lecture de chaque base d'après la formule $q = -10 \times \log_{10}(p)$. q est le score Phred, et p est la probabilité d'erreur du Base calling. Plus ce score est haut, meilleur est la qualité. Ainsi, un score de 20 correspond à une probabilité d'erreur de un sur cent. Cette probabilité tient compte uniquement de la forme et de la position des pics sur les électrophorogrammes en sortie de séquenceur et ne prend pas en compte les erreurs de séquences liées à d'éventuelles PCR en cours de clonage, par exemple.

3 E-value : l'E-value définit la probabilité que l'alignement de 2 séquences soit dû au hasard. Sa formule est $E = Kmne^{-\Lambda S}$ avec K et Λ paramètres de Karlin et Altschul (Karlin et Altschul, 1990), m et n la taille des banques comparées et S le score. Plus faible est l'E-value, meilleur est l'alignement.

4 Northern électronique : Le Northern électronique consiste à déterminer le niveau d'expression d'un gène dans une condition donnée (tissu, stade de développement, etc..) d'après l'abondance relative des Est correspondantes dans une banques de cDNA réalisées

dans la même condition. Les banques d'EST ne doivent pas avoir subi de modifications pouvant biaiser la représentativité des gènes.

Tableau 2 : Liste des logiciels réalisés au cours de la thèse. Tout ces logiciels sont écrits en Perl. Les logiciels se terminant par le suffixe *_for_ace* sont utilisés pour rentrer des informations dans la base de données. * Logiciel nécessitant les modules Bioperl.

<i>Nom</i>	<i>Fonction</i>	<i>Fichier d'entrée</i>
autoannot_for_ace*	Génère et reporte les catégories fonctionnelles d'après Arabidopsis	Résultats de blast de données à annoter sur les protéines d'Arabidopsis, fichier d'attribution des catégories fonctionnelles des protéines d'Arabidopsis.
mananot_for_ace	Reporte les catégories fonctionnelles attribuées manuellement	Fichier tabulé
parseblast_for_ace*	Reporte les dix plus proches homologues de la séquence	Résultat de blast
Nmultigenic_for_ace	Reporte la famille multigénique « nucléique »	Fichier tabulé
proc_iso	Reporte la famille multigénique « protéique »	Fichier tabulé
proc_fcat_for_ace	Reporte le catalogue des catégories fonctionnelles	Fichier catalogue des catégories fonctionnelles
singleton_cDNA	Reporte la séquence des singlets non contigués	Fichier fasta
proc_ace_for_ace	Génère et reporte le Northern électronique, reporte la séquence des contigs et leur longueur. Localise la position des séquences sur le contig.	Fichier de sortie de CAP3 .ace
singleton_length	Reporte la taille des singlets	Fichier fasta
Concaten35.pl	Assemble les séquences 3' et 5' issues du même clone	Fichier de sortie du logiciel MuseqBox + fichiers fasta séquences 3' et 5'
PT-HP assembler*	Inclue dans une séquence protéique les traductions conceptuelles d'EST	Fichier de résultats de blast, fichier fasta de séquences protéiques

IV Résultats

IV.1 Caractérisation physiologique des fruits

Les fruits utilisés pour les expériences de microarray, protéomique et PCR quantitative ont été prélevés au cours de la campagne 2002. L'ensemble de ces techniques ont été appliquées pour l'étude de la variété Bergeron. Les variétés Moniqui et Blanc ont uniquement fait l'objet d'investigations par microarray.

IV.1.1 Émission d'éthylène

La forte différence en terme d'émission d'éthylène (Fig. 10) a principalement motivé le choix de ces trois variétés comme matériel d'étude. Le début de l'induction du dégagement d'éthylène correspond à l'initiation de la maturation des abricots. Après induction, le dégagement est exponentiel. Il n'y a pas, à proprement parler, de variété type pour le dégagement d'éthylène, aussi bien en terme de quantité dégagée, qu'en date d'entrée de maturation. Moniqui est une variété précoce présentant un dégagement d'éthylène très important. Bergeron est une variété plutôt tardive, à dégagement d'éthylène modéré. Blanc ne produit pas d'éthylène, ou très peu même après récolte ou après blessure (Gouble B. communication personnelle).

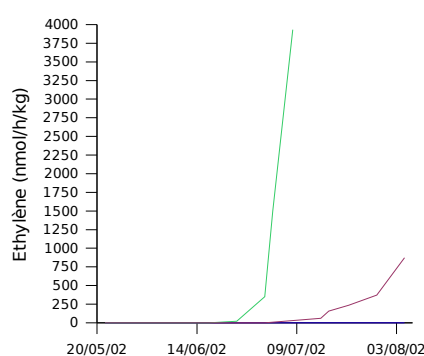


Figure 10 : Evolution du dégagement d'éthylène au cours de la maturation chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

IV.1.2 Sucres

Les 3 variétés présentent une évolution identique de l'indice réfractométrique (résultats non présentés). Les courbes d'accumulation des sucres sont présentées figure 11. Dans les trois variétés, l'accumulation de saccharose (Fig. 11A) précède le dégagement d'éthylène, elle est la plus rapide chez la variété Moniqui. Comparée aux deux autres variétés, la variété Blanc accumule des quantités plus importantes de glucose (Fig. 11B) et de fructose (Fig. 11C) au cours du développement. Blanc contient à maturité deux fois plus de glucose et près de quatre fois plus de fructose que Bergeron et Moniqui.

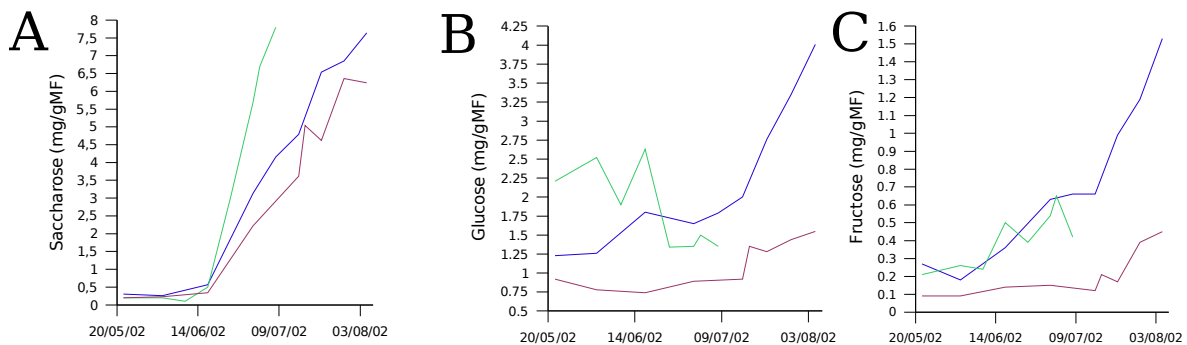


Figure 11 : Evolution des concentrations en sucres (**A** : saccharose ; **B** : glucose ; **C** : fructose) au cours de la maturation chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

IV.1.3 Poids et fermeté

Les prélèvements des fruits ont été effectués à partir du milieu de leur phase stationnaire de croissance (Fig. 12A). La deuxième sigmoïde de la courbe de croissance du poids des fruits (Fig. 3) est visible pour Bergeron et Moniqui. En revanche, la croissance de la variété Blanc est plutôt linéaire. A maturité, les poids des fruits de Blanc et Moniqui sont inférieurs à ceux de Bergeron. Les cinétiques de perte de fermeté des trois variétés ont les mêmes allures et sont légèrement décalées dans le temps (Fig. 12B). le début du ramollissement précède l'initiation du dégagement d'éthylène.

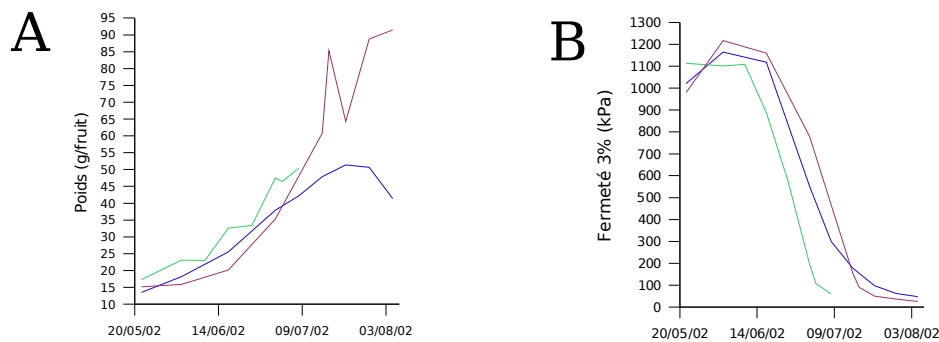


Figure 12 : Evolution (A) du poids des fruits et (B) de la fermeté au cours de la maturation chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

IV.1.4 Couleur

Les figures 13A et 13B présentent l'évolution de la coordonnée a^* de l'espace colorimétrique $L^*a^*b^*$, respectivement pour la couleur de fond (A) et la couleur de sur-impression (B). Ce paramètre correspond au changement de couleur du vert (-) au rouge (+). Les autres coordonnées (L^*b^*) de l'espace colorimétrique sont influencées dans une moindre proportion par la maturation. La luminance (L^*) des variétés blanches (Moniqui et Blanc) est plus élevée. Les valeurs de a^* des chairs sont proches de celles de la couleur de fond (données non présentées). La valeur de b^* (axe bleu-jaune) est toujours aux alentours de cinquante chez les trois variétés pour la couleur de fond et de quarante pour la couleur de sur-impression. Aux stades immatures, les couleurs de fond et de sur-impression sont jaune-vert. La couleur de fond à maturité est crème pour Moniqui et Blanc et jaune-orangé pour Bergeron. Pour la sur-impression, les couleurs sont jaune-orangé pour Moniqui et rouge-orangé pour les autres variétés. Pour les trois variétés, les couleurs de fond et de sur-impression tendent vers le rouge pendant la maturation. Bergeron accumule du β carotène (Fig. 14A) et du phytoène (Fig. 14B) en quantité importantes. Dans les autres variétés seules des traces de ces composés. Par contre, du phytofluène est retrouvé à la fois chez Moniqui et Bergeron, mais peu dans Blanc (Fig. 14C). La chlorophylle a été dosée à 92, 105 et 107 JAA dans Bergeron, sa concentration décroît pendant cette période (respectivement 633, 282 et 86 $\mu\text{g/g}$ de matière fraîche).

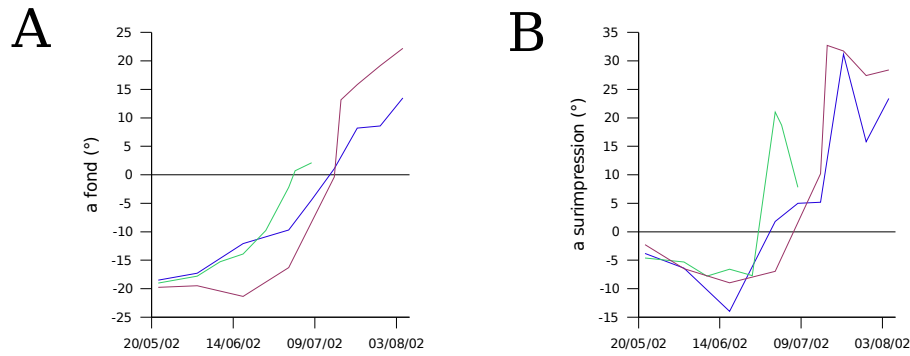


Figure 13 : Evolution des paramètres liés à la couleur. (A) coordonnée a^* de la couleur de fond ; (B) coordonnée a^* de la couleur de sur-impression, chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

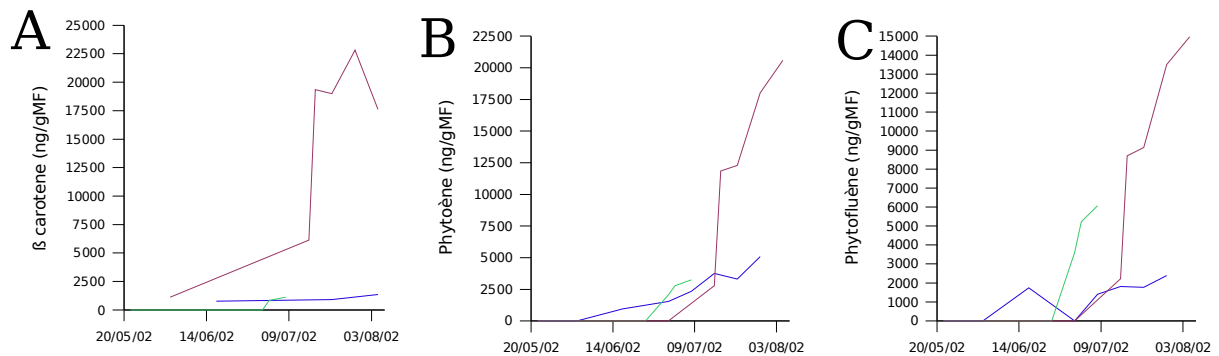


Figure 14 : Evolution des caroténoïdes au cours de la maturation. (A) concentration en β carotène ; (B) concentration en phytoène ; (C) concentration en phytofluène ; chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

IV.1.5 Acidité

L'acidité titrable exprimée en méq pour 100 g de matière fraîche diminue de façon comparable dans les trois variétés au cours de la maturation (Fig. 15A). Exprimées par fruit, les acidités sont de l'ordre de .20 méq/fruit à maturité. Les fruits verts de toutes les variétés contiennent principalement de l'acide malique (Fig. 15B) dont la concentration diminue pendant la maturation. L'acide citrique (Fig. 15C) est accumulé jusqu'à la crise éthylénique, puis sa teneur diminue jusqu'en fin de maturation.

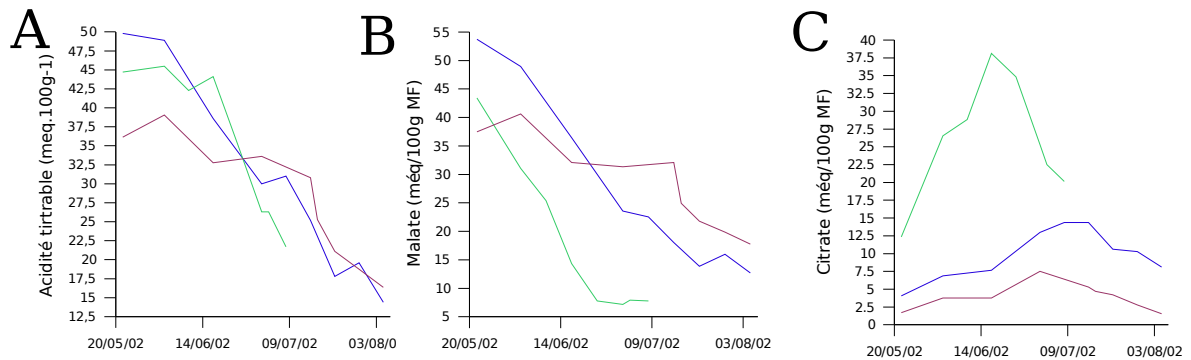


Figure 15 : Evolution des concentrations en acides (**A** : Acidité titrable ; **B** : malate ; **C** : citrate) au cours de la maturation chez les variétés Bergeron (rouge), Moniqui (vert) et Blanc (bleu).

IV.1.6 Profils aromatiques

Les profils aromatiques ont été étudiés chez Bergeron et Moniqui. L'ACP a permis de différencier les profils aromatiques de ces deux variétés. Les composés suivants sont retrouvés préférentiellement chez Moniqui : butyl acétate, hexyl acétate, hexanal, 2-hexanal, hexanol, β -ionone, γ -octolactone, γ -décylactone, γ -dodécylactone, γ -hexylactone, δ -décylactone. Les composés suivants sont préférentiellement retrouvés chez Bergeron : 6-méthyl-5heptèn-2-one, géranylacétone, β -cyclocitral, β -ionone-5,6-époxy, dihydroactinidiolide, benzaldéhyde.

En conclusion, les paramètres physico-chimiques des fruits sont affectés par :

(i) le développement (dégagement d'éthylène, teneurs en sucres, acidité, fermeté, couleur).

(ii) le fond génétique, principalement :

- au niveau de la couleur, Bergeron est différent de Blanc et Moniqui ;
- Moniqui dégage plus d'éthylène que Bergeron, et aucun dégagement n'est observé chez Blanc ;
- la concentration en citrate est plus élevée chez Moniqui que chez les deux autres variétés ;
- la concentration en malate est moins élevée chez Moniqui que chez les deux autres variétés.

Afin de rechercher les mécanismes moléculaires responsables de ces variations développementales et variétales de composition, une stratégie haut débit d'étude du transcriptome et du protéome a été développée.

IV.2 Obtention et traitement des EST

IV.2.1 Caractérisation des banques d'ADNc

(Résultats obtenus à partir des banques construites au cours de la thèse de Didier Mbéguié-Mbéguié, 2000)

Les banques d'ADNc unidirectionnelles ont été réalisées à partir d'abricots de la variété Bergeron récoltés à 89 (stade vert), 107 (stade mi-mûr) et 117 (stade mûr) jours après anthèse. Ces stades ont été choisis afin de cibler les points critiques de l'évolution des fruits. Leurs principaux paramètres physiques et biochimiques sont résumés dans le tableau 3. Le stade vert correspond au durcissement du noyau, le stade mi-mûr au départ de la production autocatalytique d'éthylène et le stade mûr à la fin de l'accumulation en saccharose.

Tableau 3 : Caractérisation physico-chimique des fruits utilisés pour la construction des banques d'ADNc. Les données pour le saccharose, l'éthylène, le malate et le citrate sont exprimées par kilogramme de fruits frais. La fermeté est mesurée par compression des fruits à 3 % de leur diamètre.

	Vert	Mi-Mûr	Mûr
Jours après anthèse	89	107	117
Poids frais (g/fruit)	38	64	88
Saccharose (mmol.kg ⁻¹)	93	1099	1319
Malate (meq.kg ⁻¹)	463	337	313
Citrate (meq.kg ⁻¹)	61	102	85
C ₂ H ₄ (nmol.h ⁻¹ .kg ⁻¹)	0	0	26
Fermeté (pa, 3 %)	1217	154	50

Les banques primaires d'ADNc contiennent 10⁶ plaques formant unités (pfu) par µg d'ADN de vecteur, ce qui est théoriquement suffisant pour inclure les ARN de faible abondance (Sung *et al.*, 1998). La taille moyenne des inserts, estimée par PCR, est de 1,2 kb. Cette valeur est comparable à ce qui a été observé pour d'autres banques construites avec le vecteur λzap chez d'autres fruits (e.g. 0,8-1,3 kb pour la pomme, Sung *et al.*, 1998 ; 1,3 kb pour le raisin, Terrier *et al.*, 2001a).

IV.2.2 Étude comparative des stratégies de séquençage en 3' et en 5'.

Il a fallu définir l'orientation de leur séquençage, au départ du programme de séquençage d'EST. La stratégie 5' était, en 2001, de loin la plus courante dans les programmes EST car la queue poly-A était difficilement franchissable par les techniques de séquençage, et si tel était le cas, la longueur des séquences obtenues ne dépassait que rarement la longueur du 3' non codant. Il était donc quasiment impossible d'effectuer l'annotation fonctionnelle par BlastX après séquençage en 3'. Ces problèmes n'étaient pas rencontrés lors du séquençage en 5', car le 5' non codant est beaucoup plus court et qu'il est d'ailleurs absent de nombreux clones, si des précautions particulières ne sont pas prises pour obtenir des banques dites de pleine longueur.

Avant le démarrage du séquençage systématique d'EST, 96 clones par banques ont été séquencés à partir des deux extrémités 3' et 5' afin de comparer la longueur de séquences, la longueur des séquences traduites, la capacité de discriminer les différents isogènes.

IV.2.2.1 Critères de séquençage

IV.2.2.1.1 Nombre de séquences par boîte

Le nombre de séquences effectivement lues est sensiblement le même dans toutes les boîtes (Tab. 4) et il est relativement élevé. Ceci peut s'expliquer par le fait que les clones ont subi très peu d'étapes de repiquage (ce sont les banques entières et non pas les clones qui ont été transportées chez le prestataire de service). Ce facteur n'est donc pas influencé par l'orientation 3' ou 5' des séquences.

Tableau 4 : Nombre de séquences obtenues par boîte 96 puits pour chacune des banques suivant l'orientation du séquençage.

	3'	5'	Total
Vert	93	90	183
Mi-mûr	95	94	187
Mûr	96	91	189
Total	282	277	559

IV.2.2.1.2 Taille moyenne des séquences brutes

La longueur moyenne des séquences effectuées en 5' est supérieure de 62 pb (Tab. 5). Cependant cette différence tient surtout à la présence d'une boîte (3', stade vert) qui ne remplissait pas les critères du cahier des charges du prestataire de service (75 % des séquences de longueurs supérieure à 250 pb, Tab. 7) et n'a pas été facturée.

Le test de Student pratiqué en intégrant cette boîte suggère que l'écart entre les moyennes 3' et 5' serait significatif ($p=95\%$). Cependant, l'écart est alors à peine significatif si on ne tient pas compte de la banque verte en 3'.

Tableau 5 : Taille moyenne des séquences brutes (pb) pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'	Total
Vert	459	579	519
Mi-mûr	559	610	584
Mûr	572	588	580
Moyenne	530	592	561

IV.2.2.1.3 Comparaison de la taille des séquences avec les données de la bibliographie

La longueur moyenne des séquences sans tenir compte de l'orientation est de 561 pb (Tab. 6). Cette valeur est inférieure à celles obtenues chez le raisin (691 pb, Terrier *et al.*, 2001a) mais supérieure celles de à *Criptomera japonica* (450 pb, Ujino-Ihara *et al.*, 2000), de *Porphyra yezoensis* (470 pb, Nikaido *et al.*, 2000) de *Citrus* (320 pb, Hisada *et al.*, 1997), de peuplier (418 pb, Sterky *et al.*, 1998), de maïs (445 pb, Fernandes *et al.*, 2002) et équivalente à celles de *Medicago* (601 pb, Covitz *et al.*, 1998). La taille des séquences est dans l'ordre des grandeurs des données actuellement publiées et peut donc être considérée comme acceptable en 5' comme en 3'.

Tableau 6 : Taille moyenne et orientation des séquences publiées, comparaison avec *Prunus armeniaca*

Espèces	Longueur en pb	Orientation	Référence
<i>Vitis vinifera</i>	691	3'	Terrier <i>et al.</i> , 2001a
<i>Medicago truncatula</i>	601	5'	Covitz <i>et al.</i> , 1998
<i>Prunus armeniaca</i>	561	3' et 5'	
<i>Porphyra yezoensis</i>	470	5'	Nikaido <i>et al.</i> , 2000
<i>Cryptomeria japonica</i>	450	3' et 5'	Ujino-Ihara., 2000
<i>Zea mays</i>	445	3' et 5'	Fernandes <i>et al.</i> , 2002
<i>Populus</i>	418	5'	Sterky <i>et al.</i> , 1998
<i>Citrus unshiu</i>	320	3' et 5'	Hisada <i>et al.</i> , 1997

IV.2.2.1.4 Répartition de la taille des séquences

La plupart des séquences a une taille comprise entre 500 et 800 pb (76 % pour les séquences en 3' et 91 % pour celles en 5') (Fig. 16), cependant une fraction non négligeable des séquences 3' a une taille inférieure à 250 pb (14 %).

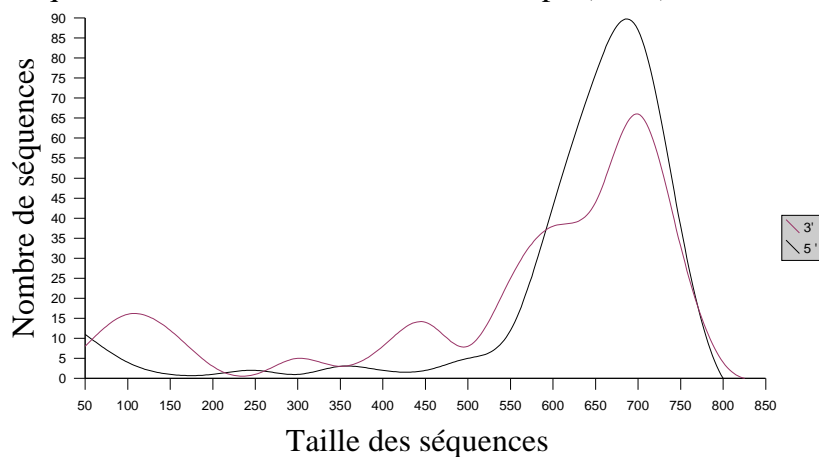


Figure 16 : Répartition de la taille des séquences en 3' et en 5'

Le pourcentage moyen de séquences dépassant les 250 pb est supérieur en 5' (86 contre 80) (Tab. 7). L'écart n'est pas significatif ($p=95\%$), il le serait encore moins, si la boîte verte avait été séquencée correctement.

Tableau 7 : Pourcentage de séquences supérieures à 250 pb pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'	Total
Vert	74	85	80
Mi-mûr	81	88	85
Mûr	84	84	84
Moyenne	80	86	83

IV.2.2.1.5 Taille des séquences supérieures à 250 pb

La taille moyenne des séquences supérieures à 250 pb (Tab. 8) et significativement plus élevée en 5' (p=95 %), il ne l'est plus si on retire la banque issue du stade vert en 3'. Cette boîte contient non seulement une plus forte proportion de séquences courtes, de plus les séquences supérieures à 250 pb sont plus courtes que la moyenne. Dans toutes les autres boîtes, la taille des séquences est similaire.

Tableau 8 : Tailles moyennes des séquences supérieures à 250 pb pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'
Vert	550	620
Mi-mûr	630	639
Mûr	619	638
Moyenne	600	632

IV.2.2.1.6 Séquences exploitables

Nous avons considéré comme exploitable une séquence qui obtient un score minimal de 100 bits en auto alignement (E-value proche de $3e^{-20}$), ce qui correspond à une longueur d'au moins 100 pb, queue polyA et régions de faible complexité exclues. La proportion de séquences exploitables est plus élevée en 5' (+5 %) (Tab. 9), mais n'est pas significativement différente (p=95 %).

Tableau 9 : Pourcentage de séquences exploitables pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'
Vert	83	90
Mi-mûr	87	97
Mûr	92	89
Moyenne	87	92

IV.2.2.2 Caractéristiques des séquences

IV.2.2.2.1 Partie non codante des séquences identifiées en 3'

La longueur de la partie non codante (255 pb) n'a été calculée que sur les séquences 3' car elle n'apparaît qu'exceptionnellement sur les clones séquencés en 5' (Tab. 10). Cette valeur est équivalente à celle observée chez le raisin (249 pb). La longueur de la partie codante est

donc de $530 - 255 = 275$ pb soit 92 aa en moyenne. La longueur de la partie codante séquencée diminue donc à peu près de moitié lors du séquençage en 3', mais devrait rester suffisante puisque 33 aa (100 pb) permettent en général d'identifier une séquence.

Tableau 10 : Longueur de la partie non codante des séquences 3' identifiées (en pb) pour chacune des banques.

	Longueur du non codant (pb)
Vert	250
Mi-mûr	264
Mûr	251
Moyenne	255

IV.2.2.2.2 Séquences supérieures à 250 pb identifiées

Nous considérons qu'une séquence est identifiée si elle présente un score minimal de 50 bits avec une séquence publique en BLASTX. 81 % en moyenne des séquences trouvent au moins un homologue (Tab. 11). Le taux d'identification est plus élevée de 4 % pour le séquençage à partir de l'extrémité 5', mais cet écart ne serait pas significatif ($p = 95\%$).

Tableau 11 : Pourcentages des séquences supérieures à 250 pb identifiées pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'	Total
Vert	75	79	77
Mi-mûr	78	84	81
Mûr	83	85	84
Moyenne	79	83	81

IV.2.2.2.3 Séquences identifiées indépendamment de leur longueur

Le seuil précédent de 250 pb avait été utilisé pour exclure l'effet de la taille des séquences sur le taux d'identification., sachant que c'est la taille moyenne de l'extrémité 3' non codante. L'écart n'est toujours pas significatif ($p = 95\%$), même s'il est plus élevé qu'entre les séquences supérieures à 250 pb en 5' (9 % d'écart) (Tab. 12).

Tableau 12 : Pourcentage de séquences identifiées sur l'ensemble des clones pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'
Vert	57	74
Mi-mûr	78	80
Mûr	77	85
Total	71	80

L'écart de taux d'identification entre 3' et 5' est relativement faible sur les séquences supérieures à 250 pb, la différence entre les deux réside principalement sur les séquences inférieures à 250 pb (Fig. 16) présentes en quantité non négligeable dans les séquences en 3'. Cette longueur seuil que nous nous sommes fixée (précisément égale à la longueur moyenne du 3' non codant) est critique pour l'identification. Ainsi une seule séquence inférieure à 250 pb a pu être identifiée en 3' et 4 séquences en 5'.

IV.2.2.2.4 Score des séquences identifiées

Le score moyen d'homologie des séquences identifiées est plus élevé en 5' (+32 bits) (Tab. 13), cependant ce critère n'a qu'une importance relativement secondaire. En effet, pour valider une homologie entre deux séquences, un score entre 50 et 100 est suffisant en fonction de la qualité de l'alignement. La différence entre les scores en 3' et 5' est due à la présence de la région non codante qui n'intervient pas dans l'alignement en BLASTX, donc dans le score. Cependant celle-ci n'est pas pénalisante à partir du moment où la séquence est assez longue. On note des scores moins bons pour les séquences de la banque verte, ce qui s'explique par la taille moyenne très faible des séquences supérieures à 250 pb dans cette boîte (Tab. 8).

Tableau 13 : Score moyen des séquences identifiées (en bits) pour chacune des banques en fonction de l'orientation de séquençage.

	3'	5'	Ensemble des séquences
Vert	130	162	146
Mi-mûr	155	181	168
Mûr	149	188	168
Ensemble des séquences	145	177	161

IV.2.2.2.5 Redondance : nombre de clones redondants

L'analyse de la redondance a été réalisée sur les séquences exploitables définies dans le paragraphe IV.2.2.1.6.

Tableau 14 : Nombre de séquences redondantes en fonction du sens de séquençage.

	3'	5'
Séquences redondantes	46	36
Moyenne par boîte	15	12

On obtient une moyenne de 13 séquences redondantes par boîte en 3' et 10 en 5' mais l'écart ne serait pas significatif ($p=95\%$). Cependant, certains clones ne sont redondants que dans un sens de séquençage (11 séquences en 3' et 4 séquences en 5', Tab. 14). Les causes peuvent être multiples : les clones sont chimériques ; les séquences ne démarrent pas au même endroit en 5' empêchant un recouvrement ; enfin, les séquences n'ont été obtenues que dans un sens de séquençage.

On obtient au total 91 % de séquences indépendantes en 3' représentées par 214 singlets et 14 familles, et 92 % de séquences indépendantes en 5' représentés par 234 singlets et 13 familles. On obtient deux séquences isogènes en 3' et deux autres en 5'. Chaque paire d'isogènes n'apparaît que dans un sens de séquençage car, dans les deux cas, dans l'autre sens l'un des deux isogènes est court et non exploitable. Les problèmes de non détection de la redondance sont pour la plupart causés en 3' par des séquences courtes et en 5' par des séquences ne se chevauchant pas, les autres phénomènes sont minoritaires.

Les données acquises concernant les isogènes sont relativement restreintes car seulement deux familles multigéniques ont été mises en évidence sur ce jeu de données restreint (ubiquitine et métallothionéine). Sur ce point il n'y a pas de différence dans le sens de séquençage au niveau comptable, mais le séquençage en 3' permet de disposer de séquences assez différentes sur une grande longueur dans le non codant pour constituer des sondes isogènes spécifiques, ce qui n'est pas le cas en 5'.

IV.2.2.2.6 Conclusion

Les arguments en faveur de chaque sens de séquençage sont récapitulés dans le tableau 15. L'identification des séquences 3' n'est pas aussi performante qu'en 5', mais la perte d'information est minime. Cependant la différence de longueur n'est significative que du fait d'une boîte qui a été séquencée en 3' dans des conditions non optimales. Les séquences en 3' sont moins bien identifiées à cause de la présence d'un certain nombre de séquences n'atteignant pas les 250 pb. Par contre la présence de la partie non codante n'est pas pénalisante pour l'identification, et permet de différencier de façon explicite les isogènes nécessaires à l'étude de l'expression, et la co-localisation avec les QTL. La plupart des

isogènes pourrait être détectée par séquençage en 5', cependant il existe un risque de ne pas détecter les isogènes strictement identiques sur leur partie codante. Par contre, d'après l'expérience acquise par la suite, ¼ des séquences obtenues à la fin du séquençage ont un homologue dans la banque suffisamment proche pour rendre inefficace toute construction de sonde dans la partie codante, ce chiffre est en dessous de la réalité biologique car le transcriptome de l'abricot n'est que partiellement connu. Ainsi chez *Arabidopsis*, le taux de gènes appartenant à des familles multigéniques est estimé à 65 % (*Arabidopsis* Genome Initiative, 2000). Enfin, le séquençage en 5' induit une sous estimation de la redondance dans la banque.

Tableau 15 : Récapitulatif des avantages de chaque sens de séquençage.

<i>Avantages 3'</i>	<i>Avantages 5'</i>
- Meilleure identification des isogènes	- Séquences plus longues
- Les isogènes sont utilisables pour les études d'expression (3' non codant= sonde)	- Meilleure identification des fonctions probables
- Les banques sont absolument non redondantes	

Le séquençage en 5' semble plus approprié dans le cadre d'un programme uniquement basé sur l'identification de séquences. Par contre, notre projet contient un volet 'mesure de la variation d'expression des transcrits' pour lequel le séquençage en 3' est plus approprié.

Au moment du choix (2001), le séquençage en 3' était peu répandu. Cependant le temps nous a donné raison, puisqu'à l'heure actuelle, le séquençage en 3' est préférentiellement utilisé (Rudd, 2003).

Environ un quart des clones a cependant été séquençé dans les deux orientations, ces séquences ont été uniquement choisies au hasard. Il eût été plus logique de re-séquencer en 5' les séquences non redondantes mais les séquences 5' nous ont été offertes par le prestataire de service à la suite d'une incompréhension.

IV.2.3 Traitement des EST

IV.2.3.1 Production du jeux d'unigènes

Bien qu'il existe un certain nombre de pipe-line « clé en main » pour réaliser le traitement des séquences, nous avons choisi de mettre au point notre propre stratégie pour

plusieurs raisons : (i) tous les logiciels utilisés ici sont sous licences publiques ou de conception locale. Ils sont cependant considérés comme les meilleurs dans leurs catégories pour la précision des résultats et sont largement utilisés par la communauté scientifique. (ii) Les chaînes des traitements sont des boîtes noires où il est difficile de modifier des paramètres et sont principalement axées sur la convivialité. (iii) Certains points de cette chaîne sont spécifiques à nos besoins (détermination du critère séquence courte, assemblage des extrémités 3' et 5' issues du même clone).

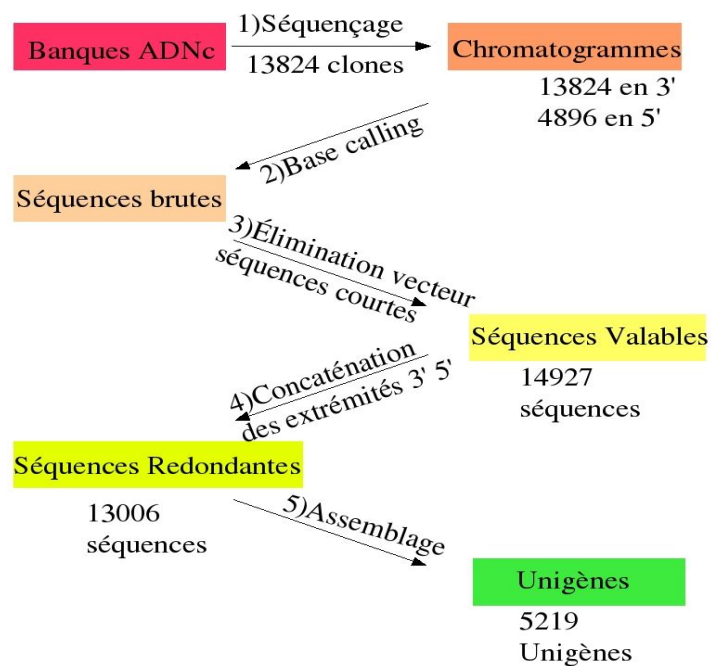


Figure 17 : Descriptif de la chaîne de traitement des EST avec à chaque étape le nombre de séquences obtenues.

L'obtention et l'analyse des EST en vue de produire le jeu d'unigènes (séquences non redondantes) sont réalisées en 5 étapes distinctes récapitulées dans la figure 17. Chaque étape est décrite dans les paragraphes suivants.

IV.2.3.1.1 Séquençage

Le séquençage, réalisé par la société MWG Biotech à Ebersberg, consiste à déterminer sous forme de chromatogramme l'ordonnancement des bases de l'insert d'ADNc à partir de la séquence de l'amorce (T7 en 3' et T3 en 5'). 13824 clones ont été séquencés à partir de l'extrémité 3' et 4896 d'entre eux ont été aussi séquencés à partir de l'extrémité 5', soit 18720 chromatogrammes au total. Le nom de chaque séquence représente les informations de traçabilité : pour l'exemple suivant BGA001A01P, les deux premières lettres (BG)

représentent la variété (uniquement Bergeron) ; la troisième, représente le stade de maturité (A : banque 'vert' ; E, banque 'mi-mûr' ; H, banque 'mûr') ; les trois chiffres suivants (001) représentent le numéro de boîte ; les trois caractères suivants représentent la position dans la boîte 384 puits (de A à P en vertical et de 1 à 24 en horizontal) et la dernière lettre représente l'orientation de séquençage (P en 3', Q en 5'). A ce stade les noms des séquences contiennent encore des informations inutiles, utilisées par le prestataire de service, qui seront éliminées plus tard (paragraphe IV.2.3.1.3).

IV.2.3.1.2 Base calling

Le Base calling est réalisé avec le logiciel PHRED (Ewing *et al.*, 1998). Cette étape consiste à lire les chromatogrammes automatiquement pour (i) produire les séquences des EST au format Fasta et attribuer à chaque base une probabilité d'erreur et (ii) éliminer les bases où cette probabilité est forte. Pour l'élimination des bases l'algorithme trim_alt a été utilisé comme recommandé par les auteurs dans le manuel de PHRED. Cet algorithme est plus conservateur que l'algorithme trim et évite de perdre trop de bases de bonne qualité. Le seuil a été fixé au score phred minimal de 20 (une chance sur 100 au maximum qu'une base ne soit pas correcte) pour être conforme aux standards internationaux en vue de la soumission des séquences. A cette étape toutes les séquences sont considérées, même celles de longueur 0.

IV.2.3.1.3 Élimination du vecteur et des séquences courtes

Le vecteur correspondant à la partie comprise entre l'amorce et le site de clonage de l'insert est éliminé en retrouvant et effaçant les régions homologues à la séquence de lambda zap II et du site de restriction. Les séquences courtes sont ensuite éliminées avec pour critère principal la possibilité de pouvoir construire une sonde de 50 pb dans la séquence. Nous avons ainsi choisi de faire un BlastN pour ne pas tenir compte des séquences répétées dans la taille minimum de la séquence. A cette étape 14927 séquences ont été considérées comme valables, soit 80 % des séquences théoriques, sans aucune distinction en fonction du sens de séquençage (80 % en 3' comme en 5'). Cette valeur est plus faible de 7 points par rapport au jeu test (chapitre IV.2.2), mais reste cependant élevée. Le nom des séquences a été débarrassé des informations inutiles. Ces 14927 séquences ont été déposées dans les bases de données publiques avec leur nom au format de type ' BGA001A01P'. Ce format est aussi celui utilisé dans le manuscrit. En revanche, dans la base de données, disponible en annexe B (CD), ce

sont les séquences après assemblage des deux extrémités (décrit au paragraphe suivant) qui ont été utilisées. Le nom des séquences comporte en plus la longueur de la séquence et la présence ou l'absence de vecteur détecté aux extrémités (afin de savoir rapidement si le clone est complètement séquencé ou non). Dans le cas des clones assemblés, la position de la partie commune aux séquences 3' et 5' est aussi indiquée.

IV.2.3.1.4 Concaténation des extrémités 3' et 5'.

A ma connaissance, il n'existait pas de logiciel publiquement disponible capable de réaliser l'assemblage d'EST 3' et 5' provenant de même clone, le but d'un tel assemblage était (i) de pouvoir parer à l'éventualité de réaliser des microarrays à partir de produits PCR et donc de pouvoir identifier les clones chimériques plus simplement, (ii) de connaître quels clones ont été séquencés en totalité, cette information est utile pour éviter les doublons si un re-séquencage est prévu dans le futur. Cet assemblage n'a pas d'effet sur la taille finale du jeu d'unigènes puisque ces séquences auraient été, pour la plupart, assemblées dans l'étape de contigage. L'assemblage des extrémités est uniquement possible pour des gènes relativement petits (environ 1500 pb), puisqu'il faut que les séquences 3' et 5' soient suffisamment longues au niveau de leurs extrémités pour se concaténer. Les données pour chaque banque sont présentées dans le tableau 16. Un insert complet signifie que la totalité de l'insert a été séquencé, mais le protocole utilisé pour la construction des banques ne permet pas de garantir que la totalité de l'ARNm ait été introduit.

Tableau 16 : Nombre d'EST pour chacune des banques avant et après les étapes 2 à 4 de la figure 17 (les étapes de nettoyage), en fonction de l'orientation des séquences. Les inserts complets correspondent aux clones qui ont été entièrement séquencés.

	Banque stade vert		Banque stade mi-mûr		Banque stade mûr	
	Total EST séquencées	Après nettoyage	Total EST séquencées	Après nettoyage	Total EST séquencées	Après nettoyage
Séquences 3'	4608	2694	4608	2967	4608	3461
Séquences 5'	2016	914	2016	820	672	229
Inserts complets		830		858		233

A cette étape, il reste 13006 séquences redondantes, soit 69 % des séquences théoriques, réparties en 9122 séquences 3', 1963 séquences 5' et 1921 inserts complets. Après ces étapes, la taille moyenne des EST est de 527 pb. Aucune différence de taille significative entre les banques n'a été observée (Student, $p=0,95$).

IV.2.3.1.5 Assemblage

L'assemblage des séquences redondantes pour obtenir le jeu d'unigènes a été réalisé avec le logiciel TGICL (Perteau *et al.*, 2003). Cette étape consiste à regrouper les séquences identiques et les assembler pour obtenir le jeu d'unigènes. Dans un premier temps le logiciel regroupe les séquences homologues ; les séquences d'un même pré-cluster contiennent au moins une partie commune et sont donc potentiellement issues du même transcrit, mais l'homologie peut avoir une autre origine (chimères, variant d'épissage, région conservée...). A l'intérieur d'un même pré-cluster, on peut retrouver plusieurs unigènes (contigs ou singlets). 1646 pré-clusters ont été obtenus. Cette procédure permet d'obtenir des séquences consensus (contig) plus longues et plus complètes. 5219 unigènes (1793 contigs et 3426 singlets) sont obtenus avec ce logiciel. Sans l'utilisation de l'étape de pré-clustering, un millier d'unigènes supplémentaires avaient été obtenu en utilisant CAP3 seul, correspondant à des séquences redondantes non assemblées. L'analyse fine effectuée pour la détection des isogènes (chapitre IV.2.3.2.5) a permis de voir qu'il ne reste apparemment pas de séquence redondante avec cette approche, certains variants d'épissage sont cependant assemblés dans les contigs les plus grands.

5219 séquences sont donc considérées comme uniques, ce qui correspond à 40 % des séquences totales de départ. Par rapport aux projets de tailles équivalentes disponibles sur le site web du TIGR, ce taux se situe dans la moyenne : *Haplochromis chilotes* (20000 séquences totales, 30 % d'indépendantes) *Haplochromis sp* red tail sheller (13752, 45 %), *Onchocerca volvulus* (12264, 33 %). Sur l'ensemble des unigènes, 4546 contiennent au moins une séquence ancrée en 3' et sont donc des transcrits uniques correspondant à la définition des Unigènes par le NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/build1.html>). Une partie de la redondance n'a pas pu être identifiée et éliminée. Ceci est dû aux séquences 5' qui ne recouvrent pas les contigs ancrés en 3', bien que correspondant aux transcrits du même gène. Tant que les régions intermédiaires ne seront pas séquencées, cette redondance est inestimable, elle est de toute façon présente dans toutes les banques d'EST. Ainsi, il est envisageable que les 673 transcrits non ancrés en 3' correspondent au même gène qu'un transcrit ancré. La taille moyenne des unigènes est de 619 pb.

En ce qui concerne la nomenclature des unigènes, les singlets (séquence apparue une seule fois) conservent la même nomenclature que les EST. Dans ce manuscrit, les contigs sont présentés sous la forme suivante, PA[nombre]TC[nombre] (exemple : PA10TC1), le premier nombre correspond au numéro de cluster auquel il appartient et le second est le numéro de

contig à l'intérieur du cluster. Dans la base de données, la nomenclature est celle produite par le logiciel TGIC1 : CL[n°de cluster]Contig[n°de contig] (exemple CL10Contig1). Pour chaque contig cité dans le manuscrit, une EST représentative est indiquée en annexe A.

IV.2.3.2 Informations relatives aux EST

IV.2.3.2.1 Analyses de la redondance des EST

Le nombre d'EST par contig a été déterminé par analyse du fichier de sortie de CAP. 26 % (3426 séquences) des EST totales sont considérées comme singlets (Fig. 18).

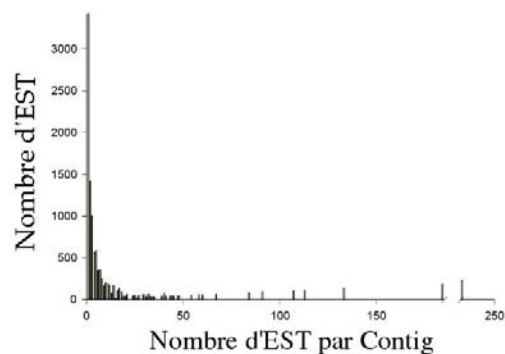


Figure 18 : Distribution des EST en fonction de la taille du contig.

Les séquences restantes ont été regroupées en 1793 contigs. Le nombre de séquences dans les contigs s'étend de 2 (712 contigs) à 283 (1 contig). Quatre contigs regroupent plus de 1 % du total des EST redondantes et les 2 plus abondants représentent respectivement 2,58 % et 2,53 %. Le gène le plus exprimé dans les fruits d'abricot (PA1TC12, 283 EST) dans les trois stades a déjà été décrit comme étant potentiellement une protéine allergène (accession GenBank : AAD32205, déposée par Mbéguié-Mbéguié *et al.*, 1999). Deux correspondent à deux isoformes d'une protéine liée au stress abscissique (PA2TC1 272 EST, PA2TC7 139 EST) dont une isoforme (PA2TC7) a déjà été identifiée chez l'abricot (ASR, Mbéguié-Mbéguié *et al.*, 1997). Ces deux isoformes sont extrêmement proches, et il semble que l'assemblage des séquences identiques n'ait pas été optimal, certaines séquences d'une isoforme se retrouvent à l'intérieur de l'autre contig. Cependant compte tenu des données de protéomique (chapitre IV.5), leur forte abondance respective ne fait pas de doute.

Le quatrième transcrit (PA4TC2, 149 EST) fortement exprimé dans le fruit est homologue à une protéine tumorale (translationally controlled tumor protein), liée au grandissement cellulaire. Sa présence a déjà été décrite chez les mammifères, *Saccharomyces cerevisiae* et d'autres plantes supérieures (Sanchez *et al.*, 1997 ; Sage-Ono *et al.*, 1998). Cette

protéine n'a jamais fait l'objet d'étude chez les fruits, cependant il semblerait qu'elle soit aussi fortement exprimée chez la tomate (D. Page, communication personnelle).

IV.2.3.2.2 Redondance intra et inter-banque

Pour cette partie, les séquences 5' n'ont pas été incluses dans les données statistiques des contigs à cause du taux de séquençage inégal entre les banques (Tab. 16).

La redondance est légèrement inférieure dans la banque du stade vert puisque qu'elle contient 2214 unigènes dont 1366 sont spécifiques de cette banque (Fig. 19), par comparaison aux stades mi-mûr et mûr (respectivement 1914 et 1995 unigènes dont 961 et 1042 spécifiques). Ceci indique que la transcription serait plus homogène au stade vert, avec moins de gènes fortement exprimés. Seulement 359 unigènes (7,9 %) sont communs à l'ensemble des trois banques et les intersections entre la banque du stade vert et les banques des stades mi-mûr (583 séquences) et mûr (624 séquences) sont plus faibles qu'entre les banques des stades mi-mûr et mûr (688 séquences). Les représentations du transcriptome des banques des stades mi-mûr et mûr semblent être plus proches entre elles (329 unigènes communs) que celle de la banque du stade vert. Par contre, il est étonnant que la représentation du transcriptome dans la banque du stade mûr soit plus proche de celle de la banque du stade vert (265 unigènes communs) que celle de la banque du stade mi-mûr (265 unigènes communs).

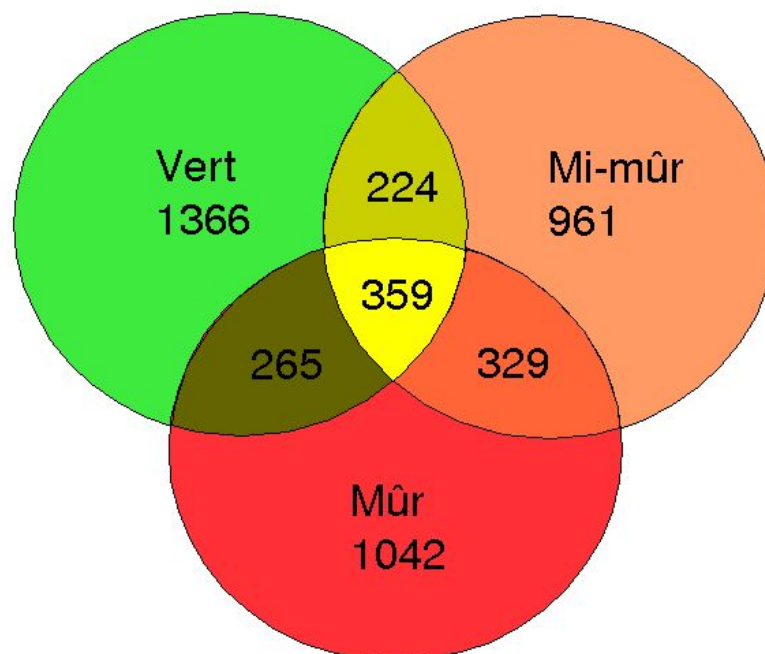


Figure 19 : Distribution des unigènes en fonction de la banque dans laquelle ils ont été détectés

IV.2.3.2.3 Homologie par rapport aux bases de données publiques

Sur les 5219 unigènes analysés, 77 % présentent une homologie avec les bases de données publiques (Fig. 20) dont 56 % pour lesquels la fonction de l'homologue est connue. La proportion d'unigènes avec un homologue dont la fonction est connue est identique dans les trois banques. 13 % des unigènes ont une similarité relativement faible ($80 > \text{score BLASTX} > 50$) avec les séquences publiques. Ces faibles scores sont certainement liés à une moyenne de la taille des séquences significativement plus courte dans cette catégorie (t-test, $p=0,95$, 527 pb vs 706 pb).

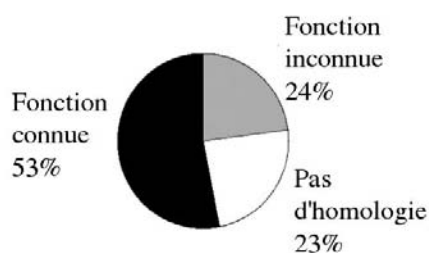


Figure 20 : Distribution des homologies pour les unigènes : (i) présentant une homologie pour des protéines de fonctions connues (noir) ; (ii) présentant une homologie pour des protéines hypothétiques de fonctions inconnues (gris) ; et (iii) ne présentant aucune homologie dans les bases de données (blanc).

23 % des unigènes ne présentent pas d'homologie avec les bases de données publiques protéiques. La présence de gènes de fonctions connues dans ce groupe ne peut pas être exclue car leur longueur moyenne est significativement plus courte que le reste des unigènes (t-test, $p=0,95$, 434 ± 175 pb). Étant donné que la région 3' est en moyenne longue de 259 pb, il reste seulement 53 acides aminés pour l'identification en BLASTX, ce qui est souvent trop juste selon nos critères (au moins 50 aa identiques).

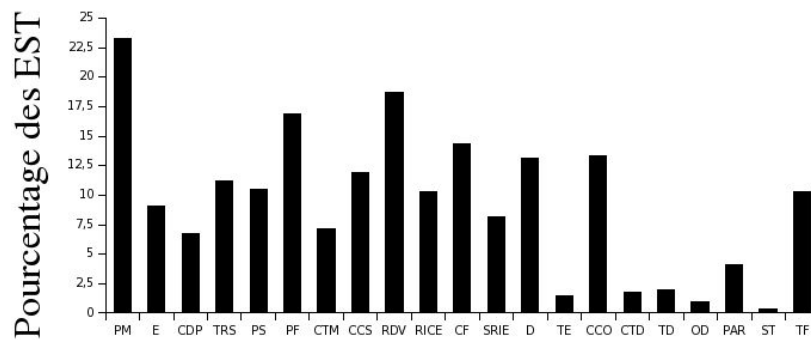
37 % (1939 séquences) des unigènes d'abricot ont un homologue en BLASTN dans les bases de données de séquences de *Prunus* existantes (pêche et amande). Seulement 14 % (200 séquences) des séquences sans homologue dans les bases de données publiques ont un homologue dans les bases de données *Prunus*.

IV.2.3.2.4 Distribution des unigènes en catégories fonctionnelles

La procédure d'annotation automatique utilisée pour déterminer les catégories fonctionnelles possède un certain nombre d'avantages par rapport à une stratégie manuelle reposant sur la bibliographie. (i) Le gain de temps est considérable, alors qu'il faut compter environ un mois pour 1000 gènes, cette opération est réalisable en quelques jours. (ii) L'assignation d'un gène à plusieurs catégories fonctionnelles est possible. (iii) Le niveau de

précision de l'annotation est beaucoup plus élevée et permet de classer les gènes dans des voies métaboliques très précises. (iii) D'autres informations sont disponibles, comme la localisation cellulaire. (iiii) Les annotations des protéines d'*Arabidopsis* effectuées au MIPS sont toutes validées par rapport à la bibliographie et sont très fiables. Un aperçu des possibilités est présenté en annexe C. La partie automatique de cette procédure est utilisée pour l'annotation d'autres projets d'EST de plantes dans la base de données Sputnik (Rudd *et al.*, 2003). 3177 unigènes ont été annotés avec cette procédure et 248 unigènes sans homologue chez *Arabidopsis* l'ont été à partir d'études bibliographiques.

65 % (3425 séquences) des unigènes ont été assignés à des catégories fonctionnelles. Cette valeur est plus élevée que le nombre d'unigènes homologues à des protéines de fonctions connues car l'analyse des motifs des protéines inconnues d'*Arabidopsis* permet de les assigner à des catégories fonctionnelles. La figure 21 présente la répartition en catégories fonctionnelles, comptée en nombre d'EST (somme des unigènes * abondance respective). La répartition des unigènes est essentiellement la même, à l'exception de la catégorie Protection Défense et Virulence, qui contient beaucoup de protéines très redondantes (ASR, métallothionéines, HSPs, protéines de tolérance au sel). Six catégories fonctionnelles regroupent plus de 12,5 % : Métabolisme, Adressage des Protéines, Développement, Organisation Cellulaire, Destin des Cellules et Protection, Défense et Virulence (protéines de réponse aux stress). Les catégories fonctionnelles ne sont pas significativement variables pendant le développement. Du point de vue des cibles potentiellement intéressantes pour la qualité des fruits, la sous-catégorie Métabolisme Secondaire (couleur, arôme), comprise dans la catégorie Métabolisme, contient 292 unigènes dont, par exemple, 13 impliqués dans la biosynthèse des dérivés de l'Isopentényl Pyrophosphate (IPP) (caroténoïdes). La sous-catégorie Réponse à la Régulation Hormonale (catégorie Régulation Systémique / Interaction avec l'Environnement) chez les plantes contient 50 unigènes, cependant un certain nombre de Facteurs de Transcription (334 unigènes) et de protéines liées à la Transduction du Signal (599 unigènes), peuvent intervenir dans cette catégorie. La sous-catégorie Paroi Cellulaire (liée à la perte de fermeté) (catégorie Contrôle de l'Organisation Cellulaire) représente 141 unigènes. La catégorie Transport Facilité (liée au transport des assimilats, sucres/acides) représente 503 unigènes. La plupart de ces unigènes a été déposée sur les lames de microarray (chapitre IV.3.2.1).



Catégories fonctionnelles

Figure 21 : Distribution des EST d'abricot en catégories fonctionnelles putative. Catégories d'après le catalogue fonctionnel du Munich Information Center for Protein Sequence (<http://mips.gsf.de>). PM: Métabolisme ; E: Énergie ; CDP: Cycle Cellulaire et traitement de l'ADN ; TRS: Transcription ; PS: Synthèse des Protéines ; PF: Adressage des Protéines ; CTM: Transports Cellulaire et Mécanisme de Transport ; CCS: Communication Cellulaire et Transduction du Signal ; RDV: Protection, Défense et Virulence ; RICE: Régulation / Interaction avec l'Environnement des Cellules ; CF: Destin des Cellules ; SRIE: Régulation Systémique / Interaction avec l'Environnement ; D: Développement ; TE: Eléments Transposables ; CCO: Contrôle de l'Organisation Cellulaire ; CTD: Différenciation des Cellules ; TD: Différenciation des Tissus ; OD: Différenciation des Organes ; PAR: Régulation de l'Activité des Protéines ; SP: Protéines de Stockage ; TF: Transport Facilité.

La localisation potentielle des protéines correspondant aux unigènes est présentée dans la figure 22. La localisation cellulaire a pu être effectuée pour 2237 unigènes, dont presque la moitié se situe dans le cytoplasme (46 %). Les autres localisations prédominantes sont le noyau, les mitochondries et les membranes. (le catalogue du MIPS regroupe toutes les protéines membranaires dans la catégorie membrane plasmique !!).

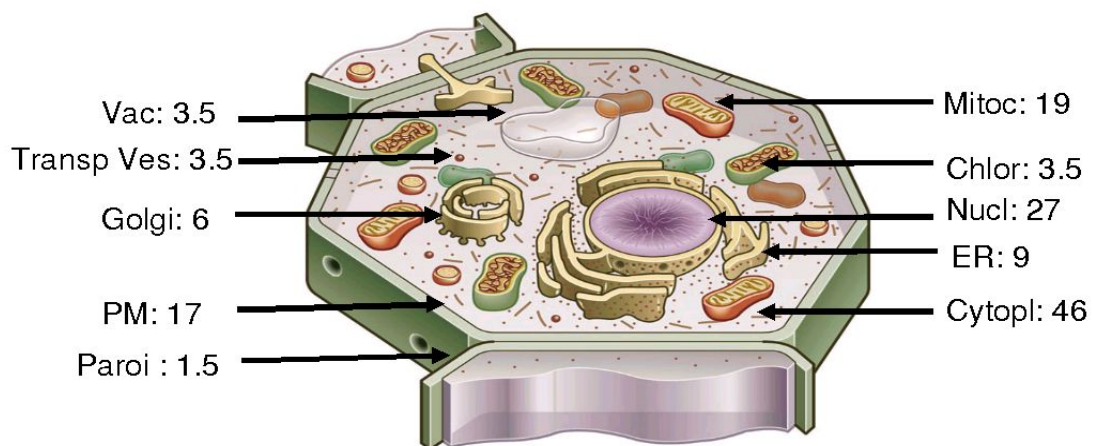


Figure 22 : Localisation cellulaire potentielle des unigènes dans les principaux compartiments cellulaires. Les valeurs sont exprimées en pourcentage des unigènes localisés. Vac: Vacuole ; Transp Ves: Vésicule de transport ; Golgi: Appareil de Golgi ; PM: Membrane plasmique ; Paroi: Paroi cellulaire ; Mitoc: Mitochondrie ; Chlor: Chloroplaste ; Nucl: Noyau cellulaire ; ER: Réticulum endoplasmique ; Cytopl: Cytoplasme.

IV.2.3.2.5 Analyse des familles multigéniques

Le but de l'identification et de la différenciation des différents membres des familles multigéniques est de pouvoir facilement discriminer leur niveau d'expression au cours des études d'expression des transcrits (Northern électronique, PCR quantitative et microarray) et des protéines (protéomique). Dans le premier cas, il est nécessaire d'avoir la possibilité de construire une sonde spécifique de chaque isogène qui se situe généralement dans la région 3' non codante, sauf dans le cas du Northern électronique qui est plus tolérant puisqu'il permet de prendre en compte l'ensemble de la séquence. Dans le second cas, la détermination exacte du transcrit correspondant à une protéine est dépendante à la fois du niveau d'homologie entre les séquences traduites des isogènes et du niveau de recouvrement de la carte massique sur la séquence. De plus, sur des génomes non entièrement séquencés cette détermination reste spéculative en utilisant la technique MALDI : il faut que l'unigène corresponde à une séquence complète et que 100 % de la séquence aient été identifiés en spectrométrie. Malheureusement, rencontrer toutes ces conditions est extrêmement rare avec le MALDI-TOF.

Deux catégories d'isogènes ont été définies pour distinguer ceux dont la construction de sondes spécifiques est critique (isogènes niveau nucléotidique : INN), des autres (isogènes niveau protéique, INP). Les INN sont homologues au niveau de leur séquence EST et sont identifiés en utilisant le BlastN avec comme seuil une E-value de $1e^{-10}$. Il semblerait que l'homologie soit suffisamment forte entre les membres de ces familles pour qu'une hybridation avec des ADNc incluant une partie codante conduise à une hybridation croisée (cette hypothèse est validée dans le paragraphe suivant). Les INP ne sont pas homologues dans ces conditions, il faut réaliser un tblastX pour faire apparaître l'homologie sur leurs séquences conceptuelles en acides aminés. Les séquences des INN sont généralement excessivement proches, les différences se situent généralement au niveau de la région 3' non codante et de la troisième base des codons. De fait les INN sont inclus dans les INP. Chaque famille multigénique a été validée en utilisant le clustalX pour identifier les chimères qui induisent une homologie erronée. 141 chimères ont été identifiées avec cette procédure.

1256 (25 %) unigènes ont été identifiés comme appartenant à des familles multigéniques (INP), ils sont regroupés en 455 familles. 617 (12,5 %) ont été groupés en 255 familles INN. 55 familles multigéniques INP regroupent les deux types d'isogènes, les autres sont constituées soit uniquement d'INN soit uniquement d'INP.

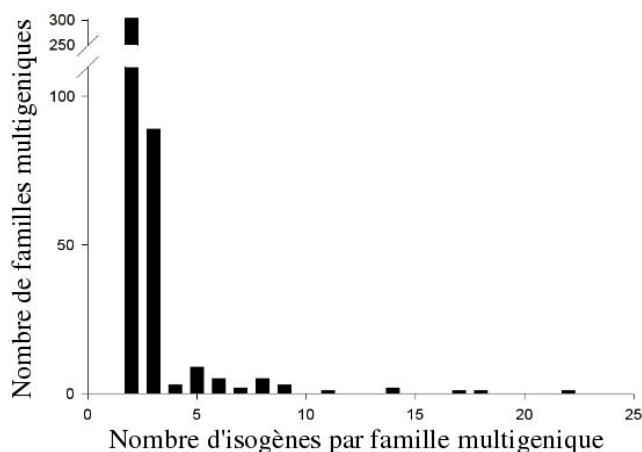


Figure 23 : Répartition des isogènes dans les familles multigéniques

La taille des familles multigéniques varie de 2 membres (300 familles) à 22 membres (1 famille) (Fig. 23). Les dix familles multigéniques majoritaires sont présentées dans le tableau 17. On retrouve une certaine similitude avec les familles identifiées chez la tomate (Van der Hoeven *et al.*, 2002). Cinq des dix familles multigéniques prédominantes chez la tomate sont aussi fortement représentées chez l'abricot (protéine kinases, cytochrome P450s, RAS related GTP binding protéines, Glucosyltransférases, Ubiquitine conjuguée), ce recouvrement serait peut être augmenté si on ne prenait en compte que les EST de fruits dans la tomate. Ce résultat confirme la conservation des familles multigéniques entre les espèces végétales et met en évidence le caractère ancestral de la multigénicité pour ces familles. Les cinq autres familles sont plus spécifiques à l'abricot et l'étude de la chronologie ainsi que la localisation de la duplication de ces gènes pourrait aider à comprendre l'évolution et la radiation des *Prunus*.

Tableau 17 : Les dix plus grandes familles multigéniques chez l'abricot et la tomate, nom d'une EST représentative de la famille chez l'abricot, fonction probable et nombre d'unigènes appartenant à la famille.

EST d'abricot représentative	Fonction putative de la famille	Nombre de membres	
		Abricot	Tomate
BGA006J21	Protéine kinases	22	139
BGA006B01	RAS related GTP binding	19	48
BGA001E21	Ubiquitine conjuguée	17	37
BGA006P24	Cytochrome P450s	14	60
BGA003J17	Glucosyltransférases	12	35
BGA006O05	Ubiquitine	11	
BGA002A22	Auxine-induite	9	
BGA001J15	Heat shock	9	
BGA002J07	Major Allergen	9	
BGA002C20	Flavonol Synthase	8	
	Péroxydase		53
	Myb		41
	E8		39
	NAC domaine		34
	Glutathione S transférase		22

La plupart de ces dix familles multigéniques est potentiellement impliquée dans l'évolution de la qualité du fruit : réponse aux hormones (kinase, GTP binding, Auxine-induite), ou aux stress (heat shock, allergen), et au métabolisme secondaire (Glucosyltransférase et Flavonol Synthase). Tous ces isogènes ont été retrouvés dans les 3 banques de fruits étudiées et sont probablement beaucoup plus multigéniques sur l'ensemble de la plante. Au cours des études d'expression, il sera indispensable de pouvoir discriminer leur expression.

IV.2.3.2.6 Test d'hybridation

Dans le but d'évaluer le risque d'hybridation croisée entre les membres de familles INN, deux couples de INN ont été hybridés. Les quatre séquences utilisées correspondent à des transcrits complètement séquencés. Le premier couple (bge004h17, 755 pb ; bga003p17, 607 pb) correspond à un allergène d'abricot et les séquences ont 69 % d'homologie entre elles. Le second (bgh001d20, 1363 pb ; bgh001i21, 936 pb) est similaire aux expansines et les deux isogènes ont 67 % d'homologie entre eux au niveau nucléique. L'hybridation croisée de ces deux couples a été confirmée par Southern blot (Fig. 24). Chaque unigène hybride à la fois sur lui même et sur son isogène.

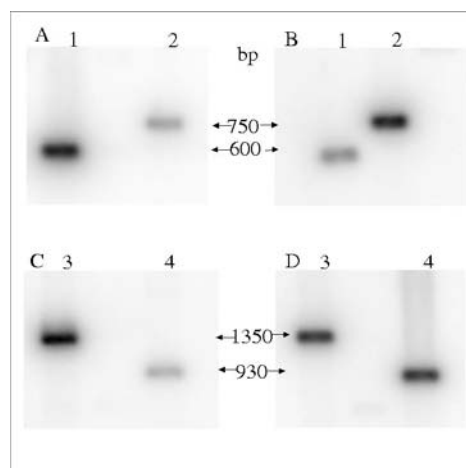


Figure 24 : Southern blot d'ADNc d'isogènes. Ligne 1 : *bga003p17* ; ligne 2 : *bge004h17* ; ligne 3 : *bgh001d20* ; ligne 4 : *bgh001l21*. **A** la sonde est l'ADNc de *bge004h17*. **B** la sonde est l'ADNc de *bge004h17*. **C** la sonde est l'ADNc de *bgh001d20*. **D** la sonde est l'ADNc de *bgh001l21*.

IV.2.3.2.7 Estimation de l'expression des gènes pendant le développement

Le Northern électronique a été effectué en considérant la redondance de chaque unigène dans chacune des banques. Les unigènes différemment exprimés ont été classés en trois catégories d'après la significativité statistique de leur induction comme décrit par Schmitt *et al.* (1999). Le profil d'expression de deux gènes représentatifs des catégories 1 et 3 a été vérifié par PCR quantitative. Seul le gène de la catégorie 1 ($p=99,1$ ratio > 10) a montré un profil d'expression identique à celui obtenu par Northern électronique (Fig. 25).

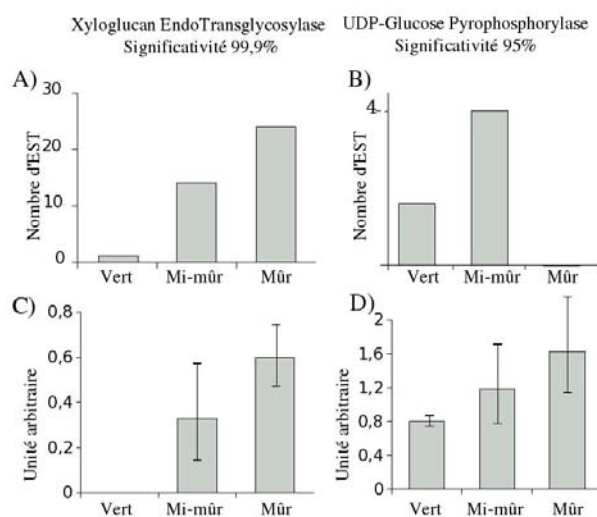


Figure 25 : Comparaison du profil d'expression des gènes des classes 1 et 3 obtenus par PCR quantitative et Northern électronique. **A** Expression de la Xyloglucan endotransglycosylase par Northern électronique dans les 3 banques d'EST. **B** Expression de l'UDP-Glycosyl Pyrophosphorylase par Northern électronique pour les 3 banques d'EST. **C** Expression de la Xyloglucan endotransglycosylase par PCR quantitative dans les 3 stades de maturité correspondant au trois banques d'EST. **D** Expression de l'UDP-Glycosyl Pyrophosphorylase par PCR quantitative pour les 3 stades de maturité correspondant au trois banques d'EST.

Les gènes de la classe 1 sont présentés dans le tableau 18 (l'ensemble des 3 classes est en annexe B (CD)). Ces 15 gènes ont un ratio d'expression supérieur à 10 et une significativité supérieure à 99,9 % (test du χ^2) entre au moins deux des trois stades.

A quelques exceptions près, ils peuvent être classés en 3 catégories : les gènes relatifs au métabolisme de la paroi cellulaire (5 unigènes), les gènes impliqués dans la défense contre les bio-agresseurs, la détoxification et la réponse aux stress (5 unigènes) et ceux de fonctions inconnues (4 unigènes).

Tableau 18 : Unigènes fortement différenciellement exprimés dans au moins un stade de développement correspondant à ceux des banques avec un ratio d'expression supérieur à 10 et une p value $<0,01$.

Unigène	Meilleur homologue dans les bases de données	Fonction potentielle	EST par banque		
			Vert	Mi-mûr	Mûr
PA17TC1	BAB78506	Xyloglucan endotransglycosylase	1	14	24
PA5TC5	AAC33530	Expansine	0	16	25
PA5TC7	AAG51534	Pectinméthylestérase inhibiteur	1	21	18
PA7TC6	AAM62643	pectinestérase inhibiteur/GRIP 28	0	23	20
PA6TC8	AAK25850	Pectate lyase	1	33	24
PA6TC3	CAA39998	Catalase	17	12	1
PA32TC1	CAB71056	DNA binding protein	14	4	0
PA92TC1	AAC28935	Polyphénol oxydase	10	0	0
PA12TC3	CAB36910	Heat shock protein 18	1	18	35
PA1TC49	CAA59472	Proline rich protein	2	32	33
PA9TC1	CAA54449	ACC oxydase	3	10	47
PA19TC3	AE005172	Inconnue	0	10	17
PA3TC4	Aucun	Pas d'homologie	2	33	49
PA62TC1	Aucun	Pas d'homologie	0	0	15
PA38TC1	Aucun	Pas d'homologie	12	1	3

- Gènes impliqués dans le métabolisme de la paroi cellulaire :

Cinq unigènes sont impliqués dans le métabolisme de la paroi, ils sont tous sur-exprimés dès le stade mi-mûr. Deux d'entre eux ont une expression qui augmente linéairement au cours du développement, la xyloglucan endotransglycosylase (PA17TC1) et l'expansine (PA5TC5). Les deux inhibiteurs de pectinestérase (PA5TC7, PA6TC8) présentent une

expression constante après le stade mi-mûr. La pectate lyase (PA6TC8) semble avoir un profil transitoire. La fonction de ces enzymes a été décrite dans l'étude bibliographique (chapitre II.4.1.1) excepté les inhibiteurs de pectinestérase, qui, comme leur nom l'indique inhibent les pectinestérases par la formation d'un complexe non covalent (Giovane *et al.*, 1995).

- Gènes relatifs à la détoxification, virulence, résistance aux pathogènes et réponse aux stress :

L'expression de la catalase (PA3TC6) décroît linéairement au cours de la maturation. Elle intervient dans le cycle du glutathion pour la détoxification des ions peroxydes (Shigeoka *et al.*, 2002). Les expressions d'une protéine liée à l'ADN (PA32TC1) et de la polyphénol oxydase (PA92TC1) diminuent après le stade vert. La première appartient à la famille des remorines sur lesquelles peu de choses sont connues, mis à part que le domaine N terminal est riche en proline, que le domaine C terminal peut interagir avec l'ADN et que son comportement présente des similitudes avec des protéines virales intervenant dans la communication cellulaire (Reymond *et al.*, 1996). La polyphénol oxydase est déjà connue chez l'abricot (Chevalier *et al.*, 1999). Elle est impliquée dans le brunissement des fruits. Une protéine de choc thermique (PA12TC3) présente une augmentation linéaire de son expression pendant le développement. Ces protéines sont généralement liées à la tolérance à la sécheresse ou au stress osmotique (Sun *et al.*, 2001). Une protéine riche en proline (PA1TC49) est induite dès le stade mi-mûr, elle est homologue à une protéine de transfert de lipide (LTP) de tomate qui semble être impliquée dans la réponse aux pathogènes (Pan *et al.*, non publié). Elle présente également une homologie avec des extensines et peut donc être liée au métabolisme de la paroi.

- Gènes de fonctions inconnues :

Deux gènes de fonctions inconnues présentent un niveau d'expression qui augmente linéairement au cours de la maturation. Le seul des quatre gènes dont un homologue existe (PA19TC3) appartient à la famille de protéines inconnues DUF 538 apparemment spécifique des plantes (entrée Interpro IPR007493). Le second gène (PA3TC4) semble spécifique des Rosacées puisqu'il est uniquement présent dans les banques d'EST de pêche, d'amande et de pomme.

Un gène de fonction inconnue (PA62TC1) est induit au stade mûr, ce gène semble être spécifique des *Prunus*, puisque hormis l'abricotier, il n'a été identifié que chez la pêche. Cependant il est étonnant qu'il n'ait pas été retrouvé dans les banques d'EST d'amande étant donné sa forte expression chez les autres *Prunus*.

Un gène de fonction inconnue (PA38TC1) est réprimé après le stade vert. Il a été retrouvé dans plusieurs espèces ligneuses, avec une plus grande homologie chez les Rosacées (pomme, amande). Il est absent des banques de pêche, mais celles-ci n'ont pas été construites à partir de fruits verts. Chez les autres espèces, il a surtout été observé dans des organes floraux.

Comme observé par Mbéguié-Mbéguié *et al.* (1999), l'ACC oxydase est fortement induite pendant la maturation.

IV.2.3.2.8 Ajustement des données de 1996 et de 2002 pour les PCR quantitatives.

Afin de valider les résultats de Northern électronique par PCR quantitative, les données 2002 de physiologie du développement des fruits de Bergeron ont été calées sur celles de 1996. Les trois stades les plus proches de ceux qui ont servi à construire les banques ont été déterminés par rapport au début du dégagement d'éthylène (banque stade mi-mûr, 1996 ; stade 107 JAA 15/07/2002) (Fig. 27). Le stade vert est situé un mois avant le dégagement d'éthylène, ce qui correspond au stade 65 JAA (17/06/2002) et le stade mûr une semaine après, soit peu avant le maximum d'accumulation des sucres, ce qui correspond au stade 114 JAA (22/07/02).

IV.2.3.2.9 Validation de l'expression des gènes par PCR quantitative

L'expression de 11 des 15 gènes identifiés comme différentiellement exprimés par Northern électronique a été évaluée par PCR quantitative (Fig. 26) sur des stades de maturité identique provenant de la campagne de récolte 2002. Les gènes non traités sont ceux dont l'expression avait déjà été validée dans des travaux antérieurs (PPO et ACO), ainsi que les deux PME1.

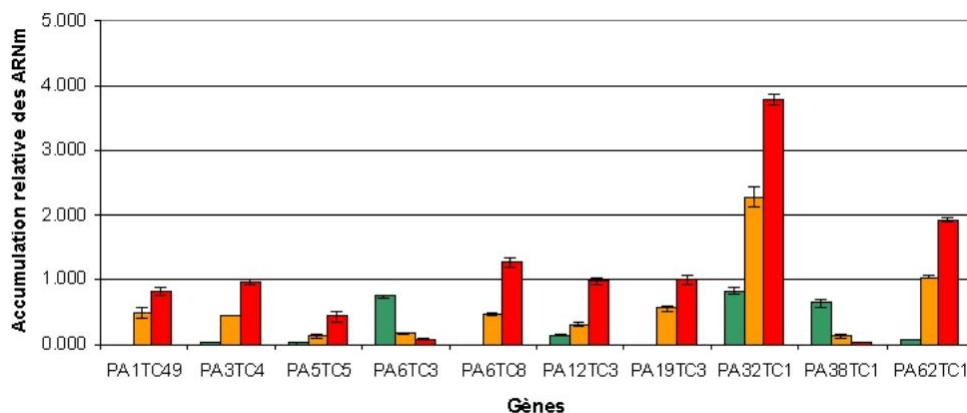


Figure 26 : Profils d'expression obtenus par PCR quantitative des unigènes dont l'expression varie significativement en Northern électronique.

Pour neuf unigènes le profil d'expression est identique à ce qui a été observé par Northern électronique. Seul le profil d'expression de PA32TC1 est à l'opposé. Le profil général d'expression de PA62TC1 est cohérent avec ce qui a été observé par Northern électronique, cependant, il est plus induit dans le stade mi-mûr utilisé en PCR quantitative. Les stades mi-mûr 1996 et 2002 ne sont pas strictement identiques, il est possible que ce gène soit fortement induit de façon très transitoire, et son expression serait fortement liée à l'explosion du dégagement d'éthylène.

IV.3 Etude du transcriptome par microarray

IV.3.1 Ajustement des stades de développement entre les différentes variétés

Deux stades ont été choisis, par variété, pour les expériences de microarray à des fins de comparaison variétale (Fig. 27).

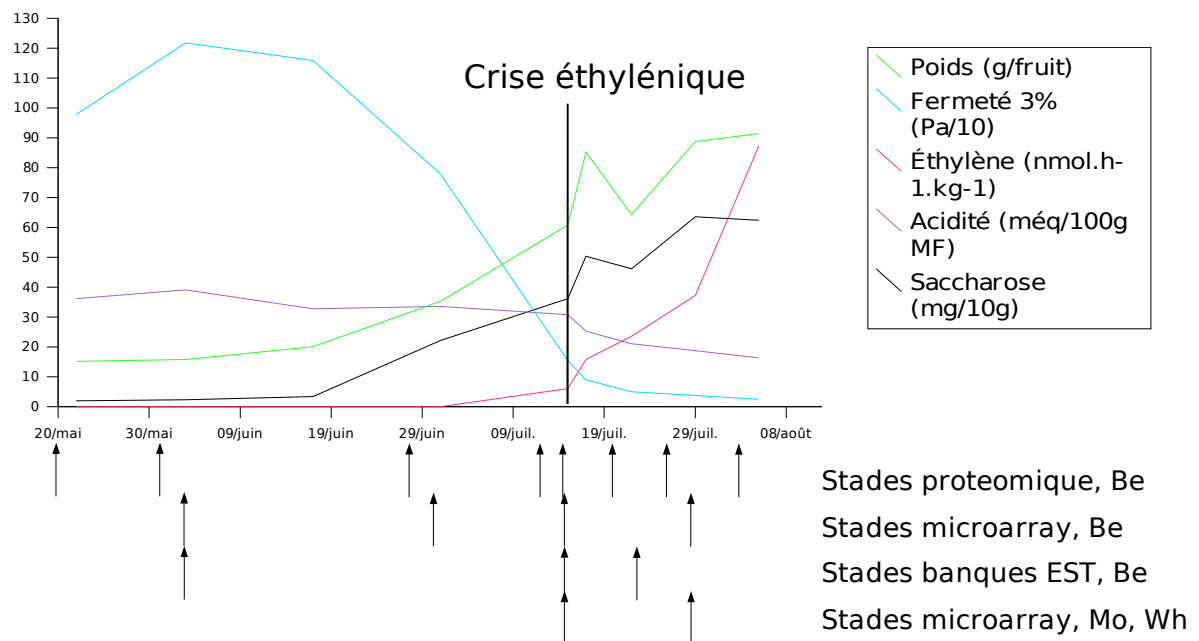


Figure 27 : Synthèse de l'évolution des principaux paramètres de qualité pendant la maturation et identification des stades utilisés dans les différentes approches. Les courbes d'évolution sont celles de Bergeron. Fermeté exprimée en Pascal. RI, indice réfractométrique, exprimé en degrés Brix. Acidité, en méq pour 100 g. Poids en gramme. Ethylène en nmol.h⁻¹ pour 100 g. Be : Bergeron, Mo : Moniqui, Wh : Blanc.

Le premier point correspond au début du dégagement d'éthylène : Bergeron 15/07/02, Moniqui 24/06/02. Il correspond aussi au stade où la moitié des sucres est accumulée, 08/07/02 pour Blanc. Le second stade est au maximum d'accumulation des sucres : Bergeron 29/07/02, Blanc 29/7/02 et Moniqui 08/07/02.

IV.3.2 Conception des lames et analyse statistique des résultats

IV.3.2.1 Conception des lames

IV.3.2.1.1 Design des oligonucléotides

Compte tenu de l'impossibilité de pouvoir déposer l'ensemble des unigènes obtenus, un choix a été effectué pour identifier les candidats à priori les plus pertinents pour le déterminisme de la qualité du fruit.

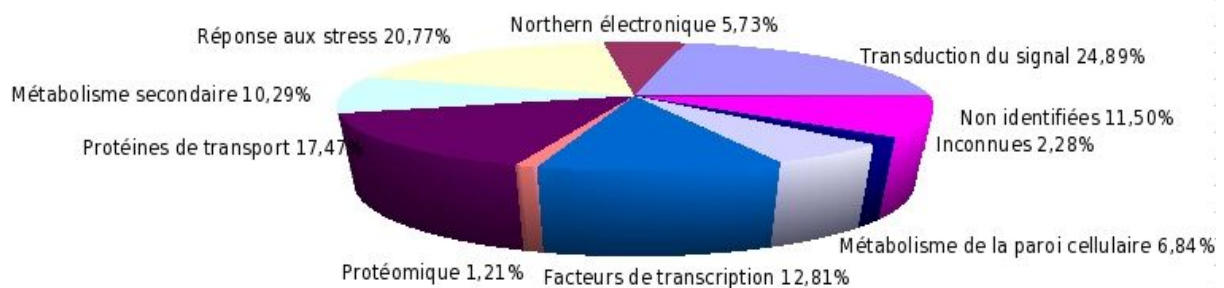


Figure 28 : Répartition en catégories fonctionnelles des 1796 unigènes sélectionnés pour la synthèse des oligonucléotides. "Protéomique" et "Northern électronique" correspondent respectivement à des unigènes dont la variation de l'abondance des protéines et des transcrits au cours du développement a été mise en évidence par les autres approches. Parmi les unigènes, 409 ont été classés dans deux ou plusieurs catégories fonctionnelles.

La figure 28 présente la répartition en catégories fonctionnelles des unigènes sélectionnés. Les gènes reliés à la biosynthèse des pigments et des composés d'arômes appartiennent à la catégorie métabolisme secondaire ; les gènes reliés à la perte de fermeté, à la catégorie métabolisme de la paroi cellulaire ; les gènes impliqués dans le transport des assimilats (sucres/acides) appartiennent à la catégorie protéines de transport ; les gènes impliqués dans la régulation hormonale ou abiotique sont classés dans les catégories facteurs de transcription, transduction du signal et réponse aux stress. Les gènes d'intérêt ont été complétés par des protéines inconnues, des transcrits non identifiés, des unigènes dont le

niveau d'expression, mesuré par approche EST (Northern électronique) et protéomique, varie et des gènes déjà identifiés chez l'abricot, dont la plupart correspond à des cibles déjà incluses dans les gènes d'intérêt.

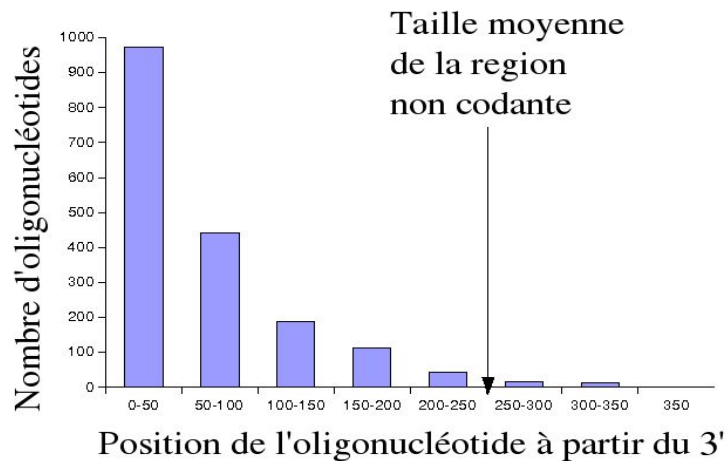


Figure 29 : Répartition des oligonucléotides en fonction de la position dans la séquence de l'unigène

Les oligonucléotides ont été définis sur la séquence du brin codant de l'ADNc dans la partie de la séquence située la plus en aval possible, ceci afin de se trouver dans la partie 3' non codante. La limite maximale a été fixée à la 400^{ème} paire de base. Cependant il n'y a jamais eu besoin d'atteindre cette limite. La moyenne de la longueur du non codant a été précédemment calculée à 256 pb, seulement 2 % des oligonucléotides ont été définis en amont de cette limite (Fig. 29). L'écart type sur la longueur moyenne du non codant est de 93 pb, la répartition suivant une loi normale :

- 65 % des oligonucléotides sont situés à une position plus en aval de la moyenne -2σ (70 pb à partir de l'extrémité 3') et d'après la loi normale (95 % des transcrits ont une taille de non codant comprise entre la moyenne et $\pm 2\sigma$, donc 97,5 % inférieur à la moyenne -2σ), 2,5 % des transcrits ont une longueur de non codant inférieure à cette valeur. Ce qui signifie que ces transcrits ont une probabilité de commencer dans le non codant de 97,5 %.

- De la même façon, les 27 % des oligonucléotides qui sont situés à une position entre la moyenne -2σ et la moyenne $-\sigma$ (entre 70 et 163 pb à partir de l'extrémité 3') ont une probabilité de commencer dans le non codant de 84 %.

- Les 8 % restant ont une probabilité inférieure à 84 %.

Les oligonucléotides sont, dans l'état actuel des connaissances sur le transcriptome de l'abricot, tous spécifiques de la séquence connue de chaque gène. Il n'est pas possible de prévoir si les oligonucléotides sont spécifiques vis à vis du génome entier de l'abricotier, bien que des garanties soient prises en ayant défini les oligonucléotides dans le 3' non codant.

IV.3.2.1.2 Analyse de l'hybridation des lames

Dans l'expérience de suivi de développement, un design en boucle sur quatre stades de maturité a été utilisé (Fig. 10), 8 lames de microarray ont donc été hybridées. Ceci signifie que l'expression de chaque unigène (ou oligonucléotide) a été mesurée 8 fois par condition : pour chacune, la valeur est mesurée deux fois sur la même lame, cette expérience est réalisée en double avec inversion de fluorophores et chaque stade est comparé à deux autres. Pour chaque unigène, il existe donc en théorie 32 valeurs d'intensité.

1566 oligonucléotides (87 %) présentent, dans au minimum 97 % des cas, une valeur d'intensité au moins supérieure à deux fois le bruit de fond. Ces gènes ont été conservés pour les analyses statistiques. Parmi les autres unigènes, on distingue trois cas :

- Les spots sur lesquels aucune hybridation n'a été détectée (47 spots) : soit les ARN cibles n'hybrident pas avec les oligonucléotides, soit pour des raisons techniques (problème de design ou de synthèse des oligonucléotides), soit pour des raisons biologiques (transcrits faiblement abondants). En effet, tous ces gènes sont faiblement représentés dans les banques d'EST : seulement 5 d'entre eux sont présents à plus de 2 copies et un seul est très abondant.

- Les spots où un répliquat sur 2 est hybridé (74) : le spot manquant appartient toujours à la partie supérieure de la lame, la non hybridation peut provenir d'un problème pendant le dépôt, ou de la présence d'un bruit de fond plus élevé dans cette région.

- Les spots où plus d'une donnée est manquante (109) : ce phénomène peut provenir de l'effet combiné de problèmes ponctuels d'hybridation, de bruit de fond local trop fort, et d'abondance de transcrits trop faible à certains stades du développement. Ce dernier point pose le problème de la viabilité statistique pour les transcrits totalement absents dans une ou plusieurs conditions. Ceux-ci n'ont donc pas pu être étudiés, cependant quelques uns seront tout de même commentés dans la discussion.

Dans l'ensemble le design effectué semble très performant puisqu'il ne peut être remis en cause que pour 2,6 % des oligonucléotides au maximum (oligonucléotides pour lesquels aucune intensité n'est détectée dans aucun des spots).

IV.3.2.2 Analyse statistique de l'expression des gènes

Après l'étape de normalisation, 1566 gènes (87 %) ont été conservés pour l'analyse du développement et 1560 pour l'analyse des différentes variétés. Les ACP et l'ANOVA ont été réalisés avec le logiciel geneANOVA (Didier *et al.*, 2002), les fichiers d'entrée correspondant aux données d'expression normalisées des gènes conservés et les fichiers de design expérimental sont en annexe B (CD).

IV.3.2.2.1 ACP

L'analyse en composantes principales (ACP) est une technique statistique permettant de réduire un système complexe de corrélations en un plus petit nombre de dimensions. Le but ici est de visualiser la corrélation entre les facteurs de variabilité dans les expériences d'hybridation. A priori, les conditions expérimentales proches doivent se retrouver à proximité les unes des autres, par exemple des stades de maturité adjacents (si la plupart des mécanismes physiologiques liés à la maturité évolue en sens unique pendant la maturation).

La figure 30A représente l'ensemble des expériences réalisées dans l'étude du développement des fruits de Bergeron, chaque condition est représentée par huit jeux de données. La distance entre les jeux de données d'une condition permet effectivement de distinguer les stades de maturité. Cependant une grande proximité entre le point correspondant à la crise climactérique (15/07, orange) et le point correspondant à un stade mûr (29/07, rouge) est observée.

La figure 30B représente l'ensemble des expériences réalisées dans l'étude des différentes variétés. Les points correspondant à des jeux de données différents permettent de distinguer clairement le facteur variété. Seuls les jeux de données des deux stades de maturité de Moniqui sont clairement séparés, mais sont tout de même proches. Il semble donc que le facteur variété soit prépondérant par rapport au facteur stade de maturité dans cette étude. Les jeux de données correspondant à la variété Blanc sont plus proches de Bergeron que de ceux de Moniqui, ce qui est en accord avec l'origine phylogénique de Blanc.

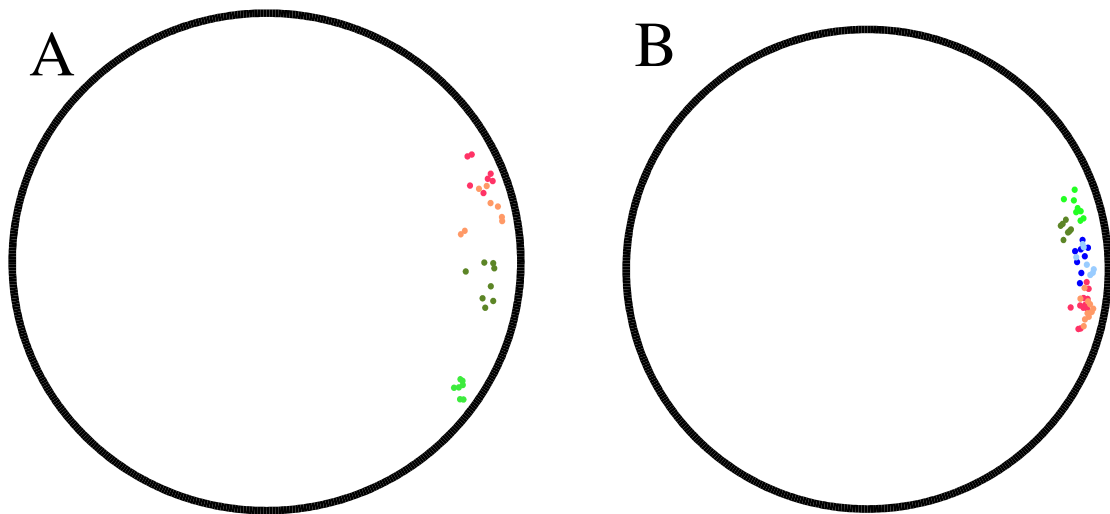


Figure 30 : Représentation par ACP des résultats d'hybridation des microarray selon les deux premières dimensions. Chaque point représente un jeu de données d'intensité d'hybridation du set d'oligonucléotides pour une lame, un fluorophore, un échantillon biologique et un réplicat. **A** Dans l'étude du développement les couleurs correspondent à: vert clair : Bergeron 03/06 ; vert foncé : Bergeron 01/07 ; orange : Bergeron 15/07 ; rouge: Bergeron 29/07. L'axe X explique 70 % de la variance et l'axe Y 11 % **B** Dans la comparaison de variété, les couleurs correspondent à: orange : Bergeron 15/07 ; rouge : Bergeron 29/07 ; vert clair : Moniqui 24/06 ; vert foncé : Moniqui 08/07 ; bleu clair : Blanc 08/07 ; bleu foncé : Blanc 29/07. L'axe X explique 80 % de la variance et l'axe Y 4 %.

IV.3.2.2.2 ANOVA globale

L'analyse statistique des résultats par des méthodes comme l'analyse de variance (ANOVA) permet de déterminer la probabilité de la régulation de chaque gène. Cette analyse permet aussi de mesurer l'effet des conditions expérimentales sur la valeur de l'intensité attribuée à chaque spot. Dans notre cas (Tab. 19), l'influence (i) de la nature du fluorophore utilisé, Cy3 ou Cy5 (effet sonde), (ii) de la position du spot sur la lame, grâce au dépôt répliqué des spots sur la lame à des positions éloignées (effet réplicat), (iii) de la lame utilisée pour une même condition et des autres facteurs liés aux conditions dans lesquelles le protocole a été réalisé (effet lame), (iiii) du stade de développement quand ce n'est pas le facteur de variabilité principal étudié (effet stade).

Dans l'expérience de suivi de développement le facteur principal qui influence le niveau d'expression des gènes est le stade de développement (effet stade, Tab. 19A). L'effet des autres facteurs est beaucoup plus faible, l'effet sonde n'est pas statistiquement significatif. Dans l'expérience de comparaison de variétés, il y a deux effets significatifs sur les niveaux d'intensité d'hybridation, l'effet du stade de développement et l'effet de la variété utilisée (Tab. 19B).

Tableau 19 : Analyse par ANOVA globale des interactions entre deux facteurs impliqués dans la variance de l'intensité des spots **A** au cours du développement de l'abricot **B** entre les variétés Bergeron, Blanc et Monique

A	Facteur	Somme des carrés	DDL	Variance	F	P-value
	gene	34899,81	1565	22,3	140,8	0,00
	gene sonde	181,21	1565	0,12	0,73	-
	gene replicat	541,45	1565	0,35	2,18	0,00
	gene lame	392,86	1565	0,25	1,58	0,00
	gene stade	7866,11	4695	1,68	10,58	0,00
	Résiduel	6198,56	39156	0,16	-	-
	Total	50080	50111	1	-	-

B	Facteur	Somme des carrés	DDL	Variance	F	P-value
	gene	69361,67	1548	44,81	286,26	0
	gene sonde	191,29	1548	0,12	0,79	-
	gene replicat	190,92	1548	0,12	0,79	-
	gene lame	275,27	3096	0,09	0,57	-
	gene stade	1310,24	1548	0,85	5,41	0
	gene variete	3723,6	3096	1,2	7,68	0
	Résiduel	11635,01	74359	0,16	-	-
	Total	86688	86743	1	-	-

IV.3.2.2.3 ANOVA locale

L'ANOVA locale est réalisée pour mettre en évidence la part du facteur étudié (stade ou variété) dans la variance totale de l'expression d'un seul gène. Le but final de cette analyse est d'identifier les gènes qui sont significativement différenciellement exprimés dans au moins deux conditions. La figure 31 présente la répartition des gènes en fonction de la significativité de la variation liée à la condition étudiée, ainsi que la part de cette variation dans la variation totale pour l'expérience de suivi de développement (Fig. 31A) et pour l'expérience de différenciation variétale (Fig. 31B).

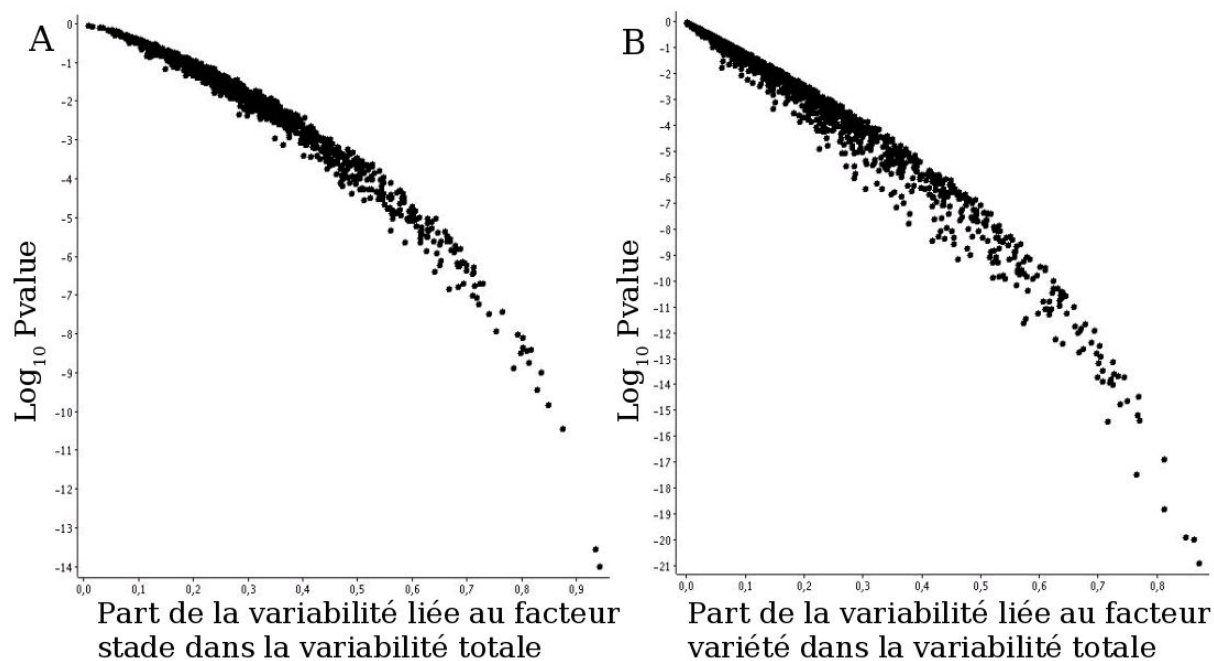


Figure 31 : Représentation graphique de l'ANOVA locale en fonction de la P-value pour **A** l'expérience de suivi de développement et **B** l'expérience de différenciation variétale. Chaque point correspond à un oligonucléotide. L'abscisse correspond à la valeur de la somme des carrés associée à la variation liée au facteur stade de maturité **A** ou au facteur variété **B** divisé par la somme des carrés de la variation totale. L'ordonnée correspond au Log_{10} de la P-Value.

Les gènes considérés comme différentiellement exprimés sont ceux dont la P value est inférieure à 10^{-3} . L'axe des abscisses permet de déterminer la part du facteur d'intérêt (stade de maturité ou variété) dans la variation totale de l'intensité d'hybridation des oligonucléotides. Plus cette valeur est proche de 1, plus sa part est importante. Les valeurs des abscisses des points de l'expérience de suivi de développement sont plus élevés que ceux de l'expérience de différenciation variétale. En effet, le facteur stade de maturité dans la seconde expérience n'est pas négligeable.

Dans l'expérience d'étude de développement des fruits de Bergeron, 553 gènes sont significativement différentiellement exprimés dans au moins deux stades et 425 d'entre eux ont une variance liée au facteur stade, supérieure à 0,1. Cette valeur a été choisie arbitrairement pour correspondre à des gènes régulés d'au moins un facteur 2. Dans l'expérience de l'étude des variétés, 532 gènes sont significativement différentiellement exprimés dans au moins 2 variétés, 345 d'entre eux ont une variance liée au facteur variété supérieure à 0,1.

Les gènes identifiés comme différentiellement exprimés grâce au crible statistique ont été utilisés pour les analyses suivantes. Les valeurs d'intensité, après normalisation, ont été reprises et moyennées pour obtenir une seule valeur d'expression pour chaque oligonucléotide

dans chaque condition. Pour réaliser le clustering des profils d'expression des gènes, il est nécessaire d'exprimer ces valeurs en intensité relative par rapport à une référence, afin de s'affranchir de l'intensité totale des oligonucléotides.

Dans les représentations graphiques des figures 33 à 56, l'intensité de la couleur verte représente le niveau de sous-expression d'un gène par rapport à la moyenne sur une échelle logarithmique de -1,5 (2,8 fois moins exprimé) à 0. Si la sous-expression est plus forte, la couleur verte est saturée. L'intensité de la couleur rouge représente le niveau de sur-expression d'un gène par rapport à la moyenne sur une échelle logarithmique de 1,5 (2,8 fois plus exprimé) à 0. Si la sur-expression est plus forte, la couleur rouge est saturée.

IV.3.3 Évolution des niveaux d'expression des transcrits par microarray

Dans notre démarche pour mettre en évidence les gènes potentiellement impliqués dans les critères de qualité, la suite logique à la production de masse des séquences est la mesure du niveau d'expression des gènes pendant des études comparatives de conditions où les critères de qualité vont varier. Si on respecte notre hypothèse de départ, les gènes qui sont les facteurs limitants d'un critère de qualité verront leur niveau d'expression varier conjointement au critère de qualité. Cependant, pour des raisons uniquement financières, nous avons été obligés de ne pas utiliser une approche complètement systématique, puisqu'un criblage, qui reste tout de même très exhaustif, a été effectué pour le choix des oligonucléotides déposés sur les lames.

Nous avons donc fait varier les critères de qualité au cours de deux expériences. La première consiste à évaluer l'expression des gènes pendant le développement du fruit. La seconde consiste à faire varier certains critères de qualité (l'intensité du dégagement d'éthylène et la couleur des fruits) sur du matériel mûr en utilisant des variétés de phénotype contrasté pour ces critères.

IV.3.3.1 Gènes régulés au cours du développement

IV.3.3.1.1 Catégorie fonctionnelle des gènes différenciellement exprimés pendant le développement

La figure 62 représente la répartition en catégories fonctionnelles des 425 gènes régulés au cours de la maturation. Chaque gène pouvant être représenté dans plusieurs catégories

fonctionnelles, elles sont toutes représentées, et pas exclusivement les catégories qui ont été sélectionnées pour être déposées sur les lames (chapitre IV.3.2.1). Les gènes de la catégorie métabolisme secondaire sont inclus dans la catégorie métabolisme.

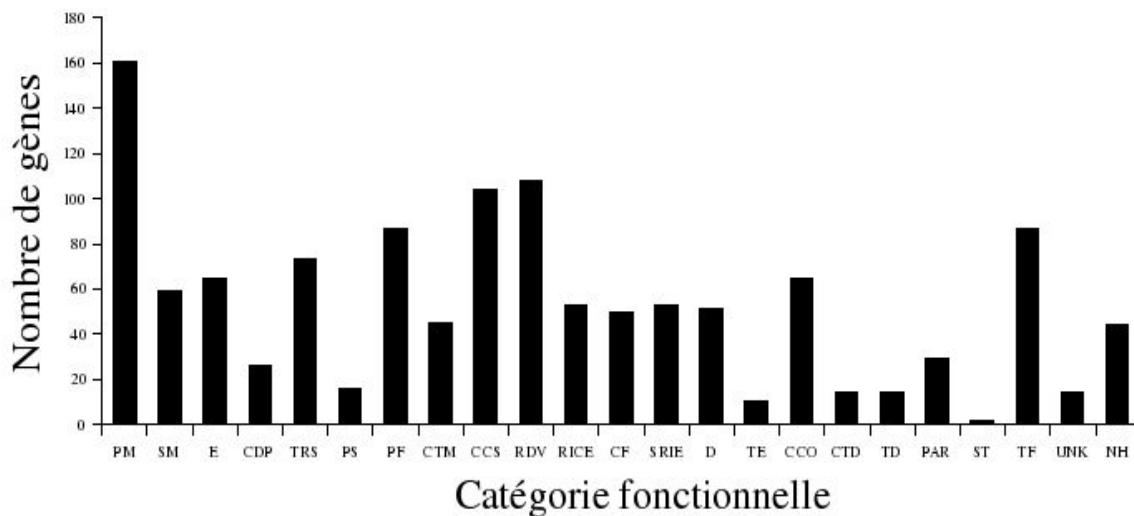


Figure 32 : Distribution en catégories fonctionnelles des gènes régulés pendant le développement. Les catégories sont définies selon le catalogue fonctionnel du Munich Information Center for Protein Sequence (<http://mips.gsf.de>). PM: Métabolisme ; SM Métabolisme secondaire ; E: Énergie ; CDP: Cycle Cellulaire et traitement de l'ADN ; TRS: Transcription ; PS: Synthèse des Protéines ; PF: Destinée des Protéines ; CTM: Transports Cellulaires et Mécanismes de Transport ; CCS: Communication Cellulaire et Transduction du Signal ; RDV: Protection, Défense et Virulence ; RICE: Régulation / Interaction avec l'Environnement des Cellules ; CF: Destin des Cellules ; SRIE: Régulation Systémique / Interaction avec l'Environnement ; D: Développement ; TE: Éléments Transposables ; CCO: Contrôle de l'Organisation cellulaire ; CTD: différenciation des cellules ; TD: différenciation des Tissus ; PAR: Régulation de l'Activité des Protéines ; SP: Protéines de Stockage ; TF: Transport facilité ; UNK: Protéines inconnues ; NH: Protéines non identifiées.

La présélection des gènes déposés sur les lames en fonction de leur fonction putative interdit de comparer les données de la figure 62 avec la répartition générale des EST. Ceci explique l'augmentation de la représentation des gènes impliqués dans les catégories fonctionnelles sélectionnées : Métabolisme Secondaire qui représente 8 % de plus sur l'ensemble des gènes (multiplié par 2,4) ; Transcription représente 5,8 % de plus (X 1,5) ; Transduction du Signal représente 11,4 % de plus (X 2) ; Protection, Défense et Virulence représente 6,3 % de plus (X 1,3) ; Métabolisme de la Paroi Cellulaire (inclus dans Organisation Cellulaire) représente 5 % de plus (X 2,8) ; les protéines de Transport représentent 10 % de plus (X 2). On ne constate cependant pas que la proportion de gènes régulés soit différente d'une catégorie à une autre : entre 20 et 28 % des gènes sont régulés par catégorie, alors que 24 % des gènes sont régulés sur l'ensemble de la lame.

IV.3.3.1.2 Profils d'expression globaux

La méthode de clustering utilisée pour regrouper les gènes est le « Self-Organising Tree Algorithm » (SOTA, Herrero *et al.*, 2001). Cette méthode est basée sur les réseaux de neurones et la classification hiérarchique. D'autres méthodes ont été utilisées, mais celle-ci :

- s'est avérée stable (les résultats sont identiques lorsqu'on réitère la même opération) ;
- permet, de plus, d'obtenir une évaluation statistique de la distance entre les gènes d'un cluster et son centroïde ou valeur de la diversité du cluster. Une valeur inférieure à 0,82 limite à 5 % le risque de regrouper abusivement deux profils différents dans le même cluster ;

- permet également d'estimer la distance entre deux clusters (annexe B (CD)) ;

Dans cette expérience, un nombre de 6 clusters a permis d'individualiser des groupes de profils présentant des valeurs de diversité de cluster acceptables, c'est à dire largement en dessous de 0,82 (Fig. 33).

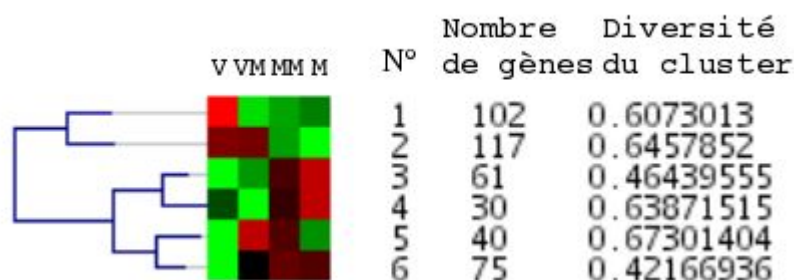


Figure 33 : Dendrogramme SOTA des clusters représentant les gènes différentiellement exprimés au cours du développement dans la variété Bergeron. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. Sur ce dendrogramme, sont présentés également le nombre de gènes inclus dans chaque cluster et les valeurs de diversité de cluster.

Les clusters 1 (102 gènes) et 2 (117 gènes) correspondent aux gènes dont l'expression diminue au cours du développement. Les gènes sont réprimés à partir du stade vert mature dans le cluster 1. Dans le cluster 2, les gènes sont fortement réprimés aux alentours de la crise climactérique (entre les stades vert immature et mi-mûr), même si leur expression peut encore diminuer par la suite.

Les clusters 3 (61 gènes) et 4 (30 gènes) regroupent des gènes induits pendant le développement. Cette induction est linéaire pour le cluster 3, tandis qu'on observe pour le cluster 4 une forte diminution de l'expression dans le stade vert mature précédant la crise climactérique. Les gènes dont l'expression est régulée par l'éthylène devraient se retrouver dans ces deux clusters.

Les clusters 5 (40 gènes) et 6 (75 gènes) représentent des gènes induits à partir du stade vert mature. Dans le cluster 5, les gènes sont fortement induits à ce stade, puis ont tendance à être réprimés dans les stades plus mûrs. Dans le cluster 6, le niveau d'expression reste constant après avoir été induit au stade vert mature.

IV.3.3.1.3 Cluster 1

Les gènes appartenant à ce cluster sont réprimés à partir du stade vert mature, certains restent réprimés, d'autres sont ensuite induits, mais ne retrouvent jamais leur niveau d'expression du stade vert (Fig. 34).

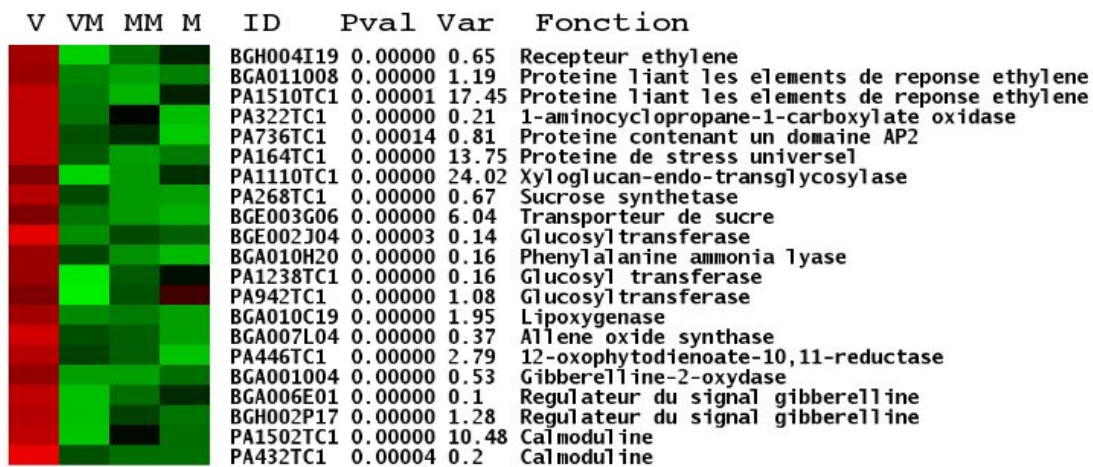


Figure 34 : Profil d'expression des gènes appartenant au cluster 1. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Six unigènes impliqués dans la réponse à l'éthylène : un récepteur à éthylène, deux facteurs de liaison aux éléments de réponse à l'éthylène, une ACC oxydase, une protéine contenant un domaine AP2 et une protéine de stress universel ;
- Un unigène impliqué dans la perte de fermeté, une XET ;
- Deux unigènes impliqués potentiellement dans le métabolisme des sucres, une saccharose synthase et un transporteur de sucre ;
- Quatre unigènes impliqués dans la biosynthèse des flavonoïdes, trois glucosyltransférases et une phényl ammonia lyase ;

- Trois unigènes impliqués dans la biosynthèse du jasmonate, une lipoxygénase, une allène oxyde synthase et une oxophytodiénoate réductase ;

- Trois unigènes impliqués dans le métabolisme (une gibbérelline-2-oxydase) et la régulation (deux gènes régulateurs du signal GA contenant un motif DELLA) de la gibbérelline.

IV.3.3.1.4 Cluster 2

Les gènes appartenant à ce cluster présentent un niveau d'expression relativement constant aux stades vert et vert mature. L'expression diminue au stade mi-mûr et pour la plupart d'entre eux, diminue encore au stade mûr (Fig. 35).

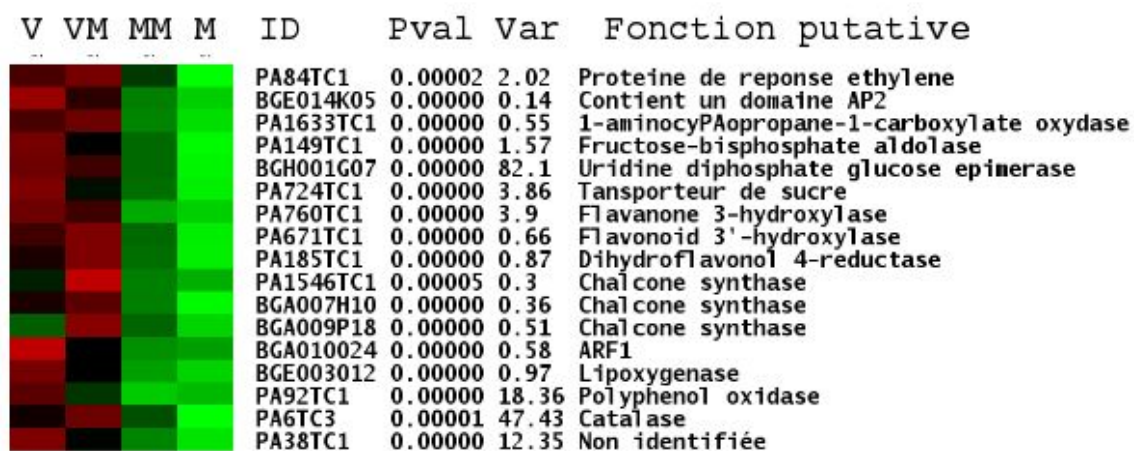


Figure 35 : Profil d'expression des gènes appartenant au cluster 2. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Trois unigènes impliqués dans le métabolisme (ACC oxydase) et la réponse à l'éthylène (protéine contenant un domaine AP2 et une ERF) ;

- Un unigène impliqué dans la photosynthèse, fructose bisphosphate aldolase chloroplastique ;

- Un unigène impliqué dans la formation de la paroi cellulaire, l'UDP glucose épimérase ;

- Un unigène impliqué dans le transport des sucres ;

- Six unigènes impliqués dans le métabolisme des flavonoïdes, deux flavonoïdes 3 hydroxylases, une dihydroflavonol 4-réductase et trois chalcones synthases ;
- Un facteur de réponse à l'auxine ;
- Un unigène impliqué dans le métabolisme du jasmonate, une lipoxygénase ;
- Une catalase, impliquée dans la détoxification ;
- Un unigène non identifié dont l'expression est confirmée par le Northern électronique (Chapitre IV.2.3.2.7).

IV.3.3.1.5 Cluster 3

Les gènes appartenant à ce cluster présentent un niveau d'expression qui augmente linéairement pendant les stades vert, vert mature et mi-mûr. Pour quelques uns l'expression continue d'augmenter au stade mûr (Fig. 36).

V	VM	MM	M	ID	Pval	Var	Fonction
Green	Green	Green	Red	PA835TC1	0.00044	0.37	Contient un domaine AP2
Green	Green	Green	Red	PA34TC1	0.00002	2.83	5-méthyltetrahydropteroyltriglutamate--homocysteine S-méthyltransferase
Green	Green	Green	Red	BGE003L14	0.00000	0.27	Calcium-dependent protein kinase
Green	Green	Green	Red	PA1496TC1	0.00012	0.15	Geranylgeranyl pyrophosphate synthase
Green	Green	Green	Red	PA229TC1	0.00001	0.92	Expansine
Green	Green	Green	Red	PA349TC1	0.00000	0.83	Sucrose-UDP glucosyltransferase
Green	Green	Green	Red	PA310TC1	0.00000	7.68	Sous unité β de canal potassique
Green	Green	Green	Red	BGH009N22	0.00000	0.72	Serine/threonine protéine kinase
Green	Green	Green	Red	PA1427TC1	0.00000	0.17	Lipoxygénase
Green	Green	Green	Red	PA19TC3	0.00000	0.82	Protéine inconnue
Green	Green	Green	Red	PA12TC1	0.00000	72.96	Protéine de choc thermique
Green	Green	Green	Red	PA1442TC1	0.00000	0.35	Reductase
Green	Green	Green	Red	PA104TC1	0.00000	25.7	Pas d'homologue

Figure 36 : Profil d'expression des gènes appartenant au cluster 3. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la réponse à l'éthylène, l'un contenant un motif AP2 et l'autre correspondant à une protéine kinase calcium dépendante ;
- Un unigène impliqué en amont de la biosynthèse de l'éthylène : 5-méthyl tetrahydropteroyl triglutamate homocystéine S-méthyltransférase ;
- Un unigène impliqué dans la biosynthèse des caroténoïdes, la géranyl-géranyl pyrophosphate synthase ;
- Un unigène impliqué dans la perte de fermeté, une expansine ;
- Un unigène impliqué dans la biosynthèse du saccharose, la sucrose synthase ;
- Un unigène impliqué dans le transport des ions, une sous unité β de canal potassique ;

- Un unigène correspondant à une sérine/thréonine kinase régulée par l'ABA ;
- Un unigène impliqué dans le métabolisme du jasmonate, lipoxygénase ;
- Un unigène correspondant à une protéine inconnue dont l'expression est confirmée par le Northern électronique (chapitre IV.2.3.2.7) et dont le profil protéique (chapitre IV.5.4.7) a été observé ;
- Une protéine de choc thermique dont l'expression est confirmée par le Northern électronique (chapitre IV.2.3.2.7).

IV.3.3.1.6 Cluster 4

Les gènes appartenant à ce cluster présentent un niveau d'expression qui diminue au stade vert mature et qui augmente à la crise éthylénique (Fig. 37).

V	VM	MM	M	ID	Pval	Var	Fonction
				PA9TC1	0.00004	101.72	1-aminocyclopropane-1-carboxylate oxidase
				PA60TC1	0.00001	9.6	S-adenosylmethionine synthetase
				BGE016B06	0.00000	0.15	Pectinesterase
				BGH008N11	0.00000	0.26	Transporteur de sulfate
				PA1592TC1	0.00001	8.64	Transporteur mitochondrial de dicarboxylate

Figure 37 : Profil d'expression des gènes appartenant au cluster 4. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la biosynthèse de l'éthylène, une ACC oxydase et une S-adenosylméthionine synthase ;
- Un unigène impliqué dans la perte de fermeté, pectinestérase ;
- Deux unigènes impliqués dans le transport des ions, un transporteur de sulfate et un transporteur de dicarboxylate.

IV.3.3.1.7 Cluster 5

Les gènes appartenant à ce cluster présentent un niveau d'expression fort aux stades vert mature et mi-mûr. Leur niveau d'expression diminue en fin de maturation (Fig. 38).

V	VM	MM	M	ID	Pval	Var	Fonction
■	■	■	■	PA150TC1	0.00000	5.42	Endo-beta-1,4-glucanase
■	■	■	■	BGH007L07	0.00000	0.62	Inhibiteur de pectin méthylesterase
■	■	■	■	PA341TC1	0.00012	3.01	Transporteur de dicarboxylate/tricarboxylate
■	■	■	■	PA1452TC1	0.00032	2.78	O-méthyltransferase
■	■	■	■	BGE003M12	0.00000	5.81	S-adenosyl-L-méthionine: acide salicylique/benzoïque carboxyl méthyltransferase
■	■	■	■	PA1355TC1	0.00028	2.27	6-phosphogluconolactonase

Figure 38 : Profil d'expression des gènes appartenant au cluster 5. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la perte de fermeté : une endo- β -1,4 glucanase et un inhibiteur de pectinestérase ;
- Un unigène impliqué dans le transport des acides organiques au niveau de la mitochondrie : le transporteur de dicarboxylate/tricarboxylate ;
- Un unigène impliqué dans le métabolisme des flavonoïdes, une O-méthyltransférase ;
- Un unigène impliqué dans le métabolisme des arômes, la S-adénosyl-L-méthionine : acide salicylique/benzoïque carboxyl méthyltransférase
- Un unigène impliqué dans la détoxification, la 6-phosphogluconate déshydrogénase.

IV.3.3.1.8 Cluster 6

Les gènes appartenant à ce cluster présentent un niveau d'expression qui augmente entre les stades vert et vert mature. L'expression est stable pendant le reste de la maturation (Fig. 39).



Figure 39 : Profil d'expression des gènes appartenant au cluster 6. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et stade. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes potentiellement régulés par l'éthylène, une protéine de stress universel et une protéine abondante à la fin de l'embryogénèse (LEA) ;
- Deux unigènes impliqués dans la biosynthèse des caroténoïdes (la phytoène synthase et la zéta carotène désaturase) et un impliqué dans la dégradation, la 9-cis époxy-caroténoïde dioxygénase ;
- Cinq unigènes impliqués dans la perte de fermeté, 4 inhibiteurs de pectinestérase et une expansine, ainsi qu'un unigène impliqué dans le grandissement cellulaire ;
- Deux unigènes impliqués dans le métabolisme des flavonoïdes, une O-méthyltransférase et une anthocyanidin-3-glucoside rhamnosyltransférase ;
- Un unigène impliqué dans la biosynthèse du malate, la malate oxydoréductase
- Deux unigènes impliqués dans le métabolisme des sucres, un transporteur de sucre et la sorbitol déshydrogénase

- Trois unigènes impliqués dans le métabolisme de l'amidon, une ADP-glucose pyrophosphorylase et deux amidon synthases ;
- Une serine/thréonine régulée par l'ABA ;
- Six unigènes impliqués dans la réponse à l'auxine, 5 AUX/AIA et un facteur de réponse à l'auxine ;
- Un unigène correspondant à une protéine inconnue dont le profil protéique a été observé (chapitre IV.5.4.7) ;
- Une protéine riche en proline dont l'expression est confirmée par le Northern électronique (chapitre IV.2.3.2.7).

IV.3.3.2 Gènes différentiellement exprimés entre les variétés.

345 gènes sont régulés significativement différemment entre au moins deux des variétés, indépendamment des stades de développement. Tout comme pour l'expérience de suivi de l'expression des gènes au cours du développement, l'algorithme utilisé pour regrouper les gènes par profil d'expression est l'algorithme SOTA. Dans cette expérience, il a fallu introduire un nombre de 16 clusters pour pouvoir individualiser des groupes de profils présentant des valeurs de diversité de cluster acceptables (inférieure à 0,82, Fig. 40). Cependant cette valeur, pour certains d'entre eux, est encore supérieure à 0,82 (clusters 2, 4 et 5).

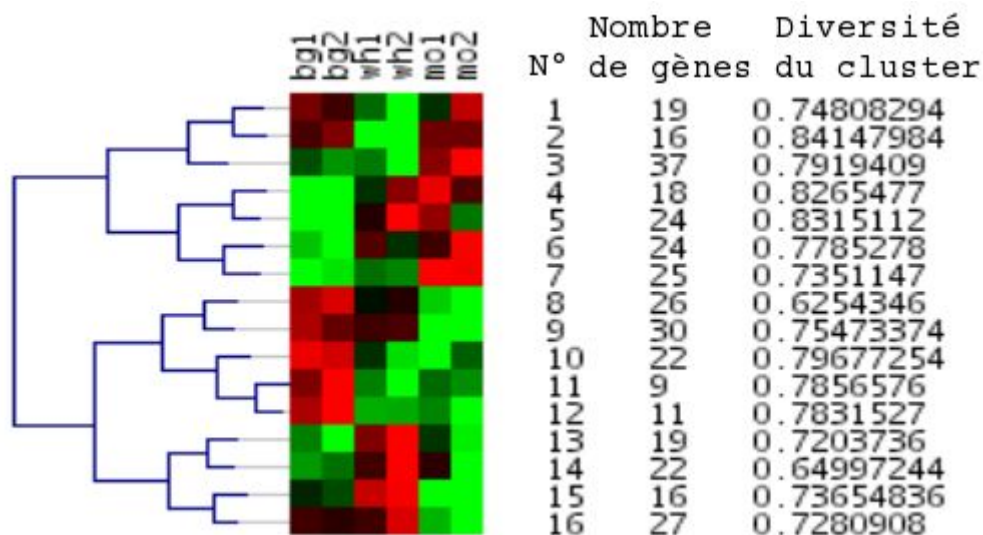


Figure 40 : Dendrogramme SOTA des clusters représentant les gènes différentiellement exprimés dans les variétés Bergeron (bg), Blanc (wh) et Monique (mo). Pour chaque variété, le stade 1 correspond à la crise éthylénique et le stade 2 au maximum d'accumulation de saccharose. Sur ce dendrogramme, sont présentés également le nombre de gènes inclus dans chaque cluster et les valeurs de diversité de cluster.

Les gènes qui sont réprimés dans des variétés différentes de celle qui a servi à la construction des lames (Bergeron) le sont sous réserve d'une hybridation moindre causée par une différence de séquence au niveau de l'oligonucléotide.

Les clusters 1 (19 gènes) et 2 (16 gènes) regroupent des gènes dont l'expression est inhibée spécifiquement dans la variété Blanc. Dans le cluster 1, les gènes sont inhibés principalement dans le stade 2, et dans le cluster 2, les gènes sont inhibés dans les deux stades. Les gènes retrouvés dans le cluster 3 sont aussi réprimés dans la variété Blanc, il le sont aussi légèrement chez Bergeron.

Les clusters 4 (18 gènes), 5 (24 gènes), 6 (24 gènes) et 7 (25 gènes) regroupent des gènes dont l'expression est inhibée dans la variété Bergeron. Dans les clusters 4 et 5, les gènes présentent une sur-expression chez Blanc au stade 2 et chez Moniqui au stade 1. Dans le cluster 6, les gènes sont sur-exprimés chez Moniqui au stade 2. Dans le cluster 7 les gènes sont légèrement sous-exprimés chez Blanc et sur-exprimés chez Moniqui.

Les clusters 8 (26 gènes), 9 (30 gènes), 10 (22 gènes), 11 (9 gènes) et 12 (11 gènes) regroupent des gènes dont l'expression est induite dans la variété Bergeron. Dans les clusters 8 et 9, les gènes sont réprimés chez Moniqui et dans le cluster 9 l'expression est équivalente au stade 2 chez Bergeron et Blanc. Dans le cluster 10, les gènes sont particulièrement réprimés chez Blanc au stade 2 et chez Moniqui au stade 1. Dans les clusters 11 et 12, les gènes sont réprimés dans Blanc et Moniqui, dans le cluster 11 la répression est plus forte chez Blanc au stade 2 et dans le cluster 12, la répression est plus forte chez Moniqui au stade 2.

Les clusters 13 (19 gènes), 14 (22 gènes), 15 (16 gènes) 16 (27 gènes) regroupent des gènes induits dans la variété Blanc, particulièrement au stade 2 et réprimés chez Moniqui, particulièrement au stade 2. Dans les clusters 13 et 14, les gènes sont réprimés chez Bergeron, surtout au stade 2 dans le cluster 13. Dans le cluster 15, la sur-expression chez Blanc et la répression chez Moniqui est visible aux deux stades. Dans le cluster 16, les gènes sont réprimés chez Moniqui à tous les stades.

IV.3.3.2.1 Cluster1

Les gènes de ce cluster sont réprimés dans la variété Blanc, principalement au stade 2 et très fortement exprimés au stade 2 chez Moniqui (Fig. 41).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
Red	Red	Green	Green	Red	Red	BGH004I19	0.00016	0.14	Recepteur a ethylene
Red	Red	Green	Green	Red	Red	BGE005K04	0.00000	1	Transporteur de dicarboxylate/tricarboxylate
Red	Red	Green	Green	Red	Red	PA1592TC1	0.00000	11.73	Transporteur de dicarboxylate/tricarboxylate
Red	Red	Green	Green	Red	Red	BGH008N11	0.00068	0.42	Transporteur de sulfate
Red	Red	Green	Green	Red	Red	PA1442TC1	0.00070	0.65	Alcool deshydrogenase
Red	Red	Green	Green	Red	Red	PA269TC1	0.00043	8.36	Serine/threonine-proteine kinase
Red	Red	Green	Green	Red	Red	PA354TC1	0.00001	0.7	Quinone oxidoreductase
Red	Red	Green	Green	Red	Red	PA19TC3	0.00003	1.75	Proteine inconnue

Figure 41 : Profil d'expression des gènes appartenant au cluster 1. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène correspondant à un récepteur d'éthylène ;
- Trois unigènes impliqués dans le transport des ions, deux transporteurs d'acides organiques et un transporteur de sulfate ;
- Un unigène régulé par l'auxine, une quinone oxydoréductase ;
- Deux unigènes régulés par l'ABA, une protéine phosphatase et une protéine kinase ;
- Un unigène impliqué dans le métabolisme des arômes, une alcool déshydrogénase ;
- Un unigène correspondant à une protéine inconnue dont l'induction pendant le développement, observé par microarray, est confirmée par le Northern électronique (chapitre IV.2.3.2.7) et dont le profil protéique (chapitre IV.5.4.7) a été observé.

IV.3.3.2.2 Cluster 2

Les gènes de ce cluster sont réprimés dans tous les stades de la variété Blanc (Fig. 42).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
Red	Red	Green	Green	Red	Red	PA1633TC1	0.00002	0.27	1-aminocyclopropane-1-carboxylate oxidase
Red	Red	Green	Green	Red	Red	PA1110TC1	0.00002	0.76	Xyloglucan endo transglycosylase
Red	Red	Green	Green	Red	Red	PA104TC1	0.00012	22.3	Non identifie

Figure 42 : Profil d'expression des gènes appartenant au cluster 2. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène homologue aux dioxygénases de la famille des ACC oxydase ;
- Un unigène impliqué dans la perte de fermeté, la xyloglucan endo transglycosylase ;

- Un unigène non identifié rosacée-spécifique dont l'expression augmente pendant la maturation (chapitre IV.3.3.1.5).

IV.3.3.2.3 Cluster 3

Le cluster 3 contient des gènes dont l'expression est plus forte chez Monique et réprimée fortement dans la variété Blanc au stade 2 (Fig. 43).

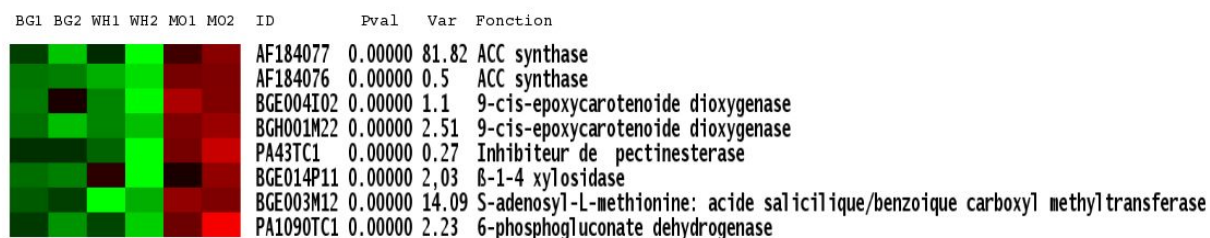


Figure 43 : Profil d'expression des gènes appartenant au cluster 3. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la biosynthèse de l'éthylène, deux isogènes d'ACC synthase ;
- Deux unigènes impliqués dans la biosynthèse de l'ABA, deux isogènes de NCED ;
- Deux unigènes impliqués dans la perte de fermeté, un inhibiteur de pectinestérase et une β -1-4-xylosidase ;
- Un unigène impliqué dans le métabolisme des arômes, la S-adenosyl-L-méthionine : acide salicylique/benzoïque carboxyl méthyltransférase ;
- Un unigène impliqué dans la détoxification dont le profil d'expression protéique est connu, la 6-phosphogluconate déshydrogénase.

IV.3.3.2.4 Cluster 4

Les gènes appartenant à ce cluster sont réprimés chez Bergeron (Fig. 44).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						BGE003G06	0.00000	5.42	Transporteur de sucre
						PA549TC1	0.00001	1.09	ADP-glucose pyrophosphorylase

Figure 44 : Profil d'expression des gènes appartenant au cluster 4. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans le transport du sorbitol ;
- Un unigène impliqué dans la biosynthèse de l'amidon, l'ADP-glucose pyrophosphorylase.

IV.3.3.2.5 Cluster 5

Les gènes appartenant à ce cluster sont réprimés chez Bergeron (Fig. 45).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						PA150TC1	0.00001	8.84	Endo-beta-1,4-glucanase
						PA23TC2	0.00055	38.26	Inhibiteur de pectinesterase
						BGA009P06	0.00002	0.14	AUX/IAA
						BGH011D24	0.00000	2.82	Transporteur de sodium-dicarboxylate

Figure 45 : Profil d'expression des gènes appartenant au cluster 5. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la perte de fermeté, une endo- β -1,4 glucanase et un inhibiteur de pectinestérase ;
- Un unigène impliqué dans le transport des ions, un transporteur de sodium-dicarboxylate ;
- Un unigène impliqué dans la réponse à l'auxine, un AUX/IAA.

IV.3.3.2.6 Cluster 6

Les gènes appartenant à ce cluster sont réprimés dans Bergeron, principalement au stade 2 et légèrement sur-exprimés chez Monique à ce même stade (Fig. 46).

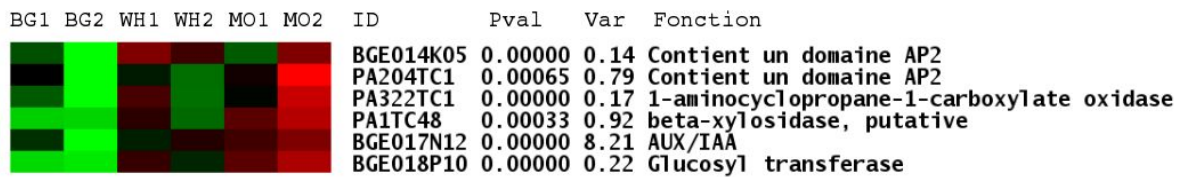


Figure 46 : Profil d'expression des gènes appartenant au cluster 6. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans la réponse à l'éthylène contenant un motif AP2 ;
- Un unigène homologue aux dioxygénases de la famille des ACC oxydases ;
- Un unigène impliqué dans la réponse à l'auxine, un AUX/IAA ;
- Un unigène correspondant à une glucosyltransférase.

IV.3.3.2.7 Cluster 7

Les gènes appartenant à ce cluster sont réprimés chez Bergeron, principalement au stade 1, il sont clairement sur-exprimés chez Monique (Fig. 47).

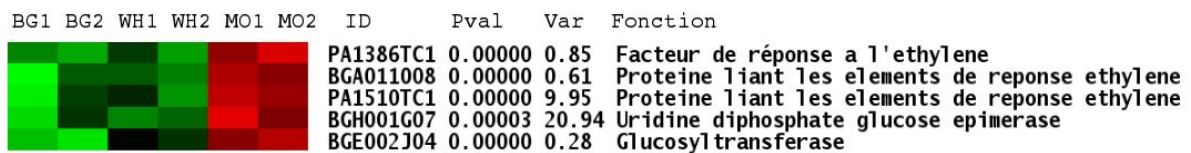


Figure 47 : Profil d'expression des gènes appartenant au cluster 7. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Trois gènes impliqués dans la réponse à l'éthylène, un facteur de réponse ERF et deux protéines liant les éléments de réponse à l'éthylène (EREBP) ;
- Un unigène impliqué dans la biosynthèse de la paroi cellulaire, l'UDP-glucose épimérase ;
- Un unigène correspondant à une glucosyltransférase.

IV.3.3.2.8 Cluster 8

Les gènes appartenant à ce cluster sont réprimés chez Moniqui au stade 2 (Fig. 48).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						PA703TC1	0.00000	0.4	1-aminocyclopropane-1-carboxylate deaminase
						PA190TC1	0.00007	0.12	Endo β 1,4 glucanases
						BGH009N22	0.00000	0.57	Serine/threonine protein kinase
						PA446TC1	0.00034	0.18	12-oxophytodienoate-10,11-reductase

Figure 48 : Profil d'expression des gènes appartenant au cluster 8. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène rentrant en compétition avec l'ACC oxydase pour l'ACC, l'ACC désaminase ;
- Un unigène impliqué dans la perte de fermeté, une endo- β -1,4 glucanase ;
- Un unigène régulé par l'ABA, une protéine kinase ;
- Un unigène impliqué dans la biosynthèse du jasmonate, une oxophytodiénoate réductase.

IV.3.3.2.9 Cluster 9

Les gènes appartenant à ce cluster sont réprimés chez Moniqui (Fig. 49).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						AF000952	0.00005	0.11	Transporteur de sucre
						PA1452TC1	0.00023	2.4	O-méthyltransférase
						PA1427TC1	0.00002	0.27	Lipoxygénase

Figure 49 : Profil d'expression des gènes appartenant au cluster 9. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans le transport des sucres ;
- Un unigène impliqué dans le métabolisme des flavonoïdes, une O-méthyltransférase ;
- Un unigène impliqué dans le métabolisme du jasmonate, lipoxygénase.

IV.3.3.2.10 Cluster 10

Les gènes appartenant à ce cluster sont réprimés chez Moniquei, principalement au stade 1 et plutôt sur-exprimés chez Bergeron (Fig. 50).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						BGH008H16	0.00000	0.12	Contient un domaine AP2
						PA60TC1	0.00000	11.69	S-adenosylmethionine synthetase
						PA1498TC1	0.00004	2.58	Calcium dependante proteine kinase
						PA614TC1	0.00000	6.07	Acetyl-CoA:benzylalcoholo acetyltransferase
						PA46TC1	0.00000	0.15	Chalcone synthase

Figure 50 : Profil d'expression des gènes appartenant au cluster 10. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans la réponse à l'éthylène contenant un motif AP2 ;
- Un unigène impliqué dans la biosynthèse de l'éthylène, la SAM synthétase ;
- Un unigène impliqué dans la réponse à l'éthylène, une protéine kinase calcium dépendante ;
- Un unigène impliqué dans le métabolisme des arômes, l'acétyl-CoA:benzylalcool acétyltransférase ;
- Un unigène impliqué dans le métabolisme des flavonoïdes, la chalcone synthase.

IV.3.3.2.11 Cluster 11

Les gènes appartenant à ce cluster sont induits chez Bergeron (Fig. 51).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						BGA006007	0.00000	0.15	Chalcone isomerase
						PA4TC5	0.00000	19.65	Anthocyanidin-3-glucoside rhamnosyltransferase

Figure 51 : Profil d'expression des gènes appartenant au cluster 11. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Deux unigènes impliqués dans le métabolisme des flavonoïdes, une chalcone synthase et une anthocyanidin-3-glucoside rhamnosyltransférase.

IV.3.3.2.12 Cluster 12

Les gènes appartenant à ce cluster sont induits chez Bergeron (Fig. 52).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						PA84TC1	0.00000	1.94	Protéine de stress universelle
						PA373TC1	0.00000	10.97	AUX/IAA

Figure 52 : Profil d'expression des gènes appartenant au cluster 12. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans la réponse à l'éthylène, une protéine de stress universel ;
- Un unigène impliqué dans la réponse à l'auxine, un AUX/IAA ;

IV.3.3.2.13 Cluster 13

Les gènes appartenant à ce cluster sont sur-exprimés dans la variété Blanc (Fig. 53).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						BGH007L07	0.00000	10.24	Inhibiteur de pectinesterase
						PA760TC1	0.00000	2.9	Flavanone 3-hydroxylase
						PA147TC2	0.00000	5.36	O-methyl transferase
						PA697TC1	0.00000	0.63	Facteur de reponse a l'auxine
						PA389TC1	0.00001	2.51	Monodehydroascorbate reductase

Figure 53 : Profil d'expression des gènes appartenant au cluster 13. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans la perte de fermeté, un inhibiteur de pectinestérase ;
- Deux unigènes impliqués dans le métabolisme des flavonoïdes, une flavonoïde 3'-hydroxylase et une O-méthyltransférase ;
- Un unigène impliqué dans la réponse à l'auxine, ARF ;
- Un unigène impliqué dans la détoxification dont le profil d'expression protéique est connu (chapitre IV.5.4.7), une déshydroascorbate réductase.

IV.3.3.2.14 Cluster 14

Les gènes appartenant à ce cluster sont réprimés chez Moniqui au stade 2 et induits chez Blanc à ce même stade. Chez Bergeron ces gènes sont sous-exprimés (Fig. 54).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						PA149TC1	0.00017	0.3	Fructose-bisphosphate aldolase

Figure 54 : Profil d'expression des gènes appartenant au cluster 14. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans la photosynthèse, une fructose bisphosphate aldolase chloroplastique.

IV.3.3.2.15 Cluster 15

Les gènes appartenant à ce cluster sont sur-exprimés chez Blanc et sous-exprimés chez Moniqui (Fig. 55).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						BGE005M19	0.00000	11.93	Inhibiteur de pectinesterase
						PA118TC1	0.00000	6.22	Expansine
						PA2TC4	0.00004	17.57	Malate oxidoreductase
						PA1425TC1	0.00000	1.17	Alcool acyltransferase
						PA529TC1	0.00000	6	Quinone oxidoreductase

Figure 55 : Profil d'expression des gènes appartenant au cluster 4. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans la perte de fermeté, un inhibiteur de pectinestérase ;
- Un unigène impliqué dans le métabolisme du malate, la malate déshydrogénase ;
- Un unigène impliqué dans le métabolisme des arômes, l'alcool acyltransférase ;
- Un unigène régulé par l'auxine, une quinone oxydoréductase.

IV.3.3.2.16 Cluster 16

Les gènes appartenant à ce cluster sont sous-exprimés chez Moniquei (Fig. 56).

BG1	BG2	WH1	WH2	MO1	MO2	ID	Pval	Var	Fonction
						AF139500	0.00000	6.53	Proteine E8
						PA1496TC1	0.00001	0.2	Geranylgeranyl pyrophosphate synthase
						PA310TC1	0.00001	5.76	Sous unite β de canal potassique
						PA12TC1	0.00000	58.81	Proteine de choc thermique
						PA82TC1	0.00000	29.97	Proteine inconnue

Figure 56 : Profil d'expression des gènes appartenant au cluster 16. Le stade V est un stade vert, le stade VM précède de 15 jours la crise éthylénique, le stade MM correspond à la crise éthylénique et le stade M au maximum d'accumulation de saccharose. ID correspond au nom de l'unigène. Pval: P value calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Var: variance calculée d'après l'analyse par ANOVA locale entre les facteurs gène et variété. Fonction: fonction probable obtenue par blast.

Parmi les gènes potentiellement impliqués dans les critères de qualité de l'abricot, on retrouve :

- Un unigène impliqué dans le contrôle de la biosynthèse de l'éthylène, la protéine E8 ;
- Un unigène impliqué dans la biosynthèse des caroténoïdes, la géranylgeranyl pyrophosphate synthase ;
- Un unigène impliqué dans le transport des ions, une sous unité β de canal potassique ;
- Un unigène codant pour une protéine de choc thermique dont l'expression est confirmée par le Northern électronique (chapitre IV.2.3.2.7) ;
- Un unigène correspondant à une protéine inconnue dont le profil protéique a été observé (chapitre IV.5.4.7).

IV.4 Aide à l'identification des protéines

Cette partie a fait l'objet d'un article qui est accepté dans Proteomics et correspond à la traduction française de cet article.

Le sujet porte sur la mise en place d'une technique, utilisée dans le chapitre IV.5, permettant l'amélioration de l'identification des spots protéiques obtenus sur les gels d'électrophorèse 2D, ainsi que sa validation sur plusieurs espèces dont l'abricotier.

L'identification des empreintes massiques peptidiques obtenues sur les génomes non modèles est améliorée lorsque les mutations observées sur les traductions conceptuelles d'EST sont incluses dans les séquences des protéines orthologues

Jérôme Grimplet¹, José Walter Gaspar¹, Anne-Laure Gancel², François-Xavier Sauvage¹ et Charles Romieu¹.

¹Unité Mixte de Recherche des Sciences Pour l'Oenologie, INRA, 2 place Viala, 34060 Montpellier Cedex 1, France

²Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Département FLHOR, TA 50/16, F-34398 Montpellier cedex 5, France.

Titre courant : Identification des protéines sur des génomes non modèles

Correspondance : Dr. Charles Romieu, Unité Mixte de Recherche des Sciences Pour l'Oenologie, INRA, 2 place Viala, 34060 Montpellier Cedex 1, France.

E-mail: romieu@ensam.inra.fr

Mots-clés : EST / empreintes massique / annotation / génomes non modèles.

Abréviations : nrEST : EST non redondantes

IV.4.1 Résumé

Afin d'améliorer l'identification des protéines à partir des empreintes massiques de peptides pour des espèces dont le génome n'est pas complètement séquencé, les mutations spécifiques déduites des séquences d'EST ont été incluses dans les cadres de lecture des séquences orthologues. Une base de données de protéines virtuelles a été créée. Cette méthode a été utilisée pour quatre espèces végétales. Par rapport à la procédure habituelle, le score mowse a été amélioré pour 20 % des protéines au moins, chez les espèces pour lesquelles on dispose d'un nombre moyen d'EST. Augmenter le nombre d'EST permet de générer des

contigs plus longs, ce qui améliore l'identification des protéines les plus abondantes et limite l'utilité de notre approche. Cependant, les contigs des transcrits mineurs restent plus courts, et l'identification des protéines moins abondantes, telles que celles isolées de fraction subcellulaires, est améliorée lorsque on interroge une banque de protéines chimériques type orthologue/EST avec les empreintes massiques peptidiques.

IV.4.2 Introduction

Les approches systématiques de protéomique à haut débit suscitent un intérêt croissant du fait de leur coût relativement faible et de la possibilité de visualiser des centaines de protéines sur des gels bidimensionnels. Cependant le succès de l'identification des protéines à partir d'empreintes massiques peptidiques (PMF) repose sur l'information pré-existante au niveau des séquences. Cette identification est fiable pour les génomes complètement séquencés car on peut alors calculer l'intervalle de confiance rigoureusement (Lester et Hubbard, 2002 ; Mathesius *et al.*, 2002). Cependant, elle est beaucoup plus problématique pour les organismes dont le génome n'est que partiellement séquencé. Seules les protéines bien conservées peuvent être correctement identifiées à partir des séquences orthologues, et le taux de succès dépend donc de la représentativité des espèces proches dans les bases de données protéiques (Lester et Hubbard, 2002 ; Liska et Shevchenko, 2003).

Pour cette raison, on compare également, à titre complémentaire, les PMF expérimentaux aux PMF conceptuels issus des bases des données EST. Bien que les risques d'échecs d'identification des PMF soient plus élevés puisque les séquences sont plus courtes, l'identification est généralement plus efficace que la recherche « inter-espèces », comme cela a pu être démontré chez *Medicago truncatula* (Gallardo *et al.*, 2003 ; Watson *et al.*, 2003) ou le riz (Imin *et al.*, 2001). Moins de protéines sont identifiées pour les organismes chez lesquels un minimum d'EST a été séquencé, cependant une telle stratégie a permis de confirmer que des transcrits originaux et très redondants sont effectivement traduits en protéines majoritaires (Sarry *et al.*, 2004). De plus, on a ainsi pu démontrer la traduction effective de gènes récemment dupliqués, dont les rares différences pouvaient être considérées comme des erreurs de séquençage, alors que les comparaisons avec les séquences orthologues ne laissaient pas présager de l'existence de paralogues aussi proches.

Il est admis que la procédure de clustering d'EST (Huang et Madan, 1999) génère des assemblages erronés et les PMF peuvent être utilisés comme moyen de vérification de ces assemblages (Lisacek *et al.*, 2001). Dans tous les cas, la taille de la base de données d'EST influence de façon complexe la longueur des contigs et le nombre de transcrits uniques. Notre but était de proposer une nouvelle méthode d'identification combinant ces deux approches, en substituant les régions homologues des orthologues putatifs avec les traductions conceptuelles de jeux d'EST non redondants (nrEST) assemblées à partir d'EST. Nous démontrons ici que cette stratégie peut améliorer l'identification des protéines majoritaires chez les *Prunus* et les *Citrus* pour lesquels un nombre relativement moyen d'EST est actuellement disponible.

Nos résultats suggèrent que les protéines majoritaires sont déjà complètes après assemblage des EST à partir des transcriptomes séquencés plus exhaustivement, comme *Vitis*. Dans ce cas, la combinaison avec les séquences orthologues n'améliore pas significativement le taux d'identification. Cependant les contigs assemblés à partir de transcrits minoritaires demeurent plus courts, de sorte que l'assignation des protéines moins abondantes, comme celles isolées à partir de vésicules tonoplastiques, est améliorée lorsque les PMF sont analysés contre les assemblages conceptuels orthologues / EST.

IV.4.3 Matériel et Méthodes

IV.4.3.1 Matériel biologique et production des empreintes massiques peptidiques.

Les protéines totales de raisin ont été extraites suivant Sarry *et al.* (2004). Les protéines des péricarpes d'abricot et de mangue, ainsi que des feuilles de citronnier, ont été extraites dans un tampon aqueux et précipitées avec du TCA. Les membranes tonoplastiques de baies de raisin mûr ont été extraites et purifiées sur un gradient de saccharose, comme décrit dans Terrier *et al.* (2001b). Les électrophorèses 2D, la coloration au bleu de coomassie, la digestion trypsique et les analyses par MALDI-TOF et MALDI-TOF-TOF ont été conduites comme dans Sarry *et al.* (2004).

IV.4.3.2 Assemblage des EST

Les ADNc partiels qui représentent 95164, 29788 et 11339 séquences respectives de *Vitis*, *Prunus* et *Citrus* ont été analysés avec le pipeline TGICL (Pertea *et al.*, 2003) pour

obtenir respectivement 19882, 10981 et 5748 séquences non redondantes. Les paramètres utilisés pour l'assemblage par CAP3 étaient $p=93$ et $o=100$.

IV.4.3.3 Construction de la base de données chimériques

Une analyse a été effectuée par BLASTX en local, afin de comparer la totalité des nrEST issues des espèces fruitières (en tant que séquences à identifier) contre les 243182 séquences importées de la section viridiplantae de la base de données protéines du NCBI du 12/06/03. Un logiciel Perl (PT-HT assembler) a permis de construire une base de données contenant des séquences chimériques dans lesquelles les séquences genre-spécifiques ont été incorporées dans les protéines orthologues. Ce logiciel utilise les modules bioperl (Stajich *et al.*, 2002) de traitement de sorties de Blast (algorithme en figure 57), et concatène les transcrits partiels (EST non redondantes avec les protéines orthologues).

L'e-value maximale a été fixée à $1 e^{-10}$ et seuls les dix meilleurs alignements ont été considérés. Le traitement du fichier de sortie de BLASTX avec PT-HP Assembler a ainsi produit de une à dix constructions chimériques par nrEST. Pour chaque nrEST, la région traduite remplace sa contrepartie dans la séquence des dix meilleurs alignements de la base de données des plantes. Par ce moyen, les transcrits partiels (contigs ou singlets) sont complétés par les plus proches orthologues. Les cadres ouverts de lecture probables des nrEST qui échappent à la procédure précédente, car elles ne trouvent pas d'homologues parmi les protéines de viridiplantae, ont été obtenues avec le logiciel getorf (Rice *et al.*, 2000), avant d'être incluses dans la base de données finale, sous réserve de comporter plus de 59 acides aminés. Afin d'autoriser une analyse exhaustive des PMF dans une seule recherche avec MASCOT, les protéines non redondantes de viridiplantae ont été également ajoutées à la base de données, qui contient finalement 570752 séquences.

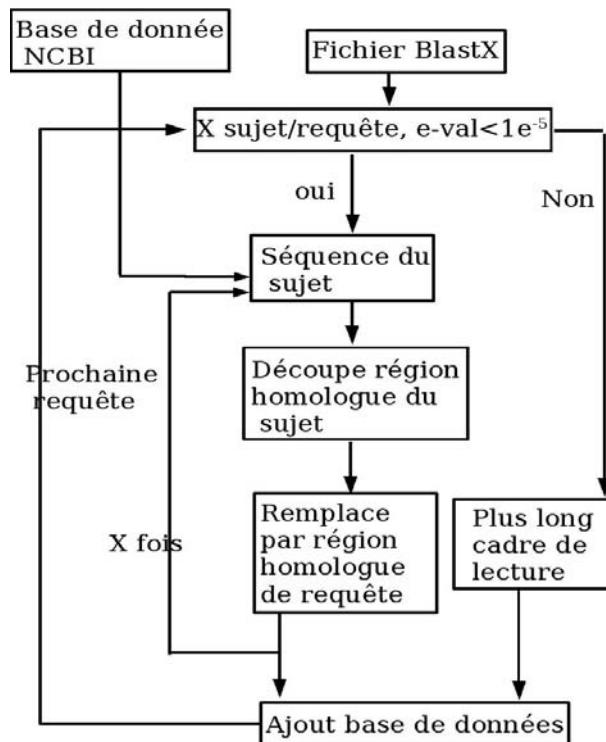


Figure 57 : Algorithme de PT-HP Assembler. X correspond à un nombre entre 1 et 10 et représente le nombre d'orthologues suffisamment proches.

IV.4.3.4 Identification des PMF

Pour l'identification des PMF, le logiciel de recherche MASCOT (Matrix Science, London, UK) a été utilisé en acceptant une probabilité maximale de 5 % pour que la correspondance avec une protéine soit due au hasard, ce qui équivaut à un score mowse de 70 pour cette banque. Les autres paramètres suivants ont été utilisés : (i) une variation maximale de la masse d'un peptide de 100 ppm, (ii) un minimum de 4 peptides retrouvés sur la séquence de la protéine, (iii) un clivage tryptique manqué toléré et (iiii) aucune modification chimique autorisée.

IV.4.4 Résultats et Discussion

IV.4.4.1 Conception de la base de données de chimères entre EST et protéines orthologues

Le pourcentage des EST non redondantes qui ne trouvent pas d'homologues parmi les protéines de viridiplantae au NCBI est de 23, 22, et 35 pour *Prunus*, *Citrus* et *Vitis*

respectivement, et le reste des nrEST présentent en moyenne 70 % d'identité avec leur plus proche homologue. Puisque l'identification des PMF suivant la stratégie « inter-espèces » devient difficile lorsque l'homologie est inférieure à 80 % d'identité (Wilkins et Williams, 1997), il peut être anticipé qu'une correction partielle des séquences orthologues avec les mutations spécifiques d'une espèce, disponibles dans les banques d'EST, pourrait sensiblement améliorer l'identification des protéines. Le PT-HP Assembler (Fig. 57) a été créé pour substituer les régions homologues dans une protéine complète du NCBI, par la séquence correspondante traduite à partir d'une nrEST de l'espèce considérée. Pour chaque nrEST, la chimérisation est réalisée sur au maximum les 10 meilleurs homologues, avec une E-value de $1e^{-10}$.

Dans le but de vérifier l'efficacité de l'identification des protéines en utilisant l'approche protéines chimériques EST/orthologue développée ici, trois banques de protéines chimériques ont été construites en complétant les produits de traduction conceptuels des nrEST de *Vitis*, *Prunus* et *Citrus*. Les genres choisis correspondent à des projets incluant un nombre variable d'EST. *Vitis* est devenue l'un des principaux genres végétaux en terme de nombre d'EST publiquement accessibles depuis le travail précédant de Sarry *et al.* (2004). Les *Prunus* et les *Citrus* correspondent à des projets intermédiaires, alors que peu de données sont disponibles sur les *Mangifera*.

La longueur moyenne des 1000 nrEST les plus abondantes par banque est de 942, 958 et 1156 pour les *Citrus*, *Prunus* et *Vitis* respectivement. Cette valeur ne dépend pas uniquement du nombre total d'EST séquencées mais aussi des stratégies de clonage (kit de clonage d'ADNc "longueur complète" ou non) et de l'orientation de séquençage (3', 5' ou les deux). La base de données protéiques a donc été particulièrement enrichie en séquences chimériques d'espèces ligneuses.

IV.4.4.2 Analyse par MALDI-TOF des spots issus des gels 2D.

Quatre jeux de PMF ont été obtenus pour les protéines totales, incluant 207 PMF de *Vitis vinifera* (raisin), 58 de *Prunus armeniaca* (abricotier), 106 de *Mangifera indica* (mangue), issus de péricarpe (pulpe et peau) et 43 de feuille de *Citrus lemon* (citron).

Le nombre maximal de spots observés sur les gels 2D est de 1000 (Fig. 58), ce qui indique que les tentatives d'identification s'effectueront sur les 10 % des protéines les plus abondantes au sein des protéomes complets. En fait, ce serait plutôt 1 % des protéines qui serait visualisé sur les gels 2D si on compare avec le protéome humain, estimé environ à

100000 protéines (Harrisson *et al.*, 2002). Pour avoir une vision de l'efficacité de la procédure sur des protéines moins abondantes (pour lesquelles, il est moins probable que les contigs d'EST correspondent à la séquence complète), 23 PMF ont été obtenus à partir de protéines de fraction tonoplastique de raisin, solubilisées dans le même tampon CHAPS/triton/urée que celui utilisé pour les protéines totales (Fig. 58E). La fraction tonoplastique obtenue après une purification sur gradient de saccharose d'un culot microsomal représente moins de 0,1 % des protéines totales. Les protéines principales présentent alors un pI approximatif de 5 sur les gels 2D, tandis que les spots de protéines totales sont distribués de façon homogène entre un pI de 4,5 à 8,5 dans toutes les espèces étudiées.

Pour chaque espèce, l'identification des PMF a été tentée sur la base de données totale incluant les protéines non redondantes du NCBI de viridiplantae, les chimères de *Vitis*, de *Citrus* et de *Prunus* (570752 séquences au total) avec le logiciel MASCOT.

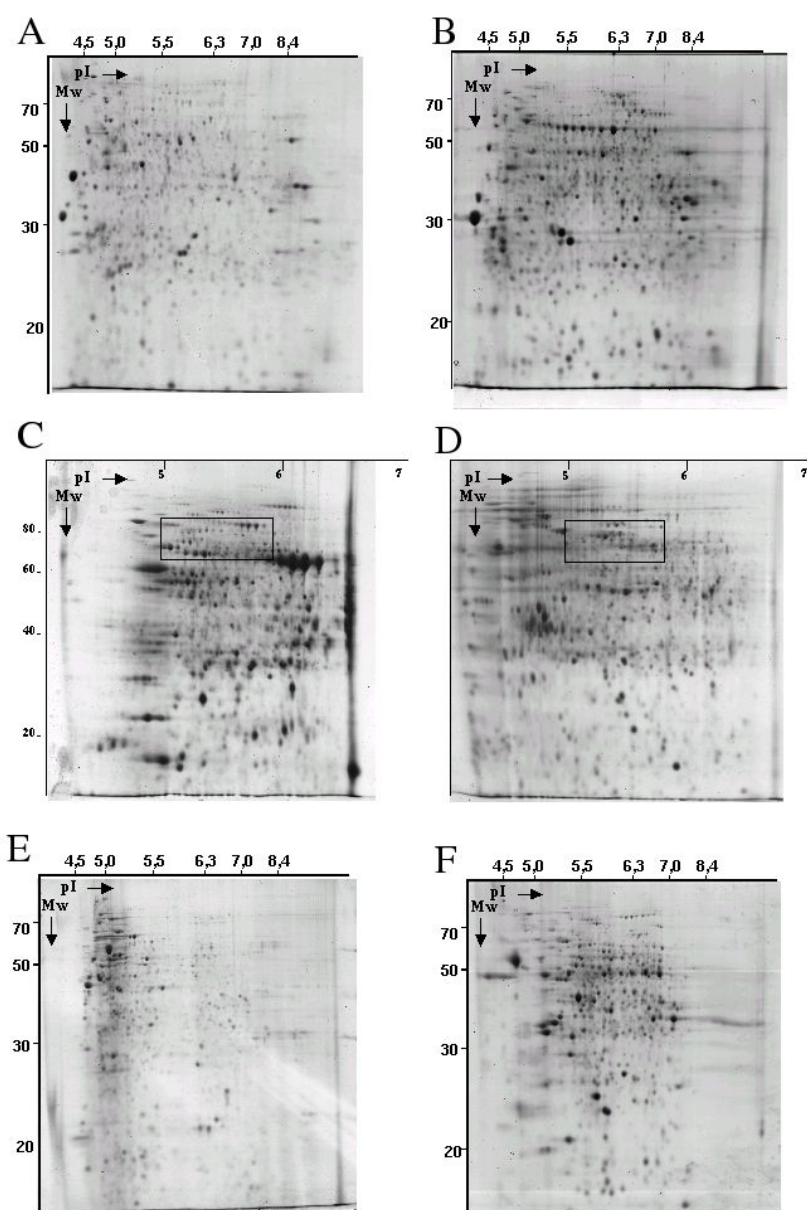


Figure 58 : Gels 2D des six extraits protéiques utilisés pour la validation de la méthode de chimérisation. **A** Gel 2D correspondant au protéome total de péricarpe de mangue. **B** Gel 2D correspondant au protéome total de péricarpe d'abricot. **C** Gel 2D correspondant au protéome total de feuilles matures de citron, la zone rectangulaire correspond à la figure 59a. **D** Gel 2D correspondant au protéome total de feuilles immatures de citron, la zone rectangulaire correspond à la figure 59b. **E** Gel 2D correspondant au protéome de la fraction tonoplastique de raisin. **F** Gel 2D gel correspondant au protéome du péricarpe de raisin. La première dimension a été réalisée pour tous les gels avec des strips de zone de pH de 3-10, excepté pour les feuilles de citron (strip de pH 4-7).

IV.4.4.3 Illustration et validation des peptides prédits par MS/MS

Des agrandissements des régions des gels de feuilles de citronnier à partir desquels les spots ont été excisés sont présentés dans la figure 59. Les spots identifiés sont présentés dans le tableau 20.

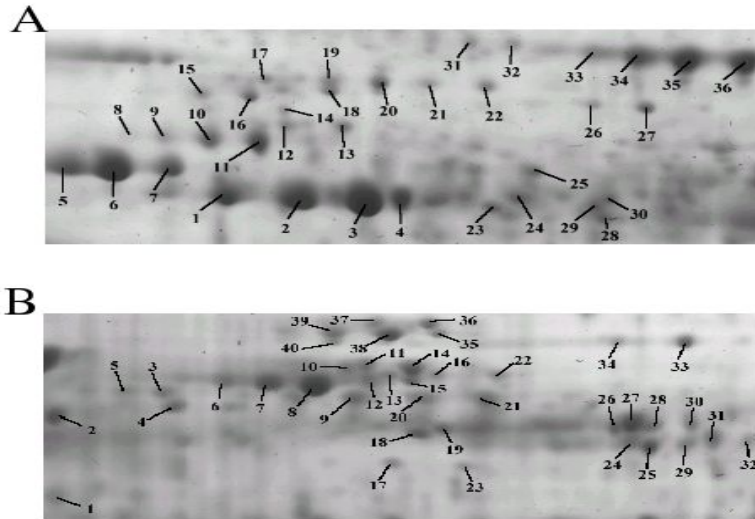


Figure 59 : Agrandissement de la zone à partir de laquelle les spots destinés aux analyses MS dans les feuilles **A**) matures et **B**) jeunes de citronnier ont été excisés.

Sur les feuilles de citronnier, cinq protéines ont été identifiées par comparaison inter-espèces, six en tant que protéines chimériques EST/orthologue et 7 en tant que séquences déjà complètes chez les *Citrus*. L'interrogation des assemblages EST/orthologue permet d'augmenter de 6,0 peptides, en moyenne, le nombre de fragments massiques appariés par rapport à l'interrogation des nrEST, tandis que l'augmentation est de 1,8 peptides par rapport aux séquences homologues dans les protéines de viridiplantae. Ces chiffres démontrent clairement que des informations complémentaires sont assemblées à partir des EST et des orthologues dans les constructions chimériques. Le tableau 20 fournit un aperçu sur les annotations fonctionnelles.

Tableau 20 : Origine de la séquence des protéines chimériques qui ont permis d'identifier les spots. Les spots pour lesquels l'identifiant commence par V proviennent de feuilles matures de citron. Les spots pour lesquels l'identifiant commence par J proviennent de feuilles jeunes de citron. Les spots pour lesquels l'identifiant commence par G proviennent de raisin. Les spots pour lesquels l'identifiant commence par M proviennent de tonoplaste de raisin. Les spots pour lesquels l'identifiant commence par S proviennent d'abricot. Les spots de citron identifiés grâce à des protéines non chimériques (d'espèces *Citrus* ou non) sont aussi présentés. Le score mowse est indiqué entre parenthèse et un score de 70 correspond à 95 % de confiance dans l'identification. ND protéines orthologues non détectée dans les cinquante premières protéines appariées. PM: poids moléculaire ; Acc#, numéro d'accession dans genbank.

	Spot #	Nbre de peptides appariés	Peptides dans la séquence de la nrEST	Peptides dans la séquence homologue	Construction chimérique			PM	Acc #	Fonction
					Espèce homologue	Genre de la NrEST	NrEST position			
Homologues	V11	10	0	10	<i>Pisum sativum</i>			63196	806808	chaperonin precursor
	J05	8	0	8	<i>Oriza sativa</i>			44662	11094192	26S proteasome subunit4
	V09	6	0	6	<i>Pisum sativum</i>			62919	806808	chaperonin precursor
	J08	10	0	10	<i>Pisum sativum</i>			62919	2506277	RuBisCO binding β subunit
	J07	6	0	6	<i>Oriza sativa</i>			29694	20043009	Transcription elongation factor
Citrus séquences complète	V01	10	10			<i>Citrus</i>	Complete	51950	6634488	ATP synthase
	V20	10	10			<i>Citrus</i>	Complete	68769	6518112	H ⁺ -ATPase catalytic subunit
	V02	13	13			<i>Citrus</i>	Complete	53553	6599365	atp synthase beta subunit
	J37	13	13			<i>Citrus</i>	Complete	68638	6518112	H ⁺ -ATPase catalytic subunit
	J02	13	13			<i>Citrus</i>	Complete	54329	11266665	H ⁺ -transporting ATPase
	J18	13	13			<i>Citrus</i>	Complete	53553	6599365	atp synthase beta subunit
	J27	6	6			<i>Citrus</i>	Complete	56299	1085960	inositol-3-phosphate synthase
Chimères	J01	7(71)	6	5(58)	<i>Z. aethiopica</i>	<i>Citrus</i>	35-389	46904	13430336	rubisco activase
	J33	8(88)	4	7(72)	<i>M. crystallinum</i>	<i>Citrus</i>	7-267	61404	602426	phosphoglyceromutase
	J17	7(146)	2	5(107)	<i>S. tuberosum</i>	<i>Citrus</i>	12-118	46453	1155334	Tat-binding homolog ATPase
	J23	12(125)	2	10(96)	<i>A. thaliana</i>	<i>Citrus</i>	5-113	46738	6652882	26S proteasome AAA-ATPase
	V17	9(96)	2	7(60)	<i>Oriza sativa</i>	<i>Citrus</i>	1-235	80306	20256	heat shock protein 82 (HSP82)
	J04	11(121)	2	9(90)	<i>Oriza sativa</i>	<i>Citrus</i>	462-506	55810	226696	ATPase alpha
	G06	10(101)	9	9(88)	<i>N. tabacum</i>	<i>Vitis</i>	81-325	170241	35571	glyceraldehyde-3-P dsgh
	G30	9(74)	6	ND	<i>Beta vulgaris</i>	<i>Vitis</i>	534-843	2369714	93484	elongation factor
	G04	7(85)	5	ND	<i>A. thaliana</i>	<i>Vitis</i>	22-158	7288027	30559	hypothetical protein
	G15	11(142)	10	ND	<i>Corylus avellana</i>	<i>Vitis</i>	11-502	18479082	59031	globulin
	G05	9(109)	8	ND	<i>Cicer arietinum</i>	<i>Vitis</i>	137-479	8894548	54374	seed storage protein
	G103	6(81)	5	ND	<i>A. thaliana</i>	<i>Vitis</i>	4-260	3157928	42355	fumarylacetoacetate hydrolase
	G25	11(125)	6	7(63)	<i>N. tabacum</i>	<i>Vitis</i>	90-246	4539545	25194	Proteasome subunit
	G49	12(70)	2	11(63)	<i>A. thaliana</i>	<i>Vitis</i>	1-190	30695996	71439	cytosolic factor
	G66	10(111)	4	8(88)	<i>Oryza sativa</i>	<i>Vitis</i>	335-453	18001149	48657	translational elongation factor Tu
	G84	8(75)	7	ND	<i>A. hypochondriacus</i>	<i>Vitis</i>	23-495	558149	60177	seed storage protein
	M23	10(116)	9	4(57)	<i>P. armeniaca</i>	<i>Vitis</i>	53-421	4887133	49234	calreticulin
	M24	7(103)	6	5(63)	<i>Ricinus communis</i>	<i>Vitis</i>	49-415	1658197	48795	calreticulin
	M29	10(88)	6	ND	<i>A. thaliana</i>	<i>Vitis</i>	182-466	9758328	53173	xylose isomerase
	M30	13(151)	7	11(118)	<i>M. crystallinum</i>	<i>Vitis</i>	349-485	26986106	54203	vacuolar ATPase subunit B
	M34	7(81)	6	6(65)	<i>A. thaliana</i>	<i>Vitis</i>	56-230	4850294	26160	Vacuolar ATPase chain E
	M38	9(108)	8	ND	<i>A. thaliana</i>	<i>Vitis</i>	8-428	30691729	49169	aminoacylase
	M40	7(80)	2	5(55)	<i>A. thaliana</i>	<i>Vitis</i>	2-93	9958054	36797	G-3-phosphate dehydrogenase
	S01	5(70)	3	2(NA)	<i>Castanea sativa</i>	<i>Prunus</i>	1-65	17647	3819743	HSP17
	S06	12(145)	8	5(64)	<i>Manioli liliiflora</i>	<i>Prunus</i>	1-184	37114	66014	G-3-P dehydrogenase
	S08	22(156)	16	7(NA)	<i>Vitis vinifera</i>	<i>Prunus</i>	1-356	65514	515759	Malate dehydrogenase
	S12	9(72)	8	7(48)	<i>Medicago sativa</i>	<i>Prunus</i>	268-805	92365	6683114	Sucrose synthase
	S16	12(109)	11	9(71)	<i>Medicago sativa</i>	<i>Prunus</i>	268-805	92365	6683114	Sucrose synthase
	S22	9(98)	6	6	<i>Glycine Max</i>	<i>Prunus</i>	380-485	56363	2529229	6-P gluconate dehydrogenase
	S54	12(102)	10	7(45)	<i>A. thaliana</i>	<i>Prunus</i>	1-417	75362	6715645	Unknown

L'alignement ClustalW (Thompson *et al.*, 1994) des trois séquences relatives au spot J01 (Tab. 20), c'est à dire la chimère et ses parents, la nrEST et l'orthologue de *Zantedeschia aethiopica* (Rubisco Activase) est présenté dans la figure 60 pour illustrer la méthode. Les parties N-terminale et C-terminale de la chimère proviennent de la séquence de *Zantedeschia*

aethiopica, et la région centrale est spécifique des *Citrus*. Les peptides N° 2, 4 et 5 sont conservés entre les deux espèces et auraient donc été identifiés aussi bien dans une approche inter-espèces que dans une approche EST. Le peptide N° 6 aurait également été identifié par les deux approches, puisqu'il inclut une mutation (I/L) qui conserve la masse. Il est raisonnable de supposer que le peptide provenant de *Citrus* est celui qui existe réellement, cependant cette mutation pourrait apparaître dans de nouvelles isoformes après complémentation du transcriptome des *Citrus*. Les peptides N° 1 et 3 sont *Citrus* spécifiques, avec une mutation (I/V) dans le peptide N° 1 et deux mutations (E/D, I/V) dans le peptide N° 3. Le peptide N° 3 inclut le peptide N° 2 du fait d'un clivage manqué. Le peptide N° 7 est présent uniquement dans la séquence de l'orthologue, dans une région non séquencée chez les *Citrus*.

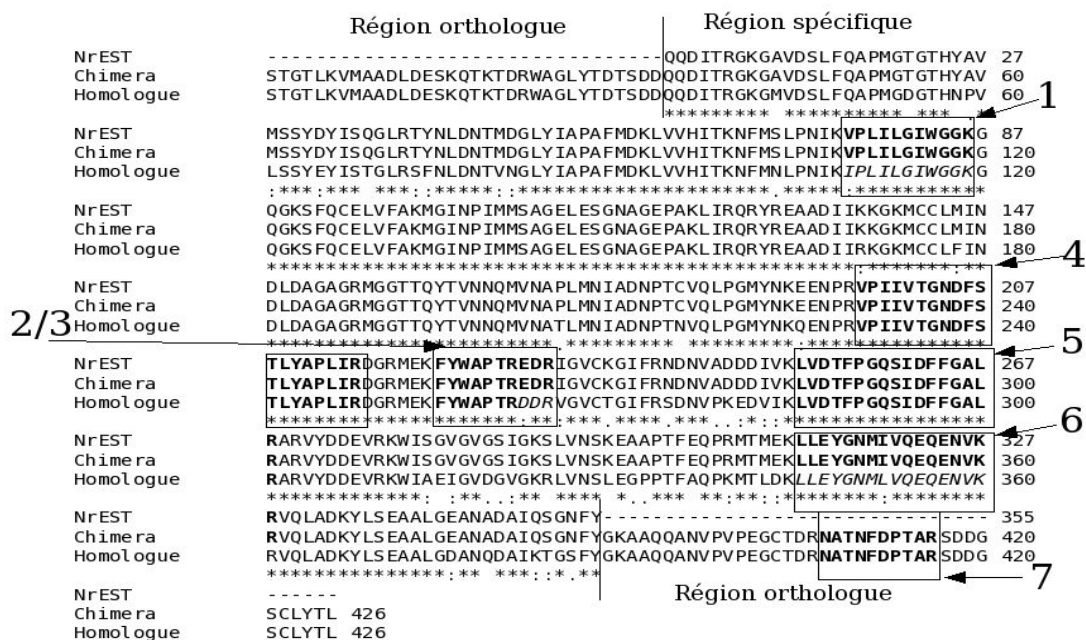


Figure 60 : Alignement multiple de la séquence chimérique identifiant le spot J01 de feuille de citron avec ses parents (la nrEST et la Rubisco activase de *Zantedeschia aethiopica*). Chaque numéro correspond aux peptides identifiés dans le PMF de J01, avec leur localisation sur les trois séquences. Les séquences où le peptide conceptuel correspond au peptide expérimental sont représentées en gras. Les séquences où le peptide conceptuel ne correspond pas au peptide expérimental sont représentées en italique.

La spectrométrie de masse MS/MS a été réalisée sur un fragment tryptique d'une protéine d'abricot de 56kD identifiée comme une 6-phosphogluconate-déshydrogénase (spot S22, 6PDH). Parmi les 7 peptides appariés avec des fragments tryptiques de ce spot, 4 ont été retrouvés dans les deux espèces, un dans la séquence de la nrEST et deux dans la protéine de soja. L'identification fonctionnelle a été confirmée par le séquençage du fragment

GFPISVYNR, attendu dans les deux espèces. D'une façon plus intéressante, puisque situé dans une région non séquencée de la nrEST d'abricot, le peptide DYFGAHTYER a aussi été validé par séquençage. Une telle confirmation expérimentale que la séquence prédite *in silico* existe dans la 6PDH de l'abricot souligne l'aspect heuristique de cette méthode d'analyse de PMF contre la base de données de protéines de chimères EST/orthologue.

Le nombre de peptides appariés dans les PMF augmente fortement dans certaines chimères par comparaison à l'identification en inter-espèce ou dans la même espèce avec les séquences d'EST (Tab. 20). Par conséquent, la fiabilité de l'identification augmente et de nouvelles protéines peuvent être identifiées. L'identification *de novo* de l'assignation des PMF est restreinte aux chimères pour lesquelles les parents partagent au moins 81 % d'identité, indépendamment de la position des espèces sur l'arbre génétique. Ce qui confirme la prédiction théorique de Wilkins et Williams (1997) et montre clairement que 80 % d'identité serait fréquemment insuffisant pour une identification fiable des PMF. Quoi qu'il en soit, le jeu complet de données confirme que les appariements avec les EST autorisant une identification significative sont sévèrement contraints par la taille des contigs, comme détaillé dans le paragraphe suivant.

IV.4.4.4 Valeur ajoutée de la procédure par rapport aux approches EST et inter-espèces

La figure 61 montre l'augmentation de l'identification en fonction de la taille de la base de données EST pour des PMF obtenus sur des spots de protéome total. Le pourcentage d'acides aminés homologues dans les chimères augmente avec la redondance de la banque d'EST : les protéines conceptuelles dérivées de contigs formés à partir de moins de 10 EST représentent en moyenne 48 % de la séquence complète, et cette valeur monte à 81 % pour les contigs assemblés avec plus de 10 EST. Trente pour cent des PMF identifiés l'ont été à partir de chimères chez les *Prunus* et les *Citrus*. Ceci correspond à un accroissement d'au moins 20 % de PMF identifiés pour les espèces dont les bases de données de nrEST contiennent entre 3000 et 16000 nrEST. Comme des banques d'EST comprenant plus de 32000 EST n'existent actuellement que pour 50 organismes eucaryotes, cette approche améliorerait ainsi substantiellement l'identification présomptive des peptides pour beaucoup de protéomistes travaillant sur d'autres espèces non modèles.

Tous les PMF identifiés à partir de mangue s'apparient avec des protéines orthologues, mais il faut noter qu'un quart d'entre eux sont des constructions chimériques d'autres espèces fruitières étudiées dans ce travail. Ce résultat inattendu pourrait indiquer que des isoformes

exprimées spécifiquement dans les tissus de péricarpe puissent partager des mutations conservatives parmi différentes plantes ligneuses, actuellement sous-représentées dans les bases de données de *viridiplantae*.

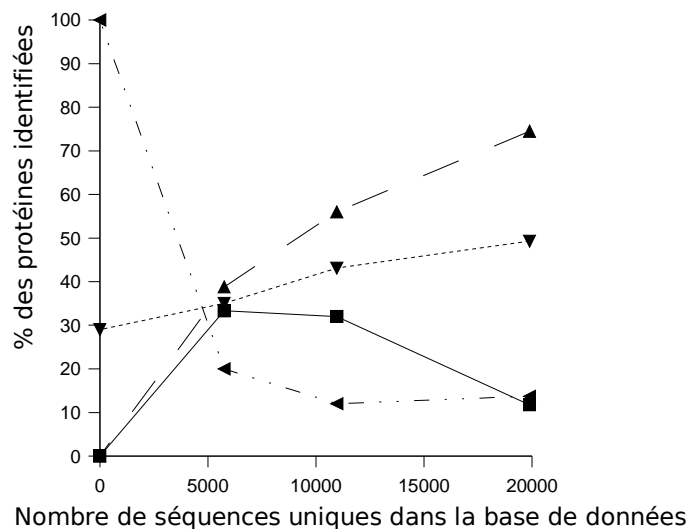


Figure 61 : Efficacité relative de l'identification par l'approche "chimère" pour les protéines majoritaires. Le nombre total de nrEST disponibles dans les différents genres de fruits est (axe X): 0 pour les *Mangifera*, 5746 pour les *Citrus*, 10981 pour les *Prunus* et 19882 pour *Vitis*. Le pourcentage relatif de protéines identifiées comme meilleur apparié (i) dans les protéines en espèces croisées (▲), (ii) dans les ORF complètes pour le même genre (▲), et (iii) dans les protéines chimériques EST/orthologues (■), sont comparés et exprimés en pourcentage des PMF identifiés. Le pourcentage d'identification (▼) est exprimé par rapport aux PMF totaux.

Peu de PMF ont été identifiés grâce à des chimères dans le raisin, puisque la redondance des transcrits majeurs permet l'assemblage de séquences presque complètes. Cependant les protéines totales colorées sur des gels 2D ne représentent que le millier des protéines les plus abondantes. L'un des avantages de l'approche protéomique est que les protéines minoritaires peuvent être étudiées après un fractionnement cellulaire et purification d'organites. La figure 32 montre la comparaison entre les protéines liées à la membrane du tonoplaste et les protéines totales en terme de meilleur appariement. Trente pour cent des meilleurs appariements correspondent à des chimères sur la fraction tonoplastique, tandis que ce pourcentage n'est que de 5,8 % sur les protéines totales. Ce résultat montre l'utilité de la base de données chimériques pour étudier les protéines moins abondantes, pour tous les organismes chez lesquels le génome complet est inconnu.

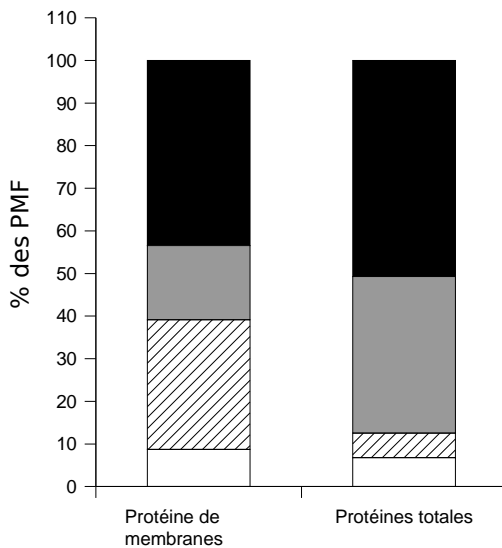


Figure 62 : La stratégie de recherche dans les protéines EST/orthologues compense le manque d'ORF complètes pour les protéines liées à la membrane du tonoplaste. Pourcentage de : (i) protéines non identifiées (noir), protéines identifiées avec (ii) des ORF complètes de raisin (gris), (iii) des protéines chimériques (hachurées), et (iiii) par l'approche espèces croisées (blanc).

IV.4.5 Conclusion

L'approche « chimère » est particulièrement efficace sur des projets EST de taille moyenne et peut être utile pour tous les groupes de recherche travaillant sur des espèces où approximativement cinq à trente mille EST ont été séquencées. Les données présentées ici montrent aussi que même des programmes de séquençage d'EST plus exhaustifs ne permettent pas de produire assez de séquences complètes pour identifier sans ambiguïté des protéines non prévalentes détectées après fractionnement cellulaire. Ces résultats indiquent aussi une conservation significative des isoformes exprimées dans différentes plantes fruitières ligneuses, ce qui devrait être confirmé à l'issue du séquençage du premier génome d'une espèce ligneuse fruitière.

Un biais statistique est toutefois introduit en utilisant la base de données de chimères avec le logiciel MASCOT, biais déjà existant dans les bases de données hétérogènes dans lesquelles les protéines de ménage les plus abondantes sont sur-représentées. Le nombre total de protéines dans la base de données chimériques excède le nombre probable de transcrits cellulaires d'un facteur deux, et les protéines majoritaires sont sur-représentées par rapport aux minoritaires. L'accumulation, dans une base de données de protéines orthologues bien au delà du nombre réel de membres d'une famille multigénique accroît artificiellement la probabilité qu'un fragment de masse conservé apparaisse comme un événement aléatoire.

Il en découle que le score Mowse ne fournit plus une indication claire de l'identification d'une protéine, et la fiabilité dans l'identification doit être accrue par d'autres techniques (Wilkins et Williams, 1997). La vérification que la masse prédite des protéines est cohérente avec les observations sur les gels 2D reste obligatoire. Cette vérification est clairement impossible sur des protéines partielles provenant d'EST seules. De plus, la stratégie de chimérisation permet de prendre en compte les fragments de séquence ayant disparus entre fragments orthologues. Une analyse en temps réel des PMF contre la base de données chimériques et la validation par MS/MS des peptides en dehors de la séquence de l'EST apparaît comme la stratégie la plus prometteuse, puisqu'elle facilite la définition d'amorces pour accélérer le séquençage d'ADNc des gènes candidats parmi les protéines chimériques identifiées *de novo*.

IV.5 Étude du protéome

IV.5.1 Introduction

L'analyse des protéines sur gel 2D permet d'obtenir un complément à celle des transcrits sur microarray. Elle fournit une information sur l'abondance des protéines, et permet donc d'avoir une vision plus précise des régulations dans la cellule. Les ARN messagers transcrits sont éventuellement traduits en protéines, mais la corrélation entre les abondances de ces deux types de molécules n'est pas absolue et peut être gène-spécifique (Greenbaum *et al.*, 2002). La mesure de la corrélation de leurs abondances fournit également une idée des niveaux de régulation et des étapes où elles se mettent en place.

Nous avons dans ce travail analysé l'abondance des protéines sur des gels à deux dimensions à huit stades du développement du fruit, dont certains stades clés ont été également analysés au niveau transcriptomique (Fig. 64, chapitre IV.5.3). Nous nous sommes particulièrement intéressés aux protéines dont l'abondance varie le plus au cours du développement, mais également aux protéines majoritaires. Nous avons pu en identifier certaines, afin d'évaluer la corrélation entre l'abondance des protéines et celle des ARNm en partie grâce à la méthode de chimérisation (chapitre IV.4).

IV.5.2 Mise au point du protocole d'extraction des protéines totales d'abricot.

IV.5.2.1 Evaluation des protocoles existants

Il existe peu de protocoles dans la littérature concernant l'extraction des protéines totales des fruits adaptés à l'électrophorèse 2D. Les protocoles utilisés pour extraire les protéines de fruits charnus sont basés sur des séparations biphasiques d'un mélange phénol/tampon aqueux (Lelievre *et al.*, 1995 ; Abdi *et al.*, 2002). Ce type de protocole nous a permis d'obtenir un rendement de 100 µg de protéine/g de poudre d'abricot, ce qui s'avère relativement faible. De

plus, les gels ne contenaient que très peu de protéines différentes (Fig. 63A), et la reproductibilité de la méthode était problématique. Compte tenu (i) du danger de l'utilisation du phénol, (ii) du caractère aléatoire de l'étape de séparation des phases et de son influence probable sur la reproductibilité du contenu protéique, nous n'avons pas essayé d'optimiser ce protocole, bien qu'il permette d'obtenir des gels de meilleure qualité que ceux basés sur la précipitation des protéines par le TCA sur la tomate (Saravanan et Rose, 2004). Le second protocole, mis au point par Sarry *et al.* (2004) sur le raisin a été testé (Fig. 63B). Cette méthode utilise la technique très répandue pour l'extraction des protéines végétales de précipitation des protéines par acétone/TCA, elle présente cependant l'inconvénient de ne pas fonctionner sur des raisins congelés, ce qui a pu être vérifié sur l'abricot. Ainsi un nouveau protocole a été mis au point.

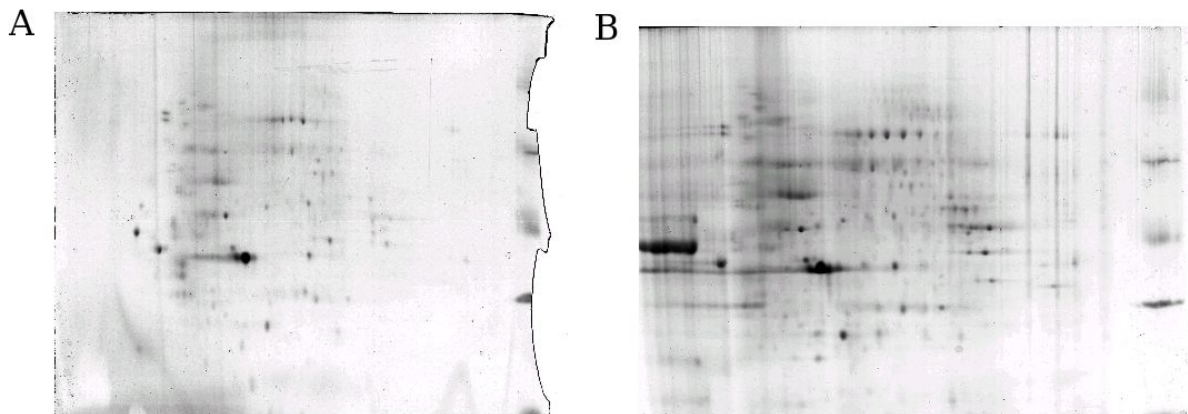


Figure 63 : Gel d'électrophorèse 2D de protéines d'abricot après extraction des protéines **A** par la méthode de séparation biphasique au phénol et **B** par précipitation des protéines dans un mélange TCA/Acétone.

IV.5.2.2 Mise au point du protocole.

Les protéines sont solubilisées par un tampon aqueux contenant divers protecteurs efficaces pour l'obtention d'extraits enzymatiques actifs à partir de tissus riches en tanins. Il contient : un tampon (MES), une antiprotéase (PMSF), trois réducteurs (DTT, β -mercaptoéthanol, métabisulfite), et un chélateur de métaux (EDTA). Du triton X-100 a été ajouté pour solubiliser les protéines hydrophobes et entrer en compétition avec les polyphénols dans les interactions avec les protéines.

La force ionique du tampon a été augmentée à 0,6 Eq. Ceci afin de limiter les variations relatives de force ionique dans le tampon provoquées par l'évolution de l'acidité pendant le mûrissement (0,3 Eq/kg), et qui pourraient affecter la solubilisation des protéines en fonction du stade de développement. Compte tenu de la dilution du fruit dans le tampon, la force

ionique ne varie plus que de 20 % entre les stades extrêmes. Il est possible que certaines protéines ne soient pas solubilisées à cette force ionique élevée, mais elle devrait être favorable à l'extraction des protéines pariétales. Le pH final est fixé à 6,1 pour tous les stades, valeur déterminée empiriquement et permettant la précipitation des pectines qui, sinon, interdisent une bonne focalisation des protéines pendant la première dimension. Le rendement du protocole est d'environ 1 mg de protéine par gramme de poudre.

IV.5.3 Analyse des gels, estimation de la reproductibilité et de l'abondance des protéines

Les gels ont été analysés avec le logiciel Progenesis (NonLinear Dynamics, Newcastle). Il permet la détection des spots, l'appariement sur différents gels des spots correspondant à la même protéine et la quantification de l'abondance des spots. Pour chaque stade un gel moyen a été obtenu à partir de tous les spots détectés sur au moins un des trois gels (Fig. 64), l'intensité des spots composant les gels moyens correspond aux valeurs moyennes d'intensité de chaque spot. Les 8 gels moyens ont été comparés entre eux pour obtenir les profils d'abondance des protéines.

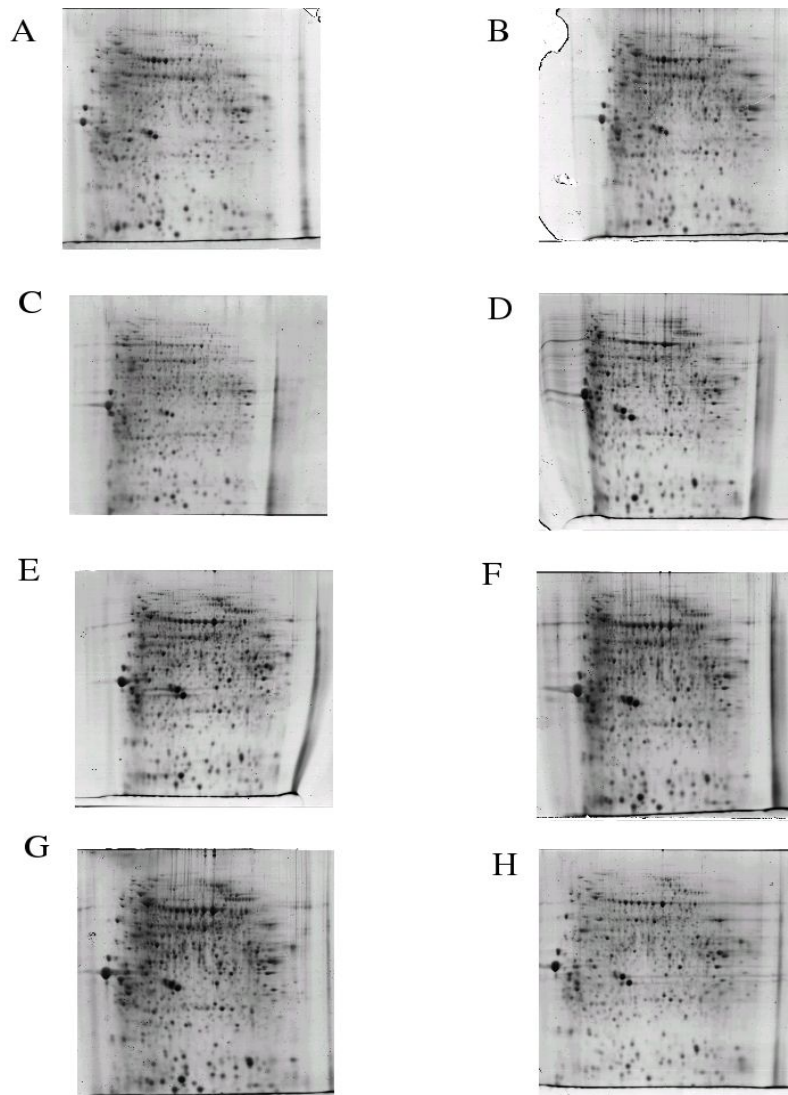


Figure 64 : Gels d'électrophorèse bidimensionnelle correspondant aux gels moyens obtenus à chaque stade de développement de l'abricot. **A**: stade 53 JAA. **B** stade 65 JAA. **C**: stade 92 JAA. **D**: stade 107 JAA. **E**: stade 109 JAA. **F**: stade 114 JAA. **G**: stade 121 JAA. **H**: stade 128 JAA.

En moyenne, la détection au bleu de Coomassie colloïdal a permis de repérer 909 spots par gel moyen. Au total 1412 spots différents ont été identifiés sur l'ensemble du développement. Le volume (intégration de l'intensité pour tous les pixels) correspond à la quantité de protéine dans le spot. Afin d'éliminer les variations expérimentales dues (i) à la coloration des protéines et du gel (bruit de fond) ; (ii) à la quantité de protéines absorbées lors de l'hydratation du strip d'IEF ; (iii) aux protéines bloquées sur le strip et ne migrant pas lors de la seconde dimension. Le volume de chaque spot est exprimé en pourcentage du volume total de tous les spots de chaque gel. Les protéines ne sont pas correctement focalisées en première dimension, dans une zone correspondant aux pI inférieurs à 5, pour la moitié des gels environ (exemple Fig. 64D). Cet événement survient aléatoirement, aussi bien dans des

extraits provenant du même stade (n'est donc pas lié au stade de maturité) qu'à l'intérieur d'une même série d'extraction (n'est pas lié à l'effet « jour d'extraction ». D'après sa localisation dans les gels et son occurrence plus fréquente dans les stades immatures, il est probable qu'il soit lié à la présence résiduelle de polymères d'acide galacturonique (pectines). Toutes les protéines dont les pI sont inférieurs à 5, soit 285 spots ont donc dû être éliminées de l'analyse. Les protéines de très faible abondance (moins de 0,001 % du volume totale) ont aussi été éliminées de l'analyse (135 spots). 992 spots sur 1412 ont donc été finalement conservés pour l'analyse de leur évolution.

Dans le but d'évaluer la reproductibilité globale, l'erreur standard relative a été déterminée pour chaque spot à partir des 3 réplicats. 64 % des spots présentent une erreur standard relative inférieure à 0,8 fois le volume du spot (Fig. 65) et sont donc considérés comme exploitables. Ce seuil, relativement élevé permet de conserver des spots dont l'abondance varie assez fortement dans un stade mais qui sont totalement absents dans d'autre. Au delà de cette valeur, il est systématiquement impossible de distinguer statistiquement un spot absent d'un spot fortement variable. La figure 66 présente l'erreur standard relative en fonction de l'intensité des spots, tous les spots fortement exprimés (Intensité > 0,6 %) sont extrêmement reproductibles ($\sigma < 0,4$) et la reproductibilité semble corrélée avec l'intensité du spot. A quelques exceptions près, les spots très peu reproductibles ont une intensité relative inférieure à 0,1 %. L'erreur standard et le niveau de variation entre deux stades (ratio > 2) nous ont permis d'effectuer un premier criblage de 91 spots potentiellement régulés au cours du développement à cette étape. Tous ces spots ont ensuite été vérifiés (i) visuellement sur les gels afin d'éliminer les spots non clairement individualisés et (ii) au niveau de la significativité de leur différence d'abondance entre les deux stades extrêmes (en terme d'abondance), par le test de Student ($p = 95$). Après cette étape, 42 spots ont été identifiés comme régulés pendant le développement. Parmi ces 42 protéines, des spots non significatifs, mais absents dans les 3 gels pour deux stades consécutifs au moins et présents dans les trois gels réplicats d'au moins deux autres stades consécutifs, ont cependant été conservés. Parmi ces 42 protéines, 18 ont été identifiées comme induites pendant le développement, 10 présentent une expression transitoire et 14 sont réprimées.

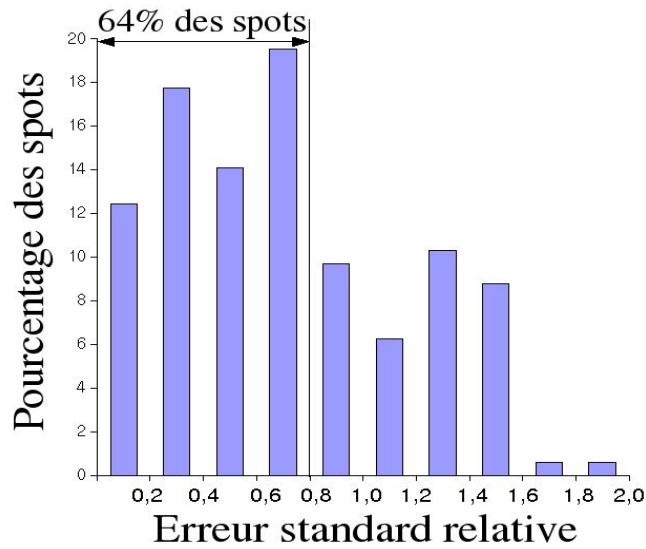


Figure 65 : Pourcentage des spots par classe d'erreur standard relative. L'erreur standard relative est le rapport entre l'écart type obtenu sur les 3 réplicats et le volume relatif moyen du spot.

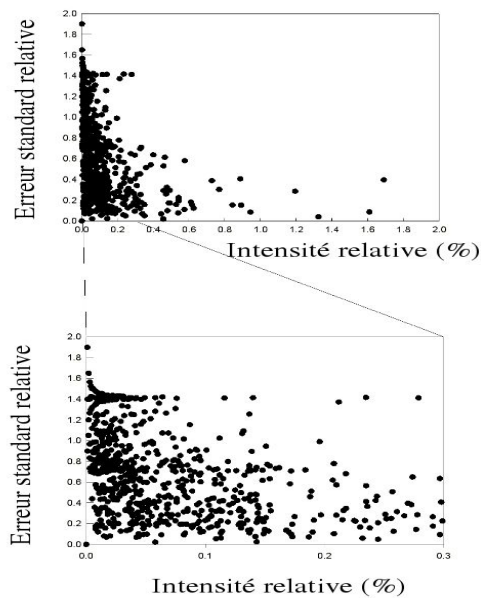


Figure 66 : Erreur standard relative en fonction de l'intensité des spots. L'erreur standard relative est le rapport entre l'écart type obtenu sur les 3 réplicats et le volume relatif moyen du spot. La figure du bas correspond à l'agrandissement de la figure du haut pour les spots d'intensité relative inférieure à 0,3 % de l'intensité totale des protéines.

IV.5.4 Les protéines identifiées

Une tentative d'identification des protéines a été effectuée sur les 42 spots régulés pendant le développement, les 10 spots majoritaires. Un spot supplémentaire, présumé significativement induit, s'est révélé contenir un mélange de protéines après identification, rendant impossible la détermination du profil d'expression. La localisation sur les gels d'électrophorèse de l'ensemble des protéines dont l'identification a été tentée, est présentée dans la figure 67. Trente trois des protéines régulées et 8 des protéines majoritaires ont pu être identifiées. Ce taux d'identification de 76 % est plus fort que celui qui est évoqué dans la partie validation de la technique d'aide à l'identification des protéines, car des modifications chimiques ont été autorisées et une analyse par MALDI-TOF supplémentaire a été réalisée pour les protéines non identifiées dans un premier temps. De plus, certaines protéines ont été identifiées par MS/MS.

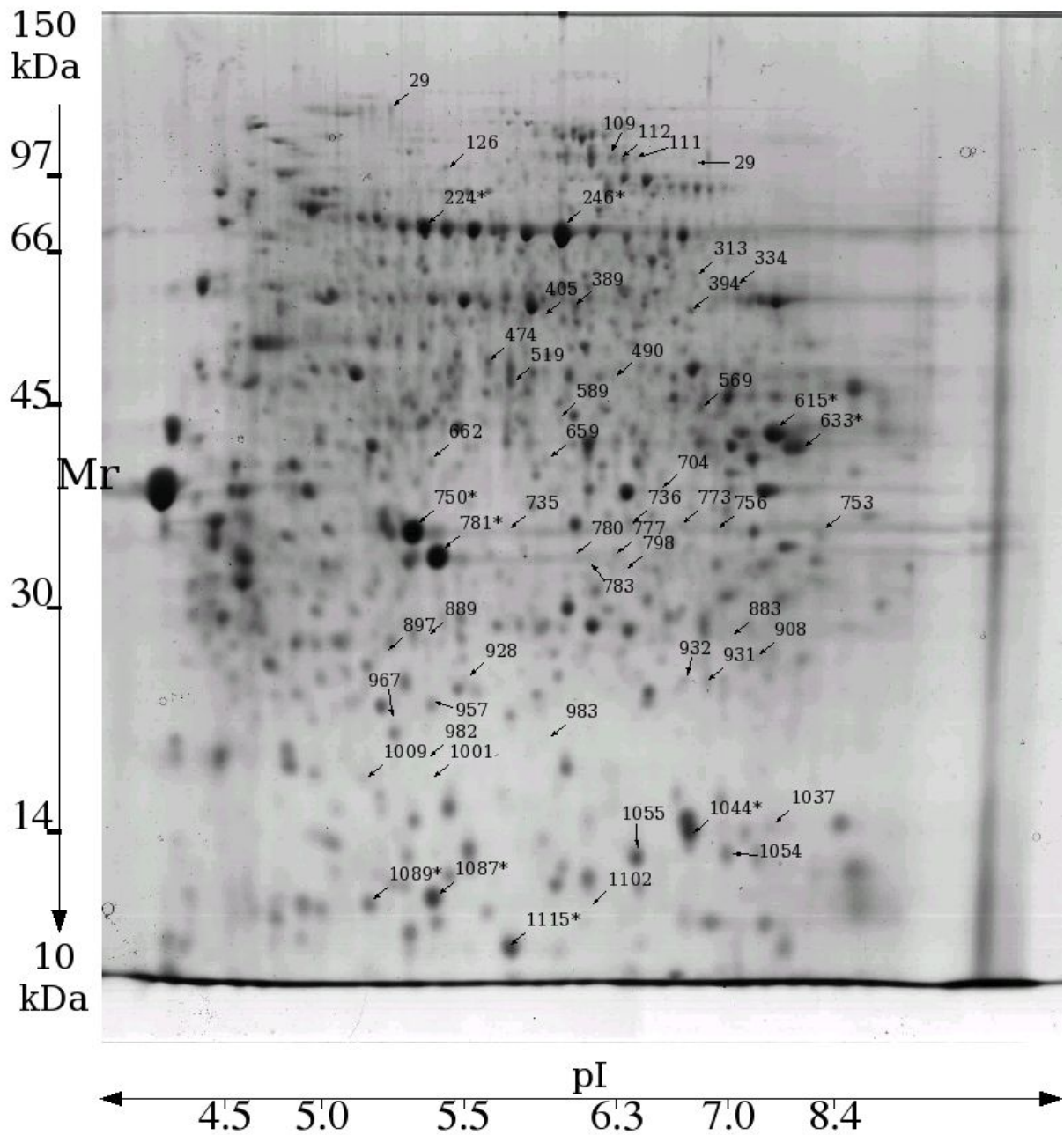


Figure 67 : Localisation des protéines dont l'identification a été tentée sur le gel 2D correspondant au stade 128 JAA. *protéines majoritaires et constitutives.

Près de la moitié des protéines régulées (44 %) sont reliées, à différents niveaux, au métabolisme du carbone, incluant des protéines impliquées dans la photosynthèse (21 %), le métabolisme du saccharose (12 %) et la glycolyse (12 %) (Fig. 68).

Les protéines induites sont impliquées principalement dans le métabolisme et la réponse aux hormones, l'expansion cellulaire, le métabolisme du saccharose. Les protéines dont l'intensité décroît sont majoritairement impliquées dans la photosynthèse. La fonction de 11 protéines n'a pu être identifiée (9 régulées et 2 constantes). Parmi ces protéines on retrouve

principalement des protéines de faible poids moléculaire (<25 kD) ou de très haut PM (>100 kD) qui sont difficiles à identifier. Pour les premières il faut un pourcentage de recouvrement très élevé pour que l'identification soit significative (exemples : spot 1087 Tab. 27 ; spot 1115 Tab. 26), ce qui nécessite la présence de presque tous les peptides dans la carte massique. Pour les secondes il faut un nombre de peptides appariés très élevé (exemple : spot 112, Tab. 22). L'absence d'identification peut donc être liée à la technique et indépendante de la présence ou de l'absence de la protéine dans la base de données.



Figure 68 : Classement par catégories fonctionnelles des protéines différentiellement régulées pendant le développement de l'abricot.

IV.5.4.1 Les protéines impliquées dans la photosynthèse

Neuf des protéines identifiées sont reliées à la photosynthèse ou à la fixation du carbone (Tab. 21). Elles incluent cinq isoformes de ribulose-bisphosphate carboxylase (Rubisco), une isoforme chloroplastique de fructose bisphosphate aldolase (FBPase), une triose phosphate isomerase (TPI), une protéine du complexe d'émission de l'oxygène du photosystème II (PsbP), la sous unité β de l'ATP synthase. Toutes ces protéines présentent une expression qui diminue pendant le développement.





Tableau 21 : Spots identifiés correspondant à des protéines impliquées dans la photosynthèse. * Spot identifié à partir de chimères, le spot 897 correspond à une EST de pêche, BU039696. ** Spot identifié par MS/MS. Le spot 659 a été identifié grâce à l'unigène complet PA149TC1. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Especies	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
897*		Triosephosphate isomérase	29	9	gi17839391	Arabidopsis thaliana	NS	28	33
1009		Rubisco	17	7	gi14100327	Theophrasta americana	NS	17	47
957**		Protéine du complexe d'émission de l'O ₂	10	2	gi1226314	Moutarde	99	21	26
405		Rubisco	28	12	gi118477662	Moraea veridickii	95	57	47
490		Rubisco	30	16	gi120257368	Ternstroemia gymnanthera	NS	45	51
982		ATP synthase, sous unité beta	23	8	gi120269422	Lissocarpa guianensis	95	23	46
798		Rubisco	24	10	gi15991250	Acrostichum aureum	NS	33	47
983		Rubisco	15	8	gi19719184	Scadoxus cinnabarinus	NS	21	49
659		Fructose bisphosphate aldolase, chloroplastique	25	9	gi129951422	Prunus Armeniaca EST	95	40	42

IV.5.4.2 Protéines impliquées dans le métabolisme des sucres

Cinq protéines impliquées directement dans le métabolisme des sucres ont été identifiées (Tab. 22), trois isoformes de saccharose synthase, une xylulose kinase et une UDP-glucose pyrophosphorylase. L'abondance exacte de cette dernière n'a pas pu être évaluée correctement, certains spots dans les trois stades les plus précoces correspondaient à de la Rubisco, il semblerait tout de même qu'elle soit induite pendant la maturation, mais ceci reste à confirmer. Les autres protéines sont plus clairement induites. Le niveau des quatre protéines identifiées augmente pendant le développement à partir de la crise climactérique.





Tableau 22 : Spots identifiés correspondant à des protéines impliquées dans le métabolisme des sucres. * Spots identifiés à partir de chimères, les spots 109, 112, 111 correspondent à l'unigène PA349TC1, le spot 589 correspond à l'unigène BGA010G05. Le spot 389 a été identifié grâce à l'unigène complet PA91TC1. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Especies	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
109*		Saccharose synthase	12	10	gil6683114	Citrus unshiu	90	110	92
112*		Saccharose synthase	21	19	gil6683114	Citrus unshiu	90	110	92
111*		Saccharose synthase	14	12	gil6683114	Citrus unshiu	95	110	92
589*		Xylulose kinase	20	8	gil15292819	Arabidopsis thaliana	95	42	61
389		UDP-glucose pyrophosphorylase	29	11	gil29951065	Prunus armeniaca		60	51

IV.5.4.3 Protéines impliquées dans le grandissement cellulaire et la perte de fermeté

Quatre protéines ont été identifiées à ce niveau, deux annexines et deux xyloglucane endotransglycosylases. Toutes présentent une abondance qui augmente pendant le développement de l'abricot et correspondent à des unigènes qui sont apparemment complets (Tab. 23).

Tableau 23 : Spots identifiés correspondant à des protéines impliquées dans le grandissement des cellules et la perte de fermeté. Les spots 1055 et 704 ont été identifiés à partir de l'unigène complet PA1TC15. Les spots 736 et 735 ont été identifiés à partir de l'unigène complet PA17TC1. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Especies	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
1055		Annexine	19	6	gil40715469	Prunus armeniaca EST	NS	13	35
704		Annexine	54	19	gil40715469	Prunus armeniaca EST	95	38	35
736		Xyloglucane endo transglycosylase	19	8	gil29959021	Prunus armeniaca EST	95	36	3
735		Xyloglucane endo transglycosylase	13	5	gil29959021	Prunus armeniaca EST	95	35	32

IV.5.4.4 Protéines impliquées dans le métabolisme et la réponse aux hormones

Cinq protéines impliquées dans le métabolisme (S-adénosyl méthionine synthétase) et la réponse aux hormones (une protéine régulée par l'auxine AUX/IAA, une sérine/thréonine kinase et deux protéines reliées au stress abscissique, ASR) ont été identifiées (Tab. 24). Les

trois premières citées présentent un profil protéique qui augmente pendant la maturation. Les deux ASR sont constitutives.

Tableau 24 : Spots identifiés correspondant à des protéines impliquées dans le métabolisme et la réponse aux hormones. * Spots identifiés à partir de chimères, le spot 908 a été identifié à partir d'une EST de pêche (BU044863), le spot 773 a été identifié à partir de l'unigène BGA010A12. Le spot 474 a été identifié à partir de l'unigène complet PA60TC1. Les spots 750 et 781 ont été identifiés à partir des unigènes complet PA2TC1 et PA2TC7 indistinctement. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Espece s	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
908*		Régulée par l'auxine	16	4	gil20385508	Populus tremula	95	27	27
773*		serine/threonine kinase	15	9	gil3021270	Arabidopsis thaliana	NS	35	71
474		Sam synthétase	30	12	gil40715736	Prunus armeniaca EST	99	42	50
750		Reliée au stress abscissique	56	10	gil2677824	Prunus armeniaca		35	21
781		Reliée au stress abscissique	61	8	gil40715761	Prunus armeniaca EST		34	21

IV.5.4.5 Protéines impliquées dans le métabolisme de l'azote

Trois protéines impliquées dans le métabolisme de l'azote ont été identifiées (Tab. 25). L'inosine monophosphate déshydrogénase, dont l'induction est spécifique de la crise climactérique, l'arginase dont le niveau diminue pendant la maturation et une UMP/CMP kinase *a* induite dans les stades tardifs.

Tableau 25 : Spots identifiés correspondant à des protéines impliquées dans le métabolisme de l'azote. * Spot identifié à partir de chimère, le spot 334 a été identifié par l'unigène PA1614TC1. ** Spot identifié par MS/MS. Le spot 928 a été identifié à partir de l'unigène complet PA130TC1. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'Expression	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Espece s	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
334*		Inosine Mono Phosphate déshydrogenase	26	8	gil26451248	Arabidopsis thaliana	95	66	37
662**		Arginase		2	gil13182957	Brassica napus	95	29	36
928		UMP/CMP kinase a	33	6	gil29953970	Prunus Armeniaca EST	95	23	25

IV.5.4.6 Protéines impliquées dans la défense vis à vis du stress oxydatif

Quatre enzymes impliquées dans le mécanisme de détoxification des formes réactives de l'oxygène ont été identifiées (Tab. 26). Après induction à 65 JAA, le niveau de la catalase et la déshydroascorbate réductase diminue pendant la maturation, avec un léger pic d'induction à la crise climactérique. La 6-phosphogluconate déshydrogénase est spécifique de la crise éthylénique. La superoxyde dismutase est quand à elle constitutive pendant le développement du fruit.








Tableau 26 : Spots identifiés correspondant à des protéines impliquées dans la lutte contre le stress oxydatif. * Protéines identifiées à partir de chimères, le spot 394 a été identifié à partir de l'unigène BGA010I02, le spot 889 a été identifié à partir de l'unigène BGA002B03, le spot 777 a été identifié à partir de l'unigène PA145TC1. ** Spot identifié par MS/MS. Le spot 1115 a été identifié à partir de l'unigène complet PA37TC2. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Espece s	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
394* **		6-phosphogluconate déshydrogenase	18	9	gi2529229	Glycine max	95	56	56
889*		Déshydroascorbate reductase	27	6	gill1761812	Arabidopsis thaliana	90	27	27
777		Catalase	17	9	gi536787	Nicotiana glumbaginifolia	95	35	57
1115		Superoxyde dismutase	40	4	gi51514501	Prunus Armeniaca EST		14	15

IV.5.4.7 Protéines diverses

13 protéines identifiées ne sont pas reliées aux mécanismes précédemment étudiés (Tab. 27). Environ la moitié d'entre elles sont des protéines abondantes et constitutives. Parmi les autres, on trouve 3 protéines inconnues ou putatives, toutes trois induites pendant la maturation.










Tableau 27 : Spots identifiés correspondant à des protéines de catégories fonctionnelles autres ou sans fonction identifiée. * Spot identifiés à partir de chimères, le spot 313 a été identifié à partir de l'unigène PA330TC1, le spot 569 a été identifié à partir de l'unigène PA443TC1, le spot 753 a été identifié à partir de l'unigène PA1382TC1, le spot 1087 a été identifié à partir de l'unigène PA12TC1, le spot 1044 a été identifié à partir de l'unigène PA329TC1, le spot 246 a été identifié à partir de l'unigène PA339TC1. ** Spots identifiés par MS/MS. Le spot 1037 a été identifié à partir de l'unigène complet PA19TC3. Le spot 519 a été identifié à partir de l'unigène complet PA82TC1. Le spot 615 a été identifié à partir de l'unigène complet PA72TC1. Le spot 633 a été identifié à partir de l'unigène complet PA50TC1. Le pourcentage de recouvrement est le rapport entre la taille totale des peptides identifiés et la taille de la protéine totale théorique identifiée. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel. Masse Theo : masse théorique de la protéine dans la base de données.

Spot #	Profil d'abondance	Identification	% recouvrement	#Peptides trouvés	Numero d'Accession	Especies	Pvalue >	Masse Exp. (kda)	Masse Theo. (kda)
29**		Protéine de stockage		1	gil170696	Triticum aestivum	90	100	721
313*		Pyruvate kinase	16	10	gil18409740	Arabidopsis thaliana	NS	58	66
569*		Formate déshydrogenase	36	9	gil6681408	Arabidopsis thaliana	95	42	42
753*		Protéine inconnue	18	6	gil4972058	Arabidopsis thaliana	95	34	35
931		Resistance maladie Pt4	31	6	gil22947662	Citrus grandis	NS	25	20
1037**		Protéine inconnue	23	4	gil29959115	Prunus armeniaca EST	95	18	14
519		Protéine inconnue	41	14	gil29956028	Prunus armeniaca EST	95	45	40
1087*		Choc thermique	40	5	gil3819743	Castanea sativa		14	17
615		glyceraldehyde-3-phosphate déshydrogénase	38	10	gil29955866	Prunus armeniaca EST		40	36
1044*		Agglutinine	21	5	gil2781232	Amaranthus caudatus		14	34
246*		malate déshydrogénase (NADP+)	40	22	gil515759	Vitis vinifera		70	65
224		polyphénol oxydase	32	17	gil3282505	Prunus armeniaca		70	67
633		fructose-bisphosphate aldolase	46	16	gil29955149	Prunus armeniaca EST		38	38

IV.5.4.8 Protéines non identifiées

Onze protéines n'ont pas pu être identifiées dont 9 régulées au cours du développement (Tab. 28).

Tableau 28 : Protéines non identifiées. La Pvalue correspond à la significativité de la différence entre les deux stades extrêmes en terme d'abondance. Masse Exp : masse expérimentale de la protéine sur le gel.

Spot #	Profil d'abondance	Identification	Pvalue >	Masse Experimentale (kda)
118		Non identifié	95	100
756		Non identifié	NS	37
783		Non identifié	95	33
932		Non identifié	NS	23
967		Non identifié	99	25
1001		Non identifié	95	18
126		Non identifié	95	120
780		Non identifié	99	35
883		Non identifié	NS	34
1102		Non identifié		10
1089		Non identifié		14

V Discussion

Ce chapitre a pour but de mettre en commun les observations faites à l'aide d'outils transcriptomiques (EST et microarray) et protéomiques, et de les discuter par voies métaboliques impliquées dans l'élaboration de la qualité du fruit.

V.1 Éthylène

V.1.1 Biosynthèse de l'éthylène

V.1.1.1 Amino-cyclopropane-1-carboxylate synthase (ACS)

Les trois gènes correspondant aux ACC synthases (ACS) isolés par Mbéguié-Mbéguié (2000) n'ont pas été retrouvés dans les banques d'EST mais ils ont cependant été déposés sur les microarrays. Aucun d'entre eux ne présente de variation de niveau d'expression significative pendant le développement. ACS1 (AF184076) et ACS2 (AF184077) sont sur-exprimés chez Moniqui, ACS1 dans les deux stades et ACS2 uniquement au stade 2 (mûr). Le niveau d'expression de ACS1 n'est pas beaucoup plus élevé chez Bergeron que chez Blanc.

Cependant, ces travaux confirment le profil d'ACS1 précédemment observé chez Bergeron et Moniqui (Mbéguié-Mbéguié, 2000), qui semble induit à la crise éthylénique. Le profil d'expression d'ACS2 chez Bergeron semble être identique à celui observé chez Moniqui par Mbéguié-Mbéguié (2000) avec une induction post crise éthylénique. L'expression de ce gène n'a pas pu être détectée lors des travaux antérieurs chez Bergeron.

Dans l'expérience de différenciation variétale, la sur-expression d'ACS1 dès la crise l'éthylénique chez Moniqui, est beaucoup plus précoce que lors de l'étude précédente puisqu'elle n'a été observée que dans des stades mûrs, cependant les techniques de mesure sont différentes (PCR quantitative vs microarray).

L'expression de ACS3 (AF184078) ne montre pas de différence significative entre les trois variétés.

D'après les résultats sur le différentiel variétal, le caractère non climactérique de la variété Blanc ne peut pas être attribué à l'absence d'une des trois ACS.

D'autre part, la production d'ACS1 semble corrélée avec le dégagement d'éthylène et son intensité lorsque celui-ci a lieu ; ACS2 est induit nettement plus tardivement. L'hypothèse

de Mbéguié-Mbéguié (2000) selon laquelle ACS2 serait induit par des facteurs différents de l'éthylène semble aussi être envisageable pour Bergeron puisque que l'orthologue de tomate le plus proche d'ACS2 est induit par blessure (Tatsuki et Mori, 1999).

V.1.1.2 Amino-cyclopropane-1-carboxylate oxydase (ACO)

Le niveau de l'expression de l'unigène PA9TC1 correspondant au transcrit déjà identifié chez l'abricot (AF026793, Mbéguié-Mbéguié, 2000) augmente au cours du développement, conjointement avec le dégagement d'éthylène (observations par Northern électronique et microarray). Contrairement à Mbéguié-Mbéguié (2000) aucune différence n'a été observée entre les variétés. Cependant, l'hybridation présente une forte variabilité expérimentale dans l'expérience de différenciation variétale. Il existe dans la base de données un variant d'épissage à cet unigène (BGH014I09) plus court pour lequel il n'a pas été possible de construire une sonde spécifique (annexe B (CD), famille multigénique N181).

V.1.1.3 La protéine E8

L'homologue de la protéine E8 de tomate (identifié précédemment par Mbéguié-Mbéguié (2000), AF13950) impliquée dans le rétrocontrôle négatif de l'expression des gènes ACS et ACO (Kneissl et Deikman, 1996), ne montre pas d'évolution significative au cours du développement. Bien que non-significative, l'induction va tout de même dans le sens de l'augmentation des transcrits obtenue par Mbéguié-Mbéguié en Northern blot (2000).

La répression de ce gène chez Monique reste cohérente avec les observations précédentes, et il apparaît ici qu'il serait sur-exprimé chez la variété Blanc, ce qui permet de penser qu'il pourrait être impliqué dans le caractère non climactérique de cette variété.

Ce gène présente une forte homologie avec les protéines de la famille des dioxygénases dont fait partie l'ACO (Prescott et John, 1996) et pourrait donc entrer en compétition avec l'ACO vis à vis du substrat ACC.

V.1.1.4 Autres membres de la famille multigénique des dioxygénases

La famille multigénique protéique 11 (annexe B (CD), base de données) regroupe des unigènes dont les 8 membres posséderaient une activité dioxygénase. On retrouve dans cette famille les ACO, mais aussi des isoformes de leucoanthocyanidine dioxygénase parmi d'autres dioxygénase de spécificité de substrat inconnue. La protéine E8 appartient aussi à cette famille. Parmi les dioxygénases de fonctions inconnues, certaines ont été observées

comme différentiellement exprimées dans des conditions qui permettent de supposer qu'elles ont une expression éthylène-dépendante.

Ainsi, l'unigène PA322TC1 présente un niveau d'expression qui diminue au cours du développement du fruit, en revanche, il est sur-exprimé chez Monique en stade mûr. Son extrémité C terminale intègre un motif référencé sur la base de données PFAM (PB007475) commun à 11 protéines, mais dont la fonction est inconnue. Étant donné son homologie avec E8 et ACO, sa sur-induction chez Monique, et sa répression pendant le développement de Bergeron, ce gène présente des signes de contrôle positif sur la production d'éthylène. Il est probable qu'il joue un rôle plus important dans le fruit mûrissant chez Monique que chez Bergeron, dans lequel son expression pourrait être uniquement liée au développement précoce.

Le niveau d'expression de PA1633TC1 chute fortement aux environs de la crise éthylénique, il est particulièrement inhibé chez Blanc. Il est probable qu'il n'intervienne pas dans la biosynthèse d'éthylène au cours de la maturation.

V.1.1.5 S-adenosylméthionine synthétase (SAMS)

La SAMS est une enzyme clé du métabolisme des plantes, catalysant la biosynthèse de S-adenosylméthionine (SAM) à partir de méthionine et d'ATP (Horikawa *et al.*, 1990). La SAM est un donneur de méthyle ubiquiste, impliqué dans la biosynthèse de l'éthylène. Cinq isoformes différentes de SAMS ont été identifiées parmi les EST, le profil d'expression d'une seule d'entre elles évolue pendant la maturation (PA60TC1). L'abondance des transcrits correspondant à cette SAMS augmenterait conjointement à l'abondance des protéines, pendant la maturation. Les membres de cette famille multigénique sont très proches (annexe B (CD) famille multigénique N 2). Les informations obtenues sur la séquence protéique ne permettent pas d'affirmer si la protéine observée en protéomique correspondait à la traduction de PA60TC1 ou à celle de l'un de deux autres isogènes dont l'expression est constitutive et qui serait alors soumise à une importante régulation post- transcriptionnelle.

Bien que les séquences de SAMS soient fortement conservées intra- et inter-espèces, des isoformes ont pu être identifiées comme différentiellement régulées à la fin du pic d'éthylène dans le kiwi (Whittaker *et al.*, 1997). Cependant, les différences entre les séquences de SAMS d'abricot et de kiwi ne permettent pas d'effectuer une corrélation phylogénique entre les isoformes induites.

La SAMS est plus fortement exprimée chez Bergeron que chez Moniqui au stade crise éthylénique. Chez d'autres espèces, la SAMS n'a jamais été observée comme étant le facteur limitant dans la production d'éthylène (Pandey *et al.*, 2000).

La 5-méthyltétrahydroptéroyltriglutamate-homocystéine S-méthyltransférase (PA34TC1) qui catalyse la réaction précédente (synthèse de la L-méthionine) dans la voie de biosynthèse de la SAM, voit son profil d'expression augmenter aussi pendant la maturation. Il est donc probable que la quantité de méthionine augmente pendant la maturation, mais son devenir n'est probablement pas réservé à la synthèse de l'éthylène.

V.1.1.6 Protéine kinases calcium dépendantes

Il a été montré que l'ACS2 de tomate possède un site de phosphorylation intervenant dans la régulation post-traductionnelle des ACS et dans la stabilité de la protéine (Tatsuki et Mori, 2001). Plus récemment, Sebastia *et al.* (2004) ont identifié des membres d'une famille multigénique de protéines kinases dépendantes du calcium (CDPK), capables de phosphoryler *in vitro* l'ACS sur ce site. Les trois isoformes testées étaient toutes capables de phosphoryler l'ACS et il n'existe aucune indication expérimentale en faveur de la spécificité des isoformes vis à vis de cette réaction *in vivo*.

Il existe dans la bases de données d'EST d'abricot 18 protéines homologues à des CDPK regroupées en trois familles multigéniques protéiques (annexe B (CD), base de données, 12, 37, 335). Certaines d'entre elles sont différentiellement exprimées pendant le développement ou entre les variétés et peuvent donc être potentiellement impliquées dans la régulation de l'activité ACS : BGE003L14 qui est apparue induite pendant le développement au travers de l'expérience de microarray, et PA1498TC1 qui est spécifique de Bergeron. Il est cependant envisageable que ces unigènes possèdent un autre rôle.

V.1.2 Transduction du signal éthylène

V.1.2.1 Récepteurs d'éthylène

Parmi les trois récepteurs d'éthylène identifiés dans la base de données EST, un seul présente une variation du niveau d'expression significative pendant le développement. Le niveau d'expression de BGH004I19 est inhibé fortement au stade vert mature, et on observe une augmentation de son niveau d'expression à partir de la crise éthylénique. LeETR5 qui est

le plus proche homologue de BGH004I19 identifié chez la tomate, présente un profil d'expression identique (Tieman et Klee, 1999). Comme LeETR5 appartient à la famille des récepteurs de type 2 dont le domaine histidine kinase n'est pas fonctionnel, il est envisageable que ce soit aussi le cas pour BGH004I19. Sous-exprimé chez la variété Blanc au stade mûr, son profil d'expression devra être caractérisé tout au long du développement chez cette variété également, car il est probable que son expression diminue pendant le développement, contrairement à Bergeron (et probablement Moniqui).

Deux caractéristiques additionnelles de LeETR5, permettent de supposer que BGH004I19 serait impliqué dans le caractère non climactérique de la variété Blanc : la forme mutée de LeETR5 confère une insensibilité dominante à l'éthylène dans les transgènes d'*Arabidopsis*. Dans la tomate, il est exprimé préférentiellement dans les organes reproductifs : ce qui expliquerait que les différences majeures de phénotype entre Bergeron et Blanc soit restreintes au fruit, et ne touchent pas l'arbre.

L'homologue de LeETR1 (PA615TC1), déjà caractérisé chez la pêche (Basset *et al.*, 2002), a été déposé sur les lames mais aucun signal d'hybridation n'a été obtenu.

Une gène relatif à Pc-ERS1a (BGE002E21) obtenu chez la poire (El-Sharkawy *et al.*, 2003) est non significativement induit, alors qu'il l'est plus clairement chez la poire.

V.1.2.2 Facteurs de réponse à l'éthylène (ERF)

Au total, 12 séquences indépendantes sont homologues à des séquences contenant le domaine AP2. Ce domaine est relativement proche du domaine retrouvé dans les facteurs de réponse à l'éthylène (ERF) qui permet la liaison à la GCC box du promoteur des transcrits induits par l'éthylène. Le domaine commun de liaison à l'ADN des ERF a été séquencé sur 4 unigènes qui ont tous été déposés sur les lames. Les 8 autres sont seulement homologues à des protéines contenant un domaine AP2 sans que celui-ci ait été séquencé dans les EST. Le séquençage incomplet des extrémités 5' de ces trois unigènes ne permet pas d'observer leur domaine putatif ERF ou AP2. Cinq d'entre eux ont été déposés sur les lames.

Les quatre unigènes pour lesquels nous disposons du domaine ERF ont été comparés au niveau de ce domaine aux 4 protéines ERF précédemment identifiées chez la tomate (Fig. 69) (Tournier *et al.*, 2003). Il faut noter qu'en dehors du domaine, les 8 protéines sont fortement divergentes.

L'un d'entre eux était déjà connu chez l'abricot : AF071893 (Mbéguié-Mbéguié, 2000) ou PA835TC1 dans les EST. Pour cette raison la séquence antérieure a été utilisée pour la

construction de la sonde, cependant l'oligonucléotide a été sélectionné dans la région codante car la séquence n'était pas complète. Son niveau d'expression augmente linéairement pendant la maturation. PA835TC1 est fortement homologue à LeERF2 dans le domaine ERF (Fig. 69), et relativement conservé sur le reste de la séquence. On retrouve chez les deux protéines, le motif MCGGAI/L dans leur région N terminale spécifique des ERF de classe IV (Tournier *et al.*, 2003). Elles partagent toutes deux la particularité d'être exprimées préférentiellement dans les fruits mûrs.

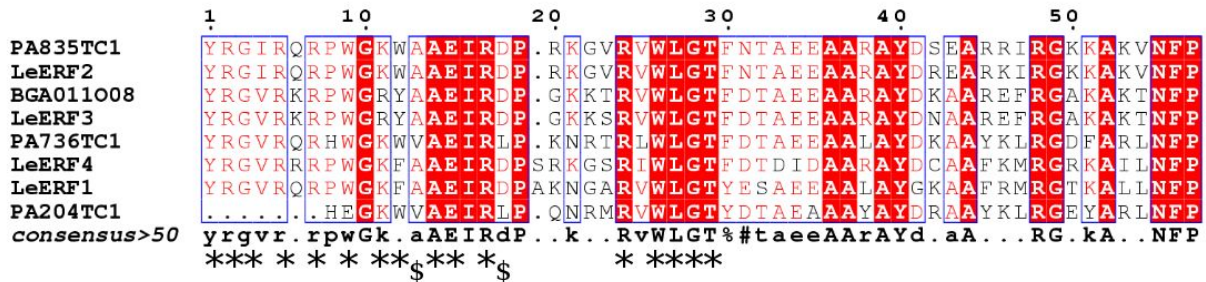


Figure 69 : Alignement multiple du domaine ERF des séquences protéiques conceptuelles des gènes codant pour des facteurs de réponse à l'éthylène d'abricot et de tomate contenant ce domaine. % correspond à un acide aminé aromatique, # correspond à un acide aminé chargé, * acide aminé interagissant avec l'ADN, \$ acide aminé différent entre les ERF et les DREB. LeERF1 correspond à l'accension genbank AAO34703 ; LeERF2 correspond à l'accension EMBL AAO34704 ; LeERF3 correspond à l'accension EMBL AAO34705 ; LeERF4 correspond à l'accension EMBL AAO34706 (Tournier *et al.*, 2003).

BGA011008 est homologue à LeERF3. Chaque séquence contient le motif ERF-associated amphiphilic repression (EAR) retrouvé à l'extrémité N terminale des ERF de classe II (Ohta *et al.*, 2001 ; Fujimoto *et al.*, 2000). Contrairement à LeERF3, dont l'expression n'a pas été détectée dans le fruit, BGA011008 est induit dans le stade le plus vert, puis réprimé par la suite. Le niveau d'expression de BGA011008 est cependant beaucoup plus fort dans la variété Monique. Il conviendrait de déterminer son évolution au cours du développement dans cette variété, afin de tester l'hypothèse d'une corrélation entre l'expression de ce gène, la précocité et la surproduction d'éthylène spécifique de Monique.

En analysant de façon détaillée le motif ERF (Fig. 69), il semble que PA204TC1 et PA736TC1 correspondent plutôt à des DREB (Drought response element binding), protéines induites par la sécheresse, le froid ou le sel et présentant un mode d'action identique à celui des ERF (Sakuma *et al.*, 2002), en se liant aux promoteurs de gènes sur un motif A/GCCGAC (à la place de la GCC box pour les ERF).

PA204TC1 et PA736TC1 possèdent des parties C terminales plus longues que les ERF classées et ne contiennent pas les deux acides aminés spécifiques des ERF les différenciant

des DREB. Ils ne semblent donc pas appartenir à une classe d'ERF, de plus il leur manque la partie N terminale.

Les séquences possèdent cependant un motif, encore inconnu, commun avec 3 protéines (Fig. 70) dont deux sont assignées comme des DREB, 21 bases après le motif ERF (Domaine inconnu 1, Fig. 70). Tous les acides aminés indispensables à la liaison avec l'ADN (Allen *et al.*, 1998) sont cependant conservés dans cette possible nouvelle classe. Ce motif est référencé sur le site PFAM PB052400 et inclut également 3 séquences d'autres espèces non représentées ici. Il existe un second motif à l'extrémité C terminale (Domaine inconnu 2, Fig. 70) référencé PB021635 sur PFAM, retrouvé sur 2 séquences supplémentaires. Pour l'instant, la fonction de ce motif, est encore inconnue mais il semblerait être spécifique des DREB. Toutes les protéines appartiendraient à la famille A6 de DREB dont le facteur régulateur, la fonction et le motif de liaison à l'ADN sont inconnus (Sakuma *et al.*, 2002). Cependant, les membres de cette famille sont les seules DREB qui possèdent une Leucine à l'emplacement du second acide aminé, qui permet de différencier les DREB des ERF. De plus, bien qu'elles ont été annotées comme des DREB par Sakuma *et al.* (2002), dans l'arbre phylogénique réalisé par ces mêmes auteurs, elles sont pratiquement équidistantes des DREB et des ERF. Tant que l'interaction d'un membre de cette famille avec le motif de liaison aux DREB n'a pas été démontré, il n'est pas à exclure que ces protéines possèdent une fonction différente.

Cependant les deux gènes d'abricot ne présentent pas le même profil au cours du développement, l'un est constitutif, (PA204TC1) et le second (PA736TC1) est réprimé bien avant la crise éthylénique.

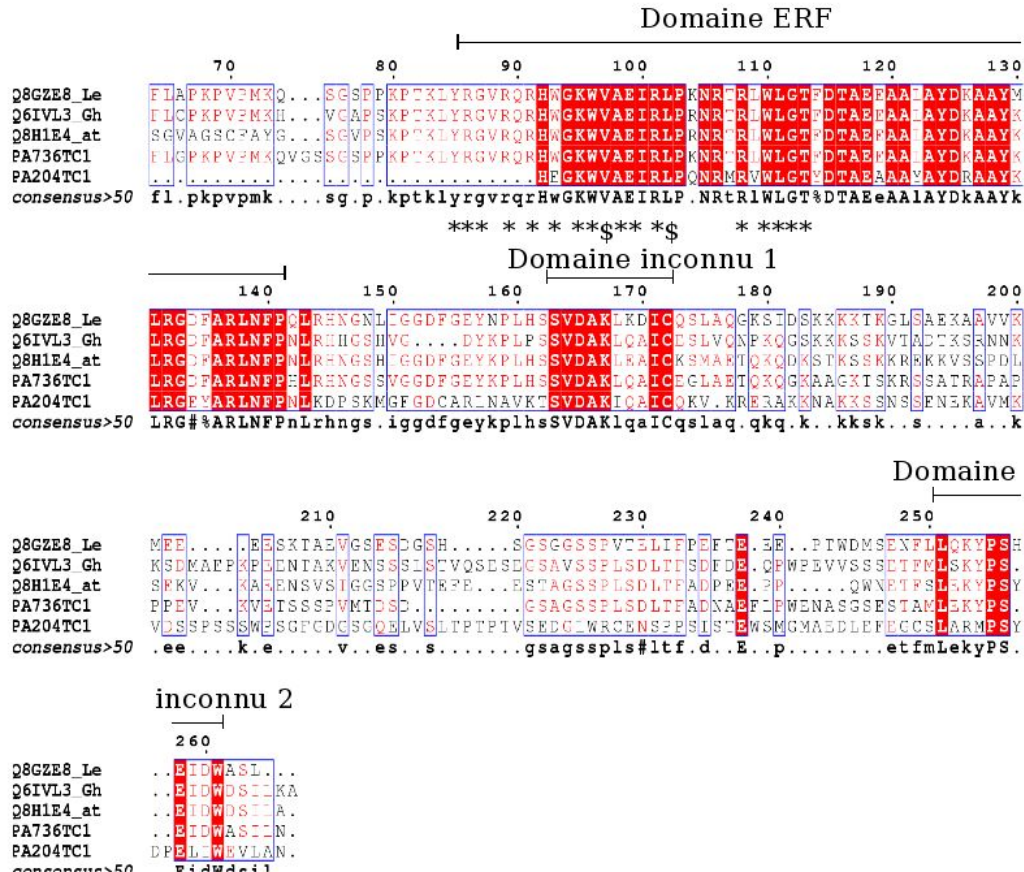


Figure 70 : Alignement multiple des protéines appartenant à la nouvelle classe A6 de DREB. % correspond à un acide aminé aromatique, # correspond à un acide aminé chargé, * acide aminé interagissant avec l'ADN, \$ acide aminé différent entre les ERF et les DREB. Q8H1E4 correspond à l'accension EMBL AAN12993 d'Arabidopsis thaliana (Yamada et al., 2002, non publié). Q8GZE8 correspond à l'accension EMBL AF506825 de tomate (Chen et Hou, 2002, non publié). Q6IVL3 correspond à l'accension EMBL AY619718 de coton (Huang et Liu, 2004, non publié)

Il y a peu de preuves pour que l'une des 4 protéines, chez lesquelles un motif ERF a été identifié, réponde à l'éthylène, deux d'entre elles présentent un profil d'expression non relié au dégagement d'éthylène, une autre est constitutive dans le fruit, bien que PA835TC1 possède un profil d'expression concomitant avec le dégagement d'éthylène, son homologue chez la tomate ne répond pas à l'éthylène. En revanche, l'induction de BGA011O08 chez Monique mérite une attention particulière. Une étude phylogénique plus approfondie des deux unigènes identifiés comme étant des DREB reste nécessaire, les investigations n'ayant pu être poussées dans ce sens faute de temps dans cette thèse.

Trois des 5 nouveaux unigènes identifiés comme appartenant potentiellement à la famille des ERF/AP2 et déposés sur les lames de microarray présentent une expression différentielle. PA1510TC1 est réprimé dans le stade précédant la crise éthylénique et BGE014K05 est réprimé linéairement au cours du développement, alors que le niveau d'expression de PA1386TC1 reste constitutif pendant le développement du fruit. Les trois transcrits sont sur-

exprimés chez Monique, BGE014K05 l'est également chez Blanc. PA1510TC1 est homologue à une protéine de tomate induite pendant la maturation (Kannan et Giovannoni, non publié, cité par Giovannoni *et al.*, 1999) dont le gène co-localise proche de la mutation *rin* (ne mûrit pas par addition d'éthylène exogène) chez la tomate (Giovannoni *et al.*, 1999). Ce gène pourrait donc être impliqué dans la réponse à l'éthylène, cependant, il n'est pas en cause dans le caractère non climactérique de la variété Blanc puisque son profil d'expression transitoire est identique à celui de Bergeron. La séquence de l'ADNc exprimé dans la variété Blanc devra être vérifiée car une mutation pourrait cependant avoir rendu ce gène non-fonctionnel.

V.1.3 Autres gènes potentiellement reliés au signal éthylène

BGA005A18 correspond à une protéine, reliée à l'éthylène, obtenue par differential display chez la tomate (Zegzouti *et al.*, 1999), nommée ER33 mais dont la fonction reste à préciser. Son profil d'expression n'a pas pu être clairement défini, il semblerait qu'elle soit induite, comme chez la tomate, mais il y a trop de variabilité expérimentale pour l'affirmer.

Un homologue à l'un des CTR de tomate, qui se lie aux récepteurs d'éthylène, a été identifié, BGA006B03. Malheureusement, il est homologue à LeCTR2 (Acc AJ318955) qui n'est pas relié à la signalétique éthylène chez la tomate (Adams-Phillips *et al.*, 2004b), et ne présente pas d'évolution au cours de la maturation de ce fruit.

PA84TC1 est homologue à une protéine de stress universelle (UPS), impliquée dans la réponse au stress transmise *via* l'éthylène chez le riz (Sauter *et al.*, 2002). Les microarrays et le Northern électronique indiquent que sa transcription est réprimée au cours du développement de l'abricot. Elle est sur-exprimée chez Bergeron par rapport aux deux autres variétés. Un second isogène est aussi réprimé au cours du développement (PA164TC1). Contrairement à PA84TC1, il est réprimé dès le stade vert mature et ne présente pas de différence d'expression dans les trois variétés. Un troisième isogène (BGE004M12) est induit pendant le développement. Ces protéines présentent quelques homologies avec ER6 (Zegzouti *et al.*, 1999).

PA498TC1, homologue à la protéine de tomate ER43 (Zegzouti *et al.*, 1999), qui correspond à une petite protéine rab qui lie le GTP, est constitutive chez l'abricot, contrairement à la tomate chez laquelle elle est réprimée pendant le développement.

PA1627TC1, correspond à l'accession AF071894 d'abricot, homologue aux transcrits de tomate ER5 (Zegzouti *et al.*, 1999). Elle est induite pendant le développement des deux fruits, cependant elle ne semble pas avoir un profil éthylène dépendant chez l'abricot. PA1627TC1 et ER5 sont homologues à des protéines de la famille LEA (abondant à la fin de l'embryogénèse). Chez la tomate, ER5 est aussi régulé par la blessure, l'acide abscissique et la sécheresse (Zegzouti *et al.*, 1999).

Deux fragments correspondant aux extrémité 5' (BGH001J06Q) et 3' (PA1581TC1) de EIN3 ont été retrouvés dans les banques d'EST mais n'ont pas été déposés sur les lames.

V.1.4 Conclusion

L'étude de l'expression des gènes potentiellement reliés à la biosynthèse et au signal éthylénique avait pour but d'apporter des éléments de réponse à deux questions principales : quels sont les gènes responsables des caractères de maturation précoce, tardive, ou non climactériques des différentes variétés d'abricot et quels sont les gènes (ou isogènes) impliqués dans la voie éthylène pendant la maturation.

Parmi les gènes étudiés, deux candidats sérieux ont été identifiés pour amorcer une explication sur le caractère non climactérique de la variété Blanc : le gène codant pour la protéine E8 (AF13950) qui est sur-exprimé dans cette variété, et le récepteur BGH004I19, qui est sous-exprimé. Compte tenu de leur position respective dans la cascade de transduction du signal éthylène, BGH004I19 semble être l'élément clé. La caractérisation de la quantité de protéines de EIN3 (BGH001J06Q et PA1581TC1) permettrait d'apporter une réponse quand à l'état d'activation de la cascade éthylène dans cette variété.

Bien que plusieurs gènes soient sur-exprimés chez Monique, leur implication dans les caractères précoce et sur-productif en éthylène n'est pas très évidente. Seule la sur-expression dès la crise éthylénique de ACS1 (AF184076) chez Monique, est assez clairement liée à la surproduction d'éthylène. La signification de la sur-expression (i) d'une dioxygénase, dont le lien avec l'oxydation de l'ACC n'est pas fait, (ii) de l'ERF BGA011O8, et (iii) des unigènes PA1510TC1, BGE014K05, PA1386TC1 codant pour des protéines possédant peut-être un domaine AP2/ERF, n'est pas établie.

Parmi les gènes impliqués dans la biosynthèse de l'éthylène pendant la maturation, nos travaux confirment ceux de Mbéguié (2000), ACS1 et l'ACO AF026793 semblent être les

isoformes impliquées. Il serait nécessaire d'évaluer si la dioxygénase PA322TC1 et la CDPK BGE003L14 jouent aussi un rôle dans la biosynthèse et le contrôle de l'éthylène pendant la maturation. Au niveau de la transduction du signal, il serait nécessaire d'évaluer le rôle de l'ERF PA835TC1 et de la protéine de réponse au stress BGE004M12.

V.2 Métabolisme, régulation et perception des autres hormones

V.2.1 Auxine

Parmi les unigènes codant pour des enzymes potentiellement impliquées dans la biosynthèse de l'auxine chez les plantes (Bartel, 1997), seulement 3 ont été identifiés dans cette étude : deux monooxygénases constitutives et une nitrilase qui n'a pas été déposée sur les lames. En revanche 11 unigènes impliqués dans la réponse à l'auxine évoluent pendant le développement : 8 sont différentiellement exprimés entre les variétés, deux d'entre eux sont à la fois induits pendant le développement et différentiellement exprimés entre les variétés. Une protéine (spot 908) de ce type est aussi induite pendant le développement ; elle est homologue à une EST de pêche traduite, l'unigène correspondant d'abricot n'a pas été déposé sur les lames.

V.2.1.1 AUX/IAA

Dix unigènes d'abricot appartenant à la même famille multigénique protéique (famille 6) correspondent à des AUX/IAA. Les membres de cette famille ont un rôle de régulation dans l'expression des gènes dont l'expression est auxine-dépendante (Liscum et Reed, 2002) ; il sont eux-même induits par l'auxine (Abel *et al.*, 1995).

Cinq AUX/IAA sont induites pendant la maturation (PA790TC1, PA373TC1, BGE017N12, PA196TC1, PA173TC1) ; elles sont toutes apparemment co-régulées, avec une forte augmentation (en moyenne d'un facteur 8) entre les stades vert et vert mature, qui se prolonge moins rapidement jusqu'au stade mi-mûr, et stagne ensuite, sauf pour PA790TC1 et PA373TC1 dont l'expression diminue après le stade mi-mûr. La protéine identifiée (spot 903) suit le même profil ; elle correspond à une EST de pêche (BU044863) homologue à un sixième transcrit à 83 %, PA210TC1, non déposé sur les lames. En ce qui concerne les quatre

autres transcrits, BGA009P06 n'est pas significativement différentiellement exprimé (il n'est pas présenté dans la figure 71, mais il est relativement proche de BGE017N12) ; BGE002L23 n'a pas été déposé sur les lames, PA803TC1 ne présente pas de signal et PA799TC1 est réprimé pendant le développement. Trois d'entre eux sont différemment exprimés dans les variétés : BGE017N12 est sur-exprimé chez Monique au stade mûr par rapport à Bergeron, PA373TC1 est spécifique de Bergeron et BGA009P06 est sous-exprimé chez Bergeron.

Afin d'anticiper une fonction probable des AUX/IAA chez l'abricot, les traductions conceptuelles des 10 unigènes ont été alignées avec les protéines de la même famille chez *Arabidopsis* (Fig. 71). Cependant il est difficile de relier des orthologues clairement. En effet, les AUX/IAA se situent dans des régions du génome d'*Arabidopsis* où ont lieu des duplications à grande échelle et récentes (Liscum et Reed, 2002). Il n'est possible de dégager des orthologues potentiels que pour peu de séquences d'abricot, seuls PA373TC1 et BGE002L23 peuvent être associés à IAA15 et IAA33 respectivement (Fig. 71), à propos desquels rien n'est connu. Les autres unigènes d'abricot ont plus d'un homologue proche. On observe que tous les unigènes d'abricot, excepté PA790TC1, sont retrouvés dans une des trois branches principales (Fig. 71) et ne sont donc homologues qu'à 16 des 29 AUX/IAA d'*Arabidopsis*. Les associations structure/fonction des protéines de type AUX/IAA seront probablement éclaircies pour un plus grand nombre d'entre elles dans les prochaines années.

Des séquences de AUX/IAA ont été identifiées chez la tomate (Balbi et Lomax, 2003). Cependant les séquences sont partielles (Jones *et al.*, 2002) et il manque notamment la région C terminale des protéines conceptuelles, ce qui empêche le repérage des orthologues potentiels avec les unigènes d'abricot. Quoiqu'il en soit, cette étude sur la tomate (Jones *et al.*, 2002) a montré la régulation différentielle des différents membres de la famille dans le fruit et le rôle des AUX/IAA, affectant la taille des fruits et le nombre de graines. Il en ressort également une possible régulation de la transcription des ACS par les AUX/IAA, car l'expression des ACS et des AUX/IAA est modifiée dans le mutant *dgt* insensible à l'auxine. Quatre AUX/IAA de tomate sont aussi induits pendant la maturation (Jones *et al.*, 2002), mais dans les dernières étapes leur niveau d'expression est indétectable et leur expression est inhibée par l'éthylène. Par comparaison avec l'abricot, cette répression se situe à partir d'un stade intermédiaire entre le stade mi-mûr et le stade mûr et ce profil d'expression semble correspondre à celui observé pour PA790TC1 et PA373TC1, pour lesquels une répression a été observée au stade mûr.

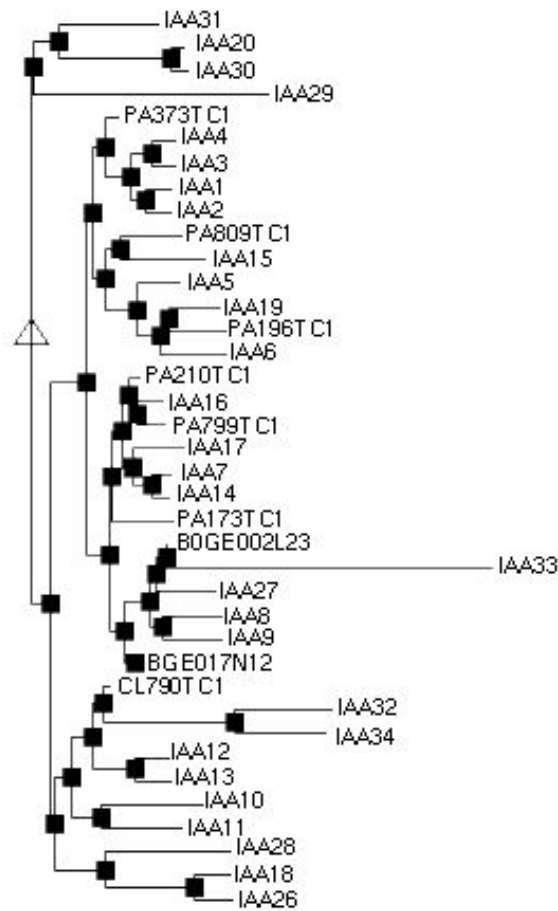


Figure 71 : Arbre phylogénique des protéines d'*Arabidopsis* et des traductions conceptuelles des gènes d'abricot correspondant à des AUX/IAA. Les séquences des IAA d'*Arabidopsis* sont issues de Liscum et Reed (2002).

V.2.1.2 Facteurs de réponse à l'auxine

Les niveaux d'expression de deux facteurs de réponse à l'auxine (ARF) évoluent pendant la maturation, l'un augmente (BGH04I20), alors que l'autre diminue (BGA010O24). Les ARF sont des répresseurs ou activateurs transcriptionnels qui se lient au site AuxREs contenant le motif TGTCTC dans les régions promotrices des gènes de réponse à l'auxine (Guilfoyle et Hagen, 2001). Les AUX/IAA contiennent cette région (Reed, 2001). Les ARF sont cependant exprimés indépendamment de l'auxine (Tiwari *et al.*, 2003). BGH04I20 est homologue seulement à ARF17 d'*Arabidopsis*, qui est assez éloigné phylogéniquement des autres ARF (Liscum et Reed, 2002) et ne possède pas le domaine C terminal propre aux ARF (Hagen et Guilfoyle, 2002); la fonction de ce gène n'a pas encore été caractérisée. BGA010O24 est homologue à ARF1 et ARF2, qui sont très proches. Un troisième unigène

(PA697TC1) correspondant à un ARF semble être induit dans la variété Blanc. Chez les fruits un ARF, nommé DR12, a pour l'instant été étudié chez la tomate (Jones *et al.*, 2002). Cependant aucun des unigènes d'abricot n'est homologue à DR12. Comme chez *Arabidopsis*, son niveau d'expression semble être indépendant de l'auxine, et il est induit pendant la maturation, mais il est réprimé par application d'éthylène exogène.

V.2.1.3 Autres gènes régulés par l'auxine

Parmi les autres familles de gènes connus pour être régulés par l'auxine, deux unigènes correspondant à des SAUR (petits ARN induit par l'auxine) ont été retrouvés dans la base de données EST. Mais ils n'ont pas été déposés sur les lames (BGA011L18 et BGE018B23) tout comme les 4 unigènes correspondant à des GH3 (familles de protéines cytoplasmique de fonction inconnue induites par l'auxine) (PA176TC1, BGE002I02, BGE017F23, BGE001J19).

PA354TC1, réprimée dans la variété Blanc, correspond à une quinone oxydoréductase, induite par l'auxine chez *Arabidopsis*, cette protéine protégerait les cellules contre le stress oxydatif en détoxifiant les semiquinones (Laskowski *et al.*, 2002). Cet unigène appartient à une famille multigénique dont les membres sont différemment régulés, puisque l'on retrouve PA529TC1 induit dans la variété Blanc. L'isogène BGA010K22 n'est pas différentiellement exprimé entre les variétés.

V.2.1.4 Conclusion

Un certain nombre de gènes, potentiellement impliqués dans la réponse à l'auxine et dont l'expression de certains varie pendant le développement, a été identifié dans ce travail. La fonction de ces gènes pendant la maturation reste à mettre en évidence. D'après les travaux réalisés chez la tomate (Jones *et al.*, 2002 ; Balbi et Lomax, 2003), il semblerait qu'il existe un lien étroit entre leur expression et le dégagement d'éthylène. Les AUX/IAA et ARF pourrait aussi jouer un rôle dans la taille des fruits, ainsi que dans la composition en pigments, principalement au niveau de la dégradation des chlorophylles (Jones *et al.*, 2002).

V.2.2 Cytokines

Peu d'éléments de la voie de transduction du signal des cytokines ont été mis en évidence à travers les banques d'EST ; un récepteur de cytokines (PA1575TC1), qui n'a pas été déposé sur les lames, et une protéine de phosphotransfert à histidine (BGA009P21) qui est constitutive. Aucun équivalent des ARR (*Arabidopsis* Response Regulator) d'*Arabidopsis* n'a été mis en évidence (voir Hutchison et Kieber, 2002 pour revue).

V.2.3 Méthyl jasmonate

V.2.3.1 Biosynthèse

Au moins un isogène codant pour chaque enzyme de la voie de biosynthèse du jasmonate (Cheong et Choi, 2003) et de son dérivé le méthyl jasmonate ont été retrouvés dans la base de données EST et la plupart a été déposée sur les lames de microarray (excepté pour l'allène oxyde cyclase, PA541TC1 PA1366TC1).

Quatre unigènes sont homologues à des lipoxygénases, la première enzyme de la voie de synthèse du méthyl jasmonate, réalisée dans le chloroplaste. PA456TC1 est homologue à une lipooxygénase de tomate fruit spécifique (LOXC, Heitz *et al.*, 1997). Cependant bien qu'elle soit chloroplastique, il n'y a pas de preuve qu'elle soit reliée au métabolisme du jasmonate chez la tomate. Elle semble être réprimée pendant le développement de l'abricot d'un facteur 10. Cependant, la variation de son niveau d'expression n'est pas significative car aucun signal n'a été détecté sur un des deux réplicats de la lame de microarray, un problème de dépôt semble quasi évident. Le niveau d'expression de PA1427TC1 augmente pendant le développement et présente un profil d'expression similaire à un de ses homologues chez la tomate avec un maximum d'expression au stade mi-mûr. Bien que le profil d'expression de PA1427TC1 soit synchrone à la production de jasmonate dans la pomme (Fan *et al.*, 1998), espèce la plus proche de l'abricot pour laquelle l'évolution du jasmonate au cours de la maturation est connue, il n'y a pas d'évidence pour qu'elle soit reliée au métabolisme du jasmonate, car aucun orthologue n'a été identifié comme chloroplastique. BGA010C19 est fortement homologue (81 %) à la lipooxygénase LOXD induite par la blessure et le méthyl jasmonate chez la tomate, et elle est *a priori* chloroplastique (Heitz *et al.*, 1997). L'expression diminue au cours du développement de l'abricot entre le stade vert et le stade vert mature puis

se stabilise. BGE003O12 diminue linéairement pendant la maturation, sa séquence est relativement proche de celle de PA1427TC1.

Le profil d'expression de l'allène oxyde synthase (AOS), déposée sur les lames de microarray, BGA007L04 est identique à celle de BGA010C19, la lipoxygénase potentiellement impliquée dans la synthèse du jasmonate. Le profil d'expression d'un homologue chez la tomate est similaire, les transcrits ont été à peine détectés dans les fruits verts puis disparaissent pendant la maturation (Howe *et al.*, 2000).

Le seul unigène identifié comme une oxophytodiénoate réductase, PA446TC1, présente un niveau d'expression qui diminue aussi au cours du développement. A la différence de l'AOS et de LOXD, on observe une seconde phase de diminution entre le stade mi-mûr et le stade mûr. Tout comme les deux transcrits précédents, son plus proche homologue chez la tomate code pour l'isoenzyme impliquée dans la biosynthèse du jasmonate (Schaller *et al.*, 2000 ; Strassner *et al.*, 2002). Contrairement aux enzymes précédentes, elle serait à priori, péroxysomale (Strassner *et al.*, 2002), en revanche, elle est induite en concomitance avec LOXD et AOS en réponse à la blessure dans les feuilles. Cependant son expression n'a pas été évaluée dans le fruit de tomate.

Un transcrit codant pour une jasmonate carboxyl méthyltransférase (BGE014G10) a été identifié dans la base de données EST. Son expression décroît légèrement de façon non significative. Cette enzyme n'a pour l'instant été identifiée que chez *Arabidopsis* (Seo *et al.*, 2001). Des transcrits homologues à la séquence nucléique d'abricot, ont été retrouvés chez la vigne et la pomme mais pas chez la tomate. Son profil d'expression n'est pas encore connu dans le fruit. Cependant ce type de transcrit est homologue à la plupart des méthyltransférases, et il serait nécessaire de mesurer la spécificité de substrat pour confirmer son rôle.

Sur les cinq enzymes impliquées dans la biosynthèse du jasmonate et du méthy jasmonate, les profils d'expression de 4 des transcrits correspondant ont été évalués. Leur expression diminue au cours du développement et ne serait pas synchronisée à la production de jasmonate et de méthy jasmonate observée chez la pomme (Fan *et al.*, 1998) ; cependant ces données, pour la LOX et l'AOS, sont en accord avec les observations effectuées chez la tomate (Heitz *et al.*, 1997 ; Howe *et al.*, 2000). Il est envisageable que la forte expression relative des transcrits au stade vert soit à relier à la production de jasmonate dans les fleurs et les stades précoces : chez la pomme la concentration en jasmonate est environ dix fois supérieure au pic observé avant la crise éthylénique (Fan *et al.*, 1997 ; Fan *et al.*, 1998), il est

donc envisageable que la quantité d'acide jasmonique est peut être encore supérieure dans le stade vert d'abricot qu'au pic éthylénique. Il est aussi possible que ces enzymes soient régulées au niveau post-transcriptionnel ou que l'accumulation du jasmonate dans l'abricot soit différente de la pomme.

V.2.3.2 Transduction du signal jasmonate

Parmi les protéines impliquées dans le signal jasmonate, deux isogènes correspondant à l'E3 ubiquitine lyase (Devoto et Turner, 2003) ont été retrouvés dans la banque d'EST mais n'ont pas été déposés sur les lames (PA272TC1 et PA65TC1), tout comme une probable AXR1 (BGE001L04), impliquée dans la médiation croisée de la transduction du signal auxine et jasmonate (Schwechheimer *et al.*, 2002).

V.2.4 Acide gibbérellique

V.2.4.1 Biosynthèse

Trois transcrits impliqués dans la synthèse de la gibbérelline (GA) ont été obtenus dans les banques d'EST et leur niveau d'expression a été évalué par microarray. Malheureusement, pour les deux premiers d'entre eux le signal est beaucoup trop faible dans la majorité des lames. BGA004I13 correspond à une acide ent-kaurénoïque oxydase, cette enzyme catalyse la formation de l'acide ent-kaurénoïque qui est précurseur du GA₁₂-aldéhyde (Hedden et Phillips, 2000). La seconde est une probable gibbérelline β-hydrolase (PA739TC1). Bien qu'elle possède les acides aminés nécessaires à la liaison au fer et au 2-oxoglutarate communs à toutes les GA oxydases, il est tout de même possible que son substrat ne soit pas la GA. La troisième séquence identifiée (BGA001O04) est homologue à plus de 80 % à des gibbérelline-2-oxydase, en particulier celle de pois (Lester *et al.*, 1999). L'expression des transcrits de BGA001O04 diminue fortement entre le stade vert et le stade vert mature. Cette enzyme catalyse la dégradation des formes active de GA (GA₁ et GA₄) et leurs précurseurs (GA₉ et GA₂₀) en leur produit correspondant, le 2β-hydroxy (Martin *et al.*, 1999). Chez le pois, ce gène est responsable de la mutation *sln* (SLENDER), pour laquelle les plantes sont hyper-allongées. La mutation bloque la conversion de GA₂₀ en GA₂₉ (Ross *et al.*, 1995), suggérant que ce gène est impliqué dans la régulation de GA. Martin *et al.* (1999) avaient précédemment détecté les transcrits de gibbérelline-2-oxydase dans les fruits jeunes de pois, mais c'est la

première fois qu'est effectuée l'évaluation de l'expression au cours de la maturation de cette enzyme de désactivation de la gibbérelline (Thomas *et al.*, 1999). L'expression des transcrits correspondant à la gibbérelline-2-oxydase est inhibée par l'auxine chez le pois, ce qui a pour effet d'augmenter la quantité de GA₁ (Ross *et al.*, 2001).

V.2.4.2 Régulation du signal GA

Trois unigènes (BGA006E01, BGE004N09, BGH002P17) codant pour des protéines appartenant à la famille GRAS, dont font partie les protéines RGA et GAI, régulateurs négatif de GA (Silverstone *et al.*, 1998), ont été mis en évidence dans la banque d'EST. Ces EST sont cependant trop courtes pour obtenir le domaine DELLA dans la région N terminale spécifique des RGA/GAI vis à vis des autres membres de la famille GRAS. En revanche le motif RVER qui est retrouvé dans toutes les RGA, mais aussi dans d'autres membres de la famille, a été identifié dans les trois unigènes. L'un des trois unigènes est constitutif pendant la maturation (BGE004N09), tandis que BGA006E01 et BGH002P17 ont un profil d'expression similaire à celui de la gibbérelline-2-oxydase vue précédemment, avec une forte inhibition entre les stades vert et vert mature. Le mécanisme d'action entre les protéines DELLA et GA n'est pas encore connu, mais l'augmentation de la concentration en GA induit la dégradation de ces protéines, pour cela le motif DELLA est nécessaire (Itoh *et al.*, 2002). Ces protéines sont aussi probablement impliquées dans la réponse à l'éthylène puisque l'éthylène inhibe la croissance des racines d'*Arabidopsis* d'une manière DELLA-dépendante et qu'il retarde la dégradation des protéines DELLA induite par la GA (Achard *et al.*, 2003).

V.3 Biosynthèse des caroténoïdes et de l'acide abscissique

Onze gènes relatifs au métabolisme des caroténoïdes ont été retrouvés dans la banque d'EST. Cependant les 3 gènes impliqués dans la voie de synthèse DOXP n'ont pas été déposés sur les lames (PA879TC1 et BGH10N05, isogènes de 4-hydroxy-3-méthylbut-2-én-1-yl diphosphate synthase (GcpE) et BGH008P13 correspondant à la 4-hydroxy-3-méthylbut-2-ényl diphosphate réductase (LytB)) car ils n'ont pas été identifiés comme appartenant à la catégorie fonctionnelle métabolisme secondaire, comme d'ailleurs l'IPP isomérase (PA579TC1). Les sept unigènes restant plus deux isogènes de zéaxanthine époxydase identifiés précédemment (Mbéguié-Mbéguié, 2000) ont été déposés sur les lames de microarray.

V.3.1 Biosynthèse des caroténoïdes

Seules trois enzymes de cette voie de synthèse n'ont pas été trouvées parmi les EST, à savoir la phytoène désaturase, la caroténoïde isomérase, et la lycopène epsilon-cyclase. Une lycopène epsilon-cyclase existerait cependant parmi les EST publiques de pêche (BU046272).

Les quatre enzymes identifiées, qui sont impliquées dans la synthèse du β -carotène, le pigment majeur de l'abricot, ont toutes un niveau d'expression qui augmente pendant la maturation (Fig. 72) et leurs inductions sont significatives, excepté pour la lycopène β cyclase (BGH008E09). Ces observations sont identiques à celles effectuées sur différentes variétés de *Citrus* (Kato *et al.*, 2004). Ces enzymes de la voie de biosynthèse du β -carotène subissent une augmentation coordonnée, conjointement avec l'accumulation du β -carotène, ce qui peut laisser supposer une co-régulation. La géranylgeranyl pyrophosphate synthase (PA1496TC1) est la seule des quatre enzymes qui présente une variation de l'expression entre les trois variétés, puisqu'elle est significativement plus exprimée chez Blanc que chez Monique au stade mûr. Bien que significatif, le ratio reste tout de même inférieur à 2.

Chez le maïs (Wong *et al.*, 2004) et les Solanacées (Thorup *et al.*, 2000), la phytoène synthase (chez l'abricot PA534TC1), la Zéta carotène désaturase (chez l'abricot PA1519TC1) (uniquement dans le maïs) et la lycopène β cyclase colocalisent avec des QTL de quantité en carotène ou de couleur, ce qui peut laisser supposer que ces gènes sont des facteurs limitants de la teneur en carotène. Si les données d'expression chez Bergeron peuvent laisser supposer que c'est également le cas chez l'abricot, les données de comparaisons variétales

l'infirmieraient. En effet, en l'absence de carotène (chapitre IV.1.4) et de couleur chez Monique et Blanc, il n'y a aucun bouleversement de l'expression de ces gènes.

Il y a trois possibilités pour expliquer cette observation :

(i) D'autres isoformes correspondant réellement aux gènes étudiés chez le maïs et les Solanées seraient réprimées dans les variétés blanches ; cependant seule la phytoène synthase semble actuellement multigénique chez d'autres espèces (Fraser *et al.*, 1999).

(ii) Il est probable, que les caroténoïdes soient synthétisés « normalement » chez les variétés blanches et soient dégradés plus rapidement.

(iii) Les transcrits sont soumis à des régulations post traductionnelles ou des mutations qui rendent les enzymes inactives dans le cas des variétés blanches. La protéomique est une bonne alternative pour observer les modifications de pI. Malheureusement ces enzymes ne semblent pas être abondantes et une purification de chloroplastes semble nécessaire.

Cette expérience montre que les transcrits codant pour les enzymes de la biosynthèse du β carotène sont régulés conjointement et que le facteur de régulation n'est probablement pas le produit, puisque sa quantité est différente entre les variétés. L'examen de la co-localisation éventuelle de ces gènes avec des QTLs couleur ou caroténoïdes est possible puisque certains de ces QTLs ont été identifiés chez les *Prunus* (Dirlewanger *et al.*, 2004).

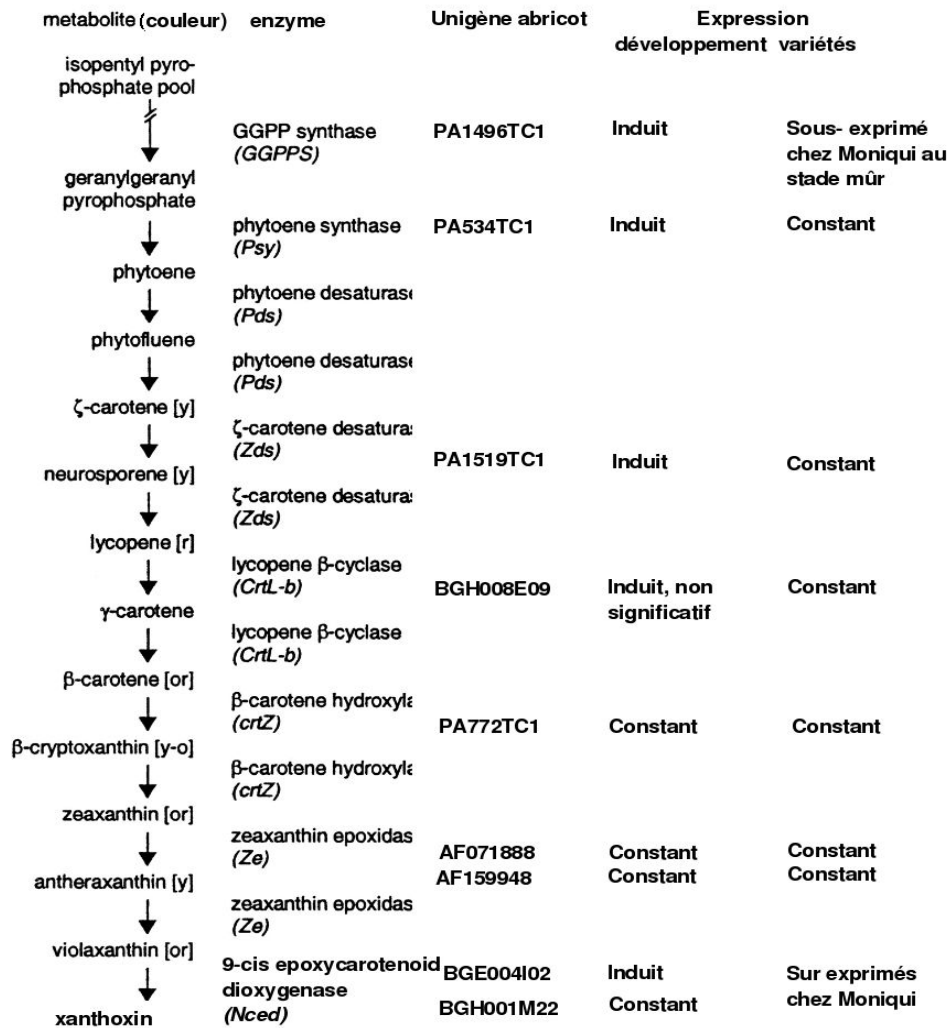


Figure 72 : Schéma simplifié de la voie de biosynthèse des caroténoïdes et des xanthophylles et évolution de l'expression des gènes correspondant aux enzymes.

V.3.2 Dégradation des caroténoïdes / biosynthèse de l'ABA

Les 3 enzymes connues impliquées dans la biosynthèse des xanthophylles ont été identifiées chez l'abricot : une β-carotène hydroxylase (PA772TC1), deux isogènes de zéaxanthine époxydase (AF071888 et AF159948) et deux isoformes de 9-cis époxy-caroténoïde dioxygénase (BGE004I02 et BGH001M22), mais aucune enzyme impliquée dans la transformation de la xanthoxine en ABA n'a été détectée. Le seul transcrit qui évolue significativement pendant le développement, est BGE004I02 qui augmente pendant la maturation. Les autres enzymes ont un profil d'expression qui montre une légère augmentation, excepté la seconde isoforme de NCED, puisque son expression diminue

légèrement. Par contre les deux isogènes de NCED sont fortement exprimées chez Moniqui, et même plus particulièrement l'isogène constitutif pendant le développement chez Bergeron. Il est assez difficile d'effectuer une comparaison phylogénique entre les deux isogènes et les NCED connues dans les autres espèces, puisque la partie codante obtenue chez chacune d'elles ne contient que 66 aa, et la séquence est très conservée à travers les espèces. Toutefois le plus proche homologue des deux isoformes d'abricot est la NCED responsable de la mutation *not* chez la tomate (déficit en acide abscissique) (Burbidge *et al.*, 1999). L'une comme l'autre sont très distantes de l'isoforme de NCED d'*Arabidopsis* qui ne serait pas impliquée dans la biosynthèse de l'ABA mais de C13 norisoprénoides (Schwartz *et al.*, 2001).

La NCED a été identifiée comme étant l'un des régulateurs clé de la biosynthèse de l'ABA, d'après les études sur différents organes et espèces. Son expression est induite par le stress hydrique dans les feuilles et les racines d'*Arabidopsis* (Iuchi *et al.*, 2001), de haricot (Qin et Zeevaart, 1999) et de tomate (Thompson *et al.*, 2000). Une seule étude rapporte actuellement l'expression de NCED dans le fruit, chez l'avocat, son niveau d'expression corrèle avec la quantité d'ABA produite (Chernys et Zeevaart, 2000). La conclusion formulée par ces auteurs semble aussi appropriée ici : il serait judicieux de déterminer quels types d'éléments de réponses aux facteurs environnementaux ou hormonaux sont retrouvés dans les régions promotrices du gène. Compte tenu du lien étroit entre l'ABA et l'éthylène (chapitre II.7.4), il n'est pas exclu de retrouver un élément de réponse à l'éthylène. Ainsi, chez l'avocat, son expression est corrélée avec le dégagement d'éthylène (Chernys et Zeevaart, 2000). Cependant chez l'abricot l'isoforme induite précède le dégagement.

Dans notre cas, il serait nécessaire de déterminer l'évolution de la synthèse d'ABA au cours du développement, en particulier chez Moniqui. Cela permettrait de tester la possibilité selon laquelle la sur-expression de NCED chez Moniqui pourrait être à l'origine des caractéristiques spécifiques de cette variété par rapport à Bergeron (couleur, précocité, dégagement élevé d'éthylène) et de confirmer ou d'infirmer les hypothèses suivantes :

- Si la synthèse d'ABA précède le dégagement d'éthylène et est très élevée chez Moniqui, alors, l'expression de NCED pourrait être sous contrôle d'un facteur abiotique (stress hydrique) ou d'une troisième hormone. L'ABA régulerait l'expression d'un facteur de transcription ou d'un récepteur d'éthylène comme observé par Beaudoin *et al.* (2000), ou inhiberait E8 (qui est effectivement sous-exprimé chez Moniqui), ce qui entraînerait une sur-production d'éthylène.

- Si l'ABA suit le dégagement d'éthylène, comme observé dans l'avocat (Chernys et Zeevaart, 2000) et est très élevé chez Moniqui, l'expression de NCED serait alors sous contrôle au moins partiel de l'éthylène, et peut être aussi induit par le stress hydrique. L'ABA jouerait un rôle catalyseur dans la production d'éthylène, par induction d'un facteur de transcription impliqué dans la régulation autocatalytique. L'induction de E8 chez Moniqui serait peut être régulée par l'ABA, ce qui limiterait l'auto inhibition de l'éthylène.

- Si aucune différence significative au niveau de la quantité d'ABA produite n'est observée, il serait alors peu probable que la NCED soit liée à la différence de phénotype entre Bergeron et Moniqui.

Dans le cas de Moniqui, l'induction des NCED suggère que l'équilibre du pool de xanthophylles serait fortement bouleversé, ce qui pourrait avoir comme effet de déplacer l'équilibre de toute la voie métabolique vers la production d'ABA, avec pour conséquence une forte diminution du pool de substrat (β -carotène) et donc une implication directe sur le phénotype du fruit. Les travaux antérieurs ayant montré un ratio caroténoïdes clivés / ABA synthétisé, de 1 dans les racines de haricot soutiennent cette hypothèse (Parry *et al.*, 1992).

Afin de valider cette hypothèse, il faudrait déterminer les effets d'une sur-expression ou d'une répression de NCED dans les fruits. Cependant le modèle tomate (*Lycopersicon esculentum*) n'est peut être pas très adapté dans ce cas, puisque les β et ϵ lycopène cyclases sont inhibées durant la maturation (Ronen *et al.*, 1999 ; Pecker *et al.*, 1996). En revanche d'autre espèces du genre *Lycopersicon* accumulent du β -carotène, comme *L. cheesmannii*, grâce à la présence d'une seconde lycopène β cyclase dans les fruits (Ronen *et al.*, 2000).

En l'état actuel, aucune piste moléculaire n'apparaît pour expliquer l'absence de caroténoïdes chez Blanc, car toutes les enzymes précédemment décrites ici ont une expression identique à celle de Bergeron. Cependant, contrairement à Moniqui, les précurseurs des caroténoïdes colorés (phytoène, phytofluène) ne sont présents qu'en quantités marginales dans le péricarpe. Il est donc beaucoup moins probable que les caroténoïdes soient effectivement synthétisés dans le fruit de la variété Blanc. Le facteur limitant dans la couleur de Blanc pourrait donc être situé en amont du phytoène dans la voie métabolique, puisque tout les gènes en aval sont transcrits comme dans Bergeron et que peu de phytoène est synthétisé. Il est aussi probable qu'une ou plusieurs enzymes ne soient pas fonctionnelles. Pour la même raison que précédemment, il serait judicieux de séquencer entièrement la phytoène synthase dans la variété Blanc et Bergeron. Comme évoqué dans l'hypothèse (iii) du paragraphe précédent, la protéomique est une autre alternative pour répondre à cette question.

V.3.3 Gènes potentiellement régulés par l'ABA

Une protéine kinase (BGH009N22) dont des homologues de riz ou de blé sont différentiellement régulés par l'ABA (Anderberg et Walker-Simmons, 1992 ; Kobayashi, Y. *et al.*, 2004), voit son niveau d'expression augmenter au cours de la maturation de Bergeron. Il est extrêmement surprenant, d'après les observations du chapitre précédent, que l'expression de ce gène soit inhibée chez Moniqui au stade mûr. Cependant, chez le riz, aucun membre de cette famille multigénique n'est contrôlé uniquement par l'ABA. Le niveau d'expression d'un autre isogène (PA269TC1) de cette protéine kinase déposé aussi sur les lames de microarray, augmente jusqu'au stade mi-mûr et chute ensuite. Il est sous-exprimé chez Moniqui au stade mi-mûr et sur-exprimé dans cette même variété au stade mûr par rapport à Bergeron, ce qui pourrait signifier qu'il est probablement constitutif dans cette variété entre ces deux stades. Les observations sur blé et riz suggèrent que l'expression de ces protéines kinases pourrait être liée au stress hydrique.

BGE005K04 correspond à une protéine phosphatase dont un homologue chez le faîne (*Fagus sylvatica*) est régulé par l'acide abscissique (Lorenzo *et al.*, 2001). Cette protéine joue un rôle dans l'arrêt de la dormance chez cette espèce, ainsi que chez la cerise sauvage (Stephen *et al.*, 2004), mais son expression n'a pas été observée dans le fruit. Elle est inhibée dans la variété Blanc et augmente pendant la maturation de Bergeron.

Les niveaux d'expression de quatre autres unigènes, dont des homologues sont sous contrôle de l'ABA chez d'autres espèces, restent constants lors du développement de l'abricot. Ce résultat a pu être vérifié pour les deux isoformes (PA2TC1, PA2TC7) de l'ASR, indépendamment par Northern et par microarray au niveau des transcrits, comme par protéomique. Les produits de ces gènes font partie des protéines les plus abondantes du péricarpe.

L'étude antérieure de Mbéguié-Mbéguié (2000) rapporte l'induction d'un transcrit d'ASR pendant le mûrissement de l'abricot mais le Northern électronique suggère ici que cette induction ne serait pas significative, alors que les microarrays font ressortir des profils constitutifs pour les deux isogènes. Deux spots protéiques (spots 750 et 781) correspondent à la ASR et sont aussi constitutifs. Il n'a pas été possible de déterminer quel isogène encodait quel spot protéique car les séquences traduites ne diffèrent que par un acide aminé. La fixation du SDS semble affectée par une richesse exceptionnelle en lysine (et autres acides aminés chargés), ce qui expliquerait l'écart important entre masse théorique et masse observée en SDS-PAGE.

Les profils d'expression constitutifs observés chez l'abricot sont cohérents avec les résultats obtenus sur le raisin, où l'ASR décroît légèrement dans le cépage Ugni blanc (Cakir *et al.*, 2003) ou bien reste constant chez Syrah (Terrier *et al.*, 2001a). Le gène est également multigénique chez la vigne (Rossi *et al.*, 1996). On observe plusieurs isoformes très abondantes sur les gels 2D de raisin, qui varient suivant le génotype (Sarry *et al.*, 2004). Plusieurs arguments montrent que l'ASR régulerait la transcription du transporteur d'hexose VvHT1 en se fixant sur un motif connu de son promoteur chez la vigne (Cakir *et al.*, 2003). L'expression constitutive du gène et de la protéine ne plaide pas pour un rôle déterminant dans l'augmentation importante des flux de sucres dans le fruit, lors de l'induction du mûrissement chez l'abricot, comme chez le raisin.

V.4 Métabolisme secondaire (hors caroténoïdes)

V.4.1 Flavonoïdes

V.4.1.1 Métabolisme général des flavonoïdes

La majeure partie des gènes impliqués dans la voie de biosynthèse des flavonoïdes ont été identifiés comme différentiellement exprimés pendant la maturation de l'abricot. Les gènes qui ne sont pas spécifiques à la biosynthèse des anthocyanes ont au moins un isogène sur-exprimé dans les fruits verts, dont le niveau d'expression décline après le stade vert mature. Seul le transcrite codant pour la première enzyme de la voie décroît linéairement pendant la maturation, la phénylalanine ammonia lyase (BGA010H20) (Fig. 73). Trois des quatre isogènes de chalcone synthase (famille protéique 13), BGA007H10, BGA009P18 et PA1546TC1 ont une expression qui décroît après le stade vert immature ; la quatrième (PA46TC1) est au moins constitutive (elle présente une augmentation non significative d'un facteur 2 entre le stade vert et le stade mi-mûr). BGA007H10, BGA009P18 et PA1546TC1 sont homologues à 81 % au moins avec une chalcone synthase de pomme dont l'expression augmente dans la pellicule des variétés rouges (Honda *et al.*, 2002), PA46TC1 n'est homologue qu'à 63 % avec cette protéine.

Deux isogènes de flavonoïde 3'-hydroxylase ont un profil d'expression similaire aux 3 chalcones synthases (PA671TC1 et PA760TC1), PA760TC1 est homologue à 94 % à celle étudiée chez la pomme dont l'expression augmente dans la pellicule des variétés rouges (Honda *et al.*, 2002) ; ce n'est pas le cas de PA671TC1. PA760TC1 est sur-exprimée chez Blanc, particulièrement au stade mûr.

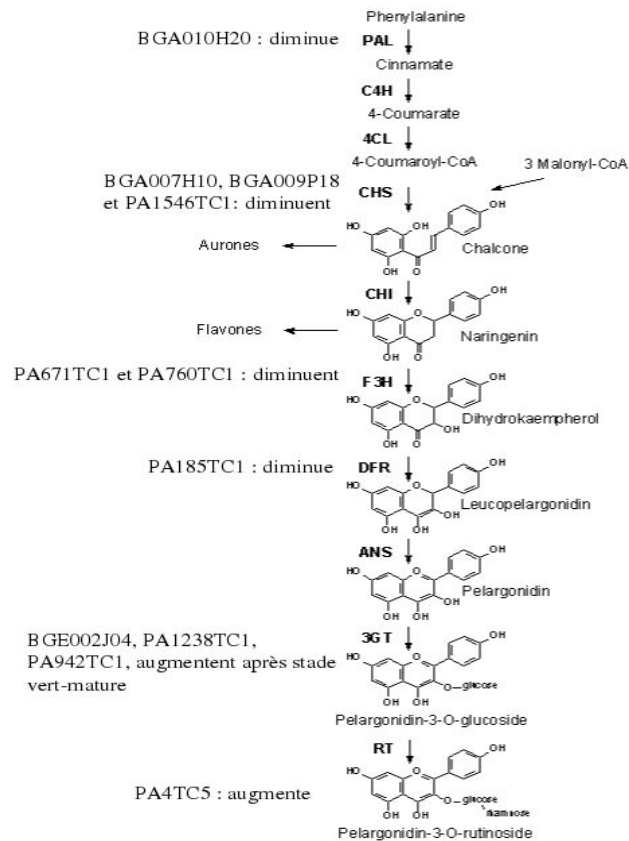


Figure 73 : Schéma simplifié de la voie de synthèse des anthocyanes et évolution des l'expression des gènes identifiés dans l'abricot correspondant aux enzymes. PAL : Phénylalanine Amonia Lyase. C4H : Cinnamate-4-Hydroxylase. CHS : Chalcone Synthase. CHI : Chalcone Isomérase. F3H : Flavanone 3-Hydroxylase. DFR : Dihydroflavonol-4-Réductase. ANS : Anthocyanidine Synthase. 3GT : UDPG-flavonoïde-3-O-GlucosylTransférase. RT : anthocyanidine-3-glucoside RhamnosylTransférase.

Une dihydroflavonol-4-réductase (PA185TC1) diminue après le stade vert immature. PA185TC1 est homologue à 85 % à celle étudiée chez la pomme dont l'expression augmente dans la pellicule des variétés rouges (Honda *et al.*, 2002). Tous ces transcrits réprimés pendant le développement des fruits d'abricot entiers, sont induits dans la pellicule de pomme (Honda *et al.*, 2002).

La variété Bergeron ne contient pas d'anthocyanes dans la pulpe, et elles ne sont accumulées que très légèrement, pendant le mûrissement, dans la pellicule et pas du tout au stade vert.

Il est probable que l'expression de ces gènes soit plutôt reliée avec le métabolisme des tannins condensés (Moyano *et al.*, 1998). Cette hypothèse est renforcée par une possible diminution du niveau d'expression d'un unigène impliqué dans ce métabolisme. PA1407TC1 est homologue à la leucoanthocyanidine réductase de *Desmodium uncinatum*, dont le rôle spécifique dans la biosynthèse des tannins vient d'être établi (formation de catéchine) (Tanner *et al.*, 2003). Le transcrit d'abricot est cependant trop faiblement exprimé, en particulier dans les stades mûrs, pour obtenir un profil d'expression significatif.

V.4.1.2 Biosynthèse des anthocyanes

Parmi les gènes spécifiques de la voie de synthèse des anthocyanes, 4 isogènes probables d'UDP-Flavonol:GlucosylTransférase ont été identifiés, 3 d'entre eux ont une expression qui est réprimée au stade vert mature (BGE002J04, PA1238TC1, PA942TC1) puis induite ensuite, sans toutefois atteindre son niveau de transcription du stade vert. La quatrième diminue après le stade vert mature (BGE018P10).

PA147TC2 correspond à une O méthyl-transférase déjà obtenue chez l'abricot (PAU82011) par Mbéguié-Mbéguié (2000). Nos résultats confirment l'analyse précédente chez Bergeron et Moniqui (Mbéguié-Mbéguié 2000), le gène est sur-exprimé aux stades vert mature et mi-mûr. Cependant il est significativement, mais légèrement sur-exprimé chez Blanc. Il est donc envisageable qu'il soit réprimé par l'éthylène. PA1452TC1 présente un profil d'expression identique, mais il est très nettement sous-exprimé chez Moniqui.

Un unigène (PA4TC5) homologue à une anthocyanidine-3-glucoside rhamnosyltransférase présente un niveau d'expression qui augmente au cours du développement et stagne après le stade mi-mûr. Il est sur-exprimé chez Bergeron. L'enzyme codée par cet unigène catalyse la réaction permettant la synthèse de pigments stables.

Il est difficile d'établir un modèle réactionnel de synthèse d'anthocyanes chez l'abricot puisque aucun de ces composés n'a été dosé dans cette étude, et que la spécificité de substrat des enzymes identifiées ici est totalement inconnue. Aucune d'entre elle ne présente d'homologie forte avec des enzymes bien caractérisées chez d'autres espèces pour pouvoir émettre la moindre hypothèse. De plus les variétés utilisées ici et la présence de mésocarpe ne sont pas adaptées à ce type d'études. Cependant il apparaît que la biosynthèse générale des

flavonoïdes semble avoir lieu bien avant la crise climactérique. Il existe d'autres isogènes codant pour ces enzymes dans la base de données d'EST dont la différence d'expression n'est pas significative, et qui pourraient être spécifiques de la synthèse d'anthocyanes dans la pellicule. Il semble tout de même que des anthocyanes soit majoritairement synthétisés pendant la maturation, puisque les transcrits codant pour l'enzyme catalysant l'étape finale de leur biosynthèse sont induits.

Quoi qu'il en soit, il a été obtenu un jeu assez conséquent d'unigènes candidats pour déterminer l'évolution et les interconversions des anthocyanes pendant la maturation chez des variétés d'abricot à forte teneur en anthocyanes, comme les variétés « double rouge ».

V.4.2 Arômes

Un transcrit (BGE003M12) qui pourrait être impliqué dans la biosynthèse des arômes augmente au stade vert mature et s'avère fortement exprimé chez Monique. Il coderait pour une S-adénosyl-L-méthionine:salicylique/benzoïque acide carboxyl méthyltransférase qui catalyse la formation de méthyl-benzoate ou de méthyl-salicylate à partir respectivement, du benzoate ou du salicylate (Murfitt *et al.*, 2000). Ces composés sont très fréquents chez les fleurs (Dudareva *et al.*, 2000), mais peu étudiés chez les fruits. Cependant une forte concentration de méthy-salicylate, qui de plus augmente pendant le développement, a été observée chez la prune (Gomez et Ledbetter, 1997). Étant donné que les quantités de benzoate, de salicylate, de méthyl-benzoate et de méthyl-salicylate n'ont pas été mesurées dans cette étude, il se peut que l'accepteur soit totalement différent. Cependant la seule donnée en notre possession sur ce point est que l'abricot est l'un des fruits contenant le plus de salicylate (Swain *et al.*, 1985). Le benzaldéhyde qui est le précurseur du benzoate est spécifique de Bergeron ; il est donc envisageable que la sous-représentation du benzaldéhyde chez Monique soit liée à l'utilisation massive de son dérivé (le benzoate) pour la synthèse de méthyl-benzoate.

Un unigène (PA614TC1) homologue à une acétylCoA: benzyl alcohol acétyltransférase est spécifique de Bergeron ; il n'est pas impossible qu'il soit induit pendant la maturation puisque son niveau d'expression augmente d'un facteur 32, mais non significativement. L'enzyme catalyse la synthèse de benzyl acétate à partir d'acool benzylique qui est aussi le précurseur du benzaldéhyde (Dudareva *et al.*, 1998). Le benzyl acétate est aussi aromatique, mais n'a pas été dosé ici.

PA1425TC1 homologue à des alcool acyltransférases est sous-exprimé chez Moniqui au stade mi-mûr. Il est non significativement induit au cours du développement de Bergeron, comme observé chez le melon (Yahyaoui *et al.*, 2002). Il est cependant trop peu homologue avec les gènes connus du melon pour prédire une spécificité de substrat. Ainsi, les paralogues du melon, avec 82 % d'homologie, n'ont pas les mêmes substrats. Le niveau auquel PA1425TC1 pourrait intervenir dans la différence aromatique entre Bergeron et Moniqui est *de facto* inconnu.

PA1442TC1 est une réductase homologue à une alcool déshydrogénase de fraise (Aharoni *et al.*, 2000). Son expression augmente pendant le développement de Bergeron et est réprimée chez Blanc. Cette enzyme, probablement reliée au métabolisme aromatique, pourrait être impliquée dans une possible différence aromatique entre Blanc et les autres variétés.

V.5 La perte de fermeté

La perte de fermeté a été largement étudiée chez les fruits compte tenu de son caractère ubiquiste pendant le mûrissement et de son importance pour la durée de vie post-récolte, particulièrement problématique chez l'abricot. La perte de la fermeté se caractérise par la dépolymérisation de substances distinctes, d'une part les pectines et d'autre part les chaînes de xyloglucanes (ou hemicellulose). Ces deux phénomènes interviennent à des moments distincts et n'impliquent pas les mêmes enzymes. Une approche génomique portant sur 32 gènes a été récemment entreprise chez la pêche (Trainotti *et al.*, 2003) mais il n'est pas possible de retrouver les séquences correspondantes dans les banques d'EST publiques car les numéros d'accèsion ne sont pas précisés dans la publication. Environ un tiers des gènes montrant des variations d'expression significatives dans l'approche Northern électronique sont reliés à ce processus, indiquant la forte abondance relative de ces gènes.

V.5.1 Dépolymérisation des pectines

V.5.1.1 Pectine méthylestérase (PME)

Trois unigènes codant pour des pectine estérases ont pu être identifiés dans la base de données EST (PA311TC1, PA717TC1 et BGE016B06), elles ne présentent d'ailleurs aucune homologie entre-elles sur leur traduction conceptuelle. Seul BGE016B06 montre une expression différentielle au cours du développement de l'abricot. Son niveau d'expression diminue au stade vert mature, puis retrouve le niveau d'expression du stade vert pendant le mûrissement. L'oligonucléotide correspondant à PA717TC1, unigène déjà répertorié dans les bases de données publiques (AF184079), n'a donné aucun signal d'hybridation. L'homologue de PA311TC1, qui s'avère constitutif chez l'abricot, est induit chez la fraise. Il est induit par l'auxine et réprimé par l'éthylène exogène (Castillejo *et al.*, 2004).

Jusqu'à neuf unigènes codant pour des inhibiteurs de pectine méthylestérases (PMEI) ont été identifiés dans la base de données EST et l'expression de 7 d'entre eux a été évaluée. Il existe également un domaine PMEI sur l'extrémité N terminale des PME, qui rend parfois délicate la distinction entre PME et PMEI. Dans notre cas, comme le séquençage a été réalisé en 3', les séquences protéiques conceptuelles « débutent » par l'extrémité C terminale, et le fait de débiter par un domaine PME ou un domaine inhibiteur permet de distinguer les deux types de gènes. Les séquences des trois PME identifiées sont d'ailleurs trop courtes pour avoir atteint la position attendue du motif PMEI.

Les PMEI inhibent les pectines méthylestérases ainsi que les invertases à travers la formation d'un complexe non covalent (Giovane *et al.*, 1995). Elles sont impliquées dans la régulation du développement du fruit chez le kiwi, le métabolisme des sucres et l'extension de la paroi cellulaire (Camardella *et al.*, 2000). Les mêmes auteurs suggèrent que les PMEI jouent un rôle physiologique plus général que la seule régulation de l'activité de la méthyestérification des pectines par les PME, puisque une possible implication des PMEI dans les mécanismes de défense des plantes contre les attaques de pathogènes a été observée (Chen *et al.*, 2000). Ce dernier rôle serait aussi très important chez l'abricot, puisque 6 des 7 PMEI sont induites pendant la maturation (PA43TC1 BGE005M19, PA23TC2, BGH007L07 observés par microarray, PA7TC6, observé par microarray et Northern électronique et PA5TC7 observé par Northern électronique), de façon anticorrélée avec l'évolution de la fermeté, mais bien liée avec la sensibilité des fruits vis à vis des attaques de pathogènes. L'induction de pectate lyase (V.5.1.2) suggère que la désestérification des pectines, qui est

une condition préliminaire à l'activité hydrolytique de cette dernière enzyme (Wakeley *et al.*, 1998) aurait effectivement lieu pendant la maturation.

PA23TC2 appartient à une famille de 3 unigènes dont la proximité interdit l'élaboration de sondes spécifiques (seule la région non codante a pu être séquencée pour PA23TC1 et BGE015D07, le premier étant un variant d'épissage de PA23TC2, et le second ne présentant que quelques mutations discrètes). En revanche, PA43TC1, BGE005M19, PA7TC6 et PA1TC31 ne présentent des homologues qu'après traduction conceptuelle.

Quatre des PMEI sont différentiellement exprimées entre les variétés : BGH007L07 et BGE005M19 sont sur-exprimés chez Blanc, PA43TC1 chez Moniqui et PA23TC2 chez Moniqui au stade mi-mûr. Une hypothétique régulation de l'expression de ces gènes par l'éthylène reste donc possible.

V.5.1.2 Pectate lyase

Trois isogènes de pectate lyase ont été identifiés dans la banque d'EST, mais le niveau d'expression n'a pu être caractérisé que pour PA6TC8, par Northern électronique et PCR quantitative. Aucun signal n'a été détecté à partir des sondes définies pour les deux autres isogènes, et le signal pour PA6TC8 n'est que rarement supérieur au bruit de fond., son niveau d'expression n'est détectable qu'à partir du stade mi-mûr, elle est homologue à 90 % à une pectate lyase (AF243475) détectée dans un stade mûr de pomme.

V.5.1.3 Polygalacturonase

Deux isogènes présents dans les banques d'EST codent pour des polygalacturonases : BGA008P12 qui semble être constitutif pendant le développement et PA274TC1. Cependant, bien que son expression soit très fortement induite, l'augmentation n'est pas significative, car aucun signal n'a été obtenu dans les expériences de comparaison entre les deux stades les plus précoces.

BGA008P12 ne possède aucun homologue connu chez les fruits. PA274TC1 présente 97 % d'identité avec la polygalacturonase dont Lester *et al.* (1994) avaient déjà démontré l'induction pendant le développement chez la pêche, aucune expression avant un stade très mûr n'ayant pu être détectée. Chez le melon, l'expression des isoformes de polygalacturonase est tissu-spécifique (Hadfield *et al.*, 1998). Il conviendra de vérifier si PA274TC1 est également fruit spécifique chez l'abricot.

V.5.2 Dépolymérisation du réseau d'hémicellulose

V.5.2.1 Xyloglucan-endotransglycosylases (XET)

Trois unigènes ont été identifiés comme correspondant à des XETs, leur rôle est d'intégrer des xyloglucanes dans le réseau de xyloglucanes-cellulose (Thompson *et al.*, 1997). Deux d'entre eux sont homologues (PA438TC1 et PA1110TC1). Le troisième, qui est un des transcrits les plus abondamment exprimés dans les banques d'EST (PA17TC1), ne présente pas d'homologie avec les deux autres.

L'induction des transcrits de PA17TC1 pendant le développement a été observée par Northern électronique, et validée par PCR quantitative. Cependant, aucun signal d'hybridation n'a été détecté sur les microarrays, ce qui est certainement lié à un problème de design, de synthèse, ou de dépôt de l'oligonucléotide correspondant. Deux spots protéiques ont été identifiés comme correspondant à cet unigène (spots 735 et 736), ce qui semble indiquer des possibles modification post-traductionnelles. Ces protéines s'accumulent simultanément avec la production d'éthylène endogène, en accord avec les observations effectuées pour son homologue chez le kiwi, induit après traitement à l'éthylène (Schroeder *et al.*, 1998). L'activité XET est d'ailleurs induite par l'éthylène chez ce fruit (Redgwell et Fry, 1993). Un homologue à PA17TC1 est également induit chez le maïs traité à l'éthylène (Saab et Sachs, 1996).

L'expression de PA1110TC1 est réprimée pendant la maturation du fruit mais elle semble augmenter légèrement à la fin de la maturation. Il semble que PA1110TC1 joue un rôle au cours de la phase de multiplication cellulaire, contrairement à PA17TC1. Elle intervient probablement ensuite pendant la phase de sénescence. Des transcrits, identiques chez la pêche (AF534870), et fortement homologues chez la pomme (89 % d'homologie AY144594) ont d'ailleurs été isolés à partir de banque d'ADNc issus de fruits surmatures. Cet isogène de XET ne serait donc pas sous contrôle de l'éthylène, contrairement au précédent. L'oligonucléotide correspondant à PA438TC1 n'a pas pu être hybridé.

Les modifications du réseau xyloglucan cellulose catalysées par les XETS seraient impliquées dans le grandissement cellulaire au cours de la maturation (Campbell et Braam, 1999), cependant un tel mode d'action ne peut pas conduire à une rupture de la paroi cellulaire. Il se peut qu'elle participe tout de même à la dépolymérisation puisque Schroder *et al.* (1998) ont démontré une activité hydrolase de la XET homologue à PA17TC1 sur les xyloglucanes, conduisant à leur dépolymérisation.

V.5.2.2 Expansine

Cinq unigènes, identifiés dans la base de données codent pour des expansines, seul PA3TC5 n'a pas été déposé sur les lames de microarray. Tous sont identiques ou homologues à plus de 92 % à des expansines précédemment identifiées dans le genre *Prunus*. Deux d'entre elles ont déjà été caractérisées au cours du développement de l'abricot (Mbéguié-Mbéguié *et al.*, 2002), PA5TC5 correspond à PaExp2 et PA118TC1 à PaExp1. Nos résultats confirment leur induction au cours du développement. Ces deux gènes ne sont cependant pas influencés par le traitement à l'éthylène (Mbéguié-Mbéguié *et al.*, 2002). Le niveau d'expression des 3 autres unigènes n'a pas été étudié précédemment chez les *Prunus*, cependant ils ont chacun un homologue chez la pêche et chez la cerise (PA3TC5 = AB047518 et AF448467 ; PA229TC1 = AB047519 et AF350939 ; PA1152TC1 = AB047519 et AF350939). Chez l'abricot, l'expression de PA1152TC2 est constitutive et celle de PA229TC1 augmente. L'expression d'orthologues potentiels des 5 expansines a été étudiée chez la poire (Hiwasa *et al.*, 2003b). Cependant nos résultats d'une part, et ceux de Mbéguié-Mbéguié d'autre part, permettent de penser que certains sont régulés de façon différente dans ces deux espèces. Pc-Exp3 qui est très nettement le plus proche homologue de PaExp1 (92 % d'homologie sur les séquences nucléiques) est induit pendant le développement de la poire, comme chez l'abricot, cependant son niveau d'expression chute après traitement au 1-MCP, à partir du stade correspondant au début du dégagement d'éthylène, ce qui n'est pas le cas chez l'abricot (Mbéguié-Mbéguié *et al.*, 2002). La même observation pourrait être effectuée entre Pc-Exp2 et PaExp2 (91 % d'homologie), mais PaExp2 est aussi fortement homologue (90 %) à Pc-Exp1, dont l'expression augmente aussi pendant la maturation, mais qui n'est pas affectée par le 1-MCP. De même, l'expression de PaExp1 est moins forte chez Monique, alors que cette variété dégage davantage d'éthylène. PaExp1 et PaExp2 ne seraient donc pas sous contrôle de l'éthylène chez l'abricot contrairement à certains de leurs orthologues potentiels chez la poire.

Pc-Exp7 est le plus proche homologue de PA229TC1 (84 %) et PA1152TC2, son niveau d'expression diminue pendant la maturation de la poire, contrairement aux deux isogènes d'abricot. Il reste bien sûr probable que toutes les isoformes d'expansines n'aient pas été isolées ni chez l'abricot ni chez la poire, ce qui rend délicate l'identification des orthologues entre les deux espèces.

V.5.2.3 Endo β 1,4 glucanases

Deux EST présentes dans les banques codent pour des endo- β -1,4-glucanases, dont l'une est constitutive au cours du développement (PA190TC1), alors que l'autre (PA150TC1) a un niveau d'expression qui augmente fortement au stade vert mature, pour diminuer ensuite. Une protéine très homologue (87 %) montre un profil identique chez la poire (Pc-EG2, Hiwasa *et al.*, 2003b). Les traitements à l'éthylène et au 1-MCP n'ont aucun effet sur son expression chez cette dernière espèce, et le profil que nous avons observé suggère qu'il en serait de même chez l'abricot. PA190TC1 est homologue à 94 % à Pc-EG1, qui est aussi constitutive pendant la maturation. Les mêmes auteurs suggèrent, d'après les travaux sur la tomate de Brummell *et al.* (1997), que cette classe de protéine qui contient un domaine transmembranaire, serait plutôt impliquée dans la synthèse de cellulose lors de l'expansion cellulaire.

Les deux endo- β -1,4-glucanases sont régulées différemment dans les variétés d'abricots : l'expression de PA150TC1 est réduite chez Bergeron, et celle de PA190TC1 est réduite chez Monique. Comme aucune différence n'est visible au niveau de la fermeté des 3 variétés, il est probable que ces deux endo β 1,4 glucanases ne soient pas limitantes dans la perte de fermeté.

V.5.3 Grandissement cellulaire

Deux isoformes d'annexine (spots 1055 et 704) ont été identifiées parmi les protéines induites pendant la maturation. Ces deux protéines semblent correspondre à un même transcrite (PA1TC15), dont le profil d'expression est parallèle à celui des protéines qu'il encode. Cependant des modifications post-traductionnelles ou des dégradations systématiques lors de l'extraction, ont probablement affectées le spot 1055 qui semble correspondre à un fragment du spot 704. Des transcrits homologues d'annexines présentent un profil identique à celui de l'abricot pendant la maturation du poivron et de la fraise (Proust *et al.*, 1996 ; Wilkinson *et al.*, 1995). C'est, à ma connaissance, le premier suivi de l'expression effective pendant le mûrissement, de cette protéine, présumée impliquée dans le grandissement cellulaire (Andrawis *et al.*, 1993). Du fait de son rôle multifonctionnel la fonction exacte des annexines n'est pas encore connue (Clark *et al.*, 2001), bien qu'elle soit impliquée dans le stress osmotique et la signalétique de l'acide abscissique par un processus dépendant du Ca^{2+} (Lee *et al.*, 2004).

V.5.4 Conclusion sur la perte de fermeté

Récemment, un modèle a été proposé pour décrire la perte de fermeté chez la pêche (Brummell *et al.*, 2004). Il a été clairement montré que le ramollissement s'effectue en plusieurs étapes indépendantes. L'évolution de la paroi cellulaire se divise en cinq étapes : (1) début du grossissement du fruit, correspondant à notre stade vert ; (2) début du ramollissement (stade vert mature), lié à l'initiation de la dépolymérisation du réseau de glucanes ; (3) période de ramollissement liée à la maturation (stade mi-mûr), intégrant la poursuite de la dépolymérisation des glucanes ainsi que l'initiation de la déméthylation et la dépolymérisation des pectines ; (4) période tardive de ramollissement (stade mûr), incluant le pic d'éthylène, et la dépolymérisation massive du réseau de pectines ; (5) entrée en sénescence.

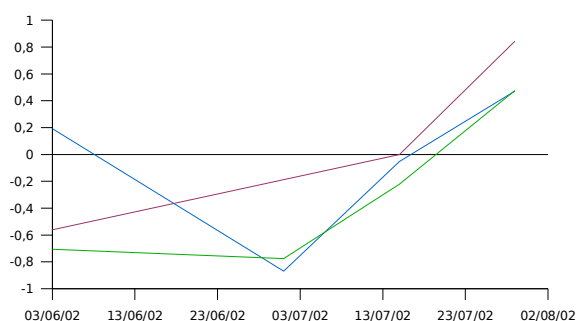


Figure 74 : Évolution du niveau d'expression des gènes impliqués dans la dépolymérisation de pectines au cours du développement du fruit (exprimé en ratio du \log_{10} du niveau d'expression par rapport à la moyenne divisé par la somme des moindres carrés). Rouge : pectate lyase (PA6TC8), les valeurs sont calculées à partir du Northern électronique, le point pour le stade vert mature est absent. Vert : polygalacturonase (PA274TC1). Bleu : pectine méthylestérase (BGE016B06).

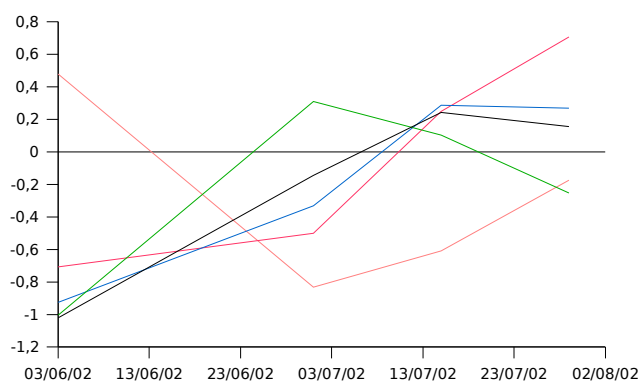


Figure 75 : Évolution du niveau d'expression (exprimé en ratio du \log_{10} du niveau d'expression par rapport à la moyenne divisé par la somme des moindres carrés) des gènes impliqués dans la dépolymérisation des xyloglucanes au cours du développement du fruit. Orange : Xyloglucan endo transglycosylase 1 (PA1110TC1). Rouge : Xyloglucan endo transglycosylase 2 (PA17TC1). Noir : Expansine 1 (PaExp1). Bleu : Expansine 2 (PaExp2). Vert : Endoglucanase (PA150TC1).

Les données obtenues ici suggèrent un mécanisme apparenté à celui observé chez la pêche. D'après l'expression relative des gènes (Fig. 74), il est très possible que les pectines soient dégradées après la crise éthylénique, car une polygalacturonase et une pectine estérase sont induites durant les phase tardives. Cela peut être aussi le cas pour la pectate lyase, bien qu'un point expérimental manque. Il existerait cependant une autre isoforme de pectinestérase, impliquée dans la déméthylestérification durant les stades plus précoces. Le profil de la PME est identique à l'activité mesurée chez la pêche, la traduction (ou l'activation) de la PG semble par contre légèrement décalée dans le temps par rapport aux observations de Brummell *et al.* (2004). Bien qu'un point important manque pour la pectate lyase, une forte induction est tout de même observée entre les stades mi-mûr et mûr. En ce qui concerne la dépolymérisation des glucanes (Fig. 75), l'induction précoce de l'endoglucanase au stade vert mature, permet de penser qu'elle précède effectivement la dépolymérisation des pectines, le profil d'expression de l'endoglucanase est identique à l'activité chez la pêche. Les deux isoformes d'expansine sont induites à partir du stade vert mature, cependant leur expression continue d'augmenter après. Par contre, les XET semblent intervenir dans un second temps, ce qui ne participe pas à éclaircir leur mode d'action déjà mal connu. Quoiqu'il en soit, il n'y a aucune raison de penser actuellement que le processus de perte de fermeté chez l'abricot soit différent de celui de la pêche.

V.6 Métabolisme du carbone

V.6.1 Photosynthèse

Contrairement aux fruits mûrs, les fruits verts sont capables d'assimiler le CO₂ par photosynthèse, principalement dans la partie externe du péricarpe, où une forte activité photosynthétique a été enregistrée (Obukosia *et al.*, 2003). Les dosages de chlorophylle effectués sur l'abricot permettent de confirmer cette tendance générale, car on observe une diminution des chlorophylles entre les stades 92 à 107 JAA.

Les données acquises concernant les enzymes impliquées dans la photosynthèse proviennent presque exclusivement de l'approche protéomique. Sur les neuf protéines

identifiées, une seule est présente dans la base de données EST (la fructose bisphosphate aldolase chloroplastique, PA149TC1, spot 659). La triosephosphate isomérase chloroplastique (spot 897) a été identifiée d'après les EST de pêche. Ceci suggère une forte régulation de toutes les enzymes au niveau post-transcriptionnel. Le niveau d'expression des transcrits PA149TC1 diminue pendant tout le développement, alors que l'abondance des protéines chute à la crise climactérique.

Parmi les protéines identifiées reliées à la photosynthèse, PsbP (spot 957) est une des protéines extrinsèques majeures du complexe d'émission de l'oxygène. Il a été démontré que le saccharose réprime l'expression des transcrits chez *Arabidopsis* (Kochhar *et al.*, 1996) et une infection par le virus de la mosaïque du tabac réprime la synthèse protéique de PsbP dans *Nicotiana benthamiana* (Perez Bueno *et al.*, 2004). D'après ces observations, en l'absence d'infection par des pathogènes et avec la concomitance entre l'accumulation de sucre et la perte de protéine, il est donc envisageable que le saccharose régule la transcription et/ou la traduction de PsbP dans les tissus du péricarpe. La seconde protéine impliquée dans la photosynthèse, la sous unité β de l'ATP synthase (spot 982) a été identifiée avec un recouvrement de 90 % dans la région correspondant au baril β , spécifique des sous unités α/β de l'ATP synthase (Shirakihara *et al.*, 1997). L'expression des ATP synthases est sous contrôle des protéines 14-3-3 (Bunney *et al.*, 2001), mais aucune de ces protéines n'a été identifiée ni par protéomique, ni différentiellement exprimée sur les lames de microarray. Cependant, la synthèse protéique d'une 14-3-3 est effectivement réprimée pendant la maturation de la mangue (Gaspar, non publié).

L'intensité de 5 isoformes de la grande sous unité de la ribulose bisphosphate carboxylase (Rubisco) décroît pendant la maturation (spots 405, 490, 798, 983 1009). La Rubisco est connue pour être fortement affectée par des modifications co- et post-traductionnelles (Houtz et Portis, 2003), ce qui pourrait expliquer en partie, son absence dans la base de données EST. De plus, il semble que la Rubisco soit aussi sujette à des modifications incontrôlées dont la raison est expérimentale. Certains spots, dont l'expression semble être différentielle, sont masqués par de la Rubisco dans certains gels pour les 3 premiers stades de développement (53, 65, 92 DAA). La Rubisco est une enzyme clé qui autorise l'entrée du carbone de la biosphère en catalysant la réaction entre le CO₂ et le ribulose 1,5-bisphosphate (RuBP) pour former deux molécules d'acide phosphoglycérique (Miziorko et Lorimer, 1983). Les abondances protéiques de la triosephosphate isomérase (TPI, spot 897) et la fructose bisphosphate aldolase (FBPald, spot 659) impliquées également

dans le cycle de Calvin-Benson (Calvin, 1962) décroissent aussi pendant la maturation. Ces deux protéines correspondent effectivement à des isoformes chloroplastiques, mais on retrouve des isoenzymes qui interviennent dans la glycolyse cytosolique chez les végétaux supérieurs. Les cinq étapes de la conversion du 3-phosphoglycérate en fructose 6 phosphate sont identiques dans les deux voies, mais fonctionnent en sens opposé. Les isoformes cytosoliques de FBPald, glycéraldéhyde-3-phosphate déshydrogénase (G3PDH) et malate déshydrogénase ont été détectées comme constitutives pendant le développement de l'abricot. Ces résultats confirment les études précédentes, où une plus grande activité photosynthétique chez l'avocat (Zaubermann et Schiffman-Nadel, 1972), ainsi que la forte abondance de transcrits de Rubisco et d'une protéine du complexe d'émission de l'oxygène, PsbA, dans le potiron (Obukosia, *et al.*, 2003) ont été observés chez les fruits jeunes par rapport aux fruits mûrs.

La concomitance observée entre l'induction de la production d'éthylène et la chute, souvent brutale, de l'intensité des spots relatifs aux enzymes impliquées dans la photosynthèse, tend à confirmer les travaux de Zhou *et al.* (1998). Le mutant *gin1* (chez lequel les enzymes photosynthétiques ne sont plus réprimées par le glucose) montre un lien inattendu entre les voies de signalisation éthylène et sucre. Le rôle exact de l'éthylène dans la photosynthèse doit être mieux caractérisé, cependant il a été précédemment démontré que l'éthylène induit la synthèse et l'activité de la chlorophyllase (Treibitsh *et al.*, 1993 ; Jacob-wilk *et al.*, 1999). Des études transcriptomiques sur *Arabidopsis* montrent aussi une répression des gènes liés à la photosynthèse après traitement par l'éthylène (Zhong et Burns, 2003).

V.6.2 Métabolisme du saccharose

Le saccharose, est le principal produit final de la photosynthèse et la forme prépondérante de sucre transporté dans le phloème des végétaux. Il représente 35 % du poids sec de l'abricot, alors que le total des carbohydrates représente 58 %. La quantité totale de sucre augmente dès la fin de la phase 1 du développement (92 JAA, stade vert mature) et le maximum d'accumulation a lieu à 121 JAA (stade mûr). L'évolution du saccharose suit la même évolution (Fig. 11). Dans ce travail, 5 protéines impliquées dans le métabolisme du saccharose ont été détectées.

V.6.2.1 Saccharose synthase

Trois isoformes de saccharose synthase (spots 109, 111 et 112), présentent une intensité qui augmente pendant la maturation. Elles semblent toutes trois provenir du même transcrit (PA349TC1), dont le niveau d'expression augmente en parallèle avec la quantité de protéines.

Un autre transcrit de saccharose synthase, non relié aux protéines observées, PA268TC1, présente une expression qui diminue pendant le développement. PA349TC1 (Fig. 76) appartient au groupe SUS1 de saccharose synthase d'Eudicotylédones (Komatsu *et al.*, 2002) et PA268TC1 appartient au groupe SUSA contenant, entre autres, CitSusA et une saccharose synthase de poire.

L'expression pendant la maturation chez les *Citrus* des plus proches homologues des saccharose synthases d'abricot n'est pas en accord avec nos observations. Pour les deux homologues à PA349TC1, CitSus2 (homologue à 84 % sur la séquence nucléique, Fig. 76) n'a pas été détecté dans les tissus comestibles et l'expression de CitSUS1 (homologue à 84 % sur la séquence nucléique, Fig. 76) diminue pendant la maturation. L'expression de CitSusA (homologue à 83 % sur la séquence nucléique à PA268TC1, Fig. 76) augmente pendant la maturation. Cependant l'expression d'une saccharose synthase de pêche dont la séquence est inconnue augmente (Trainotti *et al.*, 2003), comme celle de PA349TC1. Les auteurs ont précisé que son plus proche homologue est une saccharose synthase de pois (PEASUS2). Il est probable qu'il s'agisse de PPSUS2, transcrit retrouvé dans les bases de données publiques et effectivement fortement homologue à la saccharose synthase de pois, sa séquence est identique à celle de PA349TC1 (Fig. 76), confirmant ainsi les résultats obtenus chez l'abricot.

Ceci montre clairement (i) l'importance capitale d'une isoforme spécifique pour la régulation temporelle et spatiale de la synthèse du saccharose dans le péricarpe et (ii) que le transfert de connaissance entre espèces peut se révéler hasardeux, surtout si on ne sait pas si on a affaire à des orthologues.

Des constructions anti-sens dirigées contre deux saccharose synthases équidistantes de celles de *Citrus* et de *Prunus* (LeSUS2 et LeSUS3, Fig. 76) exprimées dans le développement précoce des fruits ont été effectuées chez la tomate et ne semblent finalement pas impliquées dans l'accumulation du saccharose (Chengappa *et al.*, 1999 ; D'aoust *et al.*, 1999).

Une autre saccharose synthase existe chez la pêche, PPSUS1 (Fig. 76), et il n'est pas polymorphique entre des variétés de pêches de phénotypes contrastés au niveau du taux de sucre (Etienne *et al.*, 2002). Aucun homologue à ce gène n'a été retrouvé chez l'abricot. Il semble donc que PA349TC1 soit un candidat mieux adapté pour ce type d'étude.

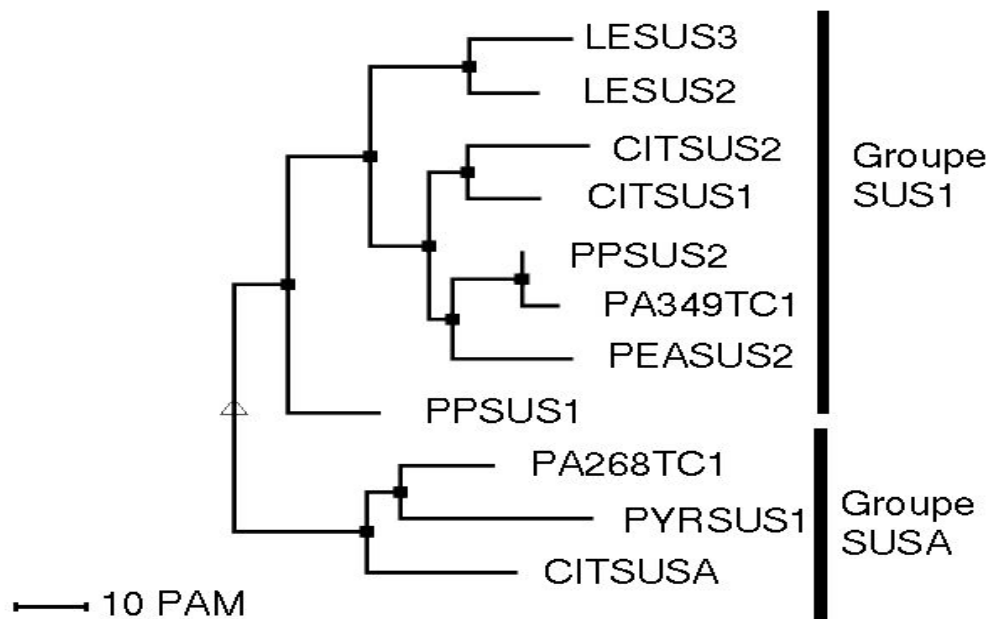


Figure 76 : Dendrogramme phylogénique des saccharose synthases représentatives identifiées chez les fruits. Les groupes SUS1 et SUSA ont été définis par Komatsu et al. (2002). CITSUS1 (espèce : mandarine, numéro d'accèsion : BAA88905), CITSUSA (mandarine, BAA88904), CITSUS2 (mandarine, BAA88902), PYRSUS1 (poire, BAB20799), PPSUS1 (pêche, AAL16966) PEASUS2 (pois, CAA09910), LESUS2 (tomate, AAA34196), LESUS3 (tomate, CAA09593), PPSUS2 (EST de pêche AJ822806).

V.6.2.2 UDP-Glucose pyrophosphorylase

L'UDP-glucose pyrophosphorylase catalyse la synthèse de l'UDP-Glucose qui sera utilisé par la saccharose synthase pour former le saccharose. La caractérisation exacte du profil protéique de l'UDP-glucose pyrophosphorylase n'a pas pu être effectuée correctement, car dans certains gels, des fragments Rubisco ont interféré dans l'identification. Il semble cependant que l'intensité protéique augmente pendant la maturation. Une induction du niveau d'expression des transcrits (PA91TC1) a également été observée par PCR quantitative ; l'augmentation observée en microarray n'est pas significative. Cette enzyme est également retrouvée en abondance dans les raisins mûrs (Sarry *et al.*, 2004).

V.6.2.3 Xylulose kinase

Une protéine de xylulose kinase (BGA010G05, non déposé sur les lames) est induite après 107 JAA, ce qui n'a jamais été observé au cours de la maturation des fruits. Cette enzyme qui catalyse la conversion de xylulose en xylulose-5-phosphate n'a presque pas été étudiée chez les plantes et son importance dans la maturation devra être mieux caractérisée.

V.6.2.4 Autres gènes

Il est très surprenant d'observer que le niveau d'expression des transcrits d'ADP-glucose pyrophosphorylase (PA549TC1) et de d'amidon synthase augmente pendant la maturation (PA134TC1, BGA001G19P). Ces enzymes sont impliquées dans la biosynthèse de l'amidon, or l'abricot n'accumule pas d'amidon. En revanche, le niveau d'expression de l'UDP-Glucose épimérase (BGH001G07) diminue fortement pendant la maturation. Cette enzyme joue un rôle dans la formation de la paroi cellulaire en catalysant la transformation de l'UDP-Glucose en UDP-galactose qui est la forme active du glucose incorporé dans les chaînes d'hémicellulose (Zabackis *et al.*, 1995). Chez *Arabidopsis*, l'UDP-galactose accumulé rentre en compétition avec l'UDP-glucose sur le site de fixation de la saccharose-phosphate synthase, et inhibe la production de saccharose (Dormann et Benning, 1998). Il est donc possible que l'UDP-galactose inhibe aussi la synthèse de saccharose dans les abricots jeunes.

Des invertases ont été retrouvées dans la base de données EST mais n'ont pas été caractérisées dans cette étude.

Un transcrit de transporteur d'hexose (PA288TC1) est induit pendant la maturation. Il présente 75 % d'homologie avec le transporteur d'orge *stp1* (Weschke *et al.*, 2003). On retrouve des homologues parmi les EST de rosacées et de vigne, il est par contre absent des EST de tomate, l'expression de cet isogène n'a jamais été caractérisée chez aucun fruit. Chez l'orge, ce transporteur est spécifique de l'endosperme. Il existe, chez cette espèce, un transporteur spécifique du péricarpe, mais qui est beaucoup plus distant de PA288TC1 (56 %).

V.6.3 Métabolisme du sorbitol

Bien que les feuilles matures d'abricotier exportent 65 à 75 % de carbone sous forme de sorbitol, celui-ci n'est qu'un composé minoritaire dans le fruit, comme chez la pêche (Brooks *et al.*, 1993) et il n'est pas accumulé chez la pomme (Marlow et Loescher, 1984).

Deux transporteurs de sorbitol ont été récemment identifiés et caractérisés chez la cerise (Gao *et al.*, 2003b), espèce qui accumule du sorbitol dans le fruit. Ces deux transporteurs ont été retrouvés dans les banques d'EST d'abricot et leurs profils d'expression ont été caractérisés. BGE003G06 est homologue à 93 % (région non codante incluse) à PcSOT1, et PA724TC1 est homologue (région non codante incluse) à 93 % à PcSOT2. Ils appartiennent tout deux à la famille multigénique nucléique 109N et ne présentent aucune homologie dans leurs régions non codantes respectives (moins de 25 %). Le profil d'expression de PA724TC1, qui chute en cours de maturation, est comparable à celui de PcSOT2. D'après les auteurs, ce transporteur serait impliqué dans l'accumulation de matière sèche pendant le grossissement des fruits. Par contre l'expression de BGE003G06 diminue pendant la maturation chez Bergeron contrairement à celle de PcSOT1 qui serait liée à l'accumulation de sorbitol dans la cerise (Gao *et al.*, 2003b). Cependant BGE003G06 est fortement sous-exprimé chez Bergeron par rapport aux deux autres variétés et il est possible que celle-ci présente un phénotype particulier pour le sorbitol. Ce sucre devrait être dosé, au moins chez Bergeron et Monique, pour confirmer l'implication de ce transporteur dans l'accumulation de sorbitol, qui n'est pas très claire chez l'abricotier. Un seul isogène de sorbitol déshydrogénase a été identifié dans les EST (PA1TC6), il est déjà connu chez la pêche (95 % d'homologie) (Yamada *et al.*, 2001) et son induction au cours du développement du fruit est identique dans les deux espèces. Cependant cette isoforme ne semble pas être responsable de toute l'activité sorbitol déshydrogénase, qui est forte dans les stades précoces chez la pêche comme chez la pomme (Nosarszewski, *et al.*, 2004). Il appartient à une famille multigénique assez large et 9 gènes de sorbitol déshydrogénase ont été identifiés dans les EST de pomme (Nosarszewski, *et al.*, 2004) dont certains exprimés dans les stades précoces.

V.6.4 Métabolisme des acides, transport des ions

Un des isogènes de malate déshydrogénase (PA2TC4) est induit pendant le développement, une protéine (spot 246) a été observée comme constitutive. Cependant il n'y a pas d'évidence pour que cette protéine soit le produit de la transcription de cet isogène.

Un unigène codant pour une sous-unité de canal potassique voltage-dépendant est induit pendant la maturation (PA310TC1). Il s'agit d'un homologue de la sous unité β du canal potassique KAB1 d'*Arabidopsis* (Tang *et al.*, 1995), dont l'expression n'a jamais été étudiée chez les fruits bien qu'on le retrouve dans les banques d'EST des principales espèces.

Ce canal potassique contient une activité oxydoréductase dont le substrat n'est pas encore connu (Gulbis *et al.*, 2000). Les sous unités β se lient à la région N-terminale cytosolique des canaux potassiques de type shaker, et exerceraient un rôle régulateur sur les propriétés des canaux potassiques (revue dans Very et Sentenac, 2003). Leur rôle précis demeure cependant largement méconnu car il a été peu étudié chez les plantes. Un canal potassique de type shaker présente un niveau d'expression qui augmente chez la vigne (Pratelli *et al.*, 2002).

Un transporteur mitochondrial de dicarboxylate/tricarboxylate est induit pendant le développement (PA341TC1). Il est très homologue (90 %) à deux membres d'une famille de transporteurs retrouvée chez le tabac (DTC3 et DTC4), qui sont capables de transporter un large spectre de molécules, dont la forme protonée du citrate et celle non protonée du malate (Picault *et al.*, 2002).

Le transporteur de dicarboxylate BGH011D24, homologue à 79 % au premier transporteur vacuolaire de malate, dont la fonction vient juste d'être identifié chez *Arabidopsis* (Emmerlich *et al.*, 2003), est présent dans les banques d'EST. L'expression des ARN codant pour ce canal anionique décroît non significativement au cours du développement de l'abricot. Il n'existe aucune référence sur son patron d'expression chez d'autres fruits, en liaison éventuelle avec l'évolution du rapport sucre/acide, mais plusieurs orthologues ont été également identifiés chez les agrumes et la vigne. Chez ces deux espèces, des modifications importantes de perméabilité membranaire aux protons accompagnent l'évolution du rapport sucre/acide (Terrier *et al.*, 2001b).

V.7 Autres gènes/protéines

Tous les gènes et les protéines dont l'expression évolue au cours de la maturation et entre les variétés ne sont pas reliés directement avec les critères de développement et de qualité du fruit que nous venons de développer. Il conviendra donc pour chaque cas particulier, de vérifier l'expression des gènes d'intérêt dans l'annexe B (CD). On retrouve par exemple des gènes reliés aux interactions avec les pathogènes, d'autres reliés à une signalétique cellulaire imprécise ou aux stress environnementaux, etc...

Dans ce chapitre, les protéines appartenant à des mécanismes où plusieurs spots protéiques ont été identifiés sont commentées, ainsi que quelques gènes de fonction inconnue dont l'expression est bien caractérisée par plusieurs approches.

V.7.1 Protéines identifiées

Deux groupes de protéines identifiées ne sont pas directement liés aux critères de qualité de l'abricot.

V.7.1.1 Antioxydation

La maturation des fruits a été décrite comme un phénomène oxydatif qui requiert une rotation des formes actives de l'oxygène (H_2O_2 , O_2^\bullet). Quatre enzymes impliquées dans ce mécanisme ont été identifiées. Après induction à 65 JAA, la catalase (spot 777, unigène PA145TC1) décroît légèrement avec un pic à la crise éthylénique. L'activité catalase mesurée chez la tomate (Jimenez *et al.*, 2002a) présente le même profil, l'expression des transcrits d'une catalase diminue aussi au cours de la maturation de l'abricot, cependant, ils correspondent à une autre isoforme (PA4TC5). Comme la catalase, la déshydroascorbate réductase (spot 889, unigène BGA002B03) décroît pendant la maturation avec un léger pic à la crise éthylénique, comme observé chez la tomate (Jimenez *et al.*, 2002a). La 6-phosphogluconate déshydrogénase (spot 394, unigène BGA010I02) est légèrement induite, avec un pic à la crise éthylénique. Il a été rapporté dans cette même étude que la quantité d' H_2O_2 augmente à la crise climactérique chez la tomate, ce qui pourrait expliquer le pic observé pour les trois enzymes régulées à ce stade. L'oxydation entraîne l'activation de Cys protéases dans les cellules de soja (Solomon *et al.*, 1999), elles interviennent dans l'exécution programmée de la mort cellulaire, qui a probablement lieu après la maturation des fruits. Le

niveau de la superoxyde dismutase détectée (SOD, spot 1115, PA37TC2) est constant, mais il n'y a aucune évidence pour que celle-ci soit reliée à la régulation des formes actives de l'oxygène, puisqu'elle semble être cytosolique et que ce mécanisme est mitochondrial. De plus une SOD mitochondriale différente de celle-ci, présente une activité qui augmente au cours de la maturation du poivron (Jimenez *et al.*, 2002b).

V.7.1.2 Métabolisme de l'azote

La présence d'arginase (spot 662) dans les stades précoces suggère que les fruits se développant peuvent mettre en place une stratégie identique à celle des graines en germination pour recruter de l'azote via l'arginine (Goldraij and Polacco, 2000). Cependant, le contenu en acides aminés des fruits est extrêmement variable entre les espèces et le stade de maturité (Boggio *et al.*, 2000), des investigations sont nécessaires pour confirmer un rôle central de l'arginine comme source majeure d'azote dans les abricots. L'inosine monophosphate déshydrogénase (spot 334, unigène PA1614TC1) catalyse l'étape limitante dans la biosynthèse *de novo* de la guanine (Weber, 1983) et a un rôle essentiel dans la fourniture des précurseurs de l'ADN et de l'ARN, ainsi que dans les voies de transduction du signal qui interviennent dans la différenciation cellulaire (Hager *et al.*, 1995). UMP/CMP kinase (UCK, spot 928, unigène PA130TC1) catalyse le transfert de phosphate entre l'ATP et les CMP, UMP, et dCMP, résultant dans la formation d'ADP et des nucléosides diphosphates correspondants qui sont requis pour la synthèse cellulaire des acides nucléiques (Sugino *et al.*, 1966). En dépit du rôle de ces deux enzymes dans la biosynthèse des acides nucléiques, il n'y a pas d'évidence d'un rôle possible dans la phase d'expansion cellulaire. De plus l'induction de l'UCK a lieu à la fin de cette phase (121 JAA).

V.7.2 Unigènes correspondant à des protéines inconnues ou non identifiées

Le transcrit inconnu PA19TC3 (protéine 1037), qui contient un domaine DUF538, et dont l'induction au cours du développement a été caractérisée par toutes les approches, montre une expression plus faible chez Blanc. Ce gène présente une forte homologie avec d'autres EST de fruit, principalement chez les rosacées, bien qu'on le retrouve aussi en quantité abondante chez les espèces ligneuses en général. Les seules protéines homologues connues sont retrouvées chez le riz et *Arabidopsis*, elle n'a par contre jamais été mise en évidence chez la tomate.

Un gène sans homologue présente une différence d'expression très élevée entre Blanc et les autres variétés. Le niveau d'expression de PA104TC1 augmente au cours de la maturation, ce gène est *a priori* totalement rosacée spécifique, et semble de plus très peu représenté en dehors des *Prunus*.

PA82TC1 (spot 519), dont l'abondance des transcrits et des protéines augmente pendant la maturation, est par contre ubiquitaire à travers le règne végétal, elle contient un domaine conservé (DUF 642) de fonction inconnue.

VI Bilan, Conclusions et Perspectives

VI.1 Bilan sur les outils mis en place

Le séquençage d'EST a permis d'acquérir une source d'information importante à propos du génome de l'abricotier. Il se situe ainsi au niveau de la 100^{ème} place chez les eucaryotes en nombre d'EST connues (15079 dont 99 % proviennent de ce travail), ce qui reflète sa position en terme de connaissance du génome. Parmi les espèces fruitières il occupe la sixième place en nombre d'EST derrière respectivement la pomme (156179 EST), la tomate (153775 EST), la vigne (142987 EST), l'orange (45192 EST) et la pêche (17812 EST). En terme de séquence unique, avec 5219 unigènes, l'abricotier occupe aussi la sixième position, mais avec 4546 unigènes absolument non redondants ancrés en 3' (ce qui correspond aux critères du NCBI et permet d'être beaucoup plus fiable pour discriminer les isogènes), il se situe à la quatrième place, loin derrière la vigne (12518 unigènes), mais dans un ordre de grandeur comparable à la pomme (6363) et la tomate (5061). Les connaissances sur le génome peuvent être ramenées à l'importance relative de la récolte. On aboutit ainsi à 168 tonnes/an par EST (484 tonnes/an/unigène TIGR, 562 tonnes/an/unigène NCBI) chez l'abricot, la seconde espèce en terme d'EST est la pomme (366 tonnes/an par EST) et en terme d'unigène NCBI, la vigne (4863 tonnes/an par unigène).

Les données relatives aux EST (séquences, catégories fonctionnelles, présence ou non de chimères dans les séquences, appartenance à une famille multigénique, occurrence par banque...) ont été regroupées dans une base de données fournie en annexe sur un support informatique. L'utilisation de cette base de données nécessite tout de même un minimum de familiarisation avec l'architecture AceDB et il serait nécessaire d'accroître sa convivialité afin de faciliter son accès aux personnes non expérimentées. Il serait de plus nécessaire de faciliter sa diffusion via internet, soit à partir d'un serveur indépendant, soit à partir d'une base de données déjà existante : il est difficilement envisageable de créer un index TIGR pour l'abricotier car un minimum de 50000 EST est nécessaire pour soumettre un projet ; la base de données sputnik (<http://sputnik.btk.fi/>) est conceptuellement très proche de celle mise au point pour l'abricotier au cours de ce travail, cependant elle est peu connue par la communauté scientifique. La solution la plus envisageable est l'intégration des EST d'abricotier sur la base de données du GDR (Genome Database for Rosacea, <http://www.genome.clemson.edu/gdr/>)

et les contacts sont déjà pris, cependant cette base de données ne prévoit pas l'intégration de la plupart des informations relatives aux séquences acquises dans ce travail.

A partir des données issues du séquençage des EST, il a été possible d'identifier un nombre élevé d'unigènes contenant un motif répété de type microsatellite (908). Ces données permettent, indépendamment du travail effectué pour identifier les gènes marqueurs de qualité, de les positionner sur les cartes génétiques existantes (travail en cours à l'UMR GAFL). Il a aussi été possible de réaliser un Northern électronique, cependant celui-ci nécessite un nombre élevé d'EST par gène pour que les profils d'expression de ces gènes puisse être quantifiés avec une précision suffisante.

Le développement de lames de microarray permet de disposer d'un outil permettant l'étude de l'expression des gènes avec un débit relativement élevé. Des motifs budgétaires ont conduit à éliminer de nombreux gènes des catégories fonctionnelles principalement liées au métabolisme de « base » des cellules végétales (métabolisme primaire, transcription, traduction...) et ces lames sont donc orientées sur la thématique qualité des fruits.

Un soin particulier a été apporté pour garantir la plus haute spécificité possible des oligonucléotides vis à vis d'un gène par rapport à la connaissance actuelle du transcriptome de l'abricot en privilégiant la zone 3' non codante. Le biais causé par les séquences homologues est donc éliminé, ainsi que l'éventualité d'avoir à re-synthétiser certains oligonucléotides lorsque s'accumuleront de nouveaux isogènes dans les familles déjà identifiées. Ce choix n'a pas été sans conséquence fâcheuse, puisque des gènes de fonction supposée importante pour la qualité n'ont pas été déposés sur les lames de microarray, pour cause d'absence de séquence non codante.

Quoi qu'il en soit cette technique a démontré une efficacité élevée (au niveau du débit et de la précision) concernant le criblage du transcriptome pour identifier les gènes potentiellement impliqués dans la qualité du fruit puisque la plupart d'entre eux a été obtenue avec cette approche. Le nombre de points retenus sur la cinétique semblent, à posteriori, devoir constituer un minimum, et il serait nécessaire de la compléter avec des répliques biologiques et une vérification par PCR quantitative.

La technique d'électrophorèse bidimensionnelle permet d'identifier des candidats dont il est alors acquis qu'ils sont effectivement traduits et qu'il existe une différence significative de leur abondance protéique entre deux conditions. Cependant, cette technique, utilisée ici pour caractériser le protéome total, est limitée à la quantification des gènes massivement traduits.

Par conséquent, elle permet de caractériser des mécanismes très « visibles » (photosynthèse, métabolisme des sucres...) d'intérêt parfois limité en ce qui concerne la qualité des fruits. Pour pouvoir identifier les régulations plus fines, il serait donc nécessaire de passer à l'étude de fractions protéiques plus restreintes (par exemple issues d'organites ou de purifications spécifiques). De plus l'identification des protéines est relativement difficile, surtout pour des organismes comme l'abricotier, pour lequel le génome n'est pas entièrement connu. La technique d'aide à l'identification mise en place permet cependant d'atténuer significativement ce problème.

VI.2 Bilan sur les candidats mis en évidence

Les technologies développées et utilisées dans cette étude (Northern électronique, microarray, gels d'électrophorèse 2D) ont permis d'obtenir un ensemble de gènes reliés potentiellement aux critères de qualité de l'abricot, soit 71 candidats. Les mutations ayant entraîné les variations d'expression de ces gènes entre les variétés portent donc sur les promoteurs, les facteurs de transcriptions ou la signalétique d'amont, plutôt que sur la séquence traduite et ses conséquences sur le plan de l'activité enzymatique. L'étape suivante consiste à valider ce lien et compte tenu des moyens disponibles sur l'abricotier (Dirlewanger *et al.*, 2004), il sera possible d'essayer de co-localiser ces candidats avec les QTL des critères de qualité respectifs.

Une partie des interprétations formulées au cours de la discussion pourrait être renforcée en réalisant des expériences complémentaires. Les unigènes repris dans les paragraphes suivants ont été retenus comme cibles potentielles de par leur profil d'expression.

VI.2.1 Dosage de composés

Le métabolome pourra être complète par le dosage de certains composés :

Métabolisme :

- Le sorbitol, afin de vérifier si l'expression différentielle de transporteurs, entre Moniqui et Bergeron, peut jouer un rôle dans une hypothétique accumulation différentielle de

ce sucre. Les marqueurs potentiellement impliqués dans le métabolisme du sorbitol sont les transporteurs BGE003G06 et PA724TC1 et la sorbitol déshydrogénase (PA1TC6).

- Des composés d'arômes, comme le méthyl benzoate, le méthyl salicylate, et le benzyl acétate, dont les transcrits correspondant aux enzymes catalysant leur synthèse sont différemment régulés entre les variétés, ou induits pendant la maturation de Bergeron. Les marqueurs identifiés comme potentiellement impliqués dans le métabolisme des arômes sont la S-adénosyl-L-méthionine:salicylique/benzoïque acide carboxyl méthyltransférase (BGE003M12), l'acétylCoA: benzyl alcohol acétyltransférase (PA614TC1), l'alcool acyltransférase (PA1425TC1) et l'alcool déshydrogénase (PA1442TC1).

Hormones :

- L'acide abscissique libre et lié, pour (i) valider le schéma proposé de la biosynthèse et la dégradation des caroténoïdes chez Monique, (ii) confirmer ou infirmer le lien possible entre une accumulation potentiellement précoce et élevée d'ABA et le dégagement précoce et élevé d'éthylène dans cette variété. Les marqueurs liés à la biosynthèse du β -carotène seraient la géranylgeranyl pyrophosphate synthase (PA1496TC1), la phytoène synthase (PA534TC1), la Zeta-carotène désaturase (PA1519TC1) et la lycopène β -cyclase (BGH008E09) et pour sa dégradation et la synthèse de l'ABA, les deux isogènes de 9-cis époxycaroténoïde dioxygénase (BGE004I02 et BGH001M22).

- L'auxine, l'acide jasmonique et la gibbérelline, afin de situer le moment exact de production au cours de la maturation. Ces données permettraient d'éclairer leur rôle respectif dans la régulation de la biosynthèse et de la transduction du signal éthylène chez l'abricot, et d'anticiper le transfert de connaissance des espèces modèles sur le fonctionnement de la communication croisée des hormones. Les marqueurs impliqués dans la régulation du signal auxine sont les AUX/IAA (PA790TC1, PA373TC1, BGE017N12, PA196TC1, PA173TC1) et les facteurs de réponse à l'auxine (BGH04I20 et BGA010O24). Les marqueurs impliqués dans la biosynthèse de l'acide jasmonique sont la lipoxigénase LOXD (BGA010C19), l'allène oxyde synthase (BGA007L04), l'oxophytodiénoate réductase (PA446TC1) et la jasmonate carboxyl méthyltransférase (BGE014G10). Les marqueurs impliqués dans la synthèse et la transduction du signal de la gibbérelline sont la gibbérelline-2-oxydase (BGA001O04) et les régulateurs RGA (BGA006E01 et BGH002P17).

Paroi :

- A un degré de priorité moindre, les pectines et les glycanes, car tout porte à croire que le mécanisme de la perte de fermeté est identique entre l'abricot et la pêche et les travaux sont

bien avancés sur cette espèce (Brummell *et al.*, 2004). L'intérêt de la connaissance de ces mécanismes chez l'abricot se situe plutôt au niveau de la rapidité de ce mécanisme, principalement après la récolte et se situerait donc au niveau de sa régulation. Chaque transcrite de fonction inconnue peut être un possible régulateur de la perte de fermeté. D'ores et déjà, les marqueurs potentiellement impliqués dans la dépolymérisation des pectines sont la pectine estérase (BGE016B06), les inhibiteurs de pectine estérase (PA43TC1 BGE005M19, PA23TC2, BGH007L07, PA7TC6, et PA5TC7), la pectate lyase (PA6TC8) et la polygalacturonase (PA274TC1). Les marqueurs potentiellement impliqués dans la dépolymérisation du réseau de xyloglucanes sont les XET (PA17TC1 et PA1110TC1), les expansines (PA5TC5, PA118TC1 et PA229TC1) et l'endo- β -1,4-glucanase (PA150TC1).

VI.2.2 Mesure de l'expression des gènes dans des variétés de phénotypes extrêmes

Des mesures d'expression globales des gènes pourront être effectuées sur d'autres variétés, afin d'exploiter au mieux la variabilité phénotypique des fruits au sein de l'espèce.

- La variété Iranien contient un taux de sucre extrêmement élevé (en particulier de monosaccharides), cela permettrait de mieux appréhender ce métabolisme. Les marqueurs potentiellement impliqués dans la biosynthèse du saccharose sont la saccharose synthase (PA349TC1), l'UDP-glucose pyrophosphorylase (PA91TC1) et le transporteur d'hexose (PA288TC1).

- L'acidité titrable de cette même variété est très faible et devient presque indétectable dans les fruits mûrs. A l'opposé, des variété comme Goldrich sont les plus acides. Les marqueurs potentiellement impliqués dans le métabolisme des acides sont la malate déshydrogénase (PA2TC4), la sous unité β du canal potassique KAB1 (PA310TC1), les transporteurs mitochondriaux de dicarboxylate/tricarboxylate (PA341TC1 et BGH011D24).

- La couleur de la pellicule des variétés dites « Double rouge », provient principalement des anthocyanes présentes et ces variétés sont des modèles bien adaptés pour étudier l'évolution et les facteurs limitant de la production d'anthocyanes dans les abricots. Les marqueurs potentiellement impliqués dans la biosynthèse des flavonoïdes sont la phénylalanine amonia lyase (BGA010H20), les chalcones synthases (BGA007H10, BGA009P18 et PA1546TC1), les flavonoïdes 3'-hydroxylases (PA671TC1 et PA760TC1), la dihydroflavonol-4-réductase (PA185TC1), la leucoanthocyanidine réductase (PA1407TC1),

les UDP-flavonol glucosyl transférases (BGE002J04, PA1238TC1, PA942TC1 et BGE018P10), la O méthyl-transférase (PA147TC2) et l'anthocyanidine-3-glucoside rhamnosyltransférase (PA4TC5).

VI.2.3 Mesure de l'expression des gènes après récolte ou après application d'hormones exogènes

La mesure de l'expression des gènes après récolte ou après application d'hormones exogènes pourrait être réalisée sur la variété Blanc. Ceci permettrait de confirmer la sous-expression du récepteur d'éthylène BGH004I19 qui, sur l'ensemble des 1800 transcrits dont l'expression a été évaluée, est le candidat le plus probable pour expliquer l'absence de réponse à l'éthylène dans cette variété. Les autres marqueurs potentiellement impliqués dans la biosynthèse et la transduction du signal éthylène sont l'ACS1 (AF184076), l'ACO (AF026793), la dioxygénase PA322TC1, la CDPK (BGE003L14), la protéine E8 (AF13950), les ERF (BGA011O8 et PA835TC1), la protéine de réponse au stress (BGE004M12) et les protéines contenant un domaine AP2 (PA1510TC1, BGE014K05, PA1386TC1).

VI.2.4 Détermination de la fonction des protéines inconnues

Il serait important de déterminer la fonction des gènes inconnus (80 sont différentiellement régulés pendant le développement dont 36 pour lesquels aucune protéine homologue n'est connue au NCBI). Au préalable d'investigations lourdes qui ne peuvent être raisonnées que sur quelques cas, il est nécessaire de d'acquérir le maximum d'information à l'aide d'approches systématiques ou faiblement coûteuses en temps afin de hiérarchiser les gènes de fonction inconnue en terme de pertinence et de « faisabilité » :

- La concomitance spatio-temporelle de l'expression d'un gène avec un phénomène d'intérêt pour la qualité du fruit devra être étendue. Dans ce travail, nous avons essentiellement focalisé les études d'expression sur le développement du fruit. Ces études d'expression pourront être étendues à d'autres expériences comme celles proposées au chapitre précédent. Examiner la spécificité d'expression tissulaire de ces gènes chez l'abricot ne pose aucune difficulté méthodologique. L'analyse en composantes principales de 486 expériences de microarrays portant sur 20000 gènes a révélé que la spécificité d'expression tissulaire

expliquait 75 % de la variance globale chez *Arabidopsis*. Des gènes très mal annotés, spécifiques des étamines, se positionnent par exemple dans un bloc chromosomique récemment dupliqué (Lohman *et al.*, 2004). La mise en évidence de tels domaines chromosomiques co-régulés pourrait revêtir une importance capitale en ce qui concerne le péricarpe charnu, du point de vue fonctionnel, mais également évolutif.

- Ce premier crible permettra de sélectionner les gènes pour lesquels il conviendra d'obtenir la séquence complète du gène, et de sa région promotrice ; l'état de sa conservation dans les espèces et les régions clé fortement conservées pourront être appréhendés. La comparaison des transcriptomes « fruits » de modèles émergents en génomique fonctionnelle (raisin, pomme, citrus) et même chez les *Prunus* (pêche, amande) permettrait d'identifier des liens entre mutations et fonctions génériques dans cet organe (Rudd *et al.*, 2003).

- Afin de maximiser les preuves d'une traduction, la protéine sera détectée à l'aide d'anticorps en protéomique, les éventuelles protéines homologues et leur profil d'abondance seront recensés. Le niveau de conservation entre des espèces éloignées devrait renseigner sur la traduction en protéine, si on suppose que la conservation d'un domaine régulateur et de la séquence pour un transcrit non traduit a peu de chance d'être maintenue.

- Pour les candidats ainsi sélectionnés, la poursuite de l'identification fonctionnelle peut s'avérer beaucoup plus lourde. Une première approche serait de phénotyper, lorsqu'ils existent et sont facilement repérables, les mutants correspondants des banques d'insertion d'ADN-T ou de Tilling des espèces dites modèles (<http://www.arabidopsis.org/links/insertion.jsp>). Cette approche présuppose que les mutations n'ont pas entraîné de néofonctionnalisation des gènes. Elle est clairement impossible pour les unigènes qui semblent spécifiques de l'abricotier ou autres plantes ligneuses. Une autre approche serait de sur-exprimer les gènes de l'abricot chez les espèces modèles rapidement transformables et régénérables, ce qui sous-entend que les produits de ces gènes trouvent des partenaires dans le métabolome ou l'interactome de l'espèce modèle considérée. Il peut être envisageable (i) d'étudier l'interactome *in vitro* par les techniques de simple ou double hybride (ii) de produire les protéines dans des systèmes hétérologues pour étudier l'activité enzymatique.

VI.3 Poursuite du travail sur les « candidats qualité »

La finalité de ce travail est

- d'une part sur le plan académique, de mieux appréhender la physiologie de la maturation de l'abricot et des fruits d'une manière générale ;
- d'autre part de poursuivre le cheminement global vers la sélection assistée par marqueurs, via la co-localisation des candidats avec les QTL de qualité.

Il apparaît donc nécessaire de souligner les limitations auxquelles ces types d'approches systématiques sont soumises.

Dans ce travail, nous avons supposé que les protéines codées par les transcrits ont la même fonction que leurs homologues, ce qui devra être vérifié au cas par cas. Toutes les annotations fonctionnelles sont réalisées par alignement des séquences d'abricot avec les bases de données publiques contenant des protéines/transcrits dont le rôle n'est pas toujours validé, et les orthologues (issus de spéciation et dont les fonctions sont conservées) sont très mal distingués des paralogues (séquences issues de duplications à l'intérieur d'un même génome, donc susceptibles de modifications fonctionnelles après divergence).

Des procédures bioinformatiques dites de phylogénomique permettent d'estimer la fiabilité de l'assignation des orthologues (Zmasek et Eddy, 2002) et de détecter les premiers représentants de nouvelles sous-familles protéiques. Ce type de procédure bénéficiera de l'intégration d'EST d'autres *Prunus* (pêche, amande, cerise...) et d'espèces fruitières au sens large. Des logiciels intègrent également des informations de cartographie physique et de syntonie (Cannon et Young, 2003) pour décrire l'évolution d'une famille génique dans un génome unique. L'ouverture au domaine public du génome du peuplier au moment où sont rédigées ces lignes (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>) permettrait de clarifier le statut d'un certain nombre de ces gènes entre génomes séquencés.

Plusieurs types de mutations peuvent affecter l'expression d'un gène supposé contrôler certains critères de qualité. Le cas le plus favorable pour la co-localisation effective d'un gène avec un QTL relié à ce phénotype est que la mutation intervienne dans son promoteur. Les mutations affectant la fonction d'une protéine (en aval du promoteur) entre deux variétés ne sont pas repérées dans notre type d'étude. Si la mutation a lieu sur une régulation située en amont dans la cascade signalétique la co-localisation devient beaucoup plus complexe. Ainsi, si l'observation que l'expression de l'UFGT est systématiquement reliée au critère baies colorées est assez ancienne (Boss *et al.*, 1996), il a fallu près de 10 ans pour mettre en

évidence le fait que le facteur de transcription qui contrôle son expression co localise avec le QTL couleur chez le raisin (Kobayashi, S. *et al.*, 2004 ; This, non publié).

En ce qui concerne les gènes régulés au cours du développement, il est probable que leur expression soit sous contrôle de régulateurs dont l'expression n'a pas été évaluée. Il est aussi possible que leur rôle dans la cascade signalétique n'ait pas encore été mis en évidence ou qu'ils soient sujets à des régulations post-traductionnelles. Ces régulateurs « inconnus » peuvent être des candidats mieux appropriés et devront être recherchés dans la mesure du possible. Du polymorphisme pourra être recherché dans les gènes de fonction et les gènes régulateurs (génotypage, recherche de SNPs).

Ces observations montrent qu'il n'existe aucune certitude en l'état actuel du travail permettant d'affirmer que tel ou tel candidat est effectivement un marqueur de qualité. La co-localisation sur des QTL pour 71 candidats est assez lourde à entreprendre, mais reste envisageable ; il est de plus assez facile de dégager une rapide hiérarchisation du niveau de pertinence entre ces 71 unigènes pour déterminer les 10 ou 15 unigènes potentiellement impliqués dans les cibles à atteindre (allongement de la conservation, modification de la couleur ou du profil aromatique...).

VII Références Bibliographiques

- Abbal, P. et Planton G** 1990. La mesure objective de la fermeté des fruits et légumes, un pénétromètre électronique automatique et programmable, le système Pénélaup. *Infos ctifl*, **62**, 27-30.
- Abdi, N., Holford, P. et McGlasson, B.** 2002. Application of two-dimensional gel electrophoresis to detect proteins associated with harvest maturity in stonefruit. *Post Biol. Tech.*, **26**, 1-13.
- Abel, S., Nguyen, M.D. et Theologis, A.** 1995. The PS-IAA4/5-like family of early auxin-inducible mRNAs in *Arabidopsis thaliana*. *J. Mol. Biol.*, **251**, 533-49.
- Achard, P., Vriezen, W.H., Van Der Straeten, D. et Harberd, N.P.** 2003. Ethylene regulates *Arabidopsis* development via the modulation of DELLA protein growth repressor function. *Plant Cell*, **15**, 2816-25.
- Adams-Phillips, L., Barry, C. et Giovannoni, J.** 2004a. Signal transduction systems regulating fruit ripening. *Trends Plant Sci.*, **9**, 331-8.
- Adams-Phillips, L., Barry, C., Kannan, P., Leclercq, J., Bouzayen, M. et Giovannoni, J.** 2004b. Evidence that CTR1-mediated ethylene signal transduction in tomato is encoded by a multigene family whose members display distinct regulatory features. *Plant Mol. Biol.*, **54**, 387-404.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B. et Moreno, R.F.** 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651-6.
- Aharoni, A., Keizer, L.C., Van Den Broeck, H.C., Blanco-Portales, R., Munoz-Blanco, J., Bois, G., Smit, P., De Vos, R.C. et O'Connell, A.P.** 2002. Novel insight into vascular, stress, and auxin-dependent and -independent gene expression programs in strawberry, a non-climacteric fruit. *Plant Physiol.*, **129**, 1019-31.
- Aharoni, A., Verhoeven, H.A., Luecker, J., O'Connell, A.P. et Van Tunen, A.J.** 2000. Fruit flavour related genes and use thereof. *Patent: WO 0032789-A*.
- Alba, R., Cordonnier-Pratt, M.M. et Pratt, L.H.** 2000. Fruit-localized phytochromes regulate lycopene accumulation independently of ethylene production in tomato. *Plant Physiol.*, **123**, 363-70.
- Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. et Suzuki, M.** 1998. A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.*, **17**, 5484-96.
- Almela, L., Fernandez-Lopez, J.A. et Roca, M.J.** 2000. High-performance liquid chromatographic screening of chlorophyll derivatives produced during fruit storage. *J. Chromatogr. A.*, **870**, 483-9.
- Alonso, J.M., Chamarro, J. et Granell, A.** 1995. Evidence for the involvement of ethylene in the expression of specific RNAs during maturation of the orange, a non-climacteric fruit. *Plant Mol. Biol.*, **29**, 385-90.
- Alonso, J.M., Hirayama, T., Roman, G., Nourizadeh, S. et Ecker, J.R.** 1999. EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science*, **284**, 2148-52.
- Alós, E., Cercos, M., Iglesias, D.J., Tadeo, F.R. et Talon, M.** 2003. Regulation of colour change in *Citrus*: effects of gibberellin, nitrate and ethylene on phytoene synthase expression in flavedo tissue during fruit ripening. *7th International congress of plant molecular biology* Barcelona, June 23-28.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. et Lipman, D.J.** 1990. Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-10.

- Anderberg, R.J.** et **Walker-Simmons, M.K.** 1992. Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. *Proc. Natl. Acad. Sci. U S A*, **89**, 10183-7.
- Andrawis, A., Solomon, M.** et **Delmer, D.P.** 1993. Cotton fiber annexins: a potential role in the regulation of callose synthase. *Plant J.*, **3**, 763-72.
- Arabidopsis Genome Initiative.** 2000. Analyse of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **403**, 503-511.
- Audergon, J.M., Duffillol, M., Souty, M., Breuils, L.** et **Reich, M.** 1991. Biochemical and physicochemical characterisation of 400 apricot varieties. Consequences in the apricot selection and improvement process. *Acta Hort.*, **293**, 111-119.
- Awad, M.A.** et **De Jager, A.** 2002. Formation of flavonoïds, especially anthocyanin and chlorogenic acid in 'Jonagold' apple skin: influences of growth regulators and fruit maturity. *Scientia Hort.*, **93**, 257-266.
- Bachem, C.W., Van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M.** et **Visser, R.G.** 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.*, **9**, 745-53.
- Badenes, M.L., Asins, M.J., Carbonell, E.A.** et **Glacer, G.** 1996. Genetic diversity in apricot, *Prunus armeniaca L.* aimed at improvement of resistance to plum pox virus. *Plant Breed.*, **115**, 133-139.
- Balbi, V.** et **Lomax, T.L.** 2003. Regulation of early tomato fruit development by the diageotropica gene. *Plant Physiol.*, **131**, 186-97.
- Bartel, B.** 1997. Auxin biosynthesis. *Annu. Rev. Plant. Physiol. Plant Mol. Biol.*, **48**, 51-66.
- Bassett, C.L., Artlip, T.S.** et **Callahan, A.M.** 2002. Characterization of the peach homologue of the ethylene receptor, PpETR1, reveals some unusual features regarding transcript processing. *Planta*, **215**, 679-88.
- Baugh, L.R., Hill, A.A., Brown, E.L.** et **Hunter, C.P.** 2001. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.*, **29**, E29.
- Beaudoin, N., Serizet, C., Gosti, F.** et **Giraudat, J.** 2000. Interactions between abscisic acid and ethylene signaling cascades. *Plant Cell*, **12**, 1103-15.
- Benitez-Burraco, A., Blanco-Portales, R., Redondo-Nevado, J., Bellido, M.L., Moyano, E., Caballero, J.L.** et **Munoz-Blanco, J.** 2003. Cloning and characterization of two ripening-related strawberry (*Fragaria x ananassa* cv. Chandler) pectate lyase genes. *J. Exp. Bot.*, **54**, 633-45.
- Bird, C.R., Ray, J.A., Fletcher, J.D., Boniwell, J.M., Bird, A.S., Teulieres, C., Blain, I., Bramley, P.M.** et **Schuch, W.** 1991. Using antisense RNA to study gene function: inhibition of carotenoid biosynthesis in transgenic tomatoes. *BioTechnology*, **9**, 635-639.
- Boggio, S.B., Palatnik, J.F., Heldt, H.W.** et **Valle, E.M.** 2000. Changes in amino acid composition and nitrogen metabolizing enzymes in ripening fruits of *Lycopersicon esculentum* Mill. *Plant Sci.*, **159**, 125-133.
- Boguski, M.S., Lowe, T.M.** et **Tolstoshev, C.M.** 1993. dbEST--database for expressed sequence tags. *Nat. Genet.*, **4**, 332-3.
- Boss, P.K., Davies, C.** et **Robinson, S.P.** 1996. Expression of anthocyanin biosynthesis pathway genes in red and white grapes. *Plant Mol. Biol.*, **32**, 565-9.
- Botondi, R., DeSantis, D., Bellincontro, A., Vizovitis, K.** et **Mencarelli, F.** 2003. Influence of ethylene inhibition by 1-methylcyclopropene on apricot quality, volatile production, and glycosidase activity of low- and high-aroma varieties of apricots. *J. Agric. Food Chem.*, **51**, 1189-200.

- Bouvier, F., Backhaus, R.A. et Camara, B.** 1998. Induction and control of chromoplast-specific carotenoid genes by oxidative stress. *J. Biol. Chem.*, **273**, 30651-9.
- Bradford, M.M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248-54.
- Brooks, S.J., Moore, J.N. et Murphy, J.B.** 1993. Quantitative and qualitative changes in sugar content of peach genotypes [*Prunus persica* (L. Batsch.)]. *J. Am. Soc. Hort. Sci.*, **118**, 97-100.
- Brown, R.L., Kazan, K., McGrath, K.C., Maclean, D.J. et Manners, J.M.** 2003. A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of *Arabidopsis*. *Plant Physiol.*, **132**, 1020-32.
- Brummell, D.A., Catala, C., Lashbrook, C.C. et Bennett, A.B.** 1997. A membrane-anchored E-type endo-1,4-beta-glucanase is localized on Golgi and plasma membranes of higher plants. *Proc. Natl. Acad. Sci. U S A*, **94**, 4794-9.
- Brummell, D.A., Dal Cin, V., Crisosto, C.H. et Labavitch, J.M.** 2004. Cell wall metabolism during maturation, ripening and senescence of peach fruit. *J. Exp. Bot.*, **55**, 2041-52.
- Bunney, T.D., Van Walraven, H.S. Et de Boer, A.H.** 2001. 14-3-3 protein is a regulator of the mitochondrial and chloroplast ATP synthase. *Proc. Natl. Acad. Sci. U S A*, **98**, 4249-54.
- Burbidge, A., Grieve, T.M., Jackson, A., Thompson, A., McCarty, D.R. et Taylor, I.B.** 1999. Characterization of the ABA-deficient tomato mutant notabilis and its relationship with maize Vp14. *Plant J.*, **17**, 427-31.
- Cakir, B., Agasse, A., Gaillard, C., Saumonneau, A., Delrot, S. et Atanassova, R.** 2003. A grape ASR protein involved in sugar and abscisic acid signaling. *Plant Cell*, **15**, 2165-80.
- Calvin, M.** 1962. The path of carbon in photosynthesis. *Science*, **135**, 879-89.
- Camardella, L., Carratore, V., Ciardiello, M.A., Servillo, L., Balestrieri, C. et Giovane, A.** 2000. Kiwi protein inhibitor of pectin methylesterase amino-acid sequence and structural importance of two disulfide bridges. *Eur. J. Biochem.*, **267**, 4561-5.
- Campbell, P. et Braam, J.** 1999. Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. *Trends Plant Sci.*, **4**, 361-6.
- Cannon, SB. et Young, N.D.** 2003. Orthoparamap: Distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. *BMC bioinformatics*, **4**, 35.
- Carpita, N.C. et Gibeaut, D.M.** 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.*, **3**, 1-30.
- Castillejo, C., de la Fuente, J.I., Iannetta, P., Botella, M.A. et Valpuesta, V.** 2004. Pectin esterase gene family in strawberry fruit: study of FaPE1, a ripening-specific isoform. *J. Exp. Bot.*, **55**, 909-18.
- Chairote, G., Rodriguez, F., Seck, S. et Crouzet, J.** 1981. Characterization of additional volatile flavor components of apricot. *J. Food Sci.*, **45** 1898.
- Chambroy, Y., Souty, M., Audergon, J.M., Jacquemin, G. et Gomez, R.M.** 1995. Researches on the suitability of modified atmosphere packaging for shelf-life and quality improvement of apricot fruit. *Acta Hort.*, **384**, 633-638.
- Chen, M.H., Sheng, J., Hind, G., Handa, A.K. et Citovsky, V.** 2000. Interaction between the tobacco mosaic virus movement protein and host cell pectin methylesterases is required for viral cell-to-cell movement. *EMBO J.*, **19**, 913-20.

- Chen, Y.F., Randlett, M.D., Findell, J.L. et Schaller, G.E.** 2002. Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of *Arabidopsis*. *J. Biol. Chem.*, **277**, 19861-6.
- Chengappa, S., Guilleroux, M., Phillips, W. et Shields, R.** 1999. Transgenic tomato plants with decreased sucrose synthase are unaltered in starch and sugar accumulation in the fruit. *Plant Mol. Biol.*, **40**, 213-21.
- Cheong, J.J. et Choi, Y.D.** 2003. Methyl jasmonate as a vital substance in plants. *Trends Genet.*, **19**, 409-13.
- Chernys, J.T. et Zeevaart, J.A.** 2000. Characterization of the 9-cis-epoxycarotenoid dioxygenase gene family and the regulation of abscisic acid biosynthesis in avocado. *Plant Physiol.*, **124**, 343-53.
- Chevalier, T., de Rigal, D., Mbéguie-A-Mbéguie, D., Gaillard, F., Richard-Forget, F. et Fils-Lycaon, B.R.** 1999. Molecular cloning and characterization of apricot fruit polyphenol oxidase. *Plant Physiol.*, **119**, 1261-70.
- Choi, C., Toivonen, P., Wiersma, P.A. et Kappel, F.** 2004. Effect of gibberellic acid during development of sweet cherry fruit: physiological and molecular changes. *Acta Hort.*, **636**, 489-495.
- Civello, P.M., Powell, A.L., Sabehat, A. et Bennett, A.B.** 1999. An expansin gene expressed in ripening strawberry fruit. *Plant Physiol.*, **121**, 1273-80.
- Clark, G.B., Sessions, A., Eastburn, D.J. et Roux, S.J.** 2001. Differential expression of members of the annexin multigene family in *Arabidopsis*. *Plant Physiol.*, **126**, 1072-84.
- Cleveland, W.S.** 1979. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.*, **74**, 829-836.
- Cohen, J.D.** 1996. In vitro tomato fruit cultures demonstrate a role for indole-3-acetic acid in regulating fruit ripening. *J. Am. Soc. Hort. Sci.*, **121**, 520-524.
- Costa, M.L., Civello, P.M., Chaves, A.R. et Martínez, G.A.** 2002. Characterization of Mg-dechelataase activity obtained from *Fragaria × ananassa* fruit. *Plant Physiol. Biol.*, **40**, 111-118.
- Covitz, P.A., Smith, L.S. et Long, S.R.** 1998. Expressed sequence tags from a root-hair-enriched *medicago truncatula* cDNA library. *Plant Physiol.*, **117**, 1325-32.
- Creelman, R.A. et Mullet, J.E.** 1997. Biosynthesis and action of jasmonates in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **48**, 355-381.
- Curtin, C., Zhang, W. et Franco, C.** 2003. Manipulating anthocyanin composition in *Vitis vinifera* suspension cultures by elicitation with jasmonic acid and light irradiation. *Biotechnol. Lett.*, **25**, 1131-5.
- Czapski, J. et Saniewski, M.** 1992. Stimulation of ethylene production and ethylene-forming enzyme in fruits of the non-ripening nor and rin tomato mutants by methyl jasmonate. *J. Plant Physiol.*, **139**, 265-268.
- D'Aoust, M.A., Yelle, S. et Nguyen-Quoc, B.** 1999. Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. *Plant Cell*, **11**, 2407-18.
- Davies, C. et Robinson, S.P.** 2000. Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. *Plant Physiol.*, **122**, 803-12.
- Davies, C., Boss, P.K. et Robinson, S.P.** 1997. Treatment of grape berries, a nonclimacteric fruit with a synthetic auxin, retards ripening and alters the expression of developmentally regulated genes. *plant Physiol.*, **115**, 1155-1161.

- Del Campillo, E.** 1999. Multiple endo-1,4-beta-D-glucanase (cellulase) genes in *Arabidopsis*. *Curr. Top. Dev. Biol.*, **46**, 39-61.
- De Vicente, M.C., Truco, M.J., Egea, J., Burgos, L. et Arus, P.** 1998. RFLP variability in apricot (*Prunus armeniaca* L.). *Plant Breed.*, **117**, 153-158.
- Devoto, A. et Turner, J.G.** 2003. Regulation of jasmonate-mediated plant responses in *Arabidopsis*. *Ann. Bot. (Lond)*, **92**, 329-37.
- Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. et Siebert, P.D.** 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U S A*, **93**, 6025-30.
- Didier, G., Brezellec, P., Remy, E. et Henaut, A.** 2002. GeneANOVA--gene expression analysis of variance. *Bioinformatics*, **18**, 490-1.
- Dirlewanger, E., Graziano, E., Joobeur, T., Garriga-Caldere, F., Cosson, P., Howad, W. et Arus, P.** 2004. Comparative mapping and marker-assisted selection in *Rosaceae* fruit crops. *Proc. Natl. Acad. Sci. U S A*, **101**, 9891-6.
- Dormann, P. et Benning, C.** 1998. The role of UDP-glucose epimerase in carbohydrate metabolism of *Arabidopsis*. *Plant J.*, **13**, 641-52.
- Dudareva, N., D'Auria, J.C., Nam, K.H., Raguso, R.A. et Pichersky, E.** 1998. Acetyl-CoA:benzylalcohol acetyltransferase, an enzyme involved in floral scent production in *Clarkia breweri*. *Plant J.*, **14**, 297-304.
- Dudareva, N., Murfitt, L.M., Mann, C.J., Gorenstein, N., Kolosova, N., Kish, C.M., Bonham, C. et Wood, K.** 2000. Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. *Plant Cell*, **12**, 949-61.
- Duncan, J.D. et West, C.A.** 1981. Properties of kaurene synthetase from *Marah macrocarpus* endosperm. Evidence for the participation of separate but interacting enzymes. *Plant Physiol.*, **68**, 1128-1134.
- Durbin, R. et Thierry Mieg, J.** 1991. A *C. elegans* Database. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. et Coleman, P.** 1992. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. U S A*, **89**, 3010-4.
- El-Kereamy, A., Chervin, C., Roustan, J.P., Cheynier, V., Souquet, J.M., Moutounet, M., Raynal, J., Ford, C.M., Latche, A., Pech, J.C. et Bouzayen, M.** 2003. Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. *Physiol. Plant Sci.*, **119**, 175-182.
- El-Sharkawy, I., Jones, B., Li, Z.G., Lelievre, J.M., Pech, J.C. et Latche, A.** 2003. Isolation and characterization of four ethylene perception elements and their expression during ripening in pears (*Pyrus communis* L.) with/without cold requirement. *J. Exp. Bot.*, **54**, 1615-25.
- Emmerlich, V., Linka, N., Reinhold, T., Hurth, M.A., Traub, M., Martinoia, E. et Neuhaus, H.E.** 2003. The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *Proc. Natl. Acad. Sci. U S A*, **100**, 11122-6.
- Etienne, C., Rothan, C., Moing, A., Plomion, C., Bodenes, C., Svanella-Dumas, L., Cosson, P., Pronier, V., Monet, R. et Dirlewanger, E.** 2002. Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. *Theor. Appl. Genet.*, **105**, 145-159.

- Ewing, B.** et **Green, P.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186-94.
- Ewing, B., Hillier, L., Wendl, M.C.** et **Green, P.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175-85.
- Ewing, R.M., Ben Kahla, A., Poirot, O., Lopez, F., Audic, S.** et **Claverie, J.M.** 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.*, **9**, 950-9.
- Fan, X., Mattheis, J.P.** et **Fellman, J.K.** 1997. Effect of methyl jasmonate on ethylene and volatile production by summered apples depends on fruit developmental stage. *J. Agric. Food Chem.*, **45**, 208-211.
- Fan, X., Mattheis, J.P.** et **Fellman, J.K.** 1998. Responses of apples to postharvest jasmonate treatments. *J. Am. Soc. Hort. Sci.*, **123**, 421-425.
- Faust, M., Suranyi, D.,** et **Nyujto, F.** 1998. Origin and dissemination of apricot. *Hort. Rev.*, **22**, 225-266.
- Fernandes, J., Brendel, V., Gai, X., Lal, S., Chandler, V.L., Elumalai, R.P., Galbraith, D.W., Pierson, E.A.** et **Walbot, V.** 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol.*, **128**, 896-910.
- Fils-Lycaon, B.R., Wiersma, P.A., Eastwell, K.C.** et **Sautiere, P.** 1996. A cherry protein and its gene, abundantly expressed in ripening fruit, have been identified as thaumatin-like. *Plant Physiol.*, **111**, 269-73.
- Fisher, R. A.** 1954. *Statistical Methods for Research Workers*. 12th ed. Hafner, New York, pp. 362.
- Fonseca, S, Hackler, L., Zvara, A., Ferreira, S., Baldé, A., Dudits, D., Pais, M.** et **Puskas, G.** 2004. Monitoring gene expression along pear fruit development, ripening and senescence using cDNA microarrays. *Plant Sci.*, **167**, 457-469.
- Fraser, P.D., Kiano, J.W., Truesdale, M.R., Schuch, W.** et **Bramley, P.M.** 1999. Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. *Plant Mol. Biol.*, **40**, 687-98.
- Fry, S.C., Smith, R.C., Renwick, K.F., Martin, D.J., Hodge, S.K.** et **Matthews, K.J.** 1992. Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem. J.*, **282**, 821-8.
- Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H.** et **Ohme-Takagi, M.** 2000. *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell*, **12**, 393-404.
- Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D.** et **Burstin, J.** 2003. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.*, **133**, 664-82.
- Gao, Z., Chen, Y.F., Randlett, M.D., Zhao, X.C., Findell, J.L., Kieber, J.J.** et **Schaller, G.E.** 2003a. Localization of the Raf-like kinase CTR1 to the endoplasmic reticulum of *Arabidopsis* through participation in ethylene receptor signaling complexes. *J. Biol. Chem.*, **278**, 34725-32.
- Gao, Z., Maurousset, L., Lemoine, R., Yoo, S.D., van Nocker, S.** et **Loescher, W.** 2003b. Cloning, expression, and characterization of sorbitol transporters from developing sour cherry fruit and leaf sink tissues. *Plant Physiol.*, **131**, 1566-75.
- Genard, M.** et **Souty, M.** 1996. Modeling the peach sugar contents in relation to fruit growth. *J. Am. Soc. Hort. Sci.*, **121**, 1122-1131.

- Giovane, A., Balestrieri, C., Quagliuolo, L., Castaldo, D. et Servillo, L.** 1995. A glycoprotein inhibitor of pectin methylesterase in kiwi fruit. Purification by affinity chromatography and evidence of a ripening-related precursor. *Eur. J. Biochem.*, **233**, 926-9.
- Giovannoni, J., Yen, H., Shelton, B., Miller, S., Vrebalov, J., Kannan, P., Tieman, D., Hackett, R., Grierson, D. et Klee, H.** 1999. Genetic mapping of ripening and ethylene-related loci in tomato. *Theor. Appl. Genet.*, **98**, 1005-1013.
- Giovannoni, J.J., DellaPenna, D., Bennett, A.B. et Fischer, R.L.** 1989. Expression of a chimeric polygalacturonase gene in transgenic rin (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. *Plant Cell*, **1**, 53-63.
- Glonek, G.F. et Solomon, P.J.** 2004. Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, **5**, 89-111.
- Goldraj, A. et Polacco, J.C.** 2000. Arginine degradation by arginase in mitochondria of soybean seedling cotyledons. *Planta*, **210**, 652-8.
- Gomez, E. et Ledbetter, C.A.** 1997. Development of volatile compounds during fruit maturation: characterization of apricot and plum-apricot hybrids. *J. Sci. Food Agric.*, **74**, 541-546.
- González-Carranza, Z.H, Lozoya-Gloria, E. et Roberts, J.A.** 1998. Recent developments in abscission: shedding light on the shedding process. *Trends Plant Sci.*, **3**, 10-14.
- Gray, J.E., Picton, S., Giovannoni, J.J. et Grierson, D.** 1994. The use of transgenic and naturally-occurring mutants to understand and manipulate tomato fruit ripening. *Plant Cell Environ.*, **17**, 557-571.
- Greenbaum, D., Jansen, R. et Gerstein, M.** 2002. Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics*, **18**, 585-96.
- Gross, J., Bazak, H., Blumenfeld, A. et Ben-Arie, R.** 1984. Changes in chlorophyll and carotenoid pigments in the peel of 'Triumph' persimmon (*Diospyros kaki* L.) induced by pre-harvest gibberellin (GA3) treatment. *Scientia Hort.*, **24**, 305-314.
- Grotte, M., Cadot, Y., Poussier, A., Loonis, D., Piétri, E., Duprat, F. et Barbeau, G.** 2001. Détermination du degré de maturité des baies de raisin par des mesures physiques : aspects méthodologiques. *J. Int. Sci. Vigne Vin*, **35**, 87-98.
- Guichard, E. et Souty, M.** 1988. Comparison of the relative quantities of aroma compounds found in fresh apricot (*Prunus armeniaca*) from six different varieties. *Z. Lebensm Unters. F. A.*, **186**, 301-307.
- Guilfoyle, T.J. et Hagen, G.** 2001. Auxin response factors. *J. Plant Growth Regul.*, **10**, 281-291.
- Guis, M., Botondi, R., Ben-Amorn M., Ayub, R., Bouzayen, M., Pech, J.C. et Latché A.** 1997. Ripening-associated biochemical traits of cantaloupe charentais melons expressing an antisense ACC oxidase transgene. *J. Am. Soc. Hort. Sci.*, **122**, 748-751.
- Gulbis, J.M., Zhou, M., Mann, S. et MacKinnon, R.** 2000. Structure of the cytoplasmic beta subunit-T1 assembly of voltage-dependent K⁺ channels. *Science*, **289**, 123-7.
- Guo, H. et Ecker, J.R.** 2003. Plant responses to ethylene gas are mediated by SCF (EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. *Cell*, **115**, 667-77.
- Guo, H. et Ecker, J.R.** 2004. The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.*, **7**, 40-9.
- Hadfield, K.A., Rose, J.K., Yaver, D.S., Berka, R.M. et Bennett, A.B.** 1998. Polygalacturonase gene expression in ripe melon fruit supports a role for polygalacturonase in ripening-associated pectin disassembly. *Plant Physiol.*, **117**, 363-73.

- Hagen, G.** et **Guilfoyle, T.** 2002. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.*, **49**, 373-85.
- Hagen, S., Khadari, B., Lambert, P.** et **Audergon, J.M.** 2002. Genetic diversity in apricot revealed by AFLP markers: species and cultivar comparisons. *Theor. Appl. Genet.*, **105**, 298-305.
- Hager, P.W., Collart, F.R., Huberman, E.** et **Mitchell, B.S.** 1995. Recombinant human inosine monophosphate dehydrogenase type I and type II proteins. Purification and characterization of inhibitor binding. *Biochem. Pharmacol.*, **49**, 1323-9.
- Hames, B.O.** et **Rickwood, B.** 1990. One dimensional polyacrylamide gel electrophoresis. In: Hames, B.O. et Rickwood, B., (Eds) *Gel Electrophoresis of Proteins, a practical approach*. 2nd ed. Oxford University Press, New York, pp. 1-147.
- Harpster, M.H., Brummell, D.A.** et **Dunsmuir, P.** 1998. Expression analysis of a ripening-specific, auxin-repressed endo-1, 4-beta-glucanase gene in strawberry. *Plant Physiol.*, **118**, 1307-16.
- Harrison, P.M., Kumar, A., Lang, N., Snyder, M.** et **Gerstein, M.** 2002. A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.*, **30**, 1083-90.
- Hedden, P.** et **Phillips, A.L.** 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.*, **5**, 523-30.
- Heitz, T., Bergey, D.R.** et **Ryan, C.A.** 1997. A gene encoding a chloroplast-targeted lipoxygenase in tomato leaves is transiently induced by wounding, systemin, and methyl jasmonate. *Plant Physiol.*, **114**, 1085-93.
- Herrero, J., Valencia, A.** et **Dopazo, J.** 2001. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics*, **17**, 126-36.
- Hieter, P.** et **Boguski, M.** 1997. Functional genomics: it's all how you read it. *Science*, **278**, 601-2.
- Hirschberg, J.** 2001. Carotenoid biosynthesis in flowering plants. *Curr. Opin. Plant Biol.*, **4**, 210-8.
- Hisada, S., Masayuki, K., Endo-Inagaki, T., Omura, M.** et **Moriguchi, T.** 1997. Refinement of cDNA clone expression analysis in random sequencing from the rapid cell development phase of *Citrus* fruit. *J. Plant Physiol.*, **155**, 699-705.
- Hiwasa, K., Kinugasa, Y., Amano, S., Hashimoto, A., Nakano, R., Inaba, A.** et **Kubo, Y.** 2003a. Ethylene is required for both the initiation and progression of softening in pear (*Pyrus communis* L.) fruit. *J. Exp. Bot.*, **54**, 771-9.
- Hiwasa, K., Rose, J.K., Nakano, R., Inaba, A.** et **Kubo, Y.** 2003b. Differential expression of seven alpha-expansin genes during growth and ripening of pear fruit. *Physiol. Plant.*, **117**, 564-572.
- Honda, C., Kotoda, N., Wada, M., Kondo, S., Kobayashi, S., Soejima, J., Zhang, Z., Tsuda, T.** et **Moriguchi, M.** 2002. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiol. Biol.*, **40**, 955-962.
- Horikawa, S., Sasuga, J., Shimizu, K., Ozasa, H.** et **Tsukada, K.** 1990. Molecular cloning and nucleotide sequence of cDNA encoding the rat kidney S-adenosylmethionine synthetase. *J. Biol. Chem.*, **265**, 13683-6.
- Hormaza, J.I.** 2002. Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. *Theor. Appl. Genet.*, **104**, 321-328.

- Houtz, R.L.** et **Portis, A.R.** 2003. The life of ribulose 1,5-bisphosphate carboxylase/oxygenase--posttranslational facts and mysteries. *Arch. Biochem. Biophys.*, **414**, 150-8.
- Howe, G.A., Lee, G.I., Itoh, A., Li, L.** et **DeRocher, A.E.** 2000. Cytochrome P450-dependent metabolism of oxylipins in tomato. Cloning and expression of allene oxide synthase and fatty acid hydroperoxide lyase. *Plant Physiol.*, **123**, 711-24.
- Huang, X.** et **Madan, A.** 1999. CAP3: A DNA sequence assembly program. *Genome Res.*, **9**, 868-77.
- Huang, Y., Li, H., Hutchison, C.E., Laskey, J.** et **Kieber, J.J.** 2003. Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signaling in *Arabidopsis*. *Plant J.*, **33**, 221-33.
- Hutchison, C.E.** et **Kieber, J.J.** 2002. Cytokinin signaling in *Arabidopsis*. *Plant Cell*, **14**, 47-59.
- Imin, N., Kerim, T., Weinman, J.J.** et **Rolfe, B G.** 2001. Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics*, **1**, 1149-61.
- Isaacson, T., Ronen, G., Zamir, D.** et **Hirschberg, J.** 2002. Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. *Plant Cell*, **14**, 333-42.
- Ishiki, Y., Oda, A., Yaegashi, Y., Orihara, Y., Arai, T., Hirabayashi, T., Nakagawa, H.** et **Sato, T.** 2000. Cloning of an auxin-responsive 1-aminocyclopropane-1-carboxylate synthase gene (CMe-ACS2) from melon and the expression of ACS genes in etiolated melon seedlings and melon fruits. *Plant Sci.*, **159**, 173-181.
- Itoh, H., Ueguchi-Tanaka, M., Sato, Y., Ashikari, M.** et **Matsuoka, M.** 2002. The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant Cell*, **14**, 57-70.
- Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., Tabata, S., Kakubari, Y., Yamaguchi-Shinozaki, K.** et **Shinozaki, K.** 2001. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *Plant J.*, **27**, 325-33.
- Jacob-Wilk, D., Holland, D., Goldschmidt, E.E., Riov, J.** et **Eyal, Y.** 1999. Chlorophyll breakdown by chlorophyllase: isolation and functional expression of the Chlase1 gene from ethylene-treated *Citrus* fruit and its regulation during development. *Plant J.*, **20**, 653-61.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G.** et **Gibson, T.J.** 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **23**, 403-5.
- Jiang, Y., Joyce, D.C.** et **Macnish, A.J.** 2000. Effect of Abscisic Acid on Banana Fruit Ripening in Relation to the Role of Ethylene. *J. Plant Growth Regul.*, **19**, 106-111.
- Jimenez-Bermudez, S., Redondo-Nevado, J., Munoz-Blanco, J., Caballero, J.L., Lopez-Aranda, J.M., Valpuesta, V., Pliego-Alfaro, F., Quesada, M.A.** et **Mercado, J.A.** 2002. Manipulation of strawberry fruit softening by antisense expression of a pectate lyase gene. *Plant Physiol.*, **128**, 751-9.
- Jimenez, A., Creissen, G., Kular, B., Firmin, J., Robinson, S., Verhoeyen, M.** et **Mullineaux, P.** 2002a. Changes in oxidative processes and components of the antioxidant system during tomato fruit ripening. *Planta*, **214**, 751-8.
- Jiménez, A., Gómez, J.M., Navarro, E** et **Sevilla, F.** 2002b. Changes in the antioxidative systems in mitochondria during ripening of pepper fruits. *Plant Physiol. Biol.*, **40**, 515-520.
- Jones, C.S., Davies, H.V.** et **Taylor, M.A.** 2000. Profiling of changes in gene expression during raspberry (*Rubus idaeus*) fruit ripening by application of RNA fingerprinting techniques. *Planta*, **211**, 708-14.

- Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. et Jones, C.A.** 2002. Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta*, **216**, 255-64.
- Kamo, M., Kawakami, T., Miyatake, N. et Tsugita, A.** 1995. Separation and characterization of *Arabidopsis thaliana* proteins by two-dimensional gel electrophoresis. *Electrophoresis*, **16**, 423-30.
- Karlin, S. et Altschul, S.F.** 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U S A*, **87**, 2264-8.
- Karvouni, Z., John, I., Taylor, J.E., Watson, C.F., Turner, A.J. et Grierson, D.** 1995. Isolation and characterisation of a melon cDNA clone encoding phytoene synthase. *Plant Mol. Biol.*, **27**, 1153-62.
- Kato, M., Ikoma, Y., Matsumoto, H., Sugiura, M., Hyodo, H. et Yano, M.** 2004. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in *Citrus* fruit. *Plant Physiol.*, **134**, 824-37.
- Kieber, J.J., Rothenberg, M., Roman, G., Feldmann, K.A. et Ecker, J.R.** 1993. CTR1, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the raf family of protein kinases. *Cell*, **72**, 427-41.
- Klee, H.J.** 2002. Control of ethylene-mediated processes in tomato at the level of receptors. *J. Exp. Bot.*, **53**, 2057-63.
- Kneissl, M.L. et Deikman, J.** 1996. The tomato E8 gene influences ethylene biosynthesis in fruit but not in flowers. *Plant Physiol.*, **112**, 537-547.
- Kobayashi, S., Goto-Yamamoto, N. et Hirochika, H.** 2004. Retrotransposon-Induced Mutations in Grape Skin Color. *Science*, **304**, 982.
- Kobayashi, Y., Yamamoto, S., Minami, H., Kagaya, Y. et Hattori, T.** 2004. Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. *Plant Cell*, **16**, 1163-77.
- Kochhar, A., Khurana, J.P. et Tyagi, A.K.** 1996. Nucleotide sequence of the psbP gene encoding precursor of 23-kDa polypeptide of oxygen-evolving complex in *Arabidopsis thaliana* and its expression in the wild-type and a constitutively photomorphogenic mutant. *DNA Res.*, **3**, 277-85.
- Komatsu, A., Moriguchi, T., Koyama, K., Omura, M. et Akihama, T.** 2002. Analysis of sucrose synthase genes in *Citrus* suggests different roles and phylogenetic relationships. *J. Exp. Bot.*, **53**, 61-71.
- Kostina, K. F.** 1969. The use of varietal resources of apricot for breeding. *Trud. Nikit. Bot. Sad.*, **41**, 45-63.
- Kostina, K.F.** 1931. The cultivation of apricots in Fergana Valley (in Russian). *Bull. Appl. Bot. Genet. Plant Breed.*, **27**, 3-158.
- Kramer, M., Sanders, R., Bolkan, H., Waters, C., Sheehy, R.E. et Hiatt, W.R.** 1992. Postharvest evaluation of transgenic tomatoes with reduced levels of polygalacturonase: processing, firmness and disease resistance. *Post Biol. Tech.*, **1**, 241-255.
- Lange, T., Hedden, P. et Graebe, J.E.** 1994. Expression cloning of a gibberellin 20-oxidase, a multifunctional enzyme involved in gibberellin biosynthesis. *Proc. Natl. Acad. Sci. U S A*, **91**, 8552-6.
- Lashbrook, C.C., Tieman, D.M. et Klee, H.J.** 1998. Differential regulation of the tomato ETR gene family throughout plant development. *Plant J.*, **15**, 243-52.

- Laskowski, M.J., Dreher, K.A., Gehring, M.A., Abel, S., Gensler, A.L. et Sussex, I.M.** 2002. FQR1, a novel primary auxin-response gene, encodes a flavin mononucleotide-binding quinone reductase. *Plant Physiol.*, **128**, 578-90.
- Layne, R.E.C., Bailey C.H. et Hough L.F.**, 1996. Apricots. In: Janick J. et Moore J.N. (Eds), *Fruit breeding, volume 1: tree and tropical fruits*. John Wiley and Sons, New York, pp 79-111.
- Lee, S., Lee, E.J., Yang, E.J., Lee, J.E., Park, A.R., Song, W.H. et Park, O.K.** 2004. Proteomic identification of annexins, calcium-dependent membrane binding proteins that mediate osmotic stress and abscisic acid signal transduction in *Arabidopsis*. *Plant Cell*, **16**, 1378-91.
- Lelièvre, J.M., Tichit, L., Dao, P., Fillion, L., Nam, Y.W., Pech, J.C. et Latche, A.** 1997. Effects of chilling on the expression of ethylene biosynthetic genes in Passe-Crassane pear (*Pyrus communis* L.) fruits. *Plant Mol. Biol.*, **33**, 847-55.
- Lelièvre, J.M., Tichit, L., Fillion, L., Larrigaudière, C., Vendrell, M. et Pech, J.C** 1995. Cold-induced accumulation of 1-aminocyclopropane 1-carboxylate oxidase protein in Granny Smith apples. *Post Biol. Tech.*, **5**, 11-17.
- Lester, P.J. et Hubbard, S.J.** 2002. Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics. *Proteomics*, **2**, 1392-405.
- Lester, D.R., Speirs, J., Orr, G. et Brady, C.J.** 1994. Peach (*Prunus persica*) endopolygalacturonase cDNA isolation and mRNA analysis in melting and nonmelting peach cultivars. *Plant Physiol.*, **105**, 225-31.
- Li, Y., Darley, C.P., Ongaro, V., Fleming, A., Schipper, O., Baldauf, S.L. et McQueen-Mason, S.J.** 2002. Plant expansins are a complex multigene family with an ancient evolutionary origin. *Plant Physiol.*, **128**, 854-64.
- Liang, P. et Pardee, A.B.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967-71.
- Lisacek, F. C., Traini, M. D., Sexton, D., Harry, J.L. et Wilkins, M.R.** 2001. Strategy for protein isoform identification from expressed sequence tags and its application to peptide mass fingerprinting. *Proteomics*, **2**, 186-93.
- Liska, A. J. et Shevchenko, A.** 2003. Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implication. *Proteomics*, **3**, 19-28.
- Liscum, E. et Reed, J.W.** 2002. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.*, **49**, 387-400.
- Lohmann, J., Henz, S., Davison, T., Pape, U., Vingron, M., Schölkopf, B., Weigel, D. et Schmid, M.** 2004. AtgenExpress-Expression atlas of *Arabidopsis* development conference. Plant Gems, Lyon, 22-25 Septembre.
- Lorenzo, O., Piqueras, R., Sanchez-Serrano, J.J. et Solano, R.** 2003. ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *Plant Cell*, **15**, 165-78.
- Lorenzo, O., Rodriguez, D., Nicolas, G., Rodriguez, P.L. et Nicolas, C.** 2001. A new protein phosphatase 2C (FsPP2C1) induced by abscisic acid is specifically expressed in dormant beechnut seeds. *Plant Physiol.*, **125**, 1949-56.
- Lu, C., Zainal, Z., Tucker, G.A. et Lycett, G.W.** 2001. Developmental abnormalities and reduced fruit softening in tomato plants expressing an antisense Rab11 GTPase gene. *Plant Cell*, **13**, 1819-33.

- Maclachlan, G.** et **Brady, C.** 1994. Endo-1,4-[beta]-glucanase, xyloglucanase, and xyloglucan endo-transglycosylase activities versus potential substrates in ripening tomatoes. *Plant Physiol.*, **105**, 965-974.
- Manning, K.** 1998. Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. *Planta*, **205**, 622-31.
- Marin-Rodriguez, M.C., Smith, D.L., Manning, K., Orchard, J.** et **Seymour, G.B.** 2003. Pectate lyase gene expression and enzyme activity in ripening banana fruit. *Plant Mol. Biol.*, **51**, 851-7.
- Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., Frey, A.** et **Marion-Poll, A.** 1996. Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. *EMBO J.*, **15**, 2331-42.
- Marlow, G.** et **Loescher, W.H.** 1984. Watercore. *Hort. Rev.*, **6**, 189-251.
- Martelli, M.R., Sabina, A., Sciancalepore, F., Sunseri, I.** et **Greco, G.** 2003. Molecular characterization of ripening fruit processes in strawberry starting from a transcribed genomic DNA fraction. *Acta Hort.*, **625**, 113-117.
- Martin, D.N., Proebsting, W.M.** et **Hedden, P.** 1999. The SLENDER gene of pea encodes a gibberellin 2-oxidase. *Plant Physiol.*, **121**, 775-81.
- Mathesius, U., Imin, N., Chen, H., Djordjevic, Weinman, J.J., Natera, S.H., Morris, A.C., Kerim, T., Paul, S., Menzel, C., Weiller, G.F.** et **Rolfe B.G.** 2002. Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics*, **2**, 1288-303.
- Matile, P., Hortensteiner, S.** et **Thomas, H.** 1999. Chlorophyll degradation. *Annu. Rev. Plant. Physiol. Plant Mol. Biol.*, **50**, 67-95.
- Mbéguie-A-Mbéguie, D., Gouble, B., Gomez, R.M., Audergon, J.M., Albagnac, G.** et **Fils-Lycaon, B.** 2002. Two expansin cDNAs from *Prunus armeniaca* expressed during fruit ripening are differently regulated by ethylene. *Plant Physiol. Biol.*, **40**, 445-452.
- Mbéguié-Mbéguié, D., Chahine, H., Gomez, R.M., Gouble, B., Audergon, J.M., Souty, M., Albagnac, G.** et **Fils-Lycaon, B.** 1999. Molecular cloning and expression of a cDNA encoding 1-aminocyclopropane-1-carboxylate (ACC) oxidase from apricot fruit (*Prunus armeniaca* cv. Bergeron). *Physiol. Plant.*, **105**, 294-303.
- Mbéguie-Mbéguie, D., Gomez, R.M.** et **Fils-Lycaon, B.** 1997. Molecular Cloning and Nucleotide Sequence of an Abscisic Acid Stress Ripening Induced (ASR) Like Protein From Apricot Fruit (Accession No. U93164) Gene Expression During Fruit Ripening (PGR97-166). *Plant Physiol.*, **115**, 1288.
- Mbéguié-Mbéguié, D.** 2000. Isolement, identification et caractérisation de gènes impliqués dans la maturation de l'abricot (*Prunus armeniaca* L.). *Thèse*. <http://www.inra.fr/theses/these-integrale/Theses/mbeguie/html/these.html>. pp. 196.
- McQueen-Mason, S.** et **Cosgrove, D.J.** 1994. Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proc. Natl. Acad. Sci. U S A*, **91**, 6574-8.
- Mehlenbacher S.A., Cociu V.** et **Hough L.F.** 1990. Apricots (*Prunus*), in genetic resources of temperate fruit and nut crops. *Acta Hort.*, **290**, 65-107.
- Michalczuk, L., Ribnicky, D.M., Cooke, T.J.** et **Cohen, J.D.** 1992. Regulation of indole-3-acetic acid biosynthetic pathways in carrot cell cultures. *Plant Physiol.*, **100**, 1346-53.
- Micheli, F.** 2001. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.*, **6**, 414-9.

- Miziorko, H.M.** et **Lorimer, G.H.** 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu. Rev. Biochem.*, **52**, 507-35.
- Mohnen, D.** 1999. Biosynthesis of pectins and galactomannans. In: Pinto, B.M. (Ed), *Comprehensive Natural Products Chemistry, Vol. 3, Carbohydrates and their Derivatives Including Tannins, Cellulose, and Related Lignins*, Elsevier, Oxford, pp. 497–527.
- Moyano, E., Portero-Robles, I., Medina-Escobar, N., Valpuesta, V., Munoz-Blanco, J.** et **Caballero, J.L.** 1998. A fruit-specific putative dihydroflavonol 4-reductase gene is differentially expressed in strawberry during the ripening process. *Plant Physiol.*, **117**, 711-6.
- Murfitt, L.M., Kolosova, N., Mann, C.J.** et **Dudareva, N.** 2000. Purification and characterization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methyl benzoate in flowers of *Antirrhinum majus*. *Arch. Biochem. Biophys.*, **382**, 145-51.
- Murray, A.J., Hobson, G.E., Schuch, W.** et **Bird, C R.** 1993. Reduced ethylene synthesis in EFE antisense tomatoes has differential effects on fruit ripening processes. *Post Biol. Tech.*, **2**, 301-313.
- Nam, Y.W., Tichit, L., Leperlier, M., Cuerq, B., Marty, I.** et **Lelievre, J.M.** 1999. Isolation and characterization of mRNAs differentially expressed during ripening of wild strawberry (*Fragaria vesca* L.) fruits. *Plant Mol. Biol.*, **39**, 629-36.
- Neuhoff, V., Arold, N., Taube, D.** et **Ehrhardt, W.** 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, **9**, 255-62.
- Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N.** et **Tabata, S.** 2000. Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res.*, **7**, 223-7.
- Nishitani, K.** et **Tominaga, R.** 1992. Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J. Biol. Chem.*, **267**, 21058-64.
- Normanly, J., Slovin, J.P.** et **Cohen, J.D.** 1995. Rethinking auxin biosynthesis and metabolism. *Plant Physiol.*, **107**, 323-329.
- Nosarszewski, M., Clements, A.M., Bruce Downie, A.** et **Archbold, D.D.** 2004. Sorbitol dehydrogenase expression and activity during apple fruit set and early development. *Physiol. Plant.*, **121**, 391-398.
- O'Farrell, P.H.** 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, **250**, 4007-21.
- Obukosia, S.D., Richards, C.M.** et **Boyer, C.D.** 2003. Expression of plastid-encoded photosynthetic genes during chloroplast or chromoplast differentiation in *Cucurbitae pepo* L. fruits. *Phytochemistry*, **64**, 1213-21.
- Ohmiya, A.** 2000. Effects of auxin on growth and ripening of mesocarp discs of peach fruit. *Scientia Hort.*, **84**, 309-319.
- Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H.** et **Ohme-Takagi, M.** 2001. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell*, **13**, 1959-68.
- Okada, K., Saito, T., Nakagawa, T., Kawamukai, M.** et **Kamiya, Y.** 2000. Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in *Arabidopsis*. *Plant Physiol.*, **122**, 1045-56.
- Ouaked, F., Rozhon, W., Lecourieux, D.** et **Hirt, H.** 2003. A MAPK pathway mediates ethylene signaling in plants. *EMBO J.*, **22**, 1282-8.

- Palejwala, V.A., Parikh, H.P. et Modi, V.V.** 1985. The role of abscisic acid on ripening of grapes. *Physiol. Plant.*, **65**, 498-502.
- Pandey, S., Ranade, S.A., Nagar, P.K. et Kumar, N.** 2000. Role of polyamines and ethylene as modulators of plant senescence. *J. Biosci.*, **25**, 291-9.
- Pappin, D.J.C., Hojrup, P. et Bleasby, A.J.** 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.*, **3**, 327-32.
- Payton, S., Fray, R.G., Brown, S. et Grierson, D.** 1996. Ethylene receptor expression is regulated during fruit ripening, flower senescence and abscission. *Plant Mol. Biol.*, **31**, 1227-31.
- Pecker, I., Gabbay, R., Cunningham, F.X. et Hirschberg, J.** 1996. Cloning and characterization of the cDNA for lycopene beta-cyclase from tomato reveals decrease in its expression during fruit ripening. *Plant Mol. Biol.*, **30**, 807-19.
- Perez-Bueno, M.L., Rahoutei, J., Sajjani, C., Garcia-Luque, I. et Baron, M.** 2004. Proteomic analysis of the oxygen-evolving complex of photosystem II under biotec stress: Studies on *Nicotiana benthamiana* infected with tobamoviruses. *Proteomics*, **4**, 418-25.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J. et Quackenbush, J.** 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, **19**, 651-2.
- Peter, G.F. et Thornber, J.P.** 1991. Biochemical composition and organization of higher plant photosystem II light-harvesting pigment-proteins. *J. Biol. Chem.*, **266**, 16745-54.
- Phimister, B.** 1999. Going global *Nat. Genet.*, **21**, 1.
- Picault, N., Palmieri, L., Pisano, I., Hodges, M. et Palmieri, F.** 2002. Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J. Biol. Chem.*, **277**, 24204-11.
- Powell, A.L., Kalamaki, M.S., Kurien, P.A., Gurrieri, S. et Bennett, A.B.** 2003. Simultaneous transgenic suppression of LePG and LeExp1 influences fruit texture and juice viscosity in a fresh market tomato variety. *J. Agric. Food Chem.*, **51**, 7450-5.
- Pratelli, R., Lacombe, B., Torregrosa, L., Gaymard, F., Romieu, C., Thibaud, J.B. et Sentenac, H.** 2002. A grapevine gene encoding a guard cell K(+) channel displays developmental regulation in the grapevine berry. *Plant Physiol.*, **128**, 564-77.
- Prescott, A.G. et John, P.** 1996. Dioxygenases: Molecular Structure and Role in Plant Metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **47**, 245-271.
- Proust, J., Houlne, G., Schantz, M.L. et Schantz, R.** 1996. Characterization and gene expression of an annexin during fruit development in *Capsicum annuum*. *FEBS Lett.*, **383**, 208-12.
- Puskas, L.G., Zvara, A., Hackler, L., et Van Hummelen, P.** 2002. RNA amplification results in reproducible microarray data with slight ratio bias. *Biotechniques*, **32**, 1330-4,1336,1338,1340.
- Qin, X. et Zeevaart, J.A.** 1999. The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean. *Proc. Natl. Acad. Sci. U S A*, **96**, 15354-61.
- Rabilloud, T.** 1999. 2-D Proteome Analysis Protocols. *Methods Mol. Biol.*, **112**, 9-19.
- Redgwell, R.J. et Fry, S.C.** 1993. Xyloglucan Endotransglycosylase Activity Increases during Kiwifruit (*Actinidia deliciosa*) Ripening (Implications for Fruit Softening). *Plant Physiol.*, **103**, 1399-1406.

- Reed, J.W.** 2001. Roles and activities of Aux/IAA proteins in *Arabidopsis*. *Trends Plant Sci.*, **6**, 420-5.
- Religio, A., Schwager, C., Richter, A., Ansoerge, W. et Valcarcel, J.** 2002. Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res.*, **30**, e51.
- Reymond, P., Kunz, B., Paul-Pletzer, K., Grimm, R., Eckerskorn, C. et Farmer, E.E.** 1996. Cloning of a cDNA encoding a plasma membrane-associated, uronide binding phosphoprotein with physical properties similar to viral movement proteins. *Plant Cell*, **8**, 2265-76.
- Rice P., Longden I. et Bleasby A.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, **16**, 276-7.
- Ridley, B.L., O'Neill, M.A. et Mohnen, D.** 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, **57**, 929-67.
- Riechmann, J.L. et Meyerowitz, E.M.** 1998. The AP2/EREBP family of plant transcription factors. *Biol. Chem.*, **379**, 633-46.
- Riov, J., Dagan, E., Goren, R. et Yang, S.F.** 1990. Characterization of abscisic acid-induced ethylene production in *Citrus* leaf and tomato fruit tissues. *Plant Physiol.*, **92**, :48-53.
- Roberts, J.A., Elliott, K.A. et Gonzalez-Carranza, Z.H.** 2002. Abscission, dehiscence, and other cell separation processes. *Annu. Rev. Plant Biol.*, **53**, 131-58.
- Rock, C.D. et Zeevaart, J.A.** 1991. The aba mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis. *Proc. Natl. Acad. Sci. U S A*, **88**, 7496-9.
- Roman, G., Lubarsky, B., Kieber, J.J., Rothenberg, M. et Ecker, J.R.** 1995. Genetic analysis of ethylene signal transduction in *Arabidopsis thaliana*: five novel mutant loci integrated into a stress response pathway. *Genetics*, **139**, 1393-409.
- Rombaldi, C.V., Chaves, A.L., Silva, J.A., Lemos, M.R.B., Ayub, R.A., Balagué, C., et Pech, J.C.** 1996. Expressão da enzima ácido 1-carboxílico-1-aminociclopropano oxidase em tomate (*Lycopersicon esculentum* Mill.) cultivar Kadá. *Revista Brasileira de Agrociência*, **2**, 5-10.
- Ronen, G., Carmel-Goren, L., Zamir, D. et Hirschberg, J.** 2000. An alternative pathway to beta -carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato. *Proc. Natl. Acad. Sci. U S A*, **97**, 11102-7.
- Ronen, G., Cohen, M., Zamir, D. et Hirschberg, J.** 1999. Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J.*, **17**, 341-51.
- Rose, J.K. et Bennett, A.B.** 1999. Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. *Trends Plant Sci.*, **4**, 176-183.
- Rose, J.K., Lee, H.H. et Bennett, A.B.** 1997. Expression of a divergent expansin gene is fruit-specific and ripening-regulated. *Proc. Natl. Acad. Sci. U S A*, **94**, 5955-60.
- Ross, J.J., O'Neill, D.P., Wolbang, C.M., Symons, G.M. et Reid, J.B.** 2001. Auxin-gibberellin interactions and their role in plant growth. *J. Plant Growth Regul.*, **20**, 336-353.
- Ross, J.J., Reid, J.B., Swain, S.M., Hasan, O., Poole, A.T., Hedden, P. et Willis, C.L.** 1995. Genetic regulation of gibberellin deactivation in *Pisum*. *Plant J.*, **7**, 513-52.
- Rossi, M., Lijavetzky, D., Bernacchi, D., Hopp, H.E. et Iusem, N.** 1996. Asr genes belong to a gene family comprising at least three closely linked loci on chromosome 4 in tomato. *Mol. Gen. Genet.*, **252**, 489-92.

- Rouillard, J.M., Zuker, M. et Gulari, E.** 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.*, **31**, 3057-62.
- Rozen, S. et Skaletsky, H.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365-86.
- Rudd, S.** 2003. Expressed sequence tags: alternative or complement to whole genome sequences?. *Trends Plant Sci.*, **8**, 321-9.
- Rudd, S., Mewes, H.W. et Mayer, K.F.** 2003. Sputnik: a database platform for comparative plant genomics. *Nucleic Acids Res.*, **31**, 128-32.
- Saab, I.N. et Sachs, M.M.** 1996. A flooding-induced xyloglucan endo-transglycosylase homolog in maize is responsive to ethylene and associated with aerenchyma. *Plant Physiol.*, **112**, 385-91.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. et Quackenbush, J.** 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374-8.
- Sage-Ono, K., Ono, M., Harada, H. et Kamada, H.** 1998. Dark-induced accumulation of mRNA for a homolog of translationally controlled tumor protein (TCTP) in *Pharbitis*. *Plant Cell Physiol.*, **39**, 357-60.
- Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K. et Yamaguchi-Shinozaki, K.** 2002. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.*, **290**, 998-1009.
- Sambrook, J., Fritsch, E.F. et Maniatis, T.** 1989. *Molecular Cloning A Laboratory Manual*, 2nd Ed., Cold Spring Harbour Laboratory Press, New York.
- Sanchez, J.C., Schaller, D., Ravier, F., Golaz, O., Jaccoud, S., Belet, M., Wilkins, M.R., James, R., Deshusses, J. et Hochstrasser, D.** 1997. Translationally controlled tumor protein: a protein identified in several nontumoral cells including erythrocytes. *Electrophoresis*, **18**, 150-5.
- Saniewski M. et Czapski, J.** 1983. The effect of methyl jasmonate on lycopene and β -carotene accumulation in ripening red tomato. *Experientia*, **39**, 1373-1374.
- Saravanan, R.S. et Rose, J.K.C.** 2004. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, **4**, 2522-2532.
- Sarry, J.E., Sommerer, N., Sauvage, F.X., Bergoin, A., Rossignol, M., Albagnac, G. et Romieu, C.** 2004. Grape berry biochemistry revisited upon proteomic analysis of the mesocarp. *Proteomics*, **4**, 201-15.
- Sauter, M., Rzewuski, G., Marwedel, T. et Lorbiecke, R.** 2002. The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *J. Exp. Bot.*, **53**, 2325-31.
- Schaller, F., Biesgen, C., Mussig, C., Altmann, T. et Weiler, E.W.** 2000. 12-Oxophytodienoate reductase 3 (OPR3) is the isoenzyme involved in jasmonate biosynthesis. *Planta*, **210**, 979-84.
- Schena, M., Shalon, D., Davis, R.W. et Brown, P.O.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-70.
- Schmitt, A.O., Specht, T., Beckmann, G., Dahl, E., Pilarsky, C.P., Hinzmann, B. et Rosenthal, A.** 1999. Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucleic Acids Res.*, **27**, 4251-60.

- Schroder, R., Atkinson, R.G., Langenkamper, G. et Redgwell, R.J.** 1998. Biochemical and molecular characterisation of xyloglucan endotransglycosylase from ripe kiwifruit. *Planta*, **204**, 242-51.
- Schwartz, S.H., Qin, X. et Zeevaart, J.A.** 2001. Characterization of a novel carotenoid cleavage dioxygenase from plants. *J. Biol. Chem.*, **276**, 25208-11.
- Schwartz, S.H., Tan, B.C., Gage, D.A., Zeevaart, J.A. et McCarty, D.R.** 1997. Specific oxidative cleavage of carotenoids by VP14 of maize. *Science*, **276**, 1872-4.
- Schwechheimer, C., Serino, G. et Deng, X.W.** 2002. Multiple ubiquitin ligase-mediated processes require COP9 signalosome and AXR1 function. *Plant Cell*, **14**, 2553-63.
- Schwender, J., Seemann, M., Lichtenthaler, H.K. et Rohmer, M.** 1996. Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*. *Biochem. J.*, **316**, 73-80.
- Sebastian, C.H., Hardin, S.C., Clouse, S.D., Kieber, J.J. et Huber, S.C.** 2004. Identification of a new motif for CDPK phosphorylation in vitro that suggests ACC synthase may be a CDPK substrate. *Arch. Biochem. Biophys.*, **428**, 81-91.
- Seo, H.S., Song, J.T., Cheong, J.J., Lee, Y.H., Lee, Y.W., Hwang, I., Lee, J.S. et Choi, Y.D.** 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc. Natl. Acad. Sci. U S A*, **98**, 4788-93.
- Seo, M. et Koshiba, T.** 2002. Complex regulation of ABA biosynthesis in plants. *Trends Plant Sci.*, **7**, 41-8.
- Seymour, G.B., Taylor, J.E., et Tucker, G.A.** 1993. *Biochemistry of Fruit Ripening*. Chapman & Hall, London. pp. 454.
- Shigeoka, S., Ishikawa, T., Tamoi, M., Miyagawa, Y., Takeda, T., Yabuta, Y. et Yoshimura, K.** 2002. Regulation and function of ascorbate peroxidase isoenzymes. *J. Exp. Bot.*, **53**, 1305-19.
- Shirakihara, Y., Leslie, A.G., Abrahams, J.P., Walker, J.E., Ueda, T., Sekimoto, Y., Kambara, M., Saika, K., Kagawa, Y. et Yoshida, M.** 1997. The crystal structure of the nucleotide-free alpha 3 beta 3 subcomplex of F1-ATPase from the thermophilic Bacillus PS3 is a symmetric trimer. *Structure*, **5**, 825-36.
- Silverstone, A.L., Ciampaglio, C.N. et Sun, T.** 1998. The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell*, **10**, 155-69.
- Smith, C.J., Watson, C.F., Morris, P.C., Bird, C.R., Seymour, G.B., Gray, J.E., Arnold, C., Tucker, G.A., Schuch, W. et Harding, S.** 1990. Inheritance and effect on ripening of antisense polygalacturonase genes in transgenic tomatoes. *Plant Mol. Biol.*, **14**, 369-79.
- Snyder, A.M., Clark, B.M., Robert, B., Ruban, A.V. et Bungard, R.A.** 2004. Carotenoid specificity of light-harvesting complex II binding sites. Occurrence of 9-cis-violaxanthin in the neoxanthin-binding site in the parasitic angiosperm *Cuscuta reflexa*. *J. Biol. Chem.*, **279**, 5162-8.
- Solano, R., Stepanova, A., Chao, Q. et Ecker, J.R.** 1998. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes. Dev.*, **12**, 3703-14.
- Solomon, M., Belenghi, B., Delledonne, M., Menachem, E. et Levine, A.** 1999. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell*, **11**, 431-44.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J.,**

- Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D. et Birney, E.** 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.*, **12**, 1611-8.
- Stephen, J.R., Dent, K.C. et Finch-Savage, W.E.** 2004. Molecular responses of *Prunus avium* (wild cherry) embryonic axes to temperatures affecting dormancy. *New Phytol.*, **161**, 401-413.
- Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. et Lundeberg, J.** 1998. Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci. U S A*, **95**, 13330-5.
- Stintzi, A. et Browse, J.** 2000. The *Arabidopsis* male-sterile mutant, opr3, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. *Proc. Natl. Acad. Sci. U S A*, **97**, 10625-30.
- Strassner, J., Schaller, F., Frick, U.B., Howe, G.A., Weiler, E.W., Amrhein, N., Macheroux, P. et Schaller, A.** 2002. Characterization and cDNA-microarray expression analysis of 12-oxophytodienoate reductases reveals differential roles for octadecanoid biosynthesis in the local versus the systemic wound response. *Plant J.*, **32**, 585-601.
- Sugino, Y., Teraoka, H. et Shimono, H.** 1966. Metabolism of deoxyribonucleotides. I. Purification and properties of deoxycytidine monophosphokinase of calf thymus. *J. Biol. Chem.*, **241**, 961-9.
- Sun, W., Bernard, C., van de Cotte, B., Van Montagu, M. et Verbruggen, N.** 2001. At-HSP17.6A, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression. *Plant J.*, **27**, 407-15.
- Sun, Z., Cunningham, F.X. et Gantt, E.** 1998. Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. *Proc. Natl. Acad. Sci. U S A*, **95**, 11482-8.
- Sun, Z., Gantt, E. et Cunningham, F.X.** 1996. Cloning and functional analysis of the beta-carotene hydroxylase of *Arabidopsis thaliana*. *J. Biol. Chem.*, **271**, 24349-52.
- Sung, S.K., Jeong, D.H., Nam, J., Kim, S.H., Kim, S.R. et An, G.** 1998. Expressed sequence tags of fruits, peels, and carpels and analysis of mRNA expression levels of the tagged cDNAs of fruits from the Fuji apple. *Mol. Cells*, **8**, 565-77.
- Swain, A.R., Dutton, S.P. et Truswell, A.S.** 1985. Salicylates in foods. *J. Am. Diet. Assoc.*, **85**, 950-60.
- Takeoka, G., Flath, R., Mon, T., Teranishi, R. et Guentert, M.** 1990. Volatile constituents of Apricot (*Prunus armeniaca* L.). *J. Agric. Food Chem.*, **38**, 471-477.
- Tan, B.C., Schwartz, S.H., Zeevaart, J.A. et McCarty, D.R.** 1997. Genetic control of abscisic acid biosynthesis in maize. *Proc. Natl. Acad. Sci. U S A*, **94**, 12235-40.
- Tang, C. et Jennings, W.** 1967. Volatile components of apricots. *J. Agric. Food Chem.*, **15**, 24-28.
- Tang, H., Vasconcelos, A.C. et Berkowitz, G.A.** 1995. Evidence that plant K⁺ channel proteins have two different types of subunits. *Plant Physiol.*, **109**, 327-30.
- Tanner, G.J., Franckin K.T., Abrahamsn S., Watson, J.M., Larkin, P.J. et Ashton A.R.** 2003. Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *J. Biol. Chem.*, **278**, 1647-56.
- Tatsuki, M. et Mori, H.** 1999. Rapid and transient expression of 1-aminocyclopropane-1-carboxylate synthase isogenes by touch and wound stimuli in tomato. *Plant Cell Physiol.*, **40**, 709-15.

- Tatsuki, M.** et **Mori, H.** 2001. Phosphorylation of tomato 1-aminocyclopropane-1-carboxylic acid synthase, LE-ACS2, at the C-terminal region. *J. Biol. Chem.*, **276**, 28051-7.
- Terrier, N., Ageorges, A., Abbal, P.** et **Romieu, C.** 2001a. Generation of ESTs from grape berry at various developmental stages. *J. Plant Physiol.*, **158**, 1575-1583.
- Terrier, N., Sauvage, F.X., Ageorges, A.** et **Romieu, C.** 2001b. Changes in acidity and in proton transport at the tonoplast of grape berries during development. *Planta*, **213**, 20-8.
- Tesniere, C.**, et **Vayda, M.E.** 1991. Method for isolation of high-quality RNA from grape berry tissues without contaminating tannins or carbohydrates. *Plant Mol. Biol. Rep.*, **9**, 242-251.
- Tesniere, C., Pradal, M., El-Kereamy, A., Torregrosa, L., Chatelet, P., Roustan, J.P.** et **Chervin, C.** 2004. Involvement of ethylene signalling in a non-climacteric fruit: new elements regarding the regulation of ADH expression in grapevine. *J. Exp. Bot.*, **55**, 2235-40.
- Theologis, A.** 1992. One rotten apple spoils the whole bushel: the role of ethylene in fruit ripening. *Cell*, **70**, 181-4.
- Thomas, S.G., Phillips, A.L.** et **Hedden, P.** 1999. Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc. Natl. Acad. Sci. U S A*, **96**, 4698-703.
- Thompson, A.J., Jackson, A.C., Parker, R.A., Morpeth, D.R., Burbidge, A.** et **Taylor, I.B.** 2000. Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-cis-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid. *Plant Mol. Biol.*, **42**, 833-45.
- Thompson, A.J., Tor, M., Barry, C.S., Vrebalov, J., Orfila, C., Jarvis, M.C., Giovannoni, J.J., Grierson, D.** et **Seymour, G.B.** 1999. Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. *Plant Physiol.*, **120**, 383-90.
- Thompson, J.D., Higgins, D.G.** et **Gibson T.J.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673-4680.
- Thompson, J.E.** et **Fry, S.C.** 2000. Evidence for covalent linkage between xyloglucan and acidic pectins in suspension-cultured rose cells. *Planta*, **211**, 275-86.
- Thompson, J.E., Smith, R.C.** et **Fry, S.C.** 1997. Xyloglucan undergoes interpolymeric transglycosylation during binding to the plant cell wall in vivo: evidence from ¹³C/³H dual labelling and isopycnic centrifugation in caesium trifluoroacetate. *Biochem. J.*, **327**, 699-708.
- Thorup, T.A., Tanyolac, B., Livingstone, K.D., Popovsky, S., Paran, I.** et **Jahn, M.** 2000. Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proc. Natl. Acad. Sci. U S A*, **97**, 11192-7.
- Tieman, D.M.** et **Klee, H.J.** 1999. Differential expression of two novel members of the tomato ethylene-receptor family. *Plant Physiol.*, **120**, 165-72.
- Tieman, D.M., Harriman, R.W., Ramamohan, G.** et **Handa, A.K.** 1992. An antisense pectin methylesterase gene alters pectin chemistry and soluble solids in tomato fruit. *Plant Cell*, **4**, 667-679.
- Tieman, D.M., Taylor, M.G., Ciardi, J.A.** et **Klee, H.J.** 2000. The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. *Proc. Natl. Acad. Sci. U S A*, **97**, 5663-8.
- Tiwari, S.B., Hagen, G.** et **Guilfoyle, T.** 2003. The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell*, **15**, 533-43.
- Tournier, B., Sanchez-Ballesta, M.T., Jones, B., Pesquet, E., Regad, F., Latche, A., Pech, J.C.** et **Bouzayan, M.** 2003. New members of the tomato ERF family show specific

- expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett.*, **550**, 149-54.
- Trainotti, L., Spinello, R., Piovan, A., Spolaore, S. et Casadoro, G.** 2001. beta-Galactosidases with a lectin-like domain are expressed in strawberry. *J. Exp. Bot.*, **52**, 1635-45.
- Trainotti, L., Spolaore, S., Pavanello, A., Baldan, B. et Casadoro, G.** 1999. A novel E-type endo-beta-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. *Plant Mol. Biol.*, **40**, 323-32.
- Trainotti, L., Zanin, D. et Casadoro, G.** 2003. A cell wall-oriented genomic approach reveals a new and unexpected complexity of the softening in peaches. *J. Exp. Bot.*, **54**, 1821-32.
- Trebith, T., Goldschmidt, E.E. et Riov, J.** 1993. Ethylene induces de novo synthesis of chlorophyllase, a chlorophyll degrading enzyme, in *Citrus* fruit peel. *Proc. Natl. Acad. Sci. U S A*, **90**, 9441-5.
- Tsuchiya, T., Ohta, H., Okawa, K., Iwamatsu, A., Shimada, H., Masuda, T. et Takamiya, K.** 1999. Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: finding of a lipase motif and the induction by methyl jasmonate. *Proc. Natl. Acad. Sci. U S A*, **96**, 15362-7.
- Ujino-Ihara, T., Yoshimura, K., Ugawa, Y., Yoshimaru, H., Nagasaka, K. et Tsumura, Y.** 2000. Expression analysis of ESTs derived from the inner bark of *Cryptomeria japonica*. *Plant Mol. Biol.*, **43**, 451-7.
- Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. et Tanksley, S.** 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell*, **14**, 1441-56.
- Vavilov, N.I.** 1949. The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot.*, **13**, 1-384.
- Velculescu, V.E., Zhang, L., Vogelstein, B. et Kinzler, K.W.** 1995. Serial analysis of gene expression. *Science*, **270**, 484-7.
- Vendrell, M.** 1985. Effect of abscisic acid and ethephon on several parameters of ripening in banana fruit tissue. *Plant Sci.*, **40**, 19-24.
- Verica, J.A., Maximova, S.N., Strem, M.D., Carlson, J.E., Bailey, B.A. et Gultinan, M.J.** 2004. Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. *Plant Cell Rep.*, (sous presse).
- Very, A.A. et Sentenac, H.** 2003. Molecular mechanisms and regulation of K⁺ transport in higher plants. *Annu. Rev. Plant Biol.*, **54**, 575-603.
- Visser, E., Cohen, J.D., Barendse, G., Blom, C. et Voesenek, L.** 1996. An ethylene-mediated increase in sensitivity to auxin induces adventitious root formation in flooded rumex *Palustris* sm. *Plant Physiol.*, **112**, 1687-1692.
- Vogel, J.P., Woeste, K.E., Theologis, A. et Kieber, J.J.** 1998. Recessive and dominant mutations in the ethylene biosynthetic gene ACS5 of *Arabidopsis* confer cytokinin insensitivity and ethylene overproduction, respectively. *Proc. Natl. Acad. Sci. U S A*, **95**, 4766-71.
- Wakeley, P.R., Rogers, H.J., Rozycka, M., Greenland, A.J. et Hussey, P.J.** 1998. A maize pectin methylesterase-like gene, ZmC5, specifically expressed in pollen. *Plant Mol. Biol.*, **37**, 187-92.
- Wan, C.Y. et Wilkins, T.A.** 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal. Biochem.*, **223**, 7-12.

- Wang, W., Hall, A.E., O'Malley, R. et Bleecker, A.B.** 2003. Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from *Arabidopsis* is not required for signal transmission. *Proc. Natl. Acad. Sci. U S A*, **100**, 352-7.
- Wang, Z., Dooley, T.P., Curto, E.V., Davis, R.L. et VandeBerg, J.L.** 2004. Cross-species application of cDNA microarrays to profile gene expression using UV-induced melanoma in *Monodelphis domestica* as the model system. *Genomics*, **83**, 588-99.
- Watson, B.S., Asirvatham, V.S., Wang, L., et Sumner, L.W.** 2003. Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol.*, **131**, 1104-23.
- Weber, G.** 1983. Biochemical strategy of cancer cells and the design of chemotherapy: G. H. A. Clowes Memorial Lecture. *Cancer Res.*, **43**, 3466-92.
- Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H. et Wobus, U.** 2003. The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. *Plant J.*, **33**, 395-411.
- Whittaker, D.J., Smith, G.S. et Gardner, R.C.** 1997. Expression of ethylene biosynthetic genes in *Actinidia chinensis* fruit. *Plant Mol. Biol.*, **34**, 45-55.
- Wilkins M. R. et Williams K. L.** 1997. Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. *J. Theor. Biol.*, **186**, 7-15.
- Wilkinson, J.Q., Lanahan, M.B., Conner, T.W. et Klee, H.J.** 1995. Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. *Plant Mol. Biol.*, **27**, 1097-108.
- Wills, R.B.H.** 1987. Composition of Australian fresh fruit and vegetables. *Food Technol. Aust.*, **39**, 523-26.
- Wong, J.C., Lambert, R.J., Wurtzel, E.T. et Rocheford, T.R.** 2004. QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. *Theor. Appl. Genet.*, **108**, 349-59.
- Wright, A.D., Moehlenkamp, C.A., Perrot, G.H., Neuffer, M.G. et Cone, K.C.** 1992. The maize auxotrophic mutant orange pericarp is defective in duplicate genes for tryptophan synthase beta. *Plant Cell*, **4**, 711-9.
- Xie, C., Zhang, J.S., Zhou, H.L., Li, J., Zhang, Z.G., Wang, D.W. et Chen, S.Y.** 2003. Serine/threonine kinase activity in the putative histidine kinase-like ethylene receptor NTHK1 from tobacco. *Plant J.*, **33**, 385-93.
- Xie, C., Zhang, Z.G., Zhang, J.S., He, X.J., Cao, W.H., He, S.J. et Chen, S.Y.** 2002. Spatial expression and characterization of a putative ethylene receptor protein NTHK1 in tobacco. *Plant Cell Physiol.*, **43**, 810-5.
- Xing, L. et Brendel, V.** 2001. Multi-query sequence BLAST output examination with MuSeqBox. *Bioinformatics*, **17**, 744-5.
- Yahyaoui, F.E., Wongs-Aree, C., Latche, A., Hackett, R., Grierson, D. et Pech, J.C.** 2002. Molecular and biochemical characteristics of a gene encoding an alcohol acyl-transferase involved in the generation of aroma volatile esters during melon ripening. *Eur. J. Biochem.*, **269**, 2359-66.
- Yamada, K., Niwa, N., Shiratake, K. et Yamaki, S.** 2001. cDNA cloning of NAD-dependent sorbitol dehydrogenase from peach fruit and its expression during fruit development. *J. Hort. Sci. and Biotechnol.*, **76**, 581-587.
- Yanagisawa, S., Yoo, S.D. et Sheen, J.** 2003. Differential regulation of EIN3 stability by glucose and ethylene signalling in plants. *Nature*, **425**, 521-5.
- Yang, S.F. et Hoffmann N.E.** 1984. Ethylene biosynthesis and its regulation in higher plants. *Annu. Rev. Plant Physiol.*, **35**, 155-189.

- Yang, Y.H.** et **Speed, T.** 2002. Design issues for cDNA microarray experiments. *Nature Rev. Gen.*, **3**, 579-588.
- Yang, Y.H., Dudoit, S., Luu, P.** et **Speed, T.P.** 2001. Normalization for cDNA Microarray Data. SPIE BIOS 2001, San Jose, California, January.
- Yip, W.K., Moore, T.** et **Yang, S.F.** 1992. Differential accumulation of transcripts for four tomato 1-aminocyclopropane-1-carboxylate synthase homologs under various conditions. *Proc. Natl. Acad. Sci. U S A*, **89**, 2475-9.
- Zablackis, E., Huang, J., Muller, B., Darvill, A.G.** et **Albersheim, P.** 1995. Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. *Plant Physiol.*, **107**, 1129-38.
- Zauberman, G.** et **Schiffman-Nadel, M.** 1972. Respiration of whole fruit and seed of avocado at various stages of development. *J. Am. Soc. Hort. Sci.*, **97**, 313-315.
- Zegzouti, H., Jones, B., Frasse, P., Marty, C., Maitre, B., Latch, A., Pech, J.C.** et **Bouzayen, M.** 1999. Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J.*, **18**, 589-600.
- Zhang, Z., Schwartz, S., Wagner, L.** et **Miller, W.** 2000. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, **7**, 203-214.
- Zhebentyayeva, N., Reighard, L., Gorina, M.** et **Abbott, G.** 2003. Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. *Theor. Appl. Genet.*, **106**, 435-44.
- Zhong, G.V.** et **Burns, J.K.** 2003. Profiling ethylene-regulated gene expression in *Arabidopsis thaliana* by microarray analysis. *Plant Mol. Biol.*, **53**, 117-31.
- Zhou, J., Hartmann, S., Shepherd, B.K.** et **Poulton, J.E.** 2002. Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. *Plant Physiol.*, **129**, 1252-64.
- Zhou, L., Jang, J.C., Jones, T.L.** et **Sheen, J.** 1998. Glucose and ethylene signal transduction crosstalk revealed by an *Arabidopsis* glucose-insensitive mutant. *Proc. Natl. Acad. Sci. U S A*, **95**, 10294-9.
- Zmasek, C.M.** et **Eddy, S.R.** 2002. RIO: Analysing proteomes by automated phylogenomics using resampled inference of orthologs. *BMC Bioinformatics*, **3**, 14.

VIII Annexe

VIII.1 Annexe A

Cette annexe présente une EST représentative pour chaque contig décrit dans le manuscrit. La plus longue EST a été choisie, ce n'est pas forcément la séquence la mieux adaptée pour des études d'expression (présence de non codant non vérifiée).

PA1TC6	BGA003N23P	PA288TC1	BGA002L10P
PA1TC15	BGH001K20P	PA310TC1	BGH014N10P
PA1TC48	BGH001O15	PA322TC1	BGH010O13P
PA1TC49	BGH008I12P	PA341TC1	BGH006P19P
PA2TC4	BGE018F13P	PA349TC1	BGH001M05
PA3TC4	BGE016L19Q	PA354TC1	BGH006C19P
PA4TC5	BGE018N21	PA355TC1	BGH002I04P
PA5TC5	BGE005K06P	PA373TC1	BGH014E18P
PA5TC7	BGH001N16	PA389TC1	BGA003M21P
PA6TC3	BGE001N06P	PA432TC1	BGE006C21P
PA6TC8	BGH001F18	PA446TC1	BGE017L10Q
PA7TC6	BGE015C17	PA529TC1	BGA005F18P
PA9TC1	BGE016O19Q	PA534TC1	BGE018D17
PA12TC1	BGE016O06Q	PA549TC1	BGE015I20
PA12TC3	BGH001O01	PA614TC1	BGH002B04P
PA17TC1	BGE018K22	PA671TC1	BGE005N16P
PA19TC3	BGE018E24	PA697TC1	BGA001H23P
PA23TC2	BGE018I18	PA703TC1	BGH006I06P
PA32TC1	BGA010J15	PA724TC1	BGE007C20P
PA34TC1	BGE014I07	PA736TC1	BGA005I06P
PA38TC1	BGA010J20	PA760TC1	BGA003P11P
PA43TC1	BGE016G08	PA790TC1	BGE006M16P
PA46TC1	BGA002I08P	PA835TC1	BGA010A05Q
PA60TC1	BGH006P11P	PA942TC1	BGH004M09P
PA62TC1	BGH001P17	PA1090TC1	BGH009M19P
PA82TC1	BGE005A07P	PA1110TC1	BGA011B21P
PA84TC1	BGA002O04P	PA118TC1	BGH001L21
PA92TC1	BGA019L20P	PA1238TC1	BGH014E22P
PA104TC1	BGH001E20P	PA1386TC1	BGH004F20P
PA134TC1	BGE003I11P	PA1425TC1	BGH006J19P
PA147TC2	BGE001J15P	PA1427TC1	BGH011P01P
PA149TC1	BGA011J13Q	PA1442TC1	BGE018C10
PA150TC1	BGH005N08P	PA1452TC1	BGA010F23P
PA164TC1	BGH009C06P	PA1496TC1	BGE015B14
PA173TC1	BGE017B16	PA1498TC1	BGE015B06P
PA185TC1	BGA010B24	PA1502TC1	BGA011L04
PA190TC1	BGE001E14P	PA1510TC1	BGE003I01P
PA196TC1	BGE016C18Q	PA1519TC1	BGH001M01
PA204TC1	BGE014P20	PA1546TC1	BGA009P18P
PA229TC1	BGE002F07P	PA1592TC1	BGH001A15
PA268TC1	BGE014B15Q	PA1627TC1	BGH001J10Q
PA269TC1	BGH005P15P	PA1633TC1	BGE018L06P

VIII.2 Annexe B

L'annexe sur le support CD contient :

- dans le répertoire isogene, la liste (fichier .lst), les séquences (fichier .txt) et les alignements (fichier .aln) des membres des familles multigéniques INN.
- dans le répertoire base de données, les éléments nécessaires à l'installation de la base de données à l'exception du logiciel AceDB.
- dans le répertoire Northern électronique, le résultat du Northern électronique.
- dans le répertoire script_perl les différents logiciels « locaux » présentés.
- dans le répertoire PT_HP_assembler, le logiciel PT_HP_assembler et sa procédure d'utilisation.
- dans le répertoire microarray, les fichiers nécessaires pour lire les lames, le fichier A114 Grimplet custom array.xls contient le nom des oligos, leur séquence et l'unigène correspondant, le fichier geneID A114 Grimplet Axon.gal est le fichier utilisé par le logiciel genepix pour la détection des spots.

Dans le sous répertoire dev, le fichier expressiontotdev.txt correspond aux résultats totaux de l'expérience de suivi de développement, chaque cluster de gènes différentiellement exprimés est présenté dans le fichier clust.xls. Le fichier du design utilisé par le logiciel geneANOVA est le fichier design_developpement.txt. Le fichier dev.anl est le fichier de l'analyse effectuée par Mev et il est exécutable avec ce même programme.

Dans le sous répertoire var, le fichier expressiontotvar.txt correspond aux résultats totaux de l'expérience de différenciation variétale, chaque cluster de gènes différentiellement exprimés est présenté dans le fichier clust.xls. Le fichier du design utilisé par le logiciel geneANOVA est le fichier design_variete.txt. Le fichier var.anl est le fichier de l'analyse effectuée par Mev et exécutable avec ce même programme.

VIII.3 Annexe C

VIII.3.1 Installation de la base de données

1) Télécharger AceDB pour Windows à l'adresse : <http://www.acedb.org/WinAce/CurrentWinAce.shtml>, utiliser la version 4.9u (j'ai rencontré des problèmes d'installation avec la version 4.9t).

2) Après le téléchargement, cliquer sur l'icône Acedb et suivre les instructions.

3) Aller dans le répertoire acedb/bin et cliquer sur l'icône makedir pour créer une nouvelle base de données et suivre les instructions.

4) Aller dans le répertoire de la nouvelle base de données :

-dans le répertoire database, remplacer les fichiers block1 et database.map par ceux disponibles sur le CD et ajouter les fichiers block2 et block3.

-dans le répertoire Wspec, remplacer les fichiers layout et models par ceux disponibles sur le CD.

5) Cliquer sur l'icône nouvellement créée sur le bureau pour lancer la base de données ou celui dans le répertoire database/nouvelle_base/new_database.

6) Pour faire des recherches avec le Query Builder ou faire des tableaux il est nécessaire de réindexer à chaque mise en route : aller dans Edit et cliquer sur read model and full reindex.

VIII.3.2 Utilisation de la base de données

Au démarrage la fenêtre suivante (Fig. 77) est affichée à l'écran :

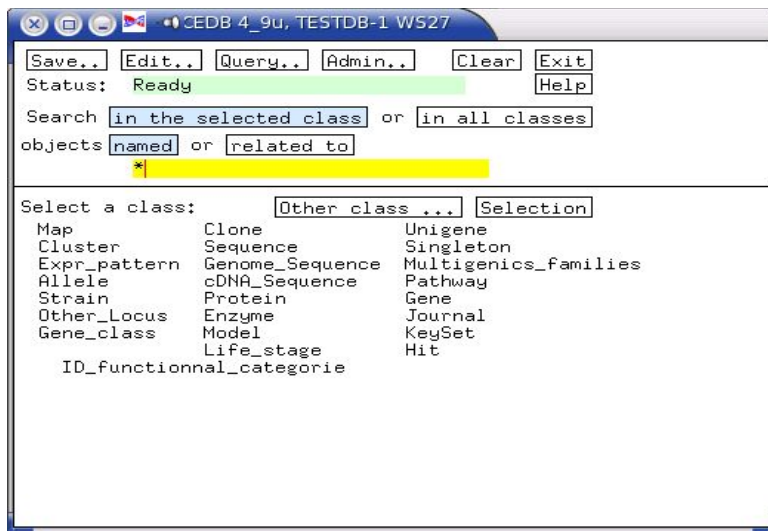


Figure 77 : Fenêtre de démarrage de AceDB

- la barre jaune sert à rentrer les requêtes (nom d'un unigène, une fonction...)
- les classes représentent les différentes classes associées aux requêtes, toutes ne servent pas, elles font partie de la construction de base de AceDB. Les classes suivantes sont utilisées :

Clone. Requête : nom d'EST contigué. Permet de voir une représentation très schématique de l'assemblage effectué par CAP3.

Unigene. Requête : nom d'un unigène. Permet de voir le détail des informations sur un unigène.

Cluster. Requête : nom d'un contig unigène. Permet de voir le détail des informations sur le contig.

Sequence. Requête : nom d'EST ou unigène. Affiche la séquence.

Singleton. Requête : nom d'un singlet unigène. Permet de voir le détail des informations sur le singlet.

Multigenic families. Requête : numéro de famille multigénique INN ou INP. Affiche les membres de la famille multigénique

Hit. Requête : fonction biologique ou nom de protéine. Affiche les noms des protéines des bases publiques homologues à un ou plusieurs unigènes.

ID_functionnal_categories. Requête : numéro de catégorie fonctionnelle suivant le catalogue du MIPS. Affiche les unigènes homologues à des protéines d'*Arabidopsis* appartenant à la catégorie recherchée.

Si on clique directement sur une classe sans rentrer de requête, on obtient l'ensemble des éléments de la classe.

Dans la classe Unigène, on retrouve les éléments présentés dans la figure 78.

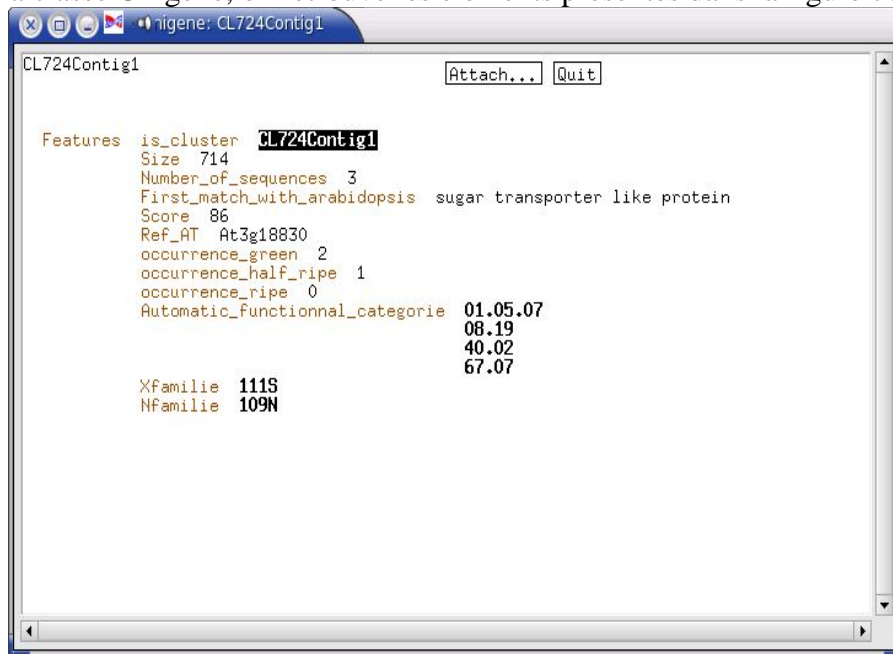


Figure 78. Fenêtre de la classe Unigène

- Is_cluster : permet d'aller à la classe cluster, dans le cas d'unigènes singlets, on a is_singleton.

- Size : la taille de l'unigène.

- Number_of_sequences : nombre de séquences dans l'unigène.

- First_match_with_arabidopsis : fonction du plus fort homologue dans *Arabidopsis*, si first_match est indiqué, cela signifie que la fonction de l'unigène a été déterminée par homologie à une autre espèce.

- Score : score de l'alignement avec le meilleur homologue.

- Ref_at : Numéro de référence MIPS du meilleur homologue chez *Arabidopsis*.

- Occurrence : nombre de séquences 3' retrouvées pour chaque banque dans l'unigène.

- Automatic_functional_categorie : liste des catégories fonctionnelles auxquelles appartient le meilleur homologue chez *Arabidopsis*, les catégories fonctionnelles attribuées par recherche bibliographique sont présentées, le cas échéant.

- Xfamille : numéro de la famille multigénique INP.

- Nfamille : numéro de la famille multigénique INN.

- la présence de chimères détectées est indiquée pour les unigènes concernés.

Chaque élément correspondant à une classe renvoie à la classe concernée.

Les dix meilleurs homologues dans les bases de données publiques sont accessibles via la classe Cluster.