

# Identification des déterminants moléculaires de la qualité du café par des approches de génomique fonctionnelle. Une revue

Thierry Joët<sup>1</sup>  
David Pot<sup>2</sup>  
Lucia Pires Ferreira<sup>3,4</sup>  
Stéphane Dussert<sup>1</sup>  
Pierre Marraccini<sup>2,5</sup>

<sup>1</sup> IRD  
UMR Diade  
911 Av. Agropolis  
BP64501  
34394 Montpellier  
France  
<thierry.joet@ird.fr>  
<stephane.dussert@ird.fr>

<sup>2</sup> Cirad  
UMR AGAP  
34398 Montpellier  
France  
<david.pot@cirad.fr>  
<marraccini@cirad.fr>

<sup>3</sup> IAPAR  
LBI-AMG  
CP 481  
86001-970 Londrina – PR  
Brésil  
<luciapiresferreira@gmail.com>

<sup>4</sup> Institut agronomique de Campinas (IAC)  
Centro de Recursos Genéticos Vegetais  
Av Theodureto de Almeida Camargo, n°  
1500  
Vila Nova  
13075-630 Campinas – SP  
Brésil

<sup>5</sup> EMBRAPA  
Genetic Resources and Biotechnology  
CP 02372  
70770-900 Brasilia – DF  
Brésil

## Résumé

La saveur du café implique le développement d'arômes lors de la torréfaction à partir de précurseurs biochimiques présents dans la graine de caféier. La caractérisation des voies de biosynthèse des principaux composés de réserve de la graine, ainsi que l'identification des gènes clés associés, permet donc d'identifier les déterminants moléculaires de la qualité. Cette revue décrit le développement récent des outils moléculaires et des connaissances génomiques chez le caféier. A travers la description des travaux de recherche ciblés sur la régulation du métabolisme des sucres, des acides chlorogéniques et de la caféine, les différentes approches de génomique fonctionnelle, et leurs complémentarités, sont illustrées.

**Mots clés :** caféier ; expression des gènes ; remplissage du grain ; voie biochimique.

**Thèmes :** amélioration génétique ; métabolisme ; productions végétales ; qualité et sécurité des produits.

## Abstract

**Using functional genomics approaches in identifying molecular determinants of coffee quality. A review**

The coffee seed contains all the necessary precursors to generate the full complement of coffee aromas and flavours. Coffee seed reserves are composed mainly of cell wall polysaccharides, lipids, proteins, sucrose and secondary metabolites including chlorogenic acids, caffeine and trigonelline. Because all of these compounds are involved in the complex chemistry of coffee roasting, understanding coffee quality ultimately requires a detailed characterization of the metabolic pathways dedicated to the synthesis of these aroma/flavour precursors which includes the transcriptional regulation for the genes encoding the relevant enzymes. This review focuses on the latest advances regarding functional genomics and its use in deciphering metabolic pathways that occur during coffee seed development.

**Key words:** biochemical pathways; *coffee*; gene expression; grain filling.

**Subjects:** genetic improvement; metabolism; product quality and security; vegetal productions.

## Composition de la graine et qualité du café

La graine des espèces cultivées de caféier (*Coffea arabica* et *C. cane-*

*phora*) est une semence à albumen (appelé endosperme dans les revues scientifiques de langue anglaise) cellulaire dont le développement requiert 8 à 10 mois (*figure 1*). Après la fécondation et jusqu'à mi-développement, la graine est exclusivement constituée

Pour citer cet article : Joët T, Pot D, Ferreira LP, Dussert S, Marraccini P, 2012. Identification des déterminants moléculaires de la qualité du café par des approches de génomique fonctionnelle. Une revue. *Cah Agric* 21 : 125-33. doi : 10.1684/agr.2012.0548

## Le point sur les approches de génomique fonctionnelle chez le caféier

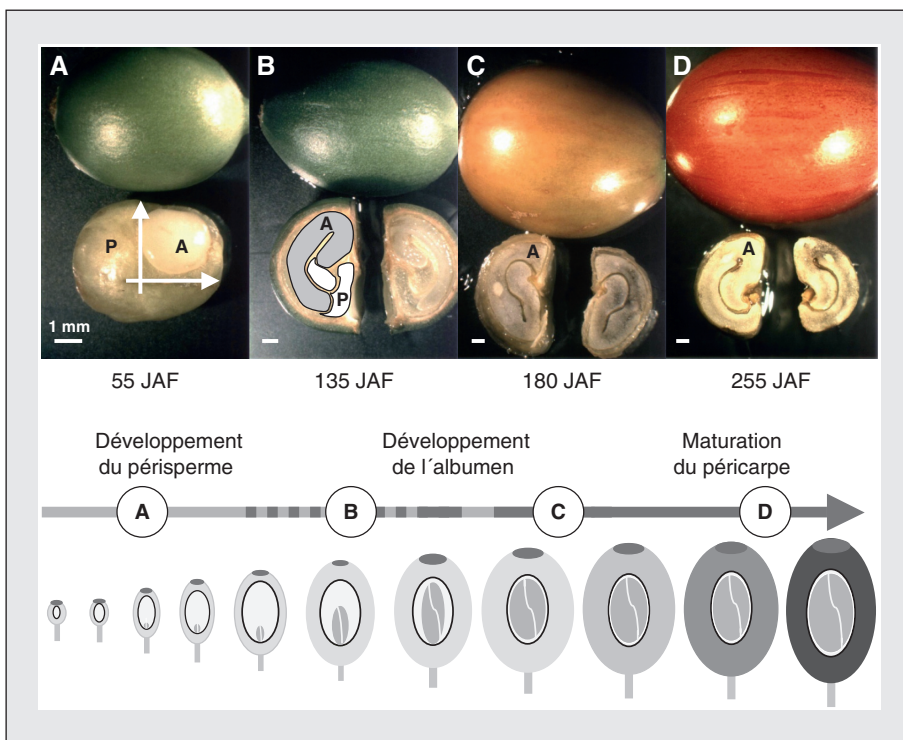


Figure 1. Phénologie de la graine de caféier.

### Figure 1. Phenology of the coffee seed.

Les fruits de *C. arabica* ont été récoltés à différents stades de développement après la floraison (JAF : jours après floraison) et présentés en coupe dorsale (A : points de coupe indiqués par les flèches) et en coupe transversale (B-D) pour montrer la croissance de l'albumen (A) à l'intérieur du péricarpe (P). Les positions de l'albumen et du péricarpe sont schématisées respectivement en gris et en blanc (135 JAF). Les stades de développement des grains illustrés dans la partie supérieure (A-D) sont reportés sur le schéma de maturation du fruit, adapté d'après De Castro et Marraccini (2006).

d'un tissu nourricier maternel diploïde, le péricarpe, qui est ensuite remplacé progressivement par l'albumen triploïde, un tissu de réserves issu de la fusion de la cellule centrale du sac embryonnaire avec un des deux gamètes mâles délivré par le tube pollinique. D'abord présent à l'état liquide, l'albumen va durcir et accumuler des réserves nécessaires au développement de l'embryon lors de la germination. L'albumen constitue ainsi la majeure partie de la graine mature (appelée « grain de café vert ») et ses réserves sont principalement constituées de polysaccharides pariétaux, de lipides, de protéines et de saccharose, ainsi que de métabolites secondaires comme les acides chlorogéniques, la caféine, et la trigonelline (Clifford, 1985). Chacun de ces composés joue un rôle essentiel dans les réactions chimiques caractéristiques de la torréfaction qui libèrent plus de 800

molécules chimiques (Flament, 2002), dont certaines, volatiles, participent à des degrés divers au bouquet d'arômes du café (Grosch, 2001), et d'autres non volatiles interviennent dans sa saveur (Homma, 2001). Comme la participation de chacun de ces composés aromatiques sur la qualité organoleptique du café torréfié est difficile à évaluer, une approche consiste à caractériser leurs précurseurs dans le grain de café vert, qui sont considérés comme des précurseurs de la qualité (Leroy *et al.*, 2006). Ainsi, l'étude des voies de biosynthèse de ces précurseurs permet de mieux comprendre les déterminants de la qualité du café. En cela, les approches de génomique fonctionnelle ont permis de réaliser des avancées significatives dans l'étude des voies de biosynthèse et de leur régulation au cours du développement de la graine de caféier.

Une des premières étapes nécessaires pour la mise en œuvre des approches de génomique fonctionnelle consiste à établir un catalogue des gènes exprimés chez l'espèce étudiée, en couvrant le plus grand nombre d'organes, de tissus et de conditions de culture possible. À cette fin, les populations de transcrits (ARN messagers) sont isolées à partir des différents matériels collectés, rétrotranscrits en ADN complémentaires, puis séquencés massivement à partir de leur extrémité 5' pour obtenir une étiquette de séquence unique pour chaque gène. Depuis 2005, plusieurs bibliothèques de séquences exprimées (EST, pour « Expressed Sequence Tag ») ont été générées pour les espèces du genre *Coffea*, notamment au sein du projet Cornell-Nestlé (Lin *et al.*, 2005) et par l'Institut de recherche pour le développement (IRD, [Poncet *et al.*, 2006]) (tableau 1). La compilation de ces séquences (environ 58 000 EST) a abouti à la génération de 15 721 unigènes. Ces séquences ont été utilisées pour la construction de puces à ADN qui ont déjà servi à effectuer une comparaison préliminaire entre les graines, les feuilles et les fleurs de caféiers : elles ont permis la mise en évidence de 1 565 gènes spécifiquement exprimés dans la graine mature par rapport aux deux autres tissus (Privat *et al.*, 2011). Récemment, les données EST du projet brésilien « Genoma café » ont été rendues publiques (Mondego *et al.*, 2011). Ainsi, en quelques années, la construction de bibliothèques d'EST a non seulement permis l'établissement d'une liste des gènes exprimés dans la graine, mais aussi le développement d'outils pour le suivi de son transcriptome. Cette phase de découverte et d'annotation des gènes bénéficiera également du séquençage du génome de *C. canephora*, dont la publication est attendue d'ici fin 2012.

L'étape suivante consiste à mettre en évidence les répertoires de gènes

**Tableau 1. Ressources EST disponibles pour différentes espèces du genre *Coffea*.**

Table 1. *Coffea* EST resources available.

	EST	Bibliothèques	Références
<i>C. canephora</i>	537	1	Raveendra <i>et al.</i> (2010) <sup>a</sup>
	46 914	1, 2	Lin <i>et al.</i> (2005)
	8 778	1, 2	Poncet <i>et al.</i> (2006)
	12 805	1, 4	Vieira <i>et al.</i> (2006), Vidal <i>et al.</i> (2010), Mondego <i>et al.</i> (2011)
<b>Total</b>	<b>69 034</b>		
<i>C. arabica</i>	618	1	Fernandez <i>et al.</i> (2004)
	292	1	Joët <i>et al.</i> (2007) <sup>a</sup> , Lecolier <i>et al.</i> (2008) <sup>a</sup>
	41 985	1, 2, 3	Moncada <i>et al.</i> (2009) <sup>a</sup>
	448	1	Cristancho et McCouch (2002) <sup>a</sup>
	139	1, 4	De Nardi <i>et al.</i> (2006)
	130 792	1-5, 6	Vieira <i>et al.</i> (2006), Vidal <i>et al.</i> (2010), Mondego <i>et al.</i> (2011)
<b>Total</b>	<b>174 274</b>		
Autres <i>Coffea</i>	± 12 500	6	Lashermes <i>et al.</i> (2008)
<b>Total</b>	<b>255 808</b>		

Le nombre de séquences EST a été obtenu par consultation de la banque dbEST (GenBank Release 178.0, 18/06/2010, <http://www.ncbi.nlm.nih.gov/nucest>). Les bibliothèques d'ADNc utilisées (1, feuille ; 2, fruit ; 3, fleur ; 4, racine ; 5, cal embryogène ; 6, autres) et les séquences publiques non publiées (<sup>a</sup>) sont indiquées. D'après Lashermes *et al.* (2008) et de Kochko *et al.* (2010).

directement impliqués dans la synthèse des arômes et de leurs précurseurs. À cette fin, les travaux réalisés chez les plantes modèles sont essentiels pour deux aspects. D'une part, ils ont permis d'établir que plusieurs voies de biosynthèse peuvent parfois coexister pour un même produit. D'autre part, la plupart des gènes impliqués dans les voies de biosynthèse des principaux composés de réserves ont été identifiés. Ces connaissances facilitent tout d'abord l'identification rapide des gènes correspondants (orthologues) chez le caféier où leur recherche dans les bibliothèques d'EST s'effectue par homologie de séquence. Ensuite, le suivi du niveau d'expression des gènes candidats au cours du développement de la graine permet d'inférer l'importance et les modes de régulation dans les différents tissus de la graine de chacune des voies de biosynthèse connues. La RT-PCR (*Real-time polymerase chain reaction*) en temps réel reste une méthode de référence pour réaliser cette étape. La mise en relation des profils d'accumulation des composés biochimiques et des profils d'expression des gènes (ou des activités enzymatiques correspondantes)

permet enfin une reconstruction plus précise des voies métaboliques actives. Le couplage des données transcriptomiques et biochimiques a ainsi permis la reconstruction de plusieurs voies métaboliques d'intérêt dans la graine de caféier, notamment la description des gènes impliqués dans la biosynthèse des acides chlorogéniques, des acides gras, des triglycérides, et des sucres, dont l'exemple sera développé dans cette revue (Geromel *et al.*, 2006 ; Lepelley *et al.*, 2007 ; Privat *et al.*, 2008 ; Joët *et al.*, 2009).

Le lien entre l'activité transcriptionnelle et l'accumulation du métabolite d'intérêt n'est cependant pas toujours facile à établir car d'autres niveaux de régulation peuvent intervenir. Une fois l'ensemble des gènes d'une même voie décrit, il s'agit donc de hiérarchiser leur importance relative et de déterminer si une régulation transcriptionnelle influe sur l'accumulation des métabolites. Pour cela, une nouvelle approche, appelée biologie des systèmes, vise à intégrer différents niveaux d'informations, par exemple la quantification des changements dans le transcriptome et le métabolome en réponse à une perturba-

tion donnée. Cette approche permet notamment d'identifier les gènes dont le niveau d'expression est corrélé quantitativement au métabolite d'intérêt, et ainsi d'identifier les étapes clés du contrôle transcriptionnel. Récemment, une approche de ce type a été menée pour l'étude de la voie des phénylpropanoïdes et la synthèse des acides chlorogéniques du caféier (Joët *et al.*, 2010a). Cette voie de biosynthèse sera donc utilisée dans cette revue pour illustrer les approches de biologie des systèmes.

Une fois les gènes clés identifiés, leur rôle dans la synthèse des principaux métabolites d'intérêt reste à vérifier *in vivo*. La fonction des gènes peut être validée par transformation génétique. Cette étape de validation a pu être réalisée pour la voie de biosynthèse de la caféine, relativement simple, grâce à l'existence de mutants naturels et l'emploi de la transgénèse. Ce métabolite secondaire est donc utilisé dans cette revue pour illustrer la démarche de validation fonctionnelle.

## Exemple d'utilisation du couplage de données transcrits-métabolites : le métabolisme du saccharose

Parmi les composés de réserve majoritaires, les polysaccharides pariétaux et le saccharose jouent un rôle de premier plan dans la genèse de la flaveur et des arômes du café. En effet, lors de la torréfaction, leur dégradation thermique génère des sucres réducteurs qui, en interaction avec les acides aminés, initient les réactions de Maillard. De plus, la proportion de saccharose non dégradé lors de la torréfaction contribue à la douceur du breuvage.

Chez le caféier, comme pour la plupart des plantes supérieures, le saccharose représente la forme majoritaire de transport des photoassimilats vers les organe-puits tels que les graines (Franck *et al.*, 2006). Il joue un rôle pivot dans le métabolisme carboné des graines, les hexoses résultant de son clivage étant utilisés à la fois

comme source d'énergie pour le métabolisme cellulaire et de carbone pour la biosynthèse de nombreux composés de réserve, notamment les polysaccharides pariétaux.

Dans un premier temps, les études biochimiques réalisées sur le péricarpe et l'albumen ont permis de montrer que ces deux tissus présentent des métabolismes glucidiques particulièrement contrastés (Rogers *et al.*, 1999). En effet, le glucose et le fructose représentent la majeure partie des sucres solubles du péricarpe durant les premiers stades de développement de la graine, alors que le saccharose est le sucre majoritaire de l'albumen en croissance et dans la graine mature (*figure 2*). Comme chez d'autres graines modèles, il semblerait que le rapport entre hexoses (H) et saccharose (S) exerce un contrôle métabolique sur l'activité de division cellulaire (ratio H/S élevé du péricarpe) et sur la biosynthèse des réserves (baisse progressive du ratio H/S au cours du développement de l'albumen) (Weber *et al.*, 1998).

Récemment, le couplage des approches biochimiques et génomiques a permis d'identifier certains gènes (correspondant à des isoformes enzymatiques particulières) directement impliqués dans le métabolisme du saccharose (*figure 2*). Deux types d'enzymes, les invertases et la saccharose synthase (SuSy), sont capables de cliver la liaison glycosidique du saccharose. Au sein du péricarpe, ce sont les invertases qui jouent un rôle primordial, notamment les invertases acides dont l'activité est spécifique des premiers stades de développement de la graine (Geromel *et al.*, 2006). Le clivage du saccharose semble être principalement effectué dans l'apoplasme, les gènes codant pour des isoformes pariétales de l'invertase étant fortement exprimés durant cette phase (Privat *et al.*, 2008). Contrairement à ce qui se passe dans le péricarpe, le saccharose est véhiculé à l'intérieur des cellules de l'albumen avant son clivage. En effet, plusieurs gènes codant pour des transporteurs de la membrane plasmique nécessaires au chargement intracellulaire du saccharose présentent un niveau d'expression élevé lors de la phase de croissance de l'albumen (Joët *et al.*, 2009). Ensuite, l'activité saccharolytique est prise en charge par la saccharose synthase. Le gène *SUS1*,

codant pour une de ses isoformes (Leroy *et al.*, 2005) présente un profil d'expression concordant avec la phase de biosynthèse des réserves (De Castro et Marraccini, 2006). Parallèlement à l'import de saccharose et à son utilisation, une néosynthèse de saccharose a également été mise en évidence en utilisant du fructose marqué au <sup>14</sup>C (Geromel *et al.*, 2006). Ainsi, l'activité enzymatique de la saccharose phosphate synthase (SPS), impliquée dans la synthèse du saccharose, est maximale lors de cette même phase. D'autre part, SuSy peut être active dans le sens du clivage ou celui de la biosynthèse. Ainsi, l'expression tardive du gène *SUS2*, codant pour une seconde isoforme de SuSy, semble corrélée avec la synthèse et l'accumulation de saccharose lors de la maturation de la graine (Salmona *et al.*, 2008).

L'utilisation de la RT-PCR en temps réel, couplée à celle des méthodes de chimie analytique classique, a ainsi permis la description des voies de biosynthèse des principaux composés de réserve de la graine de caféier. La disponibilité récente de puces à ADN (Privat *et al.*, 2011), ainsi que l'émergence de nouveaux outils d'analyse des transcrits par séquençage haut débit (RNA-Seq) devraient permettre d'étendre rapidement ces approches à l'ensemble du transcriptome. De plus, les techniques de métabolomique, telles que la chromatographie couplée à la spectrométrie de masse (GC-MS et LC-MS) ou la détection par résonance magnétique nucléaire (<sup>1</sup>H-RMN), permettent de couvrir un large éventail de molécules. Ces techniques d'analyse du transcriptome et du métabolome seront d'une aide précieuse pour décrire et décortiquer en parallèle de nombreuses voies métaboliques, mais restent à valider chez le caféier.

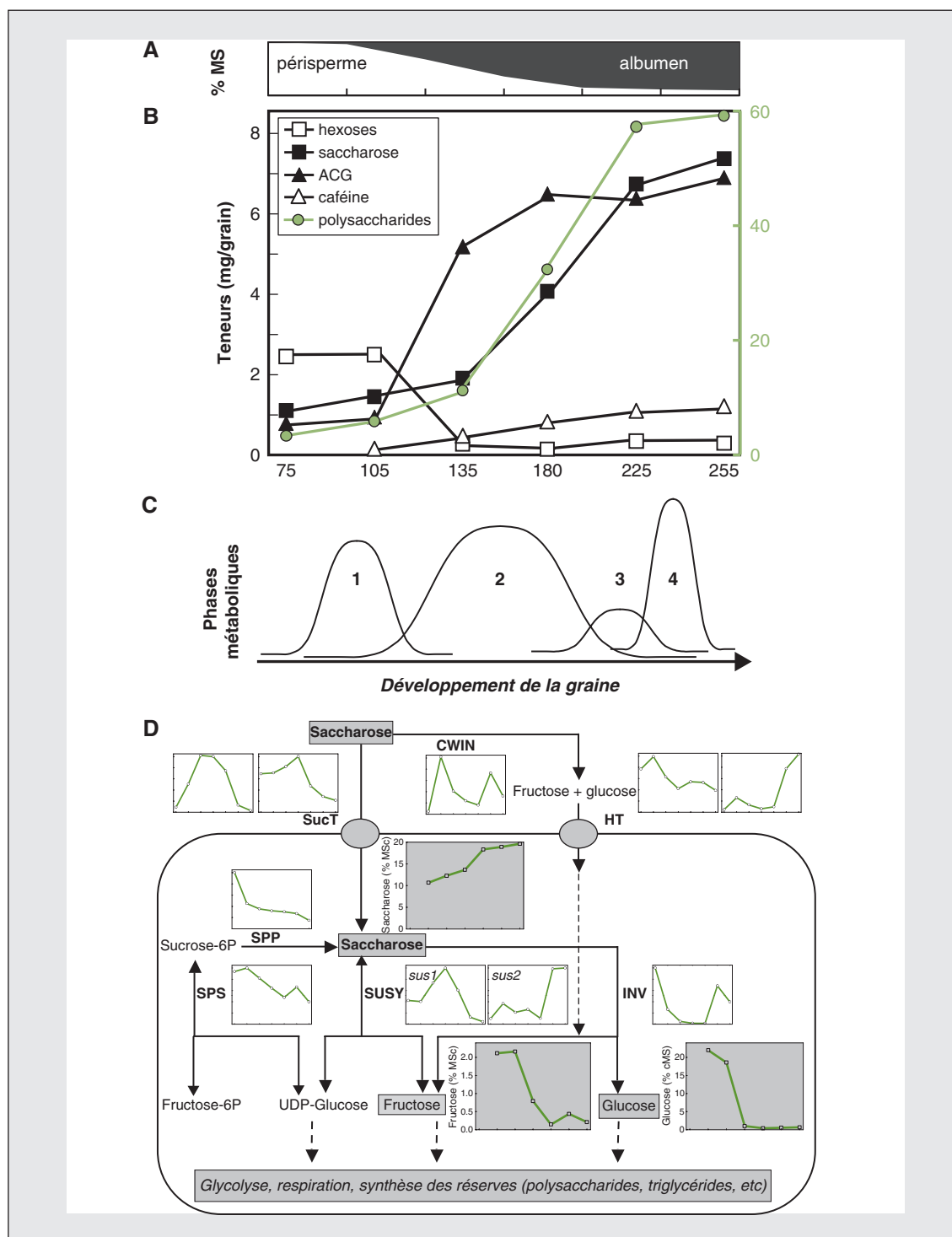
## Exemple d'approche de biologie des systèmes : la synthèse des acides chlorogéniques

La graine de caféier présente de fortes teneurs en acides chlorogéniques

(ACG), une famille d'esters formés à partir de certains dérivés trans de l'acide cinnamique, en particulier l'acide caféique, et de l'acide quinique (Hanson, 1965). Comme la caféine, les ACG contribuent à l'astringence et l'amertume du breuvage. Les ACG sont synthétisés de manière précoce dans le péricarpe et s'accumulent lors du développement rapide de l'albumen (De Castro et Marraccini, 2006). En l'absence de connaissances sur la nature moléculaire des enzymes impliquées dans les processus de transestérification entre les différents isomères d'ACG, les études de génomique fonctionnelle se sont focalisées sur la synthèse de l'ACG majoritaire, le 5-cafféoylquinique (5-CQA), qui est le précurseur de tous les autres ACG (*figure 3*). Ce composé est également un intermédiaire dans la synthèse des lignines (Aerts et Baumann, 1994) et sa synthèse emprunte les premières étapes de la voie des phénylpropanoïdes. Dans un premier temps, le couplage de données transcriptomiques et biochimiques ainsi que la caractérisation fonctionnelle de certaines protéines recombinantes ont permis l'identification de gènes et isoformes enzymatiques directement impliqués dans le métabolisme du 5-CQA (Lepelley *et al.*, 2007 ; Joët *et al.*, 2009).

Il a ensuite été montré que la teneur et la composition en ACG dans la graine mature sont sous la dépendance forte de facteurs environnementaux, notamment la température moyenne de l'air (Joët *et al.*, 2010b). La mise en évidence récente de ces effets, couplée à une description précise de la voie de biosynthèse du 5-CQA, a permis de mettre en œuvre des approches de biologie des systèmes et l'identification de plusieurs points de contrôle (Joët *et al.*, 2010a). En effet, les conditions environnementales peuvent être utilisées comme source de variation pour analyser les relations quantitatives entre transcrits et métabolites au cours du développement de la graine et identifier les gènes dont le niveau d'expression est relié à un trait quantitatif (QTT, pour « *Quantitative Trait Transcript* »). Les deux premières étapes de la voie des phénylpropanoïdes, pilotées par la phénylalanine ammonia-lyase (PAL) et la cinnamate 4-hydroxylase (C4H) sont souvent décrites chez les plantes modèles comme pivots de la régulation de cette





**Figure 2.** Synthèse des composés de réserve au cours du développement de la graine et métabolisme des sucres.

**Figure 2.** Synthesis of storage compounds and sugar metabolism in developing coffee seed.

A) les différentes phases tissulaires du développement du grain sont schématisées en proportion du poids sec de la graine ; B) les évolutions des teneurs moyennes en hexoses, saccharose, acides chlorogéniques (ACG), caféine et polysaccharides pariétaux sont exprimées en mg par grain au cours du développement (JAF : jours après la floraison). Les coefficients de variation, non représentés, sont inférieurs à 10 % ; C) représentation des principales phases métaboliques : 1) biosynthèse des ACG ; 2) des polysaccharides pariétaux (galactomannanes, cellulose) et des protéines de réserve ; 3) remobilisation des ACG ; et 4) synthèse et accumulation du saccharose ; D) représentation des profils d'expression des gènes liés au métabolisme des sucres.

CWINV : invertase pariétale ; HT : transporteur d'hexose ; INV : invertase acide ; MS : matière sèche ; SucT : transporteur de saccharose ; SPS : saccharose phosphate synthase ; SPP : saccharose phosphate phosphatase ; SUSY : saccharose synthase ; SUS1 : isoforme 1 de la saccharose synthase ; SUS2 : isoforme 2 de la saccharose synthase. Adapté d'après Joët *et al.* (2009).

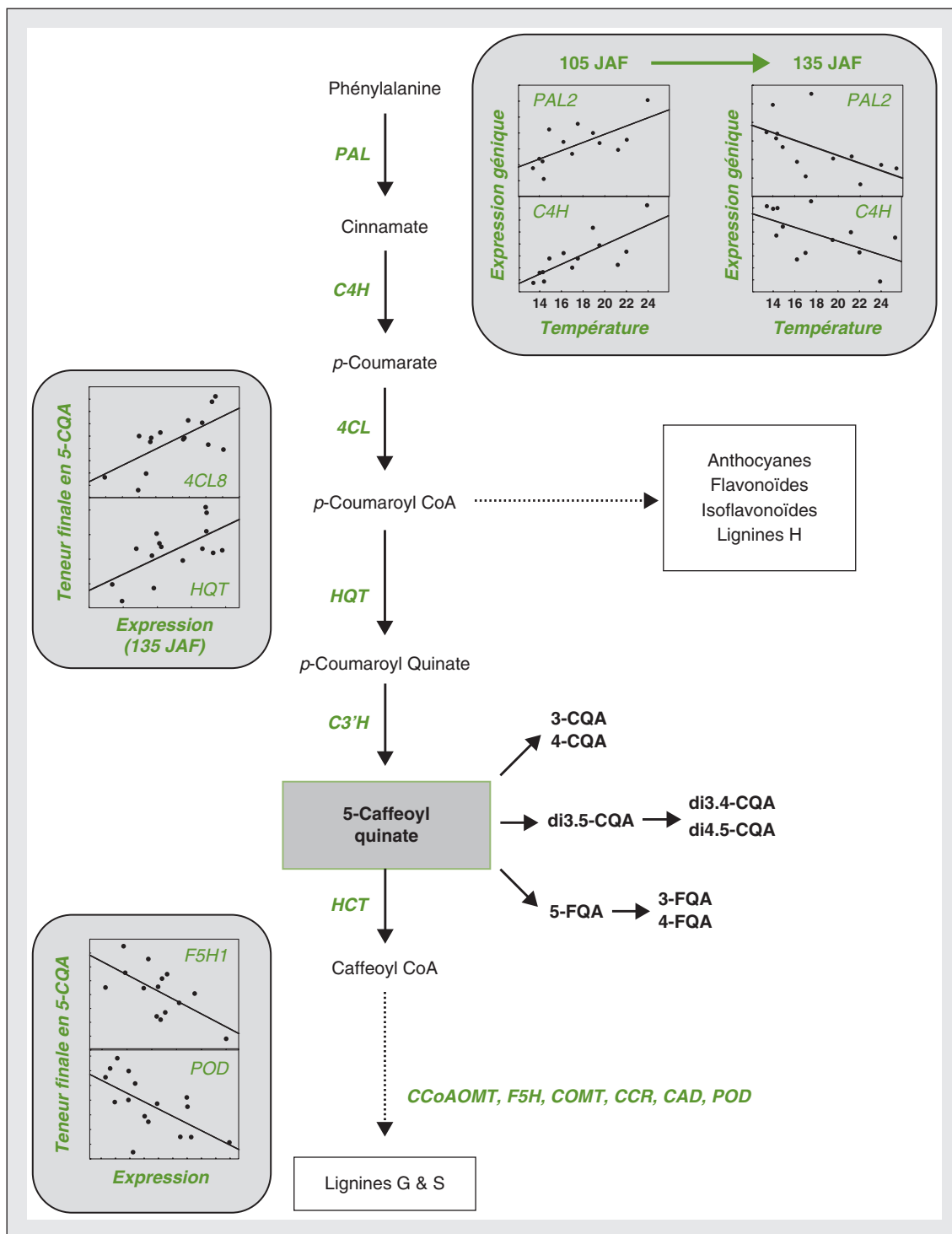


Figure 3. Voie des phénylpropanoïdes et points de contrôle pour la synthèse des acides chlorogéniques.

Figure 3. The phenylpropanoid pathway and its major control points for the biosynthesis of chlorogenic acid.

Les gènes dont les niveaux d'expression sont corrélés à la température moyenne (contrôle transcriptionnel lié à un trait quantitatif, QTT) figurent dans les encadrés. C3'H, p-coumaroyl CoA 3-hydroxylase ; C4H, trans-cinnamate 4-hydroxylase ; CAD, cinnamyl alcohol dehydrogenase ; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase ; 4CL, 4-coumarate : CoA ligase ; CCR, cinnamoyl-CoA reductase ; COMT, caffeate O-methyltransferase ; F5H, ferulate 5-hydroxylase ; HCT, hydroxycinnamoyl-CoA : shikimate/quinatate hydroxycinnamoyl transferase ; HQT, hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase ; PAL, phenylalanine ammonia lyase ; POD, secretory peroxidase. Adapté d'après Joët *et al.* (2010a).

voie. Chez le caféier, un fort contrôle transcriptionnel s'exerce sur ces étapes : d'une part l'expression des gènes codant ces enzymes est pilotée par le programme génétique de développement de la graine, mais elle est également modulée par des variables climatiques telles que la température moyenne de l'air (*figure 3*). Les étapes suivantes de la voie des phénylpropanoïdes menant au 5-CQA sont réalisées par la coumarate ligase (4CL), l'hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transférase (HQT/HCT) et la coumaroyl-CoA 3-hydroxylase (C3'H). Plusieurs gènes codant pour ces enzymes sont exprimés au cours du développement de la graine (Lepelley *et al.*, 2007 ; Joët *et al.*, 2009). Mais parmi eux seuls les gènes *4CL8* et *HQT1* montrent un lien quantitatif entre leur niveau d'expression et l'accumulation des ACGs (*figure 3*). Cette approche, basée sur les variations induites par l'environnement, a également permis de montrer que l'accumulation des ACG est aussi contrôlée en aval par la répression des gènes impliqués dans leur remobilisation vers les monolignols, en identifiant les gènes codant pour la ferulate 5'hydroxylase (*F5HI*) et peroxydase (*POD*) comme QTT. Cette approche de biologie des systèmes pourrait être appliquée à d'autres voies métaboliques. En effet, des études récentes ont montré que la température (ombrage et altitude) a également une influence sur les teneurs en hexoses des graines matures (Geromel *et al.*, 2008 ; Joët *et al.*, 2010b), suggérant qu'il serait possible d'utiliser la variance induite par l'environnement pour caractériser les points de contrôle du métabolisme des sucres. Cependant, une autre source de variation semble également prometteuse, celle induite par la diversité génétique. En effet, des approches de biologie des systèmes sur différentes variétés de maïs présentant des métabolismes contrastés ont récemment permis l'identification de QTT pour l'accumulation des caroténoïdes (Vallabhaneni *et al.*, 2009). Ces approches pourraient être utilisées chez les caféiers pour le métabolisme des sucres où, à une échelle restreinte, la comparaison des niveaux de transcrits et d'activités enzymatiques pour les invertases et saccharose synthases entre les deux espèces de caféiers cultivés suggère déjà que les faibles

teneurs en saccharose des graines de *C. canephora* sont corrélées à une plus forte activité saccharolytique par rapport à celle mesurée chez *C. arabica*, avec des niveaux de contrôle à la fois transcriptionnel et posttranscriptionnel (Privat *et al.*, 2008)

## Validation fonctionnelle des gènes : exemple de la biosynthèse de la caféine

La caféine est un alcaloïde purique synthétisé à partir de la xanthosine, et ceci à travers seulement trois étapes successives de méthylation et une étape d'élimination du ribose. Les gènes *CaXMT1*, *CaMXMT2* et *CaDXMT1*, codant pour différentes N-méthyltransférases ont été identifiés (Ogawa *et al.*, 2001 ; Uefuji *et al.*, 2003). La caractérisation fonctionnelle des protéines recombinantes a montré que les trois enzymes utilisent la S-adenosyl-L-méthionine (SAM) comme donneur de groupement méthyl et a permis d'établir un rôle pour ces gènes respectivement dans la synthèse de 7-méthylxanthosine, de théobromine (3,7-diméthylxanthine), et de caféine (1,3,7-triméthylxanthine) (Uefuji *et al.*, 2003). Ces gènes sont exprimés de manière précoce au cours du développement de la graine (Salmona *et al.*, 2008). Leur fonction peut être validée par transformation génétique, dont le protocole a été récemment optimisé pour améliorer l'efficacité (Ribas *et al.*, 2011). Ainsi, l'inactivation par transgénése du gène *CaMXMT1* codant la théobromine synthase aboutit à une réduction drastique de la teneur en théobromine et caféine dans les feuilles (Ogita *et al.*, 2003 ; Ashihara *et al.*, 2006). De même, un crible biochimique de différentes lignées d'Arabica sauvages d'Éthiopie a permis la découverte d'un mutant naturel sans caféine (Silvarolla *et al.*, 2004). Chez ce mutant l'activité caféine synthase n'est pas détectable, et des études récentes montrent la présence de mutations ponctuelles dans le gène *CADXMT1*, ainsi que

des perturbations dans la régulation de l'activité transcriptionnelle de ce gène (Maluf *et al.*, 2009). En définitive, c'est certainement pour ce métabolite secondaire que les travaux sont les plus avancés, avec notamment la structure tridimensionnelle de la xanthosine méthyltransférase (XMT) et de la caféine synthase (DXMT) récemment décrites par cristallographie (McCarthy et McCarthy, 2007).

## Perspectives

Les prochaines années devraient voir la généralisation de l'utilisation chez le caféier des outils d'analyse du transcriptome et du métabolome, ainsi que de leur couplage. Cette étape est essentielle pour décrypter les voies de biosynthèse du large répertoire de molécules présentes dans le grain de café vert. À ce titre, il est utile de rappeler que certains métabolites peuvent être présents en quantité marginale et cependant influencer de manière prépondérante les qualités organoleptiques du café. Ainsi, une part des composés volatiles sont présents avant torréfaction et sont synthétisés au cours du développement de la graine. Principalement issus des voies des phénylpropanoïdes et des isoprénoïdes, leur seuil de perception peut être de l'ordre du nanomolaire. Le couplage des approches de transcriptomique avec celles de chimie analytique des composés volatiles par GC-MS, par exemple, devrait permettre l'étude de leurs voies de biosynthèse.

La généralisation de ces approches devrait aboutir au dépôt d'un grand nombre de données dans des bases d'accès public, offrant ainsi des perspectives intéressantes en termes de méta-analyse. En effet, la généralisation des puces à ADN permet d'étudier, sur un très grand nombre d'échantillons et dans différentes conditions, les relations de covariance qui existent entre les différents gènes exprimés. L'analyse par régression linéaire des relations entre tous les gènes pris deux à deux dans différentes conditions expérimentales facilite la mise en évidence de régulateurs, groupes de gènes coexprimés. Ces unités transcriptionnelles, pouvant contenir jusqu'à plusieurs centaines de gènes, sont

souvent coréglées par un même facteur de transcription. L'identification de ces facteurs est une étape importante dans la compréhension des modes de régulation des voies de biosynthèse.

À terme, l'identification de gènes clés liés à la synthèse des composés de réserves fournira des marqueurs diagnostiques pour le suivi de la qualité, l'accumulation des composés liés à la qualité aromatique du café et à la typicité de terroir. L'exploration de la diversité nucléotidique de ces gènes et de leur régulation au sein des populations servant de base aux programmes de sélection, constituera le fondement de la mise en place de la sélection assistée par marqueurs pour la qualité du café. Les bénéfices de la mise en place d'une telle sélection sont évidents compte tenu de la complexité (coût, temps...) de l'évaluation phénotypique de ce caractère. Ces approches nécessitent une bonne connaissance des métabolites du café vert, de leur conversion lors de la torréfaction, et leurs interactions dans la genèse des propriétés sensorielles de la boisson. Les études futures devront donc intégrer la biologie de la graine, la chimie de la transformation, ainsi que des procédures standardisées d'analyse gustative. ■

## Références

- Aerts RJ, Baumann TW, 1994. Distribution and utilization of chlorogenic acid in *Coffea* seedlings. *Journal of Experimental Botany* 45 : 497-503. doi : 10.1093/jxb/45.4.497
- Ashihara H, Zheng XQ, Katahira R, Morimoto M, Ogita S, Sano H, 2006. Caffeine biosynthesis and adenine metabolism in transgenic *Coffea canephora* plants with reduced expression of N-methyltransferase genes. *Phytochemistry* 67 : 882-6. doi : 10.1016/j.phytochem.2006.02.016
- Clifford MN, 1985. Chemical and physical aspects of green coffee and coffee products. In : Clifford MN, Wilson KC, eds. *Coffee botany, biochemistry and production of beans and Beverage*. London (Grande-Bretagne) : Croom Helm.
- De Castro RD, Marraccini P, 2006. Cytology, biochemistry and molecular changes during coffee fruit development. *Brazilian Journal of Plant Physiology* 18 : 175-99. doi : 10.1590/S1677-04202006000100013
- De Kochko A, Akaffou S, Andrade AC, Campa C, Crouzillat D, Guyot R, et al., 2010. Advances in *Coffea* genomics. In : Kader JC, Delseny M, eds. *Advances in botanical research*. Oxford (Grande-Bretagne) : Academic Press.
- De Nardi B, Dreos R, Del Terra L, Martellosi C, Asquini E., Tornincasa P, et al., 2006. Differential responses of *Coffea arabica* L. leaves and roots to chemically induced systemic acquired resistance. *Genome* 49 : 1594-1605, doi : 10.1139/G06-125
- Fernandez D, Santos P, Agostini C, Bon MC, Petitot AS, Silva MC, et al., 2004. Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). *Molecular Plant Pathology* 5 : 527-36. doi : 10.1111/J.1364-3703.2004.00250.X
- Flament I, 2002. *Coffee flavor chemistry*. Chichester : John Wiley and sons Ltd.
- Franck N, Vaast P, Génard M, Dauzat J, 2006. Soluble sugars mediate sink feedback down-regulation of leaf photosynthesis in field-grown *Coffea arabica*. *Tree Physiology* 26 : 517-25. doi : 10.1093/treephys/26.4.517
- Geromel C, Ferreira LP, Guerreiro SMC, Cavalari AA, Pot D, Pereira LFP, et al., 2006. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *Journal of Experimental Botany* 57 : 3243-58. doi : 10.1093/jxb/erl084
- Geromel C, Ferreira LP, Davrieux F, Guyot B, Ribeyre F, dos Santos Scholz MB, et al., 2008. Effects of shade on the development and sugar metabolism of coffee fruits. *Plant Physiology and Biochemistry* 46 : 569-79. doi : 10.1016/j.plaphy.2008.02.006
- Grosch W, 2001. Volatile compounds. In : Clarke RJ, Vitzthum OG, eds. *Coffee : recent developments*. Oxford : Blackwell Science.
- Hanson KR, 1965. Chlorogenic acid biosynthesis. Chemical synthesis and properties of the mono-O-cinnamoylquinic acids. *Biochemistry-US* 4 : 2719-30. doi : 10.1021/bi00888a023
- Homma S, 2001. Non-volatile compounds, part II. In : Clarke RJ, Vitzthum OG, eds. *Coffee : recent developments*. Oxford : Blackwell Science.
- Joët T, Laffargue A, Salmona J, Doubeau S, Descroix F, Bertrand B, et al., 2009. Metabolic pathways in tropical dicotyledonous albuminous seeds : *Coffea arabica* as a case study. *New Phytologist* 182 : 146-62. doi : 10.1111/j.1469-8137.2008.02742.x
- Joët T, Salmona J, Laffargue A, Descroix F, Dussert S, 2010a. Use of the growing environment as a source of variation to identify the quantitative trait transcripts and modules of co-expressed genes that determine chlorogenic acid accumulation. *Plant Cell and Environment* 33 : 1220-33. doi : 10.1111/j.1365-3040.2010.02141.x
- Joët T, Laffargue A, Descroix F, Doubeau S, Bertrand B, de Kochko A, et al., 2010b. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. *Food Chemistry* 118 : 693-701. doi : 10.1016/j.foodchem.2009.05.048
- Lashermes P, Andrade CA, Etienne H, 2008. Genomics of coffee, one of the world's largest traded commodities. In : Moore PH, Ming R, eds. *Genomics of tropical crop plants*. New York (Etats-Unis) : Springer.
- Lepellet M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J, 2007. Chlorogenic acid synthesis in coffee : an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. *Plant Science* 172 : 978-96. doi : 10.1016/j.plantsci.2007.02.004
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, et al., 2005. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* 111 : 1032-41. doi 10.1007/s00122-005-0018-z
- Leroy T, Ribeyre F, Bertrand B, Charmetant P, Dufour M, Montagnon C, et al., 2006. Genetics of coffee quality. *Brazilian Journal of Plant Physiology* 18 : 229-42. doi : 10.1590/S1677-04202006000100016
- Lin C, Mueller LA, Mc Carthy J, Crouzillat D, Pétiard V, Tanksley S, 2005. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theoretical and Applied Genetics* 112 : 114-30. doi : 10.1007/s00122-005-r-0112-2
- Maluf MP, da Silva CC, de Oliveira MDA, Tavares AG, Silvarolla MB, Guerreiro O, 2009. Altered expression of the caffeine synthase gene in a naturally caffeine-free mutant of *Coffea arabica*. *Genetics and Molecular Biology* 32 : 802-10. doi : 10.1590/S1415-4752009005000090
- McCarthy AA, McCarthy JG, 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiology* 144 : 879-89. doi : 10.1104/pp.106.094854
- Mondego JMC, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, Costa GGL, et al., 2011. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biology* 11 : 30. doi : 10.1186/1471-2229-r11-30
- Ogawa M, Herai Y, Koizumi N, Kusano T, Sano H, 2001. 7-Methylxanthine methyltransferase of coffee plants. Gene isolation and enzymatic properties. *Journal of Biological Chemistry* 276 : 8213-8. doi : 10.1074/jbc.M009480200
- Ogita S, Uefuji H, Yamaguchi Y, Koizumi N, Sano H, 2003. RNA interference : Producing decaffeinated coffee plants. *Nature* 423 : 823. doi : 10.1038/423823a
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, et al., 2006. SSR mining in coffee tree EST databases : potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics* 276 : 436-49. doi : 10.1007/s00438-006-r-0153-5
- Privat I, Focquier S, Prins A, Epalle T, Eychenne M, Kandalaf L, et al., 2008. Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. *New Phytologist* 178 : 781-97. doi : 10.1111/j.1469-8137.2008.02425.x
- Privat I, Bardil A, Gomez AB, Severac D, Dantec C, Fuentes I, et al., 2011. The 'PUCE CAFE' Project : the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits. *BMC Genomics* 12 : 5. doi : 10.1186/1471-2164-r12-5
- Ribas AF, Dechamp E, Champion A, Bertrand B, Combes MC, Verdeil JL, et al., 2011. *Agrobacterium*-mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established embryogenic callus cultures. *BMC Plant Biology* 11 : 92. doi : 10.1186/1471-2229-r11-92
- Rogers WJ, Michaux S, Bastin M, Bucheli P, 1999. Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (*Coffea canephora*) and Arabica (*C. arabica*) coffees. *Plant Science* 149 : 115-23. doi : 10.1016/S0168-9452(99)00147-8



Salmona J, Dussert S, Descroix F, de Kochko A, Bertrand B, Joët T, 2008. Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Molecular Biology* 66 : 105-24. doi : 10.1007/s11103-007-r9256-6

Silvarolla MB, Mazzafera P ; Fazuoli LC, 2004. A naturally decaffeinated arabica coffee. *Nature* 429 : 826. doi : 10.1038/429826a

Uefuji H, Ogita S, Yamaguchi Y, Koizumi N, Sano H, 2003. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in

coffee plants. *Plant Physiology* 132 : 372-80. doi : 10.1104/pp.102.019679

Vallabhaneni R, Wurtzel ET, 2009. Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiology* 150 : 562-72. doi : 10.1104/pp.109.137042

Vidal RO, Mondego JM, Pot D, Ambrosio AB, Andrade AC, Pereira LF, *et al.*, 2010. A high-throughput data mining of SNPs in *Coffea* spp ESTs suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant*

*Physiology* 154 : 1053-66. doi : 10.1104/pp.110.162438

Vieira LGE, Andrade AC, Colombo CA, Moraes AHA, Metha A, Oliveira AC, *et al.*, 2006. Brazilian coffee genome project : an EST-based genomic resource. *Brazilian Journal of Plant Physiology* 18 : 95-108. doi : 10.1590/S1677-04202006000100008

Weber H, Heim U, Golombek S, Borisjuk L, Wobus U, 1998. Assimilate uptake and the regulation of seed development. *Seed Science Research* 8 : 331-45. doi : 10.1017/S096025850000426